ISBN: 978-81-920451-8-4

February 7-8, 2019

THE INSTITUTION OF ENGINEERS (INDIA)
PUNJAB & CHANDIGARH STATE CENTRE

UNDER THE AEGIS OF
COMPUTER ENGINEERING DIVISION, IEI

ALL INDIA SEMINAR
ON
"RECENT TRENDS IN IMAGE PROCESSING FOR AI IMPLEMENTATION"

IN ASSOCIATION / COLLABORATION WITH

BABA BANDA SINGH BAHADUR ENGINEERING COLLEGE

FATEHGARH SAHIB, PUNJAB

VENUE

CSE SEMINAR HALL
BABA BANDA SINGH BAHADUR ENGINEERING COLLEGE, FATEHGARH SAHIB, PUNJAB

The Institution of Engineers (India)

[Established 1920, Incorporated by Royal Charter 1935]
PUNJAB & CHANDIGARH STATE CENTRE
Madhya Marg, Sector 19-A, Chandigarh-160019

"99 Years of Relentless Journey Towards Engineering Advancement for Nation Building"

On behalf of Punjab & Chandigarh State Centre of the Institution of Engineers (India), I feel highly privileged to welcome you all on **Two Days All India Seminar on "Recent Trends in Image Processing for AI Implementation"** organized by the Institution of Engineers (India), Punjab & Chandigarh State Centre in association with Baba Banda Singh Bahadur Engineering College, Fatehgarh Sahib under the aegis of Computer Division Board of IEI during **February 7 - 8, 2019** at BBSBEC Campus, Fategarh Sahib.

The Seminar is aimed to provide a platform that brings together researchers, innovators and students alongwith industrial professionals focusing on aspects of Artificial intelligence (AI). Artificial Intelligence has become a very popular expression in recent years. The last century was rich in innovation, especially in mathematics, giving hope of reaching this goal, but many technical or conceptual gaps have buried the various initiatives. Today, however, thanks to tremendous progress over the past twenty years in terms of computational capacity, data accumulation, improvements in mathematical tools, all at reduced cost, we actually can see the birth of AI. I am confident that the deliberations of the Seminar will provide a unique forum to the participants to exchange their opinion on the topic.

My best wishes to all the organizers and participants for seeking this path towards gaining and spreading knowledge.

I eagerly look forward for the fruitful discussions during the various technical sessions.

I wish the Seminar a great success.

Janak Raj Garg FIE
Chairman
Punjab & Chandigarh State Centre
The Institution of Engineers (India)

The Institution of Engineers (India)

[Established 1920, Incorporated by Royal Charter 1935]
PUNJAB & CHANDIGARH STATE CENTRE
Madhya Marg, Sector 19-A, Chandigarh-160019

"99 Years of Relentless Journey Towards
Engineering Advancement for Nation Building"

I deem it my pleasure and privilege for being a part of the **Two Days All India Seminar on "Recent Trends in Image Processing for AI Implementation"** organized by the Institution of Engineers (India), Punjab & Chandigarh State Centre in association with Baba Banda Singh Bahadur Engineering College, Fatehgarh Sahib under the aegis of Computer Division Board of IEI during **February 7 - 8, 2019** at BBSBEC Campus, Fategarh Sahib.

The topics of the Seminar i.e. "Recent Trends in Image Processing for AI Implementation" defined as intelligence exhibited by machines, has many applications in today's society. More specifically, it is Weak AI, the form of AI where programs are developed to perform specific tasks, that is being utilized for a wide range of activities including medical diagnosis, electronic trading, robot control, and remote sensing. AI has been used to develop and advance numerous fields and industries, including finance, healthcare, education, transportation, and more.

Brainstorm of the learning community with the learned experts' community will surely be helpful in creating new ideas with best solutions in Research. I congratulate the Chairman, Honorary Secretary, dedicated team of the centre and convener for their best initial moves towards benefitting the young brains for a better India in future through this event and I wish you all Success.

Professor (Dr.) T.S. Kamal
FIE, FIETE, Life Sr. MIEEE (USA)

The Institution of Engineers (India)

[Established 1920, Incorporated by Royal Charter 1935]
PUNJAB & CHANDIGARH STATE CENTRE
Madhya Marg, Sector 19-A, Chandigarh-160019

"99 Years of Relentless Journey Towards
Engineering Advancement for Nation Building"

It is my pleasure and privilege for being a part of the **Two Days All India Seminar on "Recent Trends in Image Processing for AI Implementation"** organized by the Institution of Engineers (India), Punjab & Chandigarh State Centre in association with Baba Banda Singh Bahadur Engineering College, Fatehgarh Sahib under the aegis of Computer Division Board of IEI during **February 7 - 8, 2019** at BBSBEC Campus, Fategarh Sahib.

Areas of artificial intelligence deal with autonomous planning or deliberation for robotical systems to navigate through an environment. A detailed understanding of these environments is required to navigate through them. Information about the environment could be provided by a computer vision system, acting as a vision sensor and providing high-level information about the environment and the robot.

Artificial intelligence and computer vision share other topics such as pattern recognition and learning techniques. Consequently, computer vision is sometimes seen as a part of the artificial intelligence field or the computer science field in general.

I am hopeful that this Seminar will be a great success in our effort to aware and exchange the recent technologies and innovations in the field of "Image processing for AI Implementation".

Sukhvir Singh Mundi FIE
Honorary Secretary
Punjab & Chandigarh State Centre
The Institution of Engineers (India)

BABA BANDA SINGH BAHADUR ENGINEERING COLLEGE

Approved by AICTE, Govt. of Punjab, Affiliated to IKGPTU
(Courses Accredited by NBA (AICTE) and IE(I))

BBSBEC

Ref. No. BBSBEC/CSE/2019/001 Date 5-2-19

Principals' Message

It gives me immense pleasure to welcome all the delegates to the All India Seminar on Recent Trends in Image processing for AI implementation being organized by The Institution of Engineers (India) Punjab and Chandigarh State Centre in association with Department of Computer Science and Engineering of Baba Banda Singh Bahadur Engineering College, Fatehgarh Sahib, Punjab from 7-8 February 2019.

This seminar would serve as a perfect platform to bring together distinguished scholars to interact and share their expertise. The plenary sessions incorporating presentations and discussions by experts will certainly create awareness of the intricacies of the subject.

The participation of number of eminent scholars, scientists, educationists, and students on this occasion shall have a significant impact on the society through deliberations on a range of articles on image processing in a wide spectrum of applications.

I am certain that thoughts emerged during the Seminar will end up with key recommendations in recent trends in image processing applications and in particular for their implementation in AI.

I would like to make this Seminar a memorable event by offering sufficient opportunities and hope that you will enjoy the hospitality.

I wish the Seminar a great success.

Major General (Dr.) Gurcharan Singh Lamba, VSM (Retd)
Principal

Chandigarh Road, Fatehgarh Sahib, Punjab - 140 407 (India)
Ph : 01763 503056, 503067, 1800-137-2015 (Tool-free) **Fax:** 01763 503139
Website : www.bbsbec.edu.in **E-mail:** principal@bbsbec.ac.in

BABA BANDA SINGH BAHADUR ENGINEERING COLLEGE

Approved by AICTE, Govt. of Punjab, Affiliated to IKGPTU
(Courses Accredited by NBA (AICTE) and IE(I))

Ref. No. BBSBEC/CSE/2019/002 Date 5-2-19

Message from Convener

It is indeed a matter of pride to warmly welcome you all to All India Seminar on Recent Trends in Image processing for AI implementation being organized by Department of Computer Science and Engineering of Baba Banda Singh Bahadur Engineering College, Fatehgarh Sahib, Punjab in collaboration with The Institution of Engineers Punjab and Chandigarh state centre from 7-8 February 2019.

Research in all the fields of Engineering and Technology has to be undertaken earnestly, with the objective of original findings in new realms.I trust that the seminar will not only provide a useful forum to the participants to share their expertise in the field of image processing but will also be professionally beneficial to them.

It's our proud privilege to acknowledge our sincere gratitude to all the eminent experts who have spared their valuable time for this seminar. The deliberations would certainly lead to emergence of new ideas to find new pathways related to image processing applications and its implementation in AI.

I hope the delegates will have a comfortable stay and they will enjoy the traditional hospitality offered by the city.

We look forward to personally meeting each of you. Once again, welcome!

Dr. Baljit Singh Khehra
Convener

Chandigarh Road, Fatehgarh Sahib, Punjab - 140 407 (India)
Ph : 01763 503056, 503067, 1800-137-2015 (Tool-free) Fax: 01763 503139
Website : www.bbsbec.edu.in E-mail: principal@bbsbec.ac.in

BABA BANDA SINGH BAHADUR ENGINEERING COLLEGE

Approved by AICTE, Govt. of Punjab, Affiliated to IKGPTU
(Courses Accredited by NBA (AICTE) and IE(I))

Ref. No. BBSBEC/CSE/2019/003 Date. 5-2-19

Message from Organizing Secretary

It is my great pleasure to welcome you to the All India Seminar on Recent Trends in Image Processing using Artificial Intelligence on 7-8 February 2019. It has been a real honor and privilege to serve as the Organizing Secretary of the event. Seminar has provided a cross-disciplinary venue for researchers to address the rich space of Image Processing and Artificial Intelligence. The program spans two days and includes four keynote lectures followed by technical presentations. It will provide ample opportunities for discussions, debate, and exchange of ideas and information among participants.

I would like to express my appreciation to the Convener of the event, Dr. Baljit Singh for his valuable contribution in assembling the high quality conference program. I would like to acknowledge the efforts of Dr. Navneet Kaur Sehgal, convener paper review committee, and committee members for their invaluable help in the review process.

I am also grateful to all the authors who trusted the conference with their work. Special thanks to the Keynote Speakers, Dr. Sujata Pal, Dr. Gurleen Ahluwalia, Dr. Sarabjeet Singh and Dr. Akashdeep Sharma for sharing their views on current research topics. I appreciate the support of The Institution of Engineers Punjab and Chandigarh state centre. Likewise, I would also like to express my appreciation to the program and advisory committee, as well as to the Session chairs for their extraordinary preparation of the invited sessions. I look forward to an exciting event of insightful presentations, discussions, and sharing of technical ideas with colleagues from around the region. I thank you for attending the seminar and I hope that you enjoy your visit to BBSBEC.

Puneet Mittal

Puneet Mittal
Organizing Secretary

Chandigarh Road, Fatehgarh Sahib, Punjab - 140 407 (India)
Ph : 01763 503056, 503067, 1800-137-2015 (Tool-free) **Fax:** 01763 503139
Website : www.bbsbec.edu.in **E-mail:** principal@bbsbec.ac.in

Editorial Board

COMMITTEES

All India Seminar

On

Recent Trends in Image processing for AI implementation

NATIONAL ADVISORY COMMITTEE

Chairman: Mr. Sisir Kumar Banerjee, FIE, President, IEI
Co-Chairman: Mr. B S Patel, FIE, Chairman, CPDB, IEI
Convener: Dr.Baljit Singh, FIE, Professor and Head, CSE, BBSBEC, Fatehgarh Sahib

Members

Mr. Timothy Dkhar, FIE, Member, CPDB, IEI
Mr. Nirmal Das, FIE, Member, CPDB, IEI
Mr. K Rajanikanth, FIE, Member, CPDB, IEI
Dr. Hari Shankar Sharma, FIE, Member, CPDB, IEI
Mr. P Hirabhai Tandel, FIE , Member, CPDB, IEI
Mr. Parminder Singh Bhogal, fIE, Imm. Past Chairman, IEI
Mr. G.S. Mast, FIE, Past, Honorary Secretary, IEI
Mr. H.C.S. Berry, FIE, Past President, IEI
Mr. Jagroop Singh, FIE, Vice President, IEI
Mr. A.S. Bakshi, FIE, Council Member, IEI
Mr. S.D. Singh, MIE, IEI
Dr. Mithilesh Kumar Jha, FIE, IEI
Mr. Narinder Pal Singh, MIE, IEI
Mr. Bakshish Singh, MIE, IEI
Mr. Sohan Lal, MIE, IEI
Mr. R.P. Singla, FIE, IEI
Mr. Paramjeet singh, MIE, IEI
Dr. J.S. Johal, FIE, IEI
Mr. Swinder Singh, FIE, IEI
Dr. Niraj Bala, MIE, IEI
Mr. Jaswant Singh Gill, MIE, IEI
Mr. U.S. Kooner, FIE, IEI
Brig. Mastinder Singh, FIE, IEI
Brig. K.V. Prasad, FIE, IEI
Dr. Ajay K. Sharma, Vice Chancellor, IKGPTU
Dr. Sarit K. Das, Director, IIT Ropar
Dr. N.P. Singh, Dean(Planning & External Prog.), IKGPTU
Dr. D.S. Hira, Chairman, ISTE (PB., Chd., HP and J&K section)
Dr. Gurdeep Singh Hura, Professor, UMES, Maryland, USA
Dr. Renu Cheema Vig, Director, UIET, Chandigarh
Dr. Maninder Singh, Professor, TIET, Patiala
Dr. Arvind Kalia, Professor, HP University, Shimla
Dr. Rajesh Bhatia, Professor, PEC, Chandigarh
Dr. Savina Bansal, Professor and Dean (R&D), MRSPTU, Bathinda
Dr. Kanwaljit Singh, Professor & Director Computer Center, Pbi, University, Patiala
Dr. R.K. Bansal, Professor & Dean (Student Welfare), MRSPTU, Bathinda
Dr. A.P. Singh, Professor, SLIET, Longowal

Honorary Secretary

Er. Sukhvir Singh Mundi, Honorary Secretary, Punjab and Chandigarh State Centre

Organizing Committee

Chairman: Mr.Janak Raj Garg, FIE, Chairman, Punjab & Chandigarh State Centre, IEI
Co-Chairman: Maj.Gen.(Dr) Gurcharan Singh Lamba, VSM(Retd), Principal, BBSBEC, Fatehgarh Sahib
Organizing Secretary: Er.Sukhvir Singh Mundi, FIE, Honorary Secretary, Punjab and Chandigarh State Centre
Er. Puneet Mittal, MIE, Assistant Professor(Sr. Scale), BBSBEC, Fatehgarh Sahib

Members

Dr. Kanwalvir Singh Dhindsa, Professor, CSE, BBSBEC
Er. Amanjot Kaur, AP(Sr.Scale), CSE, BBSBEC
Er. Khyati Marwaha, AP(Sr.Scale), CSE, BBSBEC
Er. Jatinder Singh, AP(Sr.Scale), CSE, BBSBEC
Er. Deepak Aggarwal, AP(Sr.Scale), CSE, BBSBEC
Er. Gurleen Kaur Dhaliwal, AP, CSE, BBSBEC
Er. Sandeep Kaur Dhanda, AP, CSE, BBSBEC
Er. Gurleen Sidhu, AP, CSE, BBSBEC
Dr. Navneet Kaur , AP, CSE, BBSBEC
Er. Manpreet Kaur, AP, CSE, BBSBEC
Er. Rupinder Kaur, AP, CSE, BBSBEC
Er. Gurbahar Singh, AP, CSE, BBSBEC
Er. Sukhwinder Sharma, AP, CSE, BBSBEC
Er. Gurpreet Kaur, AP, CSE, BBSBEC
Er. Rupinder Kaur, AP, CSE, BBSBEC

Technical Committee

Chairman: Er. T.S. Kamal, FIE, IEI
Co-Chairman: Maj. Gen. (Dr) Gurcharan Singh Lamba, VSM (Retd), Principal, BBSBEC, Fatehgarh Sahib

Members

Dr. Lakhwinder Singh , Dean Academics, BBSBEC, FGS
Dr. A.P.S. Sethi, Dean Student affairs, BBSBEC, FGS
Dr. J.S. Oberoi, Head T&P, BBSBEC, FGS
Dr. H.S. Bhatti, Dean Examination, BBSBEC, FGS
Dr. Lakhvir Singh, Head, Mechanical Engg., BBSBEC, FGS ,
Dr. Virinder Singh, Chief Security Officer, BBSBEC, FGS
Dr. Amritbir Singh, Head Applied Sciences, BBSBEC, FGS
Dr. G.S. Brar, Head Electrical Engg, BBSBEC, FGS
Dr. R.S. Uppal, Head ECE, BBSBEC, FGS
Dr. B.S. Bhullar, Head MBA, BBSBEC, FGS
Dr. Niraj Bala, Head RIC., BBSBEC, FGS
Dr. Anshu Tomar, Head Civil Engg., BBSBEC, FGS

Technical Papers

ALL INDIA SEMINAR

on

RECENT TRENDS IN IMAGE PROCESSING FOR AI IMPLEMENTATION

Organized by

THE INSTITUTION OF ENGINEERS (INDIA)

PUNJAB AND CHANDIGARH STATE CENTRE

in association with

Baba Banda Singh Bahadur Engineering College, Fatehgarh Sahib-140407

Date: 07-02-2019		
11:00 am - 12:30 pm	**Key Note Lecture by Dr. Sujata Pal AP, IIT Ropar**	
01:30 pm - 03:00 pm	**Key Note Lecture by Dr. Akashdeep Sharma, AP, UIET, Chandigarh**	
Technical Session – I (03:00 pm – 5:00 pm)		
Session Coordinator		**Prof. Sukhwinder Sharma, Assistant Professor, BBSBEC, Fatehgarh Sahib**
Session Chair		**Dr. Navdeep Singh, Associate Professor, MGC, Fatehgarh Sahib**
Paper No.	**Name of Paper**	**Author Name(s)**
AI_001	Two Step Image Denoising	Parvinder Kaur, Computer Science &Engineering Shaheed Udham Singh College of Engineering & Technology Mohali, India. Manpreet Kaur, Computer Science & Engineering Shaheed Udham Singh College of Engineering & Technology Mohali, India.
AI_002	Segmentation of Speech into Syllables	Samanjeet Kaur, Department of Computer Science and Engineering Baba Banda Singh Bahadur Engineering College Fatehgarh Sahib, India. Puneet Mittal, Department of Computer Science and Engineering Baba Banda Singh Bahadur Engineering College Fatehgarh Sahib, India.
AI_003	A study of Deep Learning Architectures for object detection	Payal, Research Scholar UIET, Panjab University, Chandigarh. Akashdeep Sharma, Assistant Professor UIET, Panjab University, Chandigarh.
AI_004	Fruit Image Enhancement method based on Bacterial Foraging Optimization using Fuzzy Logic	Harmandeep Singh Gill, Research Scholar Punjab Technical University Jalandhar. Baljit Singh Khehra, Professor & Head Department of CSE Baba Banda Singh Bahadur Engineering College Fatehgarh Sahib.
AI_005	A Brief Summary of Generative Adversarial Networks	Pratik Joshi, B.E Student, UIET Panjab University, Chandigarh.
AI_006	K- Nearest Neighbor Classification for the cancer patient data for the disease detection	Gulshan Bleem, Department of Computer Science and Engineering Baba Banda Singh Bahadur Engineering College Fatehgarh Sahib, India. Ishpreet Singh Virk, Department of Computer Science and Engineering Baba Banda Singh Bahadur Engineering College Fatehgarh Sahib, India.
AI_007	Soft Computing Approach to Image denoising	Ravneet Kaur, Computer Science & Engineering ShaheedUdham Singh College of Engineering & Technology Mohali, India. Parvinder Kaur, Computer Science & Engineering ShaheedUdham Singh College of Engineering & Technology Mohali, India
AI_008	Performance Comparison of Homogenous and Heterogeneous Clustering Techniques in WSNs	Harjit Kaur, Department of Computer Science and Engineering Baba Banda Singh Bahadur Engineering College Fatehgarh Sahib, Punjab, India. Sukhwinder Sharma, Department of Computer Science and Engineering Baba Banda Singh Bahadur Engineering College Fatehgarh Sahib, Punjab, India.
AI_009	Intelligent Vision Extraction: A Review	Ramanjot Kaur, Assistant Professor Computer Science & Engineering Baba Banda Singh Bahadur Engineering College Fatehgarh Sahib, India. Sandeep Kaur, Assistant Professor Computer Science & Engineering Baba Banda Singh Engineering Bahadur College Fatehgarh Sahib, India. Nishu Chaubey, Student Computer Science & Engineering Baba Banda Singh Engineering Bahadur College Fatehgarh Sahib, India.
AI_010	Network Security Assessment using Penetration Testing in a Rural Banking Environment: The Case of Akuapem Rural Bank	Wellington Amponsah, Computer Science and Engineering Department ShaheedUdham Singh College of Engineering and Technology Tangori Punjab, India.

ALL INDIA SEMINAR
on
RECENT TRENDS IN IMAGE PROCESSING FOR AI IMPLEMENTATION
Organized by
THE INSTITUTION OF ENGINEERS (INDIA)
PUNJAB AND CHANDIGARH STATE CENTRE
in association with
Baba Banda Singh Bahadur Engineering College, Fatehgarh Sahib-140407

Date: 07-02-2019		
Technical Session – II (03:00 pm – 5:00 pm)		
	Session Coordinator	**Prof. Gurpreet Kaur, AP, BBSBEC, Fatehgarh Sahib**
	Session Chair	**Dr. Baljit Singh Professor, BBSBEC, Fatehgarh Sahib**
Paper No.	**Name of Paper**	**Author Name(s)**
AI_011	Perspectives of Green Computing.	Taranpreet Singh Computer Science and Engineering, Student Baba Banda Singh Bahadur Engineering College Fatehgarh Sahib, Punjab. Arshdeep Singh Computer Science and Engineering, Student Baba Banda Singh Bahadur Engineering College Fatehgarh Sahib, Punjab.
AI_012	Actions on Google: Developer Platform for Google Assistant.	Arshdeep Singh CSE, Student Baba Banda Singh Bahadur Engineering College, Fatehgarh sahib, Punjab, India. Taranpreet Singh CSE, Student Baba Banda Singh Bahadur Engineering College Fatehgarh sahib, Punjab, India.
AI_013	Review Paper on Skin Cancer Detection.	Shelly Research scholar, Department of Electronics and communication Punjabi University Patiala,Punjab, India Balkrishan A.P., Computer Engineering Section, Yadavindra College of Engineering, GKC, Punjabi University Talwandi Sabo, Bathinda, Punjab, India
AI_014	Flow Oriented Channel Assignment for Multi Radio Wireless Mesh Networks.	Jaswinder Kaur Department of Computer Science and Engineering Baba Banda Singh Bahadur Engineering College Fatehgarh Sahib, Punjab. Jatinder Singh Saini Department of Computer Science and Engineering Baba Banda Singh Bahadur Engineering College Fatehgarh Sahib, Punjab
AI_015	Studying Effects of training a neural net after expected results are obtained.	Harjot Singh , Computer Science & Engineering, Chandigarh Engineering College, Mohali, India Gurleen Kaur, Computer Science & Engineering, Chandigarh Engineering College, Mohali,India
AI_016	A Review of Digital Image Processing Application Areas.	Manjot Kaur, Department of Computer Science Engineering, Chandigarh University,Mohali , Punjab (India) Navneet Kaur, Department of Computer Science Engineering, BBSBEC, Fatehgarh Sahib,Punjab (India)
AI_017	Attendance System using Face Recognition.	Anju Bala, Assistant Professor, CSE Department Chandigarh University Chandigarh, India. BhavnaUpadhyay, Student BE, CSE Department Chandigarh University Chandigarh, India.
AI_018	Emergence of Artificial Intelligence in Defense Sector.	Neelam Rana, CSE Department Baba Banda Singh Bahadur Engineering College Fatehgarh Sahib.
AI_019	A Self-tuning System for Big Data Analytics.	Arshpreet Kaur, Associate Professor, Computer Science Engineering, Chandigarh University, Mohali, India. Amritpal Singh, Student BE, Computer science engineering Chandigarh University Mohali, India.
AI_020	Impact of Inheritance Metrics on Fault Prediction using Clustering.	Deepak Kalia, Department of Computer Science and Engineering Baba Banda Singh Bahadur Engineering College Fatehgarh Sahib, Punjab. Puneet Mittal, Department of Computer Science and Engineering Baba Banda Singh Bahadur Engineering College Fatehgarh Sahib, Punjab.

ALL INDIA SEMINAR

on

RECENT TRENDS IN IMAGE PROCESSING FOR AI IMPLEMENTATION

Organized by

THE INSTITUTION OF ENGINEERS (INDIA)

PUNJAB AND CHANDIGARH STATE CENTRE

in association with

Baba Banda Singh Bahadur Engineering College, Fatehgarh Sahib-140407

Date: 08-02-2019		
09:30 am - 10:45 am		**Key Note Lecture by Dr. Gurleen Ahluwalia, AP, University College, Ghanour**
11:00 am - 12:30		**Key Note Lecture by Dr. Sarabjeet Singh, UIET, Chandigarh**
Technical Session – III (10:00 PM – 2:00 PM)		
	Session Coordinator	**Prof. Ishpreet Singh, AP, BBSBEC, Fatehgarh Sahib**
	Session Chair	**Dr. Williamjeet Singh, AP, Punjabi University, Patiala**
Paper No.	**Name of Paper**	**Author Name(s)**
AI_021	Impact of Coupling and Cohesion Metrics on Fault Prediction using Clustering.	Ashmeet Kaur, Student, Department of Computer Science and Engineering, Baba Banda Singh Bahadur Engineering College Fatehgarh Sahib Puneet Mittal, AP, Department of Computer Science and Engineering, Baba Banda Singh Bahadur Engineering College Fatehgarh Sahib
AI_022	Digital Image Watermarking Techniques: A Comprehensive Survey.	Harneel Kaur Cheema, Department of Computer Science and Engineering Baba Banda Singh Bahadur Engineering College Fatehgarh Sahib
AI_023	Wireless Sensor Networks And Its Applications: A Review.	Sandeep Kaur Saini, PG Research Scholar, Baba Banda Singh Bahadur Engineering College Fatehgarh Sahib
AI_024	Information Extraction with the help of Machine Learning.	Hema, Assistant professor Department of Computer Science and Engineering Chandigarh University, Gharuan Shourya, Student BE Department of Computer Science and Engineering Chandigarh University, Gharuan
AI_025	Benefits of P2P technologies.	Geetika, Assistant Professor, CSE, Chandigarh University, Gharuan Chandigarh, India Abhinay Pathania, Student, BE CSE, Chandigarh University, Gharuan Chandigarh, India
AI_026	Unfolding Smartness of Neural Nets.	Tarandeep Singh, B.Tech Scholar, Department of CSE Baba Banda Singh Bahadur Engineering College Fatehgarh Sahib Perminder Kaur, Assistant Professor, Department of CSE Baba Banda Singh Bahadur Engineering College Fatehgarh Sahib Ravneet Kaur, Assistant Professor, Department of CSE Baba Banda Singh Bahadur Engineering College Fatehgarh Sahib
AI_027	Security In Applications of Blockchain and their Enhancement.	Ramanjit Kaur, B.Tech Scholar, Department of CSE Baba Banda Singh Bahadur Engineering College Fatehgarh Sahib Perminder Kaur, Assistant Professor, Department of CSE Baba Banda Singh Bahadur Engineering College Fatehgarh Sahib Latashi Bector, B.Tech Scholar, Department of CSE Baba Banda Singh Bahadur Engineering College Fatehgarh Sahib Ravneet Kaur, Assistant Professor, Department of CSE Baba Banda Singh Bahadur Engineering College Fatehgarh Sahib

AI_028	Review Paper on Automatic Braking System using fuzzy logic.	Ramandeep Singh, Computer Science & Engineering, Baba Banda Singh Engineering Bahadur College, Fatehgarh Sahib Shivam Gupta, Computer Science & Engineering, Baba Banda Singh Engineering Bahadur College, Fatehgarh Sahib Harshdeep Kaur, Computer Science & Engineering, Baba Banda Singh Engineering Bahadur College, Fatehgarh Sahib Ramanjot Kaur, Computer Science & Engineering Baba Banda Singh Engineering Bahadur College, Fatehgarh Sahib
AI_029	An Deep Insight into IOT Technologies and Applications.	Amrita Preet Kaur, M.Tech Scholar, Dept of Computer Engineering, Baba Banda Singh Bahadur Engineering College, Fatehgarh Sahib, India
AI_030	Introduction to Cloud Computing Paradigm.	Anmoldeep Singh, Student CSE, Baba Banda Singh Bahadur Engineering College Fatehgarh Sahib Sandeep Kaur, CSE Dept, Baba Banda Singh Bahadur Engineering College Fatehgarh Sahib Preetinder Kaur, CSE Dept, Baba Banda Singh Bahadur Engineering College Fatehgarh Sahib
AI_031	A Review Paper on Problem Solving Methods in Artificial Intelligence using Searching.	Amandeep Kaur, Student, CSE, Baba Banda Singh Bahadur Engineering College, Fatehgarh Sahib Amandeep Kaur, Student, CSE, Baba Banda Singh Bahadur Engineering College, Fatehgarh Sahib Preetinder Kaur, CSE, Baba Banda Singh Bahadur Engineering College, Fatehgarh Sahib Harshdeep Kaur, CSE,Baba Banda Singh Bahadur Engineering College, Fatehgarh Sahib
AI_032	HMM Based POS Tagger for Punjabi.	Simrat Walia, Department of CSE, Baba Banda Singh Bahadur Engineering College, Fatehgarh Sahib
AI_033	Android Application Development to Detect Threads using Application Programs Interface	Harpreet Singh, Student, Department of CSE, Baba Banda Singh Bahadur Engineering College, Fatehgarh Sahib Sandeep Kaur Dhanda, AP, Department of CSE, Baba Banda Singh Bahadur Engineering College, Fatehgarh Sahib

Two Step Image Denoising

Parvinder Kaur
Computer Science & Engineering
Shaheed Udham Singh College of Engineering & Technology
M ohali, India
parvinderkaurcse@sus.edu.in

Manpreet Kaur
Computer Science & Engineering
Shaheed Udham Singh College of Engineering & Technology
M ohali, India
manpreetk7322@gmail.com

***Abstract*—Images are often corrupted with noise during acquisition, transmission, and retrieval from storage media. Many dots can be spotted in a Photograph taken with a digital camera under low lighting conditions. The images are often used further for some applications. S o noise removal is a pre - processing step in most of the application. In this paper at first step the non-local means filter is used for noise removal. In the second step a hybrid approach is used to optimize the result of first step. Ant colony optimization is used in combination with bacterial foraging optimization. The proposed method is compared with the state of the art technique and performs better.**

Keywords—non-local, ant-colony, bacterial foraging, denoising.

I. INTRODUCTION

Digital images play an important role in many applications such as geographical information systems as well as it is the most vital part in the field of medical science such as ultrasound imaging, X-ray imaging, Computer tomography and MRI. A very large portion of digital image processing includes image restoration. Image restoration is a method of removal or reduction of degradation that are incurred during the image capturing. Degradation comes from blurring as well as noise due to the electronic and photometric sources. Blurring is the form of bandwidth reduction of images caused by imperfect image formation process such as relative motion between camera and original scene or by an optical system that is out of focus. Noise is unwanted signal that interferes with the original signal and degrades the visual quality of digital image. The main sources of noise in digital images are imperfect instruments, problem with data acquisition process, interference natural phenomena, transmission and compression [10]. Image denoising forms the pre-processing step in the field of photography, research, technology and medical science, where somehow image has been degraded and needs to be restored before further processing. Image denoising is still a challenging problem for researchers as image denoising causes blurring and introduces artifacts. Denoising method tends to b e problem specific and depends upon the type of image and noise model.

A. Noise modelling

Noise (n) may be modeled either by a histogram or a probability density function which is superimposed on the probability density function of the original image (s)[2]. Noise represents unwanted information which deteriorates image quality. Noise is defined as a function (n) which affects the acquired image (f) and is not part of the scene (initial signal – s). Using the additive noise model, this process can be written as:

$$f(i, j) = s(i, j) + n(i, j) \quad (1)$$

Digital image noise may come from various sources. The acquisition process for digital images converts optical signals into electrical signals and then into digital signals and is one processes by which the noise is introduced in digital images. Each step in the conversion process experiences fluctuations, caused by natural phenomena, and each of these steps adds a random value to the resulting intensity of a given pixel.

B. Types of Noises

- *Salt and pepper noise:* In the salt & pepper noise model only two possible values are possible, a and b, and the probability of obtaining each of them is less than 0.1 (otherwise, the noise would vastly dominate the image). For an 8 bit/pixel image, the typical intensity value for pepper noise is close to 0 and for salt noise is close to 255. The salt & pepper noise is generally caused by malfunctioning of camera's sensor cells, by memory cell failure or by synchronization errors in the image digitizing or transmission.
- *Gaussian Noise:* The Gaussian noise has a normal (Gaussian) probability density function. Gaussian noise is useful for modeling natural processes which introduce noise (e.g. noise caused by the discrete nature of radiation and the conversion of the optical signal into an electrical one – detector/shot noise, the electrical noise during acquisition – sensor electrical signal amplification, etc.).
- *White Noise:* When the Fourier Spectrumof noise is constant the noise is called White Noise. The terminology comes from the fact that the white light contains nearly all frequencies in the visible spectrum in equal proportions. The Fourier Spectrum of a function containing all frequencies in equal proportions is a constant.
- *Rayleigh Noise:* It arises in Range Imaging. The noise that follows Rayleigh distribution is called Rayleigh noise. Rayleigh distribution is used in medical imaging science, to model noise variance in magnetic resonance imaging.

C. Spatial Filtering

Spatial filtering is an image processing technique for changing the intensities of a pixel according to the intensities of the neighboring pixels. Using spatial filtering, the image is transformed (convoluted) based on a kernel H which has certain height and width (x,y), defining both the area and the weight of the pixels within the initial image that will replace the value of the image. The corresponding process is to convolve the input image I(i,j) with the filter function H(x,y),to produce the new filtered image:

$$I'(i,j) = I(i,j) \circledcirc H(x,y) \quad (2)$$

Low pass filtering also known as smoothing removes high spatial frequency noise from a digital image. The low-pass

filters usually employ moving window operator which affects one pixel of the image at a time, changing its value by some function of a local region (window) of pixels. The operator moves over the image to affect all the pixels in the image.A high-pass filter can be used to make an image appear sharper. These filters emphasize fine details in the image - the opposite of the low-pass filter. High-pass filtering works in the same way as low-pass filtering; it just uses a different convolution kernel. Mean filtering is a simple, intuitive and easy to implement method of smoothing images, i.e. reducing the amount of intensity variation between one pixel and the next. It is often used to reduce noise in images. [4]The idea of mean filtering is simply to replace each pixel value in an image with the mean (`average') value of its neighbors, including itself. This has the effect of eliminating pixel values which are unrepresentative of their surroundings. Mean filtering is usually thought of as a convolution filter. Like other convolutions it is based around a kernel, which represents the shape and size of the neighborhood to be sampled when calculating the mean. There are some Problem with mean filtering.A single pixel with a very unrepresentative value can significantly affect the mean value of all the pixels in its neighborhood. When the filter neighborhood straddles an edge, the filter will interpolate new values for pixels on t he edge and so will blur that edge. This may be a problem if sharp edges are required in the output. Both of these problems are tackled by the median filter, which is often a better filter for reducing noise than the mean filter, but it takes longer to compute. The median filter is normally used to reduce noise in an image, somewhat like the mean filter. However, it often does a better job than the mean filter of preserving useful detail in the image. Like the mean filter, the median filter considers each pixel in the image in turn and looks at its nearby neighbors to decide whether or not it is representative of its surroundings. Instead of simply replacing the pixel value with the mean of neighboring pixel values, it replaces it with the median of those values. The median is calculated by first sorting all the pixel values from the surrounding neighborhood into numerical order and then replacing the pixel being considered with the middle pixel value. (If the neighborhood under consideration contains an even number of pixels, the average of the two middle pixel values is used). The median is a more robust average than the mean and so a single very unrepresentative pixel in a neighborhood will not affect the median value significantly. Since the median value must actually be the value of one of the pixels in the neighborhood, the median filter does not create new unrealistic pixel values when the filter straddles an edge. For this reason the median filter is much better at preserving sharp edges than the mean filter. The Gaussian smoothing operator is a 2-D convolution operator that is used to `blur' images and remove detail and noise. In this sense it is similar to the mean filter, but it uses a different kernel that represents the shape of a Gaussian (`bell-shaped') hump [11]. The Gaussian outputs a `weighted average' of each pixel's neighborhood, with the average weighted more towards the value of the central pixels. This is in contrast to the mean filter's uniformly weighted average. Because of this, a Gaussian provides gentler smoothing and preserves edges better than a similarly sized mean filter.

There is a property of noise. Noise is generally considered to be a random variable with zero mean. Consider a noisy pixel, p=p0+n where p0 is the true value of pixel and n is the noise in that pixel. You can take large number of same pixels (say N) from different images and computes their average. Ideally, you should get p=p0 since mean of noise is zero.

You can verify it yourself by a simple setup. Hold a static camera to a certain location for a couple of seconds. This will give you plenty of frames, or a lot of images ofthe same scene. Then write a piece of code to find the average of all the frames in the video (This should be too simple for you now). Compare the final result and first frame. You can see reduction in noise. Unfortunately this simple method is not robust to camera and scene motions. Also often there is only one noisy image available. So idea is simple, we need a set of similar images to average out the noise. Consider a small window (say 5x5 windows) in the image. Chance is large that the same patch may be somewhere else in the image. Sometimes in a small neighborhood around it. The blue patches in the image look the similar. Green patches looks similar. So we take a pixel, take small window around it, search for similar windows in the image, average all the windows and replace the pixel with the result we got. This method is Non-Local Means Denoising. It takes more time compared to blurring techniques we saw earlier, but its result is very good [12].

II. LITERATURE REVIEW

A. Patel et al. (2016) explained that a fundamental step in image processing is the step of removing various kinds of noise from the image. The important property of a good image denoising model is that it should completely remove noise as far as possible as well as preserve edges. The image denoising technique will be mainly depending on the type of the image and noise in cooperating with it. There have been several published algorithms and each approach has its assumptions, advantages and limitations [6].

X. Wang et al. (2016) proposed that basic principle of non - local means is to denoise a pixel using the weighted average of the neighborhood pixels, while the weight is decided by the similarity of these pixels. The key issue of the non -local means method is how to select similar patches and design the weight of them. There are two main contributions of this paper: The first contribution is that we use two images to denoise the pixel. These two noised images are with the denoising process author get a pre-denoised image and a residual image [7].

A. Buades et al. (2015) defined a general mathematical and experimental methodology to compare and classify classical image denoising algorithms, second to propose an algorithm (Non Local Means) addressing the preservation of structure in a digital image. The mathematical analysis is based on the analysis of the “method noise”, defined as the diff erence between a digital image and its denoised version. The NL-means algorithm is also proven to be asymptotically optimal under a generic statistical image model[1].

J. patil et al. (2015) has described that visual information transmitted in the form of digital images is becoming a major method of communication in the modern age but the image obtained after transmission is often corrupted with noise. The received image needs processing before it can be used in applications. Image denoising involves the manipulation of the image data to produce a visually high quality image. The author has reviewed that the Noise models, Noise types and classification of image denoising techniques. The author presented a comparative analysis of various noise suppression algorithms [8].

S. Kaur et al. (2014) has mentioned the main challenge in digital image processing is to remove noise from the original image. The author has reviewed the existing denoising algorithms and performs their comparative study. Different noise models including additive and multiplicative types are discussed. Selection of the denoising algorithm is application dependent. Hence it is necessary to have knowledge about the noise present in the image so as to select the appropriate denoising algorithm. Author puts results of different approaches of wavelet based image denoising methods using several thresholding techniques such as Bayes Shrink, Sure Shrink and Visu Shrink. A quantitative measure of comparison is provided by SNR (signal to noise ratio) and mean square error (MSE) [9].

S. shreshtha et al. (2014); presented a new decision based technique which shows better performances than those already being used. The comparisons are made based on visual appreciation and further quantitatively by Mean Square error (MSE) and Peak Signal to Noise Ratio (PSNR) of different filtered images [19].

J. Caia et al. (2013) introduced Scarcity based regularization methods for image restoration assumes that the underlying image has a good sparse approximation under a certain system. Such a system can be a basis, a frame or a general over-complete dictionary. One widely used class of such systems in image restoration are wavelet tight frames. Such an adaptive tight frame construction scheme is applied to image denoising by constructing a tight frame tailored to the given noisy data.

M. kaur et al. (2013) has proposed an adaptive threshold estimation method for image denoising in the wavelet domain based on the generalized Gaussian distribution (GGD) modeling of sub band coefficients. The proposed method called Normal Shrink is computationally more efficient and adaptive because the parameters required for estimating the threshold depend on sub band data. The threshold is computed by $\beta\sigma 2 / \sigma y$ where σ and σy are the standard deviation of the noise and the sub band data of noisy image respectively. β is the scale parameter which depends upon the sub band size and number of decompositions. Experimental results on several test image are compared with various denoising techniques like wiener Filtering [15].

R. ahmadi et al. (2013) discussed image denoising called as a mean filter that acts on an image by smo othing it. It reduces the intensity variation between adjacent pixels. The mean filter is nothing but a simple sliding window spatial filter that replaces the center value in the window with the average of all the neighboring pixel values including itself. Image corrupted with salt and pepper noise is subjected to mean filtering and it can be observed that the noise dominating is reduced[18].

T.L. Sahu et al. (2012) stated that digital images are noisy due to environmental disturbances. To ensure image quality image processing of noise reduction is a very important step before analysis or using images. Data sets collected by image sensors are generally contaminated by noise. Imperfect instruments, problems with the data acquisition process and interfering with natural phenomena can all degrade the data of interest. The importance of the image denoising could be a serious task for medical imaging, satellite and a real image processing robot vision, industrial vision systems, micro vision systems, space exploring etc [24].

S. Ruikar et al. (2011) proposed different approaches of wavelet based image denoising methods. The main aim is to modify the wavelet coefficients in the new basis, the noise can be removed from the data. The author extended the existing technique and providing a comprehensive evaluation of the proposed method. Results based on different noise, such as Gaussian, Poisson's, Salt and Pepper and Speckle performed signal to noise ratio as a measure of the quality of denoising was preferred.

L. Yan et al. (2011) has used the noisy chaotic neural network (NCNN) that has proposed earlier for image denoising as a constrained optimization problem. The experimental results show that the NCNN is able to offer good quality solutions [13].

V. Laparra et al. (2010) explained a successful class of image denoising methods is based on Bayesian approaches working on wavelet representations. The author proposed an alternative non-explicit way to take into account the relations among natural image wavelet coefficients fo r denoising, use of support vector regression (SVR) in the wavelet domain to enforce these relations in the estimated signal. Since relations among the coefficients are specific to the signal, the regularization property of SVR is exploited to remove the noise which does not share this feature [25].

F. Palhano et al. (2010) the author presented a method for real-time denoising of ultrasound images: a modified version of the NL-means method is presented that incorporates an ultrasound dedicated noise model as well as a GPU implementation of the algorithm. Results demonstrate that the proposed method is very efficient in terms of denoising quality and is real-time [3].

III. THE PROPOSED METHOD

The proposed method is a two-step technique. In the first step the nonlocal means denoising is applied. The results of the first step are passed to second step i.e. hybrid optimization.

AN N on Local Means Filtering:

Non Local mean filter si a speacial kind of non linear filter.The basic principle of nonlocal means is to denoise a pixel using the weighted average of the neighborhood pixels, while the weight is decided by the similarity of these pixels. The key issue of the nonlocal means method is how to select similar patches and design then weight of them. Assumptions for Non Local Means Denoising

1. Natural images have repetitive textures.

2. Pixels with similar textures will probably have similar values.
3. More discriminative than bilateral filtering.

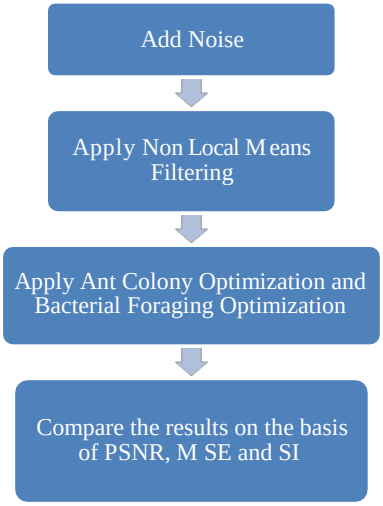


Figure.1. Flow Chart for the Proposed Method

A. *Ant Colony Optimization*

Ant colony optimization increases the contrast of images. In this way, contrast enhancement is obtained by global transformation of the input intensities. Ant colony optimization is used to generate the transfer functions which map the input intensities to the output int ensities. The results indicate that the new method achieves images with higher contrast than the previously presented methods from the subjective and objective viewpoints. Further, the proposed algorithm preserves the natural look of input images. Ant colony optimization is established on the movement of artificial ants in the search space. Each ant deposits a substance called pheromone when searching. More good results achieved by an ant lead to stronger pheromone deposition. Pheromone causes that other ants be disposed to choose previous ant's way with more probability. Therefore ACO uses the convergence property of pheromone trails while pheromone rate and its deviation determine the expected convergence time. On the other side, evaporation of pheromone assures exploration and tries to prevent immature convergences. Ant colony optimization is the core of the hybrid algorithm. It achieves near optimum transfer functions. The basic component of ACO is the dynamic evaluation value (pheromone) which affects ant's decision for moving. Solution construction and pheromone update are two common stages in ACO iterations. In addition to Heuristic values of ant's neighborhood, pheromone, the static evaluation value (heuristic value) is another effective parameter in ant's decision [16].

B. *Bacterial Foraging Optimization*

In next stage to remove the error between original image and noisy image BFO is used. This optimization technique minimizes the mean square error (MSE) between recovered image g(x, y) and the Noisy image. The Mean Square Error is considered as cost function for BFO to optimize Peak Signal to Noise. The Bacterial Foraging Optimization Algorithm is inspired by the group foraging behavior of bacteria such as E.coli and M.xanthus. Specifically, the BFOA is inspired by the chemotaxis behavior of bacteria that will perceive chemical gradients in the environment (such as nutrients) and move toward or away from specific signals.

I V. RESULT S & DISCUSSIONS

Figure.2. De-Noised Image after Non local Means and Hybrid Optimization

Figure.3. De-Noised Image after Non local Means and Hybrid Optimization

TABLE 1: PERFORMANCE EVALUATION (MEAN SQUARE ERROR)

Images	***MSE (Base Method)***	***MSE (Proposed Method)***
Image 1	.148	.120
Image 2	.140	.118
Image 3	.142	.122
Image 4	.140	.121
Image 5	.144	.124

TABLE 2: PERFORMANCE EVALUATION (PEAK SIGNAL TO NOISE RAT IO)

Image s	***PSNR (Base Method)***	***PSNR (Proposed Method)***
Image 1	35.3	55.28
Image 2	36.4	61.62
Image 3	34.4	57.83
Image 4	32.3	54.39
Image 5	36.7	56.63

TABLE 3: PERFORMANCE EVALUATION (STRUCTURAL SIMILARIT Y INDEX MEASURE)

Image s	***SSIM (Base Method)***	***SSIM (Proposed Method)***
Image 1	.7010	.7207
Image 2	.7201	.7507
Image 3	.6990	.7206
Image 4	.6787	.6997
Image 5	.7017	.7407

V. CONCLUSION & FUTURE SCOPE

Image denoising is an applicable issue found in diverse image processing and computer vision problems. There are various existing methods to denoise image. The important property of a good image denoising model is that it should completely remove noise as far as possible as well as preserve edges. Non Local Means filter performs well for Gaussian noise filtering while preserving edges and details of the original image. Contrast enhancement is obtained by ant colony optimization. Bacterial Foraging Optimization Technique proves to be a robust approach for automatically suppressing the noise yielding a better signal to noise ratio. The results of proposed method indicate that the new method achieves images with higher contrast than the previously presented methods from the subjective and objective viewpoints. Further, the proposed algorithm preserves the natural look of input images. In the future the techniques for colored image noise removal can be proposed.

REFERENCES

[1] A.Buades, "Image denoising methods," IEEE Trans, vol.32, n0. 18, pp. 321-327, 2015.

[2] Pizurica and W. Philips, "Estimating the probability of the presence of a signal of interest in multiresolution single- and multiband image denoising," IEEE Trans. Image Process., vol. 15, no. 3, pp. 654–665, 2006.

[3] F. Palhano., " Real time ultrasound image denoisingT o cite this version:HAL Id: inria-00476122 Real time ultrasound image denoising," 2010.

[4] H. Guo, J.E. Odegard, and C.S. Burrus,"Nonlinear processingofa shift invariant DWT for noise reduction," SPIE, MathematicalImaging: Wavelet Applications for Dual Use, vol. 10,no. 4, pp.333-339, 2007.

[5] J. F. Cai, H. Ji, Z. Shen, and G. B. Ye, " Data-driven tight frame construction and image denoising," Appl. Comput. Harmon. Anal., vol. 37, no. 1, pp. 89–105, 2014.

[6] J. Patil and S. Jadhav, " A Comparative Study of Image Denoisin g T echniques," Int. J. Innov. Res. Sci. Eng. T echnol., vol. 2, no. 3, pp. 787–794, 2015.

[7] J. Mohan, V. Krishnaveni, and Y. Guo, " A survey on the magnetic resonance image denoising methods," Biomed. Signal Process. Control, vol. 9, no. 1, pp. 56 –69, 2014.

[8] K. Dabov, A. Foi, and V. Katkovnik, " Image denoising by sparse 3D transformation-domain collaborative filtering," IEEE T rans. Image Process., vol. 16, no. 8, pp. 1 –16, 2007.

[9] K. Mohan, "Fundamentals of digital image processing," Prentice-Hall, vol. 5, no.8, pp. 661-667, 2004.

[10] SendurandI.W.Selesnick,"Bivariateshrinkagefunctionsforwaveletbas ed denoising exploiting interscale dependency," IEEE T rans. Signal Proc., vol. 50, no. 11, pp. 2744 –2756, 2002.

[11] L.David andM.Iain, "Adapting to unknown smoothness via wavelet shrinkage", Journal of the American Statistical Association, vol.90, no.432, pp.1200-1224, 2002.

[12] L. Sendur and I. W. Selesnick, " Bivariate shrinkage with local variance estimation," IEEE Signal Process. Lett., vol. 9, no. 12, pp. 438–441, 2002.

[13] L. Yan and L. Wang, " Advances in Cognitive Neurodynamics (V)," no.Ii, pp. 659–662, 2016.

[14] M. Haritopoulos, H. Yin, and N. M. Allinson, " Image denoising using self-organizing map-based nonlinear independent component analysis," Neural Networks, vol. 15, no. 8–9, pp. 1085–1098, 2002.

[15] M. Kaur, K. Sharma, and N. Dhillon, " Image Denoising Using Wavelet Thresholding," Int. J. Eng. Comput.Sci., vol. 2, no. 10, pp. 2932–2935, 2013.

[16] M. Kazubek, " Wavelet domain image de-noising by thresholding and Wiener filtering," Signal Processing Letters IEEE, vol. 10, no. 11, pp. 34-9,2003.

[17] Rajni and Anutam, "Image Denoising T echniques-An Overview," Int. J. Comput. Appl., vol. 86, no. 16, pp. 13 –17, 2014.

[18] R. Ahmadi, J. KangaraniFarahani, F. Sotudeh, A. Zhaleh, and S. Garshasbi, " Survey of image denoising techniques," Life Sci. J., vol. 10, no. 1, pp. 753–755, 2013.

[19] S. shreshtha, " Signal & Image Processing," An International Journal (SIPIJ). vol. 5, no. 4, pp. 11-24, 2014.

[20] S. Ruikar, " Wavelet Based Image Denoising T echnique," (IJACSA) Int ernational Journal of Advanced Computer Science and Applications, vol. 2, no.3, pp. 34 -39, 2011.

[21] S. Kaur and N. Singh, " Image Denoising T echniques: A Review," Int. J. Innov. Res. Comput. Commun.Eng., vol. 2, no. 6, p. 4578 4583, 2014.

[22] S. Dangeti, " Denoising Techniques - A Comparison," vol. 223-227, no. 22, pp. 1–59, 2003.

[23] S. Kumar," Performance Comparison of Median and Wiener Filter in Image De-noising,"International Journal of Computer Applications.Vol. 12, no.4, pp. 88 -93, 2010.

[24] T . L. Sahu, M. T . Scholar, C. Science, and E. Csit, " A Survey on Image Noises and Denoise T echniques," vol. 1, no. 9, pp. 77 –81, 2012.

[25] Laparra, J. Gutierrez, G. Camps-Valls, and J. Malo, " Image Denoising with Kernels Based on Natural Image Relations," J. Mach. Learn. Res., vol. 11, pp. 873–903, 2010.

Segmentation of Speech into Syllables

Samanjeet Kaur
Department of Computer Science and Engineering
Baba Banda Singh Bahadur Engineering College
Fatehgarh Sahib, India
samanjeetkaur@gmail.com

Puneet M ittal
Department of Computer Science and Engineering
Baba Banda Singh Bahadur Engineering College
Fatehgarh Sahib, India
puneet.mittal@bbsbec.ac.in

***Abstract* - S peech technology consists of many fields such as speech synthesis, speech recognition, speaker recognition, speech compression, speaker verification and multimodal interaction. The basic units of speech are word, syllable and phoneme. As syllable is the middle unit among three basic units. This study mainly focuses on studying the existing syllable segmentation techniques and compares the techniques on the basis of different categories of syllable. Basically two techniques of syllabification were studied. First technique is syllable segmentation using sinusoidal modeling and second is syllabification by extracting formant frequencies. In experimentation syllabification using sinusoidal mode ling technique is implemented. S yllable segmentation algorithm is applied and tested on a set of recorded words. The continuous speech signal is divided into segments. Implemented technique concludes that syllabification using sinusoidal modeling technique gives accurate results for two categories of syllable.**
Keywords—syllable, phoneme, sinusoidal modeling.

I. INT RODUCT ION

A syllable is one unit of sound. Syllables have long been regarded as units of speech perception and recognition. A syllable contains a vowel with consonants in the initial and the final margins. Automatic segmentation and labelling of speech at the phonetic level is not very accurate while syllable boundaries are more precise and well defined. In a speech recognition frame work, although the syllable as a basic acoustic unit suffers from the problem of tra ining data sparsely, techniques to improve recognizer performance even with sma ll a mount training data with longer duration units like the syllable e xist. Syllables are formed by adding vowels and consonants together. For e xa mple, the word *water* is composed of two syllables: *wa* and *ter*. A syllable is typically made up of a syllable nucleus (most often a vowel). Word that consists of a single syllable is called a monosyllable (and is said to be monosyllabic). Similar terms include disyllable (and disyllabic; also bisyllable and bisyllabic) for a word of two syllables; trisyllable (and trisyllabic) for a word of three syllables; and polysyllable (and polysyllabic), wh ich may refer e ither to a word of more than three syllables or to any word of more than one syllable.

Syllable Seg mentation is the ability to identify how many syllables (or parts) there are in a word. Segmenting spoken words into syllables is the most simple phonological awareness skill. It involves breaking words into chunks or beats and each beat will contain a vowel sound e.g. the word big contains one syllable. One of the main reason for choosing the syllable as the basic unit is that many languages are syllable-centric. Another ma jor reason for considering syllable as a basic unit for automatic speech recognition systems is its better representational and durational stability relative to the phoneme. The syllable was proposed as a unit for ASR as early as , in which irregularities in phonetic manifestations of phonemes were d iscussed. It was argued that the syllable will serve as the effective min ima l unit in the time-do main. Since then, several ASR systems have been developed for different languages, most recently for Indian languages. In those researches, it is demonstrated that segmentation at syllable-like units followed by isolated style recognition of continuous speech performs well.

A. Types of Syllables

- *Open Syllable***:** If a syllab le is open, it will end with a long vowe l sound spelled with one vowe l letter; there will be no consonant to close it and protect the vowel. For e xa mp le He , She, me , recent, protect, program, table, total.
- *Closed Syllable***:** The closed syllable is the most common spelling unit in English. When the vowel of a syllable is short, the syllable will be c losed off by one or more consonants. Therefore, if a closed syllable is connected to another syllable that begins with a consonant, two consonant letters will co me between the syllables. For e xa mp le Shot, strut, hostel, despite,common.
- *Silent-e Syllable*: These syllables end with an ‘e ’ and have another vowel. Also known as "magic **e**" syllable patterns, VC**e** syllables contain long vowels spelled with a single letter, followed by a single consonant, and a silent e. For e xa mple . Cave, home, rope, hope.
- *Vowel Team Syllable***:** A vowel team may be two, three, or four letters; thus, the term vowel digraph is not used. A vowel team can represent a long, short, or diphthong vowel sound. For exa mp le tail, stay, snow, lie, trainer, suit-able.
- *R-controlled Syllables***:** We have chosen the term "vowel-**r**" over "**r**-controlled" because the sequence of letters in this type of syllable is a vowel followed by **r (er, ir, ur, ar, or).** Fo r e xa mple Bird, turn, cu rl, herd, word, per-form, further.
- *Consonant +le Syllables***:** Also known as the *stable final syllable*, C-le co mbinations are found only at the ends of words. If a C-le syllable is co mbined with an open syllable. For e xa mp le Table , title , cradle, purple, cable

II. EXISTING TECHNIQUES

Over the last few years’ number o f researchers has been working to develop more accurate and reliable Speech Recognition and Synthesis systems. In many o f the researches in speech recognition the syllables are considered as the basic units since it contains more te mpora l informat ion than the phonemes. The syllable acts as a bridge between the lower level and the higher level representational tiers of language.

Existing technique [1] present a simple algorith m for speech syllabification. It is based on the detection of the most re levant energy ma ximu ms, using two different energy calculations: the former fro m the orig inal signal, the latter fro ma low-pass filtered version. The systemrequires setting appropriate values for a number o f para meter. The procedure to assign a proper value to each one is reduced to the minimizat ion of a variable function, for which they use either a genetic a lgorith m or simulated annealing. Different estimation of para meters for both Ita lian and Eng lish was carried out. They found the English setting was also suitable for Italian but not the reverse.

Many syllable segmentation approaches are based on minimu m phase group delay. T. Nagarajan et a l. [2], proposed a segmentation algorith m which splits the speech signal in to syllable -like units. The segmentation algorith m is based on the min imu m phase signal derived fro m the short-term energy. R. Janakira m et a l.[3], developed an syllable segmentation algorithm based on the Group Delay (GD) and vowel onset point detection.

A.Harma [4] presented a technique for syllabification they apply sinusoidal modelling to syllables of continuous speech and use obtained parameters fo r recognition of a nu mber of song bird species. This technique is purposed here as a baseline technique for bird sound identification and therefore the goal is to evaluate how well this fairly simple and low co mp le xity approach without any intelligent or conte xt-a ware p rocessing works for a nu mber of species. In this article we studied automatic sound -based identification of bird species based on syllable segmentation. They started with a hypothesis that identification of species could be done by comparing sinusoidal representations isolated syllables of bird song. Possibility to identify species on the basis of isolated syllables instead of significantly longer song patterns would be beneficial for many reasons.

A.V.Natarajan *et al*,[5] p resented a technique for the recognition of Ta mil speech by syllable segmentation. This study main ly focuses on syllable segmentation or syllabification with the aim to further develop a speech synthesis tool in Ta mil language for Hu man Co mputer Interaction. The syllable boundaries are identified using the formant frequency, F1. The proposed syllable seg mentation algorith m is applied and tested on a set of recorded continuous speech corpus. Initia lly, the continuou s speech signal is divided into segments by removing the silence regions. The silence re moval method used in this work depends on features such as signal energy and spectral centroid. After re mov ing silence portion fro m the speech signals, the speech segments are further processed using Linear Predict ive Coding (LPC) to e xtract the formant frequencies. Then the peaks in the formant frequencies are used as clue to mark the syllable boundaries in the speech. This algorith m produced an average accuracy of 89% in identifying syllab le boundaries when it is co mpared with the hand labeled syllable boundaries. When a sylla ble database is constructed using the algorithm, it can be used for developing both speech recognition system and synthesis system.

III. METHODOLOGY

The basic methodology of this technique consists of three steps. First step is to collect data, second step to apply the algorith m to the collected data and obtain results. Three steps are as follows.

Step 1 Data collection: 6 syllable categories are bein g considered in this work. 5 words under each category are identified and speech recordings will be done fo r a ll 30 words. Recordings will be done in clean and noise free environment. Noisy environment will give adverse effect to the syllable segmentation algorith m. Recordings should be in wav form. The imple mented algorith m e xtracts wav signals only.

Step 2 Sinusoidal Modeling: Apply sinusoidal modeling to each recording to segment it into syllables. Means pass each recording in the sinusoidal modeling alg orith m fo r syllabification. It will segment the word in suitable syllables that will be compared with handballed syllables.

The basic step under sinusoidal modeling is to decompose the speech recording into frequency and amp litude modulated sinusoidal puls es. Each pulse will represent one individual syllab le so that the syllables are not overlapping in time o r frequency. In sinusoidal model short time Fourier transform will be used to compute the spectrogram of a speech recording. The decomposition of a recording into N syllables will work as follows:

Algorithm

1. Co mpute a spectrogramof a song segment using FFT. We denote a spectrogram a matrix S (f,t), where f represents frequency and t is time.

2. Repeat steps 3-7 for n=1,2,3……N-1.

3. Find fn and tn such that |S(fn,tn)| is the ma ximu m value in the spectrogram. Th is position represents the ma ximu m amplitude position of nth sinusoidal syllable.

4. Store frequency parameter w(n)=fn and amplitude an(0)= 20log10|S(fn,tn)| dB.

5. Start ing fro m |S(fn, tn)|, trace the ma ximu m peak of S(f,t) For t>t 0 and for t <t0 Until an(t−t0)<an(0)−Td B, we re the stopping criteria T is typically 30 dB. It will determine how the sinusoidal syllable starts and ends at times ts and te respectively around the amplitude maximu m t0.

6. Store obtained frequency and amplitude trajectories corresponding to the nth syllable in functions wn (T) and an(T), where T=t0-ts…….t0+te.

7. Set S(f, [ts, ts + 1, · · · , te]) = 0 to delete the area of nth syllable.

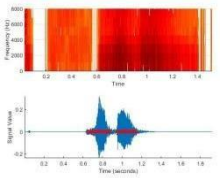
Step 3 Accuracy Detection: After imp le menting the algorith m on each speech recording, accuracy of each category will be calculated using the formula given. This will identify the category of syllab le for wh ich imple mented algorithm performs efficiently.

Accuracy:

$$\frac{\textit{Number of words correctly detected by this algorithm}}{\textit{Total Number of words}}$$

IV. IMPLEMENT ATION

In this study we limit the test a group of 5 words for each category of syllable. To evaluate the performance o f imple mented algorith m random words database was considered. Init ially the algorith m was tested for two random e xa mp les that are water and table. For these two words the output was 2 syllables detected by the implemented algorithm. The results are as under.



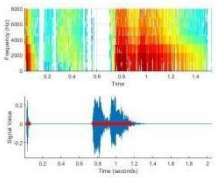


Figure 1 syllables of water Figure 2 syllables of table

According to the first step of methodology Data collection was done on the basis of studied categories of syllable. Random words were collected fro m English dictionary. There are six categories of syllable , for each category 5 words are listed with different number of syllab les. Then these words were recorded in the .WA V format. The WA V file has several major advantages when it comes to professional, high fide lity recording applicat ions. It is an accurate, lossless format; this means that the format reproduced the recording accurately without losing audio quality due to the format itself. It is a very simp le fo rmat – as a result of the files simplic ity, files are re latively easy to process and edit. Unlike WA Vfiles, MP3s are a loss format. This means that encoding audio to MP3 will reduce its quality, but also reduce its file size. All the recordings were done in noise free environ ment because noisy recordings will affect the performance of the implemented algorithm.

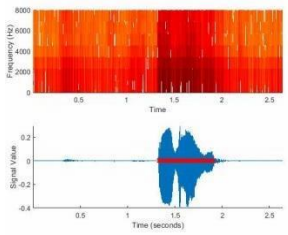
The proposed algorithm was imple mented in MATLAB and run on the entire set of collected data. The algorith m outputs a set of detected syllable peaks for each recorded input. Sinusoidal modelling a lgorith m is applied to each word recording. For each input recording imple mented algorith mproduce a frequency time graph and a signal valu e graph that shows the number of syllables for each word in the form of highest peaks.

In the third step the results were recorded for each input. A table was constructed that compares the number of syllables detected by the algorithm with the nu mber syllables hand labelled. Produced results and table of comparison shown in the ne xt section. Fina lly the accuracy of each category of syllable was calcu lated using the comparison table. First category open syllable identifies syllables of 4 words correctly fro m the 5 words and it has given 80%. Second Category closed syllable detects 4 words correctly fro m 5 words which gives same accuracy as Open syllable category. Third Category Silent-e syllable detects syllables of a ll 5 wo rds correctly, so imp le mented algorith m can be considered accurate for this category. Ne xt category Vo we l Tea m syllable detects syllables of 4 words correctly produces same accuracy as open syllable category. Ne xt category R-controlled, for this category the algorith m detects all syllables correctly for all 5 words. For last category words ending with ‘le ’, for this category the algorith m detects syllables of 4 words correctly out of 5 words. So we can conclude that the imple mented algorith m performs better for R-controlled words and Silent-e words than other categories.

V. RESULT S

Imple mented technique for syllabificat ion of speech produced following results. Results are shown category wise. Each Figure is the output of a word.

A. *Open Syllables*



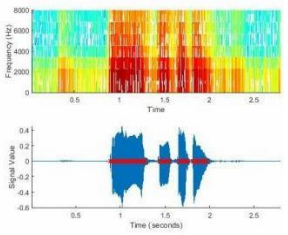
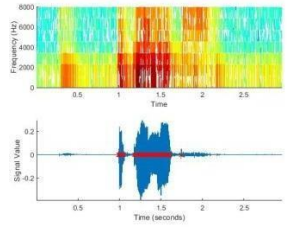


Figure 3.Eve-ning Figure 4. Elasticity



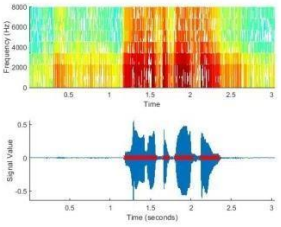


Figure 5.Equinox Figure 6. Verification

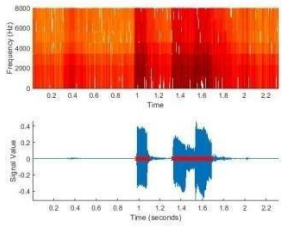
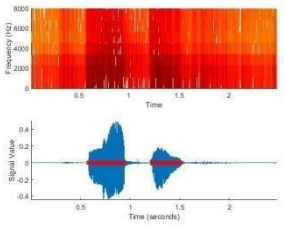


Figure 7. October

B. *Closed Syllables*



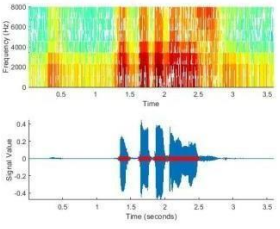


Figure 8.Napkin Figure 9. Hippopotamus

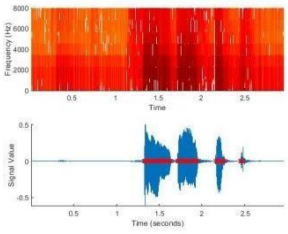


Figure 10.Fantastic

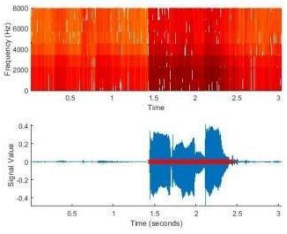


Figure 11. T ornado

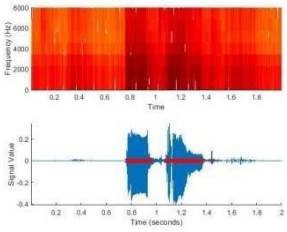


Figure 12. Absent

C. *Silente Syllables*

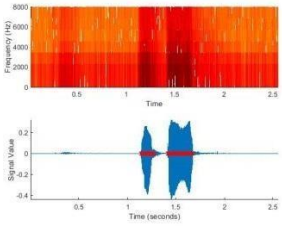


Figure 13.Mistake

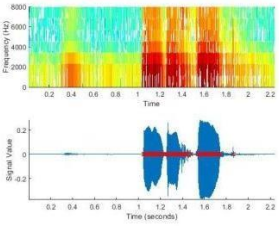


Figure 14. Interstate

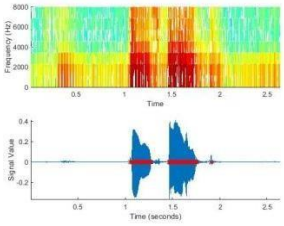


Figure 15.In-spite

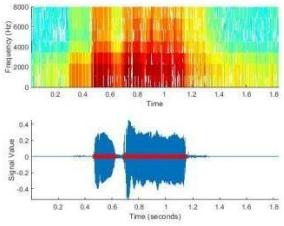


Figure 16. Enquire

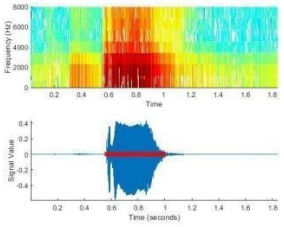


Figure 17. Craze

D. *Vowel Team Syllable*

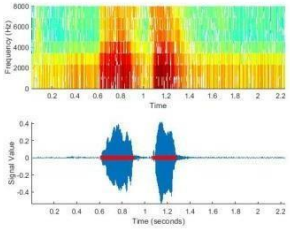


Figure 18.Moisture

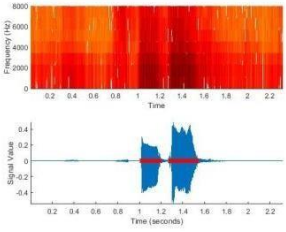


Figure 19. Speaker

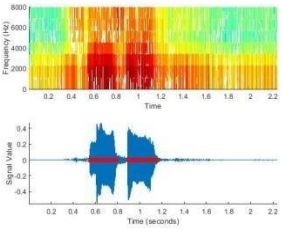


Figure 20.Snowfall

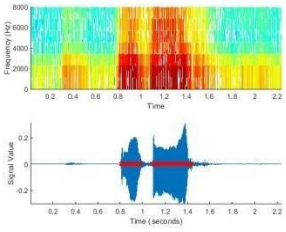


Figure 21. Receive

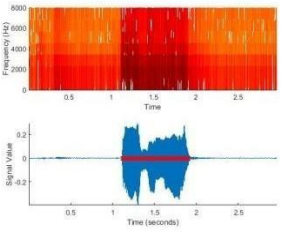


Figure 22. Avenue

E. *R-Controlled Syllable*

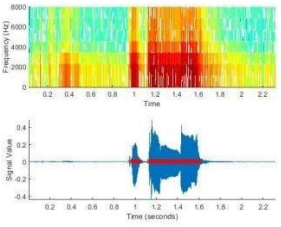


Figure 23.CuCumber

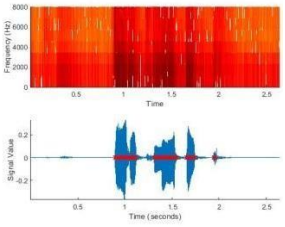


Figure 24. Arithmetic

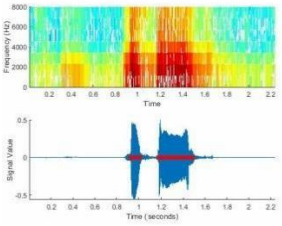


Figure 25.Dis-turb

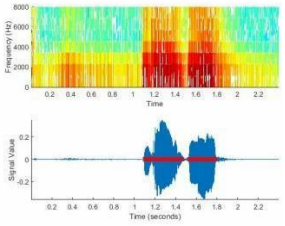


Figure 26. Gardner

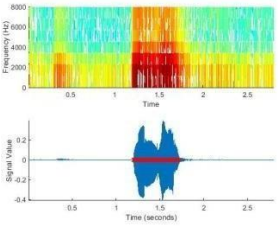


Figure 27. Mineral

F. *Le-Ending Syllable*

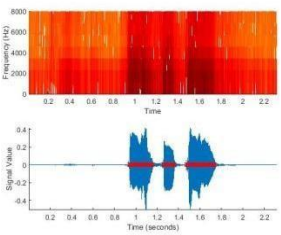


Figure 28.Bicycle

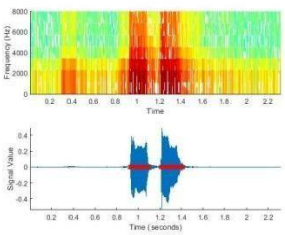


Figure 29. Bat tle

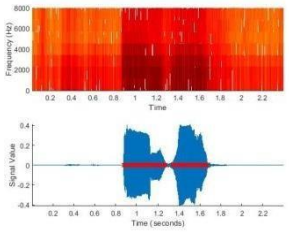


Figure 30.Armhole

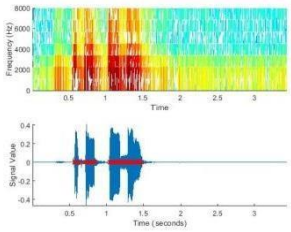


Figure 31. Acceptable

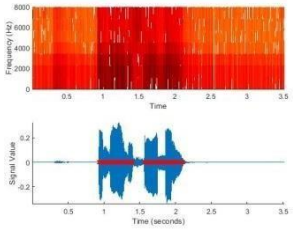


Figure 33. Abolishable

VI DISCUSSION

A technique for syllable segmentation of continuous speech signal using sinusoidal modelling is imple mented. Once the peaks are identified the speech is segmented in to various smaller units to constitute a syllable database. The syllab le database can be used for both speech recognition and speech synthesis.. A comparison table is shown below which compares the number of syllables that are hand labelled with the syllables detected by the implemented algorithm.

