ISSN : 2456-3307

TECHNO SCIENCE ACADEMY
THE INTERNATIONAL OPEN ACCESS PUBLISHER

IJSRCSEIT

The International Conference on Research Perspectives : IoT in Hybrid Grid Integrated Renewable Energy Sources

Organised by

Department of Electrical and Electronics Engineering,
New Horizon College of Engineering,
Bellandur Main Rd, Kaverappa Layout,
Bengaluru, Karnataka, India

Sponsored By

INTERNATIONAL JOURNAL OF SCIENTIFIC RESEARCH IN COMPUTER SCIENCE, ENGINEERING AND INFORMATION TECHNOLOGY

Volume 8, Issue 1, March-April-2021

Email: editor@ijsrcseit.com

The International Conference on Research Perspectives IoT in Hybrid Grid Integrated Renewable Energy Sources ICRPHGIRES – 2021

24th to 26th March, 2021

Sponsored By

In Association with

International Journal of Scientific Research in Computer Science, Engineering and Information Technology

ISSN : 2456-3307

Volume 8, Issue 1, March-April-2021

International Peer Reviewed, Open Access Journal

Organised by:

Department of Electrical and Electronics Engineering, New Horizon College of Engineering, Bellandur Main Rd, Kaverappa Layout,Near Marathahalli, Bengaluru-560103 Karnataka, India

https://newhorizonindia.edu

Published By

Technoscience Academy

(The International Open Access Publisher)

Email: info@technoscienceacademy.com

website: www.technoscienceacademy.com

Chief Patron

Dr. Mohan Manghnani, Chairman, New Horizon Educational Institution

Patrons

Dr.Karisiddappa, Vice Chancellor, Visvesvaraya Technological University, Belagavi

Dr.A.S.Deshpande, Registrar, Visvesvaraya Technological University, Belagavi

Dr. Manjunatha, Principal, New Horizon College of Engineering, Bengaluru

International Advisory Committee

- Dr. Hui Helen Li, Florida State University, USA
- Dr. Fred C. Lee, Virginia Polytechnic Institute and State University, USA
- Dr. Pedro Rodríguez, Loyola University Andalusia, USA
- Dr. John Shen, Illinois Institute of Technology, USA
- Dr. Krishna Shenai, Northwestern University, USA
- Dr. Taylor J, University of Toronto, Canada
- Dr. Yuriy V. Sentyabryov, Saint Petersburg State University, Russia
- Dr. Jian Sun, Rensselaer Polytechnic Institute, USA
- Dr. LalithaSankar, Arizona State University, USA
- Dr. Alexandra von Meier, University of California, USA
- Dr. Chester Frost, Stanford University, USA
- Dr. P. Sanjeevikumar, Aalborg University, Denmark
- Dr. Chung Shu hung Henry, City University of Hong Kong, Hong Kong
- Dr. Robert S. Balog, Texas A&M University, Qatar
- Dr. Udaya K. Madawala, University of Auckland, New Zealand
- Dr. Vassilios G. Agelidis, Technical University of Denmark, Denmark
- Dr. Petar J. Grbovic, University of Innsbruck, Austria
- Dr. Tobias Geyer, ABB Corporate Research, Switzerland, and Stellenbosch University, South Africa
- Dr. Paolo Mattavelli, University of Padova, Italy
- Dr. Ashwin M. Khambadkone, National University of Singapore, Singapore
- Dr. Ts. Chockalingam Aravind Vaithilingam,Taylor's University, Malaysia
- Dr. Huang-Jen Chiu, National Taiwan University of Science and Technology, Taiwan
- Dr. Muhammad Irfan, Najran University Saudi Arabia
- Dr. Ramesh Bansal, University of Sharjah, United Arab Emirates

National Advisory Committee

- Padmashri. Prof. R. M. Vasagam, Former Project Director, "APPLE" India's First Geo-Stationary Communication Satellite, Former VC, Anna University and Dr.M.G.R.University
- Dr. D. P. Kothari, Chairman, Board of Governors, THDC Institute of Hydropower Engineering and Technology, Tehri, Uttarakhand & Former Director i/c IIT Delhi, Former Principal VRCE, Nagpur, Former VC, VIT Vellore.
- Dr. L. V. Muralikrishna Reddy, President, Indian Technology Congress Association and University Space Engineering Consortium India
- Dr. Sri Niwas Singh, Professor, Department of Electrical Engineering, Indian Institute of Technology, Kanpur
- Dr. V. N. Mani, Principal Scientist-F. Centre for Materials for Electronics Technology(C-MET), Hyderabad
- Dr. Debapriya Das, Professor, Department of Electrical Engineering, Indian Institute of Technology, Kharagpur
- Dr. J.Ramkumar, Professor, Department of Mechanical Engineering, Indian Institute of Technology, Kanpur
- Dr. Anup Kumar Panda, Professor, Department of Electrical Engineering, National Institute of Technology,Rourkela
- Dr. Ashok S, Professor, Department of Electrical Engineering, National Institute of Technology, Calicut
- Dr. Jyoti Prakash Mishra, Associate Professor& Head, Department of Electrical Engineering, National Institute of Technology, Silchar
- Dr. U. Shripathi Acharya, Professor and Dean (Research and Consultancy), Department of Electronics and Communication Engineering, National Institute of Technology, Surathkal
- Dr. Nidul Sinha, Professor, Department of Electrical Engineering, National Institute of Technology, Silchar
- Dr. Binoy Krishna Roy, Professor, Department of Electrical Engineering, National Institute of Technology, Silchar
- Dr. Hitesh Dutt Mathur, Head, Department of Electrical and Electronics Engineering, Birla Institute of Technology & Science, Pilani
- Dr. P N Tekwani, Professor, Department of Electrical Engineering, NirmaInstitute of Technology, Ahmedabad
- Dr. B. V. Ravishankar, Professor & Principal, B.M.S. College of Engineering, Bengaluru
- Dr. C.P.S. Prakash, Professor & Principal, Dayananda Sagar College of Engineering, Bengaluru

Institution Advisory Committee

Ms. Malathi Madhusudhan, Sr.Executive Director, Accounts & Finance, NHEI

Mr.Surya Prakash, Registrar, NHEI

Prof. Gurucharan Singh, Executive Director, Training & Placements, NHEI

Dr. Girija Srinivasalu, Director, QASDC, NHEI

Institution Technical Committee

Dr. M. S. Ganesh Prasad, Dean & Head- Mechanical Engineering, NHCE

Dr. K. Gopalakrishnan, Dean - Research & Development, NHCE

Conference Organizing Chair

Dr. M. S Ganesh Prasad, Professor & Dean – Mechanical Engineering, NHCE

Publication Chair

Dr. K. Gopalakrishnan, Dean -Research & Development, NHCE

Conference Organizing Secretaries

Dr. M. Mahesh, Professor & Head - Department of EEE, NHCE

Dr. Sanjeev Sharma, Professor & Head - Department of ECE, NHCE

Steering Committee

Dr. Sridhar Kurse,Professor & Head - Automobile Engineering, NHCE

Dr. P. S Niranjan, Professor & Head - Civil Engineering, NHCE

Dr. B. Rajalakshmi, Professor & Head - Computer Science and Engineering, NHCE

Dr. R. J. Anandhi, Professor & Head - Information Science Engineering, NHCE

Institution Organizing Committee

Dr. Sheelan Misra, Professor & Head - Master of Business Administration, NHCE

Dr. Asha, Professor & Head-Master of Computer Applications, NHCE

Dr. Revathi, Professor & Head-Physics, NHCE

Dr. Anusuya Devi, Professor & Head-Chemistry, NHCE

Dr. Srinivasa.G, Professor & Head - Maths,NHCE

Dr. Sowmya, Professor & Head-Life Skills & Life Long Learning Centre, NHCE

Dr. Anitha Rai, Head-Library & Information Center, NHCE

Dr. Vijilius H Raj,Controller of Examination, NHCE

Ms. Manjula, Director -Human Resources, NHEI

Ms. Aruna, Director - Admissions, NHCE

Mr. Adarsh J Navale, Head-Branding & Marketing, NHEI

Dr. Reena Jain, Chief Counselor -Counseling, NHCE

Mr. Krishna Prakash, Head-Head Systems & Networks, NHCE

About College

New Horizon College of Engineering is an Autonomous college affiliated to Visvesvaraya Technological University (VTU), approved by the All India Council for Technical Education (AICTE) & University Grants Commission(UGC). It is accredited by NAAC with 'A' grade & National Board of Accreditation (NBA). It is one of the top Engineering Colleges in India as per NIRF rankings, ARIIA – 2020 and an ISO 9001:2008 certified Institution. New Horizon college of Engineering is located in the heart of the IT capital of India, Bangalore. The college campus is situated in the IT corridor of Bangalore surrounded by MNCs and IT giants such as Intel, Accenture, Capegemini, ARM, Symphony, Wipro, Nokia, JP Morgan and Cisco to name a few.

NHCE has a scenic and serene campus that provides an environment which is conducive for personal and intellectual growth. The infrastructure acts as a facilitator for the effective delivery of the curriculum. NHCE boasts of state -of-the -art facilities for its students. The students are given utmost encouragement in their areas of interest by providing hi-tech facilities backed by faculty support. The institute places highest priority on innovative programs of instructions that include both traditional class room theory and professional skills training. There is a strong impetus on overall personality development of the students with emphasis on soft skills. Students are supported through mentoring and counseling systems. The management offers scholarships to meritorious students. At NHCE, from the moment a student walks into the campus, he/she is well guided to know his/her strengths and choose an area of functional specialization. This enables students to concentrate their efforts and energies to gain the competitive edge. NHCE has a unique distinction of achieving 100% admissions in all its courses year after year.

NHCE has Centre of Excellence with reputed industries like Adobe, HPE, Vmware, Schneider Electric, SAP, Quest Global, CISCO.

Conference Preamble

ICRPHGIRES-2021 shares an insight into the recent research and cutting-edge technologies in" **Electrical, Electronics and Energy Conversion"**, which gains immense interest with the

exuberant presence of talented researchers, practicing engineers, and other professionals. ICRPHGIRES-2021 is the All India Council for Technical Education (AICTE) sponsored and emerging technical conference and exposition for energy conversion solutions. ICRPHGIRES-2021 includes peer-reviewed technical presentations, interactive sessions, and the exposition. Exhibitors from various industries showcase state-of-the-art technologies, products, and solutions, creating a highly interactive networking environment.

The International Conference on Research Perspectives: IoT in Hybrid Grid Integrated Renewable Energy Sources (ICRPHGIRES-2021) focus on Internet of Things (IoT) applications enabled for smart grids and smart environments, such as smart cities, smart homes, smart metering, and energy management infrastructures to investigate the development of the EI based IoT applications. These applications are promising key areas of the EI concept, since the IoT is considered one of the most important driving factors of the EI and also discuss the challenges, open issues, and future research opportunities for the EI concept based on IoT applications and addressed some important research areas.

ICRPHGIRES-2021 aims to provide an environment where the authors and participants can establish research relations and collaborations with various eminent academicians, research fellows, scientists from India and abroad.

Call for Papers

The ICRPHGIRES-2021 invites full length original research contributions not submitted to any other journal/conferences. We invite technical papers on, but not limited to, following domains.

Track 1

Smart Grid Architectures (EEE)

- National/International Energy Security
- Smart Grid Strategy and Planning
- Smart Grid Architectures and Models
- Smart Grid Security and Reliability Management
- Smart Grid Market Operations
- Smart Grid Standards

Track 2

Smart Grid Interconnection (EEE)

- Smart Grid Networking and Communications
- Integration of Distributed Resources
- Smart Grid Cyber and Physical Security
- Bi-Directional Energy Transfer
- Wide-Area Monitoring and Control

- Communication-Based Control and Protection
- Advanced Control of Grid-Connected Inverters for Power Quality and Renewable Energy Integration
- Energy Storage Systems for Resilience and Robustness Improvement in Smart Grid
- Portable and Smart Solution of Power Electronics for High Voltage Conversion with Renewable Energy and Grid Integration

Track 3

Smart Grid Implementation (ECE / EEE)

- Smart Grid Implementation and Field trials
- Energy Management Systems for Smart Grid
- Smart Grid Testing and Assessing Technologies
- Smart Sensors and Metering Infrastructure
- Smart Grid Demand Response
- Smart Grid Interactive Technologies
- Optimization Techniques in Smart Grid

Track 4

Power and Energy Engineering (ECE / EEE)