C ate gory	**W ord**	**No of Hand labe le d Syllable s**	**No of Syllables De te cte d**
Open Syllable	Equinox	2	2
	Evening	2	2
	Elasticity	5	4
	Verification	5	5
	October	3	3
Closed Syllable	Absent	2	2
	Fantastic	3	3
	Napkin	2	2
	Hippopotamus	5	4
	T ornado	3	3
Silent -e Syllable	Interstate	3	3
	Mistake	2	2
	Craze	1	1
	In-spite	2	2
	Enquire	2	2
Vowel T eam Syllable	Avenue	3	2
	Speaker	2	2
	Receive	2	2
	Moisture	2	2
	Snow-fall	2	2
R-controlled Syllable	Gardener	3	3
	Mineral	3	3
	Disturb	2	2
	Cucumber	3	3
	Arithmetic	4	4
Le- Ending syllable	Abolishable	5	4
	Accept able	4	4
	Arm-hole	2	2
	Battle	2	2
	Bicycle	3	3

T able 1.Comparison of syllables detected

Category	**Accuracy**
Open Syllable	80%
Closed Syllable	80%
Silent -e Syllable	100%
Vowel team Syllable	80%
R-controlled Syllable	100%
Le-ending Syllable	80%

T able 2. Category wise accuracy

REFERENCES

[1] M. Petrillo and F. Cutugno , Francesco"A Syllable segmentation algorithm for English and Italian" in Eurospeech –Geneva *European Conference on Speech Com m unication and Technology* 2003.

[2] Nagarajan T , Murthy HA, Hegde RM. Segementation of speech in to syllable like units.*EuroSpeech*. Geneva. 2003.

[3] Janakiram R, Kumar CJ, Murthy HA. Robust syllable segmentation its application to syllable centric cont inuous speech recognition. Proceedings of National Conference on Communications; Chennai, India. 2010. p. 276–80.

[4] A.Harma, " Automatic recognition of bird species based on sinusoidal modeling of syllables" in *IEEE Int. Conf. Acoustic, Speech, Signal Processing* (ICASSP 2003).

[5] V. Anantha Natarajan and S. Jothilakshmi " Segmentation of Continuous T amil Speech into Syllable like units " in *Indian Journal of Science and Technology*, August 2015.

A study of Deep Learning Architectures for object detection

Payal
Research Scholar
UIET, Panjab University, Chandigarh
payalmittal6792@gmail.com

Akashdeep
Assistant Professor
UIET, Panjab University, Chandigarh
akashdeep@pu.ac.in

***Abstract*-Deep learning based methods have improved the state-of-the-art in object classification, object detection, semantic segmentation and many other domains such as early stage medical diagnosis. It discovers complex structure in large training inputs by using the backpropagation algorithm to indicate how a machine automation should adapt its internal parameters. Deep learning is composed of multiple convolutional layers to understand representations of data with multiple abstraction levels. Deep convolutional nets have emerged about innovative breakthroughs in images and videos processing. In this paper, the main focus is to study various deep learning architectures for object detection. Further, a brief idea about CNNs, software required and performance parameters have also been discussed.**

Keywords- Object detection, deep learning, image, CNN

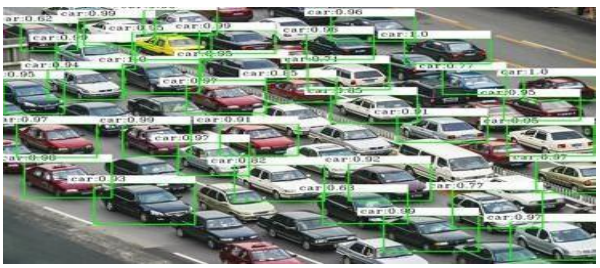
I. INTRODUCTION

Image processing in artificial intelligence has shown recent advancement in computing world [1]. Image analysis can be of many aspects namely, classification i.e. dividing the image into classes and finding class probabilities in which input is image and output is corresponding class label, object localization - recognition of an object by finding coordinates of a square type box called bounding box represented by (x, y, w, h, c), i.e. (x, y) depicts top left coordinate/centercoordinate, w stands for width, h for height and c represents class in which object belongs), object detection that represents identifying and localizing multiple objects as shown in Figure.1, and hence object segmentation that means creating pixel wise boundary of detected object. The primary objective is to study deep learning architectures for object detection. The next sections cover architecture of CNN, hardware and software required, existing CNN architectures and performance parameters used in deep learning.

II. ARCHITECTURE OF CNN

Convolution Neural Networks (CNN), a result of ILSVRC Challenge winner in 2012 [4], provides a great help in solving object detection tasks. The working methodology of CNN is shown in Figure. 2, consist of multiple layers to form class scores like convolution layer, pool layer, dense layer etc. Convolution layer takes an input in the form of raw pixel values of the image with a corresponding channel (1 for grayscale and 3 for red, green, blue) that performs dot product between image region and multiple filter or kernels to detect edges in the image, results in input volume of size say 224*224*12(use 12 filters). Further, activation function such as Rectified Linear Unit (Relu) provides elementwise activation function in non-linear way by applying max (0, x), thresholding at 0, to output of convolutional layer followed by pool layer that perform sub sampling operation into spatial dimensions (width and height), resulted in size alteration of image.

a. Cars in traffic as input asinference [2]



b. Green bounding boxes (BB) display the objects detected with label and confidence[3]

Figure. 1. Input and output of object detection method[3]

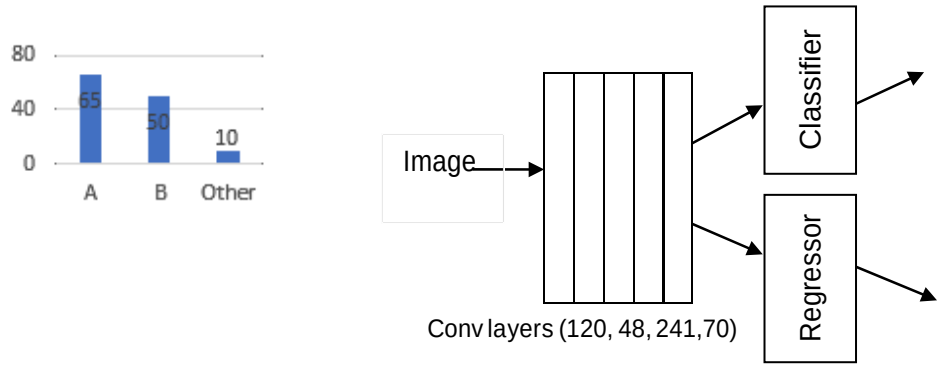


Figure. 2. Working of CNN in object detection [5]

Afterwards, final convolution feature map is sent to fully connected layers that perform one to one correspondence to all neurons in input volume to class neuron layer, then compute class scores by applying classifier like softmax, SVM etc. In real time scenarios, it is not always poss ible to have Image Net [6], a large image dataset to apply deep learning based object detection algorithm so in that case, we can use the transfer learning concept [7], in which input images are feed into pre-trained models like Alex Net, VGG16 and extract features from top or multiple network layers. After extracting features from network layers, apply classification algorithms such as linear classifier and SVM to predict the output. In classification of pre-trained networks, one can choose out of two techniques, one is image classification using feature extraction depicted in Figure. 3(a) in which a pre-trained model is used and freeze its layers means make layers non-trainable except top layers and put dataset on top of network to extract image features using classification algorithm. In second technique, image classification using fine-tuning shown in Figure.3(b), in that case, training only weights of last few layers and freeze the weights of the filters from upper layers. By freezing few layers of pre-trained model, but fine-tune or adjusting them to get high accuracy.

A. Hardware and Software

In past 2016-2018 emerging deep learning era, there has been sudden rise in deep learning application areas due to availability of GPU enabled computing libraries like CUDA and OpenCL. Fastly accessed GPUs requires fractions of second to process an input as compared to CPU because more execution threads running on parallel computing engines. Open source software packages like CUDA contributing to execute convolutions like neural network processing with the help of efficient GPUs. In solving deep learning problems, it is quite essential to choose adequate DL framework to work with technology because each framework serves different purpose and discovering perfect solution to one's problem is quite a tedious task. In this section, we will discuss most widely used DL frameworks out of dozen tools available in the market. Technology giants like Google, Facebook, Microsoft and Amazon are among the companies that invest heavily in artificial intelligence-based frameworks to solve complex DL research challenges. Some popular frameworks are: Tensor Flow [9], one of best open source library in C++ and python, created by Google, most commonly used deep learning framework by many giants like eBay, twitter, Uber, Deep Mind etc. due to flexible

Figure.3 T echniques of CNN Based T ransfer Learning [8]

architecture. It comes with effective data visualization tool known as TensorBoard for network modelling and visualizing performance; managed by a huge community of experts, technology firms and enthusiasts to manage new enhancements and features with efficient efforts; created mainly for deep learning algorithms and also provides reinforcement learning tools, Caffe[10] is developed by Berkeley Artificial Intelligence Research and supported with interfaces like Python, C, C++ and MATLAB. All pre-trained models on this framework can be hosted from Model Zoo repository; does not support fine granularity layer networks like recurrent neural networks (RNN) and language modelling; poor dealing with text, sound and time-series data, PyTorch [11], a python version of torch and Lua-based DL framework released by Facebook in early 2017; simple framework that offers high speed and flexibility so it is being considered a competitor to TensorFlow; customized GPUallocator that makes models handling quite easy; due to its developer friendly nature, Facebook, Nvidia like big companies use it to train their DL models and Keras[12], a neural network library, written in Python, provides a simplistic interface to build and run models of neural networks; supports CNN, RNN on either of TensorFlow or Theano framework; works in a straight forward manner by stacking multiple layers in a model that's why it is a part of TensorFlow's core API.

B. Existing CNN architectures for image processing

An advanced architecture used in big image recognition challenges like ImageNet used to solve a specific problem of image recognition. All existing architectures vary fromeach other in performance but fall in category of deep models. Most popular architectures have been described below:

1) *LeNet-5*, a breakthrough CNN by [13]coined in 1998, used to recognize hand-written numbers. It has ability to process higher resolution images with more convolutional layers around 60k parameters, but this architecture is constrained by the resource's availability.
2) *AlexNet,* scored 1st position in the ILSVRC 2012 [5], consisted of multiple convolutions and several add-ons such as max pooling, dropout, data augmentation, ReLU activations and Stochastic Gradient Descent (SGD) with momentum. Alex Net trained 60 million parameters that makes it complex architecture.
3) *ZFNet,* as almost same as AlexNet. The only difference is by tweaking the hyper-parameters of Alex Net with top-5 error rate of 14.8% [14].
4) *Inception***,** a CNN inspired by LeNet-5, achieved top-5 error rate of 6.7%. Additionally, it used batch normalization, and Root Mean Square (RMS) propagation [15]. Their architecture consisted of 22 deep layers but the number of parameters reduced from 60 million (Alex Net) to 4 million in this architecture.
5) V*GGNet* has very uniformarchitecture and consists of 16 convolutional layers. Most usedarchitecture for extracting features from images [16]. It consists of 138 million parameters and top-5 error rate of 7.4%.
6) *ResNet*, the residual neural networkscored 1st position in the ILSVRC 2015[17], introduced a novel architecture with "skip connections" and heavy batch normalization. This architecture is less complex than VGGNet and achieves a top-5 error rate of 3.57% that beats human-level performance.

C. Performance parameters used in Deep Learning

There are some important evaluation metrics used in deep learning methods[18]such as classification accuracy i.e. the percentage of true predictions achieved from matching of class for highest probability with target output class; Precision is the fraction of true positives (TP) from the total predicted results. Recall is the fraction of TP from the total TP and false negatives (FN); F-1 Scoreis the harmonic means of precision and recall; Intersection Over Union (IOU) is a metric that measures predicted BB, by dividing the area of overlap between the predicted and the ground bounding boxes coordinates. There exist different pooling layers such as: maxpool (most widely used pooling technique, that considers max value from theselected image subset); averagepool (that averages the neighborhood pixels); deformation [19] (extract deformable properties producing geometric constraints of the objects); Spatial Pyramid pooling[20](performs down-sampling of the image); and scale dependent pool layer [21] (which handles scale variation in object recognition).

III. SUMMARY

In summary, CNNs are especially useful for image classification and recognition. CNNs make use of pre-processing techniques that is very easy to use in deep learning frameworks . The deep network learns the weights automatically that in machine learning algorithms were handcrafted. This prior knowledge independence and minimum human effort is a major advantage in deep learning.The flexibility that deep networks provides makes it powerful in building an application oriented end-to-end model. In object detection, these deep network architectures proved successful in various application areas .

IV. CONCLUSION

Major progress in artificial intelligence comesfrom deep network systems that mergecomputational learning with innovative paradigm.While the other network architectures were consisted of stacked

Convolutional layers but modern deep architectures construct convolutional layers such that it allows for more efficient learning. Almost all deep architectures are based on a conv-pool-fully connected subset of layers that repeats throughout the network. These architectures serve as guidelines for machine learning readers to solve various computer vision applicatio ns. These architectures serve as rich classifiers that can be used for image classification, object detection, image segmentation, and many other more advanced tasks.

REFERENCES

[1] Z.-Q. Zhao, P. Zheng, S.Xu, and X. Wu, " Object detection with deep learning: A review," arXiv Prepr. arXiv1807.05511, 2018.

[2] " Objects as an Input." [Online]. Available: https://software.intel.com/sites/default/files/managed/29/bc/object -detection-drone-videos-Figure3.png. [Accessed: 04-Oct - 2018].

[3] " Detected Objects with Bounding Boxes." [Online]. Available: https://software.intel.com/sites/default/files/managed/6a/5f/object-detection-drone-videos-Figure4.png. [Accessed: 04-Oct - 2018].

[4] "CNN Classification." [Online]. Available: https://medium.com/@sidereal/cnns-architectures-lenet - alexnet -vgg-googlenet -resnet -and-more-666091488df5.

[5] A. Krizhevsky, I. Sutskever, and G. E. Hinton, " Imagenet classification with deep convolutional neural networks," in Advances in neural informationprocessingsystems, 2012, pp. 1097–1105.

[6] O. Russakovsky et al., " Imagenet large scale visual recognition challenge," Int. J. Comput. Vis., vol. 115, no. 3, pp. 211–252, 2015.

[7] A. Ahmed, K. Yu, W. Xu, Y. Gong, and E. Xing, " T raining hierarchical feed-forward visual recognition models using transfer learningfrom pseudo-tasks," in EuropeanConference on Computer Vision, 2008, pp. 69 –82.

[8] A. Ahmed et al., " Learningand transferring mid-level image representations using convolutional neural networks," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2008, pp. 69–82.

[9] M. Abadi et al., " Tensorflow: a system for large-scale machine learning.," in OSDI, 2016, vol. 16, pp. 265 –283.

[10] Y. Jia et al., "Caffe: Convolutional architecture for fast feature embedding," in Proceedings of the 22nd ACM int ernational conference on Multimedia, 2014, pp. 675 –678.

[11] R. Collobert, K. Kavukcuoglu, and C. Farabet, " T orch7: A matlab-like environment for machine learning," in BigLearn, NIPS workshop, 2011, no. EPFL-CONF-192376.

[12] F. Chollet and others, " Keras (2015)." 2 017.

[13] Y. LeCun and others, " LeNet -5, convolutional neural networks," URL http//yann. lecun. com/exdb/lenet, p. 20, 2015.

[14] M. D. Zeiler and R. Fergus, " Visualizing and understanding convolutional networks," in European conference on computer vision, 2014, pp. 818–833.

[15] K. Chatfield et al., "Return of the devil in the details: Delving deep into convolutional nets," Proc. IEEE Conf. Comput. Vis. pattern Recognit., pp. 1–9, 2015.

[16] K. Simonyan and A. Zisserman, " Very deep convolutional networks for large-scale image recognition," arXiv Prepr. arXiv1409.1556, 2014.

[17] K. He, X. Zhang, S.Ren, and J. Sun, " Deep residual learning for image recognition," in Proceedingsof the IEEE conference on computer vision and pattern recognition, 2016, pp. 770 – 778.

[18] D. PIET ROW and J. MAT USZEWSKI, " Objects Detection and Recognition System Using Artificial Neural Networks and Drones," J. Electr. Eng., vol. 6, pp. 46 –51, 2018.

[19] W. Ouyang et al., " Deepid-net: Deformable deep convolutional neural networks for object detection," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 2403 –2412.

[20] W. Ouyang et al., " Spatial pyramid pooling in deep convolutional networks for visual recognition," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 346–361.

[21] W. Ouyang et al., " Exploit all the layers: Fast and accurate cnn object detector with scale dependent pooling and cascaded rejection classifiers," in Proceedings of the IEEE Conference on Computer Vision and PatternRecognition, 2014, pp. 346–361.

Fruit Image Enhancement method based on Bacterial Foraging Optimization using Fuzzy Logic

Harmandeep Singh Gill,
Research Scholar
Punjab Technical University
Jalandhar

Dr.Baljit Singh Khehra
Professor & Head Departm ent of CSE
Baba Banda Singh Bahadur Engineering College
Fatehgarh Sahib

***Abstract*---In orde r to e fficie ntly e nhance the contrast of the fruit image s, a ne w image e nhance me nt me thod base d on Bacte ria Foraging Optimiz ation (BFO) using fuzzy logic is proposed. Initially Image s are first conve rte d into fuz z ification the n use the me mbe rship function to me asure the inte nsity value s andfinallythe de fuz z ification is pe rformed to transform the image into its original form. This me thod is fast and provide s be tte r re sults as compare d to othe r existing nature inspire d optimize d base d image e nhance me nt te chnique s such as GA,ACO, PS O,ABC, Cuckoo S e arch e tc.**

Keywords:--- Image Enhancement, BFO, Fuzzy, Fuzzification, Membership Function, De-fuzzificatio n

I. INT RODUCT ION

Agriculture is very important part of many developing countries [1]. Sorting fruits manually is very time consuming and huge labour costs [12]. Automation in fruit industry comes into play to increase productivity and economic growth of the developing countries like India. For sorting the fruits using computer vision system, it is important to detect the shape of fruit image has to be extracted [12]. Computervision strategies used to recognize a fruit based on the four main features which recognize the object intensity, colour, shape and texture [25][26][27][28][29]. During the image capturing there is a chance of noise additively, it is necessary to enhance the fruits images. Image enhancement is used to improve the image quality so that the resulted image is better than the original image. Enhancement is used to improve the interpretability or perception of information in images for image processing techniques. Image enhancement is very important phase of image processing system. An enhancement algorithmthat offers a better image quality for specific purposes. Image enhancement includes the operation that aims at change the photometric or structural characteristics of an image.

Many algorithms have been developed in the recent times for image contrast enhancement. HistogramEqualization [HE] is very popular image contrast enhancement technique, widely used in medical, agriculture, weather forecasting, X-ray, MRI etc. Image HE is exploited to increase the visibility of image contents by altering the intensity value [3]. This method is not suitable for low contrast images. CLAHE operates on small regions in the image called tiles rather than the whole image. Contrast of each tile is enhanced that the histogram of the output region approximately matches the histogram specified by the distribution parameters. Distribution string specified the described histogramshape for the image tiles. The neighbouring tiles are then combined togeth er using bilinear interpolation to eliminate artificially induced boundaries. [3] retinal images enhanced by genetic algorithm for both qualitative and quantitative performance using PSNR,AMBE and RMSE.[4] enhanced the satellite image using DWT-SVD by cuckoo search algorithm for improvement of the low contrast images and measure the performance in terms of PSNR, MSE and RMSE.[3] enhanced the fruit image by gray transformand wavelet neural network to obtain non-linear gray transform curve. Transformparameters are determined by different contrast type of input image. Changing [6] enhance the fruit images by gray transform and wavelet neural network by employing a new algorithmbased on incomplete beta transform(IBT) to obtain non-linear gray transformcurve. In their method they supposed that HE will improve the contrast of the fruit image but they do not show any results and information which needs further work and research. Alex et.al, proposed a method to enhance images based on a new artificial life model, inspired by the behaviour of a herbivore organism. This organismis in a certain environment and selects its food by travelling iteratively. This method is used for image enhancement in the spatial domain to find the intensity values and difference between the neighbours howeverthis model has some limitations that howwe can find the optimized neighbour values to evaluate the best one. The classical image processing techniques process the crisp histograms ofthe images which donot Consider the exactness of the gray-values[23]. Also the vagueness and uncertainty are not exactly measured with these existing techniques. Fuzzy logic techniques are applied to overcome these limitations. Evolutionary methods PSO, BFO and ACO used by Hanmandhu et.al, in their different research work on image enhancement and compared with Genetic Algorithm (G.A) and produced results are better.

In the fuzzy image processing, before the image processing, the image is transformed fromspatial domain to the fuzzy domain using fuzzy membership functions. The membership values are modified to enhance the contrast of the image and finally de-fuzzification will convert the enhanced image to the spatial domain.

The quality of an image captured by a camera is sometimes low due to the distortion of camera optics system, the relative motion of the photographed object and the camera. Image enhancement is an important technique used to improve the quality of the degraded image and restore the image as better image as compared to the original image.

The approach used in [23] is extended by enhance the fruit image by fuzzy logic technique using BFO(Batrerial Foraging Optimization). In this approach, fuzzy technique is used to enhance the image to analyse the image quality and improves the contrast level of the enhanced images by choosing the optimized results produced using BFO.

II. RELAT ED WORK

[16] - [23] enhances the radiographic, medical, Mammography, x-ray angiogram, infrared, microscopic, ultrasound images use fuzzy techniques based on fuzzy set theory, fuzzy IF-THEN rules, fuzzy operators and fuzzy membership functions. [20]enhances the medical images using fuzzy set theory by intuitionist fuzzy set to handle the uncertainty in the form of fuzzy membership function and produced results suggests that the enhanced images are better. Medicalimages are enhanced using the following formula:

$$winA(g) = \frac{g - g_{min}}{g_{max} - g_{min}} \quad (1)$$

Where g is the gray level of that window, ranging from 0 to L-1. gmin, gmax are the minimum and maximum values of the gray intuitionistic fuzzy generator [20].

R.C et.al [19] enhance the contrast by using a contrast enhancement algorithm which maps elements from pixel plane to membership plane and to enhancement plane. For image contrast enhancement using the formula:

$$I(i,j) = \sum_{i=1}^{m}\sum_{j=1}^{n} \frac{([\mu I(i,j)])}{I(i,j)}$$

(2)

[μ I(i,j)/I(i,j)] represents each pixel and μI(i,j) is the degree of intensity level of the image with values between 0 and 1.

Xiwen L et.al [17] used fuzzy enhancement algorithm to eliminate the noise and preserves the image intelligently by dividing the image into close points for close point detection using fuzzy characteristics, can be calculated as follows:

$$P_{ij} = X_{ij} = \left[1 + \frac{(X^{max} - X_{ij})^2}{F_p}\right] - 1$$

(3)

Where Xmax is the largest gray value, Xij is the gray valu e of pixel (i,j). Fp is fuzzy parameter.

When Xij ⟶ Xmax, Pij ⟶ 1: when Xij decreases Pij also decreases. All Pij Consist of Fuzzy characteristic plane.[16]

Farhang S. Et.al[16] enhance the contrast of mammography images using a fuzzy approach. Intensity modification method is used and fuzzification procedure is performed to transform the modified image back into the spatial domain. Madasu h.et al [21] is presented a approach for enhancement of color image using fuzzy logic.

The Image Normalization Equation

$$G(ij) = \frac{g_{min} + (g_{max} - g_{min}) * (go_{min}(i,j) g0_{max} - 1)}{go_{max} - go_{min}}$$

(4)

Where gomin and gomax are the maximum and minimum gray scale values of the original images.[22] while the go(i,j) is the gray scale value of the original images and g(i,j) is the gray scale values of the normalized image at the point(i,j).

III. FUZZY LOGIC

A fuzzy set is a class of objects with a continuum of grades of membership. [24] Fuzzy logic(FL) refers to a logical system that generalizes classical two-valued logic for reasoning under uncertainty. In broad sense, FL refers to all of the theories and technology that employ fuzzy sets, which are classes with un-sharp boundaries. FL was introduced in 1965 by Lofti A. Zadeh, is a mathematical tool for dealing with uncertainty. FL operates on the concept of Membership Function.

The fuzzy image enhancement system can be organized into the following categories[1].

(i). Fuzzificaton

(ii). Membership

(iii). De-fuzzification

Fuzzification is the process of transforming a crisp set to a fuzzy set. Membership Function defines the fuzziness in a fuzzy set irrespective of the elements in the set, which are discrete or continuous. The membership functions are generally represented in the graphicalform. Defuzzification is a mapping process from a space of fuzzy control actions defined over an output universe of discourse into a space of crisp control actions. Image Fuzzification and Enhancement Fruit image is converted into fuzzy domain. Fuzzy set theory is used to handle the uncertainty [30]. Gaussian membership function is used to convert the fruit image into the fuzzy plane.

The standard Gaussian is defined as:

$$g(x,y) = e - \frac{[f(x,y) - L_{max}]^2}{2b^2}$$

(5)

where f(x,y) is the input fruit image, g(x,y) is fuzzy image , Lmax is the maximum gray level of input fruit image and parameter b is bandwidth of the Gaussian function that is calculated from the following equation[30].

$$b = [(t_{opt} - k), (L_{max} - t_{opt})]$$

(6)

The selection of the membership function is very important for the image fuzzification to effectively utilize the fuzzy system, for better results and better information [22].

Formula of Gaussian membership Function are as follows:

$$f(x) = exp\frac{x - 0.5(x-c)^2}{\sigma_2}$$

(7)

Where c is the mean and σ is variance

The notation for fuzzy membership function as follows:

$$X = \frac{\mu_m \mu_n}{g_{mn}}$$

(8)

where μ(g) denotes the degree of compatibility of the g with an appropriate image properties[1]. Transformation of spatial domain into fuzzy domain using the Gaussian membership function by the equation as follows:

$$\mu g(_{mn}) = exp\left[\frac{-(g_{max} - g_{min})^2}{2fh^2}\right]$$

(9)

where fh and gmax are fuzzifiers and the maximum contrast level in the image[1]. By changing the fh values one can obtain the maximum values of fuzzification that can result in the maximum enhanced values of an image.where fh is called a fuzzifier and can be calculated using the following formula:

$$fh^2 = \frac{\sum_{x=0}^{L-1}(x_{max} - x)^4 P(x)}{\sum_{x=0}^{L-1}(x_{max} - x)^2 P(x)}$$

(10)

The MSE and PSNR can be calculated as follows: [13]

$$MSE = \frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}(I(i,j) - Ir(i,j)^2))$$

(11)

$$PSNR = 10log_{10} = \frac{MAX^2}{MSE}$$

(12)

IV. PROPOSED METHOD

An image of size I*J having intensity level mxy in the range [0, A-1] can be considered.

L = U{μ(mxy)}= {μxy/mxy}

Where μxy represents the membership of mxy, with mxy being the intensity at the point(x,y)th pixel.

For clearly understanding the intensity values for enhancement the (V) gray levelarea of the selected fruit image is split into two regions underexposed and overexposed respectively but in the reality, most images are of mixed type. A parameter "exposure" is involved here to detect the percentage of gray level in the image as overexposed or underexposed. The amount of intensity exposition is given by:

$$Exposure = \frac{1}{A}\frac{\sum_{m=1}^{A} p(m).m}{\sum_{m=1}^{A} p(m)} \quad (13)$$

Where m denotes the gray level values in the image, and p(m) represents the histogramof the selected fruit image and A represents the gray level values.

When an image is divided into underexposed and overexposed regions by using the value K. A Gaussian membership function is used to fuzzify the sleceted fruit image as follows:

$$\mu m_{\mu}(m) = exp\left\{ -\left[\frac{m_{max} - (m_{avg} - m)}{\sqrt{2fh}}\right]^2 \right\} \quad (14)$$

where m indicates the region gray level in the selected fruit image. M maxis used to denote the maximumintensity value in the image and m avg is the average gray level value in the selected fruit image. Fh is called a fuzzifier and the initial value for fuzzifier is found from the following equation;

$$fh^2 = \frac{1}{2}\frac{\sum_{m-0}^{A-1}(m_{max})^4 p(m)}{\sum_{m-0}^{A-1}(m_{max} - m)^2 p(m)} \quad (15)$$

For transforming the intensity level of a selected fruit image, from spatial domain to the fuzzy domain, A Triangular Membership Function is derived for the fuzzification of a selected fruit image for m>=k is given by:

$$\mu_0(m) = 0, if 0 < m$$

$$\frac{m-k}{a-k} if m \geq k \quad (16)$$

Finally the modified Membership Function of the Selected region in a fruit imageis transformed back to the spatial domain i.e, defuzzify using the inverse Membership Functions respectively.

For enhancing the Membership values of the original gray level of the selected fruit image is given by:

$$\mu' m_{\mu} m = \frac{1}{1 + e^{-t}(\mu m_m)(m) - \mu_c} \quad (17)$$

Where t is called the intensifierparameter.μc is the crossover point.

For calculating the contrast measures, a set of fuzzy measures are defined for calculation of the objective functions based on contrast and visualfactors. For evaluating the quality of the enhanced fruit image, the entropy and visual factors are used for inspecting and measuring the fruit image quality. The crossover point from the membership values are derived for computing the fuzzy contrasts of the selected fruit image based on the following:

$$Df\mu = \frac{1}{k}\sum_{m=0}^{k-1}(\mu' m\mu(m))^2 \quad (18)$$

The average fuzzy contrast of the selected fruit image is given by:

$$D_k f\mu = \frac{1}{k}\sum_{m=0}^{k-1}(\mu' m\mu(m)) \quad (19)$$

PROPOSED IMAGEENHANCEMENT ALGORITHM

Step1: Convert the color RGB image into HSV format.

Step2: Calculate the histogramh(m) from V component of HSV.

Step3: Calculate fh using

$$fh^2 = \frac{1}{2}\frac{\sum_{m-0}^{A-1}(m_{max})^4 p(m)}{\sum_{m-0}^{A-1}(m_{max} - m)^2 p(m)}$$

Step4: Compute the values of Exposure and K using A from the eq.(13)

Step5: Fuzzify V to get μmμ(m) and μm0(m) using

Step6: Calculate the mofified membership values μmμ(m) and μm0(m) for the selected fruit images.

Step7: Defuzzify the modified membership values values μmμ(m) and μm0(m) using the inverse membership values.

Step8: Enhance the selected fruit image using W(m) = [W(m)] and, then finally display the enhanced fruit image.

BFO (Bacterial Foraging Optimization) :

BFO is proposed in 2002by Kevin Passion, a nature inspired optimization algorithm based on a strategy of a swarm of Escherichia coli. Bacteria search for nutrients also communicates with others by s ending signals . BFO mainly consist of the following foursteps:

- Chemotaxis
- Swarming
- Reproduction
- Elimination/Dispersal
- Using BFO, reviewed in literature, we have experimented based on the selected parameters for fruit image enhancement and noticed the huge saving in computation time with less number of iterations also. Approximately 50 fruit images are selected for experimental work and results produced using BFO is best as compared to other optimized algorithms. Following steps are performed during the enhancement process by taking the standard parameters [21].
- Number of bacteria Nb =12
- Swimming Length Ns = 4
- Number of iterations in a chemotactic loop Nc is set to 25 (Nc>Ns).

- Number of reproduction steps Nre is set to 4.
- Number of elimination and dispersal events Ned is set to 2.
- Probability of dispersal is set to 0.25
- Location of each bacteriumwhich is a function of several parameters, that is, f (Ped,Nb,Nc,Nre,Ned)is specified by a random number in the range[0-1]

V. RESULTS AND DISCUSSIONS

The Proposed approach has been implemented using MATLAB 2013 on intel core i3-4005U 1.7 GHz Processor. Approximately 50 images are considered and results of 5 images are presented here. Seven image enhancement methods have been used. The performance shows that the proposed method gives better results as compared to other enhancement techniques. The results are produced based on MSE, PSNR, BER and RMSE. The proposed method results are better for all the input images.

Figure 1 (a) Input apple image (b) Output apple Image

Figure 2. (a) Input Orange Image (b) Output Orange Image

Figure 3 (a) Input Mango Image (b) Output Mango Image

Figure 4: (a) Input Honeydew Image (b) Output Honeydew Image

Figure 5: (a) Input Guava Image (b) Output Guava Image

T able 1

Values obtained after applying various enhancement techniques on the apple.jpg image

Me thod	ABC	MAP_MBD	C UC KO O	LAB-C LAHE	MIX-C LAHE	conte nt adaptive	BFO	FUZZY ENHANC EMENT
MSE	10.88	3583	3462	4081	26935.44	1.49	7.55	7076.65
PSNR	5776.37	1327.61	4275.74	1210.14	2382.76	6639.72	5934.87	2963.25
BER	1.73	7.53	23.4	8.26	4.20	1.51	1.68	3.37
RMSE	32.99	5530.15	185.64	6330.97	1641.20	12.21	27.49	841.23

T able 2

Values obtained after applying various enhancement techniques on the orange.jpg image

Me thod	ABC	MAP_ MBD	C UC KO O	C LAHE	MIX-C LAHE	C onte nt Adavtive	BFO	FUZZY ENHANC EME NT
MSE	3.45	110.97	609.43	139.35	41528.70	4.16	3.12	33348.08
PSNR	6318.30	1767.86	40.2816	16.6898	21.9473	76.0675	62.7495	22.9001
BER	1.59	5.66	2.48	5.99	4.56	1.71	1.55	4.37
RMSE	17.68	3331.26	246.87	3732.93	2037.86	4.01	3.58	1826.15

T able 3

Values obtained after applying various enhancement techniques on the mango.jpg image

Me thod	ABC	MAP_MBD	C UC KOOO	C LAHE	MIX-C LAHE	C onte nt adaaptive	BFO	FUZZY ENHANC E MENT
MSE	6.94	450.52	354.23	582.99	30260.47	7..65	4.40	32780.18
PSNR	59.7169	11.5937	42.6380	10.4742	63.3220	65.9479	55.8880	22.9747
BER	1.67	8.63	2.35	9.55	4.29	1.52	1.70	4.35
RMSE	26.34	6712.08	188.21	7635.39	1739.55	12.86	18.98	1810.53

T able 4

Values obtained after applying various enhancement techniques on the honeydew.jpg image

Me thod	ABC	MAP_MBD	C UC KO O	C LAHE	MIX-C LAHE	C onte nt Adavtive	BFO	FUZZY ENHANC EME NT
MSE	2.32	433.29	358.86	478.20	11710.33	0.0	00.24	43838.65
PSNR	6447.87	1176.30	4258.16	1133.48	2744.51	9900	7424.80	2171.22
BER	1.55	8.50	2.35	8.82	3.64	1.01	1.35	4.61
RMSE	15.23	658249	189.44	6915.17	1082.14	0	4.94	2093.77

T able 5

Values obtained after applying various enhancement techniques on the guava.jpg image

Me thod	ABC	MAP_MB D	C UC KO O	C LAHE	MIX-C LAHE	C onte nt Adavtive	BFO	FUZZY ENHANC EM ENT
MSE	00.41	65887.46	422.92	104.09	226.1822	7183.55	0.0115	5812.72
PSNR	7201.36	1994.28	4186.82	1795.66	2458.62	1.39	6752.95	3048.70
BER	1.39	5.01	2.39	5.57	4.07	6.53	1.08	3.28
RMSE	6.40	2566.86	205.65	3226.33	1503.94		3.97	

REFERENCES

[1] Farhang Sahba, Anastastios Venetsanopoulos. Contrast Enhancementof Mammography Images using a Fuzzy Approach", 30th Annual International IEEE EMBSConference, Vancouver-BC (Canada)August 20-24.(2008)

[2] G.Raju, Madhu S.Nair (2013) " A fast and efficient color image enhancement method based onfuzzy-logic and histogram"International Journal of Electronics and Communications (AEÜ)

[3] A.K.Bhandari, V.Soni, A.Kumar, G.K.Singh. Cuckoo Search algorithm based satellite image contrast and brightness enhancement usingDWT-SVD" ISA T ransactions (ELSEVIER) (2014)

[4] T rung Pham, Per Bro, Jose Luis T roncoso. Fuzzy Analysis of Color Images of Fruits. IEEE(2010)

[5] Farzan Farbiz, Mohammad Bagher Menhaj, Seyed Ahmad Motamedi (1998) " Fixed point design for Image Enhancement usingFuzzyLogic". IEEE

[6] Wei Jia, Jinyuan Yang, Yanbin Liu, Lilue Fan, Qingqiang Ruan(2014)" Improved Fast Image Enhancement Algorithm Based on Fuzzy Set T heory" 6th IEEE International Conference on intelligent Human-Machine System and Cybernetics.

[7] Apurba Roy, Santi P.Maity, sarat Kumar Yadav (2014)"Onsegmentation of MR images using Curvlet and fuzzy C-Means under Compressed Sensing" . IEEE

[8] S.Khehra, A.P.Singh(2011) " Digital Mammogram Segmentationusing N0n-shanon Measures of Entropy" Proceedings ofWCE-2011,London, U.K. July 6-8

[9] Fabrizio Russo. A Fuzzy Approach to Performance Measurement of Grayscale Image Denoising Algorithms" IEEE

[10] R.Sudhakar. Image Enhancement T echniques for Palm Vein Images" ICACCS, Coimbatoe, INDIA. Januray, 05 -07 (IEEE). (2015)

[11] Amer Draa, Amria Bouaziz. An Artificial bee colony algorithmforimage contrast enhancement" Swarm and EvolutionaryComputation16(2014) 69-84

[12] V.Gurunathan, S.Bharathi. Image Enhancement T echniques for palm Vein Images (ICACCS) Coimbatore, INDIA. Jan 05-07, (2015)

[13] Xiangtao Chen, Yujuan Chen. An Improved Edge Detection in Noisy Image Using Fuzzy Enhancement. 978-1-4244-5316-0/10IEEE.(2010)

[14] Shweta Narnaware, Roshni Khedgaonkar. Image Enhancement using Artificial Neural Network and Fuzzy Logic". IEEE Sponsored 2nd International Conference on Innovations in InformationEmbeddedand Communication Systems ICIIES'2015.

[15] Wang Zaifu. Generalized Fuzzy Enhancement Based Recognizing Method for Planar Objects". Proceedings of the 26th Chinese control Conference. Zhangjijie, Hunan, china. July 26-31 (2007)

[16] Chanjiang Zhang, Xiaodong Wang, Haoran Zhang. Contrast Enhancement for Fruit image by Gray T ransform and Wavelet Neural Network. IEEE.(2006)

[17] Xiwen Liu, Guorong Wang, Yonghua Shi. Application of Fuzzy Enhancement Algorithm and KSW Entropy AlgorithminSegmentation of Pre-welding Seam Image. IEEE, ISDA'06 (2006)

[18] Zhang Xiao-guang, Gao Ding, Xu Jian-Jian. Generalized Fuzzy Enhancement of Image for Radiographic T esting Weld" Proceedingsof the 4th International Symposium on Image and Signal Processing and Analysis. (2005)

[19] Alex F.Araujo, Chritos E.Constaninou, Joao Manuel R.S.T avares.New artificial Life model for image enhancement. Expert systems with Applications 41 (2014) 5892-5906.

[20] T amalika Chaira. Medical Image Enhancement UsingIntuitionisticFuzzy set". 1st International conference on Recent Advances in Information T echnology.(IEEE).(2012)

[21] Madasu Hanmandhu,Om Parkash Verma, Nukala Krishna Kumar, Murlidhar Kulkarni. A Novel Optimal Fuzzy System for Color Image Enhancement Using Bacterial Foraging. IEEE T ransactions On Instruments and Measurement, Vol.58, No.8, August (2009).

[22] Lo T zer Yuan, Sim Kok Swee, T so Chih Ping. Infrared image enhancement using adaptive trilateral cont rast enhancement . Recognition Letters (ELSEVIER) (2015)

[23] M.Hanmandhu, S.Arora, G.Gupta, L.Singh. A novel Fuzzy ColorImage Enhancement using Particle Swarm Optimization.978-1-4799-0192-0/13 (2013)

[24] L.A.Zadeh. Fuzzy Sets. INFORMAT ION AND CONT ROL8, 338-353(1965)

A Brief Summary of Generative Adversarial Networks

Pratik Joshi
B.E Student
UIET Panjab University, Chandigarh
pratikjoshi.work@gmail.com

***Abstract*-This paper summarizes the work done in the field of generative modelling in the recent years using Generative Adversarial Networks(GANs). The paper describes the problem of generative modeling and introduces GANs. It also discusses how Generative Networks work and its training. The working mechanism of GANs is also elaborated. The paper also throws light on applications and limitations of GANs.**

I. INTRODUCTION

This report summarizes the work done in the field of generative modelling in the recent years using Generative Adversarial Networks (GANs). The report is intended as an introduction to the field, it does not aim to provide a comprehensive review on the work done on GANs, thus many excellent papers are not described here.

The report describes (i) The problem of generative modelling (ii) Introduction to GANs (iii) How do GANs work? (iv) Applications of GANs (v) Limitations of GANs .

II. THE GENERATIVE MODELLING PROBLEM

Generative modelling is concerned with estimating the data distribution pdata. The data distribution is usually given to the model in the form of training data. The aim is to minimize the difference between pdata and the estimated distribution pmodel.

Generative modelling is particularly useful is designing simulation environments for reinforcement learning, providing missing data in semi-supervised learning, generating multi-modal outputs for a given scenario, representing high dimensional data distributions and generating realistic samples from the data distribution Generative modelling can be done in an implicit or an explicit manner. In explicit modelling pmodelestimated pdataby acting as function that can be evaluated at any point x to yield an estimate of probability density of pdata. Implicit modelling involves generating data samples from pmodel that aim to represent pdata.

A. Explicit Modelling

Explicit modelling involves designing a model to obtain the estimate of the target distribution using an explicitly defined function and then using gradient based optimization to reduce the difference between the estimated distribution and the target distribution.

While explicit modelling presents more flexibility in the results that are obtained, it is in practice harder to achieve. The main problems associated with explicit modelling are (i) designing a model that can capture all the complexity of data (ii) maintaining computation tractability. Optimizing for one of these problems increases the other problem.

Thus, explicit models are designed in the following manner (i) constructing models that are tractable e.g. Fully Visible Belief Nets (ii) approximation - based models e.g. variational autoencoders and markov chains.

B. Implicit Modelling

Implicit models are trained by sampling data from pmodel and then these samples are analyzed whether pdata or not. No explicitly defined density function is needed. GANs are a form of implicit modelling.

III. INTRODUCTION TO GANs

Introduced in 2014 by Goodfellow Et Al, GANs consists of two networks i.e. the generator and the discriminator respectively. The generator produces samples using some random data distribution. The discriminator judges the generated samples and samples from the original data distribution and classifies them as real and fake. possible. The generator and the discriminator act in an adversarial manner causing GANs to work on the high-level idea of generating images that are indistinguishable from reality.

In GANs the problem of evaluating the sampled images in Implicit modelling is solved using the discriminator. This property has led to the GANs being ubiquitous in image generation tasks, as this process of evaluating the image by the discriminator is simple, highly automated, scalable and is flexible enough to incorporate changes according to problem requirements.

I V. HOW DO GANs WORK?

Generally, GANs have the same cost function for the discriminator i.e. the discriminator minimizes the J-S divergence between pdata and pmodel. The

different GANs differ in the cost function of the generator. The objective of the simplest GAN is mentioned in this section.

The objective of a GAN is given as

$$\mathcal{L}_{gan}(G,D) = E_y[\log D(y)] + E_z[\log(1 - D(G(z))]$$

Where y denotes the real image and z denotes the input from the random distribution used for sampling. G denotes the generator and D denotes the discriminator.

The min-max game between the generator and the discriminator is denoted as:

$$G^* = min_G max_D \mathcal{L}_{gan}(G,D)$$

GANs work on the principle of a min-max game between two adversaries, the generator and the discriminator. The generator generates images that are designed to fool the discriminator thus maximising the error produced by the discriminator. The discriminator has to classify the images correctly thus to minimise it produces. Both the networks can see each other's parameters but cannot modify them. The discriminator and the generator both provide their gradients to each other during training.

Thus, the task of judging the images is automated and the generator can learn how the images are being judged to generate more realistic images and the discriminator can learn how the images are being produced so as to judge them more effectively. This eliminates the need for specialised cost functions and human intervention in the process.

The solution to the game is the Nash Equilibrium which in this case is the point where discriminator classifies each image with 50% probability of being real or fake, thus the discriminator cannot distinguish between images produced by the generator.

V. APPLICATIONS OF GANs

Since their introduction GANs have been applied to many problems spanning different domains.Their ubiquity lies in their generality and simplicity ; GANs perform specially well in tasks related to images, as they can minimize high structural loss using the general cost function. GANs unlike CNNs do not need specialized cost functions to apply to different domains, the general cost function solves mos t of the problems to a well degree. This is due to the implicit nature of GANs as the adversarial nature of the two networks trains them without the need of human intervention of designing specific cost functions each time for new problems.

They are the most ubiquitous solution to image generation tasks. Most GANs are today based on the DCGAN architecture. DCGAN stands for deep convolutional GAN. DCGAN and LAPGAN was the first GAN architectures that were used to generate high resolution images.These GAN architectures are heavily studied and modified to be used for image generation tasks.

GANs can also be used in a conditional setting. Introduced by Mirza Et Al in 2014 Conditional GANs allow the generator output to be conditioned on the information provided by the user. They have been proven to be very useful when constrained outputs are required. Pix2Pix framework (Isola Et Al 2016) improved conditional GANs further for paired data by giving a consistent approach for semantic label translation between different data domains. Cycle GAN architecture extends image translation between different domains in the absence of paired data by using a cycle loss. Stack GAN generate images according to the description given.

GANs are also being used in the synthesis of hig h-resolution images.SRGAN or Super Resolution GAN synthesizes images of very high-resolution. ProGAN or Progressively Growing of GAN provide a general method to generate high resolution images given a distribution of images. Pix2PixHD modifies the original Pix2Pix framework to generate high resolution images that are conditionally dependent on an input image.

GANs can also be applied in a 3D setting, 3D-GAN generated 3D images from a given 2D image. Also, GANs such as MidiNet can be used for composing music. GANs such as Vid2Vid can enable semantic label translation in a video thus eliminating the need of hardware-based renderer in graphics industries. AnoGAN can detect anomalies in Medical Data.

GANs have given birth to an entire new field of Adversarial Machine Learning where the existing machine learning models are given malicious data in order to test their robustness and reduce their biases. These adversarial examples are mostly generated using GANs.

The applications where GANs are used keep on growing every month due to a lot of ongoing research in the field. Many existing applications in the field of object detection andimage-to-image translation are using GANs to enhance their already state of the art methods.

V I. LIMITATIONS OF GANs

Being ubiquitous and simple as they are GANs still have some limitations. GANs are quite unstable during training due to the indeterministic nature of

the min-max game. Also, it is difficult for GANs to converge as finding Nash Equilibrium is a difficult computational problem than optimising a cost function

GANs also suffer from a property called mode collapse where the generator finds a weakness in the discriminator and keeps on generating images to exploit that weakness, and the discriminator is unable to cope with that weakness and thus continues to get exploited by the generator this causes the generated images to be undesirable due to increase in a particular feature of the image than the other features.

Improving GANs against these limitations remains an active area of research today.

V II . CONCLUSION

GANs are generative models that use supervised learning to implicitly estimate a given data distribution. They have been more successful in some aspects than other generative models but there are still some limitations that are present in the training process.

GANs have many applications in the field of image generation often becoming the state-of-the-art solution in many domains in the field. The emergence of adversarial machine learning is an important development towards the security and robustness of machine learning models. Improvement in GANs would enable progress in many fields in the future and would enable many further applications using them.

REFERENCES

[1] Christian Ledig, Lucas T heis, Ferenc Huszar, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, AlykhanTejani, Johannes Totz, Zehan Wang, Wenzhe Shi (2016). " Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network". arXiv preprint arXiv:1609.04802.

[2] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, SherjilOzair, Aaron Courville, YoshuaBengio(2014). Generative Adversarial Networks. *arXiv preprint arXiv: 1406.2661*.

[3] Ian Goodfellow (2017). NIPS 2016 T utorial: Generative Adversarial Networks. *arXiv preprint arXiv: 1701.00160*.

[4] Jiajun Wu, Chengkai Zhang, T ianfanXue, William T . Freeman, Joshua B. Tenenbaum (2016). Learning a Probabilistic Latent Space of Object Shapes via 3D Generative-Adversarial Modeling. *arXiv preprint arXiv:1610.07584*.

[5] Jun-Yan Zhu, T aesung Park, Phillip Isola, Alexei A. Efros (2017). Unpaired Image-to-Image T ranslation using Cycle-Consistent Adversarial Networks. *arXiv preprint arXiv:1703.10593*.

[6] Li-Chia Yang, Szu-Yu Chou, Yi-Hsuan Yang (2017). MidiNet: A Convolutional Generative Adversarial Network for Symbolic-domain Music Generation. *arXiv preprint arXiv:1703.10847*.

[7] Mehdi Mirza, Simon Osindero(2014). Conditional Generative Adversarial Nets. *arXiv preprint arXiv: 1411.1784*.

[8] Phillip Isola, Jun-Yan Zhu, T inghui Zhou, Alexei A. Efros (2016). Image-to-Image T ranslation with Conditional Adversarial Networks.*arXiv preprint arXiv: 1611.07004*

[9] Radford, A., Metz, L., and Chintala, S. (2015). Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434*.

[10]] T ing-Chun Wang, Ming-Yu Liu, Jun-Yan Zhu, Andrew T ao, Jan Kautz, Bryan Catanzaro (2017). High-Resolution Image Synthesis and Semantic Manipulation with Conditional GANs. *arXiv preprint arXiv:1711.11585*.

[11]] T eroKarras, T imo Aila, Samuli Laine, Jaakko Lehtinen (2017). Progressive Growing of GANs for Improved Quality, Stability, and Variation. *arXiv preprint arXiv:1710.10196*. [1 2] T homas Schlegl, Philipp Seeböck, Sebastian M. Waldstein, Ursula Schmidt - Erfurth, Georg Langs (2017). Unsupervised Anomaly Detection with Generative Adversarial Networks to Guide Marker Discovery. *arXiv preprint arXiv:1703.05921*.

[13] Zhang, H., Xu, T ., Li, H., Zhang, S., Huang, X., Wang, X., and Metaxas, D. (2016). Stackgan: T ext to photo-realistic image synthesis with stacked generative adversarial networks. *arXiv preprint arXiv:1612.03242*.

K- Nearest Neighbor Classification for the cancer patient data for the disease detection

Gulshan Bleem
Departm ent of Com puter Science and Engineering
Baba Banda Singh Bahadur Engineering College
Fatehgarh Sahib, India
mumtazbleem@gmail. com

Ishpreet Singh virk
Departm ent of Com puter Science and Engineering
Baba Banda Singh Bahadur Engineering College
Fatehgarh Sahib, India
Ishpreet.virk@bbsbe c.ac.in

Abstract - Enlarge amount of he alth re late d data is be ing produce s in various le ve ls of he alth syste m. Due to the siz e of the data it will be difficult to proce ss the data and the n e xtract the analysis. But using diffe re nt machine learning base d approache s data can be proce sse d. Machine le arning base d approach will provide the e fficie nt data for curing of the patie nts. Even he lps in fore cast the future dise ase . Patie nts past history for diffe re nt parame te rs can contribute to his probability of various he alth re late d issue s. But using Association cluste ring and Time Se rie s base d data mining in C ontinuous data e arly warning syste m can be de ve lope d. This pre diction base d syste m can de fine the dise ase while analyz ing his e xisting parame te rs. Some le ve l of care can prote ct the patie nt from dise ase .

Keywords- KNN, Manhatten, Euclidean, Data mining

I.INTRODUCTION

With the ever-increasing data in every field, there is a need to utilize the obtained information in a productive way and the healthcare sector specifically, has been slow in adopting technology to clout these assets. Focusing on usage and advantages of various technologies and their applications, they can result in refined patient data analysis and better overall healthcare predictions. These technologies and platforms can provide provisions for improved solutions, individual health plans and development of clinical support systems. Nowadays, there has been an extensive increase in the number of diabetic [1] patients. This is due to the failure of pancreas to produce enough insulin that is required. In order to identify the high risk patients, advanced technologies scan be used. For disease predictions, the patient data can be collected from various sources like Physicians' notes, imaging devices, computerized entries, etc. All this collected data is referred to as Big Data [1]. The major characteristics of big data are Volume (amount of data available), Variety (formats of available data), Velocity (speed of data generation), Veracity (trustworthiness of data) and Value (worth of extracted data). This data can be in structured, unstructured or semi-structured format depending upon the source from where the data has been collected [1]. So, delaying healthcare data for the sake of obtaining valuable insights is at oil some task to achieve. The emerging technologies play a pivotal role in addressing these challenges and Data mining proves to be an appropriate field. The term "Data mining" was first coined in 1990's and is often considered as a synonym of Knowledge discovery in databases. But data mining is actually a crucial part of KDD where the data is analyzed to obtain connections and find patterns within the dat a and using them predictions can be made or there center ends can be uncovered and analyzed whereas KDD is the overall process of knowledge extraction. These tasks are broadly classified into supervised, unsupervised and reinforcement learning. Various search engines and social networking sites collect huge amounts of data and by using different data mining techniques; they try to find the hidden patterns within the data [2]. Data mining is widely used in diverse as such as banking, e-commerce, health and medicine, genetics, education, stock exchange and various other fields. It is an emerging technology that is designed to efficiently handle big data and provide useful implications in various fields including health care. The data mining techniques are broadly classified into two main categories, namely, predictive and descriptive. The predictive data

mining techniques focus on understanding the future, making predictions using the available datasets whereas the descriptive techniques summarize and analyze the past data and properties to make it useful in predicting new ones. Predictive analysis considers the various symptoms, their effects on health and can help in early detection and prevention of these diseases. Different data mining techniques have been implemented to obtain the results regarding breast cancer, heart disease and diabetes to analyze the current status and make future predictions regarding the occurrence of disease, early detection and preventable patient deaths.

A. Data Mining Process

Information mining is a procedure of example revelation from a possibly enormous measure of information. It is a multi-disciplinary subject that is imagined in light of rationales in database frameworks. It can likewise be viewed as a factual strategy. Cases of information mining methods are machine learning, visual information mining, neural systems, design acknowledgments, flag handling, and so on. Information mining can likewise be viewed as data innovation advancing and thusly fanning out into sub-forms that comprise of gathering information, making database and management, examining information lastly translating information [1].

This procedure includes six key steps

- choosing the information,
- separating undesirable information,
- adding value to the filtered information,
- programming,
- information mining and creating report

B. Classification and clustering

Clustering in information mining is utilized to find design dispersion in fundamental information. The point of grouping techniques to subdividing the information points in a class in a way that focuses which has a place with same class are similar to various classes. These classes are named as bunch and number of classes can be controlled by calculation reassigned. The use of grouping is fluctuated in the field as therapeutic, correspondence, business and so forth [4].

C. KNN based technique

K-NN based technique is to identify the k-nearest neighbor. KNN is the technique used for classification. The output will be the membership function. The object in question will be put into that class which is closest to the k-nearest neighbor. KNN can also be considered as the regression. Output is the value of the object [11].

II. LITERATURE SURVEY

In the recent years, more efforts have been made to improve the accuracy of prediction. Some of the recent works have been discussed in this section.

Salama *et al.*[1] proposed a comparison among the different classifiers decision tree (J48), Multi-Layer Perception(MLP), Naïve Bayes(NB), Sequential Minimal Optimization (SMO), and Instance Based for K-Nearest neighbor (IBK) on three different databases of breast cancer (Wisconsin Breast Cancer (WBC), Wisconsin Diagnosis Breast Cancer (WDBC) and Wisconsin Prognosis Breast Cancer (WPBC) by using classification accuracy and confusion matrix based on 10-fold cross validation method.

Han Wu et al. [4] proposed a model that implemented k-means algorithm and logistic regression using WEKA toolkit to predict type-2 diabetes. The main aim was to improve the accuracy and implement the model using various datasets. This model produced satisfying results and was less time consuming. Accuracy of the proposed model was 3.04% more than the existing ones.

Isaac et al.[5] compared the accuracy obtained by implementing k-means algorithm and decision tree for diagnosis of breast cancer. Results showed that both the techniques resulted in high accuracy but statistical results show that k-means algorithm has higher performance than decision tree. On the other hand, hybrid model was proposed by Chen et al. for prediction of type-2 diabetes [6].This model implemented k-means for data reduction and decision tree for classification which resulted in better accuracy and promising results.

Kanika et al. [7] proposed an optimized hierarchical clustering method which aimed at reducing the computationcost.GeneticalgorithmandSVMwerealsou

sedforfeatureselectionandclassification respectively. PIMA Indians Diabetes dataset was used and the accuracy was improved by 1.351%. Another hybrid approach used k-means clustering using feature similarity and SVM for classification was implemented using PIMA Indian Diabetes dataset [8].The results showed that this combination has high prediction accuracy.

Nahato et al. [9] proposed a hybrid approach in which features of the dataset were divided into fuzzy sets and then extreme learning machine was used to classify them. Three different datasets were used to implement the proposed algorithm, namely, Cleveland heart disease dataset (CHD), Statlog heart disease dataset (SHD)and Pima Indian diabetes dataset(PID).The algorithm was run by varying the number of hidden layer neurons and the ones that produced the most efficient results were selected. This model performed better in terms of accuracy and training time. Another hybrid approach was proposed that used Adaptive Neuro-Fuzzy Inference System[10]. This approach had better detection rate than individual algorithm implementations.

Patricia et al.[11]proposed a variant of fuzzy k-nearest neighbor that used four different metrics to calculate the distance between vector and its neighbors.

Vaishali et al.[12] proposed a combination of genetic algorithm and Multi Objective Evolutionary Fuzzy Classifier. Feature selection was done using genetic algorithm and Fuzzy logic was used to classify the datavalues which resulted in increased performance. For the prediction of pre-diabetes and identification of risk factors,

Rajeswari et al.[13] used fuzzy logic based association rules. It eliminated the problem of boundary value misinterpretation and predicted the different types of outliers.

Suchethaet al.[14]used huma nbreath for disease detection by implementing modified convolution neural network. The signals acquired were fed to the neural network implemented in Matlab environment. The performance evaluation showed that the mean square error value reduced and the overall classification performance improved.

Deeraj et al.[15] used k-nearest neighbor and naive bayes on a large dataset to increase the efficiency of the overall system. Deepika et al. [16] classified the Pima Indian Diabetes dataset and breast cancer dataset using different data mining techniques. Performance analysis showed that naive bayes produced best accuracy for breast cancer dataset while SMO was best suitable for diabetes dataset.

Khalil et al. [17] used support vector machine, F-C mean, K-Mean and the Probabilistic Neural Network to predict depression namong diabetes patients. Same data set was used to implement these algorithms in Matlab. Comparison of accuracy showed that SVM produced more precise results than others.

Liu [18] implemented logistic regression to classify the dataset associated to breast cancer by using a combination of features. Varying these to selected attributes affected the performance of classifier. In case of breast cancer diagnosis, logistic regression proved to be an easy and efficient way.

Sun et al. [19] used a combination of logistic regression and random forest algorithm to select differently expressed genes of breast cancer on microarray dataset. The prediction accuracy rates were analyzed by varying the threshold value. Top 20 genes were recognized that are expected to influence the development of breast cancer and a maximum accuracy obtained was 95.57%.

III. PROPOSED TECHNIQUE

1. Collect the dataset for Breast cancer prediction
2. Build a training set and testing set from the available data values
3. Normalize the dataset
4. Establish the relationship between testing set and training set using Euclidean distance and Manhatten Distance

   $\sqrt{\sum_{k=1}^{m}(X_{ik}-X_{jk})}$ (Euclidean distance)

   $D1=\sum_{k=1}^{m}(X_{ik}-X_{jk})$ (Manhatten Distance)
5. Build a range using the obtained minimum and maximum values for each test data point
6. Arrange the data on the basis of distance from highest distance to the lowest distance.
7. Pick the top k distances for the top of arrange list.

IV. ALGORITHM

1. Load the data
2. Initialize the value of k
3. For getting the predicted class, iterate from 1 to total number of training data points

   I. Calculate the distance between test data and each row of training data. Here we will use Euclidean distance and Manhatten distance as our distance metric since it's the most useable method.

   II. Sort the calculated distances in ascending order based on distance values

   III. Get top k rows from the sorted array

   IV. Get the most frequent class of these rows

   V. Return the predicted class

V. FLOWCHART

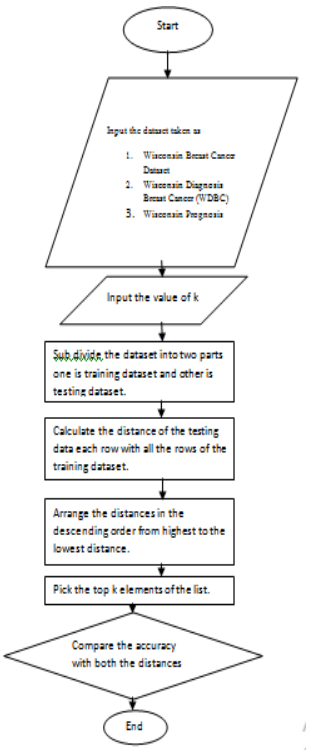


Figure 1.1 Flowchart

VI. RESULTS AND DISCUSSIONS

Current research is performed on dataset of Wisconsin Breast Cancer Dataset, Wisconsin Diagnosis Breast Cancer (WDBC), and Wisconsin Prognosis. Different datasets are classified foe disease detection using KNN based technique. two types of distance metrics are prepared. One is using Euclidean distance and other is Manhatten distance. The accuracy for the classification is compared for both the techniques.

Accuracy

Table 1.1 shows the accuracy for the KNN with Euclidean distance and Manhatten distance. The

Accuracy has been calculated using three types of datasets like Wisconsin Breast Cancer Dataset, Wisconsin Diagnosis Breast Cancer (WDBC), and Wisconsin Prognosis. For all the datasets KNN using Euclidean is performing better compared to the Manhatten.

Table 1.1 Accuracy comparison

Dataset Names	Accuracy using Using Euclidean	Accuracy using Manhatten
WisconsinBreastCancer	82.4558%	78.0701%
Wisconsin Diagnosis Breast Cancer	82.1514%	88.4955%
Wisconsin Prognosis Breast Cancer	83.5414%	70.0338%

Accuracy Comparison Graph for Wisconsin Breast Cancer Dataset

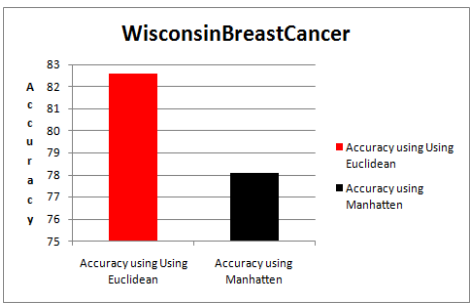


Figure 1.2 Accuracy comparisons for Wisconsin Breast Cancer Dataset

This graph shows the accuracy comparison for Wisconsin Breast Cancer Dataset. KNN using Euclidean distance performs better than the KNN with Manhatten distance.

Accuracy C omparison graph for W isconsin Diagnosis Bre ast C ance r

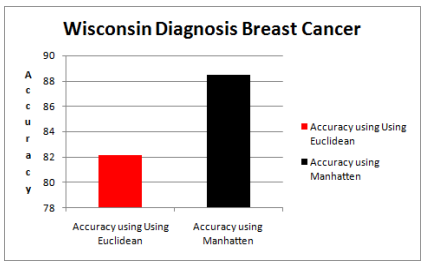


Figure 1.3 Accuracy comparisons for Wisconsin Diagnosis Breast Cancer Dataset

KNN using Manhatten distance is performing better for this dataset compared to KNN using Euclidean distance.

Accuracy Comparison graph for Wisconsin Prognosis Breast Cancer

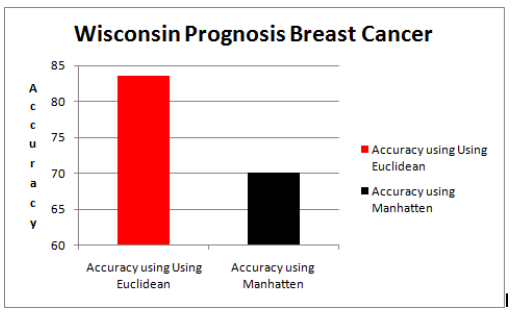


Figure. 1.4 Accuracy comparisons for Wisconsin Prognosis Breast Cancer

The accuracy comparison for the Wisconsin Prognosis Breast Cancer dataset for the KNN using Euclidean distance is performing better compared to KNN using Manhatten distance.

VII. CONCLUSION

In current time there are wide number of applications in various fields which generates large number of data of different nature. There requires a processing machine which can process this much large data. For the processing purpose first classification of different data items will be required. In current research KNN using two different distance techniques has been evaluated. The performance has been tested on three different datasets. These datasets are Wisconsin Breast Cancer Dataset, Wisconsin Diagnosis Breast Cancer (WDBC), and Wisconsin Prognosis. KNN using Euclidean distance is performing better than the KNN using Manhatten. The accuracy has been tested for different values of k. so that testing and training set for all the three dataset items are being performed on fixed basis.

VIII. FUTURE WORK

Current research has been performed on three different datasets Wisconsin Breast Cancer Dataset, Wisconsin Diagnosis Breast Cancer (WDBC), Wisconsin Prognosis. KNN using Manhatten and Euclidean distance has been evaluated. In future genet ics based optimization based technique can be evaluated.

REFERENCES

[1] Gouda I. Salama, M.B.Abdelhalim, and Magdy Abd-elghany Zeid,. Breast Cancer Diagnosis on Three Different Datasets Using Multi-Classifiers. International Journal of Computer and Information Technology ,September 2012 ,volume 01,issue 01, pp. 36-43.

[2] A.Kunjir, H.Sawant, and N.F.Shaikh,“ Data mining and visualization for prediction of multiple diseases in healthcare, "in Big Data Analytics and Computational Intelligence(ICBDAC),2017 International Conference on, pp.329-334,IEEE,2017.

[3] I.Kavakiotis,O.Tsave,A.Salifoglou,N.Maglaveras,I.Vlahavas, and I.Chouvarda,“ Machine learning and datamining methods india bêtes research, "Computational and structural biotechnology journal,vol.15,pp.104-116,2017.

[4] H.Wu,S.Yang,Z.Huang,J.He,andX.Wang,“ Type2diabetes mellitus prediction model based on data mining, "Informatics in Medicine Unlocked,vol.10,pp.100-107,2018.

[5] L.D.Isaac and C.Sureshkumar,“ Diagnosis prognosis and prevention of breast cancer based on present scenario of human life,"in2018International Conference on Communication information and Computing Technology (ICCICT), pp.1-7,IEEE,2018.

[6] W.Chen,S.Chen,H.Zhang,andT.Wu,“ Ahybridpredictionmodelf ortype2diabetesusing k-means and decision tree,"inSoftware Engineering and Service Science(ICSESS),20178thIEEE International Conferenceon,pp.386-390,IEEE,2017.

[7] K.Bhatia and R.Syal,“ Predictive analysis using hybrid clustering india bêtes diagnosis,"in Control, Automation &Power Engineering(RDCAPE), 2017Recent Developments in,pp.447-452, IEEE,2017.