- 5G technologies
- Electronic systems & technologies
- Printable and planar electronics, 3-D electronics and miniaturization
- Electricity Networks for the Future
- Energy Storage and Distributed Energy Resources
- Power Generation, Transmission and Distribution
- Intelligent Monitoring and Outage Management
- Fault Monitoring and Predictive Maintenance
- Power Systems and Automation
- Energy Management, VPP (Virtual Power Plant) and ERAB (Energy Resource Aggregation Businesses) for RESSs
- Condition Monitoring of Power Apparatus
- High Voltage Engineering
- Power Semiconductor Devices, Renewable Energy
- Inverter and Converter Technology
- Advanced AC / DC Drives
- AI Techniques & Advanced Algorithms in Power Electronics
- Power Quality, HVAC, HVDC and FACTS
- Electrical Machines and Industrial Drives
- Smart Inverters using IOT

- VLSI
- Embedded / VLSI Systems

Track 5

Clean and Renewable Energy (All Sciences / MECH / AUTO)

- Photovoltaic Systems and Solar Energy Engineering
- Plug-in Vehicles and Low-C Transportation Alternatives
- Energy-Efficient Technologies
- Wind Energy Generation
- Renewable Energy Utilizations
- New Energy Materials and Devices
- Nano Technology in Energy Sector
- Optimization Techniques in Renewable Energy Systems
- PV systems and other Distributed Energy Resources in grids: challenges, monitoring, and power quality assessment
- Integration of EV's to Grid
- Material for energy Storage and Conversion
- Intelligent Transport systems
- Automation in Manufacturing of Construction Materials
- Smart Materials & Structures
- Smart Materials in Solar PV Cells and Challenges in Grid Integration of RES

Track 6

IoT (ISE / CSE / MCA)

- IoT network design and Cloud networks
- Network virtualization technologies
- Wireless networks for IoT and Cloud
- IoT-enabled home networks
- IoT and Smart cities
- IoT enabled Software Architectures and Middleware
- Software architecture for IoT and cloud
- Services provisioning and management
- Hybrid cloud infrastructure for IoT
- IoT based Application development
- IoT Service integration
- IoT based data acquisition, knowledge management and semantics
- Mathematical modeling for IoT

- Signal processing for IoT
- Sensor Systems for IoT
- IoT applications in Building Automation

Track 7

Smart Cities (CIVIL / MECH / AUTO / MBA)

- Smart Cities Solutions Management
- Smart Buildings and Home Automation
- Renewable Energy Systems in Smart Cities
- Hybrid Electric Vehicle / Autonomous Vehicle
- Advanced Safety Technologies
- Advanced 3D Modeling Techniques and Simulation
- Green Building Technologies
- Android / IOS Applications
- Net Zero Energy Building Technologies
- New Approaches in Lightings
- Materials for environmental remediation

Track 8

Automation (MECH / ECE / EEE)

- Control System, Distributed Control System in Smart Grid
- Intelligent Automation, Medical Imaging
- Multi-Objective Optimization, Multivariable Control in Smart Grid
- Non-Linear Control , PLC, SCADA based Systems in Smart Grid
- Process Control & Sensors in Smart Grid
- Robotics and Automation in Renewable Energy Systems
- Sensor Networks, Smart Sensors in Renewable Energy Systems
- Virtual Instrumentation & Biomedical Applications in Renewable Energy Systems
- Path Prediction and Estimation for Autonomous Vehicles

Track 9

Artificial Intelligence (CSE / ISE / MCA)

- AI Techniques in Renewable Energy System (RES)
- AI based Power Quality Improvement in RES
- Advanced Control of RES in Micro grid
- AI Based Battery Energy Management system
- AI Based Power system stability improvement
- AI based Renewable Energy Power management system.

- Machine Learning / Deep Learning for RES

Track 10

Management & Emerging Technologies (MBA / ECE / BSH / CSE / CIVIL)

- Business Sustainability
- Agile Management
- Green Management
- Sustainability challenges in turbulent environments
- Design thinking
- Education and technical skills to build smart cities
- Block chain and Artificial intelligence.
- General management related topics
- Satellite, space and airborne applications
- Remote sensing and GIS Applications
- Precast and Prefabricated Structures
- Data logging and Data Mining
- Big Data, Data Sciences, Cloud computing
- Machine Vision, Informatics & Intelligent Systems
- Bio-MEMS/NEMS
- MMIC & CMOS Circuits, Nano-fabrication & Computing
- Materials for Electronics, Integration of Electronic Systems
- Cryptography
- E-mobility

CONTENTS

The International Conference on Research Perspectives : IoT in Hybrid Grid Integrated Renewable Energy Sources
In association with International Journal of Scientific Research in Computer Science, Engineering and Information Technology
ISSN : 2456-3307 (www.ijsrcseit.com)

Survey on Real-Time Face Mask Detector

Pranav KN[1], Pramod Mahajan H[1], Praveen V[1], Sai Kiran[1], Dr. M Vinoth Kumar[2]

[1]BE Student, Department of Information Science and Engineering, Dayananda Sagar Academy of Technology and Management, Bangalore, Karnataka, India

[2]Associate Professor, Department of Information Science and Engineering, Dayananda Sagar Academy of Technology and Management, Bangalore, Karnataka, India

ABSTRACT

The COVID-19 pandemic has quickly influenced our everyday lives, disrupting commerce, and movements around the world. The wearing of a mask to cover the face has become a modern normal. In the near future, several public utility providers will be asking consumers to wear masks appropriately to make use of their facilities. Face mask identification has therefore become a key factor in helping global society. Using basic Machine Learning packages such as Tensor Flow, Keras, and OpenCV, this paper offers a simpler approach to achieving this objective. The proposed method correctly recognizes the face from the image and then identifies whether or not it has a mask on it. As a surveillance task officer, a face may also be identified along with a mask in motion.

Keywords: Face Mask Detection, Tensor Flow, Keras, and OpenCV.

I. INTRODUCTION

Globally, 78,604,532 confirmed cases of COVID-19 were reported to the WHO as of 5:03 pm CET, 26 December 2020, including 1,744,235 deaths. Individuals with COVID19 have reported a broad spectrum of symptoms, from mild signs to extreme illnesses. One of them is respiratory conditions, such as shortness of breath or breathing difficulties. Elderly people with lung disease may have significant COVID-19 disease complications because they tend to be at higher risk. Some common human coronaviruses, which infect the public around the world, are 229E, HKU1, OC43, and NL63. Viruses such as 2019-nCoV, SARS-CoV, and MERS-CoV infect animals and grow into human coronaviruses before affecting individuals. Persons with respiratory problems may expose contagious beads to anyone (who is in close contact with them). Surroundings of a stained person can induce touch transmission as the virus-bearing droplets can arrive on its adjacent surfaces.

It is necessary to wear a clinical mask to curb such respiratory viral ailments, like COVID-19. The public should be aware of whether the mask should be put on for source control or COVID-19 aversion. Potential points of significance for the use of masks are to reduce the vulnerability of the noxious

individual during the "pre-symptomatic" time and to stigmatize discreet people wearing masks to restrict the spread of the virus. WHO emphasizes that health care assistants prioritize surgical masks and respirators. Therefore, in today's global society, face mask identification has become a critical mission.

Detection of face masks includes detecting the face's location and then identifying whether or not it has a mask on it. The issue is proximately cognate to general object detection to detect the classes of objects. Face recognition deals categorically with the differentiation of a specific group of entities, i.e. Face. It has various applications, such as Autonomous driving, schooling, surveillance, etc. This paper provides a brief overview of the approach to the above motive using the basic packages of Machine Learning (ML) such as Tensor Flow, Keras, and OpenCV.

The remainder of the paper is structured as follows: Section II discusses related work associated with the detection of face masks. Section III addresses the nature of the dataset used. Section IV presents the details of the packages incorporated to build the proposed model. Section V gives an overview of our method. Section VI concludes and draws the line towards future works.

II. RELATED WORK

In various application domains for object detection and recognition, there have been several advances in machine learning over the years. Most of the works typically focus on the restoration of photographs and facial recognition for identity verification. But the main objective of this work is to identify people who do not wear masks in public places to monitor the further transmission of COVID-19. In the face detection process, a face is detected from an image with several attributes. According to [1], research into face identification involves expression recognition, face tracking, and pose estimation. Since the size, shape, colour, etc. of the faces differ and are not eternal, face recognition is a complicated errand.

Authors in [2] claim that there are two major challenges to facial detection: 1) the absence of significantly large datasets comprising both masked and unmasked faces and 2) the exclusion of facial expression in the region protected.

According to the work published in[3], convolutional neural networks (CNNs) in computer vision have a strict limit on the size of the input picture. The prevalent technique reconfigures the images prior to fitting them into the network to overcome the inhibition. The main task challenge here is to correctly recognize the face from the picture and then decide whether it has a mask on it or not. In order to perform surveillance activities, the method proposed should also detect a face along with a mask in motion. Nag and others built a face recognition based door access control in the IoT area[4]. The OpenCV functionality is used to detect and classify the faces of known people automatically and thus to monitor door access. P. Hu proposed a fog computing-based face detection and recognition system that aims to offload the face recognition task from the cloud to fog nodes[5].

The Principal Component Analysis (PCA) algorithm was introduced by Md.Sabbir Ejaz et al.[6] for masked and un-masked face recognition. It was noted that PCA is effective with an accuracy of 96.25 percent in recognizing faces without a mask, but its accuracy is reduced to 68.75 percent in identifying faces with a mask. Li et al.[7] used YOLOv3 for face recognition, which is focused on the darknet-19 deep learning network architecture, in which the WIDER FACE and Celebi databases were used for training, and the FDDB database was used later for evaluation. An accuracy of 93.9% was achieved by this model. Few other works are designed to distinguish individuals with or without a face mask. Two types of facial images have been considered for the training of

the model by these methods: with mask and without mask. In terms of the framework and model chosen to construct the model, the built systems differ.

Based on the above context, it is clear that very limited numbers of research papers have been published to date, particularly for mask detection, while further improvements on existing methods are needed. Therefore, we propose a method to achieve this objective by using some simple Machine Learning packages such as TensorFlow, Keras, OpenCV, in order to contribute to further improvements in face mask recognition in the battle against COVID-19.

III. DATASET

To experiment with the current system, two datasets were used. Dataset 1[16] consists of 1376 images containing 690 images of individuals wearing face masks and the remaining 686 images of individuals not wearing face masks. Fig 1 largely includes front face pose with single face in the frame and with same kind of mask having white color only.

Fig. 1. Dataset 1 samples including faces without masks and masks

Kaggle's Dataset 2[17] consists of 853 images and its presence is explained either by a mask or without a mask. Some face collections in Fig 2 are head turn, tilt, and slant with several faces in the frame and also various kinds of masks with different colors.

Fig. 2. Dataset 2 samples including faces without masks and masks

IV. INCORPORATED PACKAGES

A. TensorFlow

Tensorflow: In order to incorporate ML systems into production across a variety of computer science fields, an interface for expressing machine learning algorithms is used, including sentiment analysis, speech recognition, spatial knowledge extraction, computer vision, text summary, information retrieval, computational drug discovery and fault detection studies. In the proposed model, TensorFlow is used at the backend of the entire Sequential CNN architecture (consisting of many layers). It is also used to reshape the data(image) in the data processing

B. Keras

Keras offers fundamental reflections and construction units with high iteration velocity for the formation and transportation of ML arrangements. It takes full advantage of TensorFlow's scalability and cross-platform capabilities. Layers and models are the central data structures of Keras. All the layers used are implemented using Keras in the CNN model. It helps to compile the overall model alongside the conversion of the class vector to the binary class matrix in data processing.

C. OpenCV

OpenCV (Open Source Computer Vision Library), an open-source software library for computer vision and ML, is used to differentiate and recognize faces, recognize objects, group movements in videos, trace progressive modules, follow eye motions, monitor camera actions, expel red eyes from flash images, locate comparative images from an image database, perceive the landscape and set up markers to overlay it with increased reality and so forth. The proposed approach allows the use of these OpenCV features in data image resizing and color conversion.