[8] A.H.Osman and H.M.Aljahdali,“ Diabetes disease diagnosis method based on feature extraction using k-svm,"IntJ Adv Comput SciAppl,vol.8,no.1,2017.

[9] K.B.Nahato,K.H. Nehemiah, and A.Kannan,“ Hybrid approach using fuzzy sets and extreme learning machine for classifying clinical datasets,"Informatics in Medicine Unlocked, vol.2,pp.1-11, 2016.

[10] H.F atemidokht and M.K.Rafsanjani, “ Development of a hybrid neuro-fuzzy system as a diagnostic tool for type2

diabetes mellitus,"in Fuzzy and Intelligent Systems (CFIS), 20186thIranian Join Congresson, pp.54-56,IEEE,2018.

[11] P.Melin,E.Ramirez,andG.Prado-Arechiga,“ A new variant of fuzzy k-nearest neighbor using intervaltype-2 fuzzy logic,"in2018 IEEE International Conference on Fuzzy Systems(FUZZ-IEEE), pp. 1-7,IEEE,2018.

[12] R.Vaishali,R.Sasikala,S.Ramasubbareddy,S.Remya,andS.Nalluri,“ Genetic algorithm based features election and moe fuzzy classification algorithm on pima Indians diabetes dataset, "in ComputingNetworkingandInformatics(ICCNI),2017Internationalconferen ceo n,pp.1-5, IEEE, 2017.

[13] A.Rajeswari,M.S.Sidhika,M.Kalaivani,andC.Deisy,“ Predictio n of prediabetes using fuzzy logic based association classification, "in 2018 Second International Conference on Inventive Communication and Computational Technologies(ICICCT),pp.782-787,IEEE,2018.

[14] S.LekhaandM.Suchetha,“ Real-time non-invasive detection and classification of diabetes using modified convolution neural network,"IEEE journal of bio medical and health informatics, vol. 22,no.5,pp.1630-1636,2018.

[15] D.Shetty, K.Rit,S. Shaikh, and N.Patil,“ Diabetes disease prediction using datamining,"in Innovations in Information ,Embedded and Communication Systems (ICIIECS), 2017 International Conferenceon, pp.1-5,IEEE,2017.

[16] D.Verma and N.Mishra,“ Analysis and prediction of breast cancer and diabetes disease datasets using data mining classification techniques, "in2017 International Conference on Intelligent Sustainable Systems (ICISS),pp.533-538, IEEE, 2017.

[17] R.M.KhalilandA.Al-Jumaily,“ Machine learning based prediction of depression among type2 diabetic patients," in Intelligent Systems and Knowledge Engineering (ISKE), 2017 12th International Conferenceon,pp.1 -5,IEEE,2017.

[18] L.Liu,“ Research on logistic regression algorithm of breast cancer diagnose data by machine learning,"in2018 International Conference on Robots & Intelligent System(ICRIS),pp.157-160,IEEE, 2018.

[19] M.Sun,T.Ding,X.-Q.Tang,andK.Yu,“ An efficient mixed- model for screening differentially expressed genes of breast cancer based onlr-rf,"IEEE/ACM Transactions on Computational Biology and Bio-informatics,2018.

Soft Computing Approach to Image denoising

Ravneet Kaur
Computer Science & Engineering
Shaheed Udham Singh College of Engineering & Technology
M ohali, India
er.ravneet89@gmail.com

Parvinder Kaur
Computer Science & Engineering
Shaheed Udham Singh College of Engineering & Technology
M ohali, India
parvinderkaurcse@sus.edu.in

***Abstract*— Image denoising is an important task in image processing. In the past three decades, a lot of denoising methods have been developed by researchers. Most of the existing methods have the same drawbacks such as over smoothing and edge sharpening. In research of image denoising, in order to remove noise effectively and preserve edges and key details, an effective image denoising algorithm based on fuzzy logic is proposed. The fuzzy rule -based approach is a powerful method for formulation of many image processing techniques. This method can be very useful for segmentation, registration, identification and sharpening of objects in a complex scene. Adaptive fuzzy means filtering approach is used with a triangular membership function. S imulated experiments show that algorithm can effectively reduce the noise of the image, and its results need not to be adjusted, which can enhance the quality of the image.**

Keywords— denoising, fuzzy, bilateral, convolution, kernel, fuzzification, defuzzification.

I. INTRODUCTION

With the encroachment in computers and digital imaging technologies, the costs of digital cameras and computers are decreasing each year, and thus these equipments are becoming affordable these days. The usage of digital images in our daily life is becoming common. As more information can be extracted from digital images , as compared to one dimensional signal, many research areas, including material researches, are now utilizing digital images, such as microscopic images and X-ray images, as one of their evaluation tools. As the information from digital images are easier to be evaluated as compared with one dimensional signals, digital images are now commonly used in many research fields[6,7]. Unfortunately, similar to other digital signals, digital images are also sometimes unintentionally corrupted by unwanted signals, called noise. Image is a powerful medium to convey visual information. In digital Image Processing, removal of noise is a highly demanded area of research. Digital images are often corrupted by noise during their acquisition and transmission. Noisy images can be found in many imaging applications. TV images are corrupted because of atmospheric interference and imperfections in the image reception. Noise is also introduced in digital artworks when scanning damaged surfaces of the originals. Impulse noise corruption often occurs in digital image acquisition or transmission process as a result of photo-electronic sensor faults or channel bit errors. Image transmission noise may be caused by various sources, such as car ignition systems, industrial machines in the vicinity of the receiver, switching transients in power lines, lightning in the atmosphere and various unprotected switches. Digital images play an important role both in daily life applications such as Satellite television, Magnetic Resonance Imaging, Computer Tomography as well as in areas of research and technology such as geographical information systems and astronomy .Digital images are often corrupted by different types of noise during its acquisition and transmission phase. Such degradation negatively influences the performance of many image processing techniques and a preprocessing module to filter the images is often required.

A. Noise modelling

Noise represents unwanted information which deteriorates image quality. Noise is defined as a function (n) which affects the acquired image (f) and is not part of the scene (initial signal – s). Using the additive noise model, this process can be written as:

$$f(i, j) = s(i, j) + n(i, j) \quad (1)$$

Digital image noise may come from various sources. The acquisition process for digital images converts optical signals into electrical signals and then into digital signals and is one processes by which the noise is introduced in digital images. Each step in the conversion process experiences fluctuations, caused by natural phenomena, and each of these steps adds a random value to the resulting intensity of a given pixel. A type of noise is salt-and-pepper noise, for which a certain amount of the pixels in the image are either black or white (hence the name of the noise). Salt-and-pepper noise can, e.g., be used to model defects in the CCD or in the transmission of the image. Given the probability r (with $0 \leq r \leq 1$) that a pixel is corrupted, we can introduce salt-and-pepper noise in an image by setting a fraction of r/2 randomly selected pixels to black, and another fraction of r/2 randomly selected pixels to white[8].

Most denoising algorithms make two assumptions about the noisy image. These assumptions can cause blurring and loss of detail in the resulting denoised images. The first assumption is that the noise contained in the image is white noise. This means that the noise contains all frequencies, low and high. Because of the higher frequencies, the noise is oscillatory or non -smooth. The second assumption is that the true image (image without the noise) is smooth or piecewise smoothes. This means the true image or patches of the true image only contain low frequencies [9].

Previous methods attempt to separate the image into the smooth part (true image) and the oscillatory part (noise) by removing the higher frequencies from the lower frequencies. However, not all images are smooth. Images can contain fine details and structures which have high frequencies. When the high frequencies are removed, the high frequency content of the true image will be removed along with the high frequency noise because the metho ds cannot tell the difference between the noise and true image. This will result in a loss of fine detail in the denoised image. Also, nothing is done to remove the low frequency noise from the image. Low frequency noise will remain in the image even after denoising[10].

B. *Soft Computing Techniques*

Soft computing, as opposed to traditional computing, deals with approximate models and gives solutions to complex real-life problems. Unlike hard computing, soft computing is tolerant of imprecision, uncertainty, partial truth, and approximations. In effect, the role model for soft computing is the human mind. Soft computing is based on techniques such as fuzzy logic, genetic algorithms, artificial neural networks, machine learning, and expert systems. The concept of fuzzy logic was introduced by Zadeh as a method for representing human knowledge that is imprecise by nature. Fuzzy Logic based systems have four stages. The fuzzification interface transforms the crisp input value into a fuzzy linguistic value. The fuzzification is always necessary in a fuzzy logic system since the input values from existing sensors are always crisp numerical values. The inference engine takes the fuzzy input and the fuzzy rule base and generates fuzzy outputs. The fuzzy rule base is in the form of "IF-THEN" rules involving linguistic variables. The last processing element of a fuzzy logic system is the defuzzification which has the task of producing crisp output actions [2].

C. *Denoising artifacts*

Denoising often adds its own noise to an image. Some of the noise artifacts created
By denoising are as follows:

- Blur: attenuation of high spatial frequencies may result in smooth edges in the image. Ringing/Gibbs Phenomenon: truncation of high frequency transform coefficients may lead to oscillations along the edges or ringing distortions in the image.
- Staircase Effect: aliasing of high frequency components may lead to stair-like structures in the image.
- Checkerboard Effect: de-noised images may sometimes carry checkerboard structures.
- Wavelet Outliers: these are distinct repeated wavelet-like structures visible in the de-noised image and occur in algorithms that work in the wavelet domain.

The paper is organized in five sections. Section 2 introduces the literature. Section 3 introduces the steps of the proposed algorithm. Simulation and performance analysis is made in Section 4. Conclusions are given in Section 5.

II. EXISTINGWORK

It is well known that the nonlinear filters have been widely exploited due to their much improved filterin g performance, in terms of impulse noise attenuation and details preservation. One of the most popular and robust nonlinear filters is the standard median (SM) filter [1], which exploits the rank-order information of pixel intensities within a filtering window and replaces the center pixel with the median value. The median filter and many of its variants, however, have limitations like blurring of image details due to simultaneous noise and signal suppression of median filter affecting the image fidelity badly. Due to its effectiveness in noise suppression and simplicity in implementation, various modifications of the SM filter have been introduced, such as the adaptive median (AM) [2] filter and the adaptive fuzzy multilevel median (AFMM) [3] filter. Although these methods have been improved, the quality of image filtering is still not satisfactory. Most of these algorithms provide good outputs at smaller percent of noise levels and find difficulty with higher level impulse noises. To addres s this issue, the boundary discriminative noise detection (BDND) filter was proposed in [4], which can handle image corruption even up to 90% noise density. But it is too time-consuming to be used in real application. Another powerful open -close sequence (OCS) filter was proposed in [5], which can restore images that are corrupted by 30%-80% probability impulse noise. But it performs a worse result when the noise ratio is not high.

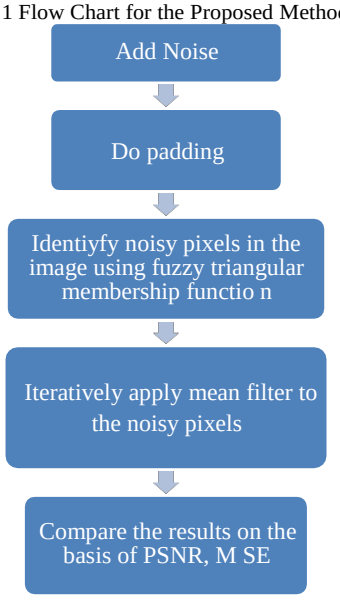
III. THE PROPOSED METHOD

The proposed method uses adaptive fuzzy mean filtering.

A. *Adaptive Fuzzy Mean Filtering*

The first step is to add the salt and pepper noise to the image. Padding is done so that the noise can also be removed from the corners of the image. A fuzzy triangular membership function is used to help decide whether the pixel is corrupted or uncorrupted. For the reconstruction of image from noisy image the fuzzy mean filtering is used. The process is repeated until we get the consistent denoised image.

Fig. 1 Flow Chart for the Proposed Method



B. *Triangular membership function*

It is defined by a lower limit **a**, an upper limit **b**, and a value **m**, where **a** < **m** < **b**.

$$\mu_A(x) = \begin{cases} 0, & x \le a \\ \frac{x-a}{m-a}, & a<x\le m \\ \frac{b-x}{b-m}, & m<x<b \\ 0, & x\ge b \end{cases}$$

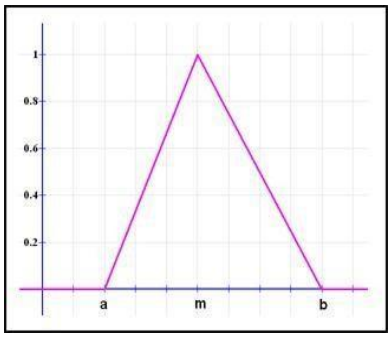


Fig. 2 Fuzzy T riangular Membership Function

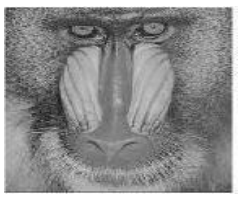
1. Consider a : variable containing all parameters x,z : feet of a triangle y : peak of a triangle
2. if (a<x) Tr = o;
3. if (x ≤a ≤ y) Tr = (a-x)/(y-x)
4. if (y ≤ a ≤ z) Tr =(y-a)/(z-b)
5. if (Tr (r : m , a ,M) < Threshold) pixel = noisy pixel else pixel = image pixel

Fig. 3 Fuzzy Membership Function

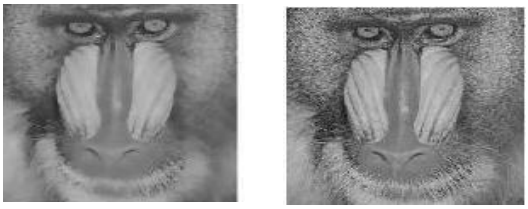
Triangular Membership Functions are specified by three parameters as follows: Here a is the variable which stores all the values of the parameters. The parameter x and z locate the "feet" of the triangle and the parameter y locates the peak. Triangular membership functions are easy to use and they are best to use for impulse noise removal. A pixel is either the noise pixel or the image pixel. Threshold value is represented by Tr. Noise pixel can be obtained by using the triangular membership function and the correction term is added to the noise pixels where the image pixels are left alone so that the image can be enhanced efficiently.

I V. RESULT S & DISCUSSIONS

Below are the proposed result and discussions:

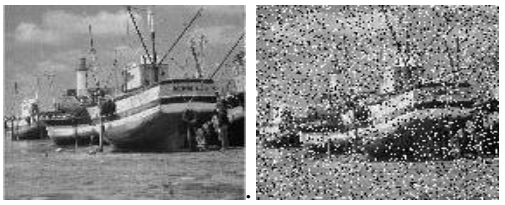


(a) **(b)**



(c) **(d)**

Fig 4: (a) Original image , (b) Noisy image
(c) Existing Technique (d) Proposed Technique



(a) **(b)**

(c) **(d)**

Fig 5: (a) Original image, (b) Noisy image
(c) Existing Technique (d) Proposed Technique

C. QUALITY MEASURES

It is important at this point to define the quality measures used compare our proposed filter with other state-of-the-art filters. In the field of image reconstruction several quality measures have been identified to determine whether the filter used is more or less efficient. Some of these measures include the following.

(i) Mean Squared Error (MSE). Better reconstruction gives lower values and zero indicates perfect reconstruction.

(ii) Peak Signal to Noise Ratio (PSNR). This is a measure related to the previous one. It is a logarithmic measure widely used in image processing. Higher values indicate better quality.

T ABLE 1: PERFORMANCE EVALUATION (MEAN SQUARE ERROR)

Image s	MSE (Base Me thod)	MSE (Proposed Me thod)
Image 1	.0016978	.0004782
Image 2	.0009436	.0002983
Image 3	.0004716	.0001603
Image 4	.0009135	.0003439
Image 5	.0004962	.0001195

TABLE 2: PERFORMANCE EVALUATION (PEAK SIGNAL TO NOISE RATIO)

Images	PSNR (Base Me thod)	PSNR (Proposed Me thod)
Image 1	27.093	33.203
Image 2	30.271	35.25
Image 3	32.212	37.94
Image 4	31.009	34.63
Image 5	33.008	39.22

V. CONCLUSION & FUTURE SCOPE

Image denoising is a very important step in image processing. In many significant applications it is considered as a mandatory preprocessing step. The literature survey shows that most of the existing techniques do over smoothing and edge sharpening. In order to remove noise effectively and preserve edges and key details, an effective image denoising algorithm based on fuzzy logic is proposed. Intelligent fuzzy means filtering with a triangular membership function is used in this paper. Simulated experiments show the algorithm can effectively reduce the noise of the image, and its results need not to be adjusted, which can enhance the quality of the image.

.

REFERENCES

[1] Pitas and A. N. Venetsanopoulos, "Order statistics in digital image processing," Proc. IEEE, vol. 80, no. 12, Dec. 1992, pp. 1893-1921.

[2] Hwang. H and Haddad. R.A, "Adaptive Median Filters: New Algorithms and Results", IEEE Trans. on Image Processing, vol. 4, no. 4, 1995, pp. 499-502.

[3] Xiahua Yang and Peng Seng Toh, "Adaptive fuzzy multilevel median filter", Image Processing, IEEE Transactions on Volume 4, Issue 5, May 1995, pp. 680-682.

[4] Pei-Eng Ng and Kai-Kuang Ma, "A Sw itching Median Filter With Boundary Discriminative Noise Detection for Extremely Corrupted Images", IEEE Trans. Image Processing, VOL. 15, NO. 6, June 2006, pp. 1506-1516.

[5] Ze-Feng, D. Zhou-Ping, Y. and You-Lun, X, "High Probability Impulse Noise-Removing Algorithm Based on Mathematical Morphology", Signal Processing Letters, IEEE Volume 14, Issue 1, 2007, pp. 31 - 34

[6] A. Buades, "image denoising methods," IEEE Trans, vol. 32, n0. 18, pp. 321-327, 2015.

[7] A. Pizurica and W. Philips, "Estimating the probability of the presence of a signal of interest in multiresolution single- and multiband image denoising," IEEE Trans. Image Process., vol. 15, no. 3, pp. 654–665, 2006.

[8] F. Palhano et al., "Real time ultrasound image denoisingTo cite this version : HAL Id : inria-00476122 Real time ultrasound image denoising," 2010.

[9] H. Guo, J.E. Odegard, and C.S. Burrus,"Nonlinear processingofa shift invariant DWT for noise reduction," SPIE, MathematicalImaging: Wavelet Applications for DualUse, vol. 10,no. 4, pp.333-339, 2007.

[10]] J. F. Cai, H. Ji, Z. Shen, and G. B. Ye, "Data-driven tight frame construction and image denoising," Appl. Comput. Harmon. Anal., vol. 37, no. 1, pp. 89–105, 2014.

[11] J. Patil and S. Jadhav, "A Comparative Study of Image Denoising Techniques," Int. J. Innov. Res. Sci. Eng. Technol., vol. 2, no. 3, pp. 787–794, 2015.

[12] J. Mohan, V. Krishnaveni, and Y. Guo, "A survey on the magnetic resonance image denoising methods," Biomed. Signal Process. Control, vol. 9, no. 1, pp. 56–69, 2014.

[13] K. Dabov, A. Foi, and V. Katkovnik, "Image denoising by sparse 3D transformation-domain collaborative filtering," IEEE Trans. Image Process., vol. 16, no. 8, pp. 1–16, 2007.

[14] K. Mohan, "Fundamentals of digital image processing," Prentice-Hall, vol. 5, no.8, pp. 661-667, 2004.

Performance Comparison of Homogenous and Heterogeneous Clustering Techniques in WSNs

Harjit Kaur
Departm ent of Com puter Science and Engineering
Baba Banda Singh Bahadur Engineering College
Fatehgarh Sahib, Punjab, India
k.harjitcse@gmail.com

Sukhwinder Sharma
Departm ent of Com puter Science and Engineering
Baba Banda Singh Bahadur Engineering College
Fatehgarh Sahib, Punjab, India
sukhwinder.shar ma @bbsbec.ac. in

Abstract*—Wi re le ss Se nsor Ne twork (W SN) is compose d of multiple numbe r of node s e ach of which consists of se nsing de vice s to colle ct data from e nvironme nt. The se se nsing de vice s are provide d with proce ssing unit to pe rform ope rations on data. Se nsing de vice s are de plo ye d randomly in re mote e nvironme nt, due to which batte ry charging or re place me nt is not conside red practical. C luste ring has been prove n as one of the most e ffe ctive te chnique for re ducing e ne rgy consumption of the wire le ss se nsor ne tworks. In this se nsor node s are groupe d into se parate cluste rs. W ire le ss Se nsor Ne twork have be e n paid huge atte ntion for the ir pote ntial use in obse rving e nvironme nt, he alth, military surve illance , home applications and many more . The de sign of se nsor ne twork is influe nce d by factors like scalability, e ne rgy consumption, e nvironme nt e tc. and de pends on the application. O f the thre e activitie s: se nsing, proce ssing and communication, most of the e ne rgy is spe nt on communication purpose s. Ene rgy conse rvation is thus a dominant factor in wire le ss se nsor ne tworks. Routing strate gy se le ction is ve ry important for prope r de live ry of packe ts. O n -going re se arch aims in e xte nding ne twork life time by de signing protocols that re quire s le ss e ne rgy communication. This pape r provide s a surve y on e ne rgy efficie nt routing in wirele ss sensor ne tworks.

I. INTRODUCTION

A Wireless Sensor Network or WSN as shown in Figure. 1 is supposed to be made up of a large number of sensors and at least one base station. The sensors are autonomous small devices with several constraints like the battery power, computation capacity, communication rang and memory. They also are supplied with transceivers to gather information from its environment and pass it on up to a certain base station, where the measured parameters can be stored and available for the end user.

In most cases, the sensors forming these networks are deployed randomly and left unattended to and are expected to perform their mission properly and efficiently.

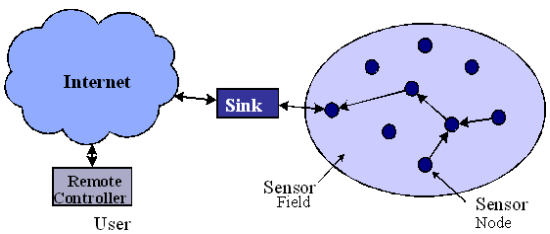


Figure 1. Sensor Network Architecture

As a result of this randomdeployment, the WSN has usually varying degrees of node density along its area.

Sensor networks are also energy constrained since the individual sensors, which the network is formed with, are extremely energy-constrained as well. The communication devices on these sensors are small and have limited power and range.

Both the probably difference of node density among some regions of the network and the energy constraint of the sensor nodes cause nodes slowly die making the network less dense.

Also it is quite common to deploy WSNs in harsh environment, what makes many sensors inoperable or faulty. For that reason, these networks need to be fault-tolerant so that the need for maintenance is minimized. Typically the network topology is continuously and dynamically changing, and it is actually not a desired solution to replenish it by infusing new sensors instead the depleted ones. A real and appropriate solution for this problemis to implement routing protocols that perform efficiently and utilizing the less amount of energy as Possible for the communication among nodes.

The development of wireless sensor networks (WSNs) has recently opened up a new and interesting area for the creation of new types of applications. WSNs

consist of a large number of small sensing nodes that monitor their environment, process data if necessary and send/receive processed data to/fromother sensing nodes. These sensing nodes, distributed in the environment, are connected to a sink node. In centralized networks, the sink collects sensor data to be used by the end user. In many cases, the sink is also capable of activating sensing nodes via broadcasting, by sending network policy and control information.

1) Sensor nodes: Sensor node as shown in Figure. 2 is a tiny device that includes three basic components: a sensing subsystemfor data acquisition from the physical surrounding environment, a processing subsystem for local data processing and storage, and a wireless communication subsystem for data transmission. In addition, a power source supplies the energy needed by the device to perform the programmed task. This power source often consists of a battery with a limited energy budget.

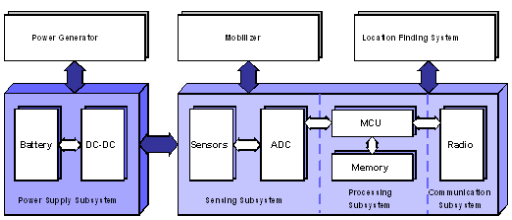


Figure. 2. Architecture of wireless sensor node

Thus architecture of sensor node consists of four main components:

1) A sensing subsystem including one or more sensors (with associated analog-to-digital converters) for data acquisition;
2) A processing subsystem including a micro - controller and memory for local data processing;
3) A radio subsystem for wireless data communication; and
4) A power supply unit.

2) Sink : Sink is a base station used to collect and process data. It acts as a gateway or interface between the sensor field and to the user. In wireless sensor networks (WSNs), all the data collected by the sensor nodes are forwarded to a sink node. Therefore, the placement of the sink node has a great impact on the energy consumption and lifetime of WSNs.

II. Applications of Wireless Sensor Networks

Wireless Sensor Networks have broad applications like environment monitoring, target tracking and surveillance. Unlike Mobile Ad-hoc Networks (MANETs), WSNs are usually application-specific. The areas where WSNs finds its major applications are:

Area Monitoring: Area monitoring is a common application of WSNs. In area monitoring, the WSN is deployed over a region where some phenomenon is to be monitored. An example in the military application is the use of sensors to detect enemy intrusion; a civilian example is the geo-fencing of gas or oil pipelines. When the sensors detect the event being monitored (heat, pressure) the event is reported to one of the base stations, which then takes appropriate action (e.g., send a message on the internet or to a satellite). Similarly, wireless sensor networks can use a range of sensors to detect the presence of vehicles ranging from motorcycles to train cars.

Air Pollution Monitoring: Wireless sensor networks have been deployed in several cities to mo nitor the concentration of dangerous gases causing air pollution. These can take advantage of the ad-hoc wireless links rather than wired installations, which also support mobility for testing readings in different areas.

Forest Fire Detection: A network of sensor nodes can be installed in a forest to detect the occurrence of forest fire. The nodes can be equipped with sensors to measure temperature, humidity and gases which are produced by fires in the trees or vegetation. The early detection is crucial for a successful action of the fire-Figurehters; the fire brigade will be able to know how the fire spreads.

Greenhouse Monitoring: Wireless sensor networks are also used to control the temperature and humidity level inside commercial greenhouses. When th e temperature and humidity drop below specific levels, the greenhouse manager must be notified via e-mail or cell phone text message, or host systems can trigger misting systems, open vents, turn on fans, or control a wide variety of system parameters.

Landslide Detection: A landslide detection system makes use of a wireless sensor network to detect the slight movements of soil and changes in various parameters that may occur before or during a landslide. With the data gathered it may be possible to know t he occurrence of landslides long before it actually happens.

Machine Health Monitoring: Wireless sensor networks have been developed for machinery condition-

base maintenance as they offer significant cost savings and enable new functionalities. In wired systems, the installation of enough sensors is often limited by the cost of wiring. Previously inaccessible locations, rotating machinery, hazardous or restricted areas, and mobile assets can now be reached with wireless sensors.

Water/Waste-Water Monitoring: There are many opportunities for using wireless sensor networks within the water/waste water industries. Facilities not wired for power or data transmission can be monitored using industrial wireless Input/output (I/O) devices and sensors powered using solar panels or battery packs and also used by pollution control board.

Agriculture: Using wireless sensor networks within the agricultural industry are increasingly common; using a wireless network frees the farmer from the maintenance of wiring in a difficult environment. Gravity feed water systems can be monitored using pressure transmitters to monitor water tank levels, pumps can be controlled using wireless I/O devices and water use can be measured and wirelessly transmitted back to a central control centre for billing. Irrigation automation enables more efficient water use and reduces waste.

Structural Monitoring: Wireless sensors can be used to monitor the movement within buildings and infrastructure such as bridges, flyovers, embankments and tunnels and thus enable monitoring of assets remotely without the need for costly site visits, as well as having the advantage of daily data, whereas traditionally this data was collected weekly or monthly, using physical site visits, involving either road or rail closure in some cases. It is also far more accurate than any visual inspection that would be carried out.

III. Classific ation of Wireless Sensor Networks

A. *Proactive Network s*

The nodes in this sort of network periodically switch on their sensors and transmitters, sense the environment and transmit the data of interest. Hence, they provide a snapshot of the relevant parameters at regular intervals. They are well suited for applications requiring periodic data monitoring. Some known instances of this kind are the LEACH protocol, some improvements on LEACH and PEGASIS

B. *Reactive Network s*

The nodes of the networks according to this scheme react immediately to sudden and drastic changes in the value of a sensed attribute. They are well suited for time critical applications.

C. *Hybrid Network s*

The nodes in such a network not only react to time-critical situations, but also give an overall picture of the network at periodic intervals in a very energy efficient manner. Such a network enables the user to request past, present and future data from the network in the form of historical, one-time and persistent queries respectively.

IV. Routing Models

All known routing protocols may be included into one of the following three models. This classification will facilitate the analysis of the protocols that have been taken into account in this work.

A. *One-hop / Single hop model*

This is the simplest approach and represents direct transmission. In these networks every node transmits to the base station directly as shown in Figure. 3. This communication implies not only to be too expensive in terms of energy consumption, but it is also infeasible because nodes have limited transmission range. Most of the nodes in networks with large area coverage usually are far enough thus their transmissions cannot reach the base station. Direct communication is not a feasible model for routing in WSN.

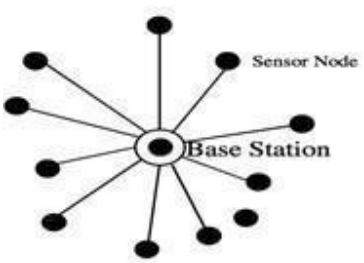


Figure 3. Single-Hop Model

B. *Multi-hop Planar Model*

In this model, a node transmits to the base station by forwarding its data to one of its neighbours, which is closer to the base station as shown in Figure. 4. The latter passes on it to a neighbour that is even closer to the base station. Thereby the information travels from source to destination by hop from one node to another until it

Reaches the destination. One of these techniques is data aggregation used in all clustering-based routing protocol. Even though these optimization techniques improve the performance of this model, it is still a planar model.

In a network composed by thousands of sensors, this model will exhibit high data dissemination latency due to the long time needed by the node information to arrive to the base station.

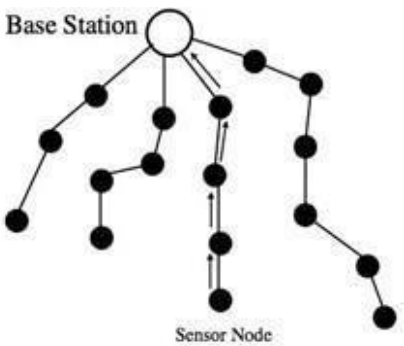


Figure 4. Multi-Hop Planer Model

V. Clustering -B as ed Hierarchic al Model

A hierarchical approach shown in Figure. 5 for the net- work topology breaks the network into several areas called clusters. Nodes are grouped depending on some parameter into clusters with a cluster head, which has the responsibility of routing the data from the cluster to other cluster heads or base stations. Data travels from a lower clustered layer to a higher one. Data still hops from one node to another, but since it hops from one layer to another it covers larger distances and moves the data faster to the base station than in the multi-hop model.

The latency in this model is theoretically much less than in the multi-hop model. Clustering provides inherent optimization capabilities at the cluster heads, what results in a more efficient and well-structured network topology. This model is more suitable than the one-hop or multi-hop model.

The remainder of this subsection deals with the characteristics and challenges of this model and the suitability of all sorts of model.

For several reasons direct communication is infeasible for a large sensor network that is formed by thousands of sensors. It is a model that wastes energy and even worse, nodes far from base station do not have enough transmission power to reach the base station what would

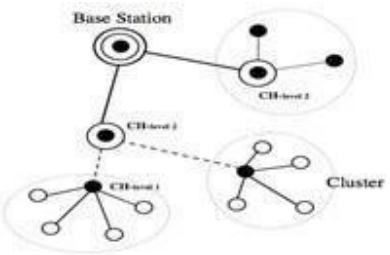


Figure 5. Hierarchical Clustering-based Model

Turn into unreachable the most part of the network. Even though the sensors would be close to the base station, the density of it would create such number of collisions that would seriously degrade the network efficiency. The multi-hop model is a more practical approach than the one-hop. In this case, data is forwarded by hops from one node to another until it reaches the base station.

Taking into account the energy constraint nodes that comprise sensor networks, it is a feasible approach. The coverage area is improved over the one-hop model since most nodes are able to connect the network and the amount of collisions is reduced.

Some drawbacks of this model are the high latency in networks comprised of thousands of sensors and the serious delay that data experiences. Perhaps the most important drawback is that the closest nodes to the base station would have to act as intermediaries to all traffic being sent to the base station by the rest of the network. As they have to handle all the traffic, they will die first creating a black hole around the base station for incoming traffic. This situation will appear another time with the new closest nodes to the base station causing in the mid-term that no data arrives to the base station and rendering the network useless.

In the clustering-based hierarchical model, data is aggregated in the cluster and sent to a higher-level cluster head, thus travelling greater distances than in both other models explained and reducing time and latency. LEACH and LEACH-C use a one level clustering whereas Chain-based 3 level PEGASIS uses three-level approach. Some advantages of this model comparing with multi-hop are that data moves faster to the base station thus reducing latency and that only cluster heads perform data aggregation unlike multi-hop model where every intermediate node perform this optimization tech - nique. Therefore, the clustering-based model is more suitable for time-critical applications than the multi-hop model. Common clustering routing protocols are: LEACH and SEP etc.

VI. Homogeneous and Heterogeneous Clustering Techniques

A. Homogeneous Clustering Techniques

In the energy homogeneous wireless sensor networks the energy level of each node is same. In homogeneous networks all the sensor nodes are identical in terms of battery energy and hardware complexity. With purely static clustering (cluster heads once elected, serve for the entire lifetime of the network) in a homogeneous network, it is evident that the cluster head nodes will be over-loaded with the long range transmissions to the remote base station, and the extra processing necessary for data aggregation and protocol co-ordination. As a result the cluster head nodes expire before other nodes. However it is desirable to ensure that all the nodes run out of their battery at about the same time, so that very little residual energy is wasted when the system expires. One way to ensure this is to rotate the role of a cluster head randomly and periodically over all the nodes as proposed in LEACH. However the downside of using a homogeneous network and role rotation is that all the nodes should be capable of acting as cluster heads, and therefore should possess the necessary hardware capabilities.

B. Heterogeneous Clustering Techniques

On the other hand, in a heterogeneous sensor network, two or more different types of nodes with different battery energy and functionality are used. However fixing the cluster head nodes mean that role rotation is no longer possible. When the sensor nodes use single hopping to reach the cluster head, the nodes that are farthest fromthe cluster heads always spend more energy than the nodes that are closer to the cluster heads. On the other hand when nodes use multi-hopping to reach the cluster head, the nodes that are closest to the cluster head have the highest energy burden due to relaying. Consequently there always exists a non-uniform energy drainage pattern in the network [1].

VII. Literature Review

The comprehensive literature survey on energy conservation [2], clustering techniques, LEACH and SEP routing protocols for WSNs; a review of research that aimed at reducing energy consumption so as to increase lifetime, stability and throughput has been also provided. In last few years WSNs have gained increasing amount of attention from both the researchers and users. As sensor nodes are generally battery-powered devices, the critical aspects to face concern how to reduce the energy Consumption of nodes, so that the network lifetime can be extended to reasonable times. So for solving the problem stated in the problem formulation, there are various energy conservation methods and techniques for WSNs. There are mainly three techniques given in survey; that are duty cycling, data-driven and mobility based techniques. Duty cycling approaches deals with topology control and power management but the problem with those techniques was that they are oblivious to data. Then the second approach was based on the data.it deals with data reduction, data prediction and data acquisition methods. But in these both (duty cycling and data driven) approaches the nodes are static. So in the third approach that is mobility approach, the nodes (either sensor nodes or sink) are movable.

The use of clusters for transmitting data to the base station leverages the advantages of small transmit distances for most nodes, requiring only a few nodes to transmit far distances to the base station. However, LEACH [3] outperforms classical clustering algorithms by using adaptive clusters and rotating cluster-heads, allowing the energy requirements of the system to be distributed among all the sensors. In addition, LEACH is able to perform local computation in each cluster to reduce the amount of data that must be transmitted to the base station. This achieves a large reduction in the energy dissipation, as computation is much cheaper than communication.

In WSNs there are basically two types of clustered WSNs that are homogeneous and heterogeneous [4]. Clustering is the key technology to extend the network lifetime through efficient resource utilization. The theory of clustering routing protocol in WSN is to divide the network into many small areas, and we regard each small area as a cluster. There are various clustering routing protocols based on homogeneity and heterogeneity of sensor nodes.

In order to increase the life- time of the WSNs the life time of sensor nodes LEAC [5][6] protocol can be used. LEACH protocol refers to Low Energy Adaptive Clustering Hierarchy, which is designed as a routing protocol for Wireless Sensor Network. Generally speaking, there are three steps in cluster routing protocol: the generation of cluster heads, the formation of clusters and the communication among clusters. Hence, LEACH algorithm also includes these three steps, just merging the first two steps into one that is the establishment of clusters and the communication among clusters. Thus, LEACH protocol algorithm contains the set-up of clusters and stable data transmission. As for the selection of

Cluster heads, LEACH adopts equal probability method, selecting cluster heads in a circle and random manner and distributing the energy of the whole network evenly on each node. Therefore, LEACH algorithm reduces the energy consumption, prolongs the lifetime of the net - work. The executive process of LEACH [7] is periodical, and each period includes the establishment of clusters and data transmission, and we call a period as a round. In order to conserve the energy, the duration of stable data transmission phrase is much longer than the time required for the establishment.

In WSNs once the nodes begin to exhaust their en - ergies (become dead), the tolerance of network goes on decreasing, until such a time when it can be no longer operational. The period of time from the beginning of network operation till the first node dies ma[y be referred to as the stability period. An endeavour to study this protocol that may increase the stability period of a WSN as one of the aims of increasing efficiency and life-time of the network to improve WSNs commercial viability. SEP [8] [9] is a protocol that assumes heterogeneity in a network, making it basis to prolong the stability period in a hierarchically clustered. It is a two-level heterogeneous proactive network protocol. SEP attempts to maintain the constraint of well-balanced energy consumption. It as - sumes that each node in the network has different energy. Therefore in SEP there are two types of nodes; normal nodes and advanced nodes. Advanced nodes have more energy than the normal nodes. SEP assigns weighted probability to each node in the network according to their energy level.

We assume that the sink is not energy limited (at least in comparison with the energy of other sensor nodes) and that the coordinates of the sink and the dimensions of the field are known. We also assume that the nodes are uniformly distributed over the field and they are not mobile. Under this model, we propose a new protocol; we call SEP, for electing cluster heads in a distributed fashion in two-level hierarchical wireless sensor networks.

Unlike prior work; SEP is heterogeneous -aware, in the sense that election probabilities are weighted by the initial energy of a node relative to that of other nodes in the network. This prolongs the time interval before the death of the first node (we refer to as stability period), which is crucial for many applications where the feedback from the sensor network must be reliable. We show by simulation that SEP provides longer stabil- ity period and higher average throughput than current clustering heterogeneous -oblivious protocols. We also Study the sensitivity of our SEP protocol to heterogeneity parameters capturing energy imbalance in the network. We show that SEP is more resilient than LEACH in judiciously consuming the extra energy of advanced (more powerful) nodes SEP yields longer stability period for higher values of extra energy.

VIII. METHODOLOGY

A. *Radio model*

We have assumed the same radio model which has been used in earlier works. For the radio hardware as shown in Figure. 6, the transmitter dissipates energy to run the transmitter radio electronics and power amplifier, and receiver dissipates energy to run the transmitter radio electronics. For the scenarios described in the project work, both the free space (d^2 power loss), and multipath fading (d^4 power loss) channel model were used depending on the distance between the transmitter and receiver, if distance is less than a threshold, the free space model is used; otherwise, the multi path model is used.

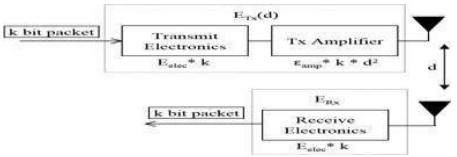


Figure 6. Radio Model

B. *Performance measures*

1) Stability Period: Stability Period is the time inter- val from the start of network operation until the death of the first sensor node. We also refer to this period as stable region.

2) Instability Period: Instability Period is the time interval from the death of the first node until the death of the last sensor node. We also refer to this period as unstable region.

3) Network lifetime: Network lifetime is the time in- terval from the start of operation (of the sensor network) until the death of the last alive node.

4) Number of cluster heads per round: This instanta- neous measure reflects the number of nodes which would send directly to the sink information aggregated from their cluster members.

5) Number of Alive (total, advanced and normal) nodes per round: This instantaneous measure reflects the total number of nodes and that of each type that has not yet expended all of their energy.

6) Throughput: We measure the total rate of data sent over the network, the rate of data sent from cluster heads to the sink as well as the rate of data sent from the nodes to their cluster heads.

IX. IMPLEMENTATION

Simulations are conducted using MATLAB R2018a and to get the plots, confidence interval is taken. Sensor nodes are deployed in random manner and made heterogeneous WSN using MATLAB. The wireless channel is used because the nodes deployed in the network are communicating wirelessly based on their distance, transmission range etc. Simulation show that SEP performs better considering metrics of throughput network lifetime and stability period etc.

The proposed work in this project is mentioned below in steps:

1) Deploy WSN by initializing the parameters.
2) After deploy the network worth-noting to use the appropriate topology for that network.
3) Selecting the cluster head in the sensor network.
4) Implementing LEACH.
5) Implementing SEP.
6) Evaluating the performance and observing the comparative analysis.

A. Simulation Results

We simulate a clustered wireless sensor network in a field with dimensions 100m 100m. The total number of sensors n = 100. The nodes, both normal and advanced, are randomly (uniformly) distributed over the field. This means that the horizontal and vertical coordinates of each sensor are randomly selected between 0 and the maximum value of the dimension. The sink is in the centre. The initial energy of a normal node is set to E0 = 0.5 Joules -Although this value is arbitrary for the purpose of this study, this does not affect the behaviour of our SEP protocol The size of the message that nodes send to their cluster heads as well as the size of the (aggregate) message that a cluster head sends to the sink is set to 4000 bits. In the next subsections we simulate the heterogeneous oblivious LEACH and our SEP protocol, in the presence of heterogeneity in the initial energy of nodes. We evaluate the behaviour of both protocols in terms of the performance measures defined in Section III. We also examine the sensitivity of SEP to the degree of heterogeneity in the network. We first summarize our general observations:

1) In a wireless sensor network of heterogeneous nodes, LEACH goes to unstable operation sooner as it is very sensitive to such heterogeneity.
2) SEP protocol successfully extends the stable region by being aware of heterogeneity through assigning probabilities of cluster-head election weighted by the relative initial energy of nodes.
3) Due to extended stability, the throughput of SEP is also higher than that of current (heterogeneous - oblivious) clustering protocols.
4) The performance of SEP is observed to be close to that of an ideal upper bound obtained by dis - tributing the additional energy of advanced nodes uniformly over all nodes in the sensor field.
5) SEP is more resilient than LEACH in judiciously consuming the extra energy of advanced nodesSEP yields longer stability region for higher values of extra energy.

1) Wireless Sensor Network simulation: As shown in Figure. 7 below, we simulate a clustered wireless sensor network in a field with dimensions 100m 100m. The total number of sensors n = 100. The nodes, both normal and advanced, are randomly (uniformly) distributed over the field where o represent no rmal nodes; + represent advanced nodes; x represent Sink; and. Represent dead nodes. Comparison between LEACH and SEP in the

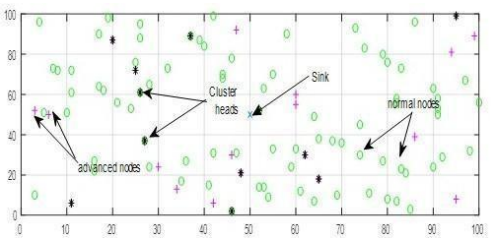


Figure. 7. WSN simulation

presence of heterogeneity m=0.2, a=3, where m=0.2 means 20% (20 out of 100 nodes) are advance no des and a=3 means advanced nodes equipped with 300% more energy than normal nodes.

2) Cluster heads in LEACH and SEP: LEACH fails to take full advantage of heterogeneity because normal nodes die fast, but advanced nodes in very slow fashion, because they are not very often elected as cluster heads after the death of normal nodes but in SEP the advanced nodes get more chances to be cluster heads than normal

nodes, thus stability period and number of cluster count is more in SEP as shown in Figure. 8.

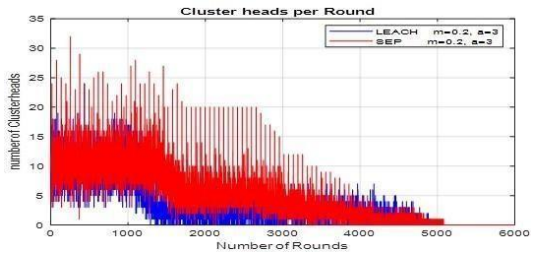


Figure 8 Cluster heads per rounds

3) Stability Period: In this subsection we compare the performance of SEP and LEACH in the same heterogeneity setting. The Figure. 9 shows that the SEP takes full advantage of heterogeneity. In LEACH the first node dies at 1034th round but in SEP the first node die at 1299th round the stable region (stability period) is increased significantly by 25.62%. The unstable region

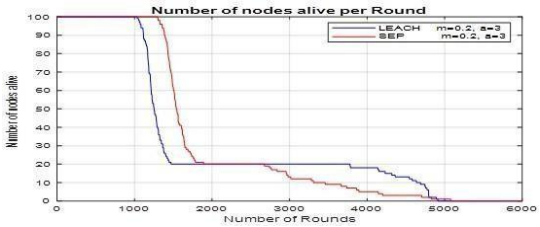


Figure. 9. Stability Period

(Instability period) of SEP is shorter than LEACH. In LEACH the first node dies at 1034th and last node dies at 4912th round but in SEP, the first node dies at 1299th round and last node dies at 5077th round. Thus in SEP the instability period decreased by 2.57%.

4) Pack ets to Cluster Head: SEP transmits more packets than LEACH as shown in Figure. 10below

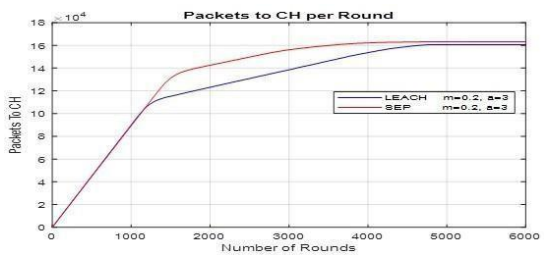


Figure 10. Packets to Cluster Head per round

5) Packets to Base Station: The cluster heads collect data packets fromnodes and then transmit these packets to base station or sink. As shown in below Figure. 11 the packets to base station transmitted in SEP are significantly high than the packet transmitted in LEACH

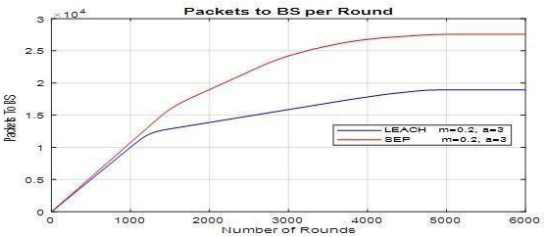


Figure 11 Packets to Base Station per round

6) Throughput: Throughput as shown in Figure. 12 of SEP is significantly larger than LEACH in stable region and for most of the unstable region .This means that because SEP guarantees cluster heads in more rounds then these cluster heads will report to sink.

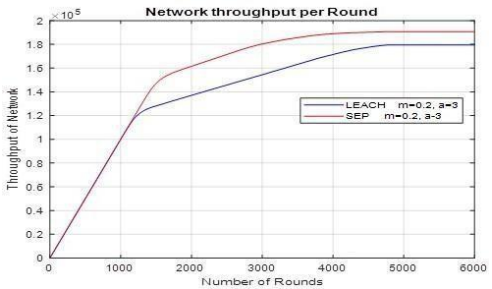


Figure. 12. Throughput of network

7) Energy Dissipation: The energy consumed by network in LEACH protocol is high with respect to SEP protocol as shown in below Figure. 13.

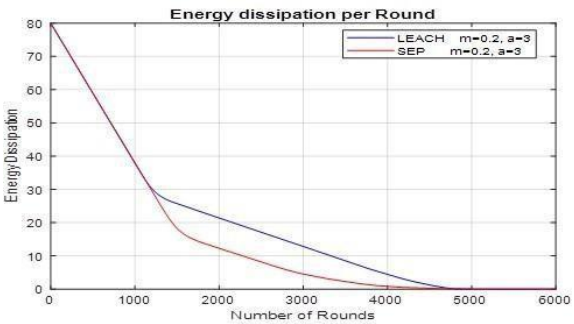


Figure 13. Energy Dissipation

8) Sensor network when all nodes are dead: When all nodes are dead as shown Figure. 14

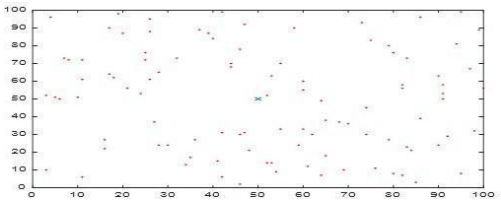


Figure. 14. Network when all nodes are dead

X. Discussion

In this paper we have surveyed the main approaches to energy conservation in wireless sensor networks. Special attention has been devoted to a systematic and comprehensive classification of the solutions proposed in the literature. We did not limit our discussion to topics that have received wide interest in the past, but we have also stressed the importance of different approaches such as data-driven and mobility-based schemes. It is worth noting that the considered approaches should not be considered as alternatives, they should rather be exploited together.

We can draw observations about the different approaches to energy management. As far as traditional techniques to energy saving, an important aspect which has to be investigated more deeply is the integration of the different approaches into a single off-the-shelf workable solution. This involves characterizing the interactions between different protocols and exploiting cross -layer interactions.

We have discussed the Low Energy Adaptive Clustering Hierarchy (LEACH) protocol and analyzed the protocol based on network lifetime, stability period and the net - work throughput. We have put light on the comparison of SEP protocol.

Acknowledgment

I express my sincere gratitude to the Maharaja Ranjit Singh Punjab Technical University, Bathinda for giving me the opportunity to work on the project during my final year of M.Tech. Project work is an important aspect in the field of engineering.

I would like to thank Dr. G.S. Lamba Coordinator, Baba Banda Singh Bahadur Engineering College, Fatehgarh Sahib for their kind support.

I would like to express my deepest appreciation to my supervisor Mr. Sukhwinder Sharma, Assistant Professor, Department of Computer Science and Engineering, Baba Banda Singh Bahadur Engineering College, Fatehgarh Sahib, who has the attitude and substance of a genius; he continually and convincingly conveyed a spirit of adventure in regard to research and scholarship, and an excitement in regard to teaching. Without his guidance and persistent help this dissertation would not have been possible. I cant say thank you enough for his tremendous support and help. I feel motivated and encouraged every time attends his meeting. Without his encouragement and guidance this project would not have been materialized.

References

[1] M. Vivek and R. Catherine *Homogeneous vs heterogeneous clustered sensor networks: a comparative study*, ICC, 2004.

[2] A. Giuseppe and C. Marco and D. Mario and P. Andrea *Energy conservation in wireless sensor networks: A survey*, Ad hoc networks, Elsevier, 2009.

[3] S. Sharma and R. Bansal and S. Bansal *Issues and challenges in wireless sensor networks*, Machine Intelligence and Research Advancement (ICMIRA), IEEE, 2013.

[4] S. Sharma and R. Bansal and S. Bansal *Heterogeneity-aware Energy-efficient Clustering (HEC) Technique for WSNs*, TIIS, 2017.

[5] X. Li and J. Wang and L. Bai *LEACH Protocol and Its Improved Algorithm in Wireless Sensor Network*, Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC), IEEE, 2016.

[6] H. Oudani and K. Salahddine and K. Mustapha and B. Kaoutar and E. Mohamed and K. Khaoula and E. Hicham and E. Lahoucine *Energy efficient in wireless sensor networks using cluster-based approach routing*, International Journal of Sensors and Sensor Networks, IEEE, 2017.

[7] W. Heinzelman and C. Anantha and B. Hari *Energy-efficient communication protocol for wireless microsensor networks*, Proceedings of the 33rd annual Hawaii international conference on, IEEE, 2000.

[8] I. Siddiq and S. Sandesh B and M. Bhagrithaya *Performance analysis of Stable Election Protocol and its extensions in WSN* , Advanced Communication Control and Computing Technologies (ICACCCT), IEEE, 2014.

[9] S. Georgios and M. Ibrahim and B . Azer*SEP: A stable election protocol for clustered heterogeneous wireless sensor networks*, Boston University Computer Science Department, 2004.

Intelligent Vision Extraction: A Review

Ramanjot Kaur
Assistant Professor
Com puter Science & Engineering
Baba Banda Singh Bahadur
Engineering College
Fatehgarh Sahib, India
ramanjot.kaur@bbsbec.ac.in

Sandeep Kaur
Assistant Professor
Com puter Science & Engineering
Baba Banda Singh Engineering
Bahadur College
Fatehgarh Sahib, India
sandeep.benipal@bbsbec.ac.in

Nishu Chaubey
Student
Com puter Science & Engineering
Baba Banda Singh Engineering
Bahadur College
Fatehgarh Sahib, India
nishuchaubey31@gmail. com

***Abstract*-In 1995 e ve ryone in te ch wante d a slice of the dot-com boom, but today, fie lds like artificial intelligence (AI), machine le arning (ML) and big data drive the te ch ve nture capital (VC) of the world to dig into the ir pocke ts. C ompute r vision is at the inte rsection of allthese data-drive n innovations. W hile use s for compute r vision are we ll -known within the te ch world, the te rm is still virtually unknown to the ge ne ral public, e ve n though many of the m are alre ady bene fiting from it.**

Keywords-intelligent, computer vision, artificial intelligence.

I. INTRODUCTION

Humans use their eyes and their brains to see and visually sense the world around them [8]. Computer vision is the science that aims to give a similar, if not better, capability to a machine or computer. As humans, we perceive the three-dimensional structure of the world around us with apparent ease [3]. Think of how vivid the three-dimensional percept is when you look at a vase of flowers sitting on the table next to you. You can tellthe shape and translucency ofeach petal through the subtle patterns of light and shading that play across its surface and effortlessly segment each flower from the background of the scene Figure. 1.

Figure. 1: T he human visual system has no problem interpretingthe subtle variations in translucency and shading in this photographand correctly segmenting the object from its background.

Computer vision is concerned with the automatic extraction, analysis and understanding of useful information from a single image or a sequence of images. It involves the development of a theoretical and algorithmic basis to achieve automatic visual understanding [1]. It is closely linked with artificial intelligence, as the computer must interpret what it sees, and then perform appropriate analysis or act accordingly. As a scientific discipline, computer vision is concerned with the theory and technologyfor building artificial systems that obtain information from images or multi-dimensional data. Other parts which sometimes are described as belonging to artificial intelligence and which are used in relation to computer vision is pattern recognition and learning techniques [6]. This technique has been used successfully to detect or to characterize abnormalities on digital images.

II. ROLE OF AI IN COMPUTER VISION

One of the most powerful and compelling types ofAI is computer vision (CV) which you've almost surely experienced in any number of ways without even knowing it. Artificial Intelligent (AI) can gather new facts, link scattered but conceptually associated data and gives higher-level algorithms essential for managing complicated systems [9]. It can be used for exploring information placed on SW through ontology that clearly determine and define what agents necessitate to know to execute a task. The combination of two innovative technologies is called Intelligent Visual Extraction (IVE). AI computer vision and image recognition is meant to achieve a specific goal by communicating with

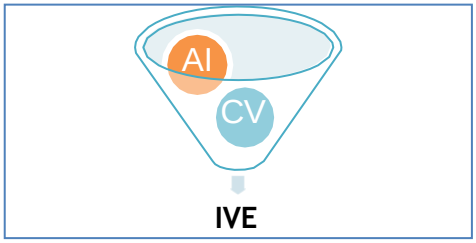


Figure. 2: Intelligent Visual Extraction (IVE)

by recognizing surroundings. Where computer vision gives it the sense of prospects, image recognition helps in understanding the physical universe around us. Another important part of AI is Visual observation and analysis. It understands to find visually exact and similar matches to an image from a database of images. AI is perfect in replicating the most natural way of providing perfect visual search results. AI can not only name the person but can also perceive their information. Thus AIdetects the information in person such as age and genre from a given static files.

III. APPLICATIONS OF COMPUTER VISION

With the rapid growth of technology the computer vision plays an important role in every field[1].

A. Image Motion Capture

A comprehensive survey of computer vision -based human motion capture literature from the past two decades is presented [5]. The focus is on a general overview based on taxonomy of systemfunctionalities, broken down into four processes: initialization, tracking, poses estimation and recognition until recently the computer graphics field has been mostly interested in visual realism3 and personalized human models, while the motion capture community has been more interested in spatial accuracy of the human models. We expect that the commercial interest in both fields will accelerate the development in human modeling and make the two fields approach and benefit from each other. The applications of human motion capture are numerous and it is expected that we will see a continuous growth in the resources devoted to this topic and hence that interesting new results in spite of everything will appear in the not too distant future.

B. Retail Security

Amazon recently opened to the public the Amazon Go store where shoppers need not wait in line at the checkout counter to pay for their purchases. Located in Seattle, Washington, the Go store is fitted with cameras specialized in computer vision. It initially only allowed Amazon employee shoppers, but later welcome to public beginning in early 2019s [5]. The technology that runs behind the Gostore is called Just Walk Out. As shown in this one-minute video, shoppers activate the IOS or Android mobile phone app before entering the gates of the store.

C. Image Reorganization

The advances and capabilities of image recognition that are now possible thanks to deep learning neural networks and cheaper and faster computing power have changed the game, but these brilliant algorithms are nothing unless they have access to reams and reams of picture data to train on [10]. From medical images to faces to hugs, the vast majority of pictures that could be used for training data is proprietary and in the hands of two behemoths: Google and Facebook. Considering the amount of sharing and uploading that's taking place on those companies' consumer platforms, this trend shows no signs of slowing down.

D. VR and AR need computer vision

Some believe virtual reality (VR) and augmented reality (AR) are over-hyped at the moment and that computer vision efforts being put into these technologies are a waste of energy [2]. However, most people I've spoken with believe both VRand AR are here for the long term and that, in order to advance, they will need high-quality computer vision capabilities (such as using image recognition to improve "interactive" forms ofVR).

E. Automotive Applications

Computer Vision could make a real difference in the automotive industry by adding important safety features to our vehicles. If a car could detect danger, it could stop before an accident happens and save countless lives and property. It is already being used in driver assistance systems to detect simple objects such as road lanes, traffic signs, and road boundaries. All autopilots include these functions, but are only usable in monotonous conditions, with little variation, such as highway traffic. Even if you don't buy a self-driving vehicle, if you care about safety, when evaluating a car, these are some of the features you should put on your mandatory list.

IV. COMPUTER VISION SYSTEM

A. StopLift

In retailsecurity specific to groceries, Massachusetts - based StopLift claims to have developed a computer-vision systemthat could reduce theft and other losses at store chains. The company's product, called StopLift, is a system that detects checkout errors or cashiers who avoid scanning, also called "sweethearting." Sweethearting is the cashier's act of fake scanning a product at the checkout in collusion with a customer who could be a friend, family or fellow employee[2].

B.Waymo

One company that claims to make driving safer is Waymo. Formerly known as the Google self-driving car project, Waymo is working to improve transportation for people, building on self-driving car and sensortechnology developed in Google Labs [3].

C.Slantrange

Slantrange claims to offer computer vision-equipped drones that are connected to what the company calls an "intelligence system" consisting of sensors, processors, storage devices, networks, an artificial intelligence analytics software and other user interfaces to measure and monitor the condition of crops. At 120 meters above ground level, the camera has a resolution of 4.8 cm/pixel. However, its website notes that flying lower provides better resolution [7]. *D.Mitek Systems*

Mitek Systems offers image recognition applications that use machine learning to classify, extract data, and authenticate documents such as passports, ID cards, driver's licenses, and checks.

E.Gauss Surgical

Gauss Surgical has developed blood monitoring solutions that are described to estimate in real-time blood loss during medical situations. This solution, the website reports, maximizes transfusions and recognizes hemorrhage better than the human eye [11].

V. CONCLUSION

Computer vision applications have emerged in more industries, although some have adopted the technology faster than others. Whatever computer vision technology exists continues to rely on the human element, to monitor, analyze, interpret, control, decide and take action.

REFERENCES

[1] Luciano M. Prevedello, Ross W. Filice, J. Raymond Geis, "Implementing Machine Learning in Radiology Practice and Research Marc Kohli", American Journal of Roentgenology. 2017;208:754-760. 10.2214/AJR.16.17224.

[2] Jay A. Baker, Eric L. Rosen, Joseph Y. Lo, Edgardo I. Gimenez, Ruth Walsh, Mary Scott Soo, " Computer-Aided Detection (CAD) in Screening Mammography: Sensitivity of Commercial CAD Systemsfor Detecting Architectural Distortion", American Journal of Roentgenology. 2003;181:1083 -1088. 10.2214/ajr.181.4.1811083.

[3] Carlos Ramos ; Juan Carlos Augusto ; Daniel Shapiro, (2018), " Ambient Intelligence: the Next Step for Artificial Intelligence", IEEE Intelligent System, Vol 23 , No 2, pp15-18.

[4] Vernon, David, (1991), " Machine vision - Automated visual inspection and robot vision", pp 1-26.

[5] T homas B.Moeslund, ErikGranum (2001), " A Survey of Computer Vision-Based Human Motion Capture", Computer Vision and Image Understanding, Vol 81, No 3, pp 231-268.

[6] Huimin LuYujie, LiMin ChenHyoungseop, (2018), " Brain Intelligence: Go beyond Artificial Intelligence", Mobile Networks and Applications, Vol 23, Issue 2, pp 368–375.

[7] Hashimoto, Daniel, (2018), " Artificial IntelligenceinSurgery: Promises and Perils", Annals of Surgery, Vol 268, pp70-76.

[8] Sunny T hukral, Vijay Rana, (2018), " VersatilityofFuzzyLogic in Chronic Diseases: A Review", Journal of Medical Hypotheses, Elsevier, Vol 122, pp 150-156.

[9] Sunny Sharma, Sunita, Vijay, (2018), " An OptimumApproach for Preprocessing of Web User Query", InternationalJournalof Informatics and Communication T echnology, Springer,Vol7, no1. pp 8-12.

[10] Sunny Sharma, Sunita, Vijay Rana, (2018), " A semantic framework for ecommerce search engine optimization ", International Journal of Information T echnology-Springer,pp 1-6.

[11] Siqi Nie, Meng Zheng, Qiang, (2018), " T he deep regression bayesian network and its applications: Probabilistic deep learning for computer vision", IEEE Signal Processing Magazine, pp 11-18.

Network Security Assessment using Penetration Testing in a Rural Banking Environment: The Case of Akuapem Rural Bank

Wellington Amponsah
Computer Science and Engineering Department
Shaheed Udham Singh College of Engineering and Technology
Tangori Punjab, India
amponsahbullet400@gmail.com

***Abstract*—Ghana's Banking sector has seen much improvement and this is evidence by the establishment of various commercial Banks to the current Rural and Community Banks, to S avings and Loans. Cyber-threats and the measures necessary to counter them in the Banking industry are the major security issue of the hour that the bank is facing. Information security issue is vital and more critical when it wants to implement and apply IT into its operations. Penetration testing uses principles as hackers to penetrate network system to verify presence of vulnerabilities. This research is to show how to perform and identify the security strength and reveal the vulnerabilities, and possible exploits in the security of the internal network and misconfiguration of firewall of the Bank using penetration testing.**

Keywords—network security, cyber security, penetration testing

I. INTRODUCTION

Ghana's Banking sector has seen much improvement and this is evidence by the establishment of various commercial Banks to the current Rural and Community Banks, to Savings and Loans. The Rural and Community Banks (RCBs) were first established in Ghana in 1976 to provide banking services to the rural population, providing credit to small-scale farmers and businesses and supporting development projects. These Banks are locally owned and managed [1]. In 1981 about 30 existing RCBs formed an Association of Rural Banks (ARB) to serve as a networking forum and as a network of institutions sharing a common mission, the ARB promoted and represented the RCBs and also provided training services to member RCBs [1]. The Rural and Community Banks are supervised by the clearing ARB Apex Bank under the regulation of the Bank of Ghana, which owns shares in the Banks [1][2]. According to [3] the country currently has 144 Rural and Community Banks across the country with their respective branches spread across the country. [4][5] Rural Bank is a unit bank which is limited by liability under the company's code which can be defined as rural financial institution/cooperative/ community bank which is to mobilize savings that provides customized financial services to rural communities within its radius of operation.

Technology has become the main driving tool for our national development and now applied in all forms of our lives including the financial, health, educational sector, etc. In information technology data protection or information security is one of the great challenges for the world. In the Banking industries network and data security is one of the serous issues that are inevitable.

Ghana is now maturing in the field of cyber security and this has been described by World Bank in collaboration with the Global Cyber Security Capacity Centre (GCSCC) as being at the formative level. A finding fromthe assessment is evidenced in the implementation of a number of initiatives including the formal adoption of a National Cyber Securit y Policy & Strategy (NCSPS) in 2016, and the National Cybersecurity Institutional Framework (NCIF). In addition, a National Cybersecurity Advisor was appointed to coordinate cybersecurity issues in government and across non - governmental sectors; the National Cyber Security Inter-Ministerial Advisory Council (NCSIAC) and the National Cyber Security Technical Working Group (NCSTWG) were also inaugurated. The establishment of the National Cyber Security Centre (NCSC) to oversee and coordinate national cybersecurity programs, the development of the Computer Emergency Response Team (CERTs). Ghana is one of the countries in Africa where the internet and the use of technology are now penetrating all aspects of our public lives. In the Ghanaian banking industry, Information technology has now been recognized as the life wire of the Banks since it simplifies and supports the performance which has made it very demanding of the Banks to adopt in full Information Technology in her operation for customer satisfactions and the reputation and performance the financial market. And this demand for a high network and information security in the banking services to protect and secures its systems to maintain information vital to its operations.

II. PROBLEM STATEMENT

Information security issue is vital and more critical when it wants to implement and apply IT into its operations. The demand for the application of information communication and technology in the banks is to create more business and ease of operations. Inadequate security measures and insecure network of the Bank affect the beliefs of customers and their ability and wiliness to transact any kind of transaction or business with the Bank. And it is a major threat to the growth of the banks in the country as a whole. The security of customer information is very vital information that is treated as a valuable. Hackers with information about the network of the Bank, staff information and even customer information can perform many transactions in the name of these staff and customers. Some information which can be stolen includes their date of birth, social security number, tax identification numbers, address, staff login details, etc.

Cyber-threats and the measures necessary to counter them in the Banking industry are the major security issue of the hour that the bank is facing. This is because there have been a

number of attacks on the various networks in the various Banks which are done in an organized and in a very dangerous way and this calls for a serious attention and considerations.

Priority and attention have not been given to security of the information system of the bank. Hacking attacks on Bank's network in Ghana recent days has increased drastically in a very sophisticated manner. The attacks on the bank's information system have become more frequent in a well-organized and a more dangerous way. Protecting the digital assets of the bank has not been a priority or concern and the access of this information by hackers will affect the reputations and also compromises intellectual property of the Bank.

This is because ports on the network have been left open since they really on firewalls or have no knowledge about the various ports that are opened on the network. These ports that are opened are being scanned by hackers and also exploits for vulnerabilities which are the means of locating and identifying specific weaknesses in the network of the bank and services that are being run. The Bank becomes vulnerable to hacking because they do not employ secure information security systems, as well as intrusion and detection systems to protect their data.

IT staff of the bank have no knowledge or background in security and few checks of the various vacancies declared by the banks in the country only requested for people with BSc/HND Computer science, IT, MIS, MBA background. None of the RCBs has a staff dedicated to cyber security or network security to handle security issues of the bank.

Technologies are constantly evolving and growing, at a rate that is so rapid that one can have a difficult time keeping up. Antivirus that is aimed at protecting one computer is often found to be out of date. Most of the RCBs use free and outdated VPN.

III. OBJECTIVE

This research is to show how to perform and identify the security strength and reveal the vulnerabilities, loopholes and the possible exploits in the security of the internal network and misconfiguration of firewall of the Bank using penetration testing techniques approach. This will enable identify security flaws, to understand the level of risk and vulnerabilities and exploits to secure the network or important data from outsiders like hackers, who can have unauthorized access to the application and exploit the network to access sensitive information if any kind of vulnerability is identified within it. It is to also provide evidence for the support to increase investments in the security personnel and technology

IV. ABOUT THE AKUAPEM RURAL BANK

Akuapem Rural Bank was established in 1980 for helping the rural dwellers. The Bank was awarded for climbing up the ladder and reaching the 64th position in the Ghana Investment Promotion Centre (GIPC) Ghana club 100 awards. The bank also won the maiden Association of Rural banks awards for the best Eastern region RCB. The goal of the Bank is to become the best Rural Bank in Ghana and has its corporate values Honesty, responsibility, dependability, and motivation. As part of government policy, the Bank has the main data center at Apex Bank. However, the bank also has its own internal Data center which is linked to the main data center at Apex Bank. The Bank has an application called Jboss services deployed on its server and other branches. The Bank's data center is connected to the Apex Bank data center through Wide Area Network and the communication channel is through Multiple Layer Switching (MPLS) radio and Very Small Aperture Terminal (VSAT). The Bank uses wireless communication across all the offices and its surroundings. Customer can access online account balances by the use of U-connect application we can be downloaded from google play store. Bank does most of the network and system security with regards to storage and WAN. The SUSU application is developed by the bank. The network is connecting through Comsys Ghana as ISP to Apex Bank. Comsys provide Data communication and internet communication. The Data center connects to Apex for its banking application and other auxiliaries' product whiles the internet connects to the outside world.

V. LIMITATION OF THE STUDIY

As part of this research, the objective is to present how network penetration testing of the Bank can be done. The penetration testing procedure shall not involve attacking the Bank's network structure and security system. This is only to help IT staff to be able to undertake penetration testing to test how secure the network of the bank is, adopting the penetration and testing approach to identify vulnerabilities in the system.