V. THE PROPOSED METHOD

The proposed method consists of a cascade clasiffier and a pre-trained CNN which contains two 2D convolution layers connected to layers of dense neurons. The algorithm for face mask detection is as follows:

A. Data Processing

Algorithm 1: Face Mask Detection

Input: Dataset including faces with and without masks

Output: Categorized image depicting the presence of face mask

For each image in the dataset do

- Visualize the image in two categories and label them
- Convert the RGB image to Gray-Scale image
- Resize the gray-scale image into 100 x 100
- Normalize the image and concert it into 4 dimensional array

End

For building the CNN model do

- Add a convolution layer of 200 filters
- Add the second convolution layer of 100 filters
- Insert a flatten layer to the network classifier
- Add a dense layer of 64 neutrons
- Add the final dense layer with 2 outputs for 2 categories

End

Split the data and train the model

Data preprocessing involves converting data to a much more user-friendly, desired, and meaningful format from a specified format. It can be in any form, such as tables, photos, images, maps, etc. Such ordered data fits into a model or composition of information and captures relationships between various entities. Using Numpy and OpenCV, the proposed method deals with image and video data.

a)Data Visulization:

Data visualization is the method of translating abstract information by encoding into concrete representations using knowledge exchange and insight exploration.

In both types,' with mask' and 'without mask', the total number of images in the dataset is visualized.

The statement categories = os.listdir(data path) categorizes the directory list in the specified data path. Variable types now look like: ['with mask,' 'without mask']

Then to count the amount of labels, we need to differentiate those categories using labels=[i for I in range(len(categories))]. The labels are set as: [0, 1]

Each category is now mapped to its respective label using dict=dict(zip(categories,labels)), which initially returns a tuple iterator in the form of a zip object, where the items are then paired together in each passed iterator. The mapped variable label dict looks like: {'with mask': 0,'without mask': 1}

b) Conversion of RGB image to Gray image:

Modern descriptor-based image recognition systems continue operating regularly on grayscale images, without elaborating the method used to convert from color to grayscale. This is because the color-to-grayscale approach has no impact on the use of robust descriptors. The inclusion of non-essential information could increase the amount of training data needed for successful results to be achieved. As grayscale rationalizes the algorithm and decreases the computational requirements, instead of operating on color images instantaneously, it is used to extract descriptors.

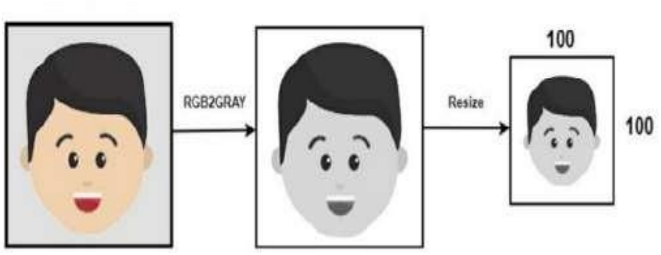


Fig. 3 Converting an RGB image to a Gray Scale Image of size :100 x 100

For changing the color space, we use the function cv2.cvtColor(input image, flag). The flag here specifies the form of the conversion. The flag cv2.COLOR BGR2GRAY is used for gray conversion in this case. Deep CNNs require a fixed-size input image. For all the images in the dataset, we therefore need a fixed, common scale.

The gray scale image is resized into 100 x 100 using cv2.resize().

c) Image Reshaping:

A three-dimensional tensor is the input during image relegation, where each channel has a prominent unique pixel. All the images must have the same tantamount size corresponding to the 3D feature tensor. Nevertheless, neither images nor their corresponding feature tensors are typically co-extensive. Most CNNs are only able to accept

Images that are fine-tuned. Throughout data collection and model implementation, this engenders many issues. However, it can help to resolve this limitation by reconfiguring the input images before augmenting them into the network.

The images are normalized such that the range of pixels between 0 and 1 converges. Then they are transformed to 4 dimensional arrays using data=np.reshape(data,(data.shape[0],img size,img size,1)) where 1 indicates the Grayscale image. Since the final layer of the neural network has 2 outcomes, the data is transformed to categorical labels with a mask and without a mask, i.e. it has categorical representation.

B. Training of Model

a)Building the model using CNN architecture:

In different computer vision functions, CNN has become ascendant. The current approach uses Sequential CNN. The First Convolution layer is followed by the Rectified Linear Unit (ReLU) and the MaxPooling layers. The Convolution layer learns from 200 filters. The size of the kernel is set to 3 x 3, which determines the height and width of the window of a 2D convolution. As the model should be aware of the expected input shape, it is important to provide information about the input shape to the first layer of the model. pInstinctive shape reckoning can be performed by following layers . In this case, the input shape is specified as data.shape[1:], which returns the data array dimensions from index 1. Default padding is “valid” where the spatial dimensions are authorised to truncate and the input volume is non-zero padded. The activation parameter is set as 'relu' for the Conv2D class. It represents an approximately linear function that possesses all the assets of linear models that can be easily optimized by gradient-descent methods.

Considering the efficiency and generalization of deep learning, it is better compared to other activation functions. In order to minimize the spatial dimensions of the output stream, Max Pooling is used. The size of the pool is set to 3 x 3 and the resulting output has the shape (number of rows or columns) of: shape of output = (input shape-pool size + 1)/strides), where the strides has the default value (1,1).

As shown in fig 4, the second layer of Convolution has 100 filters and the size of the kernel is set to 3 x 3. This is followed by the ReLu and MaxPooling layers. To insert data into CNN, a long input vector is passed through a flatten layer that transforms a matrix of features into a vector that can be fed into a fully connected neural network classifier. To reduce overfitting a Dropout layer with a 50% chance of

setting inputs to zero is added to the model. A dense layer of 64
neurons with ReLu activation function is then introduced. The final layer (Densus) with two outputs for two groups uses the Softmax activation function.

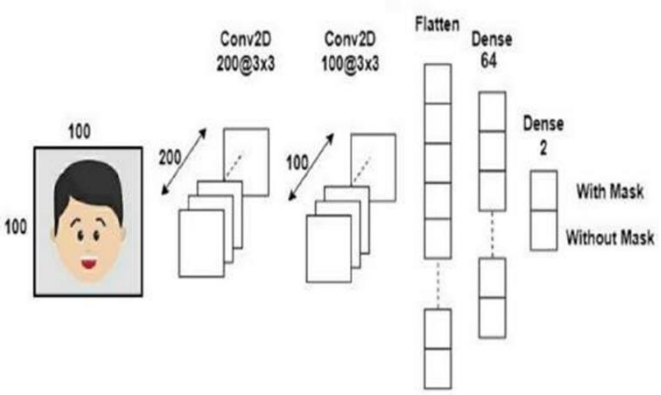


Fig. 4. Convolutional Neural Network architecture

The learning process must first be configured with the compile method. Here the "adam" optimiser is used. Categorical crossentropy, also known as multiclass log loss, is used as a loss function (the goal that the model seeks to minimize). As the problem is a classification problem, the metrics are set to "accuracy."

b) Splitting the data and training the CNN model:

The model needs to be trained using a particular dataset after setting the blueprint to analyze the data and then to be tested against a different dataset. A proper model and an optimized split of the train test help to generate precise results when creating a prediction. The test size is set to 0.1, i.e. 90% of the dataset data is trained and the remaining 10% is used for testing purposes. Using ModelCheckpoint, the validation loss is tracked. Next, the images in the training set and the test set are placed on the Sequential model. Twenty percent of the training data here is used as validation data. For 20 epochs (iterations), the model is trained, maintaining a trade-off between accuracy and chances of overfitting. Fig.