VI. NETWORK SECURITY

[6]Network security is the monitoring of network access and prevents unauthorized use of network resources and the security tasks are managed by the administrators of the network. Network security consists of the policies and practices adopted to prevent and monitor unauthorized access, misuse, modification, or denial of a computer network and network-accessible resource. It protects the usability and integrity of your network and data. It includes both hardware and software technologies. Effective network security manages access to the network. It targets a variety of threats and stops them from entering or spreading on the network. Network security combines multiple layers of defenses at the edge and in the network. Each network security layer implements policies and controls. Authorized users gain access to network resources, but malicious actors is blocked from carrying out exploits and threats. It is prudent to know that the best method of understanding the network security posture is to try to defeat it. The administrator will place himself in the mind -set of an attacker, and deploy techniques using penetration testing against the network.

VII. CYBER SECURITY

Simply put, cyber security is the process whereby you protect yourself online, as well as your entire online presence [7]. It consists of programs that you install on your computer, such as antivirus software or a virtual protected network (VPN), and practices that you may employ on a day-to-day basis, such as guarding your usernames and passwords or keeping a cover on your webcam. Cyber security is intended to protect individuals, companies, computers, networks, programs, and data from unauthorized

access of their sensitive information or corrupting files such as viruses, worms, or Trojan horses. Cyber security does not take a one-size-fit all approach. What works for one computer system may not necessarily provide full protection to another. Technologies are constantly evolving and growing, at a rate that is so rapid that one can have a difficult time keeping up. Antivirus software that may have protected an older computer that you had five years ago may not protect you adequately on the computer that you have now. An encryption program or VPN that promises to keep you safe online may leave you exposed to undetected threats, possibly those originating in other countries

V II I. PENETRATION AND T EST ING

Penetration testing refers to the testing of a network system, cyber system, or application to detect weaknesses that may be exploited by a malicious hacker. The term "penetration testing" refers to testing the security of a computer system and/or software application by attempting to compromise its security, and in particular the security of the underlying operating system and network component configurations. According to [ken] the de facto stewardship of penetration testing tools and processes by IT Security organizations is significant which has afforded IT Security people with the opportunity to test a computer and application's configurations from afar, as a sort of independent audit function. During penetration and Testing, one is essentially trying to gain access to a system without having any usernames or passwords or the credentials of the system and obtain vital inform. The aim is to see how easily it will be for one to obtain confidential information about an organization, and then increase the security of the network system or the information security system that is being tested. In simple words, penetration testing, also known as pen testing is the process of testing the network and other applications for vulnerabilities. The main purpose of this test is to secure the network or important data from outsiders like hackers, who can have unauthorized access to the application and exploit the network to access sensitive information if any kind of vulnerability is identified within it. Generally, vulnerabilities are introduced by accident during configuration of the network security, development and implementation of the system and applications. Common vulnerabilities include configuration errors, application bugs, and design errors. Testers use different sophisticated tools and advanced knowledge of IT to identify the behavior of an attacker, who penetrates the client's network and it's applications to obtain information and access to higher permissions without proper authorization. Penetration testing tools are used to identify standard vulnerabilities in the application. These tools will scan code to check whether there is any malicious code present in the system network by examining data encryption techniques and figuring out different hard -coded values like username and password.

A. Types Of Penetration Testing

The type of a penetration test selected depends on the purpose and scope of the company and organization, whether they want to simulate an attack by an employee, a network admin, or external sources. Generally, there are three different types of penetration testing namely Black box, white box and grey box penetration testing. In black-box penetration testing, the tester is not provided with much information about the application he/she is going to test and it is the tester's responsibility to gather information about the target network, system, or application.

In white-box penetration testing, the tester will get complete information about the network, system, or application along with the source code, OS details, and other required information. It can be considered a simulation of an attack by internal sources. In grey box penetration testing, the tester will have the partial knowledge of the application or the network system and it can be considered as an attack by an external hacker, who had gained illegitimate access to the Bank's network infrastructure details. [10]The test team is given information such as Hostnames, Few IP addresses and whether senior management can connect to the network remotely. The Penetration Tester is given enough common information a normal user can know. It is the combination of both the white box testing and the black box testing approach during penetration testing. The Penetration Tester is provided some basic details of the target; however, internal workings and some other privileged information is still kept from the Penetration Tester. Real attackers tend to have some information about a target prior to engaging the target. Most attackers do not choose random targets. They are motivated and have usually interacted in some way with their target before attempting an attack. Gray box is an attractive choice approach for many conducting Penetration Tests because it mimics real-world approaches used by attackers and focuses more on vulnerabilities rather than reconnaissance.

B. How Penetration testing can be done

The process of penetration testing follows a globally accepted approach and methodology. In other to perform the penetration testing on the network, the following steps but not limited to, should be adopted as explained below.

a) Reconnaissance

This is the first stage of penetration testing which involves us defining the scope and goal of the test, the systems to be addressed and the methods to be used. [8]Reconnaissance defines the target environment based on the scope of work and once the target is identified, research is performed to gather intelligence on the target such as which ports are in use and the services being run. The network diagram, IP Address, Devices on the network, applications and the services and their version, security defense systems in place such as Intrusion Detection System (IDS) and Intrusion Prevention Systems (IPS)

b) Scanning

[9]Scanning is a method for bulk target assessment. To discover the live IP addresses in the network, to discover the open ports on the machines, to fingerprint the services and to detect the vulnerabilities which is done by the vulnerability scanners. The fundamental goal of scanning is to identify potential targets for security holes and vulnerabilities of the target host or network [10]. There are many tools that are available for scanning our network such

as ncat, and nmap. The most popular scanning tool is network mapper which is called nmap. The nmap tool can be used in combination with other tools such as metasploit framework to determine the available ports, services, the services type and their version. Nmap is able to identify even if a host is alive or dead even if ICMP is completely down on the network Nmap is free in kali Linux and can also be used in windows. It is important to know that the type of scanning to perform determines the type of scanning such as scanning and bypassing firewalls will require

c) Identification of system vulnerabilities

After the scanning of all available devices on the network is completed the pen tester undertakes vulnerability assessments which are to identify vulnerabilities in the system which can be exploited. Vulnerability assessment is the process of identifying the vulnerabilities in a system which is conducted on behalf of the organization [11]. Threats available in the system will be identified. Vulnerability phase is started after some hosts are identified via nmap scans or other scanning tool after the reconnaissance [10]. One of the best tools for vulnerability scanning is Nessus, Nikto, Metasploit and Open Vulnerability Assessment System (OpenVAS). It is an open - source vulnerability scanner which employs Nessus Attack Scripting Language (NASL). Finding the vulnerability allows the users to access complete information on the network

d) Reporting

The final phase of the penetration testing is reporting. This is the stage where all the findings are conveyed to the managers in a very meaningful manner. They are being made aware of what the good things and the bad things so they are able to improve their security. This report must be in a clear language which a non-technical staff can easily comprehend. Kali Linux has several tools that are available to deliver information found during the penetration testing.

I X. CONCLUSION

It is recommended that the Bank runs a regular and consistent penetration testing of the network security and relevant applications. The penetration testing activity is to prevent hackers or intruders from tempering; compromising of information resources becomes extremely difficult in large networks and even impossible when attacks are launched using 0-day vulnerabilities. It is hereby conclude that, it is very necessary to ensure that information security breaches are detected as early as possible. Threats should be detected in a timely manner that will put the system always ahead of hacker's attack and there must also be the need for a prompt response which is to enable the Bank to prevent and mitigate the damages that may be caused by attackers of the system. This is to aid check how well the network is secured against extremely trained hackers who are attacking the network with determined stealth, as well as help train the information security team to identify attacks and react to them in real-world conditions.

REFERENCES

[1] GT samenyi Matthew, and Shazard Uddin " the caes of rural banks in ghana: corperate Governance in Less Developed and Emerging Economie*s* ". Emerald Group Publishing. pp. 311 334. ISBN 978-1-84855-252-4

[2] BAkwasi A. Boateng, Kennedy Annoh-Koranchie, andmEric Hayford " An Appraisal of rural and community banks in Ghana", www.iiste.org ISSN 2224-607X (Pper) ISSN 2225-0565 (online) Vol.6, No.6,2016

[3] Bank of Ghana report , " List of rural banks". https://www.bog.gov.gh [Accessed: Januay, 2019 online]

[4] Alok. Kumar, and Alimany Contte, "Financial and accounting manual for Rural banks," https://www.microfinancegateway.org/sites/ [Accessed: January,2019]

[5] R. Nicole, " Title of paper with only first word capitalized," J. Name Stand. Abbrev., in press.

[6] Y. Yorozu, M. Hirano, K. Oka, and Y. T agawa, " Electron spectroscopy studies on magneto-optical media and plastic substrate interface," IEEE Transl. J. Magn. Japan, vol. 2, pp. 740 –741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].

[7] Waleed Al Shehri " A Survey On Security In Wireless Sensor Networks" International Journal of Network Security & Its Applications (IJNSA) Vol.9, No.1, January 2017 DOI: 10.5121/ijnsa. 2017.9103 25

[8] Georgia Weidman " Penetration Testing: A hands-on introduction to hacking"

[9] Chaitra N. Shivayogimath " An overview of network penetration testing". IJRET eISSN: 2319-1163 | pISSN:2321-7308

[10] T imothy P. Layton, Sr. " Penetration Studies–A T echnical Overview". URL:http://www.sans.org/readingroom/whitepapers/testing/penetratio n-studies-technicaloverview-267

[11] Vulnerability assasment www.en.wikipedia.org/wiki/Vulnerabilitya ssessment

[12] Simon Parkinson and Andrews Crampton " Guide to Vulnerability Analysis for Computer Networks and Systems"

Perspectives of Green Computing

Taranpreet Singh
Computer Science and Engineering, Student
Baba Banda Singh Bahadur Engineering College
Fatehgarh Sahib, Punjab
Taranpreets1999@gmail.com

Arshdeep Singh
Computer Science and Engineering, Student
Baba Banda Singh Bahadur Engineering College
Fatehgarh Sahib, Punjab
Arsh199820@gmail.com

***Abstract* - Green computing is the study of minimizing the hazardous impact on environment and produces maximizing energy efficiency and green IT in the environment. Green computing has been in the focus since past and present years by public and private organizations. This review based study focuses on introduction of Green Computing, and describes historical and current issues and efforts made to progress in this field along with the hope for future in the field. It also mentions challenges and future works of Green Computing and its impact and its struggles for the future along with possible solutions.**

Keywords-Green Computing, Past, Present, Future Challenges.

I. INTRODUCTION

Over the past years, the debate of green technology along with the green computing is increasing day by day and is gone a concerning issue for both governments and businesses .Similar to the security issues, which IT architects were not giving proper attention and eventually they faced the consequences for it. The sustainable environmental design is also going to quickly transform into an unwelcomed architectural issue for modern projects. To be environmentally sustainable is recognized globally as the 21st century important problem. As such more IT stakeholders are moving exponentially towards green IT making the economy green as a whole in return as there also exists studies that proves the fact that the organization, which is strived for green computing and Strived for green computing and achieved it, also have resulted in economic boost [5].

Green computing is where we practice along with studying the lifecycle of computing or studies and practices where computing is achieved to be sustainable environmentally. Green Computing focuses on recyclability of products when they are produced with less hazardous materials and consume less energy during their lifetime.

When we want to increase performance of level of the computing, we face a huge challenge from processors with high-performance. The information Technology systems are based on the following mixtures:

- Networking,
- Hardware, and
- People.

Green computing comes with an objective of continuous improvement in the performance of computing with minimum consumption of energy and carbon footprints. And this is among the main objective of any Green Computing study. Green computing studies deals with manufacturing, using and then efficiently recycling the information technology equipment without leaving any negative impact on the environment. Green Technology is a continuously evolving technology and is not just a limited term for reducing the power consumption in the field of information technology. [4] Its development goals focus on following ideas:

- Creation of reusable products.
- Meeting the present as well as future society requirements.
- Applying production and consumption patterns which shall reduce materials to be wasted.
- Providing alternatives to health and environment damaging technologies.
- Forwarding towards viable economy and technology.

This concludes that the green computing is result of the following within the green form [1]:

- Use
- Design
- Disposal
- Manufacturing
- Energy Efficiency
- Technologies

The goal of green computing reduces the use of hazardous material, maximize energy efficiency during the products lifetime, and promote the recyclability of products and factory waste [6]. Research continues into key areas such as making the use of computers as energy- efficient as

possible, and designing algorithm and systems for efficiency- related computer technologies [17].

Figure.1: Main Goals of Green Computing

A. Green Concept in other Fields of Studies

When it comes to green technology, the computing or information technology is not the sole-dependent for it.[4]

It is but also widely used in other fields and as a result, we can see in the market products like:

1. Designing Of Hybrid Cars
2. Manufacturing Of Cars With:
   - Better fuel economy.
   - Lower emission
   - And use of natural materials.
   - Portable Solar Cells,
   - Making sea water able to use.
   - A Windy Day (Efficient energy generation from Wind Turbine),
   - Rapid Waste Treatment (Sewage treatment),
   - Wave-After-Wave (Harnessing ocean waves'),
   - Thermal Oceans (Harnessing different ocean water temperatures to generate energy).

Energy solar cells -Generate different types of energy levels and energy cells used in green technologies and it produces less amount of energy efficient in the in our environment,

3. Some Other Subject Areas of Green Technology. Some of the other areas where Green Technology is implemented are mentioned below [20]:

- Energy – studies about alternative fuels, and new means of efficient energy generation.
- Environment friendly purchasing preferring the purchase of products which will cause in lowest environmental impact.
- Green Building – focuses on studies about buildings from its location to choosing its materials.
- Green Chemistry – producing chemicals where hazardous substances' use and generation is eliminated or reduced.
- Green nanotechnology – the implementation of green chemistry and green engineering in manipulation of materials at a scale of one billionth of ammeter.
- Green IT fields - the implementation of green IT fields produces various types of products in industries with less harmful impact on environment and also maximizing the energy efficient in different IT fields[5].
- Green Eco- System - the implementation of green eco- system using different types of products and different fields in our green environment.

B. Material & methods

Following section will explore the past, present and future of green computing.

a) Green Computing Past

Without discussing different generations of the computing devices, computer's history remains incomplete. Further the history is equally connected with the struggle for energy efficient computing devices [8]. Each generation of computers has always brought efficiency to the technology usage into the field, such as 486-DX processor that was introduced with support for power management. Alongside, standards for Advanced Configuration & Power Interface (ACPI) with its first version were officially released. These evolutions over time resulted in more productive and at the same time, less energy eaters. There was an age when circuits were known as vacuum tubes, and memory were known as magnetic drums, and computers were huge in size [9]. It was the age when expensive operation was the first click in the mind of one who would think of acquiring a computer. It used high amount of electricity with

heat for even the machine itself as it caused malfunctions due to it. Then with an all-round development of technology and electronic switches were based on transistors and diodes. Latter on the second generation started with faster, cheaper, smaller, and greater efficient to the energy. This caused the computers also to be more reliable.

b) Efforts on the Green Technology

Energy Star in 1992 started to make products which would go to sleep when not in use and hence consuming less energy when even they are in standby mode. They also are internationally recognized standard for Energy efficient equipment. Business giants like Dell, IBM etc. have also been struggling for a greener concept of technology usage.[8] Below are some of the business giants outlined for their work in the field. As in year 2009 Dell took first position in Corporate Sustainability Index (CSI) Benchmark Report leaving the 2nd placed firm 52 point behind. Google had been committed to be environment friendly since 2007. They also claim to be using half the power compared to the average consumption of the industry. Google along with Intel in 2007 also initiated the Climate Saver Computing (CSC), a free group was purposed to show that reducing the emissions is beneficial and to promote the smart technologies. [8]IBM controls use and conserving of energy and water through consultation about the same and the implementation of Lean Six Sigma principles.

c) Efforts in The Past

In these types of efforts made not describes the approaches of sustainable development in green environments. The lack of quality approaches and depreciable hardware was also a major hurdle in advancement in this field. Some types of researchers are compiled a report on the America's market in 2008 and mentioned that 25 big companies, were struggling for a greener environment [14] and all these companies are from several different industries, such as auto industry, banking institutions, computers giants, consultants, food, garments and stock exchanges.

B. Green Computing Present

There are growing numbers of activists in the market trying to tend towards the green information technology (IT) despite the fact that IT Greenhouse Gas (GHG) Emission share is 2% of the total global GHG emission. This term means that IT should become a helping hand in reducing the GHG emission globally [19].

a) Cloud, Virtualization

Primary reduction in the energy consumption is shifting towards virtualization. The use of virtualization along with that of Cloud computing is present in our lives and plays an important role regarding green computing concept. And so it will not be fair with the green technology to exclude it from discussing here. Efforts of moving to cloud and using computers or servers of others for your computing requirements rather than purchasing a dedicated machine is considered a greener option in today's era. Similarly there are people shifting towards the virtualization concept whereby they virtualize their several computers or servers on a single physical machine and save their costs along with the energy to an amazingly noticeable amount. The move to cloud computing by Microsoft reduces from 30 to 90 percent per user the energy use and its carbon footprint. Microsoft also facilitates the hardware recycling and donation [12].

b) Educating the public

More produces green technology products is used to easiest way to studying in green environment [20]. Along with the growing literacy in this field, some still do not understand quite well the Green IT term, which includes professionals and users of this field equally. User will uses more Green IT products in educating fields in public places.

c. Greening Computing

Present era is the one where no one thinks living without computing devices. While on the other hand the producers are trying to introduce products to use less of the energy with high productivity in return, like making of software to best use the available energy of the hardware and give results for longer and accurate. The parameters considered for Information Technology to be Green are reduction of waste, efficient usage of energy and its consumption [15].

Environment issues:-Sitting in front of computer for more than five hours increases the chances to get diseases like:

- Depressions,
- Pain in different muscles of the body,
- Headaches,
- Joint pains,
- Eyestrain,
- Stress

Resulting, the employees will be no more productive [10].

d. Green Computing Future

To prevent computing to have any negative effect on humans lives in future, researchers had advised set of goals to follow. Have our digital infrastructure designed optimally so that to reduce the overall energy consumption. Sense the surrounding in a manner to keep you aware of energy consumption around the globe and its effects. While traditional computing systems will be important to execute the models on it for future, we need to predict and react to the upcoming events by responding to the future behaviours by modelling our predictions.[19]using alternatives our daily work sine digital environment like digitally reading the news personable, eliminate physical disks and use downloaded materials for materials in audio and video online shopping instead of physical onsite shopping etc.

a) Virtualization and Cloud Computing

By receiving noticeable attention that the cloud computing provides the services and solutions by using the existing infrastructure by having full use of the resources, it is contributing to green computing and is the future of the Green Computing[12].It is the greenest option to think about other's server when there is need of purchasing your specific server arises. Cloud computing is in concept of virtualization and for the virtualization technologies; companies like VMware are struggling and progressing towards the green concept in the daily computing requirements. Hence the more virtualization and cloud shift paradigms are adopted, more we will follow the green path as it will reduce the energy till up to 80%.

b) Desk tops versus Laptops

Individuals use laptops for their day to day work and even for their office work. The companies should be shifting towards the use of laptops instead of desktops for their employees who do not have heavy use from computers. This should be done because the desktop PCs use about six times higher energy than a laptop. This may not seem to be huge saving, but still every drop counts.

c) Dynamic Power Scaling and Smart Standby

Energy can be saved by green power scaling using green technologies on green networks.[6]Similarly the dynamic power scaling is the concept whereby the device will enter into a low power state whenever there is no transfer of data and will all of sudden go to the full power state when data packets are started transferring through it – this concept adoption will be a greater achievement in the greening of the environment for the computing.

C. Challenges of Green Computing

Computers use a major portion of available energy to the world. While this era has more need of computing in daily life, therefore this has become the major challenge for the field. This is one of the reasons; that the focus of green computing, is shifting from efficient computing to reducing the IT infrastructure and equipment associated costs as the natural resources are being consumed greatly. For the same purpose major IT companies are focusing in developing machines. While these issues are regularly observed by field specialists, following are still many areas which needs to be focused and some challenges faced [17].

- It is used to control the cooling equipment and other increasing the power sources and increasing the challenges of green computing.
- Equipment for heat removing also requires to be controlled as this is also a major increase in the total power consumption.
- Proper disposing and finding ways to proper recycling of the electronic items also is an issue, which needs to be resolved.
- New Optimization Techniques are needed in Performance-Energy-Temperature aware Computing [16].

a) Green Networks

The networks of today are increasing heavily and so are their energy consumptions. [12] Due to the consumption requirements, there rises the requirement of large number of devices at every provider's space. And the providers for the requirement of cutting cost and saving energy, welcomes all the innovations in this field. As the green networking is still at the beginning and is interestingly open issue for the technologists [12]. The most important issue of computing is the growing about energy conservation in computing activity. It is required to get the maximum benefit by getting a centre point in temperature, energy, and performance, by maintaining adequate designing techniques to have a trade-off among these.

II. RESULTS AND FINDINGS

Society is interested in changing towards the sustainable practices in using energy. And as such communities have introduced incentives for doing so in order to more facilitate the green practices.
[18] Researchers are working on the increased issue of reducing the space between power and data centres and the cooling systems. The companies like HP and Dell are trying to overcome challenges by introducing green computers, like HP rp5700. HP rp5700 PC is expected to have 5 years life at least, have almost all its materials recyclable. Even Dell produces greener computer ever than they were producing 10 years before. IBM in the same way is not behind any one, and is working on a solar cells technology to develop it cheaper and efficient than ever and will be sustainable at the same place.

Organizations should adopt ways to find out cooling methods and equipment which consume less energy and provide great cooling, as in this way the cooling costs will be reduced greatly. Also the temperature level of the cooling devices should be maintained in a way that the datacentres will be cool, but the cooling system will not be consuming extra energy. This can also be achieved by building the datacentres on cooler areas, rather than on hot areas, so that there will be less need of cooling devices and energy will be saved.

Organizations and institutions should adopt technology which are built on green concepts, as in such case the computing technologies, will not be producing the heat, and equipment for removing of heat will be less required. There will be a helping hand in the green technology initiative, if the computing devices once purchased are made full use of it, and should be kept using to its maximum life before disposing off or recycling. Also at equipment purchase, the one which can be easily recyclable should be looked for so that there is no worry of disposing the materials. It will be much better if there is agreement set with the seller to provide recycling at the end of life of the equipment.

While it is also the responsibility of organizations and research and educational institutions to provide opportunities and encourage people to work on the field and explore for ways in the green computing or green technology, in order to overcome the challenges of the research areas of the green technology.

III. DISCUSSION AND FUTUREWORK

Green technology is constantly growing and green computing is becoming crucial part of life, and we learned that information technology is not just an idle observer but is contributing activ ely in the concepts of green technology or green computing [8]. .IT is from its beginning always involved in efforts to reduce the energy consumption and currently also treasonable vela long with the progress made by other field so studies. Therefore being involved in IT, it becomes the duty of any one, to understand and follow the principles of green computing to be able to be survived in the future. IT industry can adopt different techniques like going for virtualization of servers, moving to cloud, using laptops in place of desktops, even virtualizing the networks and the like. There are the challenges incompletely implementing methods where we can get more production out of less energy consumption. There are more areas which still need to be observed deeply by the researchers and we consider it also the responsibilities of institutions to facilitate and encourage the researchers to overcome the exploration of the areas. And the researchers should look into methods to overcome these challenges so that proper implementation of green computing is made possible. Greening the work environment to bend your computing technology towards green concept, we need to follow some preventions like:

- Optimizing the environment including the workstation to feel more comfortable for the employees
- Guide the employees to be in a suitable posture to prevent any physical harmful
- Give your employees frequent breaks from the computers
- Create such an environment, where employee may sometimes look away from the computer screen; this will prevent eyestrain
- Typing with light touch
- Avoid strict deadlines on computer related tasks to avoid high stress situations.
- Adopting such techniques can lead to improvements in the workers' health and resulting in greater productivity and less energy consumption [19].

REFERENCES

[1] Galant "Green Computing: Barriers vs Benefits," https://medium.com/the ultimate-guide-to-green-computing-[Accessed 03 05 2015]
[2] "Environmental Sustainable Infrastructure Design," T he Architecture
[3] "Information SystemsInnovation for Environmental Sustainability," MIS Quarterly, vol.34, no. 1, pp.1-21, March2010.

[4] "Harnessing Green IT : Principles and Practices," IEEE Computer Society,pp.24-33
[5] "Green IT : Helpingto Create a Sustainable Planet," IEEE Computer Society, Sydney, 2011.
[6] S. Mittal and Z. Zhang, "Palette: Apache Leakage Energy Saving T echnique ForGreen Computing".
[7] S. Vivekananda and N. Subhash, "Services in Green Computing,"2014.
[8] "What is Green T echnology?" Green T echnology, [Online]. Available: http://green-technology.org/what.htm.[Accessed 13 052015].
[9] "Computing Now Archive," IEEE Computer association and Society, [Online].Available:[Accessed13-5-2015] http://www.computer.org/web/computingnow/0511/links.
[10] T . Ewer, "How to Stay Healthy and Productive at Your Computer" https://managewp.com/how-to-stay-healthy- and-productive-at-your-computer.[12 10 2012]
[11] "Microsoft Environment," Microsoft, [Online]. Available: [01 05 2015] http://www.microsoft.com/environment/.
[12] "T he Potential Impact of Green T echnologies in Next -Generation Wireline Networks: IsThere Room for Energy Saving Optimization?" IEEE Communications Magazine.
[13] "History of Computing," It -history.net, [Online].Available:[Accessed 2 5 2015] http://it -history.net/index.html.
[14] "25 Big Companies T hat Are Going Green," http://www.businesspundit.com/25-big- companies-that-are-going-green/. [Accessed 13 05 2015]
[15] B. E. Smith, Green Computing: T ools and T echniques for Saving Energy, Money, and Resources.
[16] R..Bolla, "Energy-Aware Performance Optimization for" ACM 2009
[17] Aditya et al., " Green Computing Research Challenges: A Review" vol. 3, no-10
[18] How Can I Overcome the Barriers to Green Infrastructure? [Online]. Available: http://water.epa.gov/infrastructure/greeninfrastructure/gi_barrier.cfm.[03/06/2016]
[19] T .R. Soomro and M. Sarwar, "Green Computing: From Current to Future Trends, " WorldAcademy of Science, Engineeringand T echnology, vol.6,2012.
[20] Prof. T .Venkat Narayana Rao, Manikonda Jhansi Rani, ChenchuSwetha, A.Satyam, " Sustainable Environment Friendly Green Computing for Futuristic T echnologies", Journal of Information Sciences and Computing T echnologies, vol. 2, no. 1, February2015

Actions on Google: Developer Platform for Google Assistant

Arshdeep Singh
CSE, Student
Baba Banda Singh Bahadur Engineering College
Fatehgarh sahib, Punjab, India.
Arsh199820@gmail.com

Taranpreet Singh
CSE, Student
Baba Banda Singh Bahadur Engineering College
Fatehgarh sahib, Punjab, India.
Taranpreets1999@gmail.com

***Abstract*-- This paper includes how to implements our speech recognition by passing through different levels with having the knowledge of node.js, JavaS cript and dialogflow. An action on Google is basically a developer platform for Google through which we can work on Google assistant.**

Actions on Google let you extend the functionality of the Google Assistant with Actions. Actions let users get things done through a conversational interface that can range from a quick command to turn on some lights or a longer conversation, such as playing a trivia game.

Keywords-- Actions on Google, Node.js, Java, Dialogflow, Intent, Action, Fulfillment, Actions project, Geeknum, Fulfillmentweebhook, and JavaScript.

I. INTRODUCTION:

Actions on Google is a developer platform that lets you create software to extend the functionality of the Google Assistant, Google's virtual personal assistant, across more than 500 million devices, including smart speakers, phones, cars, TVs, headphones, watches, and more.

Users engage Google Assistant in conversation to get things done, like buying groceries or booking a ride (for a complete list of what's possible now, sees the Actions directory). As a developer, you can use Actions on Google to eas ily create an d manage delightful and effective conversational experiences between users and your own 3rd -party fulfilment service.

II. ACTIONS FOR THE GOOGLE ASSISTANT:

Unlike with traditional mobile and desktop apps, which use computer-centric paradigms, users interact with Actions for the Assistant [14]through natural-sounding, back and forth conversation.

Key Terms:

- Intent: A goal or task that users want to do, such as ordering coffee or finding a piece of music. In Actions on Google, this is represented as a unique identifier and the corresponding user utterances that can trigger the intent.
- Action: An [1] interaction you build for the Assistant that supports a specific intent and has a corresponding fulfillment that processes the intent.
- Fulfillment: A service, app, feed, conversation, or other logic that handles intent and carries out the corresponding Action.

A. Smart home Actions:

We can build smart home [12] Actions that let users control Internet of Things (IoT) devices through the Google Assistant. Building smart home Actions lets you connect, query, and control devices through your existing cloud infrastructure.

B. Conversational Actions:

To start a conversational Action [13], the user needs to invoke your Action through the Assistant. Users say or type a phrase like "Hey Google, talk to Google IO 18". This tells the Assistant the name of the Action to talk to.

From this point onwards, the user is talking to your Action and giving it input. This conversation continues as a two-way dialog until the user's request is fulfilled or the conversation is finished.

C. How conversational Actions work :

i. When you build an Action for the Assistant, you design your conversations for a variety of surfaces, such as a voice-centric conversation for voice-activated speakers or a visual conversation on a surface that the Assistant supports. This lets users get things done quickly through either voice or visual affordances.

ii. Users can invoke your Actions from any of these surfaces:
   - Wear OS devices
   - Assistant-enabled headphones
   - Chromebooks
   - Android TV
   - Android phones and tablets
   - Smart displays and speakers

iii. Your Actions run entirely in the cloud, even when users talk to them on their phone, smart home device, or watch.

iv. The user's device sends the user's utterance to the Assistant, which routes it to your fulfillment service via HTTPS POST requests.

v. Your fulfillment Figures out a relevant response and sends that back to the Assistant, which ultimately returns it to the user.

III. HOW ACTION DISTRIBUTION WORKS

Users don't need to install anything to invoke your Actions on the Assistant. However, all Actions projects must be submitted and approved before users can discover and invoke them. Once

approved, your users can invoke and discover your Actions in a variety of ways, and we provide metrics for you to track and analyse how your Actions are being used to better serve your users.

Key Terms:

- Actions project: A project created in the Actions Console that contains your Actions. A single Actions project contains at least one Action for the welcome intent, but can include additional Actions to fulfill other relevant intents. Your Actions project also defines you r listing in the Actions Directory

A. How users invok e and discover Actions:

How users discover and invoke your Actions is very important in its long term success. You have control over these variables by defining your own invocation name and Action phrases. To better understand how this works, the following examples show how users might invoke and discover an imaginary Actions project called GeekNum.

i. Invocation methods

- By invocation name
  Users can explicitly invoke your Actions by saying things like "Ok Google, talk to Geek Num" or "Ok Google, speak to Geek Num".
- By invocation name and action phrase
  Users can explicitly invoke your Actions with an Action phrase by saying things like "Ok Google, talk to Geek Num to learn about the number 42"

ii. Discovery Methods

- By Action phrase
  Users discover your Actions by saying things like "Ok Google, learn about the number 42" . The Assistant knows you can handle this action and says "For that, try saying 'Ok Google, let me talk to Geek Num'". Action phrases give you an easy and powerful way to increase usage of your Actions by supporting specific and differentiated tasks that your Actions excel at handling.

- Actions Directory
  Users can browse and discover Actions that they might want to use on the Actions directory. Your Actions project listing shows a description of your Actions, the surfaces they're available on, sample invocations, ratings, and more.

- On your website
  You can provide a link to the web or mobile version of your Action project's listing in our directory, so you can promote your Actions and drive usage. You should write good descriptions and provide great images on what your Action's experience is like to increase traffic and ranking.

- Action links
  You can generate a link to a specific Action in your project. When a user clicks on your link on a web or mobile browser, they will be directed to the Assistant and your Action will be invoked.

IV. BUILD FULFILLM ENT WITH THE ACTIONS ON GOOGLE NODE.JS CLIENT LIBRARY[6]: The Actions on Google Node.js client library is the recommended way for accessing and interacting with the Actions on Google platform if you are creating a fulfillmentwebhook in JavaScript.

A. Introduction:

The Node.js client library is a fulfillment library for Actions on Google that provides these features:

- Supports all Actions on Google features, including text and rich multimedia responses, account sign-in, data storage, transactions, and more.
- Provides an idiomatic layer of abstraction in JavaScript that wraps the conversation HTTP/JSON webhook API.
- Handles the low-level details of communication between your fulfillment and the Actions on Google platform.
- Can be installed using familiar package management tools, such as npm or yarn.
- Let's you easily deploy your fulfillmentwebhook on serverless computing platforms such as Cloud Functions for Firebase [5] or AWS Lambda. You can als o host your fulfillmentwebhook on a cloud service provider or on a self-hosted and self-managed environment.
- Is compatible with Node.js v6.0.0 and higher.
- You can use the client library in conjunction with the Dialogflow integration for Actions on Google or with the Actions SDK.
- This document describes the basic developer workflow for using Actions on Google Node.js client library version 2. If you are migrating to version 2 from Actions on Google Node.js client library version 1, make sure to review the migration guide.
  We can use the client library in conjunction with the Dialogflow integration for Actions on Google or with the Actions SDK.

B. Understand how it work s:

Before we use the client library, it is helpful to understand how your fulfilment webhook uses the client library to process user requests that Actions on Google sends to your fulfillment.

When we create a fulfillmentwebhook in JavaScript, you can deploy and host your code on a serverless computing environment like Google's Cloud Functions for Firebase or AWS Lambda. You can also host the code yourself

without additional work using the Express web framework.

Within the runtime environment, the fulfillmentwebhook can call functions in the client library to process user requests and send responses back to Actions on Google for rendering into user output [9].

The key tasks that your fulfillmentwebhook handles with the aid of the client library are briefly summarized below:

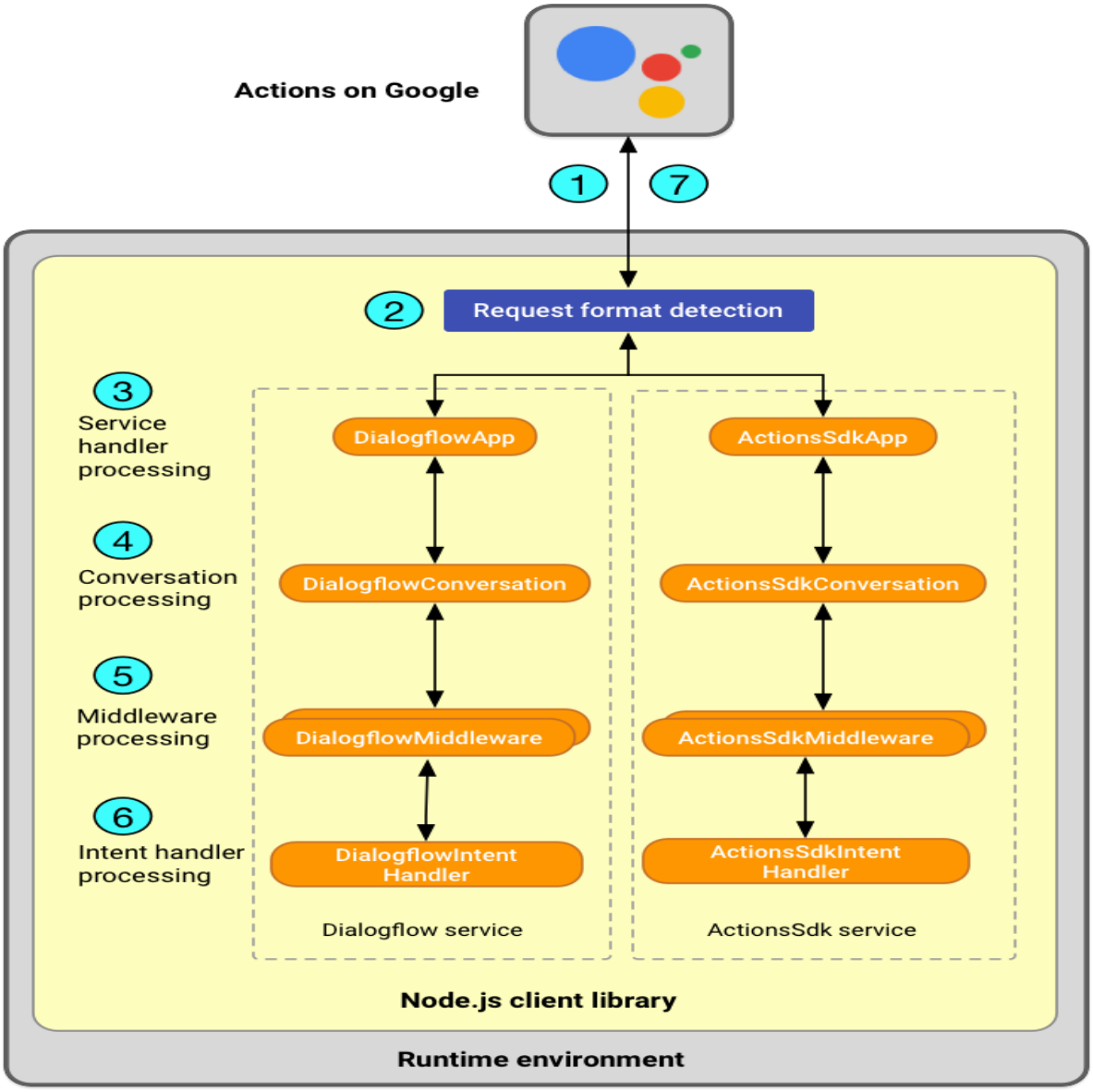


Figure no.1: Node.js client library

a) *Receiving user requests:* When a user makes a query to the Google Assistant, the Actions on Google platform sends an HTTP request to your fulfilmentwebhook; the request includes a JSON payload that contains the intent and other data such as the raw text of the user input, and the surface capabilities of the user's device. For more examples of the JSON payload content, see the Dialogflow webhook format and conversation webhook formatguides.

b) *Framework call format detection:* For supported frameworks, the client library auto - detects the framework's call format (for example, if the request came from the Express web framework or from AWS Lambda) and knows how to seamlessly handle communication with the Actions on Google platform

c) *Service handler processing:* The client library represents the conversation HTTP/JSON webhook API for Dialogflow and Actions SDK as a service function. Your fulfillmentwebhook uses the appropriate service to create a global app instance. The appinstance acts as a handler for HTTP requests and understands the service's specific protocol.

d) *Conversation processing:* The client library represents per-conversation information as a Conversation object that's attached to the app instance. Your fulfilment webhook can use the Conversation object to retrieve cross -

conversational stored data or state information, send responses to users, or close the mic.

e) *Middleware processing:* The client library lets you create your own conversation services middleware, which consists of one or more functions you define that the client library automatically runs before calling the intent handler. Your fulfillmentwebhook can use your middleware to add properties or helper classes to the Conversation object.

f) *Intent handler processing:* The client library lets you define handlers for intents your fulfillmentwebhook understands. For Dialogflow, the client library routes the request to the correct intent handler by mapping to the exact string of the intent name defined in the Dialogflow console. For Actions SDK, it's routed based on the intent property sent from Actions on Google.

g) *Sending responses to users:* To construct responses, your fulfillmentwebhook calls the Conversation#ask() function.
The ask() function can be called multiple times to incrementally build the response. The client library serializes the response into an HTTP request with a JSON payload and sends it to Actions on Google. The close() function has a similar behavior as ask() but closes the conversation.

V. DIALOGFLOW :

Dialogflow [5] lets you build conversational interfaces on top of your products and services b y providing a powerful natural language understanding (NLU) engine to process and understand natural language input. This document goes over how Dialogflow works and how it can help you create conversational user interfaces that delight users.

A. Introduction:

Traditional computer interfaces require structured and predictable input to function properly, which makes the use of these interfaces unnatural and sometimes difficult. If users can't easily Figure out this structured input, they'll have a hard time figuring out what to do.

Even with this simple question, you can see that conversational experiences are hard to implement. Interpreting and processing natural language requires a very robust language parser that's capable of understanding the nuances of language. Your code would have to handle all these different types of requests (and potentially many more) to carry out the same logic: looking up some forecast information for a time and location. For this reason, a traditional computer interface would tend to force users to input a well-known, standard request at the detriment of the user experience, because it's just easier.

However, Dialogflow [11] lets you easily achieve a conversational user experience by handling the natural language understanding (NLU) for you. When you use Dialogflow, you create agents that can understand the vast and varied nuances of human language and translate that to standard and structured meaning that your apps and services can understand. Let's take a look at how Dialogflow might handle the previous examples for weather forecast requests.

B. Agents:

Agents [10] are best described as Natural Language Understanding (NLU) modules. These modules can be included in your app, website, product, or service and translate text or spoken us er requests into actionable data. This translation occurs when a user's utterance matches intent within your agent.

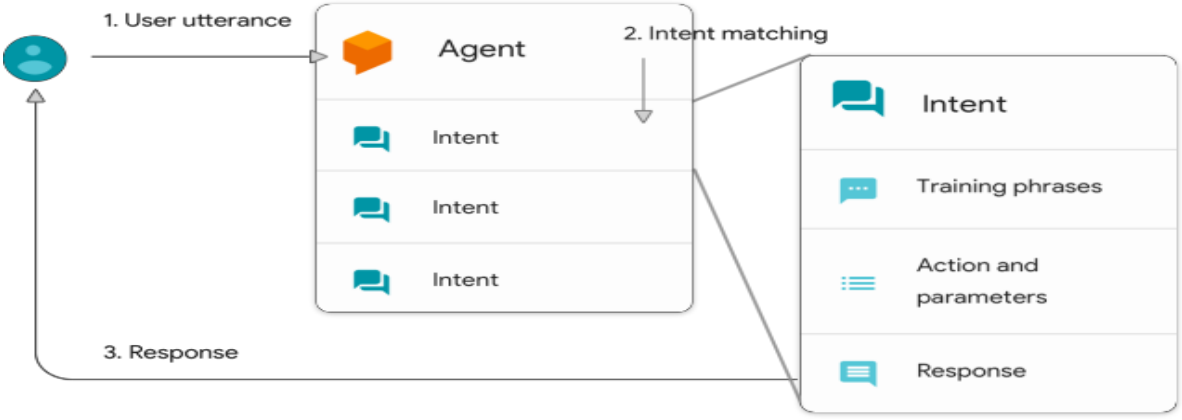


Figure no. 2: Agent

The matched intent then delivers a response back to the user. This response can be a simple text or spoken acknowledgment or a webhook response that includes information obtained from another system. Dialogflow sends a webhook request with the location and date parameters to a th ird -party weather service. This weather service returns a webhook response in JSON format. The agent's custom fulfillment parses the JSON data and

delivers a response to the user with the relevant information.

C. Intents:

In Dialogflow, the basic flow of conversation involves these steps [7]:

- The user giving input
- Your Dialogflow agent parsing that input
- Your agent returning a response to the user

To define how conversations work, you create intents in your agent that map user input to responses. In each intent, you define examples of user utterances that can trigger the intent, what to extract from the utterance, and how to respond.

Generally, intent represents one dialog turn within the conversation. For example, you could create an agent that recognizes and responds to users' input about their favouritecolour. If a user said something like "My favouritecolour is purple", your agent would match that input to its corresponding intent and return the response you defined within that intent. Your agent's response usually prompts users for another utterance, which your agent will attempt to match to another intent, and the conversation continues.

D. Entities:

Entities are Dialogflow's mechanism for identifying and extracting useful data from natural language inputs.

While intents allow your agent to understand the motivation behind a particular user input, entities are used to pick out specific pieces of information that your users mention — anything from street addresses to product names or amounts with units. Any important data you want to get from a user's request will have a corresponding entity.

a) Entity terminology: The term entity is used in this documentation and in the Dialogflow console to describe the general concept of entities. When discussing entity details, it's important to understandmore specific terms:

i. *Entity type*: Defines the type of information you want to extract from user input. For example, vegetable could be the name of an entity type. Clicking Create Entity from the Dialogflow console creates an entity type. When using the API, the term entity type refers to an EntityType object.

ii. *Entity entry*: For each entity type, there are many entity entries. Each entity entry provides a set of words or phrases that are considered equivalent. For example, if vegetable is an entity type, you could define these three entity entries:

- carrot
- scallion, green onion
- bell pepper, sweet pepper

When editing an entity type from the Dialogflow console, each row of the display is an entity entry. When using the API, the termentityentry refers to an Entity object. EntityEntry would be a better name than Entity for this object, but we need to maintain backwards compatibility when improving the API

b) System entities: Dialogflow is equipped with numerous systementities, which allow agents to extract information about a wide range of concepts without any additional conFigureuration. For example, system entities are available for extracting dates, times, and locations from natural language inputs.

c) Developer entities: If you need to extract information about concepts beyond those covered by Dialogflow's s y s t e m entities, you can define own developer entity types .For example, a brand might create an entity type to recognize its unique set of product names.

d) Session entities: It is also possible to define entity types that apply only to a specific conversation. For example, you might create an entity type to represent the time-sensitive options available to a particular user when making a booking. These are called session entity types.

e) Manage entities Dialogflow also provides tools for managing entities, including mechanisms for exporting and uploading entity data and modifying entities via API.

E. Contexts:

Contexts represent the current state of a user's request and allow your agent to carry information from one intent to another. You can use combinations of input and output contexts to control the conversational path the user takes through your dialog.

F. Events:

Events allow you to invoke intents based on something that has happened instead of what a user communicates. Dialogflow supports events from several platforms (like Google Assistant, Slack, and more) based on actions users take on those platforms. You can also create your own custom events that can be triggered via fulfillment or the detectIntent API.

G. Fulfillment:

Fulfillment [8] is code that's deployed as a webhook that lets your Dialogflow agent call business logic on an intent-by-intent basis. During a conversation, fulfillment allows you to use the

information extracted by Dialogflow's natural language processing to generate dynamic responses or trigger actions on your back-end.
Most Dialogflow agents make use of fulfillment.
The following are someexample cases where you can use fulfillment to extend an agent:

- To generate dynamic responses based on information looked up from a database.
- To place orders based on products a customer has asked for.
- To implement the rules and winning conditions for a game.

H. Training and Analytics[13]:

Dialogflow provides features that can help you build and refine your agents using real world data. As a developer, you can leverage existing sources of conversation data you might have access to, as well as usage and performance data that pertains to your running agents.
There are three main use cases for real world data within Dialogflow:

- Building a new Dialogflow agent using logs of existing customer interactions.
- Use training to add existing data to training phrases.
- Improving the performance of alive Dialogflow agent using its own logs.
- Use training to add additional data to training phrases.
- Understanding the performance of a live Dialogflow agent to inform your design decisions.
- Use analytics to assess performance of agent.

CONCLUSIONS:

By this paper, it is reported that by passing some levels, we can easily built a small project on Actions on Google with some basic knowledge of node.js, dialogflow and JavaScript. We can make a project by readingdocumentation of dialogflow and Actions on Google

REFERENCES:

[1] Internet Source:https://developers.google.com/actions/
[2] Internet Source: https://developers.google.com/actions/smarthome/
[3] Internet Source: https://developers.google.com/actions/verticals/overview
[4] Internet Source: https://developers.google.com/actions/dialogflow/
[5] Study of Google Firebase APIfor Android--International Journal of Innovative Research in Computer and Communication Engineerin g (An ISO 3297: 2007 Certified Organization) Vol. 4, Issue 9, September 2016 Copyright to IJIRCCE DOI: 10.15680/IJIRCCE. 2016. 0409133 16738
[6] Internet Source: https://developers.google.com/actions/reference/nodejsv2/overview
[7] Internet Source: https://console.dialogflow.com/api-client/#/agent/e4f07efe-e973-4f7c-95d9-c735ff97e664/intents
[8] Internet Source: https://console.dialogflow.com/api-client/#/agent/e4f07efe-e973-4f7c-95d9-c735ff97e664/fulfillment
[9] Real Time Databases for Applications-- Sonam Khedkar1, Swapnil Thube2 MCA (IMCOST), Mumbai University C-4, Wagle IndustrialEstate, NearMulund(W) CheckNaka, Thane (w) – 400604
[10] Internet Source:https://console.dialogflow.com/api-client/#/agent/e4f07efe-e973-4f7c-95d9-c735ff97e664/prebuiltAgents/
[11] Internet Source: https://dialogflow.com/docs/getting-started
[12]Internet Source:https://blog.dialogflow.com/
[13] Internet Source: https://dialogflow.com/docs/training-analytics
[14] An integrated voice recognition system ngoc c. Bui, jean j. Monbaron, and jean g. Michel, member, IEEE

Review Paper on Skin Cancer Detection

Shelly
Research scholar, Department of Electronics and communication Punjabi University
Patiala,Punjab, India
shellygarg96@gmail.com

Balkrishan
AP P., Computer Engineering Section, Yadavindra College of Engineering, GKC, Punjabi University
Talwandi Sabo, Bathinda, Punjab, India
balkrishan_76@rediffmail.com

***Abstract*— The problem of S kin cancer is increasing now these days. It is important to detect the skin cancer at early stages for curability and survival of human being. S ome clinical observation methods are unable to identify the skin cancer. S o, Automatic diagnosis methods are used to detection of skin cancer. In this paper, skin cancer detection techniques and different classification methods are described**.

***Keywords*— Automatic detection, Classification, Feature extraction, Segmentation, Skin cancer detection**

I. INTRODUCTION

Skin has a tendency to develop several types of cancers. Cancer is defined as a condition in which there is an unnatural development of certain cells that happen due to modification in gene expression. These cancerous cells further invade neighbourhood cells. In the modern world, number of cancer patients has increased because of factors affecting the human body such as life span increase, some personal habits and ultraviolet exposure. Cancer can be classified into two types that are Benign and Malignant. The difference between both categories is due to their ability to spread or metastasize to remote tissues and organs. Malignant cancer can effectively invade and destroy their neighbouring tissues. They can also spread to remote tissues as well as organs through blood stream or lymphatic system. Benign cancer is more localized type of cancer but it can affect the environment through applying pressure on neighbouring nerves or blood vessels. It can also secrete chemical compounds which may affect normal functioning of neighbour tissues. Benign cancer has less growth rate and is less life-threatening than Malignant type. If these cancers not treated properly, it may lead to harmful effects on human being. So, skin cancer detection is very necessary for proper diagnosis. For cancer patient, a technique named as biopsy, is used which is an invasive operation and gives unpleasant experience. To avoid unnecessary biopsy, Dermoscopy imaging technique is used for detailed inspection of inner layers of skin using microscope and special illumination equipment. The main problem is to determine the absence or presence of skin lesion in dermoscopic skin images and to classify the skin lesion type. Several researchers tried to build automatic cancer detection techniques for improving in accuracy. In this paper, various detection techniques are explained.

II. LITERATURE REVIEW

Barata et al. [1] described melanoma skin cancer detection based on local and global features. They used the K-NN, Support Vector Machine (SVM) and Boost classifiers for the classification of skin lesion types. They extracted the features based on color as well as texture and showed that combination of both types of these features performed better in form of sensitivity and specificity. Silva et al. [2] discussed the classification of melanoma skin cancer based on Menzies color method using Neural Network classifier applied on 28 skin lesion images. Experimental results were shown 74% accuracy in classification. Silveira et al. [3] discussed the Melanocytic skin lesion segmentation based on six segmentation methods using 100 dermoscopic images. They used the Adaptive Thresholding (AT), EM level set (EM-LS), Adaptive Snake (AS), Gradient Vector Flow (GVF), Fuzzy-Based-Split-Merge (FBSM) ,level set (LS) methods for segmentation and showed that Adaptive Snake and EM level set methods performed better than other methods. Sheha et al. [4] discussed melanoma skin cancer classification based on Multilayer Perceptron (MLP) classifier using 102 dermoscopy images. They used the Fisher score ranking method to select optimal features on the basis of GLCM features and showed that Traditional MLP classifier performed better than Automatic MLP classifier in the form of sensitivity and specificity. Shimizu et al. [5] discussed the detection of melanoma skin cancer based on single shot and double shot classifier using 658 dermoscopic images. They used the Wilk's Lambda method to select optimal features on the basis of shape, color and texture. They showed that double shot classifier performed better than single shot classifier in the form of and specificity and sensitivity. Chen et al. [6] discussed the selection of multi class features based on Recursive Feature Elimination (RFE) method using 400 skin lesion images. They used LS-SVM classifier for classification of skin lesion images on the basis of selected features and performed better in terms of accuracy. Garnavi et al. [7] designed Computer Aided Design (CAD) system for Melanoma skin cancer detection using 289 skin lesion images. Wavelet-decomposition method is used to extract texture features. Gain-ratio method is used to select optimal features on the basis of texture and border features. Experimental results were shown 91.26% accuracy in classification. Shimizu et al. [8] designed CAD system for the discrimination of Melanocytic and Non -Melanocytic skin lesions using 968 dermoscopic images. They used Wilk's Lambda method to select optimal features on the basis of GLCM features. They showed that layered model performed better than flat model in terms of detection rate. Chang et al. [9] designed CAD system for the discrimination of Melanocytic and Non-Melanocytic skin lesions using 769 conventional skin lesion images. RFE method is used to select the optimal features on the basis of new color and texture features. Experimental results were shown 90.64%

accuracy in classification. Ganster et al. [10] described the detection of melanoma skin cancer based on fusion segmentation techniques using 5393 dermoscopic images. Statistical feature subset method is used for optimal features selection on the basis of shape and radiometric features. Experimental results were shown 87% sensitivity and 92% specificity in classification. Celebia et al. [11] described the classification of pigmented skin lesions based on SVM method using 564 dermoscopic Filter method is used to select optimal features on the basis of shape, color and texture features. Experimental results were shown specificity of 92.34% and sensitivity of 93.33% in classification. Yu et al. [12] described the detection of melanoma skin cancer based on deep residual network using 1250 dermoscopic skin images. Fully Convolution Residual Network is used to segment the skin lesion. Experimental results were shown high accuracy in both segmentation and classification tasks. Li et al. [13] discussed the Melanoma skin cancer detection based on deep learning technique using ISIC database. Lesion Indexing framework is used to segment skin lesion images and then classified on the basis of CNN based framework. Experimental results were shown 91.2% accuracy in classification. Barata et al. [14] described the detection of pigment skin cancer and image quality is improved using bank of direction filters and multiple morphological operations. Ada-boost algorithm is used to classify the image on the basis of extracted features. In 2013, they also described the detection of melanoma skin cancer based on local and global techniques [15]. They also used Adaptive Thresholding method for segmentation of skin lesion image. They extracted the features based on texture as well as color. The performance of extracted color features is better than texture features in the form of sensitivity and specificity. Silveira et al. [16] discussed the melanocytic skin lesion segmentation based on six segmentation methods using 100 dermoscopic images. They showed that Adaptive Snake and EM-Level Set methods performed better than other segmentation methods. Wong et al. [17] analyzed segmentation of skin cancer based on stochastic region-merging method. They used region merging likelihood function on the basis of statistics region and showed that segmentation error rate less as compared to other methods. Yu et al. [18] described the segmentation and classification of skin lesion based on FCRN and deep residual network. They enabled the classification network to extract more specific and important features based on segmented results instead of the whole image. SVM and the Softmax classifiers are used to obtain high accuracy. Eltayef et al. [19] described the detection of melanoma skin cancer based on Markov Random field and Fuzzy C Means. Bank of directional filter is used for noise reduction to enhance the image quality. They also described the detection of pigment network based on feed forward neural [20]. Gabor and Sobel filters are used for noise reduction to enhance the image quality. They used the Barata's method to extract pigment networks from dermoscopic image and performed better than other methods. Li et al. [21] analyzed the segmentation of skin lesions based on 3D depth information using region-based active contour segmentation method on 20 test skin images. They extracted the features based on color, texture as well as depth features and showed accurate results for non-pigment skin lesions. Marques et al. [22] described the detection of melanoma skin cancer based on pattern recognition system using 163 dermoscopic images They used Binary based classifier to discriminate melanoma skin lesion. They extracted features based on colour as well as texture and showed that combination of both types of these features performed better in the form of sensitivity and specificity. Silva et al. [23] discussed the classification of melanoma skin cancer based on Menzies color method using 28 skin lesion images. Experiment results were performed in terms of Jeffries Matusita and Transformed Divergence metric. They also described the classification of melanoma skin cancer based on Menzies color method using Neural Network classifier. Nisar et al .[24] analyzed the skin lesion segmentation based on four color models using 15 skin lesion images. They used the HSI, CMY, YCbCr and CIELAB color models for segmentation and showed that HSI, CIELAB color models performed better than other models. Jiji et al. [25] described the detection of melanoma skin cancer based on SVM method using 2000 d ermoscopic images. They used the CIE-Lab space method for segmentation of skin lesion types. They extracted the features based on color, shape as well as texture and performed 91.1% accuracy in classification.

III. DESCRIPTION OF DATASET

Different data sets are used for research purposes. Details of datasets are mentioned as follows:

A. Dermofit dataset

It is a digital tool make to train medical professionals or researchers to accurately determine the malignant skin lesion and benign skin lesions. The Dermofit Library is an assembly of different 1,300 focal high quality images taken under standardized conditions. Ten different classes including melanomas, basal cell carcinomas and Seborrhea keratosis are found in this dataset which are Actinic Keratosis (AK), Dermatofibroma (DF), Haemangioma (HEM), Pyogenic Granuloma (PYO), Basal Cell Carcinoma (BCC), Squamous Cell Carcinoma (SCC), Melanocytic Nevus (ML), Melanoma (MEL), Seborrhea Keratosis (SK), Intraepithelial Carcinoma (IEC). Each and every image has been diag nosed by dermatologists.

B. Database Hospital Pedro Hispano (PH^2)

This dataset is obtained from the Dermatology Service of Hospital Pedro Hispano, Portugal developed for the research work as well as used for benchmarking. It consists of 200 dermoscopic images. This dataset has been. The different categories of skin lesions are represented in Figure 1.

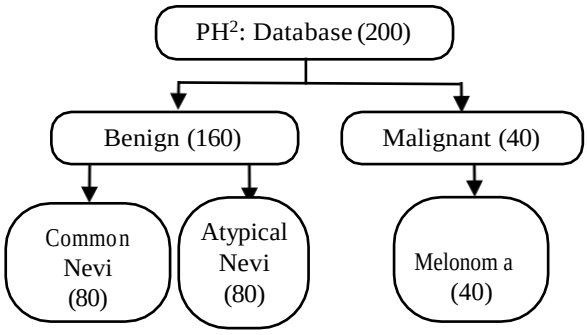


Figure 1: PH^2 Database Description

C. Database International Skin Imaging Collaboration (ISIC) Archive

The International Society for Digital Imagin g of the Skin (ISDIS) was established in 1992 and dataset are diagnosed by a group of U.S. dermatologists. In this dataset, number of images is available in range of thousands. Contributions in the field of ISIC by researchers is summarized using dataset is described in Table 3. The main aim of this society is to familiarize the members with new and advanced digital skin imaging technologies. ISDIS also collaborates with the International Dermoscopy Society (IDS) and International Confocal Group (ICG) groups.

D. Images collected from by a particular research group in collaboration with many hospital

Some researchers collected samples of different types of skin lesions from the different universities and used for research work. These different categories are s hown in Figure 2.

I V. AUT OMATIC OF SKIN CANCER DETECTION

Computer Aided Design (CAD) system design employing different schemes for diagnosis between different classes of skin lesion image and shown in Figure 3. The following steps are Pre-processing, Image Segmentation or ROI extraction, Features Extraction, feature selection methods or feature space dimensionality reduction, and classification. It is the generalized diagram that is used for the research purpose. Various skin lesions are diagnosed using CAD system. Now days, this method becomes most popular for detection of skin cancer type. So various techniques are used in this system.

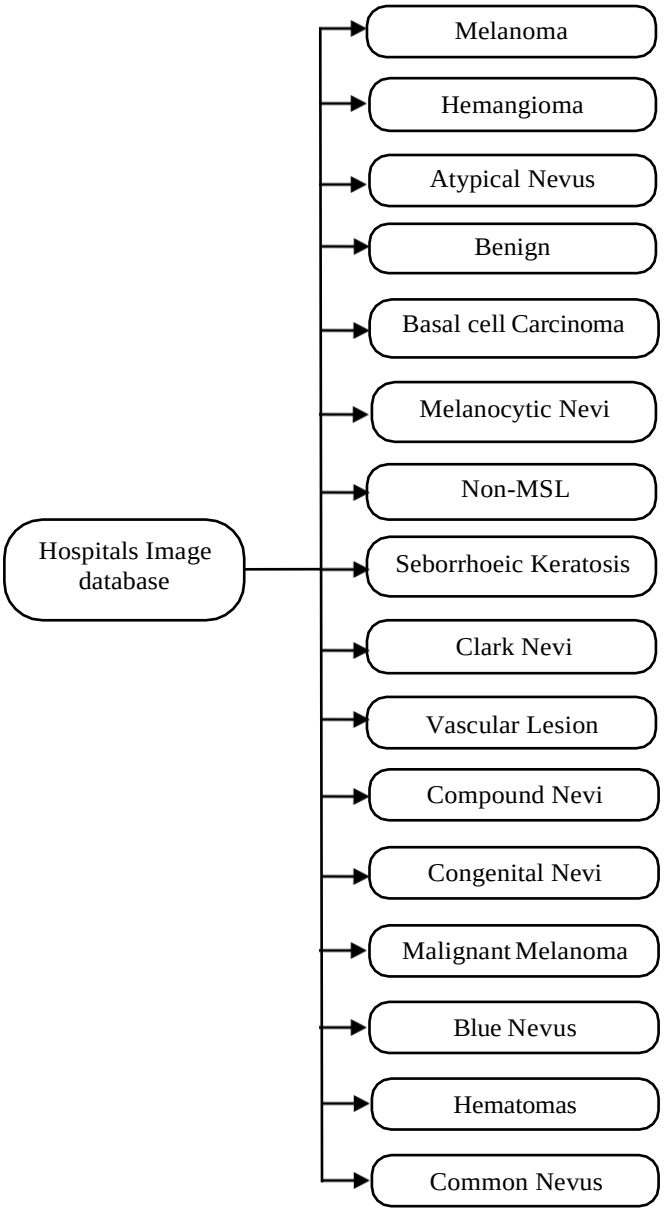


Figure 2: Hospitals Database Description

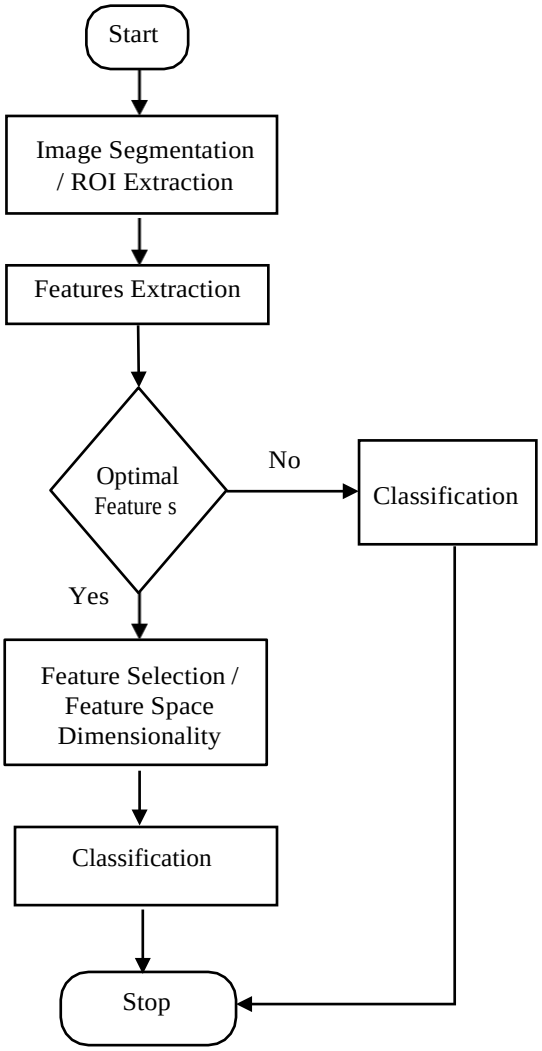


Figure 3: Automatic Detection framework

A. Pre-Processing

Because of the presence of noise in dermoscopic images, it becomes very difficult to analyze the image clearly. So, Pre-processing is the first step of detection. to improve the image quality of skin lesion, remove the unwanted noise and eliminate irrelevant information [26].

B. Image Segmenation/ROI Extraction

Segmentation can be done by Manual method/ Semi-automated method. It is the process to determine the border of skin lesion from the dermoscopic image [27-30]. This skin lesion part contains the important information of the dermoscopic image. It is very important to examine the exact defined region of lesion for diagnosis. This part differentiates the skin cancer types. Then this part is further used for features extraction. Many techniques are used for segmentation which are edge based methods, hybrid based methods ,classification based and region based method.

C. Feature Extraction

Features may be on shape- based, color based and texture based which are to be extracted as seen in Figure 4.

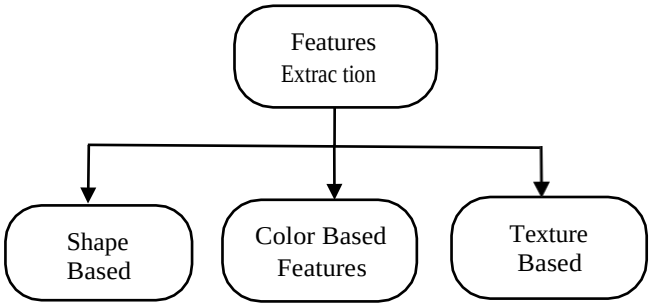


Figure 4: Features Extraction

- Shape Based Features: The shape of skin lesion examines by its physical structure. The shape can be entitled by the boundary, region, etc. These illustrations are to be used to recognize different images Shape-based features are area, perimeter, form factor, asymmetry, compactness, variance, maximum diameter, aspect ratio, solidity, rectangularity, bounding box, eccentricity, elongation and so on. Various shape-based features are mentioned in Figure 5.

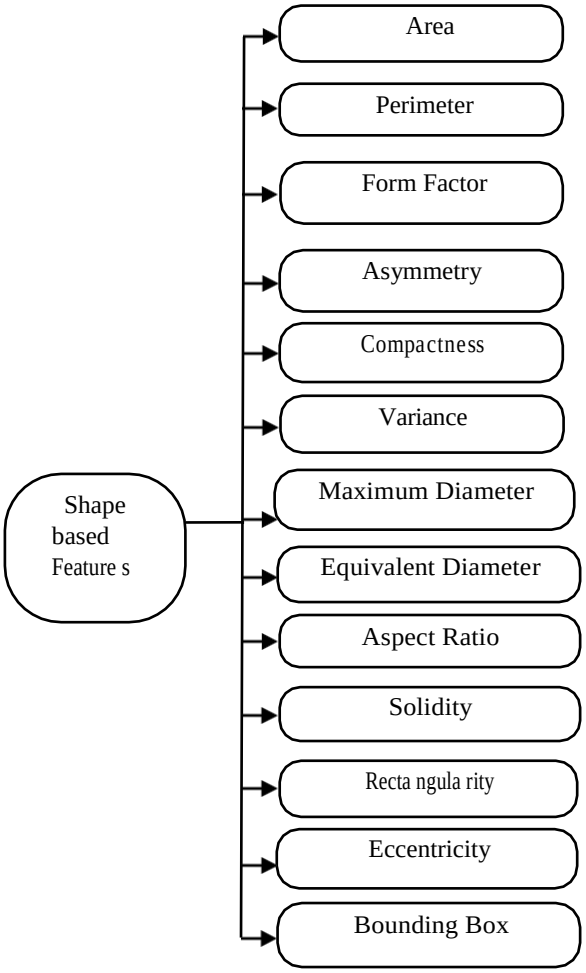


Figure 5: Shape based features

- Color Based Features: Skin Lesion representation can also be described by using color. It is also an important feature presents the skin lesion information. Proper analysis and relative distribution plays a vital role for the correct identification of different skin lesion classes. Color Analysis means to examine the pigmentation rate and color distribution in skin lesion image. Color based features are variance, entropy, correlation, maximum and minimum average, mean, standard deviation, color asymmetry and so on. Various color-based features are described in Figure 6.

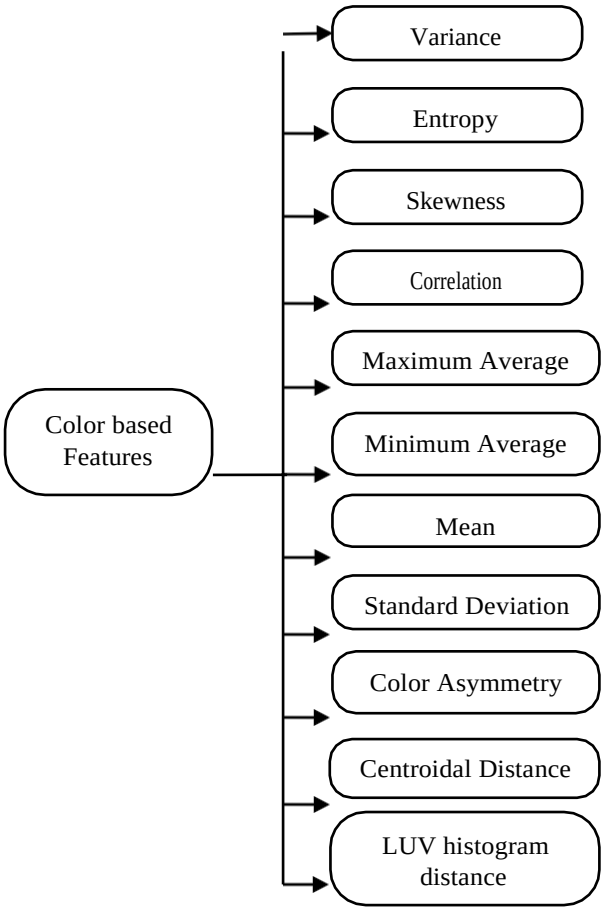


Figure 6: Color based features

- Texture based Features: The methods for texture-based feature extraction can be mainly classified into transform domain and spatial domai. nDifferent texture-based features methods are Statistical methods, Wavelet-based methods, Signal processing-based methods, Gabor based methods and so on.

D. Feature Selection methods

All the features are computed by using feature extraction techniques and these features may contain some redundant, unwanted information which can degrade the system performance. These redundant features don't give any effective information. So, feature selection methods are employed to select the optimal features. Selected features are further processed to provide accurate results. Feature Selection methods are Filter-based, Wrapper based methods and Feature Space Dimensionality Reduction method. Different types of methods for feature extraction are described in Figure 7.

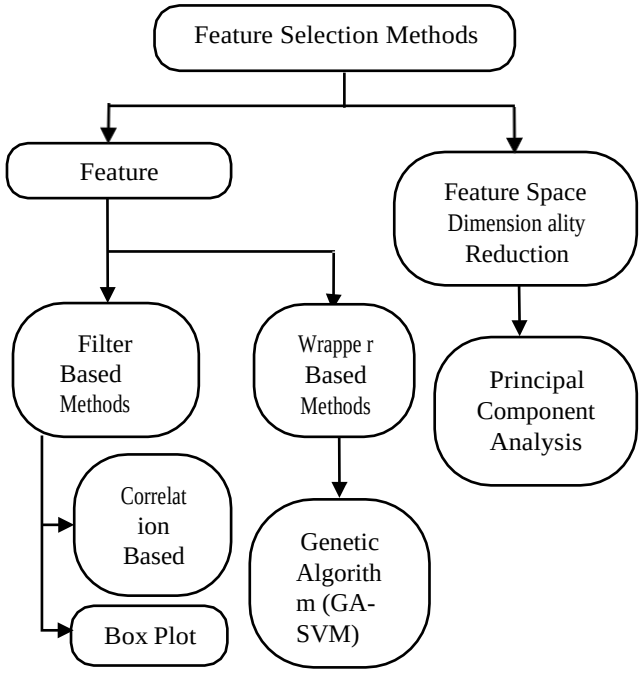


Figure 7: Feature Selection methods

E. Classification

It is a process to estimate the information of unknown class from the different types of classes. It is th e final stage to classify the skin cancer type based on extracted features. There are different types of classifiers used for the classification of images like Probabilistic Neural Network (PNN), K-Nearest Neighbor, (k-NN), Neural Network (NN), Support Vector Machine (SVM), Smooth Support Vector Machine (SSVM), Adaptive neuro-fuzzy (ANFC), Linear Discriminant Analysis (LDA) [31-34]. A suitable classifier is designed to identified whether the skin lesion is benign or Malignant and also further identify the other classes. Classification is important task to develop automatic detection system for diagnosis of skin lesion type. Different types of classifiers are mentioned in Figure 8.

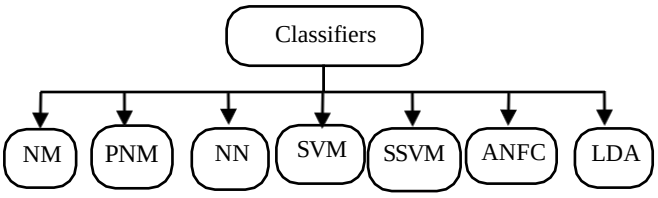


Figure 8. Different types of Classifiers

V. CONCLUSION

In this study, skin cancer is crucial problem in human being. For the prior detection of skin cancer and its treatment, a

vast variety of research papers were studied and discussed in this paper. Certain steps are used for automatic detection of the skin cancer disease and discussed. The techniques required in each step are analyzed and given to another technique for further processing. In Future, the automated skin cancer system could be developed to prior detection of skin cancer. But it is not a substitute of clinician in melanoma diagnosis .

REFERENCES

[1] Barata, C., Ruela, M., Francisco, M., Mendonça, T . and Marques, J.S., "T wo Systems for the Detection of Melanomas in dermoscopy Images using T exture and Color Features", *IEEE System s Journal*, 8(3): 965-979, 2013 .

[2] Silva, C.S.P . and Marcal, A.R.S., "Color-based dermoscopy classification of Cutaneous Lesions: an alternative Approach ", *Com puter Methods in Biomechanics and Biom edical Engineering: Imaging & Visualization, Taylor & Francis*, 1(4): 211-224, 2013 .

[3] Silveira, M., Nascimento, J.C., Marques, J., Marçal, A.R.S., Mendonça, T., Yamauchi, S., Maeda, J. and Rozeira, J., "Comparison of Segmentation Methods for Melanoma Diagnosis in Dermocopy Images", *IEEE Journalof Selected Topics in Signal Processing*, 3(1): 35-45, 2009.

[4] Sheha, M.A. and Mabrouk, M.S., "Automatic Detection of Melanoma Skin Cancer using T exture Analysis", *International Journal of Com puter Applications*, 42(20): 22-26, March 2012

[5] Shimizu, K., Iyatomi, H., Norton, K.A. and Celebi, M.E., "Extension of automated melanoma screening for Non-Melanocytic skin lesions," *International journal of Com puter Applications in Technology*, 50(1/2): 122–130, 2014.

[6] Chen X., Zeng X. and Alphen D.," Multi-class feature selection for texture classification", *Pattern Recognition Letters, Elsevier*, 27(14): 1685-1691, 2006.

[7] Garnavi, R., Aldeen, M. and Bailey, J., "Computer-aided diagnosis of melanoma using border- and wavelet -based texture analysis," *IEEE Transactions on Inform ation Technology in Biom edicine*, 16(6): 1239–1252, 2012.

[8] Shimizu, K., Iyatomi, H., Celebi, M.E., Norton, K.A. and T anaka, M.," Four-Class Classification of Skin Lesions with T ask Decomposition Strategy," *IEEE Transactions on Biom edical Engineering*, 62(1): 274-283, 2015.

[9] Chang, Y., Huang, A., Yang, C.Y., Lee, C.H., Chen, Y.C., Wu, T .Y., Chen, G.S., "Computer-Aided Diagnosis of Skin Lesions Using Conventional Digital Photography: A Reliability and Feasibility Study", CAD of *Skin Lesions with Digital Macrographs*, 8(11): 76212, 2013.

[10] Ganster, H., Pinz, A., Rohrer, R., Wildling, E., Binder, M. and Kittler, H., "Automate melanoma recognition ," *IEEE Transactions on Medical Im aging*, 20(3): 233–239, 2001.

[11] Celebi, M.E., Kingravi, H.A., Uddin, B., Iyatomi, H., Aslandogan, Y.A., Stoecker, W.V. and Moss, R.H., " A methodological approach to the classification of dermoscopy images," *Com puterized Medical Im aging and Graphic, Elsevier*, 31(6): 362–371, Sep.2007.

[12] Chen, L.Y.H., Dou, Q., Qin J. and Heng, P.A., "Automated Melanoma Recognition in Dermoscopy Images via Very Deep Residual Networks", *IEEE Transactions on medical Imaging*, 36(4): 994-1004, 2016.

[13] Li, Y. and Shen, L., "Skin Lesion Analysis towards Melanoma Detection Using Deep Learning Network", *National Centre for Biotechnology Inform ation journal*, 18(2): 556, 2017.

[14] Barata, C., Marques, J.S. and Rozeira, J., "A system for the detection of pigment network in dermoscopy images using directional filters", *Biom edical Engineering, IEEE Transactions on*, 59(10): 2744–2754, 2012.

[15] Barata, C., Ruela, M., Francisco, M., Mendonça, T . and Marques, J. S., "T wo systems for the detection of melanomas in dermoscopy images using texture and color features", *IEEE Systems Journal*, 8(3): 965–979, 2014.

[16] Silveira, M., Nascimento, J.C., Marques, J.S., Marçal, A.R., Mendonça, T., Yamauchi, S., Maeda, J. and Rozeira, J., "Comparison of segmentation methods for melanoma diagnosis in dermoscopy images", *Selected Topics in Signal Processing, IEEE System s Journal*, 3(1): 35–45, 2009.

[17] Wong, A., Scharcanski, J. and Fieguth, P., "Automatic skin lesion segmentation via it erative stochastic region merging", *IEEE Transactions on Information Technology in Biomedicine*, 15(6): 929–936, 2011.

[18] Yu, L., Chen, H., Dou, Q., Qin, J. and Heng, P.A., "Automated melanoma recognition in dermoscopy images via very deep residual networks", *IEEE transactions on medical imaging*, 36(4): 994–1004, 2017.

[19] Eltayef, K., Li, Y. and Liu, X., "Detection of melanoma skin cancer in dermoscopy images." Journal of Physics Conference Series, IOP Publishing ,787(1):012034, 2017.

[20] Eltayef, K., Li, Y. and Liu. X., "Detection of pigment networks in dermoscopy images." Journal of Physics Conference Series, IOP Publishing,787(1), 2017.

[21] Li, X., Aldridge, B., Ballerini, L., Fisher, B. and Rees, J.," Depth Data Improves Skin Lesion Segmentation", Proc. of the International conference on Medical Image Computing and Computer Assisted Intervention MICCAI Lecture Notes in Computer Science, Springer, Berlin, Heidelberg, pp. 1100-1107, 20-24 Sept. 2009.

[22] Marques, J.S., Barata, C. and Mendonça, T., "On the Role of Texture and Color in the Classification of dermoscopy Images", Proc. of the International Conference of the IEEE Engineering in Medicine and Biology Society, San Diego, California USA, pp. 4402 -4405, 1 August 2012.

[23] Silva, C.S.P., Marcal, A.R.S., Pereira, M., Mendonça, T. and Rozeira, J., " Separability analysis of color classes on dermoscopic images ", Proc. of the International Conference on Image Analysis and Recognition (ICIAR), Springer, Berlin Heidelberg, pp.268-277, June 2012.

[24] Nisar, H., Chang, Y.K. and Chew, T .Y.,"A color space study for skin lesion segmentation," Proc. of the International conference in IEEE on circuits and systems, Kuala Lumpur, Malaysia, pp.172-176, 18-19 Sept.2013.

[25] Jiji, G.W. and Raj, P.J.D., "An Extensive T echnique to Detect and Analyze Melanoma", Proc. of the International Symposium on Biomedical Imaging (ISBI), Melbourne, Australia, 18-21 April2017.

[26] T oossi, M.T.B., Pourreza, H.R., Zare, H., Sigari, M.H., Layegh, and Azimi, A., " An effective hair removal algorithm for dermoscopy images", *Skin Research and Technology*, 19(3): 230–235, 2013.

[27] Iyatomi, H., Oka, H., Saito, M., Miyake, A., Kimoto, M., Yamagami, J., Kobayashi S., T anikawa, A., Hagiwara, M., Ogawa, K., Argenziano, G., Soyer, H.P. and T anaka, M., " Quantitative assessment of tumour extraction from dermoscopy images and evaluation of computer-based extraction methods for an automatic melanoma diagnostic system," *Melanoma Research.*, 16(2): 183-190, 2006.

[28] Celebi, M.E., Kingravi, H.A., Iyatomi, H., Aslandogan, Y.A., Stoecker, W.V., Moss, R.H., Malters, J.M., Grichnik, J.M., Marghoob, A., Rabinovitz, H.S. and Menzies, S.W., " Border detection in dermoscopy images using statistical region merging," *Skin Research and Technology*, 14(3): 347-353, 2008.

[29] Garnavi, R., Aldeen, M., Celebi, M.E., Varigos, G. and Finch, S., "Border detection in dermoscopy images using hybrid T hresholding on optimized color channels," *Com puterized Medical Im aging and Graphics*, 35(2): 105-115, 2011.

[30] Celebi, M.E., Wen, Q., Hwang, S., Iyatomi, H. and Schaefer, G., "Lesion Border Detection in Dermoscopy Images Using Ensembles of T hresholding Methods," *Skin Research and Technology*, 19(1): 252-258, 2013.

[31] Rubegni, P., Cevenini, G., Burroni, M., Perotti, R., Dell'Eva, G., Sbano, P., and Andreassi, L., "Automated diagnosis of pigmented skin lesions", *International Journal of Cancer, 101*(6): 576-580,2002

[32] Blum, A., Luedtke, H., Ellwanger, U., Schwabe, R., Rassner, G. and Garbe, C., " Digital image analysis for diagnosis of cutaneous melanoma: Development of a highly effective com puter algorithm based on analysis of 837 Melanocytic lesions," *British Journal of Derm atology, 151*(5): 1029-1038, 2004.

[33] Burroni, M., Sbano, P., Cevenini, G., Risulo, M., Eva, G.D., Barbini, P., Miracco, C., Fimiani, M., Andreassi, L. and Rubegni, P., "Dysplastic Nevus in situ melanoma: Digital dermoscopy analysis," *British Journal of Dermatology, 152*(4): 679-684, 2005.

[34] Seidenari, S., Pellacani, G. and Grana, C., " Pigment distribution in Melanocytic lesion images: A digital parameter to be employed for computer-aided diagnosis," *Skin Research and Technology, 11*(4): 236-241, 2005.

Flow Oriented Channel Assignment for Multi Radio Wireless Mesh Networks

Jaswinder Kaur
Department of Computer Science and Engineering
Baba Banda Singh Bahadur Engineering College
Fatehgarh Sahib, Punjab
jassk820@gmail.com

Jatinder Singh Saini
Department of Computer Science and Engineering
Baba Banda Singh Bahadur Engineering College
Fatehgarh Sahib, Punjab
sainijatinder@gmail.com

***Abstract*-It has been seen that using multiple radios per node increases the throughput of multi -radio wireless mesh networks. However, multi-radios create several research challenges. One of the challenge is interference, which occurs due to overlapping of channels. Interference reduce the capacity of channels. Flow control are not yet considered for channel assignment. Rather this is a critical factor for proper communication in the network. We investigate channel assignment for a multichannel wireless mesh network backbone, where each router is equipped with multiple interfaces. Of particular interest is the development of channel assignment heuristics for multiple flows. In this work, flow control has been considered for the channel assignment problem. Flow control prevents the overlapping of data between different radio channels and too much unwanted interference between different radio channels.**

Keywords— Wireless mesh networks (WMNs),Multi Radio-Multi Channel WMNs, Channel Assignment.

I. INTRODUCTION

Wireless mesh networks (WMNs) are dynamically self-organized and self-conFigureured, with the nodes in the network automatically establishing an ad hoc network and maintaining the mesh connectivity. WMNs are co mprised of two types of nodes: mesh routers and mesh clients. [4].Wireless Mesh Network is an encouraging key area, supporting various types of emerging and corporate applications e.g. network neighborhood, broadband home networking and disaster management.

Wireless mesh networks (WMNs) are becoming increasingly popular as they provide mobility support, flexibility, and are easy to deployable. A WMN is a set of wireless nodes. Each node can communicate with each other and also can forward packets of each other. Every node is both a host and a wireless router. Wireless routers acts as access points (APs) for wireless mobile devices. Some of them also act as gateways to the Internet through high speed wired links. WMNs are constituted with two nodes: (1) mesh routers and (2) mesh clients. A mesh router having advanced routing functions to support multihop mesh networking. Using multi-hop communications.

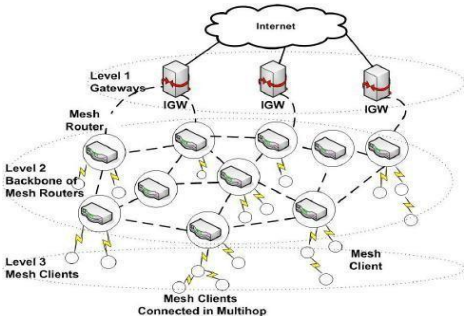


Figure 1 Wireless Mesh Networks

To understand mesh networking, we first need to obtain an appreciation for what a mesh topology represents. If we have n nodes in a network, where the term "node" refers to a communications device that can transport data from one of its interfaces to another, then the ability of each node to communicate with every other node in the network represents a mesh network topology. We can view the structure of a mesh network by simplifying the number of nodes in the network from a value of n , which is what mathematicians like to work with, to an easy -to-visualize number, such as three, four, or five.

Advantages of WMNs:

Self-Organizing and Self ConFigureuring Network: Ad hoc networking enhances network performance, such as flexible network architecture, easy deployment and conFigureuration, connectivity the mesh, i.e., MTM connectivity of mesh.

Cost of deployment reduced: Mesh routers are wireless and they have the ability to service in multi-hop data communication. Thus, using wireless routers in large areas are cheaper compared to single hop routers or access points that they have wired connections.

Increased Reliability: In a WMN, the multiple paths from source to destination nodes provide alternate paths in case of failure. Alternate paths may be selected, in order to reduce the bottlenecks in congested area of the network too. This also allows the traffic loads to be balanced in the network. Load balancing and reducing the bottleneck via alternate routing can significantly increase network reliability in WMNs.

Scalability: In wireless networks, when nodes are increased, the network performance will be affected. But, in WMNs, increasing the number of nodes will increase

transmission capacity for better load balancing and alternate routes. Usually, the local packets (generated in clients of mesh router) run faster compared to packets (generated in two or more hops away) from the neighbours.

Interoperability: WMN has a hybrid multi-point to multi-point architecture which is compatible with existing standards such as: WiMAX, Cellular, Wi-Fi, Bluetooth, Sensor, MANET, etc. All technologies mentioned above, are able or will be able soon to conFigureured a WMN and communicate with each others. Most of necessary improvements needed in any type of networks to enable them communicate with others; can augment the current standards to maintain interoperability.

A. Multi Radio Multi Channel (MRMC) WMNs

In MR-MC WMNs, each mesh router is equipped with multiple NICs and each NIC (Network Interface Card) can operate on multiple frequency channels. In the experimental MC-WMN test-beds in [3], each mesh router is equipped with two NICs. Providing up to four NICs is also considered reasonable [10]. Figureure 2 illustrates an example of an MR-MC WMN with six wireless mesh routers, three NICs per router, and five frequency channels. The label number indicates the assigned channels that are reused spatially.The MR-MC solution has attracted a lot of attention with the benefits of interference reduction and network scalability improvement in wireless mesh networks. Nevertheless, the MR-MC model also poses technical issues to be dealt with. As mentioned before, the number of available channels is limited to 3 or 12 within the IEEE 802.11 frequency bands. This implies that some logical links may be assigned the same channel. In this case, interference occurs if these logical links are close to each other, and hence the interfering links cannot be active simultaneously. Furthermore, the number of available NICs is also limited, and hence some logical links in a router need to share a NIC to transmit and receive the data packets.

When two logical links in a router share a NIC, they are required to operate over the same frequency channel, and cannot be active simultaneously. Thus, it significantly reduces their effective capacity. The effective link capacity can be increased by removing some of the links from the logical topology. However, when some of the links are not activated, the number of hops through some routing paths may be increased, and the logical topology may not even be connected. Therefore, how many logical links should be assigned between neighboring routers, how to allocate interfaces and channels, and through which logical links should the packets be forwarded need to be considered in MR-MC WMNs. Furthermore, given the physical topology of the routers and other constraints in MR-MC WMNs, four important issues that need to be addressed i.e., logical topology formation, interface assignment, channel allocation, and routing. Logical topology determines the set of logical links and network connectivity. Interface assignment decides how the logical links should be assigned to the NICs in each wireless router. Channel allocation selects the operating channel for each logical link. Finally, routing determines through which logical links the packets should beforwarded.

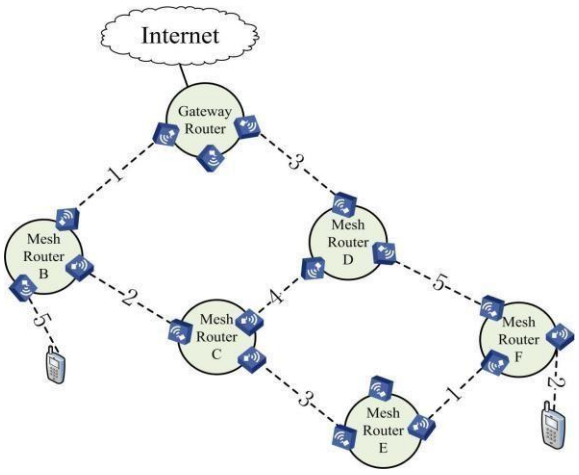


Figure 2. MR-MC WMN

Why Use Multi channel, multi radio?

- Increased Throughput.
- Decreased chance of collision/interference.
- Hardware cost is low enough to allow multiple radios (interfaces).

B. Channel Assignment

The Channel Assignment is an efficient tool to exploit multiple non overlapping channels to minimize interference and enhance the capacity of the wireless mesh network. It tries to find a feasible mapping between wireless channels and radio interfaces at each node with the aim of minimizing the interference and improving the capacity of WMN network. The frequency of IEEE 802.11b offers 3 non-overlapping channels, while IEEE 802.11a offers 12 non-overlapping channels. Multiple channels have been exploited in infrastructure-based networks by assigning different channels to adjacent access points[4] The goal of channel assignment in a multi-channel wireless mesh network is to bind each network interface to a radio channel in such a way that the available bandwidth on each virtual link is proportional to its expected load. A simple approach to the channel assignment problem is to assign the same set of channels to the interfaces of each node

C. Channel Assignment Designs Issues

Enable the node to be work on different channels from the neighbor nodes will lead to increase the node capacity. However, it may cause several problems that degrade the network performance.

Connectivity: The channel assignment has been used as the channel assignment should minimize total network interference while maintaining the network connectivity effective tool to minimize the WMN interference and increase the capacity of the multi-radio wireless mesh network.

Routing: The designing of an effective channel assignment for interference avoidance required knowledge

about traffic load for each link. The routing protocol can determine the total traffic load carried by each link in WMN which is the main factor of the traffic-aware channel assignment approaches. On the other hand, the routing decision depends on the network connectivity and the effective capacity of each link which depend on the result of channel assignment.

Mesh Client issues: The mesh client node can join and leave the mesh router dynamically. As the mesh client moves from an old mesh router and becomes close to a new mesh router, it should join the new mesh router. The process involves handoff steps and switches its interface to the channel associated with the new router. This new connectivity needs to establish a new route with a new channel in order to continue receiving the data. To minimize the handoff issues of the mesh client need to find proper network connectivity with less re-routed traffic.

Fault Tolerance: In multi-radio wireless mesh network where each node equipped with multiple radio, integrated the channel assignment with the fault tolerance issue become complex problem. Hence, the channel assignment should address the fault tolerance in such that when a link or a node is failed, the network does not suffer from the network partition and can work in self-healing fashion. This is because the channel assignment needs to select one link among the availability of multi-links per two nodes in such way that obtain the optimal solution.

II. EXISTING TECHNIQUES

S. Roy and Z.Niu [1] are investigate channel assignment for a multichannel wireless mesh network backbone, where each router is equipped with multiple interfaces. Of particular interest is the development of channel assignment heuristics for multiple flows. To maximize the aggregate useful end-to-end flow rates, both algorithms identify and resolve congestion at instantaneous bottleneck link in each iteration. The performance of our algorithm is shown to be significantly superior to best known algorithm in its class in multichannel limited radio scenarios.

Avallone et.al [2] proposed that fundamental design issues in multi-radio wireless mesh networks. They proposed a method to determine the pre-computed flow rates based on the max flow algorithm, in order to maximize the network throughput. This paper showed that not only the channels but also the transmission rates of the links have to be properly selected to make a given set of pre - computed flow rates schedulable. The availability of cost - effective wireless network interface cards makes it practical to design network devices with multiple radios which can be exploited to simultaneously transmit/receive over different frequency channels. It has been shown that using multiple radios per node increases the throughput of multi-hop wireless mesh networks.

X. Wang et al. [10] is presented to reduce inter-link interference and balance networks loads, this paper presents and evaluates a cross -layer algorithm for multi-radio multi-channel wireless mesh networks (WMNs), while conducting routing channel assignments. In channel assignment phase, we propose a greedy algorithm based on a multi-radio multi-channel interference model and select the channel with minimum of interference degree. In the routing phase, a new routing parameter which comprehensively considers interference, delay, load balancing and other factors is proposed. The new algorithm is compared with existing algorithm that is multi-radio multichannel WMN based on ad hoc on- demand distance vector routing (MRMC-AODV) to illustrate the advantages. The simulation results show that this algorithm can effectively improve the network throughput, reduce delay and decrease packet loss ratio.

P. Kyasanur et.al [6] studies the problem of improving the capacity of multi-channel wireless networks by using multiple interfaces. The paper considers the scenario when multiple interfaces are available, but the number of available interfaces is lesser than the number of available channels. They provide a classification of interface assignment strategies, and propose a new strategy that does not require modifications to IEEE 802.11. And also identify routing heuristics that are suitable for use with the proposed interface assignment strategy.

III. METHODOLOGY

The basic methodology of this technique consists of four steps that are as follows.

*Step 1.*Create a network of n nodes.
Step 2. Assign a channels based on least interference.
*Step 3.*Measure the flow control on each node.
Step 4. Assign the channel based on flow control.

I V. EVALUATION OFRESULTS

In this work, channel with the least interference will be assigned to the node. To find the Interference, channel capacity will be checked. If the channel capacity reduces to its threshold value, then it signifies that there is some interference in the channel. Channel capacity can b e evaluated using Channel capacity theorem.

Channel Capacity**:** The highest rate of information that can be transmitted through a channel is called the channel capacity, C.

C = B log (1+ S/N) bits/s

Where C is the channel capacity, B is the channel bandwidth in hertz, S is the signal power and N is the noise power.

V. RESULTS

For implementing the channel assignment algorithm, a network is created. The nodes in the network are assigned channel based on various factors. The interference among the various channels is evaluated. The channel with least interference is assigned to that node. Further, flow control is checked and again channel is assigned to that node.

In Figureure 3, Network is shown with only those nodes to which channels are assigned. Link index, Lin k rate for each link, Channel number and Interference value for each node are shown in the graph.

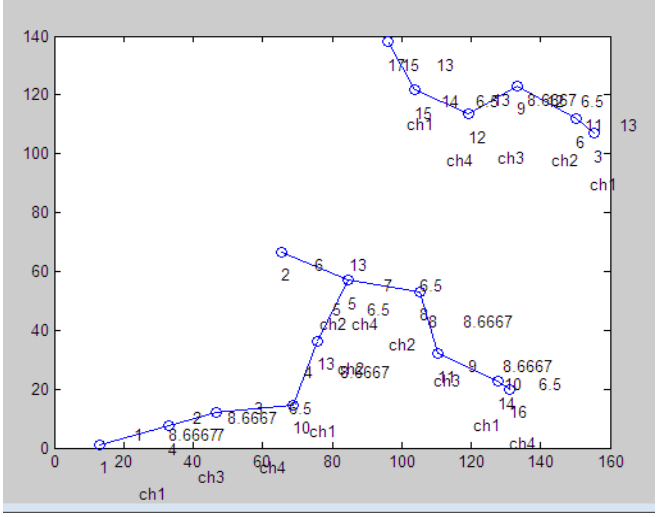


Figure 3.. Channel assignment based on flow control and interference

In Figureure.4, The interference of the channel (channel with least interference for a single node) against the number of nodes is plotted.

Figure 4. Graph for the Interference at each node

In Figureure.5, A table is formed with the number of nodes and channel assigned to each node.

nodes / channel	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
ch1	✓					✓				✓	✓					
ch2			✓		✓		✓						✓			
ch3		✓						✓								✓
ch4				✓					✓			✓		✓	✓	

Figure 5 Table for Channel assigned for each node

V I. CONCLUSION

Channel assignment is a network wide process where the allocation of non-interfering channels leads to significant throughput increase and better media access performance. The implemented algorithm allocates the channels based on less interference and flow control in multi-radio multi-channel wireless mesh network. Because only a number of channels are available, Co-channel interference exists between various radio signals which are running on the same frequency channel due to overlapping. It can severely affect the performance of WMN resulting in capacity reduction of the channel. In this algorithm, the channel assignment is done based on the least interference among the channels. Then, its flow control is evaluated and channels are allocated using the value of flow control.

V II . REFERENCES

[1] Sumit. Roy and Zhisheng Niu, " Flow Oriented Channel Assignment for Multi-Radio Wireless Mesh Networks, vol. 2010, Article ID 930414, pp. 10, Feb 2010.

[2] Avallone, S., Akyildiz, I.F. and Ventre, G., 2009. A channel and rate assignment algorithm and a layer-2.5 forwarding paradigm for multi-radio wireless mesh networks. IEEE/ACM T ransactions on Networking (T ON), 17(1), pp. 267-280.

[3] A. H. Mohsenian Rad and V. W. S. Wong, " Joint optimal channel assignment and congestion control for multi-channel wireless mesh networks," in Proc. IEEE ICC, 2006, vol. 5, pp. 1984–1989.

[4] A. Raniwala and T . Chiueh, " Architecture and algorithms for an IEEE 802.11-based multi-channel wireless mesh network," in Proc. IEEE IN FOCOM, 2005, vol. 3, pp. 2223–2234.

[5] J. T ang, G. Xue, W. Zhang, "Maximum Throughput and Fair Bandwidth Allocation in Multi-Channel Wireless Mesh Networks", INFOCOM 2006. 25th IEEE International Conference on Computer Communications, 2006.

[6] P. Kyasanur and N. H. Vaidya, " Routing and interface assignment in multi-channel multi interface wireless networks," in Proc. IEEE WCNC, 2005, vol. 4, pp. 2051 – 2056

[7] M. Kodialam and T . Nandagopal, " Characterizing the capacity region in multi-radio multi-channel wireless mesh networks," in Proc. ACM MobiCom, 2005, pp. 73–87

[8] A. Raniwala and K. Gopalan, and T . Chiueh, " Centralized channel assignment and routing algorithms for multi-channel wireless mesh networks," ACM Mobile Comput. Commun. Rev., vol. 8, no. 2, pp. 50–65, Apr. 2004.

[9] K. Jain, J. Padhye, V. Padmanabhan, and L. Qiu, " Impact of interference on multi-hop wireless network performance," in Proc. ACM MobiCom, 2003, pp. 66–80.

[10] X. Wang and L. Meng: " A joint routing and channel assignment in multi-radio multi-channel wireless mesh networks," Int. J. Sensor Networks, vol. 24, pp. 173-182, Nov 2017.

[11] P. Kyasanur and N. Vaidya, " Routing and link -layer protocols for multi-channel multi-interface ad hoc wireless networks," ACM SIGMOBILE Mobile Computing and Communications Review, vol. 10, no. 1, pp. 31–43, 2006.

[12] H. Skalli, S. Ghosh, S. K. Das, L. Lenzini, and M. Conti, "Channel assignment strategies for multiradio wireless mesh

networks: issues and solutions," IEEE Communications Magazine, vol. 45, no. 11, pp. 86–95, 2007.

[13] P. Bahl, R. Chandra, and J. Dunagan, "SSCH: slotted seeded channel hopping for capacity improvement in IEEE 802.11 ad-hoc wireless networks," in Proceedings of the 10th Annual International Conference on Mobile Computing and Networking (MobiCom '04), pp. 216–230, 2004.

[14] M. Lechery, R. Bhatia, and L. Li, "Joint channel assignment and routing for throughput optimization in multi-radio wireless mesh networks," in Proceedings of the 11th Annual International Conference on Mobile Computing and Networking (MobiCom '05), pp. 58–72, September 2005.

[15] J. T ang, G. Xue, and W. Zhang, "Cross-layer optimization for end-to-end rate allocation in multi-radio wireless mesh networks," Wireless Networks, vol. 15, no. 1, pp. 53 –64, 2009.

[16] L. Gao, X. Wang, and Y. Xu, "Multiradio channel allocation in multihop wireless networks," IEEE T ransactions on Mobile Computing, vol. 8, no. 11, pp. 1454–1468, 2009.

[17] C. Chereddi, P. Kyasanur, and N. H. Vaidya, " Design and implementation of a multi-channel multi-interface network," in Proceedings of the 2nd International Workshop on Multi-Hop AdHoc Networks: From T heory to Reality (REALMAN'06),pp. 23–30, 2006.

[18] J. So and N. Vaidya, " A routing protocol for utilizing multiple channels in multi-hop wireless networks with a single transceiver," T ech. Rep., University of Illino is at Urbana, Champaign, Ill, USA, 2004.

[19] B. Vedantham, S. Kakumanu, S. Lakshmanan, and R. Sivakumar, "Component based channel assignment in single radio, multi-channel ad hoc networks," in Proceedings of the 12th Annual International Conference on Mobile Computing and Networking (MobiCom '06), pp. 378–389, 2006.

[20] J. T ang, G. Xue, and W. Zhang, "Interference-aware topology control and qos routing in multi-channel wireless mesh networks," in Proc. ACM MobiHoc, 2005, pp. 68–77.

[21] S. Ramanathan, " Unified framework and algorithm for channelassignmentinwirelessnetworks,"WirelessNetworks,vo l.5, no. 2, pp. 81–94, 1999.

[22] A. P. Subramanian, H. Gupta, and S. R. Das, " Minimum interference channel assignment in multi-radio wireless mesh networks," in Proceedings of the 4th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks (SECON '07), pp. 481–490, 2007.

Studying Effects of training a neural net after expected results are obtained.

Harjot Singh Baidwan
Com puter Science & Engineering
Chandigarh Engineering College
(Affiliated by PTU)
Mohali,India
hsb1821@gmail.com

Gurleen Kaur
Com puter Science & Engineering
Chandigarh Engineering College
(Affiliated by PTU)
M ohali,India
gurleenkaur15@gmail.com

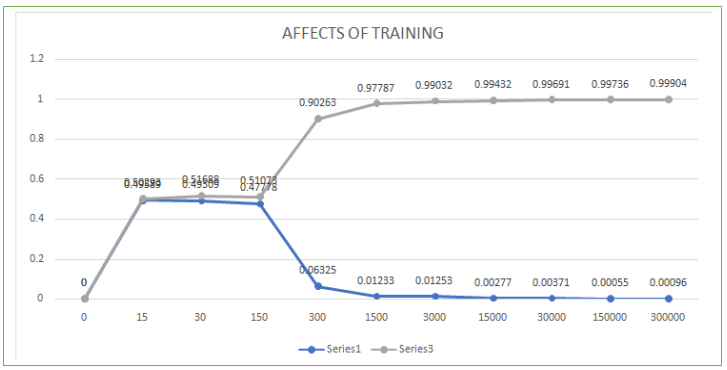
***Abstract:*-O ne of the important parts of building a working artificial ne ural ne twork mode l is to train it In this re port we trie d to train a single laye red ne ural ne twork mode l as X-NO R. Usi ng back-propagation one can corre ct the we ights to achie ve e xpe cte d re sults .O nce that point is achie ve d it is important to unde rstand to what de gre e multiple back -propagation affe ct the ne ural ne t.**

I. INTRODUCTION

Neural networks are a set of algorithms that are designed to recognize patterns, similar to the way human brain recognizes them. Artificial Neural Networks interpret sensory data by machine perception, labelling or clustering raw input. They recognize patterns in which all real-world data like images, audio, texts,are translated into numericalvectors. Neural Networks are used to extract patterns and detect trends which complex to be noticed by human beings or other computing techniques. A trainer neural network can be used to provide projections in a field of information to give solutions to questions and new situations of interests. It also provides real time operations, adaptive learning, self – organisations and fault tolerance via redundant information coding.

I. DIFFERENT PHASES IN TRAINING A NEURAL NET

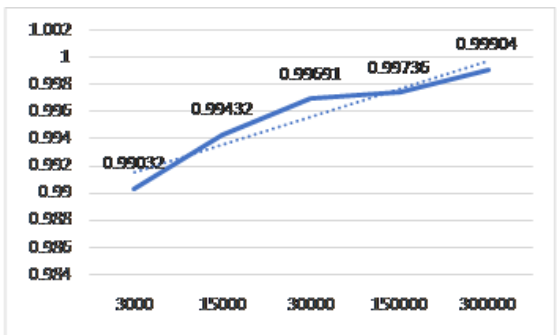
Here we try to train a model to act as an X-nor Gate i.e., it will give output one on even no. of one's.In the graph below we can see the till 150 weights the updatingmodel hasn't be accurate enough but At 300th iteration the achieves Considerable accuracy Enough for general purpose XNOR gate.



The above table represents inputs,expected output of the model. It also the actual output after training the model for certain number of times.

I. INFERENCE

Since a Single Hidden layered neural network is used the back-propagation works well .We observe that increase in accuracy of output w.r.t the training iterations is not constant .This creates ambiguous situation which makes it difficult for the user to quantify the amount accuracy gain one can have with increasing training iterations to huge numbers.



Inputs(X-inputs)			Expected output	Training(N TIMES)										
				0	15	30	150	300	1500	3000	15000	30000	150000	300000
0	0	1	0	0	0.49589	0.49309	0.47778	0.06325	0.01233	0.01253	0.00277	0.00371	0.00055	0.00096
0	1	1	1	0	0.50460	0.48707	0.51219	0.88511	0.97826	0.98346	0.99400	0.99528	0.99740	0.99872
1	0	1	1	0	0.50293	0.51688	0.51073	0.90263	0.97787	0.99032	0.99432	0.99691	0.99736	0.99904
1	1	1	0	0	0.51050	0.50957	0.51904	0.12764	0.02424	0.01447	0.00640	0.00409	0.00301	0.00115

REFERENCES

[1]. A Beginner's Guide to Neural Networks and Deep Learninghttps://skymind.ai/wiki/neural-network.

[2]. NEURAL NETWORKSby Christos Stergiou and DimitriosSiganoshttps://www.doc.ic.ac.uk/~nd/surprise_96/journal/vol4/cs11/report.html

[3]. G. Strawn and C. Strawn, "Masterminds of Artificial Intelligence: Marvin Minsky and Seymour Papert," in IT Professional, vol. 18, no. 6, pp. 62-64, Nov. Dec. 2016. doi: 10.1109/MITP.2016.116 ,URL: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7763722&isnumber=7763693

[4]. Werbos, Paul & J. (Paul John, Paul.)(1974). Beyond regression: new tools for prediction and analysis in the behavioral sciences.

[5]. Werbos, Paul. (1990). Backpropagation through time: what it does and how to do it. Proceedings of the IEEE. 78. 1550 - 1560. 10.1109/5.58337

A REVIEW OF DIGITAL IMAGE PROCESSING APPLICATION AREAS

Manjot Kaur
Department of Computer Science Engineering,
Chandigarh University
Mohali , Punjab (India)

Navneet Kaur
Department of Computer Science Engineering,
Baba Banda Singh Bahadur Engineering College
Fatehgarh Sahib,Punjab (India)

***Abstract*—this paper discusses about digital image processing applications. As we all are living in digital world, images are important part of our lives and we can extract information from these images using image processing techniques. Extracted information can be used further in various field such as medical (for detecting various diseases), biometrics (for keeping data authorized and secure) and object detections etc. so we are going to discuss most of these applications.**

I. INTRODUCTION

Digital Image is basically 2D signal and It is characterized by the mathematical function f(x,y), where x and y are the two co-ordinates in plane. This mathematical function gives us pixel value at that point of image. Below a digit image has been shown which is nothing but 2D array whose values lie between 0 to 255.

Figure.1. Digital Image [1].

II. IMAGE PROCESSING TECHNIQUES

- A digital image is preprocessed before extracting information to make it more clearly for fetching information.

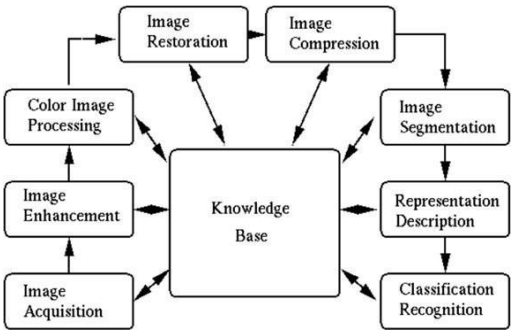


Figure.2. Image Preprocessing Steps [2].

- Image Acquisition is foremost step of image preprocessing it involves scaling.it begins with the catching of a picture by the sensor, (for example, a monochrome or shading TV camera) and digitized. On the off chance that, the yield of the camera or sensor isn't in computerized frame then a simple to - advanced converter (ADC) digitizes it. In the event that the picture isn't appropriately gained, you won't have the capacity to accomplish assignments that you need to. Modified equipment is utilized for cutting edge picture securing strategies and techniques. 3D picture procurement is one such propelled strategy picture obtaining technique [3].

- Image Enhancement is used to find out information that is unclear, hiddenor to highlight specific features according to the requirements of an image. Such as changing brightness & contrast etc. Basically, it involves manipulation of an image to get the desired image than original for specific applications. Many algorithms have been designed for the purpose of image enhancement in image processing to change an image's contrast, brightness, and various other such things. Image Enhancement aims to change the human perception of the images. Image Enhancement techniques are of two types: Spatial domain and Frequency domain [4].

- Image Restoration is basically objective image enhancement which is based on mathematical or probability based models. It can also be used to restore information hidden due to blurring of an image [4].
- Color Image Processing has been turned out to be of incredible intrigue in light of the critical increment in the utilization of computerized pictures on the Internet. It incorporates shading displaying and handling in an advanced area and so forth. There are different shading models which are utilized to determine a shading utilizing a 3D organize framework. These models are RGB Model, CMY Model, HSI Model, and YIQ Model. The shading picture preparing is done as people can see a large number of hues. There are two territories of shading picture preparing full-shading handling and pseudo shading preparing. In full-shading handling, the picture is prepared in full hues while in pseudo shading preparing the grayscale pictures are changed over to hued pictures. It is an intriguing theme with regards to picture preparing [4].
- Wavelets and Multi Resolution Processing is used to represent images with varying resolution. Images are divided into smaller parts for data compres sion. Wavelets are mathematical functions which are used to divide image on different frequencies [4].
- Compression is used for compress the size of image to easily store on computer or transmit over internet there are various compression algorithms [4].
- Morphological Processing is set of tools used to describe the shapes of objects in images and these operations can be applied to greyscale images [4].
- Segmentation is a process of dividing images into parts. it is a rough division methodology that takes far toward a fruitful arrangement of imaging issues that expect items to be distinguished exclusively. In basic terms, picture division implies dividing a picture into different sections for improvement and changing the portrayal of the picture [4].
- Representation and Description image depiction relies upon the yield of a division stage and it incorporates pixel information, comprising either every one of the focuses in the rule Picking an image is a piece of the answer for change information into an appropriate frame that permits consequent PC preparing. As description manages extricating traits that yield quantitative data of intrigue or fundamental to isolate one class from another [4].
- Object Recognition it involves assigning tags to objects based on descriptors. Object can be recognized either in image or video. It mainly includes machine learning techniques.
- Knowledge Base is collection of data requires to process an image and it is complex to create a knowledge base. It interprets or deduce high level information from low level information [5].

III. IMAGE PROCESSING APPLICATIONS

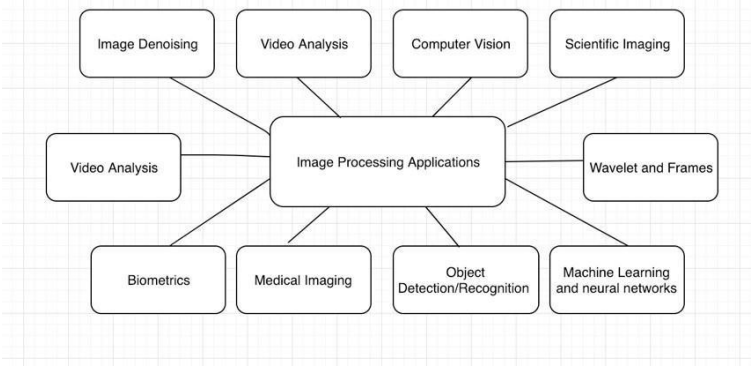


Figure.3.Image Processing Applications.

Biometricsis widely used image processing area. Here physical features are used to construct secure devices such as fingerprint, iris etc. here data of physical features has been stored in database then some computation is performed and matching is performed. It is mainly used for security related applications. Some times more than one physical feature is combined to make more secure systems.

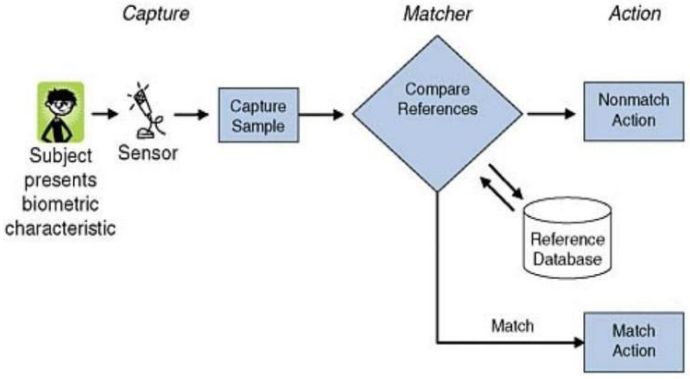


Figure.4. Biometric System [6].

Machine Learning &Neural networks are the fields of Artificial intelligence. Machine learning is a model where system learns itself based on training data sets and examples. The system adapts itself and solves the real world problems. Neural network is based on biological model of neurons which emulates the human brain and works in a real time environment and solves the problem like human being solves. Initially it learns from the sample data from the training sets and learns and makes the system capable enough to deal with the problems [7].

Object Detection/Recognition is used to detect hidden objects from the images. Before starting detecting objects, images are preprocessed using various filters and then us ed to find hidden data such applications are mainly used for reading characters also like in case of vehicle number plates are read by traffic police.now days real time object can also be detected. Object detection can also be done using machine learning for example deep learning algorithms are used to detect objects from images, videos [8].

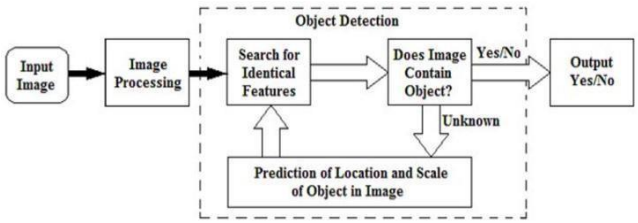


Figure.5. Object Detection Algorithm Example [8].

Image Denoising is a technique to make images noise free.it can be done by various noise models such as addictive etc. selection of algorithm is application based means algorithm varies if you want to use image for different purposes. Image is denoised to retrieve an original image.

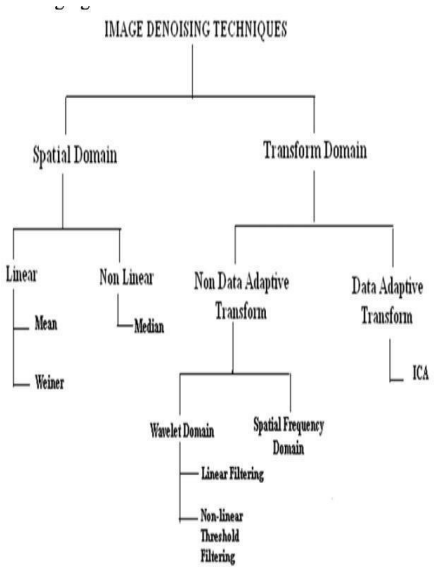


Figure.6. Image Denoising Classification [9].

Computer Vision is set algorithms which enables compu ter to perceive the images as human does. It can be applied either on single image or set of images. Its main aim to develop understanding in computers to analyze images as humans. It is a widely used in agriculture, augmented reality, robotics and many more.

Video Analysis is widely used as videos are mostly commonly developed by user and also watched by viewers. Video analysis mainly done in sports, film industry and to detect robbery. So various algorithm has been developed to analyze videos [10].

Scientific Imaging is somewhat different to normal imaging. It is mainly deal with space, galaxies, stars etc. scientific images mainly contain information. The topic of the picture is, basically, logical. It might be an image of an item, for example, the ink well from an archeological site, or a photomicrograph of a cross segment taken from that inkwell. Notwithstanding logical topic, the pictures may include logical instrumentation, for example, magnifying instruments, telescopes, or different instruments. Reg ardless of whether improved the situation tasteful reasons, for example, photomicrographs, despite everything they qualify as logical pictures. The imaging vitality can be photons, for example, with the light magnifying instrument, or electrons, as in the checking electron magnifying lens, or in ultrasound waves, which have given eager guardians such bliss to see their creating youngster [11].

Wavelet and Frames provide us a tool to obtain expansions in Hilbert spaces. However, the frame conditions are significantly weaker, which makes frames much more flexible. Frame theory has attracted much attention in recent years, especially in connection with its concrete manifestations within Gabor analysis and wavelet analysis. In this article we give a short overview of the general theory for frames in Hilbert spaces, as well as its applications in Gabor analysis and wavelet analysis. Finally, we present a method that allows to construct wavelet frames based on certain Gabor frames, and vice versa. Applying this to Gabor frames generated by exponential B-splines produces a class of attractive dual wavelet frame pairs generated by functions whose Fourier transforms are compactly supported splines with geometrically distributed knots [12].

Biomedical imaging these images deal with humans and animals. Initially it started with X-Ray to detect diseases and train the machines. It is best way to view internal organs and diseases of humans. Ultrasound imaging is fast and flexible. CT scan is alternative to X-Rays and ultimate tool to train systems. MRI are used for clinical neurology and used for detecting tumours. Mammography is used to detect early stage cancers [13].

I V. CONCLUSION

Above we have discussed what is digital image processing and their wide areas where it can be applied. So concluding these entire digital image processing is widely used in almost every area as world is full of images. But some images are distorted so first we need to apply filter to enhance the images and then it can be used various fields and every field further helping to improve our lives whether it be medical images or scientific images. These all areas are focusing on new advancement to serve society.

REFERENCES

[1] https://www.tutorialspoint.com/dip/image_processing_introduction.htm.D igital Image processing introduction.

[2] http://fourier.eng.hmc.edu/e161/lectures/e161ch1.pdf Image Processing and Related Fields.

[3] Michel P. Ekstrom.Digital Image Processing T echniques, Academic press,Inc ,1984.Chapter 1: image enhancement pp. 1-48.

[4] https://www.techsparks.co.in/thesis-topics-in-digital-image-processing/ What is image processing.

[5] Ms. Kavita Khobragade,Knowledge Based Image Processing Applications,3RD nation confrence of advanced computing technologies,2012.

[6] https://www.nap.edu/read/12720/chapter/2#2. Summary.

[7] Olivier Lezoray,Chirstophe Charrier,Hubert Cardot and Sabstein Lefevre, Machine Learning In Image Processing, EURASIP Journal On Advances in Image Processing,2008.

[8] Kartik Umesh Sharma and Nileshsingh V. T hakur,A Review And Approach For Object Detection In Images,Internation Journal of Computional Vision and Robotics,Vol.7,2017.

[9] VijayaLakshmi. A.,Titus C. and Lilly Beaulaha H.,Image Denoising For Different Noise Models By Various Filters:A Brief Survey,IJET TCS,Vol.3 ,Issue 6,2014

[10] J.R. Wang and N.Parameswaran,Survey Of Sports Video Analysis:Research Issues And Applications, Preceeding of pan sydney area workshop on Visual Information Processing,pages 87 -90,2004.

[11] https://www.mccrone.com/mm/scientific-imaging-webinar/. What is scienfic imaging why it is important.

[12] Christensen Ole,A Short Introduction T o Frames, Gober System And Wavelet System,Azerbaijan journal of mathamatics,2014.

[13] Sukru Mehmet Erturk, Cairan Johnston,Claire Tempany-Affdhal,Annick D.Van Den Abbeele,Clinical and Translation Science ,Academic Press ,Page 87-104,2009.

Attendance system using face recognition

Anju Bala
Assistant Professor, CSE Department
Chandigarh University
Chandigarh, India
Anju.cse@cumail.in

Bhavna Upadhyay
Student BE, CSE Department
Chandigarh University
Chandigarh, India
cu.16bcs1367@gmail.com

***Abstract*— Face identification and acknowledgment has picked up a great deal of notoriety in the space of Image Processing in the most recent couple of years and specialists have possessed the capacity to actualize it in diffe rent fields of our everyday life including security purposes, sexual orientation arrangement, human PC communication, biometric control and so forth. This paper points toward another effective execution of Attendance framework utilizing face identification and acknowledgment. Physically taking participation by educators cause a great deal of control and a considerable measure of disarray on occasion. The mechanized framework for participation outlined by us keeps the additional intermediaries and control of the information. It likewise spares the time spent on tallying the understudies and taking participation, which thusly can be utilized for improvement of understudies. Our Attendance framework bargains of four steps i.e. making a database, taking a gathering photograph of class, confront division, coordinating the face with the database. This paper comprises a fundamental presentation of our work, past chips away at a similar theme, a proposed architecture, simulation and results and end.**

KEYWORDS: PCA-LDA, segmentation, Ada-Boost, cascading classifier, background regions.

I. INTRODUCTION

Physically taking participation causes a bunches of inconsistencies in the participation report and chances are there of control of information. Furthermore, it makes the educator's activity dull and some measure of time out of aggregate dispensed time for showing results in an aggregate wastage. Face discovery and acknowledgment can be effectively actualized here to check the participation. The picture is caught through a top quality camera amid the class hours. Countenances of understudies are distinguished, divided and put away for confirmation with database. After the fruitful match of countenances, the participation is set apart without anyone else.

Our framework chips away at 3 divisions which we will be clarified in later part. A presentation for all the 3 divisions is being given to have an appropriate comprehension.

The first is making a database of the understudies; framework utilizes python code utilizing opencv to set up a database of the individual sitting before the camera in different temperaments and positions. This is just a single methodology. The pictures caught and divided will be added to the database for the nonstop refresh of the database.

The second one is confront location; a superior quality camera is utilized to catch the photograph of the entire class. The caught picture is then divided into faces utilizing an outstanding calculation i.e. haarcascades planned by viola jones for eyes and confront location. This calculation disposes of the issues of light, revolution and scaling to some degree.

The third one is confront acknowledgment; framework utilizes nearby twofold example (LBP) to perceive the face. In this the face is caught and separated into various squares. Histograms of each square is made and square histogram are moved into Confront picture. The first LBP administrator labs the pixels of picture by thresholding the pixels with the middle pixel. In the event that the estimation of the neighbor pixel is more prominent than or equivalent to the Centre pixel, the neighbor pixel is appointed as 1 else 0.

.

II. EARLIER WORKS

Face recognition utilizing LBP was proposed by Jo Chang - yeon. His paper portrays how the LBP functions and how it is utilized for confront perceiving. The result of his paper was that LBP is more straightforward than haar-like highlights and quicker to segregate among faces and non faces.

Face location and acknowledgment for programmed participation framework was proposed by Dr. Nita Thakare, Meghna Shrivastava,NidhiKumari, Neha Kumari, Darleen Kaur, Rinku Singh which goes for the effectiveness of the system.Automated Attendance Management System utilizing Face Recognition was proposed by Mrunmayee Shirodkar, Varun Sinha, UrviJain, Bhushan Nemade. Their paper depicts the ongoing programmed participation framework. Their created calculation in taking participation is accomplishing 83.2% proficiency.

RESULT OF ESTIMATING THE ATTENDANCE BASED ON CONTINUOUS OBSERVATION

We compared the results one cycle only and continuous observation. 12 students existed in the center area, And 2 of them did not have their faces registered. In this experiment of 79 minutes, 8 scanning cycles were completed during this period. Table 2 shows face detectionrate, and Table 3 shows the result of estimating the attendance. In the case of 1 cycle only, we judge the recognized students to be present. In the case of continuous observation, the system estimates the attendance by the method of section 3.5 using the recognition data obtained during 79 minutes. This table shows that continuous observation improved the face detection rate and improved F-score of estimation of the attendance, which is the harmonic mean of precision and recall.

CONCLUSION

The brilliant and robotized participation framework can be demonstrated as a proficient framework for classroom participation. By utilizing this framework, the odds of phony participation and intermediaries can be decreased. There are a heaps of Biometrics Systems which can be utilized for overseeing participation, yet the face acknowledgment has the best execution. So we have to execute a solid and proficient participation framework for classroom participation which can work for different face acknowledgment at one time. Also to execute this framework, no any particular equipment is required. A camera gadget and an independent PC, database servers are adequate for building the shrewd participation framework. With the assistance of a unique blend of calculations, this framework causes us to accomplish wanted outcomes with better precision and less time utilization.

ACKOWLEDGEMENT

We are grateful to every one of the creators who dealt with these innovations and guided us to contribute in it. We are significantly thankful to Prof. Naveen Kumar N for his master direction and persistent support to see this task rights since its initiation.

REFEREN C ES

[1] AjinkyaPatil, MridangShukla,"Implementation of Class Room Attendance System Based on Face Recognition III Class", IJAET (International Diary of Advances in Engineering and Technology), Vol. 7, Issue 3, July 2014.

[2] Naveed Khan Baloch, M. HaroonYousaf, Wagar Ahmad, M. Iran Baig, "Calculation for Efficient Attendance Management: Face Recognition based Approach", IJCSI, Vol. 9, Issue 4, No I, July 2012.

[3] Paul Viola, Michael J. Jones, "Vigorous Real-T ime Face Detection" International Journal of Computer Vision Volume 57 Issue 2, pp. 137 – 154, May 2014.

[4] Zhang, Ligang, T jondronegoro, Dian W., Chandran, Vinod, "Discoveringthe best component extraction and choice calculations for unconstrained outward appearance acknowledgment" IEEE Conference on Multimedia and Expo, Issue 2, pp.1027 -1032, Feb-2012.

[5] J. Chatrath, P. Gupta, P.Ahuja, A.Goel, "Constant Human Face Detection and T racking" International Conference on Signal Processing and Coordinated Networks (SPIN) Volume 1, Issue 1, pp. 705-710, 2014.

[6] Abdur Rahim, Najmul Hossain, T anzillah Wahid, ShafiulAzam, "Face Recognition Using Local Binary Patterns" International Journal of Software engineering and T echnology Graphics and Computer Vision, Volume 13, Issue 4, pp.817-823, 2013.

[7] DC. He and L. Wang (1990), "Surface Unit, T exture Spectrum, And T exture Analysis", Geoscience and Remote Sensing, IEEE T ransactions on, vol. 28, pp. 509 -512.

[8] Sapna Vishwakarma, Prof .Krishan Kant Pathak, "Face Recognition utilizing LBP Coefficient Vectors with SVM Classifier" International Journal of Building T rends and Technology (IJETT), Volume 9, Issue 2, pp.96-100, Mar-2014.

[9] M. Pati!, Dr. Satish R. Kolhe, Dr. Pradeep M. Pati!, "Face Recognition by PCA Technique", Second International Conference on Emerging Patterns in Engineering and Technology(ICETET).

[10] T iwari Priti Anilkumar, Kalyani Jha, Karishma P Uchil, Naveen H., "Haar Features Based Face Detection and Recognition for Advanced Classroom and Corporate Attendance", IJIRCCE, Vol.3.

EMERGENCE OF ARTIFICIAL INTELLIGENCE IN DEFENSE SECTOR

Neelam Rana
CSE Departm ent
Baba Banda Singh Bahadur Engineering College
Fatehgarh Sahib
rana100896@gmail.com

Abstract:-This paper throws light on the emerging trends in artificial intelligence that can be equipped into military systems and weapons to empower the defense forces of a nation and make them more withstanding against numerous technologies. It highlights the key areas where artificial intelligence could be blended with the defense support systems such as to keep the military systems secure, sensitive data and information protected, border activities under surveillance, target recognition of enemies, huge volumes of data intact and prevention against cyber-attacks and much more.

The topic has been thoroughly discussed by also taking into account the downsides of this emerging technology. The viewpoints from various tech giants have been gathered that how their serious concerns towards the enormous potential of artificial intelligence that can put human existence into danger. Examples have been cited that show various nations that have already advanced in the field of automated weapons and ammunition. The status of India has also been explained by listing out the major developments made in this field and ongoing researches. Thus, the paper describes the current outlook of the world towards the incorporation of artificial intelligence into military support systems for enhanced capabilities and efficacy.

I. INTRODUCTION

Artificial intelligence, simply put, is the use of computers to perform tasks that normally require human analytical skills. There are three classes or levels of AI. The most basic class of AI is artificial narrow intelligence (ANI) which encompasses technologies that are designed for specific, limited purposes. ANI examples include: the algorithm behind Google Translate, anti-lock brake systems, or facial recognition technology. The second class of AI, which has not yet been realised, is that of artificial general intelligence (AGI), which would be capable of operating across all areas, including cognitive, that the human brain does. The most distant class of AI, and also its final incarnation, is considered artificial super intelligence (ASI). This form of AI, which will be by far the most challenging to achieve, describes intelligence that exceeds the capacity of the human brain [2].

Artificial intelligence has an immense potential to revolutionize the militaries of the nations. To every nation, the importance of its armed forces is supreme. Guarding the country's borders and territories, facing close combats, militant attacks and countering infiltrators is a demanding job. One never knows what superior technology your enemy could have. So the only option is to be prepared with every possible systems, machines, weapons and war crafts that help the military in facing every extremity. Paired with developments in robotics, we can now imagine a wholly different kind of warfare, with swarms ofdrones and autonomous ground vehicles taking the place of boots on the ground. Robotics and artificial intelligence is a combination of technology and cognitive intelligence for simulation, processing of information and knowledge to build capability in a machine to imitate human behavior. AI improves self-control, self-regulation and selfactuation of combat systems due to its inherent computing and decis ion-making capabilities. There are many reasons why AI in the battlefield presents an attractive prospect to many militaries. It can be used to save lives – both by keeping human soldiers out of the most dangerous parts of the battlefield but also avoidin g collateral damage and inadvertently targeting civilians [10].

The development in defense technology requires increased defense spending year by year for the modification or replacement of updated weapons. However the recent trend to replace conventional s ystems with intelligent robots that are adaptive to learning new methods by observation and trials provides a cost effective solution.

II. LITERATURE SURVEY

Since the advent of Artificial Intelligence in the 1980's, it has been subjected to sheer contemplation and research. A number of research papers and journals have been published to account for the progress made in this field and more upcoming researches. In a research paper by M.L. Cummings titled "artificial intelligence and future ofwarfare", he had discussed that though machines have better capabilities than humans in skill and rule based tasks, but still in situations involving judgment, knowledge and significant uncertainties, humans are superior to machines. He also states that the global defense industry is falling behind its commercial counterparts due to outmatched funding [1]. Dr Jean-Marc Rickli in his research paper "The Economic, Security and Military Implications of Artificial Intelligence for the Arab Gulf Countries" states that in terms of military and

security consequences, selforganised collective decision-making in swarms of autonomous agents will likely become a defining feature of future battlefields. He expressed his concern malicious actors has use of AI having the potential to magnify the tensions in the Gulf region and increase its instability [2]. Enn Tyugu in one of his articles "artificial intelligence in cyber defense" has discussed about the fact of achieving defense against intelligent cyber weapons only by intelligent software, and has mentioned example of the Conficker worm that affected some military and police networks in Europe. Further, he has discussed about the existing artificial intelligence applications in cyber defense and suggested new intelligent applications [14]. In a report by Michael E.O'Hanlon "The role of AI in future warfare", he predicts that by the year 2040Robotics and AI could take on a central, and very important, role in warfare even without anything resembling a terminator or a large killer robot [3].in a reportby Luukas Ilves, the Lisbon Council "Responsible, Safe and Secure AI", views have been shared on the gradualevolution of automation in weapons systems and the significant research programmes led by the world's major military powers in autonomous weapons. While the US has the lead in such research, other countries are catching up. And Vladimir Putin has said of AI that "Whoever becomes the leader in this sphere will become the ruler of the world [10]. Group Captain Atul Pant (IAF) has given his views on "aerial drone swarms: the next generation military weapon". According to him swarms are likely to reduce the monetary costs of war and could readily take on tasks in contaminated environments having radiation, chemical or biological hazards or in dangerous situations where the odds are heavily against own forces and likely casualties to own forces could be high[13]. Moreover , some thinkers and experts have emphasised on the need of India to take up the challenge of lethal autonomous weapon systems and need for controlling the development of autonomous weapons and the extent of automation so that no innocent lives are on a stake [4][11].

III. APPLICATIONS

Artificial intelligence is paving the way for many future technologies that could have major applications in the defense sector.AI is deployed in almost every military application, and increased research and development funding from military research agencies to develop new and advanced applications of artificial intelligence is projected to drive the increased adoption of AI-driven systems in the military sector. The major areas in which artificial intelligence can prove to be useful for the defense forces of a nation are listed under the subsequent headings [11].

A. Data information processing

Artificial intelligence is particularly useful for the advanced analysis of sensitive data through quick, efficient and accurate processing of huge volumes of data resulting in valuable information.AI assists in culling and gathering this information from various distributed sources as well as datasets.

B. Warfare platforms

Various nations are empowering their defense forces by inducting AI based weapons and other systems into their land, naval, airborne and space platforms. This incorporation of AI with warfare systems is leading to enhanced efficiency and performance by enabling execution of collaborative tasks and also, requiring less human input and maintenance.

C. Cyber security

Since, the armed forces need to deal with highly classified information at most of the times, Therefore, cyber security is a major field of concern. Military sys tems, being vulnerable to cyber-attacks, the threat of information loss is ever prevalent. However, AI enabled web security systems can autonomously protect networks, compu ters and data from unauthorized access as well as record the pattern of cyber-attacks and develop counter attack tools for prevention against them.

D. Logistics and transportation

The transportation of goods, ammunition, armaments and troops plays an important role in the conduct of successful military operations. Integration of AI with military transportation helps in easy detection of anomalies, component failures and thus, reducing human operational efforts. Recently, the US Army collaborated with IBM to use its Watson AI platformto help pre- identify problems in Stryker combat vehicles .

E. Target recognition

AI systems are being developed in order to enhance the accuracy of target recognition systems.AI improves the ability to locate target positions in complex combat environments. Also, Machine learning can be complied with AI for improved efficiency and support.

F. Battlefield health care

The life of the soldiers is precious. In warzones, AI equipped robotic surgical systems and robotic ground platforms can be deployed so as to provide remote medical support. The US is currently involved with the development of such technologies. For instance , a step has been put forth by the IBM based Watson company in collaboration with some US veterans administration to develop a clinical reasoning prototype known as electronic medical record analyzer(EMRA). This

preliminary technology uses machine learning to process patients' medical records. The integration of AI will beneficial by adding even more functionalities.

G. *Surveillance and reconnaissance*

Unmanned aerial systems – also known as drones - with integrated AI can patrol border areas, identify potential threats and transmit the related information to reaction teams. These AI assisted systems provide situational awareness to the defense personnel which plays a significant role in taking the necessary actions and thus increasing the safety and efficacy of the deployed army men.

V. DISCUSSIONS

In the context of artificial intelligence systems being inducted into the defense forces, there always prevail some debates and counter-attack opinions stating the promises and perils of this technology. Along with the rise of AI cum Robotics equipped machines, weapons and other systems seeming to open new ventures in the military departments, the danger towards humanity is also coming into picture. Artificial intelligence (AI) since the early 2010s has witnessed a series of technical breakthroughs due to the increase in computing power, the amount of data generated and the application of machine learning techniques. It is deployed in almost every military application, and increased research and development funding from military research agencies to develop new and advanced applications of artificial intelligence is projected to drive the increased adoption of AI-driven systems in the military sector[5].AI is no newcomer to the battlefield. Advanced weapons systems already acquire and engage targets with no human intervention. During the 2003 Iraq war, autonomous air defense missile systems were responsible for the majority of airborne friendly fire incidents [10].

Countries like US, Israel, Iran, UK and Russia have already pioneered the field of fully automated weapons. Some of the well-known weapon defensive systems from these countries are like the Missile defense systems such as the Iron Dome of Israel and the Phalanx Close-in Weapon System used by the US Navy. UK has Fire and forget systems such as the brimstone missile system. The harpy air defense suppression systemof Israel is an anti-radiation loitering weapon capable of finding and attacking radar installations autonomously. SGR-A1 is a sentry robot deployed by South Korea in the demilitarized zone with its northern adversary. Uran-9, attackedarmoured vehicle controlled remotely by an operator equipped with 2A72 30-millimeter cannon with a rate of fire of 350 to 400 rounds per minute and can shoot high explosive incendiary and armour-piercing ammunition. A Swedish Maritime Robotics Centre is developing the underwater systems of tomorrow.

When it comes to India and its ventures in AI landscape, one finds that India also does not want to lag in the race of nations equipping with artificial intelligence. The beginning has been made with Unmanned Aerial Vehicles (UAV), Unmanned Ground Systems (UGS) and guided bomb and missile systems. Indian defense services are presently using indigenously designed Remotely Operated Vehicle (ROV) Daksh for Bomb Disposal, Unmanned Autonomous Vehicles (Netra UAV, Rustom, Searcher etc.) for reconnaissance and other mini robot machines. DAC has approved approx. 544 Robots for Indian Army from indigenous source. The robots will be used for surveillance and can deliver suitable ammunition. CAIR, a DRDO lab, has been working on a project to develop a Multi Agent Robotics Framework (MARF) for more than eight months now. MARF will equip India's armed forces with an array of robots that Indian Army has already built like Wheeled Robot with Passive Suspension, Snake Robot, Legged Robot, Wall-Climbing Robot, and Robot Sentry, among others and worktogether as a team. India is also heading towards Aerial Swarm Drones as the next generation military weapons as well concentrating into building more stealth technology weapons. India is also trying to build more indigenous stealth technology so as to empower its forces even more.

The issue of LAWS (LETHAL AUTONOMOUS WEAPON SYSTEMS) is also posing a serious challenge that cannot be just wished away. LAWS raises out serious questions about ethics [5].The UN, through the Convention of Certain Conventional Weapons (UNCCW), has debated the issue ofautonomous weapons systems (AWS) since 2014.To qualify as a fully autonomous system, a weapon should fulfil at least three core functions ofits engagement cycle autonomously:the search of the objective, the decision to engage and the engagement of the target. Such a weapon should be able to move independently through its environment to arbitrary locations; select and fire upon targets in their environment and create and or modify its goals, incorporating observation of its environment and communication with other agents[2]Recently, there was a request to United Nation to cancel permission t o use 'killer robot' called as 'Pandora Box'. It can be used to conduct warfare bin future by using artificial intelligence. Once the 'Pandora Box' is opened or activated it cannot be controlled or deactivated easily. These can be weaponsof terror and beyond human control [4]. Granting full rights to machines and allowing them to take decisions at their own and responding in the similar way can be far too dangerous than even our minds can think of. Machines won't think and act like human soldiers by keepin g into consideration ethics and moral values, instead, a single disruptance in the machinery can cause huge causalities. Great pioneers in the field of technology like Bill Gates, Elon Musk, Sundar Pichai and Stephen Hawkings have also expressed their sincere opinions on Artificial

Intelligence. "I am increasingly inclined to think that there should be some regulatory oversight, maybe at the national and international level, just to make sure that we don't do something very foolish. I mean with artificial intelligence we're summoning the demon"- Elon Musk warned at MIT'S AeroAstro Centennial Symposium. Even Stephen Hawkings once said" AI is likely to be either the best or worst thing to happen to humanity". Recently, The CEO of Google Sundar Pichai also said that a malfunctioning in the artificially intelligent robots can make them violent against humanity. They may start thinking of humans as their enemies and start killing them or controlling themand this scenario could also be the end of this world. Therefore, what needs to be understood is that technology is created by humans for our assistance and not to control humans. Autonomous technologiesare of huge importance to the military department. Banning of any autonomous weapon is not a viable solution because one never knows what his adversary is carrying. So what needs to be taken into due consideration is that how the development of such autonomous weapons be controlled, so that it does not put the lives ofour own men at stake.