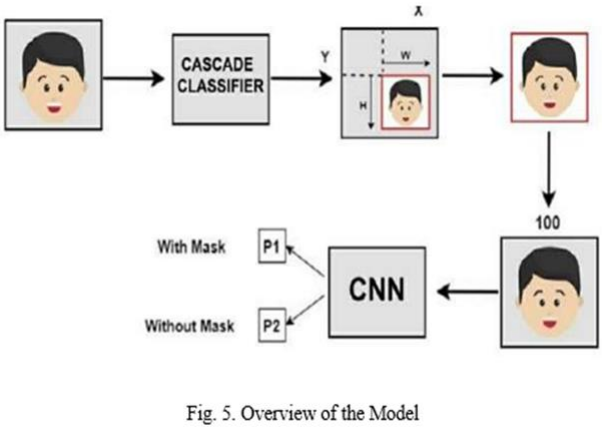


Fig. 5. Overview of the Model

Fig. 5 portrays the proposed model's visual representation.

VI. CONCLUSION

We briefly explained the motivation of the work in this paper first. Then, we discussed the learning and efficiency aspect of the model. It is possible to use it for a variety of applications. Wearing a mask may be mandatory in the immediate future, given the Covid-19 crisis. Many public service providers will require consumers to correctly wear masks in order to take advantage of their services. The model introduced would contribute greatly to the framework of public health care. It can be expanded in the future to detect whether or not a person is wearing the mask correctly. The model can be further enhanced to detect whether or not the mask is sensitive to viruses, i.e. the form of mask is surgical, N95 or not.

VII. REFERENCES

[1]. W.H.O., "Coronavirus disease 2019(covid- 19): situation report, 205".2020

[2]. "Coronavirus Disease 2019 (COVID-19) Symptoms", Centers for Disease Control and Prevention,2020.Online].Available:https://www.cdc.gov/coronavirus/2019-ncov/symptoms-testing/symptoms.html. 2020.

[3]. "Coronavirus — Human Coronavirus Types — CDC", Cdc.gov, 2020. Online]. Available: https://www.cdc.gov/coronavirus/types.html. 2020.

[4]. W.H.O., "Advice on the use of masks in the context of COVID-19: interim guidance", 2020

[5]. M. Jiang, X. Fan and H. Yan, "RetinaMask:A Face Mask detector",arXiv.org,2020. Online].Available:https://arxiv.org/abs/2005.03950. 2020.

[6]. B. Suvarnamukhi and M. Seshashayee, "Big Data Concepts and Techniques in Data Processing", International Journal of Computer Sciences and Engineering, vol. 6, no. 10, pp. 712714,2018. Available:10.26438/ijcse/v6i10.712714.

[7]. F.Hohman, M. Kahng, R. Pienta and D. H. Chau, "Visual Analyticsin Deep Learning: An Interrogative Survey for the Next Frontiers," in IEEE Transactions on Visualization and Computer Graphics, vol. 25, no. 8, pp. 2674-2693, 1 Aug. 2019, doi: 10.1109/TVCG.2018.2843369.

[8]. C. Kanan and G. Cottrell, "Color-to-Grayscale: Does the Method Matter in Image Recognition?", PLoS ONE, vol. 7, no. 1, p. e29740,2012.Available:10.1371/journal.pone.0029740.

[9]. Opencv-python-tutroals.readthedocs.io. 2020. Changing Colorspaces-OpenCV-PythonTutorials 1Documentation.online]Available at:https://opencvpythontutorials.readthedocs.io/en/latest/pytutorials/pyimgproc/pycolorspaces/py colorspaces.html. 2020.

[10]. M. Hashemi, "Enlarging smaller images before inputting into convolutional neural network: zero-padding vs. interpolation", Journal of Big Data, vol. 6, no. 1, 2019. Available: 10.1186/s40537-019-0263-7. 2020.

[11]. S. Ghosh, N. Das and M. Nasipuri, "Reshaping inputs for convolutional neural network: Some common and uncommon methods", Pattern Recognition, vol. 93, pp. 7994,2019. Available:10.1016/j.patcog.2019.04.009.

[12]. R. Yamashita, M. Nishio, R. Do and K. Togashi, "Convolutional neural networks: an overview and application in radiology", Insights into Imaging, vol. 9, no. 4, pp. 611-629, 2018. Available: 10.1007/s13244-018-0639-9.

[13]. "Guide to the Sequential model - Keras Documentation", Faroit.com, 2020. Online]. Available: https://faroit.com/keras-docs/1.0.1/getting- started/sequential-model-guide/. 2020.

[14]. Nwankpa, C., Ijomah, W., Gachagan, A. and Marshall, S., 2020. Activation Functions: Comparison Of Trends InPracticeAndResearch For Deep Learning. online] arXiv.org. Available at: https://arxiv.org/abs/1811.03378. 2020.

[15]. K. Team, "Keras documentation:MaxPooling2D layer",Keras.io,2020.Online].Available:https://keras.io/api/layers/poolinglayers/ max pooling2d/. 2020.

[16]. "prajnasb/observations", GitHub, 2020. Online].Available:https://github.com/prajnasb/observations/tree/master/experiements/data. 2020.

[17]. "Face Mask Detection", Kaggle.com, 2020. Online]. Available: https://www.kaggle.com/andrewmvd/face-mask-detection. 2020.

[18]. "TensorFlow White Papers", TensorFlow, 2020. Online]. Available:https://www.tensorflow.org/about/bib. 2020.

[19]. K. Team, "Keras documentation: About Keras", Keras.io, 2020. On- line]. Available: https://keras.io/about. 2020.

[20]. "OpenCV", Opencv.org, 2020. Online]. Available: https://opencv.org/.2020.

[21]. D. Meena and R. Sharan, "An approach to face detection and recognition," 2016 International Conference on Recent Advances and Innovations in Engineering (ICRAIE), Jaipur,2016,pp.1-6,doi: 10.1109/ICRAIE.2016.7939462.

[22]. S. Ge, J. Li, Q. Yeand Z. Luo, "Detecting Masked Faces in the Wildwith LLE-CNNs," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, 2017, pp. 426-434,doi:10.1109/CVPR.2017.53.W.H.O., "Coronavirus disease 2019 (covid19):situation report, 205".2020

The International Conference on Research Perspectives : IoT in Hybrid Grid Integrated Renewable Energy Sources
In association with International Journal of Scientific Research in Computer Science, Engineering and Information Technology
ISSN : 2456-3307 (www.ijsrcseit.com)

Design of an Optimized Neural Network Controller for Rotor Oscillations Problem

Parthasarathy.V*, Rakshitha G M, Niranjan J, Md. Farhan Khan
Cambridge Institute of Technology (North Campus) , Bangalore, Karnataka, India

ABSTRACT

In this paper, an Artificial Neural Network (ANN) based system is proposed as an alternate to the Fixed Gain Rotor Oscillations Controller (FROC) that is capable of mitigating the low frequency electro mechanical oscillations in a power system. Since our ultimate objective is to develop a Hardware Neural Network for ROC, the complete design flow is developed with due consideration of hardware optimization. The role of the developed ANN is to fix the FROC parameters for the given loading condition. An ANN block is generated in MATLAB and inserted in the SIMULINK model of FROC and tested for various operating conditions. The training pattern generation, validation, data analysis and filtering with respect to the proposed ANN model has discussed in brief. The final response and the limitations of the developed model is explained at the end. The historical database obtained from ABB-UNITROL 5000 Static Excitation system has used for training and validation of ANN performance.

Keywords : Dynamic Stability, Rotor Oscillator, Artificial Neural Network

I. INTRODUCTION

The Artificial Neural Networks can be realized for the given control problem using many real time approaches. The basic and direct implementation is the computer-based simulation .In some cases, where the ANN structure is very small and not involved with nonlinear functions can even be mathematically solved with some approximations in the desired result. The “training and recall” methodology has been successfully used for many power system applications including stability studies, Economic power generation and dispatch, reactive power control, Load forecasting, Optimum allocation of spinning reserve, location identification of compensator and many other protection related problems. The requirement of a Neuro computing system is to perform computations in parallel, scheduling appropriate data communication and specializing arithmetic or transfer operations that are performed frequently.

Prabha Kundur, John Paserbay, Venkat Ajjarapu and Göran Andersson [1] have clearly defined the concept of power system stability and explained various terminologies related to the same. The authors shared their practical industrial experiences and provided some real time examples related to

power system stability studies. R. E. Bourguet and P. J. Antsaklis [2] have given the overall view of the role of ANNs in the Electric Power System. G. Shahgholian [3] developed a qualitative study of various types of Power System Stabilizers (PSS) for rotor oscillations problem. He had compared between the conventional techniques and the intelligent techniques available for a PSS design. Also he suggested for the optimum location of controllers.

Byung Ha Lee and Kwang Y. Lee [4] have suggested the classifications of dynamic and static stabilities and examined the possibility of approximating the static stability constraints. Zhihua Cui and Xiaozhi Gao [5] have elaborated the concept of Swarm intelligence which concentrates on systems that performing their activities by adopting the real time behavior of insects and other creatures. Izni Nadhirah Sam'on, Zuhaila Mat Yasin and Zuhaina Zakaria [6] suggested a new technique using ant ion optimizer technique to solve the unit commitment problem.K. Gnana Sheela and S. N. Deep [7] recommended various techniques available for the ANN structure design.

Janardan Misra and Indranil Saha [8] suggested that HNN models will have their significant role in near future when the industry will face their demands imposed by ubiquitous computing with learning and autonomous decision making capabilities. Philip H.W. Leong [9] has summarized the advancements and the innovations in the FPGA realization techniques. Starting from the ASIC implementations, they have provided the year wise growth of the hardware development.

Parthasarathy, Muralidhara. B.[10] developed a state space model of a FROC and analyzed its performance. They have varied the gain within the prescribed levels by keeping time constants fixed. Also the Simulink model they have developed was comparatively simpler than the other existing models. The critical gain obtained from routh's array has been used to develop the mathematical modeling in this work.

II. THE PROPOSED ANN MODELING

The complete ANN design consists of fixing some important specifications by using trial and error technique. After fixing a certain parameter, the corresponding neural block is generated and the error was calculated for a randomly selected loading condition. The specification that allows a permissible error is fixed for the final model.

The ANN consists of a 3-x-3 structure where 'x' denotes the number of hidden layers and the number of neurons in every layer. Since our ultimate objective is to develop the hardware realizable ANN, the smaller the structure, lesser the chip space occupied by the algorithm. The input and output parameters and their hardware requirements are fixed, the optimization to be done only with respect to the hidden layer or processing layer requirement. For validating the training patterns, a randomly selected sample group has been compared with the real time data obtained from the ABB PSS Commissioning Report for Static Excitation System UNITROL 5000 model installed at Adani plant, Kawai.

The proposed ANN has given with the inputs of Real power, Reactive power and the Terminal voltage and it is to predict the values of ROC gain and the time constants. There are direct mathematical relations existing for the evaluation of the same. Subsequently, the ROC will produce the damping signals, which will be given to the excitation system. Out of the three output variables, it was found that the ROC gain is affected directly by the electrical torque, demagnetizing effect and terminal voltage variations. Obviously the real power output also depending on

the above parameters. The lead time constant is again depends on the electrical torque constant and hence the real power. The lag time constant is the function of reactive power injected to the system. And hence any change in the three parameters P,Q and V_t can be made approximate variations in the corresponding dynamic coefficients also.

The data size and the training duration are directly proportional to each other. It was observed that the network takes several minutes to complete the training process. Hence, the patterns were minimized to a best lower level without compromise on the core information provided by the system. Then the number of iterations was changed to different values ranging from 500 to 15000 and it was found that within the range of 1000 to 5000 iterations the data accumulation was near accurate and the corresponding data gradient stood at around 0.0025. (roughly 5 % deviation is acceptable). After the necessary training is over, the testing has been conducted with a selected data set of roughly 10% of entire population.