Nevertheless, the superiority of trained and crafted army personnel over AI enabled systems and weapons cannot be neglected on the basis of their ethics, judgment abilities, multi-sided approaches and situation analyzing capabilities.

V. CONCLUSION

As the history of technology shows that it is very difficult to stop the march of technology once it is invented. Once a technology is invented, everyone rushes out to seek its benefits, underlying the fact that it brings some perils along the way too. Similar is the case with artificial intelligence. It leads us to a new world of unimagined and unexplored possibilities, where human efforts get reduced to a great extent and chances of success increase. With the incorporation of AI into the corporate sector b y the creation of intelligent and automated robots like Siri and Cortana and other intelligent platforms assisting humans in their day to day tasks, the defense forces of the nation's worldwide have also realized the need to equip its weapons and systems with artificial intelligence.

The nations are keen to explore this field more and more so as it could strengthen and empower its armies with the best of the technology available. Nations like Israel, US, South Korea, UK have already in use some automated weapons in their armed forces. Developing nations like India are also their full modes to equip their defense services with AI and machine capabilities. The militaries all over the world are looking at LAWS with great interest. They feel that LAWS will increase the operational efficiencies and minimize harm to soldiers. But on the other side, lie serious concerns regarding ethics, moral values and threat to humanity. Surely, fully automated weapons can act as invincible shields against the enemy and protect the lives of our soldiers. But these technologies are also prone to failures and hacking, a terrorist group can manipulate the instructions and processing of these machines which can pose huge vulnerabilities to the in charge troops themselves. Even more worse is the condition where technology goes beyond human control.

The future of military weapons, systems and machinery relies in the hands of artificial intelligence and machine capabilities. The need is to devise strategies and measures to keep a check on how to control the development of these systems and prevent fromproliferation of automated weapons.

REFERENCES

[1] Cummings M.L," Artificial Intelligence and the FutureofWarfare" ,International Security Department and USand the Americas Programme January 2017 ,p no: 2-7

[2] Dr Rickli Jean-Marc, " T he Economic, Security and Military Implications of Artificial Intelligence for the ArabGulfCountries" p no: 1-2, 6-8.

[3] Hanlon' Michael," T he role of AI in future warfare"

[4] NagpalKavita, "Robotics And artificial Intelligence in Defense Sector"

[5] GuptaArvin, " India should take up challenge of Lethal Autonomous Weapons Systems"

[6] Scharre Paul, " Army of None: Autonomous Weapons and the Future of War" ,New York: W. W. Norton, 2018 , p no: 4-20

[7] Button Robert," A Survey of Missions for Unmanned Undersea Vehicles" ,Santa Monica, Calif.: RAND, p no:3-15

[8] Mizokami Kyle, " T he U.S. Navy Is Getting a More Lethal T orpedo," Popular Mechanics, December 27, 2016 , p no:3-37

[9] Jacques Bughin, " Notes from the AI Frontier: Modelling the Impact of AI on the World Economy" ,McKinsey GlobalInstitute, September 2018, p no:2-8

[10] Ilves Luukas, " Briefing note: Responsible, Safe and Secure AI",T he Lisbon Council, p no: 10-12

[11] Singh T ejaswi , Gulhane Amit," Military applications for artificial intelligence in 2018"

[12] Gonzales Daniel, Harting Sarah," Designing Unmanned Systems with Greater Autonomy Using a Federated, PartiallyOpenSystems Architecture Approach",National Defence ResearchInstituteno:3-10

[13] Group Captain Atul Pant," Aerial Drone Swarms: T he Next Generation Military Weapon"

[14] T yugu Enn, " Artificial Intelligence in Cyber Defense"pno:95-97

A Self-tuning System for Big Data Analytics

Arshpreet Kaur
Associate Professor, Computer science engineering
Chandigarh University
M ohali, India
Arshpreet.cse@cumail.in

Amritpal singh
Student BE, Computer science engineering
Chandigarh University
M ohali, India
Sandhu.lmp1996@gamil.com

***Abstract*-Opportune and financially savvy investigation over "Enormous Data" is currently a key element for achieveme nt in numerous organizations, logical and building controls, and government tries. The Hadoop programming stack—which comprises of an extensible Map Reduce execution motor, pluggable appropriated stockpiling motors, and a scope of procedural to definitive interfaces—is a mainstream decision for enormous information examination. Most specialists of huge information investigation—like computational researchers, frameworks scientists, and business examiners—do not have the mastery to tune the framework to get great execution. Lamentably, Hadoop's execution out of the crate fails to impress anyone, prompting problematic utilization of assets, time, and cash.**

I. INTRODUCTION

Opportune and financially savvy examination over "Huge Data" has risen as a key element for accomplishment in numerous organizations, logical and designing orders, and government attempts. Web crawlers and informal organizations catch and examine each client activity on their destinations to enhance webpage configuration, spam and extortion discovery, and publicizing openings. Great telescopes in stargazing, genome sequencers in science, and molecule quickening agents in material science are putting monstrous measures of information under the control of researchers. Key logical leaps forward are relied upon to originate from computational examination of such information. e.g., computational science, computational financial matters. Cohen et al. as of late authored the acronym and Depth—to express the highlights that clients anticipate from a framework for enormous information examination.

Magnetism**:** An attractive framework pulls in all wellsprings of information regardless of issues like conceivable nearness of anomalies, obscure diagram or nonattendance of structure, and missing qualities that keep numerous valuable information sources out of traditional information distribution centers.

Agility**:** A lithe framework adjusts in a state of harmony with fast information advancement.

Depth**:** A profound framework bolsters investigation needs that go a long ways.

Hadoop MAD framework that getting to be well known for enormous information examination. A whole biological community of apparatuses is being produced around Hadoop. Hadoop itself has two essential segments: a Map Reduce execution motor and an appropriated document framework. While the Hadoop Distributed File System is utilized transcendently dispersed document framework in Hadoop, other record frameworks like Amazon S3 are likewise bolstered. Examination with Hadoop includes stacking information as records into the disseminated document framework, and after those running parallel Map Reduce calculations on the information. Getting wanted execution from a MAD framework can be a nontrivial work out. The professionals of enormous information investigation like information examiners, computational researchers.

• *Information mistiness until handling :* The attraction and deftness that accompanies translating information just at handling time represents the trouble that even the outline might be obscure until the point when the moment that an investigation work must be kept running on the information.

• *File-based processing***:** Information for a Map Reduce occupation might be put away as couple of vast records, a great many little documents, or anything in the middle. Such uncontrolled information formats are a checked complexity to the precisely arranged designs in database frameworks.

• *Overwhelming utilization of programming dialects:* A sizable part of Map Reduce projects will keep on being composed in programming dialects like Java for execution reasons, or in dialects like Python or R that a client is most alright with while prototyping new investigation errands

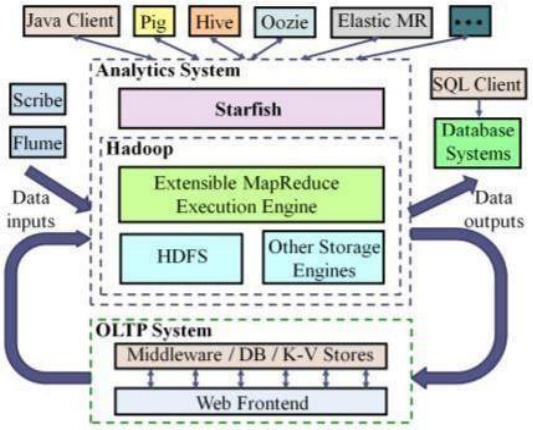


Figure 1: Starfish in the Hadoop ecosystem

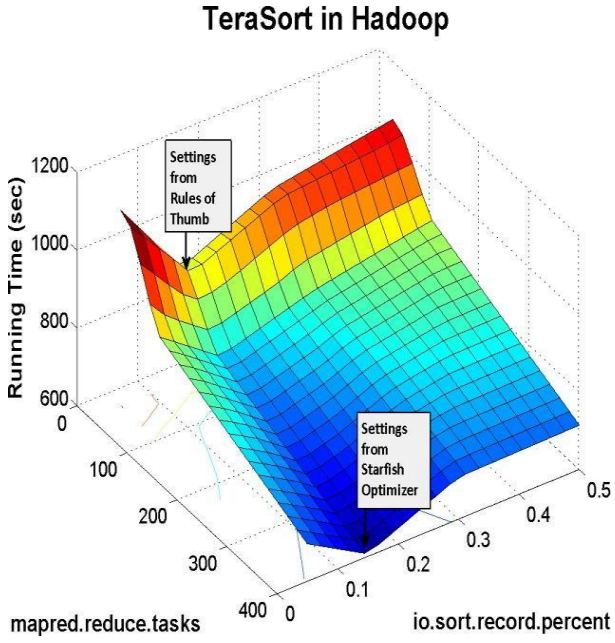
A. Self-Tuning Hadoop and MADDER

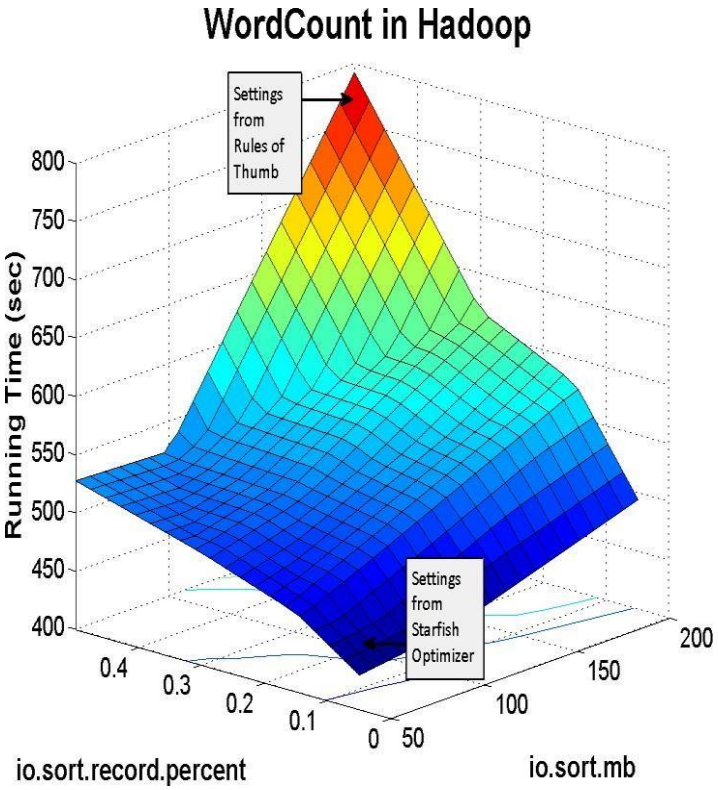
Hadoop has the center instruments to be MADDER than existing examination frameworks. In any case, the utilization of the greater part of these instruments must be overseen physically. Accept flexibility for instance. Hadoop underpins dynamic hub expansion and in addition decommissioning of fizzled or surplus hubs. Notwithstanding, these instruments don't mystically make Hadoop flexible as a result of the absence modules to choose (a) when and how to rebalance the

information design in this procedure. Starfish is a self-tuning and starfish framework for examination on huge information. An imperative plan choice we made is to manufacture Self tuning on the Hadoop stack as appeared in Figure 1. Hadoop as watched prior has valuable natives to help meet the new prerequisites of enormous information examination. Furthermore, Hadoop's selection by scholastic, government, and modern associations is developing at a quick pace. Various continuous undertakings mean to enhance Hadoop's pinnacle execution, particularly to coordinate the inquiry execution of parallel database frameworks [1, 7, 10]. Starfish has an alternate objective. The pinnacle execution a physically tuned framework can accomplish isn't our essential concern, particularly if this execution is for one of the numerous stages in the information lifecycle. Consistent clients may once in a while observe execution near this pinnacle. Starfish will probably empower Hadoop clients and applications to get great execution naturally all through the information lifecycle in examination; with no need on their part to comprehend and control the many tuning handles accessible. Area 2 gives an outline of Starfish while Sections 3– 5 portray its parts. The essential focal point of this paper is on utilizing trial results to delineate the difficulties in every segment and to persuade Starfish's answer approach.

B. Overview Starfish

The capacity to perform opportune and adaptable investigative handling of huge datasets is currently a basic element for the accomplishment of generally associations. The Hadoop programming stack is a well known decision for enormous information examination. Tragically, Hadoop's capacity and adaptability comes at a mind-boggling expense since it depends intensely on the client or framework head to improve Map Reduce occupations and the general framework at different levels. For example:





Starfish finds good settings automatically for Map Reduce jobs

- A number of tuning parameters must be set even to run a solitary occupation in Hadoop. These parameters incorporate the quantity of mappers and reducers, memory designation settings, controls for I/O and system utilization, and others. A significant number of these parameters majorly affect work execution as our examination appears.Operations like joins, declarative queries, and large workflows are specified in Hadoop using higher-level languages like HiveQL and Pig Latin. These language compilers do not support cost -based optimization to convert the declarative specifications to Map Reduce jobs; relying instead on hard -coded rules and user-specified hints.
- A all around tuned information format is regularly the way to great execution in parallel information handling since it is basic to move the calculation to the information. The information design has different measurements, for example, information square situation, dividing, replication, pressure, and decision of capacity and recovery motor. Hadoop has little help for tuning the information format because of remaining burden and asset changes.
- The Starfish venture is tending to these difficulties utilizing a blend of methods from cost-based database inquiry enhancement, strong and versatile question preparing, static and dynamic program examination, dynamic information inspecting and run -time profiling, and measurable machine learning connected to surges of

framework instrumentation information. Starfish expands on Hadoop while adjusting to client needs and framework remaining tasks at hand to give great execution naturally, with no requirement for clients to comprehend and control the many tuning handles in Hadoop. The curiosity in Starfish's methodology originates from how it centers at the same time around various remaining task at hand granularities—generally speaking outstanding burden, work processes, and employments and crosswise over different choice focuses —provisioning, enhancement, booking, and information format.

C. Employment level Tuning

The conduct of a Map Reduce work in Hadoop is controlled by the settings of in excess of 190 setup parameters. On the off chance that the client does not determine parameter settings amid work accommodation, at that point default esteems —dispatched with the framework or indicated by the framework chairman—are utilized. Great settings for these parameters rely upon occupation, information, and group attributes.. The framework presently needs to rapidly pick the join execution method—given the constrained data accessible up until this point, and from among 10+ different ways to execute participates in Starfish—and additionally the comparing settings of employment arrangement parameters. Starfish's Just-in-Time Optimizer tends to one of a kind streamlining issues like those above to consequently choose productive execution strategies for Map Reduce employments. "Without a moment to spare" catches the online idea of choices constrained on the analyzer by Hadoop MADDER highlights. The enhancer takes the assistance. The Profiler utilizes a system called dynamic instrumentation to learn execution models, called work profiles, for unmodified Map Reduce programs written in dialects like Java and Python. The Sampler gathers insights effectively about the information, moderate, and yield key-esteem spaces of a Map Reduce work. An exceptional component of the Sampler is that it can test the execution of a Map Reduce work so as to empower th e Profiler to gather estimated work profiles at a small amount of the full employment execution cost.

D. Work process level Tuning

Work process execution draws out some basic and unexpected communications between the Map Reduce undertaking scheduler and the basic appropriated file system. Noteworthy execution gain acknowledged parallel errand planning by moving calculation to information. Suggestion, the information design crosswise over hubs in the group compels how undertakings can be planned in an "information neighborhood" form. Disseminated file systems have their own strategies on how information kept in touch with them is spread out. HDFS, for instance, dependably composes the principal imitation of any square on a similar hub where the author (for this situation, a guide or diminish errand) runs. This connection between information neighborhood booking and the disseminated file system's square situation arrangements can prompt a lopsided information format crosswise over hubs in the group amid work pro cess execution; causing serious execution corruption as we will appear in Section 4. Productive booking of a Hadoop work process is additionally confused by concerns like (an) abstaining from falling reexecution under hub disappointment or information debasement [11], (b) guaranteeing power corresponding figuring, and (c) adjusting to awkwardness in load or cost of vitality crosswise over geographic districts and time at the datacenter level [16]. Starfish's Workflow-mindful Scheduler tends to such worries related to the What-if Engine and the Data Manager. This scheduler speaks with, yet works outside, Hadoop's inner assignment scheduler.

ACKNOWLEDGM ENT

I am extremely appreciative to Mr. Rohit Bajaj who supports me, and help to making this report.

REFERENCES

[1] Abouzeid, K. Bajda-Pawlikowski, D. J. Abadi, A. Rasin, and A. Silberschatz. HadoopDB: An Architectural Hybrid of MapReduce and DBMS Technologies for Analytical Workloads. PVLDB, 2(1), 2009.

[2] BTrace: A Dynamic Instrumentation Tool for Java. http://kenai.com/projects/btrace.

[3] M. J. Cafarella and C. R´e. Manimal: Relational Optimization for Data-Intensive Programs. In WebDB, 2010.

[4] M. Cantrill, M. W. Shapiro, and A. H. Leventhal. Dynamic Instrumentation of Production Systems. In USENIX Annual T echnical Conference, 2004.

[5] Cloudera: 7 tips for Improving MapReduce Performance. http://www.cloudera.com/blog/2009/12/ 7-tips-for-improving-mapreduce-performance.

[6] J. Cohen, B. Dolan, M. Dunlap, J. M. Hellerstein, and C. Welton. MAD Skills: New Analysis Practices for Big Data. PVLDB, 2(2), 2009.

[7] J. Dittrich, J.-A. Quian´e-Ruiz, A. Jindal, Y. Kargin, V. Setty, and J. Schad. Hadoop++: Making a Yellow Elephant Run Like a Cheetah. PVLDB, 3(1), 2010.

[8] P. K. Gunda, L. Ravindranath, C. A. Thekkath, Y. Yu, and L. Zhuang. Nectar: Automatic Management

Impact of Inheritance Metrics on Fault Prediction using Clustering

Deepak Kalia
Departm ent of Com puter Science and Engineering
Baba Banda Singh Bahadur Engineering College
Fatehgarh Sahib, Punjab

Puneet Mittal
Departm ent of Com puter Science and Engineering
Baba Banda Singh Bahadur Engineering College
Fatehgarh Sahib, Punjab

***Abstract*—maintaining the quality of software system is one of the major tasks of software industry. For improving quality, more focus should be done on te sting for those parts of code which have large st numbe r of faults. Faults must be che cke d at the e arly stage s of the product de ve lopme nt as it save s time and cost. Inhe ritance is one of the powe rful fe ature s of obje ct orie nted programming. Inhe ritance he lps in re usability of code but it should be use d in a prope r way so that the syste m doe sn't be come comple x. Various inhe ritance me trics are give n to che ck the fault prone ne ss and quality of a software syste m. In this pape r, we analyz e d various inhe ritance me trics and the impact of the se me trics on error occurrence in classe s is discusse d.**

Keywords- CK Metrics, Inheritance metrics, fault proneness, fault prediction, error.

I. INTRODUCTION

The Software industry is paying more attention to peremptory Error in any software system is very common and complex problem. Preventing a software system from errors is such a difficult task. Faults must be checked at the early stages of the product development as it saves time and cost. It also helps in reducing complexity at early stages, which also affects other quality attributes. If complexity is less it implies that the effort required to test a programwould be less and product would be more reliable. The Software industry is paying more attention to peremptory Error in any software system is very common and complex problem. Preventing a software system from errors is such a difficult task. Software metrics are the quantitative measurement of the complexity of the software, so they are good candidates for guiding the selection of testing techniques. In the world of object-oriented languages, software metrics has been used for many years to provide developers with additional information about their software quality. Software metrics can monitor the quality of software. Preventing a software system from errors is such a difficult task. Quality of software is increasingly important for software. For improving quality we should provide more focus on testing for those portions of code which have largest number of faults.

The object oriented paradigm provides strong support for software reuse. Inheritance helps in reusability as well as it affects other factors like complexity, testability etc.

A. Inheritance Metrics

Inheritance promotes reusability but it should be used in a proper range so that the project doesn't become complex. The inheritance metrics are as:

Depth of Inheritance Tree (DIT): DIT is the maximum length from node to the root of the tree. It is a class level metric. If more the values of DIT then more the Number of Methods (NOM) it is likely to inherit, making more complex to predict its behavior, the greater the potential reuse of inherited methods. Small values of DIT in most of the system's classes may be an indicator that designers are forsaking reusability for simplicity of understanding.

Number of Children (NOC): NOC is defined as the number of immediate subclasses subordinated to a class in the class hierarchy. If the greater is the NOC then the greater is the reuse, the greater is the probability of improper abstraction of the parent class, the greater the requirements of methods testing in that class. Small values of NOC, may be and indicator of lack of communication between different class designers.

Weighted Methods per Class (WMC): This is a weighted sum of all the methods defined in a class. A class with a low WMC usually points to greater polymorphism. A class with a high WMC indicates that the class is complex (application specific) and therefore harder to reuse and maintain.

B. Clustering techniques and algorithms [15]

Clustering algorithms can be broadly classified into three categories, in the following subsections together with specific algorithms:

- Partitioning
- Hierarchical
- Density-Based
- Meta-Heuristic OptimizationTechnique

In short, partitioning algorithms attempt to determine k clusters that optimize a certain, often distance-based criterion function. Hierarchical algorithms create a hierarchical decomposition of the database that can be presented as a dendrogram. Density - based algorithms search for dense regions in the data space that are separated from one another by low density noise regions. Meta-heuristic is a higher-level procedure or heuristic designed to find, generate, or select a heuristic (partial search algorithm) that may provide a sufficiently good solution to an optimization problem.

Partitioning Clustering Algorithms

Partitioning clustering attempts to decompose a set of N objects into k clusters such that the partitions optimize a certain criterion function. Each cluster is represented by the center of gravity (or centroid) of the cluster, e.g. k-means, or by the closest instance to the gravity center (or medoid), e.g. k-medoids. Typically, k seeds are randomly selected and then a relocation scheme iteratively reassigns points between clusters to optimize the clustering criterion. A serious drawback of partitioning algorithms is that there are a number of possible solutions. It includes the following:

- K-Means
- Fuzzy k-means clustering K-Medoids

Hierarchical Algorithms

Unlike partitioning methods that create a single partition, hierarchical algorithms [11] produce a nested sequence of clusters, with a single all-inclusive cluster at the top and singleton clusters of individual points at the bottom. The hierarchy can be formed in top-down (divisive) or bottom-up (agglomerative) fashion and need not necessarily be extended to the extremes. The merging or splitting stops once the desired number of clusters has been formed. Typically, each iteration involves merging or splitting a pair of clusters based on a certain criterion, often measuring the proximity between clusters. Hierarchical techniques suffer from the fact that previously taken steps (merge or split), possibly erroneous, are irreversible.

Density-BasedClustering Algorithms

Density-based clustering methods group neighbouring objects into clusters based on local density conditions rather than proximity between objects [14]. These methods regard clus ters as dense regions being separated by low density noisy regions. Density-based methods have noise tolerance, and can discover non-convex clusters. Similar to hierarchical and partitioning methods, density-based techniques encounter difficulties in high dimensional spaces because of the inherent scarcity of the feature space, which in turn, reduces any clustering tendency. Density-Based Spatial Clustering of Applications with Noise (DBSCAN) is an example of density based clustering algorithms.

Meta-Heuristic Optimization Technique

The SFLA is a meta-heuristic optimization technique. The concept of the SFLA is based on observing, imitating, and modelling the social behaviour of a group of frogs when they search for the location of a rich source of food. Several engineering optimization problems have been solved by the SFLA.

II. LITERATURE SURVEY

The literature survey is carried out from the designing of CK metrics to explore different techniques used for fault prediction. CK metric suit is most widely used metrics for the object- oriented (OO) software. Chidamber et al. [2] developed and implemented a new set of software metrics for Object Oriented designs. They defined sixmetrics -Weighted Methods per Class (WMC), Depth of Inheritance (DIT), Coupling between Objects (CBO), Response for a Class (RFC), Lack of Cohesion in Methods (LCOM), Number of Children (NOC). The metrics given by Chidamber and Kemerer are one of the best-known OO metrics. CK metrics explore different techniques used for the modeling of fault prediction. Basili et al. [3] investigated object-oriented (OO) design metrics introduced by Chidamber and Kemerer. They define hypotheses for each metric that represented the expected connection between the metrics and the fault-proneness of the code. They tested these hypotheses and found that some of the metrics were very good predictors. From their results, five out of the six CK OO metrics appear to be useful to predict class fault-proneness. R.Subramanyam [4] validated the WMC, CBO, and DIT metrics as predictors of the error counts in a class. Their results indicated that the CK metrics could predict error counts. Subramanyam and Krishnan chose a large e-commerce application developed in C++ and Java. They examined the effect of the size along with the WMC, CBO, and DIT values on the faults by using multivariate regression analysis. They compared the applicability of the metrics in different languages; thus, they validated their hypotheses for C++ and Java classes separately. They concluded that WMC and CBO could be validated only for C++. R.Bender [5] proposed a method for quantitative risk assessment and investigated about threshold value. These values are used to predict about the probabilities risk level. By defining acceptable levels for the absolute risk and the risk gradient the corresponding benchmark values of the risk factor can be designed. TiborGyimo et. al [6] illustrated fault-proneness detection of the source code of the open source Web. For fault

proneness detection they used regression and machine learning methods to validate the usefulness of these metrics for fault-proneness prediction. They found CBO and LOC metric seems to be the best in predicting the fault-proneness of classes but DIT metric is untrustworthy and NOC cannot be used at all for fault-proneness prediction. The LOC metric performed fairly well and because it can be easily calculated, it seems to be suitable for quick fault prediction. C.Catal et .al [9] proposed a fully automated technique which does not require an expert during the prediction process. They used X-means clustering with software metric threshold. Experiments revealed that unsupervised software fault prediction can be fully automated and effective results can be produced. Raed Shatnawi et. al [1] design threshold values of software metrics. They used area under curve to find out threshold values for each metrics. Raed Shatnawi [7] introduced methodology to produced threshold values with better classification accuracy. Threshold values provide a meaningful interpretation for metrics and provide a surrogate to identify classes at risk. The classes that exceed a threshold value can be selected for more testing to improve their internal quality, which increases the testing efficiency. Periklis Andritsos et al [11] explained cluster analysis has been widely used in several disciplines such as statistics, software engineering, biology, psychology and other social sciences, in order to identify natural groups in large amounts of data. These data sets are constantly becoming larger, and their dimensionality prevents easy analysis and validation of the results. Tian Zhang et al[12] presented a data clustering method named BIRCH (Balanced Iterative Reducing and Clustering using Hierarchies), and demonstrates that it is especially suitable for very large databases. BIRCH incrementally and dynamically clusters incoming multi-dimensional metric data points to try to produce the best quality clustering with the available resources (i. e., available memory and time constraints). BIRCH can typically find a good clustering with a single scan of the data, and improve the quality further with a few additional scans. Anna Huang [13] described clustering is a useful technique that organizes a large quantity of unordered text documents into a small number of meaningful and coherent clusters, thereby providing a basis for intuitive and informative navigation and browsing mechanisms. Partitioned clustering algorithms have been recognized to be more suitable as opposed to the hierarchical clustering schemes for processing large datasets. A wide variety of distance functions and similarity measures have been used for clustering, such as squared Euclidean distance, cosine similarity, and relative entropy. Yixin Chen et al [14] explained existing data-stream clustering algorithms such as CluStreamare based on k-means. These clustering algorithms are incompetent to find clusters of arbitrary shapes and cannot handle outliers. Further, they require the knowledge of k and user-specified time window. To address these issues, it proposes D-Stream, a framework for clustering stream data using a density-based approach. The algorithm uses an online component which maps each input data record into a grid and an offline component which computes the grid density and clusters the grids based on the density. The algorithmadopts a density decaying technique to capture the dynamic changes of a data stream. Exploiting the intricate relationships between the decay factor, data density and cluster structure, it can efficiently and effectively generate and adjust the clusters in real time. Deepti Sisodia et al [15] represented partitioning a set of objects into homogeneous clusters is a fundamental operation in data mining. The operation is needed in a number of data mining tasks. Clustering or data grouping is the key technique of the data mining. It is an unsupervised learning task where one seeks to identify a finite set of categories termed clusters to describe the data. The grouping of data into clusters is based on the principle of maximizing the intra class similarity and minimizing the inter class similarity. The goal of clustering is to determine the intrinsic grouping in a set of unlabeled data.

III. DATA COLLECTION

In this study, firstly CK Object Oriented metrics presented by Chidamber and Kemerer [2] were studied and the working of CK metrics was analyzed. As the work is on the analysis of impact of Inheritance metrics on fault prediction using clustering, then the working of inheritance metrics (i.e., NOC, WMC and DIT) and various clustering techniques were analyzed. Tool which is used to collect database for implementing this study is ckjm tool, which helps us to collect inheritance metrics values of software code of jhotdraw.

After collecting and preparing the database, error collection is done on the basis of bad smells. Bad smells are calculated using together tool. These give the values of error count for every class of jhotdraw software. Calculated values of error count are categorized into two categories i.e., either „0' or „1' for every class. 0' indicates the class is not erroneous and „1' indicates the class is erroneous.

TABLE 1. DESCRIPTION OF JHOTDRAW SOFTWARE

Software	No Of Classes	Faulty Classes	Non faulty Classes
Jhotdraw	239	44	195

After collecting and preparing the database, the implementation is done in matlab using Shuffled Frog Leap Algorithm [10].

IV. Methodology Of Work

We applied Shuffled Frog Leap algorithm [10] on the inheritance metrics calculated for jhotdraw software. The classes are divided for training and testing. Training is done on 180 classes and testing is applied on 59 classes.

TABLE 2. CLASSIFICATION OF CLASSES OF JHOTDRAW

Software	No of classes	Classes for Training	Classes for Testing
Jhotdraw	239	180	59

Shuffled Frog Leap Algorithm

The SFLA is a meta-heuristic optimization technique. The concept of the SFLA is based on observing, imitating, and modeling the social behavior of a group of frogs when they search for the location of a rich source of food. Several engineering optimization problems have been solved by the SFLA.

Work ing Algorithm for SFLA

1. The first is the formation of the initial population, followed by frog design groups: According to F frogs (solutions), randomly generated frog populations.
2. Each frog for a specific calculates the objective function value.
3. Descending order according to the objective function F frogs into S sub-groups according to.
4. For each sub-group of frogs are found in one of the best individual and the worst individual to identify the best individual groups.
5. For each subgroup, in descending order according to the objective function value of the individual, re-distribution and mixing operations.
6. Termination conditions are met.
7. The end of iteration.
8. The optimal objective function value of the output information, or turn to the original sequence.

V. Results and Conclusion

After applying SFLA on all the metrics i.e., NOC, WMC and DIT calculated for 59 classes of jhotdraw, confusion matrices are drawn for following cases in testing phase-

Case 1: Confusion Matrix between Actual error count and computed error count by NOC

Case 2: Confusion Matrix between Actual error count and computed error count by WMC

Case 3: Confusion Matrix between Actual error count and computed error count by DIT

0	43	2
1	11	3
Truth values	0	1

Figure 1. Confusion Matrix between Actual error count and computed error count by NOC

0	45	0
1	14	0
Truth values	0	1

Figure 2. Confusion Matrix between Actual error count and computed error count by WMC

0	41	4
1	11	3
Truth values	0	1

Figure 3. Confusion Matrix between Actual error count and computed error count by DIT

From above results, a truth table is shown which represents number of true and false values of all the metrics after comparing them with actual error count.

T ABLE 3. T RUTH VALUES OF NOC, WMC AND DIT

Metrics	True value	False Values
NOC	46	13
WMC	45	14
DIT	44	15

The percentage of true values is more in NOC than other two metrics which indicates it is more reliable in predicting errors than the other two metrices.

REFERENCES

[1] Raed Shatnawi1, Wei Li, James Swain," Finding software metrics threshold values using ROC curves" , JOURNAL OF SOFTWARE MAINTENANCE AND EVOLUTION: RESEARCH AND PRACTICE, Evol.: Res. Pract. 2010; 22:1–16.

[2] Shyam R. Chidamber and Chris F. Kemerer, "A Metrics Suite for Object Oriented Design," IEEE TRANSACTIONS ON SOFTWARE ENGINEERING, VOL. 20, NO. 6, JUNE 1994.

[3] V.R. Basili, L.C. Briand, and W.L. Melo, " A Validation of Object-Oriented Design Metrics as Quality Indicators," IEEE Trans.Software Eng., vol. 22, no. 10, pp. 751-761, Oct. 1996.

[4] Ramanath Subramanyam and M.S. Krishnan," Empirical Analysis of CK Metrics for Object-Oriented Design Complexity: Implications for Software Defects" IEEE TRANSACTIONS ON SOFTWARE ENGINEERING, VOL. 29, NO. 4, APRIL 2003.

[5] R. Bender, " Quantitative Risk Assessment in Epidemiological Studies Investigating Threshold Effects," Biometrical J., vol. 41,no. 3, pp. 305-319, 1999.

[6] Tibor Gyimo´ thy, Rudolf Ferenc, and Istva´n Siket," Empirical Validation of Object-Oriented Metricson Open Source Software for Fault Prediction", IEEE TRANSACTIONS ON SOFTWARE ENGINEERING, VOL. 31, NO. 10, OCTOBER 2005.

[7] Raed Shatnawi," A Quantitative Investigation of the Acceptable Risk Levels of Object-Oriented Metrics in Open-Source Systems" IEEE TRANSACTIONS ON SOFTWARE ENGINEERING, VOL. 36, NO. 2, MARCH/APRIL 2010.

[8] Sonia Chawla and Dr. Rajender Nath, " Evaluating Inheritance and Coupling Metrics" International Journal of Engineering Trends and Technology (IJETT) ,VOL. 4 Issue 7- July 2013.

[9] C. Catal, U. Sevim, and B. Diri, Member, IAENG"Software Fault Prediction of Unlabeled Program Modules" Proceedings of the World Congress on Engineering 2009 Vol I WCE 2009, July 1 - 3, 2009,London, U.K.

[10] Hany M. Hasanien, " Shuffled Frog Leaping Algorithm for Photovoltaic Model Identification", IEEE, vol 6, Issue 2, 99-509-512, April 2015

[11] Periklis Andritsos," Data Clustering Techniques", Pp 1-34, 2002

[12] Tian Zhang," BIRCH: An Efficient Data Clustering Method for Very Large Databases", ACM, vol 25(2), Pp 103-114, 1996

[13] Anna Huang," Similarity Measures for Text Document Clustering", NZCSRSC, Pp 49-56, 2008

[14] Yixin Chen et al," Density-Based Clustering for Real-Time Stream Data",Research track, Pp 133-142, 2007

[15] Deepti Sisodia," Clustering Techniques: A Brief Survey of Different Clustering Algorithms", IJLTET, Vol 1(3), Pp 82-87 , 2012

IMPACT OF COUPLING AND COHESION METRICS ON FAULT PREDICTION USING CLUSTERING

Ashmeet Kaur
Departm ent of Com puter Science and Engineering
Baba Banda Singh Bahadur Engineering College
Fatehgarh Sahib, Punjab

Puneet M ittal
Departm ent of Com puter Science and Engineering
Baba Banda Singh Bahadur Engineering College
Fatehgarh Sahib, Punjab

***Abstract*- S oftware me trics he lps in e stimating the faults. All faults pre diction te chnique s ge t a he lp in this study wi th the designing of Cluste rs. The focus of the study is to design andpredict faults for next pie ce of software. As the comple xity andthe constraints unde rwhich software is de velope d are incre asing, it is usual to produce software with se vere faults. Various coupling and cohe sion me trics are give n to che ck the fault prone ne ss and quality of a software syste m and the se me trics e xamine s the cluste ring algorithm for software fault pre diction using CK me trics. He re, fault is conside re d as de pendent variable and C K me trics suite as inde pe nde nt variable s. For this purpose , diffe re nt pre diction mode ls have be e n de velope d using cluste ring.**

Keywords- Error, Fault Prediction, Metrics reliability, CK metrics

I. INTRODUCTION

Object oriented metrics were proposed to evaluate the quality of software such as the fault-proneness, reliability and maintainability of software. In the world of object -oriented languages, software metrics has been used for many years to provide developers with additional information about their software quality. One cost-effective way to deal with this problem is apply fault occurrence based on object-oriented (OO) design metrics to detect software faults in classes in developed systems before delivery. Software development managers seeking to produce high-quality software on time and within budget and indeed the last three decades has seen proposed many software fault occurrences [13], [11], [6], [9], [7]. Software faults however may vary considerably in their severity, yet, to date no software faults have been proposed that are able to distinguish among faults. A software fault prediction is a proven technique in achievin g high software reliability. This study is an attempt to predict fault in software by applying different techniques. The accuracy of the fault-occurrence predictions of sixwidely used OO design metrics, especially a subset of the Chidamber and Kemerer (CK) metrics suite.

A. CK Metrics

In this section, we define the metrics that we extracted from the source code and used for occurrence. Six of these metrics were first presented by Chidamber and Kemerer. The discussion of Chidamber and Kemerer metrics are-

- Depth of Inheritance Tree (DIT)
- Number of children (NOC)
- Weighted Methods per Class (WMC)
- Coupling between objects (CBO)
- Response for a Class (RFC)
- Lack of Cohesion in Methods (LCOM)

B. Coupling Metrics

Coupling metrics are identified to compute complexity, reliability, testability and reusability. Coupling in a system increases complexity, reduces reusability. Coupling is a measure of interconnecting among modules in a software structure.

- *Coupling between objects (CBO)* - CBO for a class is a count of number of classes to which it is coupled. It is a class level metric and the couplings due to inheritance are also computed in CBO. CBO for a class is a count of the number of non-inheritance related couples with other classes. Two things are coupled if and only if at least one of themacts upon the other. Any evidence of a method of one object using methods or instance variables of another object constitutes coupling.

The viewpoints are: If small values of CBO then improve modularity and promote encapsulation, indicates independence in the class making easier its reuse, makes easierto maintain and to test a class.

- *Response for a Class (RFC)* -The response set of a class is defined as set of methods that can be potentially executed in response to a message received by an object of that class.

The viewpoints are: If a large numbers of methods are invoked from a class (RFC is high) then testing and maintenance of the class become more complex.

$$RFC = |RS|$$

Where, RS is the response set for the class. Response set of an object ≡ {set of all methods that can be invoked in response to a message to the object}

C. Cohesion Metrics

A cohesive module performs a single task within a software procedure, using little interaction with procedures being performed in other parts of a program. Cohesion refers to the internal consistency within the parts of the design.

- *Lack of cohesion of methods (LCOM)* - The LCOM metric is the number of pairs of methods in the class using no attributes in common (referred to as P), minus the number of pairs of methods that do (referred to as Q). The LCOMis set to zero if this difference is negative. After considering each pair of methods:

$$LCOM = (P>Q) ? (P-Q) : 0$$

The LCOM metric measures the coherence among methods in a class. The viewpoints are: If a greater value ofLCOM then increases complexity does not promotes encapsulation and implies classes should probably be split into two or more sub classes and helps to identified low quality design. The difference between Coupling and Cohesion is a distinction o n a process of change, not a static analysis of code's quality.

D. Clustering techniques and algorithms[3]

Clustering algorithms can be broadly classified into three categories, in the following subsections together with specific algorithms:

1) Partitioning
2) Hierarchical
3) Density-Based
4) Meta-Heuristic Optimization Technique

In short, partitioning algorithms attempt to determine k clusters that optimize a certain, often distance-based criterion function. Hierarchical algorithms create a hierarchical decomposition of the database that can be presented as a dendrogram. Density-based algorithms search for dense regions in the data space that are separated from one another by low density noise regions. Meta-heuristic is a higher-level procedure or heuristic designed to find, generate, or select a heuristic (partial search algorithm) that may provide a sufficiently good solution to an optimization problem.

1) *Partitioning Clustering Algorithms*

Partitioning clustering attempts to decompose a set of N objects into k clusters such that the partitions optimize a certain criterion function. Each cluster is represented by the centre of gravity (or centroid) of the cluster, e.g. k-means, or by the closest instance to the gravity centre (or medoid), e.g. k-medoids. Typically, k seeds are randomly selected *and* then a relocation scheme iteratively reassigns points between clusters to optimize the clustering criterion. A serious drawback of partitioning algorithms is that there are a number of possible solutions. It includes the following:

- K-Means
- Fuzzy k-means clustering K-Medoids

2) *Hierarchical Algorithms*

Unlike partitioning methods that create a single partition, hierarchical algorithms [8] produce a nested sequence of clusters,

with a single all-inclusive cluster at the top and singleton clusters of individual points at the bottom. The hierarchy can be formed in top-down (divisive) or bottom-up (agglomerative) fashion and need not necessarily be extended to the extremes. The merging or splitting stops once the desired number of clusters h as been formed. Typically, each iteration involves merging or splitting a pair of clusters based on a certain criterion, often measuring the proximity between clusters. Hierarchical techniques suffer from the fact that previously taken steps (merge or split), possibly erroneous, are irreversible.

3) Density-Based Clustering Algorithms

Density-based clustering methods group neighbouring objects into clusters based on local density conditions rather than proximity between objects [14]. These methods regard clusters as dense regions being separated by low density noisy regions. Density-based methods have noise tolerance, and can discover non-convex clusters. Similar to hierarchical and partitioning methods, density-based techniques encounter difficulties in high dimensional spaces because of the inherent scarcity of the feature space, which in turn, reduces any clustering tendency. Density - Based Spatial Clustering of Applications with Noise (DBSCAN) is an example of density based clustering algorithms.

4) Meta-Heuristic Optimization Technique

The SFLA is a meta-heuristic optimization technique. The concept of the SFLA is based on observing, imitating, and modelling the social behaviour of a group of frogs when they search for the location of a rich source of food. Severalengineering optimization problems have been solved by the SFLA.

II. LITERATURE REVIEW

Many researchers have carried out significant work in the area of fault prediction. The literature survey is carried out from the designing of CK metrics to explore different techniques used for the modeling of fault prediction. CK metric suit is most widely used metrics for the object- oriented (OO) software. Software Metrics are the key element for fault occurrence.

Anna Huang (2008) [1] explained Clustering is a useful technique that organizes a large quantity of unordered text documents into a small number of meaningful and coherent clusters, thereby providing a basis for intuitive and informative navigation and browsing mechanisms. Partitioned clustering algorithms have been recognized to be more suitable as opposed to the hierarchical clustering schemes for processing large datasets.

Deepti Sisodia et al. (2012) [3] proposed partitioning a set of objects into homogeneous clusters is a fundamental operation in data mining. The operation is needed in a number of data mining tasks. Clustering or data grouping is the key technique of the data mining. It is an unsupervised learning task where one seeks to identify a finite set of categories termed clusters to describe the data. The grouping of data into clusters is based on the principle of maximizing the intra class similarity and minimizing the inter class similarity. The goal of clustering is to determine the intrinsic grouping in a set of unlabeled data.

C. Catal et al. [2] proposed a fully automated technique which does not require an expert during the prediction process. They used X-means clustering with software metric threshold. Experiments revealed that unsupervised software fault prediction can be fully automated and effective results can be produced.

Shyam R. Chidamber et al. (1994) [12] developed and implemented a new set of software metrics for Object Oriented designs. They noticed that noted that Object Oriented may hold some of the solutions to the software crisis. The stud y is aimed at examining the relationships between these metrics and cost, quality, and productivity. The results show that high values of CBO and LCOM are associated with lower productivity, greater rework and greater design effort.

M.H. Tang et al. (1999) [7] analyzed C K OO metrics suite on three industrial applications developed in C++. They found none of the metrics examined to be significant except RFC and WMC and investigated the dependency between CK metric suite and the object-oriented system faults.

K. El Emam et al. (2001) [6] analyzed a large C++ telecommunication application and found that the size (i.e., SLOC) of classes confounded the effect of most OO design metrics on faults. In their experiment, WMC, RFC, CBO, and LCOM were found to be significant without size control but none of these metrics was significant after controlling for the size of the system. NOC was not investigated.

P. Yu et al. (2002) [9] examined the correlations of 10 metrics with fault proneness, where CBO and RFC were divided into two different types. They used the same LCOM metric as our study. Empirical study has been carried out in the past of fault proneness models. Though this study have raised the need to study fault proneness models at different severity levels of faults, yet less work has been done in this important area.

Periklis Andritsos et al. (2002) [8] introduced cluster analysis which has been widely used in several disciplines, such as statistics, software engineering, biology, psychology and other social sciences, in order to identify natural groups in large amounts of data. These are the data sets.

To further examine K. El Emam et al.'s findings, **R. Subramanyam et al. (2003) [11]** validated the WMC, CBO, and DIT metrics as predictors of the error counts in a class. Their results indicated that the CK metrics could predict error counts. Besides validating the usefulness of metrics, they compared the applicability of the metrics in different languages; thus, they validated their hypotheses for C++ and Java classes separately. To study the effect of the size along with the WMC, CBO and DIT on fault-proneness of classes. Their results indicated, that even after controlling for the size (i.e. SLOC) of classes, some of the CK metrics were significantly associated with faults.

T. Gyimo´thy et al. (2005) [13] illustrated fault-proneness detection of the source code of the open source Web. For fault proneness detection they used regression and machine learning methods to validate the usefulness of these metrics for fault-proneness prediction. From the review of literature, we found that earlier fault relationship analysis have only considered OO design metric and most of the studied systems were implemented in C++.

Yixin Chen et al. (2007) [14] illustrated that the existing data-stream clustering algorithms such as CluStream are based on k-means. These clustering algorithms are incompetent to find clusters of arbitrary shapes and cannot handle outliers. Further, they require the knowledge of k and user-specified time window. To address these issues, this paper proposes D-Stream, a framework for clustering stream data using a density -based approach. The algorithm uses an online component which maps each input data record into a grid and an offline co mponent which computes the grid density and clusters the grids based on the density. The algorithm adopts a density decaying technique to capture the dynamic changes of a data stream. The experimental results show that the algorithm has superior quality and efficiency, can find clusters of arbitrary shapes, and can accurately recognize the evolving behaviours of real-time data streams.

Raed Shatnawi (2010) [10] introduced methodology to produced threshold values with better classification accuracy. Threshold values provide a meaningful interpretation for metrics and provide a surrogate to identify classes at risk. The classes that exceed a threshold value can be selected for more testing to improve their internal quality, which increases the testing efficiency.

III. DATA COLLECTION

All of the six metrics were first presented by Chidamber and Kemerer [8]. In the study the metrics CBO, W MC, RFC, NOC, LCOM, DIT were investigated. Tool which is used to collect database for implementing this study is CKJM tool, which help us to collect all CK metrics value and find out the Error Count (errors). As the work is on the analysis of impact of coupling and cohesion metrics on fault prediction using clustering.

After collecting and preparing the database, error collection is done on the basis of bad smells. Bad smells are calculated using together tool. These give the values of error count for every class of jhotdraw software. Calculated values of error count are categorized into two categories i.e., either '0' or '1' for every class. 0' indicates the class is not erroneous and '1' indicates the class is erroneous.

Software	No Of Classes	Faulty Classes	Non faulty Classes
Jhotdraw	239	44	195

Table 1. Description of jhotdraw software

After collecting and preparing the database, the implementation is done in matlab using Shuffled Frog Leap Algorithm[4].

a) Dependent and independent variables:

The goal of this study is to establish the relationship between Object-Oriented metrics and fault proneness at the class level. In the study, a fault is used as a dependent variable and each of the CK metric is an independent variable. Fault is a function of WMC, NOC, DIT, RFC, CBO and LCOM and can be represented as shown in the following equation:

Faults = f (WMC, NOC, DIT, CBO, RFC, LCOM)

IV. METHODOLOGY OF WORK

We applied Shuffled Frog Leap algorithm [10] on the inheritance metrics calculated for jhotdraw software. The classes are divided for training and testing. Training is done on 180 classes and testing is applied on 59 classes.

Software	No. of classes	Classes for Training	Classes for Testing
Jhotdraw	239	180	59

Table 2. Classification of classes of jhotdraw

a) Shuffled Frog Leap Algorithm

The SFLA is a meta-heuristic optimization technique. The concept of the SFLA is based on observin g, imitating, and modeling the social behavior of a group of frogs when they search for the location of a rich source of food. Several engineering optimization problems have been solved by the SFLA.

b) Work ing Algorithm for SFLA

1) The first is the formation of the initial population, followed by frog design groups: According to F frogs (solutions), randomly generated frog populations.
2) Each frog for a specific calculates the objective function value.
3) Descending order according to the objective function F frogs into S sub-groups according to.
4) For each sub-group of frogs are found in one of the best individual and the worst individual to identify the best individual groups.
5) For each subgroup, in descending order according to the objective function value of the individu al, re-distribution and mixing operations.
6) Termination conditions are met.
7) The end of iteration.
8) The optimal objective function value of the Output information, or turn to the originalsequence.

V. RESULTS ANDCONCLUSION

After applying SFLA on all the metrics i.e., NOC, WMC and DIT calculated for 59 classes of jhotdraw, confusion matrices are drawn for following cases in testing phase-

Case 1: Confusion Matrixbetween Actual error count and computed error count by NOC

Case 2: Confusion Matrixbetween Actual error count and computed error count by WMC

Case 3: Confusion Matrixbetween Actual error count and computed error count by DIT

0	7	0
1	38	14
TRUTH VALUES	0	1

Figure 1. Confusion Matrix between Actual error count and computed error count by CBO

0	8	1
1	37	13
TRUTH VALUES	0	1

Figure 2. Confusion Matrix between Actual error count and computed error count by WMC

0	45	14
1	0	0
TRUTH VALUES	0	1

Figure 3. Confusion Matrix between Actual error count and computed error count by DIT

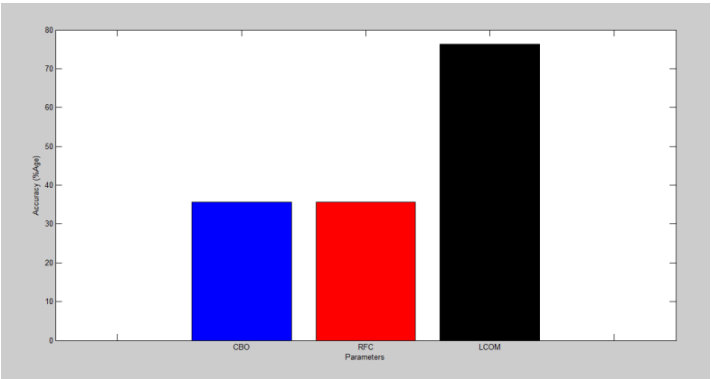


Figure 4. Accuracy (in %age) by CBO, RFC and LCOM

From the above results, it is concluded that the error count computed using shuffled flog leap algorithm for CBO and RFC gives less accuracy and error count computed for LCOM is more accurate than traditional approaches.

Also Bar Graph indicates that LCOM is more reliable than other two metrics (i.e., CBO and RFC) in finding errors in any open source software.

REFERENCES

[1] Anna Huang," Similarity Measures for T ext Document Clustering", NZCSRSC, Pp 49-56, 2008.

[2] C. Catal, U. Sevim, and B. Diri, Member, IAENG" Software Fault Prediction of Unlabeled Program Modules"

[3] Deepti Sisodia," Clustering T echniques: A Brief Survey of Different Clustering Algorithms", IJLT ET , vol. 1(3), Pp 82 -87, 2012.

[4] Hany M. Hasanien, " Shuffled Frog Leaping Algorithm for Photovoltaic Model Identification", IEEE, vol 6, Issue 2, 99-509-512, April 2015.

[5] Hitz, M. and Montazeri, B., Measuring Coupling and Cohesion in Object Oriented systems, International Symposium on Applied Corporate computing, Monterey, Mexico, 25-27, 1995.

[6] K. El Emam, S.Benlarbi, N. Goel, and S.N. Rai, "T he Confounding Effect of Class Size on the Validity of Object -Oriented Metrics," IEEE T rans. Software Eng., vol. 27, no. 7, pp. 630 -650, July 2001.

[7] M.H. T ang, M.H. Kao, and M.H. Chen, " An Empirical Study on Object - Oriented Metrics," Proc. Sixth Int'l Software Metrics Symp., pp.242-249, 1999.

[8] Periklis Andritsos," Data Clustering T echniques", Pp 1 -34, 2002.

[9] P. Yu, T . Systa, and H. Muller, " Predicting Fault -Proneness Using OO Metrics: An Industrial Case Study," Proc. Sixth European Conf. Software Maintenance and Reeng., pp. 99-107, 2002.

[10] Raed Shatnawi," A Quantitative Investigation of the AcceptableRiskLevels of Object -Oriented Metrics in Open Source Systems" IEEE T RANSACT IONSON SOFT WARE ENGINEERING, VOL. 36, NO. 2, MARCH/APRIL 2010.

[11] R. Subramanyan and M.S. Krisnan, " Empirical Analysis of CK Metricsfor Object -Oriented Design Complexity: Implications for Software Defects," IEEE T rans. Software Eng., vol. 29, no. 4, pp 297 -310, Apr. 2003.

[12] Shyam R. Chidamber andChris F. Kemerer, " A Metrics Suite for Object Oriented Design," IEEE T RANSACT IONS ON SOFT WARE ENGINEERING, VOL. 20, NO. 6, JUNE 1994.

[13] T . Gyimo´thy, R. Ference, and L. Siket, " Empirical Validation of Object-Oriented Metrics on Open Source Software for Fault Prediction," IEEE T rans. Software Eng., vol. 31, no. 10, pp. 897-910, Oct. 2005.

[14] Yixin Chen et al," Density-Based Clustering for Real-T ime Stream Data", Research track, Pp 133-142, 2007.

[15] M. M. Eusuff, K. E. Lansey, and P. Fayzul, " Shuffled frog-leaping algorithm: a memetic meta-heuristic for discrete optimization," Engineering Optimization, Vol. 38, No. 2, pp. 129 –154, March 2006.

Digital Image Watermarking Techniques: A Comprehensive Survey

Harneel Kaur Cheema
Department of Computer Science and Engineering
Baba Banda Singh Bahadur Engineering College
Fatehgarh Sahib, India
harneelcheema@gmail. com

***Abstract* - Due to expanded utilization of web and interactive media applications Image watermarking is accomplishing greater prominence. Picture watermarking connects valuable data about the host (unique) picture as logo/sound/video or content. The essential point of watermarking is to apportion information trustworthiness, copyright insurance, proprietorship recognizable proof, and substance verification. Alongside giving content validation and information trustworthiness watermarking additionally keeps the content from being adjusted. The essential prerequisites of watermarking are security, limit, indistinctness, strength that varies as indicated by various applications. Methods which are utilized in spatial domain can install more bits as they are simple and straightforward, have less multifaceted nature however they are not impervious to a few sorts of geometric ass aults. Then again, strategies which are utilized in the frequency domain are impervious to both geometric and picture preparing assaults however their vigor diminishes when the recurrence is debased because of which quality gets defiled and inserting increasingly number of bits turns into a troublesome undertaking. As these systems degenerate the nature of the picture so these can't insert various bits. So as to beat this issue these systems must be utilized with some different procedures of the spatial domain with high limit.**

Keywords- Robust, Semi-fragile, Fragile Watermarking, Digital image watermarking

I. INTRODUCTION

Due to the snappy advancement of sight and sound and web applications, copyright assurance and substance security turned into a prime concern. For settling the issue of duplicating, adulteration, and contortions, the legitimacy of substance is particularly fundamental. There are numerous strategies that are utilized for concealing the information like cryptography, steganography. These procedures are distinct from encryption and mark calculations in an accompanying way. A Digital mark is scrambled hash value that can recognize the alter in a picture yet can't distinguish the area of alteration. So also, in encryption, clients can access and utilize the information yet they can't have possession distinguishing proof. So as a productive strategy it is utilized to give copyright assurance and proprietorship distinguishing proof [1].

II. GENERAL PROCESS

The process of watermarking system produ ces additional data about the original picture. It is essentially 4-advance procedure that incorporates: First is generation of watermark, second is embedding watermark, thirdly distribution and then attacks on image and similarly extraction of watermark.

a) Generation: Generation of watermark step creates a logo in the type of sound/video/message that is novel to the substance and must be with the end goal that extraction or mutilation from various assaults is troublesome.
b) Embedding: It is a procedure that implants a logo picture into the host picture.
c) Distribution and Attacks: It is a procedure which can be viewed as watermarked information transmission. And assault is the event when somebody endeavors to adjust the substance.
d) Extraction: It is the procedure which gives data to the proposed beneficiaries and also enables the proprietor to get distinguished. Extraction is indistinguishable procedure from embedding yet happens in an invert way. Distinctive procedures are utilized for embedding of watermark and watermark extraction.

2.1 Basic Requirements

A. Robustness: The system ought to be sufficiently ready to deal with different assaults conceivable on pictures during the process of dissemination.

B. Imperceptibility: Watermark which is implanted in the picture ought not to be obvious to the eyes of human.
C. Authentication: The information substance should only be open for lawful clients following legitimate validation. All this is done by utilizing cryptography.
D. Capacity: The measure of information payload, in the form of watermark, that can be covered up in the picture is called as capacity.
E. Region of Interest (ROI) Intactness **:** The ROI must not be influenced by a watermarking procedure. Harmed ROI cannot have genuine results regarding patient's wellbeing.
F. Reversible**:** The picture ought to be corrigible in the wake of watermark extraction.
G. Complexity**:** Less mind-boggling calculation is constantly beneficial regarding client and also framework by sparing execution time.

2.2 Sorts of watermarking dependent on various ideas
Digital watermarking methods are characterized by records types as:

- Video Watermarking
- Audio Watermarking
- Image Watermarking
- Text Watermarking

In light of the human recognition they are delegated as:

- Visible Watermarking[2]
- Invisible Watermarking[2]
  - Robust Watermark- Robustness watermarking is a sort of watermarking whose principle objective is to sign copyright data, the installed watermark can hold the regular alter handling, picture preparing, and lossy pressure, and regardless of whether some assault happens, the watermark isn't wrecked. It opposes numerous sorts of assaults, geometrical or non-geometrical without influencing the implanted watermark.
  - Fragile Watermark- The primary motivation behind fragile watermarking is giving uprightness, which ought to be especially sensitive to any of the adjustment in a signal. It can be watched that the information has been changed or not.
  - Semi-fragile watermark- It can deal with just some level of the change to a watermarked picture, for example, the incorporation of clamor because of lossy compression.

2.3 Categorization of attacks

- Noise attacks: Their objective is to render the watermark by attaching mutilation to the original picture. The assault is fruitful if the recognizing watermark becomes difficult any longer yet the picture is as yet clear and can be utilized for a specific decided reason.
- Disabling detection attacks: The primary target of this assault is to interrupt the connection within the original and that of extricated watermark arrangement.
- Inversion Attacks: In inversion attacks the primary target is to confound by creating counterfeit unique information or phony watermark information.
- Removal attacks: The assaults which endeavor to evacuate watermarks, like nonlinear sifting, are known as removal attacks.

III. WATERMARKING TECHNIQUES BASED ON SPATIAL DOMAIN

Spatial domain systems install the watermark by the changing the pixels values [3]. It additionally utilizes a hash function and secret key for inserting [4]. It very well may be connected utilizing shading division too. Different spatial domain systems are as per the following:

A. Method based on least significant bits
This is the most, for the most part, used methodology for inserting and extraction in the spatial domain. Firstly, the picture is partitioned into a subset of pictures. Then the encoder first chooses the pictures subset and after that chooses the bits to be supplanted. It replaces Least significant bits (LSB) of original picture with Most significant bits (MSB) of the watermark picture. The span of the original and watermark picture must be the equivalent. Installing is done in pixels LSB's in light of the way that the modification in pixels LSB's can't be successfully recognized by eyes of a human. The permeability of watermark in the original picture depends upon the quantity of bits that are supplanted [2].

B. Repetition codes and hamming codes based Digital image watermarking [5]
In this strategy, firstly the watermark is installed utilizing distinctive size redundancy codes and hamming codes. The encoding of principal watermark picture is done by utilizing (7, 4) Hamming code. Firstly, the watermark picture is examined row-wise and isolated into different blocks. At that point, the encoded watermark is inserted in the least significant bits of the original picture. Implanting utilizing redundancy codes fills in as

pursues: Firstly, each piece is encoded utilizing (3, 1) reiteration code. At that point, LSB's of three back to back pixels of the main picture is loaded up with corresponding watermark.

C. Spatial domain digital image watermarking based on block probability concept [6]
In this technique, a binary picture is utilized as a watermark. Then it is mixed by utilizing sequence numbers. At that point installing of the encoded watermark is completed multiple times at various positions by a secret key. Watermark is installed by adjusting intensities of non-covering 8*8 blocks of the blue segments of the host picture. Extraction is the equivalent yet works in a turn around the way.

IV. WATERMARKING TECHNIQUES BASED ON FREQUENCY TRANSFORM DOMAIN

This system applies the distinctive changes on a picture and inserts the watermark into coefficients by adjusting the values of coefficients. Therefore methods in this space install the watermark as opposed to in powers values. Procedures in frequency transform domain are as perthe following:

A. Image watermarking based on Discrete Cosine Transformation (DCT) [7]
It changes over a picture from the spatial domain to transform domain and the other way around. At the point when a picture is changed utilizing DCT, it separates given picture into 8*8 blocks. At that point, it discovers low and high-recurrence parts by crisscross examining. And after that implants the watermark in low-recurrence parts. This technique gives high vigor against JPEG compression. DCT strategies need protection from solid geometric assaults.

B. Image watermarking based on Discrete Wavelet Transformation (DWT) [8]
It is a deterioration strategy that decays given picture into a lot of fundamental wavelets. It gives spatial and changes the portrayal of a picture. DWT is a reasonable system to distinguish the area in the picture that contains the secret picture. DWT break down given picture into low and high-frequency segments and discovers high-frequency segments and implants a picture into high-frequency parts. In DWT technique frequency goals depends with respect to frequency so when the frequency is tainted it diminishes power.

C. Hybrid watermarking bas ed on Discrete Wavelet Transform-Discrete Cosine Transform (DWT-DCT) [9]
In this strategy, the hybrid watermarking joins DCT and DWT. In this strategy, DWT is connected to the original picture up to various dimensions pursued by DCT and after that applies diverse kinds of assaults. As the quantity of level builds the measure of watermark diminishes and PSNR increments [9]. In this technique, check picture is increased with the deviation of host picture so quality corrupts gradually. This strategy gives high PSNR and can remove high caliber and extensive imprints. It doesn't change the perspective of the host picture. This strategy fulfills the prerequisites of heartiness.

V. ANALYSIS OF DIFFERENT DOMAINSOF WATERMARKING TECHNIQUES

From the overview of various strategies of the spatial domain, it is reasoned that the systems in the spatial domain are straightforward, has low multifaceted nature and can be utilized for the main verification. They are less exorbitant can insert the quantity of bits so limit is solid. On the off chance that LSB systems are joined with different methods, just if demonstrates great execution. Be that as it may, these techniques are less impervious to various sorts of assaults.
From the review of various systems of the frequency area, it has been presumed that methods in this space are increasingly mind-boggling, includes vast overhead, progressively expensive. In frequency area systems, more data bits can't be inserted in the coefficients as it debases picture quality. Systems in this space are utilized for copyright assurance and give solid vigor against various kinds of assaults which is the fundamental point of any watermarking method. On the off chance that twists are made into a picture it will cause a difficult issue. It has been presumed that to limit twists and to expand limit, systems in this area must be joined with another method which has a high limit and solid heartiness against various sorts of assaults.

TABLE I. COMPARISON OF WATERMARKING TECHNIQUES

Methods	Advantages	Disadvantages
LSB substitution Method	Simple, Low complexity	Less efficient, Low PSNR, modification can be done easily
Color image watermarking using repetition codes & hamming codes	Resistant to Salt & Pepper noise	Less perceptual quality

Color image watermarking in spatial domain based on block probability method	Simple, More secure with use of keys	Less resistant to JPEG compression with high values of Quantization factor
DCT based Watermarking	High robustness to JPEG compression attack	Large overhead ,Not resistant to geometric attacks
DWT based Watermarking	Strong robustness	Frequency changes decrease robustness
DCT -DWT based hybrid method	Strong robustness, Good imperceptibility	Increased Complexity of Method, Extraction Is Not Good Enough

VI. CONCLUSION

Watermark inserting and extraction calculations are required for giving copyright security and possession ID. Here, a complete overview of different digital picture watermarking procedures in various areas and their necessities. In this work, diverse procedures have been characterized by their prerequisites, advantages, and restrictions. It has been inferred that to limit mutilations and to expand limit, systems in the frequency domain must be joined with another method which has a high limit and solid strength against various sorts of assaults.

REFERENCES

[1] F. Cayre, C. Fontaine and T. Furon, " Watermarking security: theory and practice", IEEE Transactions on Signal Processing, Vol. 53, No. 10, pp. 3976–3987, Oct. 2005.

[2] D. Chopra, P. Gupta, Gaur Sanjay B.C and A. Gupta, " Lsb Based Digital Image Watermarking For Gray Scale Image", IOSR Journal of Computer Engineering, Vol. 6, No. 1, pp. 36-41, Sep-Oct. 2012.

[3] B. Surekha, Dr G.N. Swamy, "A Spatial Domain Public Image Watermarking", International Journal of Security and Its Applications, Vol. 5, No. 1, Jan. 2011.

[4] R. K. Megalingam, M. M. Nair, R. Srikumar and V. K.. Balasubramanian, " Performance Comparison of Novel, Robust Spatial Domain Digital Image Watermarking with the Conventional Frequency Domain Watermarking Techniques," Proc. IEEE International Conference on Signal Acquisition and Processing, pp. 349 – 353, 2010.

[5] S. Rohith, Dr. K. N. Bhat, " A Simple Robust Digital Image Watermarking against Salt and Pepper Noise using Repetition Codes," Proc. of Int. Joint Colloquium on Emerging Technologies in Computer Electrical and Mechanical, 2011, ACEEE,DOI: 02.CEM.2011.01.531.

[6] Nasir, Y. Weng, J. Jiang, " A New Robust Watermarking Scheme for Color Image in Spatial Domain," Third International IEEE Conference on Signal-Image Technologies and Internet-Based System, DOI 10.1109/SITIS.2007.67.

[7] S. D. Lin and C. F. Chen, " A Robust DCT -Based Watermarking for Copyright Protection", IEEE Transactions on Consumer Electronics, Vol. 46, No. 3, pp. 415-421, August 2000.

[8] S. M. Mohidul Islam, Rameswar Debnath, S. K Alamgir Hossain, " DWT Based Digital Watermarking Technique and its Robustness on Image Rotation, Scaling, JPEG compression, Cropping and Multiple Watermarking," IEEE Conference on ICICT, ISBN: 984-32-3394-8, pp. 7-9, March 2007.

[9] Akter, Nur-E-Tajnina, and M. A. Ullah, " Digital Image Watermarking Based on DWT -DCT: Evaluate for a New Embedding Algorithm", IEEE 3rd International conference on Informatics, Electrronics & Vision, pp. 1-6, 2014.

Wireless Sensor Networks And Its Applications: A Review

Sandeep Kaur Saini
PG Research Scholar
Baba Banda Singh Bahadur Engineering College
Fatehgarh Sahib, Punjab, India

***Abstract*— in real-time monitoring of a physical process, the developing field of wireless sensor network (WS N) has potential advantages. The wireless sensors persistently examine the noticeable phenomenon and carry information to the base station. It is an information technology which coordinates latest technological achievements in the network, micro-electronics and communications. S ensor nodes can interact over the smallest distance through wireless channel and combined to complete a single task. The paper presents a review of Wireless S ensor Networks**.

Keywords-Wireless Sensor Networks, Applications, Challenges

I. INTRODUCTION

Wireless sensor networks (WSN) has emerged as an upcoming technology for a large number of applications with different needs. However, before considering it a mature technology, there are few shortcomings which are needed to overcome. Energy constraint is the key obstacle in WSN by which sensor nodes undergo, where main source of the power supply are batteries. Therefore, there is a requirement to develop energy-efficient solutions which can increase the network lifetime. The objective is to present an overview of wireless sensor networks.

The rest of the paper is organized as follows, Section II contain the applications where wireless sensors play vital role, Section III contain the routing in WSN, Section IV describe the technical challenges in WSN, Section V explain the storage management in WSN with balance of subsequent goals and Section VI concludes the review with future directions.

II. APPLICATIONSOF WSN

Sensor network is generally established for remote monitoring and data acquisition purposes. Some applications where wireless sensors plays vital role are discussed below.

A. Environmental applications

Wireless sensor networks are appropriate for agricultural and leeway supervising applications seeing as wired arrangement would be costly and ineffective. A diversity of functions has been created in ecological monitoring, domestic animal scrutinizing and accuracy agriculture.

B. Health careapplications

Wireless networks have acknowledged major consideration from the research society when used in healthcare systems. The two medical care directed systems are namely, remote healthcare surveillance and vital status monitoring. Vital status monitoring is used to identify emergency situations like to administer essential parameters patients wear sensors which allow guardians to respond effectively [1]. Remote healthcare monitoring is related to the care services that are not essential and for which the patient does not require constant presence of a caregiver.

C. Agricultural applications

In accuracy agriculture, to examine significant parameters, for instance soil moisture, temperature, the humidity of leaves and hours of sunlight, sensor nodes are deployed all through a ground generating a system for analysis and decision making. An additional feature of accuracy agriculture is disease expansion calculation, resource (pesticides, fertilizer and water) optimization and frost protection.

D. Structural monitoring

This application is used to examine movement inside building, condition of building, flyovers, construction and bridges. The bridges, buildings, bridges and other constructions with the help of WSN can present their condition statistics to the administration and as per their need they can repair these buildings. That's why these structures are known as intelligent buildings.

E. Intelligent home monitoring

The self-regulating living environment gives more convenience and comfort to people. WSN is installed to manage wide range of services spontaneously and work collectively. Smart home environment is collaboration of services and technology by utilizing home networking for security, communication, safety, comfort, automation and energy savings. For example, sensor nodes are installed to interpret utility meter at homes like water, electricity, gas and then transmit readings at distant centres.

F. Military applications

To predict and handle unpredictable actions, like man -made troubles or natural disasters, wireless sensor networks are used. The systems whose nodes are involved with agent s for impermanent arrangement is referred to as active intervention and is devoted to the security of team-centred actions. Every component consists of a sensor while working so that isolated manager is capable of monitoring both the controller's position and ecological features. It is applied to miners, disaster rescue teams and armed forces.
The motionless sensors are arranged in huge areas under passive supervision like a social infrastructure or nuclear spot for durable surveillance. Significant instances of applications of passive supervision are fire detection in building, emergency navigation, surveillance and target tracking, natural calamity avoidance, for instance, during eruptions, flooding or tsunamis and structural health monitoring.

G. Industrial applications

Many utility companies have the aim of mechanization of surveillance and control systems in oil & gas refining, manufacturing, electrical power division and water treatment.

The combination of wireless network with Smart Grids and Supervisory Control & Data Acquisition (SCADA) systems is considered. With an intelligent and automated power-system administration, the monitoring of energy supply and consumption process is possible by Smart Grids. In smart grids, the possible applications of sensor networks are: control of turbines, remote recognition of defective components; underground cables and motors; residence energy management and sensing the important parameters disturbing power production (such as humidity, temperature, radiation, wind direction).

H. Vehicle detection

Detection and tracking of vehicle has become an important application in WSN. Advanced Vehicle Location system is comprised of two GPS systems, one is built-in GPS satellite receiver that is fundamentally used to compute precis ely the position of vehicle and other one is the reliable GSM network to convey the position directions to the control centre [3].

I. Congestion control

Decreasing the traffic jam on roads is a major challenge for metropolitan authority. The system is required based on sensor network which can sense the jamming on the road and the congestion information is transmitted to the drivers so that they can choose alternative route for avoiding congestion [4][5][6]. Commercial on-board applications in addition to pass enger-safety applications are being planned by commodity suppliers. They include support in finding a parking space, route supervision to stay away from smart high-speed tolling, rush-hour jams and vehicle journey information gathering.

J. Underwater and Underground sensor networks

The forthcoming forms of WSNs are underground and underwater sensor networks, which can be utilized in various applications containing ecological monitoring, community security and organization. They are different from conventional terrestrial networks in which the sensors are installed in particular surroundings which build complicated interactions, thus affect their deployment simplicity. In underground networks, there are hidden sensors and they interact with each other with the help of thick resources such as concrete or soil. These networks are utilized for reporting soil moisture in farms, intrusion detection, transportation systems and infrastructure management. The underwater networks depend on absorbed sensors and are used in different applications like disaster prevention, ocean administration, water quality monitoring, pipeline monitoring and surveillance.

III. ROUTING IN WSN

A sensing node has limited computation and sensing capacities, power and communication capability. Th ese nodes can communicate with one another for obtaining sensed data using either single-hop or multi-hop communication. Hence each node in the sensing network operates as a router within the network. In single hop communication, the data is captured by sensor nodes from environment in which they are deployed. The sensory data is transmitted from sensor nodes to the base station directly without any involvement of intermediate nodes. The topology of the sensing network varies oftenly. Since in case of direct communication the distance between the sensor nodes and base station is large, the nodes consume energy rapidly and become inoperative. In multi-hop communication, the data is captured by deployed sensor nodes from environment and transmit data to their specific cluster head nodes. Through intermediate nodes, sensed data is transmitted from nodes to the BS. The cluster nodes are the intermediate nodes also called individual nodes. A routing protocol is used which allows scattered information to pick out routes between any two nodes on the network, the selection of the route being done by routing algorithms. As the possible routes may change, dynamic routing allow routing tables in routers. In case of wireless sensor networks, dynamic routing is generally used because nodes might modify their location and die at any moment. The comparison of different routing protocols in WSN is shown below in Table 1.

Table 1. Comparison of different routing protocols in WSN

Routing Protocols	**Power usage**	**Classification**	**S calability**	**Position awareness**	**Mobility**	**Data Aggregation**	**QoS**	**Query based**
Flooding	High	Flat	Limited	No	No	No	No	No
Gossiping	High	Flat	Limited	No	No	No	No	No
Direct Diffusion	Limited	Data centric/ flat	Limited	No	Limited	Yes	No	Yes
LEACH	M aximum	Hierarchical	Good	No	Fixed BS	Yes	No	No
PEGASIS	M aximum	Hierarchical	Good	No	Fixed BS	No	No	No
TEEN & APTEEN	M aximum	Hierarchical	Good	No	Fixed BS	Yes	No	No

IV. CHALLENGES IN WIRELESS SENSOR NETWORKS

In WSN, the foremost practical challenges for realization are identified as follows [8, 9, and 10].

A. Resource constraint: The three scarce resources in WSN which requires planning while implementation are [11]:

a) Restricted energy
b) Limited memory
c) Limited computational capability.

B. Dynamic topologies and harsh environmental conditions: In wireless environment, the topology of the network could vary due to link and sensor-node failures. Moreover, sensors may additionally be affected from elevated humidness levels, sensations, dirt and mud, or alternative conditions that can degrade the performance [10].

C. Quality-of-service needs: A large variety of applications in WSNs have entirely different QoS necessities and particulars. The QoS given by wireless network refers to the exactness between the information reached at the base station and the actual information detected by sensors.

D. Pack et errors and variable-link capacity: Compared to wired networks, capability of every wireless link depends on

the interference level and high bit error rates in transmission [10].

E. Security: Security is one of the necessary features within the architecture of WSNs to make the communication secure from intrusion and denial-of-service attacks. WSNs have unique features that can alter methods of security attacks. Passive attacks are carried out by eavesdropping on broadcasting information. Active attacks encompass fabrication, modification, and interruption that could grasp flooding, node capturing or routing attacks.

F. Integration with web and alternative network s: It is necessary for the improvement of sensing network to provide several services which give access to retrieve any information at any time and thus it may be integrated with the internet for information retrieval [11].

V. STORAGE MANAGEMENT IN WSN

The main area of sensor network analysis is storage management that starts to attract awareness of researchers. For management of storage, data gathered by sensor nodes is not conveyed to the base station in limited time period. In such applications, the information should be kept among the network and later collected by the base station. Therefore, additional to energy, storage has also become a primary resource that determines time period and coverage of the network. The appliance features are discussed with trends that motivate the requirement for storage management and numerous resource constraints that can manipulate the planning of storage management techniques in sensor networks [12]. Additionally, storage management is divided into various components:

1. System support for storage management;
2. Cooperative storage; and
3. Assortment.

A storage management approach should be the balance of subsequent goals:

- Minimize range of stored information: Since sensor nodes have restricted storage, minimizing the dimensions of information that is required to be kept ends up in improved coverage because the network will maintain storage of data for extended period of time and become additional reasonable if the information is less [13].
- Minimize consumption of energy: As the sensor nodes are battery-powered and so energy is considered to be a scarce resource, therefore storage management must be energy efficient.
- Maximize information retention or coverage: Aggregated information is also the prime goal of the network. Information re-allocation should be applied efficiently if storage is affected to ensure coverage for new data. The management protocol ought to plan to retain relevant information at an acceptable quality level.

VI. CONCLUSION

The paper presented an overview of Wireless Sensor Networks. The review described the routing and storage management of wireless sensor networks. Applications areas and characteristics of WSNs are also discussed. There are still many challenges and restrictions in the sensor networks which are required to be highlighted including inadequate bandwidth and network lifetime of sensor nodes. The basic for this emerging field is enhanced to decide on a particular problem in WSN and work for an efficient outcome.

REFERENCES

[1] S. Prasanna, S. Rao, " *An Overview of Wireless Sensor Networks Applications and Security*", International Journal of Soft Computing and Engineering (IJSCE), Vol.2, Issue.2, pp. , 2012.

[2] S. Zhang, H. Zhang, " *A Review of Wireless Sensor Networks and Its Applications*", In the Proceeding of the IEEE International Conference on Automation and Logistics, Zhengzhou, China, pp.386-389 , 2012.

[3] Nilesh Ananthanarayanan ,"*Intelligent Vehicle Monitoring System using Wireless Com m unication*", In Advances in Technology and Engineering, IEEE Conference, Mumbai, India, 2013.

[4] P. Singhal, A. Yadav, " *Congestion Detection in Wireless Sensor Network using Neural Network*", International Conference for Convergence of T echnology, Pune, India, pp.1-4, 2014.

[5] M. Rahman, N.U. Ahmed, and H.T. Mouftah, "*City Traffic Management Model using Wireless Sensor Networks*". In Electrical and Computer Engineering (CCECE), IEEE 27th Canadian Conference, T oronto, Canada, pp.1-6, 2014.

[6] N.S. Nafi, and J.Y. Khan, " *A VANET based Intelligent Road Traffic Signalling System* ", Australasian T elecommunication Networks and Applications Conference (ATNAC), Brisbane, Australia, 2012.

[7] Z. Belhadi, L. Fergani, B. Fergani, and J.M. Laheurte, " *RFID Tag Indoor Localization by Fingerprinting Methods*", In Wireless Communications, Vehicular T echnology, Information T heory and Aerospace & Elect ronics Systems (VIT AE), 4th International Conference, Aalborg, Denmark, 2014.

[8] H. Chawla, "*Some issues and challenges of Wireless Sensor Networks*", International Journal of Advanced Research in Computer Science and Software Engineering, Vol.4, Issue.7, pp.2 36-239, 2014.

[9] N. Srivastava, " *Challenges of Next-Generation Wireless Sensor Networks and its Im pact on Society*", Journal of T elecommunications, Vol.1, Issue.1, pp.128-133, 2010.

[10] P. Ghune, R.N. Ghune, P. Pandey & P. Mishra, "*Application of Wireless Sensor Networks in Sm art Grid Opportunities, Challenges & Technologies Available (A Survey)*", International Journal on Advanced Computer Theory and Engineering (IJACTE), Vol.2, Issue.5, pp.11-18, 2013.

[11] P. Antoniou, A. Pitsillides, T . Blackwell, A. Engelbrecht and L. Michael, " *Congestion Control in Wireless Sensor Networks based on Bird Flocking Behavior*", Computer Networks, Vol.**57**, Issue.**5**, pp.**1167–1191**, **2013**.

[12] G. Weimin, Z. Lingzhi, " *Distributed Data Storage in Wireless Sensor Networks*", International Journal of Database T heory and Application, Vol.**8**, No.**4**, pp.**179-182**, **2015**.

[13] A.K. Luha, T . Vengattraman, M. Sathya, " *RAHTAP Algorithm for Congestion Control in Wireless Sensor Network*", International Journal of Advanced Research in Computer and Communication Engineering, Vol.**3**, Issue.**4**, pp.**6250-6255**, **2014**.

Information Extraction with the help of Machine Learning

Hema
Assistant professor
Department of Computer science and Engineering
Chandigarh University, Gharuan
Hema.cse@cumail.in

Shourya
Student BE
Department of Computer science and Engineering
Chandigarh University, Gharuan
Cu.16bcs1003@gmail.com

***Abstract*-We consider the issue of figuring out how to perform data extraction in areas where phonetic handling is dangerous, for example, Usenet posts, email, and finger plan documents. Instead of syntactic and semantic data, different wellsprings of data can be utilized, for example, term recurrence, typography, arranging, and increase. We depict learning ways to deal with this issue, each drawn from an alternate worldview: a repetition student, a term-space student in light of S upport Vector Machine, a methodology utilizing syntactic acceptance, and a social govern student. Analyses on data extraction issues characterized more than four various record ac**

Keywords: information extraction,

I. Introduction

Goal of information extraction is machine reading .We have lot of information available in web that is in unstructured format. there is no particular structure of data i.e we cannot obtain information from data in structured format by applying some naïve computer operations .So in information extraction systems we can find and understand limited part of the relevant data and produce a structured information of relevant text in relations or tables in data base form and then this information is organized in meaningful form so that later computer algorithms can draw information from this data For ex we need to buy a pc we will search the web .it will contain all the information about it. We don't need the nitty gritty details about the pc generally we need processor specification this information can be generated in tables and similarly ram information can be generated in another table.

This data generated in tables in generally is raw we need to preprocess it

Problems in Raw Data:

Data can be non atomic in tables i.e two or more values are corresponding to a single attribute this is a very big problem it can lead to a bad trained machine learning algorithm

We can apply normalization operations to segregate non atomic columns into atomic operations.

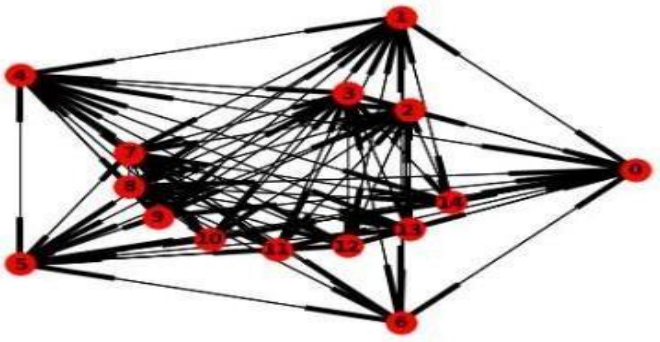
There is also another problem ie missing data how can we overcome it.by filling it with default value or last different value (i.e the most recent value of the attribute in the time series domain).

Now we can further improve our data after cleansing and normalization by aggregating attributes into single attribute by some arithmetic operation .operations can simple or complex as demand of the situation. Now the data is pure free from human bias as much as possible .Now it is being ready to trained for machine learning algorithm

Modeling the structure using graphml

Now since the data is preprocessed .we can represent the data in more meaningful form that is in graphml format.
Relationship between various structural
elements that is tables can be now somewhat predicted

In graph theory, arborescence is a directed graph in which, for a vertex *u* called the root and any other vertex *v*, there is exactly one directed path from *u* to *v*. An arborescence is thus the directed-graph form of a rooted tree, understood here as an undirected graph



For ex consider reviews about a personal computer in web
Pc is cool it has two meaning pc is good .or pc working at the low temperature .Moreover more than one inference can be drawn from these words.
We can give edge to edge relationship p c and cool in graph ml format

Spanning tree based arborescence approach

Data converted to graph ml format can be co nverted into a graph based structure using any conventional programming language. Now weights can be assigned edge edge relationship.

We can get the weights by using svm for each edge where equation of separating plane is hyperbolic Equation for the same is

$$L = \frac{1}{2} w^T w + C \sum_{i=1}^{n} \xi_i + \sum_{i=1}^{n} \alpha_i \left(1 - \xi_i - y_i \left(w^T x_i + b\right)\right) - \sum_{i=1}^{n} \mu_i \xi_i$$

$$\frac{\partial L}{\partial w_{jn}} = w_j - \sum_{i=1}^{n} \alpha_i y_i x_{ij} = 0$$

$$w = \sum_{i=1} \alpha_i y_i x_i = 0$$

Where wis the weights of separating hyperplane and these weights can be used For weight assignment in edges

These weights can also be obtained by using neural network and perceptron model .Initially weights can be assigned random value by applying the dense neural network approach weight could be obtained for each edge.

After some time algorithm will reach to convergence and each weight of edge can be then recorded

We need to apply Edmonds chu liu algorithm for maximu m spanning arborescene. It will generate a spanning tree with maximum weight and now edge to edge conflicts are resolved, Edges which are now remaining in the graph corresponds to the meaningful information.

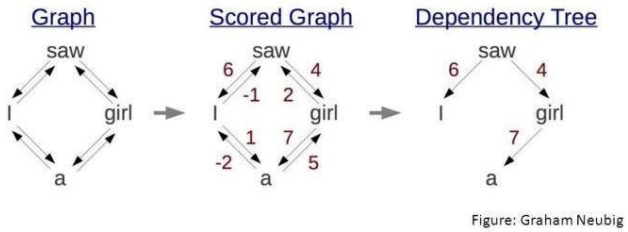


Figure: Graham Neubig

And traversal from root to leaves generate a meaning .

When whole of the data is processed through chu liu Edmonds algorithm whole lot of meaning ful information in generated

Opportunities and further scope:

Further optimize the weights and if possible reduce the running time of machine learning algorithm to assign edge weights and also if a mathematical relation to derive edge weights is predicted by using machine learning .

CONCLUSION:

Data can be extracted from web in unstructed format to tabular form .Tabular form consists of noises and missing values.These noises need to be removed so process of normalizing helps us to do so.Further this tabular form is converted into a common format graphml for better understanding.Edge to edge relationship represents all kinds of possibilities in text.

So we need to gather only the meaningful information.We assign edge weights to each edge using machine learning algorithm and further we pass this weighted graph to Edmonds chu liu algorithm which generates a maximum spanning tree and traversal from root to leaves generates a meaningful information

REFERENCES:

[1] Andrei Alexandrescu, Katrin Kirchhoff

[2] Graph-based Learning for Statistical Machine T ranslation

[3] Human Language T echnologies: The 2009 Annual Conference of the North American Chapter of the ACL, pages 119 –127,

[4] Boulder, Colorado, June 2009. c 2009 Association for Computational Linguistics

[5] [Swabha Swayamdipta and SamT homson] ,"Graph -based Dependency Parsing Chu Liu Edmonds and Camerini(k-best)",Camegie Mellon University 23 Nov 2014

[6] Graph ml Reference

[7] graphml.graphdrawing.org/xmlns/1.1/graphml-structure.xsd

[8] SVM reference

[9] https://en.wikipedia.org/wiki/Support -vector_machine

[10] Neural Network Reference

[11] https://www.doc.ic.ac.uk/~nd/surprise_96/journal/vol4/cs11/report.html

[12] Machine Learning Reference

[13] https://en.wikipedia.org/wiki/Machine_learning

Benefits of P2P technologies

Geetika
Assistant professor
Department Of Computer science and engineering
Chandigarh University, Gharuan
Chandigarh, India
Geetika.cse@cumail.in

AbhinayPathania
Student BE
Department Of Computer science and engineering
Chandigarh University, Gharuan
Chandigarh, India
abhinayp11@gmail.com

***Abstract*-Here we talk about the conditions under which shared (P2P) innovations are prudent in offering IPTV arrangements over framework structures . It has 2 critical parts. To start with, we assess two framework plans used to break down the proficiency of such a framework: a generally utilized sensible "Web as a cloud" outline and a "physical" outline that demonstrates the highlights of the real framework. In particular, we uncover that the thinking configuration investigates basic compositional parts of this technique and may altogether exaggerate the benefits of P2P innovation by a part of at least 3. Second, we offer a money saving advantage examination of P2P motion picture material accommodation concentrating on the cash exchange os for various estimating/impetus outlines as opposed to absolutely on potential expansion. Specifically, we locate that under extraordinary amount of motion picture prerequisite, a P2P worked in inspiration plan per-frames superior to anything some other outline for both top notch and standard-definition press, while the use based plan for the most part produces more benefits when the interest sum is low. The level reward configuration as a rule falls in the middle of the utilization based outline and the implicit plan as far as gainfulness.**

I. INTRODUCTION

Web convention TV guarantees to gatherings of people an inventive arrangement of decisions and authority over their TV material. Two noteworthy U.S. telecom organizations, AT&T and Verizon, have contributed essentially to supplant the birdwatcher lines in their methods with materials optic wires for conveying numerous IPTV stations to private customers.

A watcher can get IPTV video cuts in high caliber if the Authorization to make advanced or hard copies of all or part of this advantage individual or classroom utilize is conceded without expense gave that copies are not made or conveyed to profit or business advantage and that copies bear this notice and the full citation on the main page. To duplicate generally, to distribute, to post on web servers or to redistribute to records, requires earlier particular approval as well as an expense.

Accessible data exchange utilization fulfills the need of motion picture development sum for the objective goals and edge sum. To offer adequate data exchange utilization for IPTV arrangements, Internet specialist co-ops utilize fast xDSL or wire systems to convey film con-tent to watchers' set-top boxes. For instance, AT&T Light-Speed is utilizing Fiber-to-the-Neighborhood (FTTN) Networks. Its structure contains some across the nation to a great degree head -closes (SHE) and an enormous measure of territorial motion picture center point work environments (VHO). The greatly head -closes work as the across the nation material assembling or gathering focuses for communicate and motion picture when required development. A close-by motion picture center point working environments offer assembling or gathering and capacity of local material. Every motion picture center point office fills in as a Video-On-Demand (VOD) library and markets film material through provincial openness changes to the customers. We identify with this framework structure as the "physical" plan all through the record. FTTN procedures can offer 20-25Mbps data exchange utilization to every family unit, which is generally enough to back up a few the plain best TV sources and in addition rapid Online and Voice over IP (VoIP) arrangements.

A major issue in offering IPTV arrangements is its extraordinary execution and upkeep cost. What's more, ability of it web servers can without much of a stretch turn into a bottleneck. One answer for mitigate the weight on web servers is to utilize shared (P2P) methods like Skype or Kontiki. While starting P2P methods were generally utilized for data record introducing, as of late there have been a few endeavors on utilizing the distributed way to deal with back up stay stacking and VOD streaming. Current tests that assess the benefits of P2P film material accommodation as a rule don't think about the confinements of the genuine help offices. Or maybe, they point of view this technique as a "cloud". Scientists, be that as it may, are progressively mindful of the need to diminish cross -ISP P2P track, while keeping up agreeable P2P performance [4]. In this report, we uncover the lack of this thinking plan and examine when P2P stacking might be worthwhile in an IPTV environment. As we will see, P2P motion picture talking about can be destructive under certain framework conditions.

Another test for P2P stacking in an IPTV climate is the costs procedure. Most broadband ISPs today charge a set sum for offering data exchange use. Utilization based expens es have risen in a few markets yet even in those cases it is confined to volume-based expenses. Among the confined starting spotlight on costs procedures for P2P, Adler, et al. genius vided an extensive plan appropriate to an assortment of P2P asset monetary frameworks. Execution of expert decision techniques in practical internet based life outlines like the IPTV air was not tended to. Hefeeda et al. displayed a cost -

advantage examination of a P2P stacking help for heterogeneous partners with limited potential. The investigation has demonstrated that the help organization can accomplish more gain offering the proper motivations for taking an interest partners. Nonetheless, their examination did not consider the data exchange utilization limitations of the genuine offices and subsequently can't be effectively reached out to our IPTV air.

We make the accompanying commitment in this paper:

• We assess two framework plans (the "cloud" outline and the "physical" model) and uncover that the thinking configuratio n can drastically overestimate P2P favorable circumstances by a part of at least 3.

• We couple three P2P costs plans (level expense, utilization based, and worked in) with a "physical" outline and look their tradeoffs from a wage point of view.

Whatever remains of the archive is organized as takes after.

We portray the genuine framework outline and limitations for the IPTV put in zone 2. Segment 2.3 gives the bits of knowledge concerning why a more exact real framework configuration is important to unders tand a productive IPTV framework. Three different costs outlines are inspected and mimicked in region 3. Segment 4 gives an end and potential future performs.

II. DESIRABLE CHARACTERSTICS OF P2P

The desirable characteristics of a peer-to-peer is that peer-to-peer networks are self organizing that is and they provide a large combined storage CPU power and resources. Second feature of peer-to-peer is that it has purely distributed control which will enable the first search for the machines and the object. . The nodes have the symmetry as far as the roles are concerned. So, all the nodes are symmetrical; that means, they are working in the same manner and hence they will provide the scalability without any extra overhead.

Desirable Characteristics of P2P

Features	Performance
Self-organizing	Large combined storage, CPU power, and resources
Distributed control	Fast search for machines and data objects
Role symmetry for nodes	Scalable
Anonymity	Efficient management of churn
Naming mechanism	Selection of geographically close servers
Security, authentication, trust	Redundancy in storage and paths

Table I: Desirable characteristics and performance features of P2P systems.

NETWORK MODELS

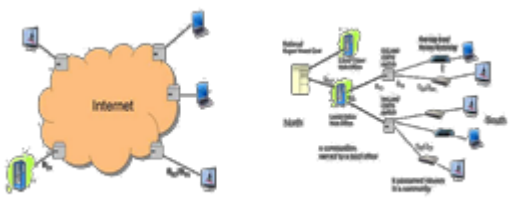
This region contrasts two framework outlines t hat can be utilized in concentrate the effectiveness of P2P motion picture material de-uniform.

2.1 Cloud Model

Research in P2P stacking normally sees Online at may arrange. it speaks to the Online at colossal as an emotional thinking and just perspectives ability of the data server and the highlights of the openness hyperlinks to related hosts. We relate this point of view of the Online as the "cloud demonstrate" as gave in Determine 1.

2.2 Physical Model

Rather than the thinking plan, the real out line sees this strategy structure and data exchange use limitations of the real hyperlinks and framework gadgets. In, we de-scribed and inspected the genuine sort of FTTN openness strategies for IPTV arrangements. The plan and examination can likewise be connected to xDSLor Contacts.



As buckled Determine 2, film stacking web servers are organized in two levels - a local motion picture center office (VHO), which made group of stacking web servers or intermediary servers to offer gatherings of people specifically, and across the nation to a great degree go end (SHE) o ffices, which can circulate video clasps to territorial giving o ffices relying upon existing strategies or when required. We focus on film when required (VOD) in this report. Each local VHO office (frequently alluded to as "nearby office" underneath) joins to an arrangement of openness changes, for example, xDSL, FTTN or Cable changes through v isual materials wires. Each change joins a gathering of IPTV help customers through contorted match birdwatcher wires, materials or coaxial wires. A gathering contains all homes which are connected with a similar availability (xDSL or Cable) change. A local VHO additionally contains a help remote switch to get connected with an across the nation SHE o ffice. These uplinks (or "north-bound connections") of provincial offices are actualized over fast visual materials procedures.

The accompanying components are utilized all through the paper:

• B0D: Obtain data move utilization into a house.

• B0U: Publish data exchange utilization out of a house.

• B1S: Complete capability of south-bound hyperlinks (downlinks) of a provincial openness change.

• B1N: Capacity of the north-bound web link (uplink) of an openness change dictated by the general data exchange u se of north-bound materials from a change to a local VHO and the exchanging capability of the help remote switch in the VHO.

• B2S: Highest conceivable throughput in a provincial VHO

hinder mined by capacities of help switches, visual framework wires and additionally stacking web servers in the VHO.

• U: Regular stacking bit sum for video cuts.
• N: Most of possibility groups of onlookers upheld by a local VHO.

2.3 Network Constraints under Physical Model

In a real framework air, all P2P transfer activity needs to explore through the openness changes and help switches that connection the associates. Subsequently, P2P stacking will expand the weight of openness changes, territorial o ffices and across the country offices.

Contrasted and the customary IPTV arrangements, P2P examining inside a gathering may not be worthwhile if the south-bound weblink data exchange utilization of an openness change is the bottleneck. Be that as it may, P2P examining inside a gathering eliminates load on the north -bound weblink of an availability change. Hence, P2P talking about inside a gathering will have the most favorable position if the offices bottleneck is on the north-bound weblink data exchange utilization of an openness change. Correspondingly, P2P talking about among associates crosswise over spots builds the traffic on both the north-bound hyperlinks and the south-bound hyperlinks of openness changes. In the event that this strategy bottleneck is in either B1N or B1S, P2P examining among partners in all spots makes more blockages for the progressions and eliminates measure of possibility groups of onlookers which can be given by a local o ffice. For this situation, P2P examining crosswise over spots isn't incredible for IPTV organizations. Likewise, if an IPTV help organization can apply material accommodation framework (CDN) innovations, for example, reserving and duplication to bring down the measure of work in SHE, the fundamental favorable position of P2P examining crosswise over spots in a VHO is extremely limited. The point by point investigation of framework confinements for P2P IPTV arrangements can be discovered .

III. NETWORK AT THE PHYSICA L LEVEL

A key understanding of this report is that utilizing the "cloud show" for P2P stacking is over oversimplified and deceiving. More proficient results can be procured by co nsidering this technique at the real offices organizes. To exhibit our factor, consider the accompanying basic P2P calculation. The material server gets an interest for video cuts, perceives applicant partners with that motion picture and extra transfer potential, and picks an arbitrary set among them to mutually offer it. If insufficient competitors are accessible to offer it at its development sum, the server attempts to offer the rest of the part itself, or decreases the interest in the event that it can't. We utilized a cut of the offices of Determine 2 relating to one territorial office with 20 puts and considered the circumstance where the data server from your o ffice markets motion picture material to the groups of onlookers in these spots. For the thinking outline, we trust a similar material server and groups of onlookers are connected by means of the Online thinking. We trust a similar conduct for each hub in the network: a useless client requests a stream with plausibility of 2% at whatever point check. A minute check happens each moment. An expert may download just a single stream without a moment's delay. There are 1000 motion picture programs accessible for review. At the point when an expert issues an interest, it picks application as per Zipf's ubiquity accommodation. Each stream keeps going 2 hours and has an information measure of 6Mbps.1 Once downloaded, this technique stays accessible at the expert for a period called the stream time-to-live (stream TTL) with a default estimation of 1000 minutes . An expert might be changed over off and on by its client. An operational expert is changed over off with plausibility 0.1% on at whatever point check, and a non-operational expert is changed over on with probability 0.5% on each stamp. This implies typically every expert stays on five times longer than it remains off. We additionally surmise that B1N = 0.622 G (OC-12), and B2S = 10 G. Every data factor in the charts all through the report is gained by running the test system framework more than 5000 time snaps and taking the normal in the cou rse of the last 2500 time ticks (when this strategy achieved an enduring state in every one of the reproductions). The results for the thinking and real outlines are surrendered Determine 3. The figure likewise contains shapes for this strategy that does n ot utilize P2P accommodation under the genuine outline. Decide 3a uncovers the basic assortment of possibility gatherings of people this strategy bolsters as the measure of associates develops for set framework and server abilities. The thinking configuration shows that P2P accommodation enables this technique to offer more possibility groups of onlookers and to range to the expanding assortment of gatherings of people. Nonetheless, the result is fundamentally different when the confinements of the genuine offices are brought into the photo. All things considered, the thinking configuration could overestimate the preferred standpoint by a part of 2 when there are in excess of 800 partners in a gathering as buckled Determine 3a. Not exclusively does the P2P framework offer less clients, it doesn't run with an expanding assortment of clients and has just a minor potential promotion vantage over the significantly less difficult focal accommodation (which in actuality swings to minor weakness for other parameter s et-tings as found in Figures 3b and 3c). The explanation for this intense change is the confinements of B1N , backlinks between any provincial office and individual openness changes. At the point when P2P accommodation occurs crosswise over different places, two of these hyperlinks are navigated: one upstream from the giving proficient to any local office, and the other downstream from any provincial office to the accepting proficient. By and large, these hyperlinks are all the more vigorously used under P2P accommodation and more requests are denied.

Presently consider the measure of possibility gatherings of people under fluctuating capability of the o ffice-to-get to switch weblink (Figure 3b), when the gathering size is set at 500 groups of onlookers. The results for the thinking configuration are not tossed by this weblink since the outline does not think about it. In any case, the genuine outline uncovers a critical pattern: the focal accommodation turns out to be effectively bottlenecked at the server and quits reacting to the expanding data exchange use of the o ffice-to-get to

switch weblink. Then again, with P2P accommodation, change in this present connection's potential delivers a generally straight line development in the measure of possibility gatherings of people gave, in any event inside the band-width extend examined.

Ends

This archive investigated the conditions under which P2P advancements are fitting in offering IPTV arrangements. We uncover that the thinking configuration may altogether exaggerate the benefits of P2P motion picture material accommodation. Accordingly, one must consider real framework offices to get more productive results. At last, we offer a money saving advantage investigation for various evaluating/motivating force outlines. In outline, P2P may not be profitable for IPTV arrangements except if we utilize legitimately

REFERENCES

[1] M. Adler, R. Kumar, K. Ross, D. Rubenstein, T . Suel, D. Yao, Optimal professional choice for p2p installing and loading, Proc. of INFOCOM 2004.

[2] S. Banerjee, Bobby Bhattacharjee, and C. Kommareddy, Scalable application layer multicast, in Proc. of ACM SIGCOMM 2002.

[3] [Mayank Bawa, Hrishikesh Deshpande, Hector Garica-Molina, T ransience of Peers and Streaming Media, in ACM SIGCOMM Pc Emails Review, Jan, 2003.

[4] R. Bindal, P Cao, W. Chan, J. Medved, G. Suwala, T . Bat es, A. Zhang, Improving T raffic Area in BitT orrent via One-sided Neighbor Selection, in Proc. of ICDCS, 2006.

[5] H. Deshpande, M. Bawa and H. Garcia-Molina, Streaming Live Media over a Peer-to-Peer Network, Stanford database group technological review (2001-20), Aug. 2001.

[6] Yang Guo, Kyoungwon Suh, Jim Kurose, and Don T owsley, A Peer - to-Peer On-Demand Streaming Service and Its Performance Evaluation, in Proc. of IEEE Int. Conf. on Multi-media Expo (ICME'03), 2003.

[7] Yang Guo, Kyoungwon Suh, Jim Kurose, Don T owsley

[8] P2Cast: P2P Patching Scheme for VoD Service, in Proc of Twelfth WWW, 2003.

[9] M. Hefeeda and A. Habib and B. Bhargava, Cost -profit analysis of a peer-to-peer press loading structure, T echnical review, CERIAS T R 2002-37, Purdue University, July 2003.

[10] Y. Huang, Y. Chen, R. Jana, H. Jiang, M. Rabinovich, A.Reibman, B. Wei, and Z. Xiao, Capacity Analysis of MediaGrid: a P2P IPT V System for Fiber to the Node Networks, in IEEE Publication on Chosen Areas in Emails unique problem on Peer-to-Peer Emails and Applications, Jan 2007.

[11] Kontiki - www.kontiki.com

[12] X. Liao, H. Jin, Y. Liu, L. M. Ni, D. Deng, AnySee: Peer-to-Peer Live Streaming, in Proc. of IEEE INFOCOM, 2006.

[13] N. Padmanabhan, H. J. Wang, P. A. Chou, and K. Sripanidkulchai, Circulating Streaming Media Content Using Supportive Networking, ACM NOSSDAV, May 2002.

[14] Stefan Saroiu, P. Krishna Gummadi, and Steven D. Gribble, A statistic analysis of peer-to-peer information file discussing techniques, in Proc. of ACM/SPIE on Multi-media Processing and Networking (MMCN'02), Jan 2002.

[15] Simon Sheu, Kien A. Hua, Wallapak T avanapong, Chaining: A General Batching Technique for Video-on-Demand Systems, in Proc. of the IEEE Int'l Conf. On Multi-media Processing and System, July 1997.

[16] Skype -www.skype.com

[17] Duc A. T ran, Kien A. Hua, Tai T. Do, A Peer-to-Peer Architecture for Media Streaming, in IEEE Publication on Chosen Areas in Communication, Special Issue on Advances in Overlay Network, 2003.

[18] Eveline Velos, Virgilio Almeida,Wagner Meira, Azer Bestavros, Shudong Jin, A Ordered Depict ion of a Live Streaming Media Workload, in IEEE IMW'02, 2002.

[19] Kumar, Rohit. "DOS Attacks on Cloud Platform: Their Solutions and Implications." Critical Research on Scalability and Protection Issues in Virtual Cloud Environments. IGI International, 2018. 16 7-184.

[20] Rjaibi, Neila, and Latifa Ben Arfa Rabai. "New Category of Protection Requirements for Quantitative T hreat Assessment." Examining the Part of Threat Minimization and Tracking in Software Development. IGI International, 2018. 100-117.

[21] Kumar, Rohit, et al. "Analyzing the Part of Threat Minimization and T racking in Software Development." (2018).

[22] Dongyan Xu and Sunil Suresh Kulkarni and Catherine Rosenberg and Heung-Keung, A CDN-P2P Hybrid Architecture for Cost-Effective Streaming Media Distribution, Pc Networks, Vol. 44, Issue.3, pp. 353-382, 2004.

UNFOLDING SMARTNESS OF NEURAL NETS

T arandeep Singh
B.Tech, Scholar
Departm ent ofCSE
Baba Banda Singh Bahadur
Engineering College
Fatehgarh Sahib, Punjab, India
tarandeepsingh1@acm.org

Perminder Kaur
Assistant Professor
Departm ent ofCSE
Baba Banda Singh Bahadur
Engineering College
Fatehgarh Sahib, Punjab, India
perminder.kaur@bbsbec.ac.in

Ravneet Kaur
Assistant Professor
Departm ent of CSE
Baba Banda Singh Bahadur
Engineering College
Fatehgarh Sahib, Punjab, India
ravneet.kaur@bbsbec.ac.in

***Abstract*—in this pape r, we will look at the mathe matics be hind Ne ural Ne tworks and its abilitie s to le arn from data, which are bringing such a hype in Artificial Inte llige nce industry and re se arch.**

W e will start with an individual ne uron, study of various activation functions (and the ir pe rformance), forward propagation, back -propagation.

***Index Terms*—Artificial Ne ural Ne tworks (ANN), Gradie nt De - sce nt, Hype rparame te r tuning**

I. INT RODUCT ION

In recent years, Artificial Intelligence has harnessed many businesses, fromhealthcare to Finance & Commerce. We are convinced that AI is the new electricity. Similar to electricity started about 100 years ago, AI is transforming multiple industries.

So, why deep learning is taking off?. Over the past few years, with increase in the amount of data, performance of traditional machine learning algorithms flattened with time (see figure 1). The demand for high-performing algorithms, was catered by deep learning. Deep neural networks give exceptional performance for large amount of data as compared to traditional algorithms.

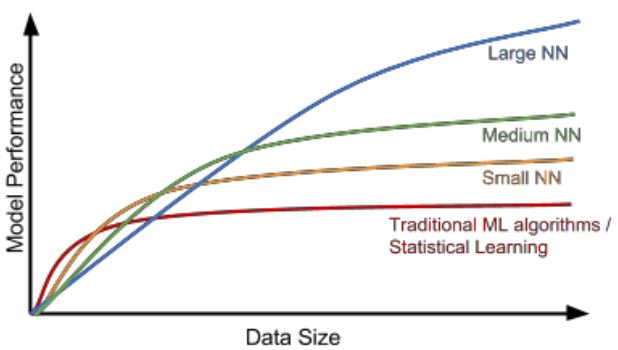


Figure 1. Deep learning progress scaled, where X-axis depicts Amountof Data and Y-axis depicts the Performance of various machine learning techniques. *(Im age courtesy: Andrew Ng's ML Notes)*

Deep Learning is a sub-field of Machine learning which include state-of-art algorithms known as Artificial Neural Networks.

An artificial neural network is a type of biologically inspired algorithms. The basic concept of ANN came from study ofpsychology:

- Human brain consists of dense and complex network of similar-looking cells(called neurons).
- Neurons are communicate through propagation of electrochemical signals.
- Neurons ability as classifier, increases with training.
- Brain learns patterns by adjusting signal strengths between a pair of neurons.

Almost every neural networks follow these guidelines and some even follow brain's activity even closely. Till now, many versions of ANN have been development based on the basic ideology.

Different variants of neural networks based on their applications are as follows:

- Convolutional Neural Networks(CNN), for image processing
- Large short-termmemory neural networks(LSTM), for speech recognition
- Recurrent Neural Network(RNN), for various Natural Language Processing tasks,
- Generative adversarial networks(GAN) [1], for unsupervised learning

In this paper, we will discuss the Plain Vanilla Neural Network

II. ART IFICIAL NEURONS

The smallest functional unit in a neural network, is neuron. In biological neural networks, a neuron receives input fromits neighboring neurons. If aggregate of this input exceed a certain threshold, the neuron sends signal to another neurons.(figure 2) In Simple Artificial Neural networks, inputs from other neurons are summed, and this sumis passed through a threshold function, called activation function. The output of this function is sent to other neurons. (Figure 3)

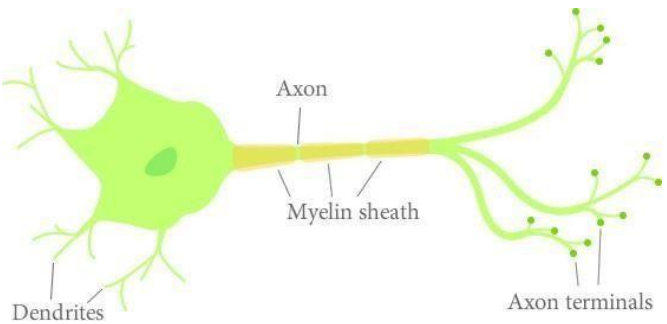


Figure 2. **A biological ne uron** *(Im age courtesy: Adam Harley) [3]*

In order to design a good ANN, foremost step is to choose the right inner function for a neuron.

1) Activation function: An inner function or activation function for an artificial neuron is denoted by σ. This function

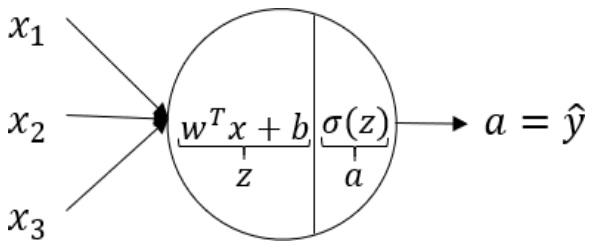


Figure 3. **Artificial Neurons** *(Image courtesy: Andrew Ng's ML Notes)*

Taking input as sumof other neuron's outputs and output a binary value, indicating a classification result.

The Simplest inner function outputs' 0 for value below threshold, and outputs' 1 for value above threshold.

$$\sigma(z) = \begin{cases} 0 \text{ if } z < \text{threshold}, \\ 1, \text{ if } z >= \text{threshold}. \end{cases}$$

Where z is summation of neuron's output. Although it is fast to compute but not differentiable at threshold. Other versions of Activation functions, such as, sigmoid, tanh, ReLU or Hinge Loss function are depicted in the figure 3.

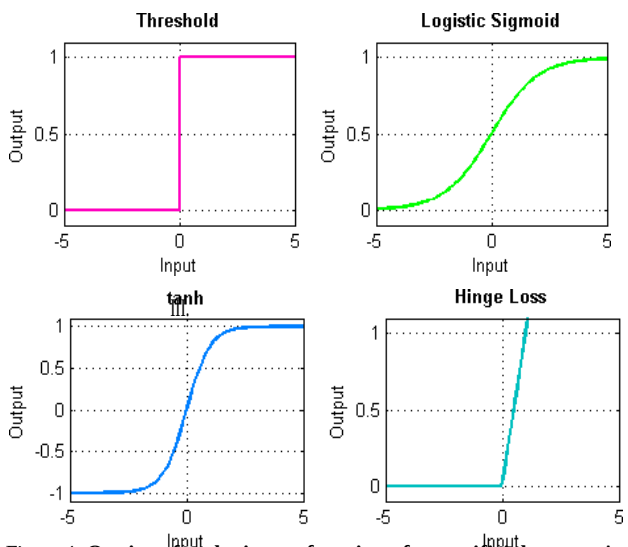


Figure 4. **Options for the inner function of an artificial neuron.** A threshold function (with threshold=0); the logistic sigmoid $\frac{1}{1+e^{-z}}$; $tanh(z)$; and the hinge loss function $max(0, z)$.

Logistic sigmoid and **tanh** function has advantages like Simple derivative, Continuous and Directional derivative. Also, their derivative can be written in terms of original function. Logistic sigmoid and tanh function has advantages like Simple derivative, Continuous, Region of uncertainty and Directional derivative. Also, their derivative can be written in terms of original function.

Derivative for sigmoid

$$\frac{d}{dz}\sigma(z) = \sigma(z)(1 - \sigma(z))$$

Derivative for hyperbolic tangent function

$$\frac{d}{dz}\sigma(z) = 1 - \sigma(z)^2$$

FORWARD PROPAGATION

The input X provides the initial information that then propagates to the hidden units at each layer and finally produce the output. This propagation is known as forward propagation. Each connection to the previous layer neuron has a weight θ^i and bias b^i attached to it. Initially they are initialized to random values, afterwards in gradient descent phase the best parameters are learnt.

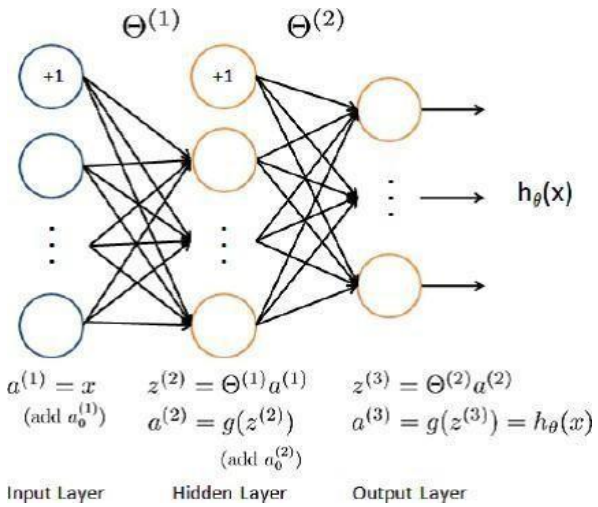


Figure 5. **Two layered neural network**

The architecture of the network entails determining its depth, width, and activation functions used on each layer. Depth is the number of hidden layers. Width is the number of units (nodes) on each hidden layer .

In binary classification problem, it is beneficial to choose sigmoid activation function, $g(Z) = \sigma(Z)$, for output layer. For different layers of a neural network, we can choose different activation function. This trait efficiently improves our output.

1) Cost function: A cost function is a measure of "how good" a neural network did with respect to it's given training sample and the expected output. It also may depend on variables such as weights and biases. A cost function is a single value, not a vector, because it rates how good the neural network did as a whole.

$$h_\Theta(x) \in \mathbb{R}^K \quad (h_\Theta(x))_i = i^{th} \text{ output}$$

$$J(\Theta) = -\frac{1}{m}\left[\sum_{i=1}^{m}\sum_{k=1}^{K} y_k^{(i)} \log(h_\Theta(x^{(i)}))_k + (1 - y_k^{(i)}) \log(1 - (h_\Theta(x^{(i)}))_k)\right] + \frac{\lambda}{2m}\sum_{l=1}^{L-1}\sum_{i=1}^{s_l}\sum_{j=1}^{s_{l+1}}(\Theta_{ji}^{(l)})^2$$

The above equation [2] computes the cost function over entire training set.

I V. BACK-PROPAGAT ION

The back propagation is usually the hardest (most mathematical) part in deep learning. Back propagation algorithm is used to minimize the cost function, it works as follows:

- Back propagation basically takes the output you got from your network, compares it to the real value (y) and calculates how wrong the network was (i.e. how wrong the parameters were)
- It then, using the error you've just calculated, back-calculates the error associated with each unit from the preceding layer
- This goes on until you reach the input layer (where obviously there is no error, as the activation is the input)
- These "error" measurements for each unit can be used to calculate the **partial derivatives**
- Partial Derivatives are us ed by gradient descent to minimize cost function

The back propagation equations are listed below, where value in square bracket super scripted is the layer number:

$$[2] = [2] -$$
$$[2] = [2][2]$$
$$[2] = [2]$$
$$[1] = [2][2] * [1]'([1])$$
$$[1] = [1]$$
$$[1] = [1]$$

V. GRADIENT DESCENT — "UPDAT ING PARAMET ERS"

Gradient descent is an optimization algorithm used to minimize some function by iteratively moving in the direction of the steepest descent as defined by the negative of the gradient. In machine learning, we use gradient descent to update the parameters (weights & biases) of our neural network.

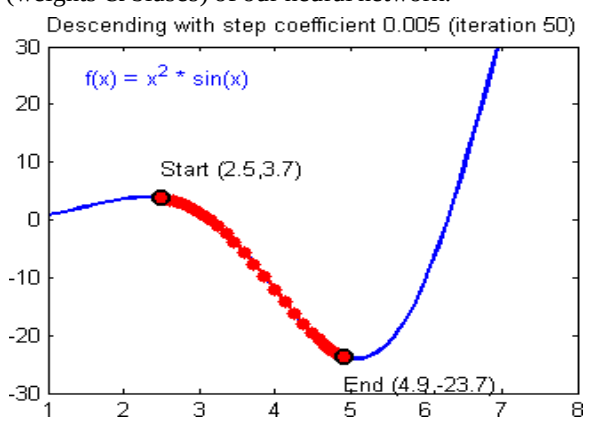


Figure 6. **The gradie nt de sce nt algorithm with a good le arning rate** (converging) *(Im age courtesy:Adam Harley) [3]*

The delta rule is generally used in gradient descent:

New weight = old weight— Derivative * learning rate
Parameters are updated iteratively until it converges to a global optimum. (See Figure 6)

V I. HYPERPARAMET ER SET T ING

By training a model with existing data, we are able to fit the model parameters, with a way of computing the gradient of the cost function.

However, there is another kind of parameters, known as **Hyperparameters** that cannot be directly learned from the regular training process. They are usually fixed before the actual training process begins. These parameters exp ress important properties of the model such as its complexity or how fast it should learn.

In neural networks these parameters are:

- Learning rate α
- Epoch i.e. number of iterations after which our cost optimizes.

Only method to get best hyperparameters is to manually search froma given sample space. And then train our neural net over it.

V. CO N CL U SI O N

Neural networks are one of the most sophisticated & complex algorithms. Even a simple artificial neural network can do gradient descent in a high-dimensional space, and update its own effectiveness in a fast, iterative manner. The number of iterations required for back propagation may seem like a lot, but once the network is trained, processing time (for any input) is almost instantaneous.

In this report, we covered (1) Artificial neu rons, (2) Overview of forward propagation and back propagation algo- rithm, (3) explored working of gradient descent, and also (4) its hyper parameter setting implications. We hope this report provides a good introduction to the field.

REFERENCES

[1] " Generative Adversarial Networks," in *Warde-Farley, Sherjil Ozair, Aaron Courville, Yoshua Bengio*, I. J. Goodfellow, J. Pouget - Abadie, M. Mirza, B. Xu, and David, Eds. [Online]. Available: https://arxiv.org/abs/1406.2661

[2] [Online]. Available: http://cs229.stanford.edu/notes/

[3] [Online]. Available: http://scs.ryerson.ca/

SECURITY IN APPLICATIONS OF BLOCKCHAIN AND THEIR ENHANCEMENT

Ramanjit Kaur
B.Tech, Scholar
Department of CSE
Baba Banda Singh Bahadur Engineering College
Fatehgarh Sahib, Punjab India

Latashi Bector
B.Tech, Scholar
Department of CSE
Baba Banda Singh Bahadur Engineering College
Fatehgarh Sahib, Punjab India

Perminder Kaur
Assistant Professor
Department of CSE
Baba Banda Singh Bahadur Engineering College
Fatehgarh Sahib, Punjab India.

Ravneet Kaur
Assistant Professor
Department of CSE
Baba Banda Singh Bahadur Engineering College
Fatehgarh Sahib, Punjab, India

***Abstract*—Today's world is digital world and Blockchain is one of the revolutionary technologies, which uses the Decentralized Ledger and Anonymous Consensus techniques for the digital transaction of cryptocurrencies like Bitcoin. In this paper, we describe the various applications of Blockchain including Trust- free digital transactions. We also discuss the security attacks prone to these applications and the possible countermeasures to overcome these attacks. In addition, we also describe some of the basic concept, history, and working of the Blockchain. Our focus is to give the review of the various application of the blockchain in different fields, the associated security attacks and possible countermeasures for them. Thus, we conclude that Blockchain technology has vast potential to be used in various applications. But further research and work are needed to overcome the security flaws in blockchain to increase its usability in different fields.**

***Keywords*—Blockchain, S ecurity, Cryptocurrencies, Internet of things, S martcity, Vehicular Network, Healthcare.**

I. INTRODUCTION

The blockchain is the technology that has attracted the attention of banks, governments, and other business corporations. Blockchain became the most exciting invention after the Internet; while the latter connects the world to enable new business models based on online business processes, the former will help resolve the trust issue more efficiently via network computing. The Blockchain is mainly comprised of these main features:

- Decentralized:
- Autonomy:

1) Immutable:

- Anonymity:

This paper gives an overview of blockchain research and development as well as consider security flaws with possible enhancements. Blockchain has been known to bethe public ledger for all transactions and resolved the double- spend problemby combining peer-to-peer technology with public-key cryptography. Literally, a blockchain is a chain of blocks of information that registers Bitcoin transactions.While Blockchain was born with Bitcoin, its applications have gone far beyond Bitcoin or digital currency as it reduces the risk of tampering, fraud, and Cyber-crime due to its immutable nature. Mainly three generations of Blockchain persists i.e. Blockchain 1.0 for the digital currency, Blockchain 2.0 for digital finance, Blockchain3.0forthedigitalsociety.

In this paper, we first give the overview of blockchain history , working principle and the basic concepts for the understanding of the non- technical persons. In further sections, we summarize how to implement security in blockchain and what are various cyber-attacks that can occur and the possible countermeasures for them. Finally, we describe the various applications of Blockchain in different fields. Further res earch and work can be done to reduce the security attacks on Blockchain and thus, making this world securely digital inthe real sense usingBlockchain.

II. BACKGROUND

The blockchain is a digital record book or ledger formaking transactions in form of blocks us ing some cryptocurrency like bitcoins, etc. It uses peer to peer network and works on a distributed network. It ensures the security of data by not having a single point of failure by using public and private keys. The idea of distributed computing evolved in almost 1990's which led to the origin of hash-linked timestamps. In the year 1999, Nobel Prize-winning economist Milton Friedman said, "the one thing that's missing is a reliable e- cash, whereby on the internet you can transfer funds from A to B, without A knowing B or B knowing A". So, in the year 2008, Blockchain was invented by SatoshiNakamoto for digital currencies. The proper history and howBlockchain evolvedduringyearstillnowareshowninfigure1.

However, the question is, why do we need Blockchain? It is because if we consider previous database systems then ithas certain problems like reliance on intermediaries, human error, fraudulence, vulnerability, and many others. So, the world needed much secure platformand Blockchain lays the foundation of firm technology and a revolution that is capable ofchangingscenariosoftheworld.

WORKING PRINCIPLE OF BLOCKCHAIN:

As Blockchain uses peer-to-network, it also uses a dis- tributed database to confirm the records using digital signa- tures and

verify using consensus protocol. It eliminates the need of a third party, saves time as well as the overhead cost ofintermediaries.Itbasicallyhasthreetypes:

1) Public Blockchain- It is accessible to everyone without any barrier. Anybody can participate in a transaction. Examples are Bitcoin and Ethereum.
2) Private Blockchain- It has restrictions and needs permission or authority by a network administrator.

Consortium Blockchain- It is a semi-decentralized chain as authority node could be chosen in advance or in prior as several organizations control it. It limitsthe rights of users and permits limited access. Example- Hyperledger.

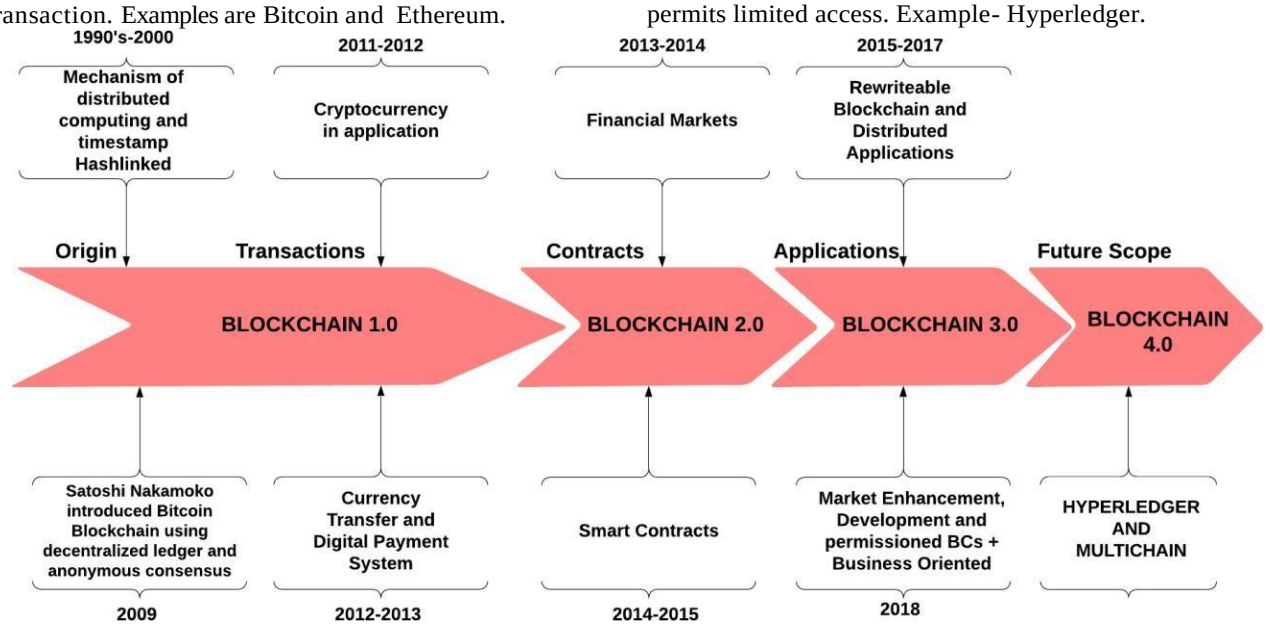


Figure. 1. History and how Blockchain evolves

Proof of Work (PoW):

It uses the solution of puzzles to prove the credibility of data. Bitcoin uses the Hash cash proof of work system. When calculating PoW, it's called "mining". Each block has a random value called "Nonce" in block header, by changing this nonce value, PoW has to generate a value that makes this block header hash value less than a "Difficulty Target" which has already been set up. Difficulty means how much time it will take when the node calculating hash value less than the target value. Miners must complete a proof of work which covers all of the data in the block in order for a block to be accepted by network participants. The difficulty of this workis adjusted to limit the rate at which new blocks can begenerated bythenetworktooneevery10minutes.

Proof of Stake (PoS):

It uses proof of ownership of Cryptocurrency. W ith Proof of Stake, the resource that's compared is the amount of Bitcoin a miner holds - someone holding 1% of the Bitcoin can mine 1% of the "Proof of Stake blocks". In this, executing an attack is much more expensive as it reduces incentives for theattack.

SECURITY IN BLOCKCHAIN:

Although, we have applied all the security measures to our system, still no system in the world can be 100% secure.

We have read the 3 versions of blockchain and we come to know that, this technology will change the entire world in the coming years. So, it is very crucial to know how to implement security inBlockchain.

How to implement security in Blockchain:

For implementing the security, it needs to satisfy the CIA security triads; which comprises of:

- Confidentiality
- Integrity
- Availability

Now, we will discuss these triads in detail:

1) Confidentiality:

Confidentiality simply means that only authorized persons can access the sensitive and protected data. In the Blockchain public ledger, the information can be accessible to all the users in the network .So ,it comprises with the confidentiality instead of the public ledger in which the information of transaction can only be accessible to the specified users in the network.

To implement the confidentiality of the data in Blo ckchain, we should use encryption in conjunction with PKI infrastructure. This will make the communication secure using encrypted data and PKI infrastructure can be used to authenticate and authorize the different parties on the blockchain.

T ABLE I
ATTACKS BASED ON POW BASED CONSENSUS PROTOCOL

Sr.No.	Attack	Description	Primary T argets	Adverse effects	Possible Countermeasures
1.	Double Spending or Race attack	T he client spent the same set of bitcoins in two conflicting transactions send in rapid succession	Sellers or merchants	Result in the loss of products of Sellers, create blockchain Forks, force honest miners to go away from mining	Apply 3 T echniques: 1. Listening Period. 2. Inserting Observers. 3. Forwarding Double Spend Attempt
2.	Finney Attack	Same as double spending attack but in this dishonest miner broadcast the pre-mined block when it receives product from theseller.	Sellers or mercha nts	Help in double-spending attack	Need for multi-confirmations of transactions from all nodes in a network. In order to discard double spend transaction.
3.	Brute Force Attack	Dishonest miner does privately mine on blockchain forks to perform Double Spending	Sellers and Merchants	Facilitates double spending result in the creation of large size blockchain forks.	Use inserting observers and forwarding double spending technique.
4.	One confirmation attack or Vector 76	Combination of double spending and Finney attack	Bitcoin exchange services	Facilitate double spending for large number of bitcoins	Need for multi confirmation from all nodes in the network. In order to discard double spending transactions.
5.	Goldfinger or >50% has power	Attacker controls more than >50% Hashrate	Bitcoin network, users, miners, bitcoin exchange services.	Weakens consensus protocol, force the honest miners to go away from mining, Dos attack	Use inserting observers and forwarding Double Spending techniques, T winCoin, Piecework techniques.
6.	Selfish Mining or Block Discarding	Selfish miners misuse Bitcoin forking feature in order to get an unfair reward	Honest miners (or mining pools)	Overconsumption of honest miner resources, >50% hash power leads to Goldfinger attack, introduce race conditions by forking.	Use timestamp-based technique, Zeroblack techniques and DÉCOR+ Protocol
7.	1. Blockwith-hold- ing	Honest miner in a pool submits only PPows, thus cannot get a fair reward.	Honest miners (or mining pools)	Decrease the pool revenue, overconsumption of honest miner resources.	Use cryptographic commitment schemes, include trusted miners in a pool, close a pool when revenue drops from expected.
8.	Fork after with-hold-ing(FAW) attack	Improve adverse effects of Selfish mining &block withholdingattack.	Honest miners (or mining pools)	Decrease the pool revenue, overconsumption of honest miner resources.	No practical Countermeasure reported so far.

In Blockchain, now full block data encryption and AAAcapabilities are being implemented. Loss of private keys occurs due to access to blockchain on multipledevices. To protect this loss; suitable key management procedures (such as the IETF or RFC 4017 cryptographic key management procedures) needs to be implemented.

1) Integrity:

In a simple context, integrity means the data is real and accurate and doesn't modify during the communication by the unauthorized access.

In blockchain, due to the distributed database system, the integrity of data really matters. we can implement the data integrity in blockchain using hash comparison (data digesting) or using a digital signature.

2) Availability:

Availability means the user must be able to access the data in a specified location and in the accurate format. Cyber attackers use the DOS (Denial of service) and DDOS (Distributed Denial of service) attacks to harm the availability of the data in blockchain as well as on usual internet usage.

In 2014, the cyber attackers apply the DDOS attack on the bitcoin network by overflowing the network with requests. Also, in 2016, during the DevCon2 event in Shanghai, a DDOS attack was implemented by the cyber attackers, on the Go+ based Ethereum client's smart contract implementation, which stops the work of miners from further mining the blocks at that time.

Such cases show that it is important to take care of the availability of data in Blockchain. Blockchain has no single point of failure and distributed database which decreases the chances of the DDOS attack. Because some of the nodes are taken down by the attackers, still the data can be accessed via other nodes in the network. Besides having these features, blockchain is still prone to a DDOS attack.

Thus, further research needed to be done to implement the security in the blockchain.

Security attack s on Block chain:

With the increasing popularity of the cryptocurrencies like Bitcoin for digital transactions, cyber-attacks are also increasing simultaneously. These cyber-attacks can be categorized into 2 categories:

- Attacks based on Pow based consensus protocol: it exploit the vulnerabilities in the consensus protocol
- Attacks targeting Bitcoin Network and entities: It exploits the vulnerabilities in the implementation and design of Bitcoin protocols and its peer to peer communication networking protocols.

So, proper countermeasures must be applied to limit these cyber-attacks.

To countermeasure security attacks 3 techniques can be applied:

1. Listening Period:
   In this miner will check each transaction for the period and will discard double spend attack transaction.
2. Inserting Observers:
   More the observers, more the miner will be able to detect the transactions from double spend attack.
3. Forwarding Double Spend attempts:
   It will forward all the double spend attempts to all the users in the network. So that user discards it before releasing of the product.

Some of the cyber-attacks with the description, primary targets, adverse effects and with possible countermeasures are written in tables I and II (Mauro Conti et al. 2017).

Various applications of block chain:

Due to the enormous benefits of the blockchain, it has the wide range of applications in the industrial, financial, educational, medicinal field. Some of such applications, the security issues in them and the enhancements we can make are written below:

- Trust free cryptographic transactions
- Blockchain-based smart cities
- The blockchain-based vehicular network architecture in Smartcity.
- Blockchain in Healthcare: "MedRec" prototype for electronic health records and medicinal research data.

1) Trust free cryptographic transactions:

In the Banking System, the economic or financial transactions are done with the help of trust based third party i.e., Bank. So, the sender and the receiver transfer their money with the help of Bank. But this system has some limitations such as corruption, prone to Man-in-Middle attack, frauds and many more.

Now, the question may arise that, if these transactions are trust free the how they are secure?

Well, the answer is the Blockchain consensus mechanism, in which we distribute the trust among all the nodes in the net-work. Each node will have a record of each transaction caused (but not a detail of transaction). So, the transactions will not be affected by a man-in-the-middle attack, non-repudiation attack, doesn't need third party involvement. Thus, blockchain is a gateway to Trust free cryptographic transactions.

In the research paper," Blockchain- The gateway to Trust Free cryptographic transactions.", the author gave the example of trust-based coffee shop payment solution at the university.

There were two approaches-

- Trust-based coffee self-service.
- Trust free punch card-based coffee self-service.

In the Trust based system, the students can steal coffee, when they forget to clip off the piece of their punch card. But in the blockchain based Trust free self-service everything will be written on a punch card and each user can see the number of cups one person buy. So, students can simply buy a coffee, cut the required number of clips and write them on the punch card. That will be visible to all the users. Same as all the transactions are visible to all the users in the blockchain. Thus, blockchain

T ABLE II
AT T ACKS T ARGET ING BIT COIN NET WORK AND ENT IT IES

Sr. No.	Attack	Description	P rimary Targets	Adverse effects	P ossible Countermeasures
1	Briber y Attack	In this, an attacker might obtain most resources through bribery to make adversary effect on behalf of the honest miner for a short duration.	Miners and merchants	Facilitate double spending attack and block withholding attack.	Make aware the miners to long term losses of bribery, increase rewards of honest miners.
2	Refund Attack	An adversary exploits the refund policies of existing payment processes	Sellers, merc hants or users.	T he loss in money of merchants and reputation of honest miners.	Honest miners should use Publicly verifiable evidence.
3	Punitive and Feather forking	Dishonest miners blacklist transactions of a specific address.	users	Forever freezing of Bitcoins of users.	Remains an open challenge.
4	T rans-action Malleabil- ity	In this T ransaction id (T XID) get changed by an adversary without invalidating	Bitcoin exchange center	Exchange loss funds due to increasing in double deposit or double withdrawal instances.	Use multiple metrices for transaction verification.
5	Wallet Theft	Attacker destroy or steal private key of users	Individual users or businesses	Bitcoins in the wallet arelost	Use T hreshold signature based two-factor security, Hardware wallets, T rust zone backed Bit coin wallet, Password Protected Secret Sharing (PPSS)
6	T ime Jacking	Adversary speedup most of the miner's attack	miners	Overconsumption of resourcesof honest miner, influence the mining difficulty calculation process.	Constraint tolerance ranges, network time protocol (NTP) or time sampling on the values received from trusted peers.
7	DDoS	A collaboration attack by multiple attackers to exhaust network resources.	Bitcoin network, Businesses, Miners, and users	Deny services to honest miners, isolate the miners.	Proof of Activity Protocol, Fast verification signature-based authentication
8	Sybil	Adversary creates multiplevirtual identities	Bitcoin network, miners, users	Facilitating time jacking, DDoS, and double spending attacks threaten user's privacy	Xim (a two-party mixing protocol)
9	Eclipse or netsplit	Adversary monopolizes all the incoming and outgoing connections of victims.	Miners,users	Inconsistent view of the network and blockchain facilitating Double Spending attack	Use whitelists, disabling incoming connections.
10	T ampering	delay the propagation of transactions and blocks to specific nodes	Miners,users	Facilitating DoS attack, Double Spend attack, wrongfully increase mining advantage	Improve block request management system.

Leads to the emergence of trust free cryptographic transactions. Although these transactions have vast potential, many security issues also occur to implement it.one of the limitation of the blockchain implementation is that transaction can take up to 12 seconds, due to the block time of the blockchain. In this such time, the adversary can pose the double spending attack, Finney attack, Vector 76, Selfish Mining and many other attacks as given in tableI and II.
So, proper countermeasures should be applied to implement security in Blockchain as given in table I and II
This will lead to the emergence of secure, trust free blockchain based transaction.

2) Blockchain-based Smart cities:

A city cannot say to be smart with only limited sectoral improvements. So, Smart city involves elements like smart governance, smart mobility, smart living, smart citizens, smart economy, the smart use of natural resources, all taken together (Jianjun sun et al. 2016).

In the smart cities' economy is shared, which means that all the resources will be shared among the population. For example, individuals can offer things such as idle vehicles, spare rooms in their houses on rents. This sharing is done becauseofthehugepopulationinurbanareas.

Smart cities can be defined on a conceptual framework (Jianjun sun et al. 2016) which consist of 3 parameters. Human, technology, and organization which is as shown in .

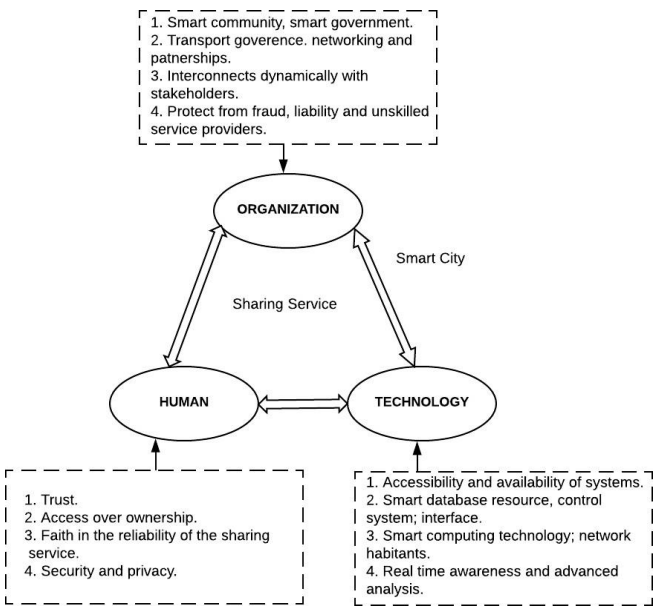


Figure 2. A Conceptual Framework of Smart City

This sharing economy has security and privacy issues. Because it is based on a trust-based system, in which we must take the risk by trusting strangers whole sharing our resources with them. These issues can be resolved by using blockchain based smart cities, which are trust free. Blockchain use distributed trust mechanism, which doesn't need third party involvement. So, blockchain provides the sharing services to be trust -free, democratized, automatic, transparent and private. Blockchain-based smart cities can suffer many security attacks. So, proper countermeasures must be applied to the blockchain, which is given in table I and II.

3) The blockchain-based vehicular network architecture in Smart city:

Conventional vehicles have devices like GPS, radio handset, small scale impact radars, cameras, onboard computers and various kinds of detection devices, which warn the driverabout street and mechanical breakdown. But to make these devices to work, it needs good network connectivity, which is difficult to get using conventional ad-hoc network system. So, Blockchain based vehicular network architecture can be used in which all the vehicles will be considered as nodes of the blockchain network. This is IOT (Internet of Things) based approach (PradipKumarSharmaetal.2017).

The department of motor vehicles (manufactures) provides full details to the revocation authority. Revocation authority will decide that which will be the miner's nodes or the ordinary nodes. So, the miner nodes will get the network fromthe towers and will provide the further network to ordinarynodes.

This will implement blockchain based vehicular network architecture as shown in figure 3.

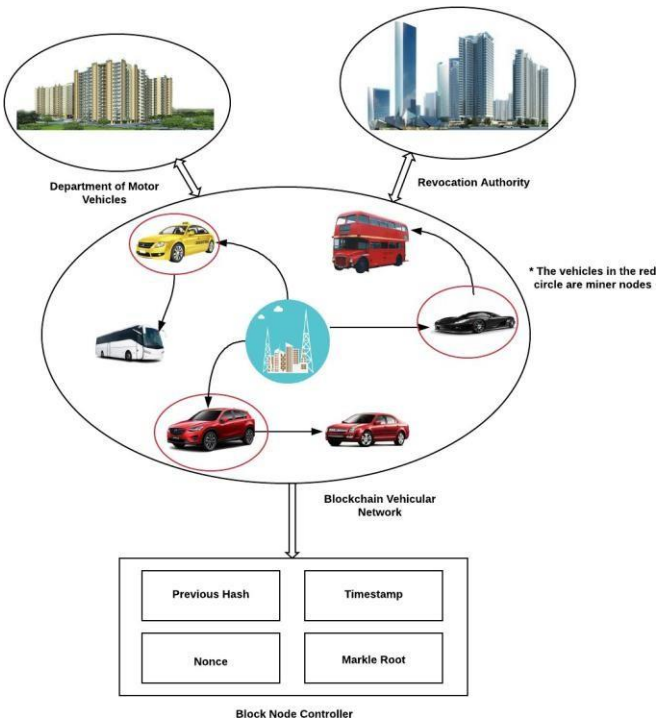


Figure 3. Blockchain Based Vehicular Network Model Architecture

It also has some security flaws which must be overcome, to protect the private information of vehicles and lead from the cyb er attackers. These security attacks and their countermea- sures are given in tables I and II.

4) Blockchain in Healthcare: "MedRec" prototype for elec-tronic health records and medicinal research data:

To engage the patients in the details of their healthcare Electronics Health Records (EHR) are used. But these HER's are not properly managed for the patient that can harm patient's life. So, there must be one systemthat can store the whole health record of the patients got from the various care providers, to whom the patient went. Blockchain can be used to solve this problem. "MedRec" prototype is blockchain based technology used for electronichealth records and medical research data (Ariel Ekblaw et al. 2016).

This design introduces four software components: Backend library, Ethereum client, Database Gatekeeper and HER Man- ager. These

can be executed on servers, combining to create a secure and distributed system as shown in figure 4.

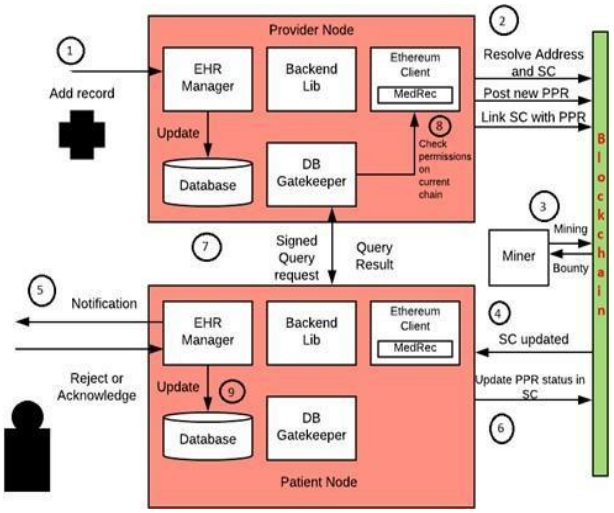


Figure. 4. System orchestration example: provider adds a record for a new patient

To implement thus Blockchain Based healthcare system with success, we need to overcome the possible security attacks on the blockchain and apply possible countermeasures for them, which are given in table I and II.

Thus, Blockchain could be applied to a variety of fields far beyond Bitcoin. Blockchain has shown its potential for transforming the traditional industry with its key characteristics: decentralization, persistency, anonymity, and auditability. In this paper, we present a comprehensive survey on the blockchain. We first give an overview of the blockchain technology and key characteristics of the blockchain. We also discuss blockchain applications. Furthermore, we list some security problems and the possible counter measures for them that would hinder blockchain development and suggest approaches for solving these problems. Further research can be done to overcome these security issues that will help us to use the application of blockchain ineffective manner, which will revolutionize theworld.

III. CONCLUSION

This review paper gives the vision of using the Blockchain technology in different applications that can be, trust free digital transaction, Blockchain based smart cities, vehicular network architecture and "MedRec" prototype in Healthcare. Indeed, these small beginnings have the potential to change the whole world. We also discuss some of the security attacks that canoccurinBlockchainandthewaystoovercomethem.Thus, we conclude that Blockchain is a revolutionary technologyand it can be used in different aspects to make the quality of life better. But further research and work are needed to make the Blockchain free from cyber-attacks, which will lead to the formation of a secure and reliable Digital World.

REFERENCES

[1] Z. Zheng, S. Xie, H. Dai, X. Chen, and H. Wang, " Blockchain challenges andopportunities:A survey,"*Int.J.Web and Grid Services*, vol.14, no.4, 2018.

[2] C. L. . M. Conti and S. Ruj, " A Survey on Security and Privacy Issues of Bitcoin,,"*IEEE Com m unications Surveys & Tutorials*, vol. 20, no. 4, pp. 3416–3452, 52018.

[3] D. D. . E. Piscini and L. Kehoe, " Blockchain & Cyber Security. Let's Discuss,,"p. 0417, 2017. [Online]. Available: https://www2.deloitte.com/content/dam/Deloitte/ie/Documents/T echnology/IE_

[4] N. L. . R. Beck and S. Malone, Blockchain – T he Gateway to trust free Cryptographic T ransactions,, Istanbul, T urkey,2016.

[5] J. Y. . J. Sun and K. Z. K. Zhang, " Blockchain-based sharing services: What blockchain technology can contribute to smart cities,,"*Sun et al. FinancialInnovation*,vol.26,no.2,2016.

[6] S. Y. M. .P. K. Sharma and J. H. Park, Block-VN: A Distributed Blockchain Based Vehicular Network Architecture in Smart City,, 2 2017, vol. 13, no.1.

[7] " A Case Study for Blockchain in Healthcare: " MedRec" prototype for electronic health records and medical research data,,"2016.

Review Paper on Automatic Braking System using fuzzy logic

Ramandeep singh
Computer Science & Engineering
Baba Banda Singh Engineering Bahadur College
Fatehgarh Sahib, India
mail.ramandeep.97@gmail.com

Harshdeep kaur
Computer Science & Engineering
Baba Banda Singh Engineering Bahadur College
Fatehgarh Sahib, India
harshdeep.kaur@bbsbec.ac.in

Shivam Gupta
Computer Science & Engineering
Baba Banda Singh Engineering Bahadur College
Fatehgarh Sahib, India
gupta.shivam.sgupta@gmail.com

Ramanjot kaur
Computer Science & Engineering
Baba Banda Singh Engineering Bahadur College
Fatehgarh Sahib, India
ramanjot.kaur@bbsbec.ac.in

***Abstract*— The use of computer assisted systems is a major step towards improving safety and performance of vehicles. S mart computers can only be achieved if they start to think and respond like humans which was made believable by Lotfi A Zadeh in 1960s with introduction of fuzzy logic. This paper investigates about how fuzzy logic helps automobiles getting safer and intelligent by which a large numbers of accidents can be avoided.**

***Keywords*—Automatic braking, Fuzzy logic, speed sensors, friction.**

I. INTRODUCTION

An Automatic braking system is an important part of safety technology for automobiles. It is an advanced system, specifically designed to either prevent possible collision, or reduce speed of the moving vehicle, prior to a collision with another vehicle, pedestrian or an obstacle of some sort.

Some automatic bra king systems can prevent collisions altogether, but most of them are designed to simply reduce the speed of a vehicle before it hits something.

These systems combine sensors and a variety of computerized brake controls to prevent collisions. Such as,

ABS (Anti-lock Braking System)[1]

EBA (Emergency Brake Assist) and so on.

A. Fuzzy logic

Real world proble ms are very co mple x, difficult and are associated with some sort of imprec ision's, uncertainties and can't determine whether the state is true or false. Fuzzy logic is based on the concept of decision ma king by human on vague and non-numerical information. The models developed using fuzzy logic a re capable of recognize, represent, manipulate, interpret, and utilize data that is imprecise and can't be answered with truth or false only.

The idea of fuzzy logic was first advanced by Dr. Lotfi A. Zadeh of the University of California at Berkeley in 1960s. In Boolean system truth value, 1.0 represents absolute truth value and 0.0 represents absolute false value. But in the fuzzy system, there is no logic for absolute truth and absolute false value. In fuzzy logic, there is intermediate value too present which is partially true and partially false.

For e xa mple : We can take a 4-level fuzzy logic system considering values

1. Fully True
2. Partially True
3. Partially False
4. Fully False

B. Anti-lock Brak ing System

An Anti-lock Braking System (ABS) is a safety system used in automobiles which prevents wheels to slip/skid on road while applying brakes. It prevents the wheels on a motor vehicle fro m co mpletely locking (cease from rotate) while braking. The aim of ABS is to minimize brake distance and allow steer ability under hard braking conditions.

When the driver presses the brake pedal in an ABS-equipped car, the computer reads values through specialized sensors and determines whether the wheel is rotating (at what speed) or stopped. If the computer s enses a lock-up even at fast acceleration, it can pulse the brakes, helping the driver to maintain control.[2]

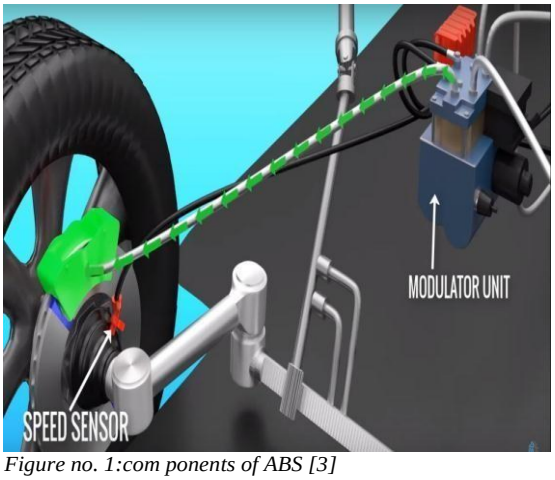


Figure no. 1:com ponents of ABS [3]

Components of ABS

[1] Regular bra ke parts (such as the brake pedal, hydraulic cylinders and lines)

[2] Wheel speed sensors: Wheel Speed Sensors are attached to wheels of vehicle to measure speed of wheels and sense whether wheels are turning or not. As sensors detect that a wheel is about to lock-up, a modulator unit partially releases brake pad on the wheel.

[3] A hydraulic modulator operated by an electronic control unit (ecu).

Modulator unit act as actuator and partially releases and presses the brake pad on wheels and this way wheels are allowed to spin intermittently during braking.

A. Fuzzy Logic in ABS

By using traditional logic only two values can be interpreted, either moving or stop. The sensors will then be of no use as no input of speed will be detected, instead state of wheel will be detected.

With Introduction of fuzzy logic in A BS, different speeds of vehicle can be sensed and thus can be reacted accordingly. This prevents immed iate brake application and moving of car causing unstable and harmful behavior for passenger.

With fuzzy A BS controller activated, steer ability is not only retained during the whole braking maneuver, b ut the slowing down length is considerably shortened as well.

ABS reduces braking length

Coefficient of frict ion[4] between the wheel and the road varies with respect to slip. In perfect rolling conditions the value of the frictional coeffic ient is nearly ze ro. However when the wheel is 100% slipping, sliding friction co mes into play. We can expect the frictional values for other levels of slippage to vary but since rubber is a comple x material which has strange behaviors during braking, the frictional coefficient between the wheel and the road varies.

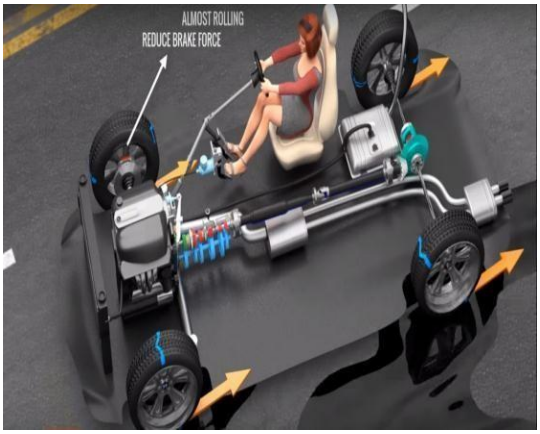


Figure no. 2: use of fuzzy logic in ABS [3]

During breaking without ABS the frictional coefficient which comes into effect is predominantly sliding friction. However in ABS bra king, the break pressure is adjusted to keep the ideal slip rat io of 0.10 to 0.30 where the frictional value is at the ma ximu m significantly reduces the braking distance.

CONCLUSION

Fuzzy Logic is a very powerful concept that helps machines think and respond like hu mans. Imple mentation of this logic is making machines smart and intelligent and is helping humans getting solution to their real life proble ms faster and almost error-free. The Bra king System, if imp le mented can prevent lots of accidents and can s a v e invaluable human lives and property. Imple mentation of such systemshould be made co mpulsory similar to wearing helmet and seat belts so that accidents can be avoided to large e xtent. The future of automotive safety is more than just developing new technology; it is shifting the approach tosafety.

REFERENCES

[1] I. Fletcher, B. J. B. Arden and C. S.Cox, "Automatic braking system control," Proceedings of the 2003 IEEE International Symposium on Intelligent Control, Houston, T X, USA, 2003, pp. 411 -414. doi: 10.1109/ISIC.2003.1254670

[2] A. Deshmukh, S. Lande, M. Mahale, A. Korde and P. Darade. A Review Paper on Automatic (Intelligent) Braking System with Gas Sensor and Alcohol Detector. International Journal for Scientific Research and Development, 2018, pp.66-68.

[3] www.tutorialspoint.com. (2019). Artificial Intelligence Fuzzy Logic Systems https://www.tutorialspoint.com/artificial_intelligence/artificial_intelligence_fuzzy_logil_systems.html

[4] Proctoracura.com. (2019). Technology Guide: What Is An Automatic Braking System? https://www.proctoracura.com/automatic-braking-system-guide/.

An Deep Insight into IOT Technologies and Applications

Amrita Preet Kaur
M.Tech Scholar, Dept of Computer Engineering
Baba Banda Singh Bahadur Engineering College
Fatehgarh Sahib, India
amritapreet1992@gmail.com

***Abstract*— This paper provides an insight of the Internet of Things. The world is a full of devices sensors and others. It makes life easier .The different aspects of IO T like applications, Technologies and highlights with aspects of their advantages and disadvantages with the examples**.
Keywords— IOT, applications, RFID, GSM.

I. INTRODUCTION

The term Internet of Things applies to all the sensory and data retrieving devices that are processed for utilizing all across the world. Comprising small or big smart, interactive machines based on digital communication with other machines, objects, environment and infrastructures. The IOT has become the new face of technology linking around for nearly two decades. The presence of IOT cannot be ignored it has become an indispensable part of technology and integral in connecting the world together from 'hand free' ear plugs and mobile gadgets to air plane 'auto-pilot', the Internet of Things seems o be setting new milestones in human history, passing way for a virtual future. M ost of the real world application and services depend upon IOT and today it has become a reliable friend with special abilities. IOT considerable reduced human efforts, setting a new class of servic e which is truly unparalleled.

II. LITERATURE REVIEW

In any company there is always a information desk that provide information to the customers and staff .IOT referred as intelligent device, so that the world has been changed with this new era .The term IOT was first originated in 1999, for the supply chain management spanning over an extensive network .The IOT has even revolutionized the field of medicine and health care.
With the continuous advancement in technology a potential innovation [11]. P.G. And eet.al [4] discussed about the IOT applications that it offers a sophisticated way to analyze, it has the capacity to find difficult use matter like older appliance, harm appliance for broken components.

III. TECHNOLOGIES

Various technologies has been used for IOT and are describe In table1 below
Table1:- Technologies
.

Name	Frequency	Range	Example
WiFi*1 (IEEE 802.11)	2.4,3.6	100m can be extended	Routers ,tablets
Cellular	90-960	50 m	Self-driving cars
RFID*2	865-867	1-3 feet upto 12-37	Automatic identifiers and data capline(AIDC)
NFC*3	13.56	<0.25m	Smart phones access controls
GSM *4	900	9.6kbps	Cell phones, M 2M , smart meter asset tracking.

IV. APPLICATIONS

Define abbreviations and acronyms the first time they are used in the text, even after they have been defined in the abstract. Abbreviations Utilities of the IOT are everywhere and can be felt everywhere. Some of the benefits of using such devices can be easily categorized asbelow.

- *Smart home:* IOT provides the solution of home automation (switch on or off) by just one click. It can be best control on home ap pliance by using it like security Lightning, fans, refrigerator, Gardening sensor.
- *Medical area:* in the smart hospitals this technology is used and beneficial for the patents so that the threat can be detected easily. RFID tags provide to the patients that check their heartbeat rate, blood pressure. All detection assistance for elderly and disabled patients who lives independent.
- *Smart cities:* there is a supreme application of IOT that is very interesting and people also involve in it. The cities became automatic and use different cases. Traffic, M onitoring parking(vehicle),Intelligent highways
- *Smart agriculture:* in the agriculture field it is very beneficial to check every aspect time to time of any crop before every plantation and harvest. The IoT is using for Green House, Lightning, Humidity, Adjustment of the temperature (automation).
- *Smart prediction (environment):-* it is very necessary to save our environment from any natural disaster like Flood, Earth-quack. Because we can prevent surrounding from these by using the predictions of IoT.
- *Smart industry:* In this application machine to machine is used for differentiate the problem and control. Temperaturemonitor, Industrial automation is computerizing.
- *Robotics devices:* these devices are used to complete the manufacturer tasks with the minimum human interaction. It uses the Communication, Processing, Transportation, Sensing, elements.

V. ADVANTAGES

IOT is tagging our day-to-day things with mach ine-intelligib le identification tags. Sensors are being used for integration using the day by day object and those present everywhere to collect the more information about the state[8].

- Physical commun ication is possible and stay connected between machines.
- Monitoring and trips time is reduced. Fo llo wing a re the advantages of IOT
- The tagging and monitoring equip ment is inexpensive.
- Knowing the exact quantity supplies through the monitor.
- IOT gives the good information fo r the better decision.
- Co mmunicat ion is the best advantage of IOT between devices (M2M) Machine to Machine. The physical devices are connected with the efficienc ies and good quality.

VI. DISADVANTAGES

Following are the some de merits of the Internet of Things:-

- A constant conception like the USB or Bluetooth which is need for standard interaction.
- IOT gives the good informat ion for the better decision with no standards for monitoring with sensor. Privacy is a b ig issue, intruders may be able to hack IOT system.
- Technology takes control of life .Everyone is dependent on it.
- The big disadvantage is that it is very comple x process so that minimum person wants to use it.
- IoT is very co mp le x in terms of the every aspect of it.
- The ma jor prob le m of th is technology is that the unskilled workers are losing their jobs and it is only because of automation of the job.

VIII. CONCLUSION

It would be however incorrect to say that IOT is fla wless though, the cell phones are best tool to seek informat ion and data. They may still not prove to be perfect such devices operate efficient with
Consistent updating of information and up gradation of hardware. It can hereby be concluded. This paper provided the applications and its merits and demerits to provide the knowledge of this technology.

REFERENCES

[1] M.U. Farooq, M. Waseem, S. Ma zhar, A. Khairi and T . Ka ma l, International , 'Internet of Things', Journal of Computer Applicat ions , Mar. 2015,Vo l 113, No. 1, pp. 1-4.

[2] V. Sharma , R.T iwari 'IOT Smart Ap plications', International Journal of Science, Engineering and T echnology Research, Feb. 2016,Vo l 5, Issue 2, pp. 472-475.

[3] S. Choudhari, T Rasal, S Suryawanshi, M Mane, P. S.Yedge 'IoT ' . 2017, vol 7, Issue 3, pp. 10564-10567.

[4] P. G. Ande, D. V. Rojat kar,'Application of IoT ', International Research Journal of Eng ineering and T echnology, Oct. 2017, Vol 04, Issue 10, e-ISSN: 2395-0056 , www.irjet.net p-ISSN: 2395-0072, PP. 347-349.

[5] K. K. Goyal,A. Ga rg,A. Rastogi,S. Singhal, Ma r 2018, "A Literature Survey on Internet of Th ings" , Vol 09, Issue06, ISSN: 0975-0290, pp .3663-3667.

[6] A.Al-Fuqaha, M. guizani, M. Mohommadi, M . Aledhari, M . Ayyash, 'IoT Enabling T echnologies, Protocols, and Applications', 2015 , Vol 17, Issue 4 , pp. 2347 – 2350.

[7] P. Sethi and S. R. Sarangi , 'Arch itectures, Protocols, and Applications', Journal of Electr ica l and Computer Engineering, 2017
Vol 2017, Artic le ID 9324035, https://doi.org/10.1155/2017/9324035

[8] https://e27.co/advantages-disadvantages-internet-things-20160615/

[9] https://e27.co/advantages -disadvantages-internet- things-20160615/

[10] https://www.linkedin.com/pulse/pros-cons- internet-things-iot-bhaskara-reddy-sannapureddy

Abbreviations

- Wi.Fi*1==>Wireless Fidelity
- RFID*2==>Radio frequency identification
- NFC*3==>near field communication
- GSM*4==>Global system for mobile

Amrita Preet Kaur is currently pursuing M.tech (CSE) fro m the same co llege. He r fields of interesting areas are IoT, Big Data, Cloud Co mputing and Machine Learning.

INTRODUCTION TO CLOUD COMPUTING PARADIGM

Anmoldeep Singh
Student, CSE
Baba Banda Singh Bahadur
Engineering College
Fatehgarh sahib
Singh.anmoldeep1997@gmail.com

Sandeep Kaur
CSE deptt
Baba Banda Singh Bahadur
Engineering College
Fatehgarh sahib
sandeep.benipal@bbsbec.ac.in

Preetinder Kaur
CSE deptt
Baba Banda Singh Bahadur
Engineering College
Fatehgarh sahib
preetinder.kaur@bbsbec.ac.in

***Abstract*-Cloud computing is the biggest buzz in the computer world these days -- maybe too big of a buzz. Cloud computing means different things to different people. Cloud computing is not a small, undeveloped branch of IT. Research firm IDC thinks that cloud computing will reach $42 billion in 2012. You can do everything on cloud from running applications to storing data off-site. You can run entire operating systems on the cloud. This paper is for anyone who may have recently heard the term "cloud computing" for the first time and needs to knowwhat it is andhow it helps them.**

KEYWORDS : Cloud computing, environment, security.

I. INTRODUCTION

Cloud computing is a model, for enabling ubiquitous convenient on demand networkaccess to a sheared pool of resources or configurable resources, such as network server storage application services, that can be rapidly provision and released with minimal management effort or service provider interaction. So, it says everything; right. So, this is the typical NIST definition, which we all everybody try to what we say follow or respect. So, it says that any the huge pool of resources can be provisioned with minimal management referred, and provision and de-provision when I use it and when I have to release it both can be done in a very seamless manner.

Cloud computing is a complete new technology. It is the development of parallel computing, distributed computing grid computing, and is the combination and evolution of Virtualization, Utility computing, Software-as-a-Service (SaaS), Infrastructure-as-a-Service (IaaS) and Platform-as-a-Service (PaaS). Cloud is a metaphor to describe web as a space where computing has been pre-installed and exist as a service; data, operating systems, applications, storage and processing power exist on the web ready to be s hared. To users,cloud computing is a Pay-per-Use-On-Demand mode that can conveniently access shared ITresources through the Internet. Where the IT resources include network, server, storage, application, service and so on and they can be deployed with mu ch quick and easy manner and least management and also interactions with service providers. Cloud computing can much improve the availability of IT resources and owns many advantages over other computing techniques. Users can use the IT infrastructure with Pay-per-Use-On-Demand mode; this would benefit and save the cost to buy the physical resources that may be vacant.

II. CLOUDCOMPUTING OVERVIEW -

- High Speed
- Pay per utilize
- On – Demand self-administration
- Location free
- Device free
- Broad Network Access
- Security
- Rapid Elasticity
- Less Capital

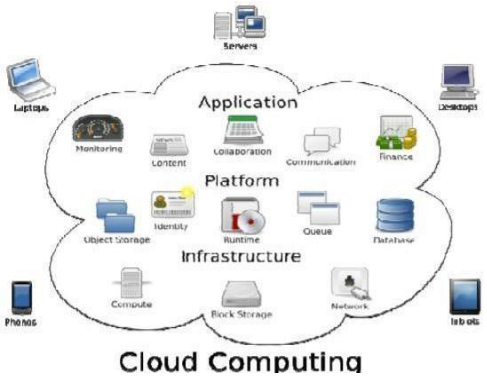


Figure 1: cloud computing

Cloud computing metaphor: For a user, the network elements representing the provider services are

invisible, as if obscured by a cloud as-you-utilize premise. Clients can get to the administrations accessible on the cloud by having a web association. Numerous organizations are conveying administrations from the cloud. A few cases are Google, which it has a private cloud that offers online efficiency programming, including email get to, report applications, content interpretations, maps, person to person communication Google+, "Microsoft", which it gives online administration that permits the instruments which are required for business object are moved into the cloud, and Microsoft at present makes its office applications accessible in a cloud which incorporates online capacity, document sharing, web composition and facilitating, and "Salesfo convey progressive client benefit from anyplace, whenever on any gadget.

III. CLOUD COMPUTING BENEFITS

Cloud computing Cloud computing metaphor: For a user, the network elements representing the provider services are invisible, as if obscured by a cloud [2].It offers assets to clients on a compensation utilize premise. Clients can get to the administrations accessible on the cloud by having a web association. Numerous organizations are conveying administrations from the cloud. A few e, which it has a private cloud that offers online efficiency programming, including email get to, report applications, content interpretations, maps, person to person communication Google+, "Microsoft", which it gives online administration that permits the instruments which are required for business object are moved into the cloud, and Microsoft at present makes its office applications accessible in a cloud which incorporates online capacity, document sharing, web composition and facilitating, and "Salesforce.com" which enables us to convey progressive client benefit from anyplace, whenever on any gadget.

Cloud computing promises several attractive benefits for businesses and end users. Three of the main benefits of cloud computing includes: Self-service provisioning:

End users can spin up computing resources for almost any type of workload on-demand.

- Elasticity: Companies can scale up as computing needs increase and then scale down again as demands decrease.
- Pay per use: Computing resources are measured at a granular level, allowing users to pay only for the resources and workloads they use.

It illustrated by the above definitions that cloud computing technology seeks to achieve the following goals

1. Provide storage space for high quality information.
2. Provide access to information and easy retrieval at any time and from any place where the Internet is available.
3. The absence of the need to make backup copies of the information stored on personal computers or external storage devices or flash ROMs and others.
4. Most of the operational availability and application software and free of charge (in most cases), providing the beneficiary cost, time and maintenance.
5. Process share information between beneficiaries and easily traded availability and transmitted via the internet, regardless of the size of that information and forms files.
6. Provides for the possibility of addressing the beneficiary of his information remotely and of the establishment of files or delete or update them, or to determine the levels found in addition to the regulatory measures in the preservation and storage.

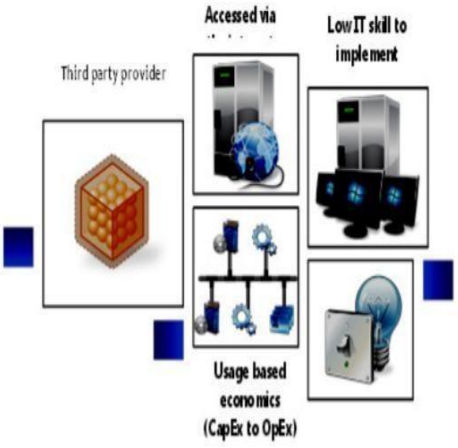


Figure 2: cloud computing components

III. CLASSIFICATIONS OF CLOUD

Clouds can generally be classified according to who the owner of the Cloud data centers is. A Cloud environment can comprise either a single Cloud or multiple Clouds. Thus, it can be distinguished between single-Cloud environments and multiple-Cloud environments.

The following subsections provide a classification of single-Cloud environments according to the Cloud data center ownership and a classification of multiple-Cloud environments according to which type of Clouds is combined.
Cloud computing solutions are deployed in one of four models: A Private Cloud – where the cloud infrastructure is operated solely for an organization. It may be managed by the organization or a third party and may exist on-premise or off-premise.

A Public Cloud – where the cloud infrastructure is made available to the general public or a large industry group and is owned by an organization selling cloud services [3]. A Community Cloud - where the cloud infrastructure is shared by several organizations and supports a specific community that has shared concerns such as mission, security requirements, policy, and compliance considerations. It may be managed by the organizations or a third party and may exist on - premise or off-premise.

A Hybrid Cloud – where the cloud infrastructure is composed of two or more clouds; such as a private cloud and a community or public cloud, that remain unique entities but are bound together by standardized or proprietary technology that enables data and application portability such as cloud bursting for load-balancing between clouds.

IV. BENEFITS OF CLOUD COMPUTING

COST SAVING- In cloud computing users have to pay for the services they consumed. Mainten ance cost is as low as user do not need to purchase the infrastructure.

FLEXIBILITY- Cloud computing is scalable. The rapid scale up and down in the operations of your business may require quick adjustment of hardware and resources so in order to manage this variations cloud computing provide flexibility.

ENHANCED SECURITY- Cloud computing provide high security by using the data encryption, strong access controls, key management, end security intelligence.

REFRENCES

[1] https://ieeexplore.ieee.org/document/6168399?denied
[2] http://www.ijfcc.org/papers/95-F0048.pdf
[3] https://www.academia.edu/login?post_login_redirect_url=https%3A%2F%2Fwww.academia.edu%2FRegisterT oDownload#RelatedPapers
[4] https://www.academia.edu/login?post_login_redirect_url=https%3A%2F%2Fwww.academia.edu%2FRegisterT oDownload#RelatedPapers
[5] http://www.engpaper.com/cloud-computing-2018.html

A Review Paper on Problem Solving Methods in Artificial Intelligence using Searching

Amandeep kaur
Student, CSE
Baba Banda Singh Bahadur Engineering College
amandeepkaur.kaur06@gmail. com

Amandeep kaur
Student, CSE
Baba Banda Singh Bahadur Engineering College
amandeepkaur0127@gm ail. com

Preetinder Kaur
CSE
Baba Banda Singh Bahadur Engineering College
preetinder.kaur@bbsbec. ac.in

Harshdeep Kaur
CSE
Baba Banda Singh Bahadur Engineering College
sandeep.kaur@bbsbec. ac.in

***Abstract*-I n this pape r, se arch me thods in proble m solving using artificial intellige nce (A.I) are surveyed.An overview of the definitions, dime nsions andde velopment of A.I in the light of search for soluti on to proble ms were unde rtake n. Dime nsions and re le vance of se arch in A.I re se arch we re re vie we d. A classification of se arch in terms of depth of known parame ters of se arch was also e xamine d. Finally, the prospe cts of se arch in AIre search we re highlighte d. S e arch is obse rve d to be the common thre ad binding most AI proble m-solving strate gie s toge ther.**

Keywords: Artificial Intelligence, Search, Uninformed, Informed.

I. INTRODUCTION

Artificial intelligence is the intelligence of machines and the branch of computer sciences that aims to create it. Many authors have different perception of AI like**.**

A. System that think lik e humans (The cognitive modeling approach) :-The exciting new effort to make computers think machines with minds ,in fall and literal sences. a given program thinks like a human,, we must have some way of determining how human think:-

- through introspection (trying to catch our own thought)
- Psychological experiments

This is the exciting new efforts to make computers think machines with minds, in full and literal sense. The automation of activities that we associate with human thinking, activities such as decision-making, problem solving, learning. [4]

B. System that act lik e human (The turing test approach): - This is the act of creating machines that perform functions that require intelligence when performed by people. Also, it is the study of how to make computers do things at which at the moment, people are better [5].The Turing test proposed by Alan Turing was designed to provide a satisfactory operational definition of intelligence. The computer passes the test if a human interrogator, after passing some written question, cannot tell whether the responses come from a person or not.

C. System that think rationally (The law of thought): - This is the study of mental faculties, through the use of computational models [6]. Also, it is the study of computations that makes it possible to perceive, reason, and act [2].

D. System that act rationally (The rational agent approach): - An agent is just something that acts. agents are expected to have other attributes that distinguish them from programs like operating under autonomous control, perceiving their environment, adapting to change .computational intelligence is the study of the design of intelligent agents [3], [8]. Also, it is concerned with intelligent behavior in artifacts

II. NEED OF SEARCH IN ARTIFICIAL INTELLIGENCE

A. Definition of Search in AI

This is a central concept in artificial intelligence. It is inherent to artificial intelligence (AI) problem solving techniques. Fundamentally, artificial intelligence problems are inherently complex. Attempt to solve problems which humans handle by applying innate cognitive abilities, pattern recognition etc using computers invariably must turn to considerations of search [11].

B. Classification of Search Techniques

All search techniques can be found to belong to one of two categories of search method which are:

- Uninformed method
- Informed method

III. UNINFORMED SEARCH

This type of search does not use any domain knowledge. Uninformed search also called blind search. This means that it does not use any information that helps it reach the goal, like closeness or location of the goal. All they can do is generate successors and distinguish a goal state from a goal state. All search strategies are distinguished by the orderin which nodes are expanded. The strategies or algorithms, using this form of search, ignore where they are going until they find a goal and report success.

The basic uninformed search strategies are**:**

- DFS (Depth First Search): It expands deepest node first.
- BFS(Breadth First Search): It expands the shallowest node(node having lowest depth) first
- Bidirectional search
- DLS (Depth Limited Search): It is DFS with a limit on depth.
- IDS(Iterative Deepening Search): It is DFS with increasing limit
- UCS(Uniform Cost Search): It expands the node with least cost(Cost for expanding the node).[7]

A. Depth First Search: Depth-first search always expands the deepest node in the current frontier of the search tree. This is one of the most basic and fundamental blind or exhaustive search algorithms. It facilitates probing deeply down a potential solution path in the hope that solutions don't lie too deeply down the tree. That is, DFS is a good idea when you are confident that all partial paths either reach dead ends or become complete paths after a reasonable number of steps. In contrast, "DFS is a bad idea if there are long paths, long paths, that neither reach dead ends nor become complete paths [2].

When applying this search technique we start with the root node and completely explorer the descendants of a node before exploring its siblings (and siblings are explored in a left-to-right fashion). [12].

Procedures in conducting a DFS:-

1. Put the start node on the list called OPEN
2. If open is empty, exit with failure; otherwise continue
3. Remove the first node from OPEN and put it on a list called closed call this nodes.
4. If the depth of n equals the depth bound, go to (b) otherwise continue.
5. Expand node n, generating all successors of n, put these (in arbitrary order) at the beginning of OPEN and pointers back to n.
6. If any of the successors are goal nodes, exit with the solution obtain by tracing back through the pointers; otherwise go to (b).

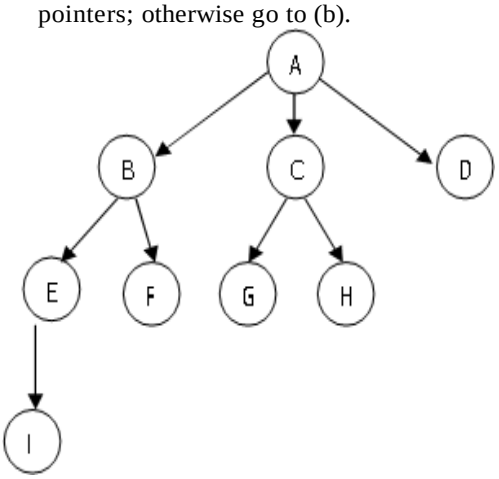


Figure 1 DFS

DFS always explores the deepest node first. That is, one which is farthest down from the root of the tree. To prevent consideration of unacceptably ling paths, a depth bound is often employed to limit the depth of search. The DFS would explore the tree in Fig 1 in the order:

A – B – E – I – F – C – G – H – D

The depth-first -search algorithm is an instance of the graph search algorithm**.** Depth first search uses a LIFO queue. A LIFO queue means that the most recently generated node is chosen for expansion. This must be the deepest unexpanded node because it is one deeper than its parent which in turn was the deepest unexpanded node when it was selected.

The properties of depth first search depend strongly on whether the graph search or tree search version. The graph search version, which avoids repeated state and redundant paths is complete in finite state spaces because it will eventually expand every node.[14]

B. Breadth First Search

This always searches closest to the root node first, thereby visiting all nodes of a given length first before moving to any longer paths. It pushes uniformly into the search tree. The problem with Breadth first search is most effective when all paths to a goal node are of uniform depth. It is a bad idea when the branching factor (average number of offspring for each node) is larger or infinite.

The algorithm for breadth first search is as follows:

i. Put the start node on a list called OPEN
ii. If OPEN is empty, exit with failure, otherwise continue

iii. Remove the first node on OPEN put it on a list called CLOSED call this node.

iv. Expand node n, generating all of its successors, if there are no successors go immediately to (ii). Put the successor at the end of OPEN and provide pointers from these successors back to **n.**

v. If any of the successors are goal nodes, exit with the solution obtain by tracing back through the pointers otherwise go to (ii).

The problem with unbounded depth-first search is that we may get lost in an infinite branch while there could be another short branch leading to a solution. The problem with depth-bounds depth-first search is that it can be difficult to correctly estimate a good value for the bound. BFS is most effective when all paths to a goal node are of uniform depth. It is a bad idea when the branching factor (average number of offspring for each node) is large or infinite. BFS is also preferred over DFS if you are worried that there may be long paths that neither reach dead ends nor become complete paths.[2].BFS is simple strategy in

C. Bidirectional Search

To this point all search algorithms discussed (with the exception of means-end analysis and backtracking) have been based on forward reasoning. Searching backwards from goal a node to predecessors is relatively easy. Combining forward and backward reasoning in searching for solution is called bidirectional search [13]. The idea is to replace a single search graph, which is likely to grow exponentially, with two smaller graphs. .one starting from initial state and one starting from goal. The search is approximated to terminate when the two graphs intersect. This algorithm is guaranteed to find the shortest solution path through a general state-space graph. A building block. The DLS overcomes the drawback of unbounded depth of A. The depth limit (l) is predefined and solve the infinite path problem [8]

- Depth-First search but with depth limit l. i.e. nodes at depth l have no successors.
- No infinite-path problem
- If l = d then optimal failure
  - No solution with in the depth limit.
  - No solution at all if goal is deeper.

Depth limited search can be implemented as a simple modification to the general tree or graph search algorithm. Alternatively, it can be implemented recursive algorithm [14].

D. Iterative deepening search

IDS is problem of unbounded trees (DFS) got removed by depth limited search. But, it als o introduces an additional source of incomers .if the goal state is beyond the depth limit their will be no solution[10]

- A general strategy to find best depth limit l.

which the node is expanded first, then all the successors of the root node are expanded next. then their successors, and so on, in general all the nodes are expanded at given depth in the search tree before any nodes at the next level are expanded.[14]

Breadth First Search is an instance of the general graph search algorithm. Which the shallowest unexpanded node is chosen for expansion .this is achieved very simply by using a FIFO queue for the frontiers. thus new nodes go to the back of the queue, and old nodes, which are shallower than the new nodes get expanded first.[14].This always searches closest to the root node first, thereby visiting all nodes of a given length first before moving to any longer paths. It pushes uniformly into the search tree. The problem with depth -first search is that it can be difficult to correctly estimate a good value for the bound. Such problems can be overcome by using breadth-first search where we explore (right-hand) siblings before children [12]

- Key idea: use Depth-limited search as subroutine, with increasing l . For d = 0 to ∞ do depth-limited-search to level d if it succeeds then return solution
- Complete & optimal: Goal is always found at depth d, the depth of the shallowest goal-node.

Like DFS its memory requirements are modest: O(bid) to be precise like BFS, it is complete when the branching factor is finite and optimal when the path cost is a non-decreasing function of the depth of the node. Iterative deepening search may seem wasteful because states are generated multiple times. It turns out this is not too costly. this reason is that in a search tree with the same branching factor at each level, most of the nodes are in the bottom level, so it does not matter much that the upper levels are generated multiple times.[14]

E. Uniform Cost Search

When all step costs are equal , breath first search is optimal because it always expands the shallowest unexpanded node. By a simple extension, we can find an algorithm that is optimal with any step -cost function. Instead of expanding the shallowest node, uniform cost search expands the node n with the lowest path cost g(n).this is done by storing the frontier as a priority queue ordered by g.[14]

In computer science, uniform-cost search (UCS) is a tree search algorithm used for traversing or searching a weighted tree, tree structure, or graph.[10]

- Optimal where all step are equal.
- Expands the node n with lowest path cost.
- It does not care about the number of steps a path has but only about their total cost
- It is guided by path cost rather than depths.

APPLICATIONS OF SEARCH IN AI PROBLEM SOLVING

1. Many real-time search methods can use heuristic knowledge to guide planning, be interrupted at any state and resume execution at a different state and improve their plan execution time, as they solve similar planning tasks, until their execution is optimal.
2. In computer science, AI researchers have created many tools to solve the most difficult problems in computer science which are search techniques. These were adopted by mainstream computers science and are no longer considered a part of AI.
3. In finance, Banks use search techniques to organize operations, invest in stocks and manage properties. In August 2001, Robots beat humans in a simulated financial trading competition. Financial institutions have long used artificial neural network systems to detect charges or claims outside the norm, flagging these for human investigation.
4. In medicine, search techniques saw application in organizing bed schedules, make a staff rotation, and provide medical information. Artificial neural networks are used as clinical decision support systems for medical diagnosis.
5. In telecommunications, many telecoms companies make use of heuristic search in the management of their workforces, for example BT group has deployed heuristic search in a scheduling application that provides the work schedules of 20,000 engineers

IV. INFORMED SEARCH

This type of search uses domain knowledge. It generally uses a heuristic function that estimates how close a state is to the goal. This heuristic need not be perfect. This function is used to estimate the cost from a state to the closest goal [7].The basic informed search strategies are:

A. *Greedy search:-* Greedy search (best first search): It expands the node that appears to be closest to goal

B. A* search:- Minimize the total estimated solution cost, that includes cost of reaching a state and cost of reaching goal from that state.

C. *Iterative Deepening A* :-* Iterative deepening version of A*. Uses f-cost limit as depth bound

- Complete and optimal
- Modification of A* search
- Storage efficient practical
- Each iteration expands all nodes for current f-cost
- Same order of node expansion
- IDA* do not remember history

Idea:

- Performs limited-cost depth-first search for the current evaluation function limit – Keeps expanding nodes in the depth-first manner up to the evaluation function limit
- Progressively increases the evaluation function limit (instead of the depth limit)[9]

D. *SMA* Search* Simplified Memory Bound A* need to make space

- Version of A*
- Memory bounded (make use of all available memory to carry out search)

Properties:

- Utilize whatever memory is made available
- Avoids repeated states as far as its memory allow
- Complete if available memory is sufficient to store the shallowest solution path the problem, that your AI is going to solve, in the right way.
- Some of the real world examples, where they are being used are:

V. CONCLUSION

Almost all artificial intelligence (AI) programs can be said to be doing some form of problem-solving. Search is one of the central issues or goals in problem-solving systems. It becomes so whenever the system, through lack of knowledge, is faced with a choice from a number of alternatives, where each choice leads to the need to make further choices, and so on until problem is solved. The amount of search involved can be reduced if there is a method of estimating how effective an operator will be in moving the initial problem state towards a solution. This is where heuristics search methods becomes more effective in finding solution. There is a tension in AI between the investigation of general purpose methods that can be applied across domains and the discovery and exploitation of special knowledge, heuristics, tricks and shortcuts that can be applied in particular domains to improve performance. The big

challenge to researchers is the ability to manipulate the various problem-solving techniques, namely, search algorithms that have, been developed for a variety of conditions. We have presented in this paper a classification that is hoped to reduce this bottle-neck.

REFERENCES

[1] "What is Artificial Intelligence", Retrieved on 15.09.2011
[2] P. H. Winston, Artificial Intelligence. Reading, Massachusetts: Addison-Wesley,1984.
[3] Poole, A. Mackworth and R. Goebel, Randy, Computational Intelligence: A logical Approach New York: Oxford University Press,1998.
[4] P. McCorduck, Machines Who Think, (2nd ed) Natick, MA: A.K. Peters Ltd,2004.
[5] http://www.cs.edu/afs/cs.cmu.edu/academic/class/15381-s06/www/introcol.pdf
[6] http://www.exampleessays.com/essay_search/Charniak_McDermott.html
[7] S. J. Russell, P. Nerving, Artificial Intelligence; A Modern Approach (2ndEd). Upper Saddle River, New Jersey: Prentice Hall,2003.
[8] http://people.cspitt.edu
[9] http://citeseer.ist.psu.edu/viewdoc/summary?doi=10.1.191.288
[10] http://www.cs.ubbcluj.ro/-csatol/log_funk/prolog/slides/7-search.pdf
[11] http://en.wikipedia.org/wiki/bidirectional.search
[12]http://www.cin.ufpe.br pp no.81-95

HMM BASED POS TAGGER FOR PUNJABI

Simrat Walia
Departm ent of CSE
Baba Banda Singh Bahadur Engineering College
Fatehgarh Sahib
simrat.walia@bbsbec.ac.in

***Abstract*--Part of spe e ch tagging is to marking the words in the te xt according to its word classe s such as nouns, pronoun, ve rbs, adje ctive s, adve rbs e tc. It is most crucial and fundame ntal proce ss in various Natural Language Proce ssing (NLP) applications such as grammar che cking, Text to S pe e ch, information re trie val, parsi ng, se mantics and machine translation. This pape r pre se nts the de ve lopme nt of Part of spe e ch (PO S) tagge r for Punjabi by using trigram HMM mode l. The de ve lopme nt proce ss comprise s of two phase s te sting and training phase . The training proce ss include s annotation of training data, pre proce ssing of training data and probability calculation. This tagge r incorporate s a Laplace Law of probability for smoothing and handling of unknown words. This pape r also pre se nts a small tagse t of 30tags. This tagge r give s pe rformance of 83% along with tagging of unknown words.**

Indexed Terms: Best first search,Corpus, HMM, Laplace Law,POS Tagger, Trigrams

I. INTRODUCTION

Part of Speech (POS) Tagging is the process of assigning the words in a text as corresponding to a particu lar part of speech, based on grammatical rules and context. The tags identify the words as nouns, verbs, adjectives, adverbs etc. The POS tagging techniques are of three types :Rule based, Probabilisticand Transformation based. Rule based approach has handwritten rules based upon a word and its contextual information. A rule based POS tagging system consists oftwo stages. First stage assigns each word a list of potential POS tags by using a dictionary. Second stage uses hand written rules to find out single POS tag for each word. Punjabi language previously has its rule based POS tagger based on handwritten rules. This tagger was developed at Punjabi University, Patiala for Punjabi grammar checking system[6].

The probabilistic taggers are of two types: supervised and unsupervised. The supervised probabilistic approaches are based on the information from the annotated corpus. The annotated corpus provides the frequency ofthe sequence of tags which will help in disambiguating words in sentence. The problem arises for languages that lacks in available training data. Brants presented an HMM based POS tagger that is called TnTTagger(trigrams)[2]. In TnT second order Markov Model is used, means trigrams were used. In this POS tagger lexical and contextual probability is maximized. The average accuracy of POS tagger is between 96% and 97%. Church gives a part-of-speech tagging program which is based on probabilistic model of tagging [4]. Probability estimates were obtained by training on the Tagged Bro wn Corpus, a corpus of approximately 1,000,000 words with part of speech tags assigned manuallyin many years.The transformation based approach combines rule based and statistical approaches. It generates the most likely tag based on a training corpus and then applies certain set of rules to disambiguate. It saves any new rules that it has learnt in the process, for future use. Brill use this tagging technique and achieves 90% accuracy [3].The present study was concerned with development of POS tagger for Punjabi using HMM model. The paper is organized as section II describes tagset design and data preparation. The section III concerned with design of POS tagger and IV with handling of unknown words. The implementation part is described in section V. Finally VI and VII sections present results and concluding remarks respectively.

II. POS TAGSET DESIGN AND DATA PREPRATION

Punjabi language spoken by inhabitants of the historical Punjab region, now split between India and Pakistan. Punjabi is written by two scripts by Gurmukhi and Shahmukhi. But, forthis POS tagger Gurmukhi script is considered. Punjabi language shares most of its vocabulary from Hindi, Urdu, Sanskrit, Brij, Sindi and Persian. Punjabi has both inflectional and derivational morphology [6]. In previous rule based system for Punjabi there was a fine grained tagset of 630 tags. But training with that tagset needs a very large sized annotated corpus. So due to limitation of corpus size a small tagset was used. In this tagset all the word categories defined for Punjabi tagset were taken but the inflections were removed

The accuracy of a tagger also depends on the corpus. A corpus was taken from Ajit Newspaper. That was initially tagged with previous rule based tagger. Then output of that preprocessed like inflections were removed and errors like double space, unwanted symbols and noise was removed manually. The available corpus was based on six domains i.e. games, news, finance, culture entertainment, Gurbani and literature. Thenentire text is converted to Unicode font. Then,an automated module was created that has been generated twoXML files for testing and training. This module placedthe lines of corpus in train.XML but every tenth line in test.XML, so that both files contained sentences from each domain.

III. DESIGN OF POS TAGGER

This POS Tagger accepts input Gurmukhi sentences in Unicode and store output in XML file. The proposed system has three components : initial tagging, unknown word handling and disambiguation. The major algorithms employed for this study are Hidden Markov Model [14] and Laplace Law to deal with unknown words [18].

Hidden Markov model is used to estimate the best sequence of tags for a sentence. It uses a tagged corpus to estimate the frequency of the occurrence of a tag with a word. It is called Hidden as the actual sequence of states i.e. tag s generated for a sentence is unknown. Hidden Markov model has five parameters [14]:

- Total number of states in the system is represented by N. In part of speech tagging, N is the total number of tags used by the system.
- Total number of output symbols and is represented by M. For part of speech tagging, M will be the number of words in the lexicon of the system.
- Probability of moving from state i to state j and is represented by aij. It is called transition probability of the states. For part of speech tagging, state transition probability will be the probability of moving from tag i to tag j, in other words probability that tag j will follow tag i. This probability is estimated from the corpus.
- Observation probability bj(k) will be the probability of having symbol k on state j. For part of speech tagging, it will be the probability of word having tag j.
- Initial state distribution πj is the probability that model will start in state i. For part of speech tagging, this is the probability that the sentence will start with tag i.

In brief Basic HMM part of speech tagging appeared as :

So the first part is known as emission probability and later part is transition probability. Both are estimated from a trained corpus . Table 1 shows formulas deduced from HMM model used in this POS tagging.

Table 1: Trigram Formulas

Sr. No.	Type	Formula
1.	Unigram	P(t3)=f(t3)/N
2.	Bigram	P(t3|t2)=f(t2,t3)/f(t 2)
3.	Trigram	P(t3|t2,t1)=f(t3,t2,t 1)/f(t1,t 2)
4.	Lexical	P(w3|t3)=f(w3,t3)/f(t3)

N is the total number of tokens (words) in the training corpus. Here f(ti) is the frequency of tag ti. Here F(t3) denotes frequency of t3 in the corpus of size N. f(t2,t3) is the frequency of t3 preceded by t2.All the unigram, trigram, bigram and lexical probabilities are stored in XML files during training phase.

IV. UNKNOWN WORDS HANDLING

The tagging of unknown words was derived from smoothing technique based Laplace's law that consists of adding one to all the observations [18]. That means that all the unseen events will have their probability computed as if they had appeared once in the training data. Then Laplace smoothing was used when estimating the transition probabilities and emission probabilities [12].The table 2 gives transition probabilities and emission probabilities:

Table 2: Trigram Formula after smoothing

Unigram	P(t3)=(f(t3)+1)/(N+QA)
Bigram	P(t3|t2)=f(t2,t3)+1/f(t 2)+QB
Trigram	P(t3|t2,t1)=f(t3,t2,t 1)+1/f(t1,t2)+QC
Lexical	P(w3|t3)=f(w3,t3)+1/f(t 3)+V

The accepting states QA are the total POS-tags, QB is total number of tag pairs and QC is total tag triplets. Where, V is total different words in corpus. The system is trained by using these new formulas and best tags sequence for sentence is chosen with best first search algorithm.

V. TAGGER IMPLEMENTATION

This POS tagger was developed in C# using .NET2005 package and all data is stored in XML format. Application of part of speech tagger takes two inputs i.e. a given input text and trained corpus. Input text is observed sentence by sentence. Tagger creates the annotated output of each sentence. Then the working of tagger is as follows:

- Take the input in Gurmukhi script.
- Divide input on the basis of sentence marker.
- Take words from sentence.
- Assign its potential tags to an ambiguous word .
- Sort out the ambiguity.
- Generate the output.

Part of speech tagger takes information from the training corpus i.e. unigram.XML, bigram.XML, trigram.XML and lexical.XML. In these files emission and transition probabilities are stored. So training is a crucial step and algorithm of training database is as follows:

- Take annotated text from a text file.
- Calculate total counts of each word tag pair i.e. total number of occurrences of each word w with tag t.
- Calculate total counts of each tag pair i.e. total occurrences of each tag pair "ti-1 ti" having previous tag t i-1.
- Calculate total counts of each tag triplet i.e. total occurrences of each tag ti having previous tag pair "ti-2 ti-1".

- Calculate total counts of each tag i.e. total occurrences of each tag t i-1, tag pair "ti-2 ti-1".
- Apply smoothing.
- Calculate probabilities .
- Save the probabilities in separate files.

Probability calculation was done using formulas given in table 3. Tagger creates the annotated output of each sentence. Algorithm of POS tagger testing phase is as follows:

- Read input from textbox in the user interface
- Divide input in tokens
- Continue until input ends
- Take a sentence
- Repeat until sentence end
- Take a word from sentence
- Load databases
- Assign its potential tags from lexicon or assign potential tags for an unknown word
- Assign word tag probabilities to the pair
- If the word has single tag then put that tag in the final taglist else repeat the following two steps until all the candidates for tag are checked.
- Assign tag bigram probabilities and trigram probabilities
- Find out the maximum probability with help of best first algorithm which is the combination of all the probabilities
- Save the output sentence by sentence
- Write output in an array known as final taglist and then show all the output and save the output in test.XML file

VI. RESULTS

The performance of POS tagger is measured in terms of number of words tagged accurately. So the accuracy of the system is defined as:

$$\text{Accuracy} = \frac{\text{Number of words correctly tagged}}{\text{Total number of words}}$$

In the present work accuracy was observed to be 83% for the total words including all the words which were present in the database and were unknown to the database. For training a database of 18000 words and for testing a data base of 2000 words were taken.

VII. CONCLUSION AND FUTURESCOPE

In this study the Statistical Part of speech tagger bas ed on HMM model is presented and it gave the accuracy of 83%.. During analysis some of observations were made:

Impact of size of training corpus: The large sized training corpus can improve accuracy because degree of similarity among statistical data will increase. But major barrier is large time and effort required for manual annotation of documents.

Impact of size of tagset: When tagset becomes large then for same accuracy large annotated corpus is required.

Impact of data sparseness: The data sparseness means a sentence has more than two unknown words. So it reduces the accuracy significantly because accuracy is based on statistical analysis of data present in the corpus.

The main problem which arose in present work was scarcity of resources such as tagged corpus due to that the accuracy of system declined. In future the main advancements would be in terms of fine grained tagset and the size of corpus. Another concern is portability of tagger so that it is able to support the different languages

REFERENCES

[1] Bharati, Akshar, Amba P. Kulkarni, and VineetChaitanya, "Challenges in Developing Word Analyzers for Indian Languages" Paper presented at the Workshop on Morphology. CIEFL, Hyderabad, India 1996.

[2] Brants and Thorsten "TNT – A Statistical Part-of-Speech Tagger", Computational Linguistics,pages 136-143 Saraland University 2001.

[3] Brill Eric, "A Simple Rule-based Part-of-Speech Tagger", In Proceedings of the Workshop on Speech and Natural Language, pages 112-116, San Mateo, CA 1992.

[4] Church and Kenneth Ward,"A Stochastic Parts Program and Noun Phrase Parser for Unrestricted Text", In Proceedings of Second Conference on Applied Natural Language Processing, pages 136-143, Austin, Texas 1998.

[5] Dalal, Aniket, Kumar Nagaraj, Uma Sawant, SandeepShelke, and Pushpak Bhattacharyya, "Building Feature Rich POS Tagger for Morphologically Rich Languages: Experiences in Hindi" In Proceedings of ICON, Hyderabad, India2007.

[6] Gill, MandeepSingh, "Development of a Punjabi grammarchecker", Ph.D. Thesis, Department of Computer Science Punjabi University, Patiala 2008.

[7] Hardie Andrew, "Developing a tagset for automated part-of-speech tagging in Urdu", Department of Linguistics and Modern English Language, University of Lancaster.

[8] Haulrich,Martin,"Different Approaches to Unknown Words in a Hidden Markov Model Part-of-Speech Tagger" Stockholm.

[9] Jiayun Han, "Building an Efficient, Scalable, and Trainable Probability-And-Rule-Based Part of Speech Tagger of High Accuracy", Master's Thesis, University Of Georgia, The United States2007.

[10] Khoja, Shereen., "APT: Arabic Part-of-Speech Tagger" Computing Department, Lancaster University, UK2001.
[11] http://www.archimedes.fas.harvard.edu/mdh/arabic/NAACL.pdf
[12] Megyesi, Beata, "Brill's POS tagger with extended lexical templates for Hungarian", Department of Linguistics University of Stockholm 1996.
[13] Padro, Muntsa and Padro, Lluıs. "Developing Competitive HMM POS Taggers Using Small Training Corpora", TALP Research Center, Polytechnic University of Catalonia.
[14] PradeepVarma, Rakesh M., RatnaSanyal, "HMM-based Language-independent POS Tagger", Indian Institute of Information Technology Allahabad, India.
[15] Sajjad, Hassan "Statistical Part-of-Speech Tagger for Urdu" , Master's Thesis, Department of Computer Science, National University of Computer and Emerging Sciences, Lahore, Pakistan 2007.
[16] Shamsi, Fatma Al and Ahmed Guessoum "A Hidden Markov Model based POS Tagger for Arabic" , In Proceedings of the Eighth International Conference on the Statistical Analysis of Textual Data (JADT 2006), pages31-42, University of Besançon,France2006.
[17] Thede andScott M., "Predicting Part-of-Speech Information about Unknown Word using Statistical Methods", Purdue University West Lafayette, pages 1505-1507, IN 47907.
[18] http://www.ldc.upenn.edu/acl/P/P99/P99-1023.pdf
[19] Tseng Huihsin, Daniel Jurafsky Christopher Manning "Morphological features help POS tagging of unknown words across language varieties", Dept. of Linguistics Stanford University Stanford.
[20] WaqasAnwar,Xuan Wang Luli and Xiaolong Wang "Hidden Markov Model Based Part of Speech Tagger for Urdu", Information Technology journal Asian Network for Scientific information, pages 1090-1098 China.

Android Application Development to Detect threads using Application programs Interface

Harpreet Singh,
Department ofCSE
Baba Banda Singh Bahadur Engineering College
Fatehgarh Sahib, Punjab.

Sandeep Kaur Dhanda
Department of CSE
Baba Banda Singh Bahadur Engineering College
Fatehgarh Sahib, Punjab.

***Abstract*—Android is a mobile ope rating syste m (O S) and it is ope n source .It has own S oftware De ve lopme nt kit (S DK). It is modifie d ve rsion of the Linux ke rne l .W ith the rapid incre ase de ve lopme nt in Android applications in last ye ars, the Android applications' se curity is incre asing more atte ntion to it. The Android malware de te ction ca n be cate goriz e d into two type s: be havior-base d malware de te ction and code -base d malware de te ction. In this pape r, we pre se nt a be havior of Application De ve lopme nt to de te ct thre ads using Application programs Inte rface . In the proce ss, the API are ge ne rate d and is in form of obje cts and arrays. API are basically work at back e nd to se rve the data that is re que ste d. The API is usually packe d in an S oftware de ve lopme nt kit(S DK) .The source code (se cre t code) in be gin and malware will be compare and cate goriz e d into APIand manage r classe s afte r that the most fre que nt API and manage r class use d in malware will be de te cted.**

Keywords: API, Android App, Malware Detection, JSON.

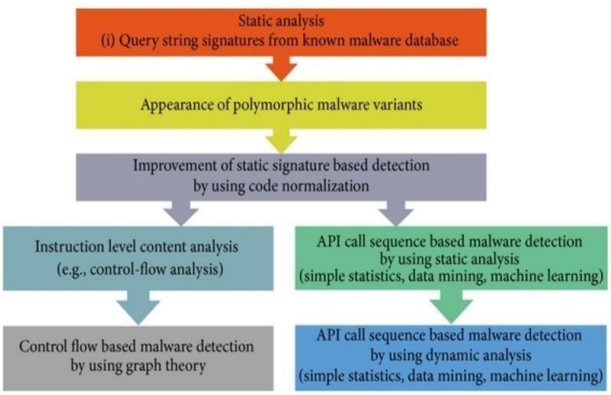


Figure. 1. Steps in malware analysis

I. INTRODUCTION

Android is the mobile operating systemdeveloped by Google. It is designed for touchscreen mobile devices such as smartphones, tablets, Android Connected Cars, and Wear OS and many more. Initially it was developed by Android Inc., which Google take in 2005.the first commercial Android device launched in September 2008. The interface of Android's interface is based on direct manipulation, using touch inputs basically screen that corresponds to real-world actions, like swipe, tap, pinch, and reverse pinching s ecurity threats.

Research from Banking-based mobile malware is rise, as hackers steal the user information who prefer conducting all of their business including money transfers and bill payments from their mobile devices. Malware are looking for ways to exploit their data using communication as a way to deliver malware. Even if users didn't open the text, the malware could insert into pc to deploy, allowing hackers to access authorized information to your mobile device. Antivirus like Trend Micro lists premium abuse the most common type of Android malware, where text messages are sent from infected phones in form of videos, images, SMS to premium-rate telephone numbers .malware displays unwanted advertisements on device and sends personal information to unauthorized third parties. Google maintains that dangerous malware is actually extremely rare and a survey is conduct by F-Secure and they showed that only 10 percent of Android malware is reported had come from the Google Play store.

Moreover, the number of API calls are used by Android applications is huge, with more than 250 high-risk APIs. Malware in general attempts to hide from the defensive mechanisms in systemin order to performits performtasks. Many of available API call techniques fail to detect malware. Some of the techniques that focus on extracting APIs that are frequently observed in malware. They monitor APIs(Application Program Interface) that are called and calculate frequency and total number of events that certain API function called. In given paper Android malicious detection method is proposed. This method is mainly categorized into two parts: training phase and detection phase. For the training part, the main tas k includes two aspects. The mutual information model is utilized to form a new collection of 20- dimensional APIs ranked by sensitive level. The final result is generated by displaying the result of two classifiers which are decision tree and k-Nearest Neighbor (kNN) classifier. Techniques are based on dynamic analysis that execute malware and trace its behaviors. Two approaches in dynamic analysis that control flow analysis and API call analysis.

II. SYSTEM DESIGN AND IMPLEMENTATION:- The Steps in malware analysis flow chart is shown in Figure. The method composes of two parts, training phase and detection Phase. Each has two sub-sections namely: Feature Generation and Ensemble Learning Model.

A. Feature Generation

when android application is run and installed an APK file is built, which is a variant of JAR file format . The necessary feature can be obtained after decompiling the class. dex file, which contains all of the Dalvik byte codes, the API calls. A regular expression pattern is then applied to get all the methods fromthe DEX file.

1. XML(extensive markup language)

XML stands for Extensible Markup Language. XML is a markup language same as HTMLused to describe data. We must define our own Tags in Xml as itself is well by machine. Also, it is scalable and simple to develop. In Android we use xml for designing our layouts because xml is lightweight language so it doesn't make our layout heavy. It has two layout that we use the most that is Linear layout and Relative layout, both has its own property. This xml file is used to replace the Hard-coded strings with a single string. We define almost all strings in this xml file and then access them in our application from this file. This use the reusability of the code. This xml is used to define many different styles that looks for the UI (User Interface) of application.

2. Manifest file

The AndroidManifest.xml file contains all information of your package, some components of the application such as activities, services, broadcast receivers, content providers etc. where to start the activity ,when it will run etc.

1. It performs some other tasks like:
2. It is responsible to protect the application providing the permissions.
3. It also use the android api that the application is going to use.
4. It lists the instrumentation classes.

3. Android virtual device (AVD)

A AVD (Android Virtual Device) is virtual Android dev ice, used for operating the Android emulator from the Android SDK. Android Virtual Device (AVD) manager enables us to launch virtual android emulators in our PC and run the app in the emulator, and at the same time we can track and debug each app activity from the Log cat in our IDE.

4. SQLite Database

SQLite is a software library that implements a self- contained, server less, zero-configuration, transactional SQL database engine. SQLite is the most widely deployed SQL database engine in the world. The source code for SQLite is in the public domain. SQLite is an in-process library that implements a self-contained, zero-configuration, transactional SQL database engine. It is a database, which is zero-configured, which means like other databases you do not need to configure it in your system.

A. Ensemble Learning Model

Ensemble methods use multiple learning algorithms to obtain better predictive performance than could be obtained from any of the constituent learning algorithms alone. Unlike a statistical ensemble which is usually infinite, a machine learning ensemble consists of only a concrete finite set of alternative

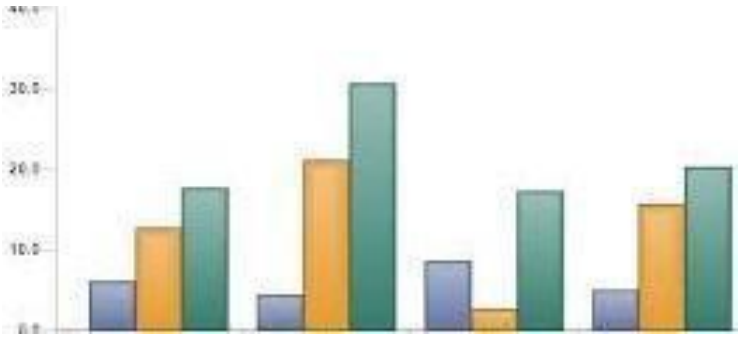


Fig. 2. Overall performance

Alternative models, but typically allows for much more flexible structure to exist among those alternatives.

There is a certain degree of deficiencies in the attribute division of the single classification algorithm. Therefore ensemble learning methods are used in this work. In this study, we use an ensemble learning model based on kNN and decision tree to classify attributes. In the ensemble learning model, the bagging method constructs multiple instances of a black box classifier on a random subset of the original training set.

1. Voting Combination Strategy

In voting methods, tactical voting occurs, in elections with more than two candidates, when a voter supports another as candidate in order to prevent an undesirable outcome. By combining multiple learners, it is possible to learn a better approximation because the corresponding hypothesis space has been expanded. In this we use two sets of two dimensional array(x, yi) to store their detection results. To detect unknown application, the threshold frequency of the detection model is set to have the value of 0.5. Because the probability of an unknown application being classified as a malicious application is the same as that of a benign application.

II. EXPERIM ENTA L RESULTS AND ANALYSIS

A. Experimental environment

In this experiment, 520 normal applications and 527 maluscious applications are selected as experimental data sets. Among them, 516 normal Android applications are collected from third-party application market and Google Android Market using the original application and some application require payment to purchase. The Requirement forenvironment is: operating system (os) of Windows 10, processor is Intel Core i5, 4GB of memory, Python 3.7.2 scripting languages.

A. Results and analysis

Overall performance: According to the sensitive Application programming interface API calls proposed in this paper, a set of further experimental is done comparison is made on the impact of the detection results. The results are shown in Fig 2.

B. CONCLUSION

Based on Application programs Interface calls supported by data from experiments, we set a new Android malware detec-

tion method. Through the extraction of the Android application API calls, the feature vector are generated. After combining with the mutual information model, a set of sensitive API calls is generated by machine learning algorithms. Unknown Android applications can be effectively classified by an ensemble learning model based on kNN classifier and decision tree classifier. The experimental results show that our method can achieve a good effect on precision, accuracy and true positive rate. In the future research, we will conduct more experiments to reduce the false positive rate to improve the ability of detecting unknown Android applications.

REFERENCES

[1] Chunlei Zhao, Wenbai Zheng, Liangyi Gong, Mengzhe Zhang, Chundong Wang "Quick and Accurate Android Mal- ware Detection"2018 IEEE International Conference on Smart Internet of T hing-2018.

[2] Bin Peng,Jinming Yue, Chen T ianzhou "T he Android Application Development College Challenge"2018.

[3] G.Ganesh Sundarkumar, Vadlamani Ravi, Ifeoma Nwogu, VenuGovindaraju "Malware Detection via API calls, T opic Models and Machine Learning"2015 IEEE International Conference on AutomationScienceand Engineering (CASE) Aug 24-28, 2015. Gothenburg, Sweden.

[4] T . Kasama, K. Yoshioka, D. Inoue , and T. Matsumoto, Malware detection method by catching their random behavior in multiple executions, In: Applications and the Internet (SAINT),2012IEEE/IPSJ12th International Symposium on, 2012, pp. 262 -266.

[5] K. S. Reddy, and A. K. Pujari, —N-gram analysis for computer virus detection, Journal in Computer Virology, 2(3), 2006, pp. 231-239.Altaher, S. Ramadass, and A. Ali, — Computer virus detection using featuresranking and machine learning, Journal of Applied Sciences Research, 7(9), 2011.

[6] Youngjoon Ki, Eunjin Kim, and Huy Kang Kim"A Novel Approach to Detect Malware Based on API Call Sequence Analysis"Hindawi Publishing Corporation Interna- tional Journal of Distributed SensorNetworksVolume 2015.

[7] H. Yin and D. Song, Automatic Malware Analysis: An Emulator Based Approach, Springer Science+Business Media LLC, 2012. wikipedia,https://en.wikipedia.org.

[8] Android Developers,https://developer.android.com.