For the basic prediction problems, a single layer network may produce the satisfactory results. In that case, the ANN processor has to merely identify the optimum output from its training database. But if the ANN is connected to a controller and it has to tune the same for any given operating condition, the response of hidden layer is to predict and also to tune the rest of the system. For example, in the proposed work, the ANN has to predict as well as tuning the ROC. In such a scenario, the processor may require two or even more hidden layers. But our ultimate aim is to develop a hardware neural structure, any additional layer may add burden to the existing system. Hence the number of hidden layers was fixed with two. Infact, we have tried with three hidden layers also without any improvement in the performance.

III. NUMBER OF PROCESSING NEURONS

Since the introduction of more number of neurons may end up with the corresponding weights allocation and the overall LUT requirement will be increased. The input is only three and having two layers for parallel processing, any number around 5 to 6 neurons is more than sufficient in this case.
We have witnessed that the ANN training has abruptly terminated and still the error was about 8%. Then we have gradually changing the number of neurons in the ascending order to pick the optimum number. Or we can go for the descending order also for the same. For both the approaches, sufficient works are available in the literature in the name of constructive and pruning methods.

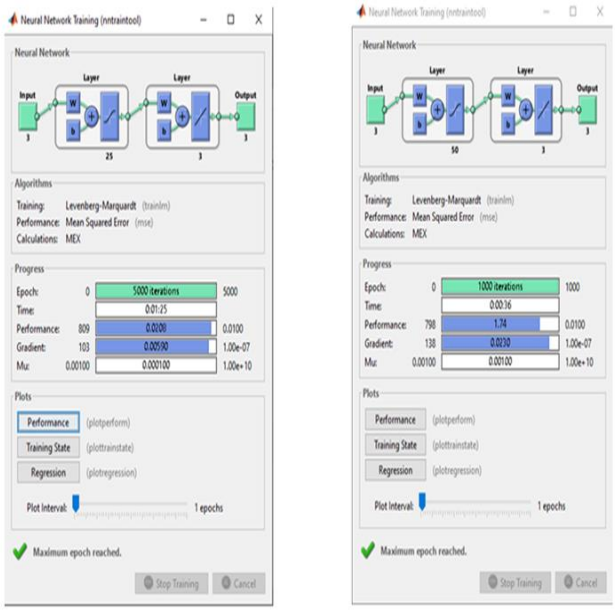


Fig. 1. ANN training for different hidden neurons

The fig (1) shows the training process for different number of hidden neurons.We have given the priority for the performance index, which may be considered as the direct measure of the ANN training efficiency. Number of iterations is not that much significant because, in several attempts we have noticed even after 10000 iterations, the desired error level has not achieved. Also for the same structure, the ANN required different number of iterations

every time. That is the number of epochs for the training patterns are not fixed.

The proposed problem of HNN based ROC is highly nonlinear in nature and the selection of the Activation Functions (AFs) for the ANN model is very crucial. We need to pick two AFs out of which one must be a nonlinear AF. For understanding the AF theory, we have selected two liner AFs (purelin and Rectified Linear Unit) and two nonlinear AFs (Tansig and Logsig).Different combinations were tried using these four AFs and some of the important conclusions were derived at the end of training the ANN with the specified training database for an uniformly fixed epoch of 1500.

IV. THE ANN-ROC SIMULINK MODEL

A Neural block generated by using the "gensim" code in .m file is able to predict the output of any trained information. The SIMULINK model of the FROC diagram shown in fig (7) has been developed with the system parameters for the actual studies about the FROC design. In the diagram the ANN block is included between the system output and the input.

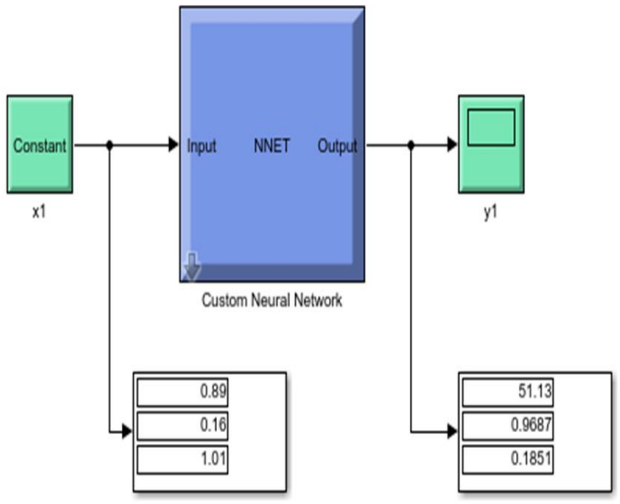


Fig: 2. The generated Neural Block

The figure shown in (2) gives one random loading condition and the corresponding ANN output. Thus, the controller is continuously monitoring the system output and able to produce the necessary positive damping signals so that the negative damping due to the exciter loop and other reasons will be nullified. After the analysis about the uncompensated system is completed, the fixed gain compensator has been introduced between the system inertia block and the summation point A.

Sl. No	System loading Condition (in p.u.)			ANN based ROC setting		
	P	Q	V_t	Gain	T_{lead}	T_{lag}
1	**0.89**	**0.16**	**1.01**	**51.13**	**0.97**	**0.18**
2	0.75	0.36	0.87	47.01	0.86	0.13
3	0.65	0.31	0.84	43.0	0.77	0.12
4	0.44	0.21	0.76	36.44	0.74	0.13
5	0.85	0.30	0.91	49.66	1.10	0.13
6	1.01	0.27	0.98	54.22	0.76	0.11
7	0.87	0.11	0.92	51.05	0.63	0.08
8	0.93	0.34	0.94	52.12	0.77	0.12
9	1.27	0.19	1.08	64.31	0.97	0.14
10	1.19	0.05	1.05	59.92	0.65	0.31

Fig.3. The ANN output for random loading conditons

The table shown in fig (3), ten samples tested on ANN for its ability of prediction. Similarly about 50 different operating conditions were considered such that some are present in the training database and some are not. And it was found that the average error created by the ANN was stood at 3.5% which is acceptable. The figure above is the ANN prediction for an input of P=0.89 pu, Q=0.16 pu and V_t=1.01 pu.

Those values were compared with the real time data developed from empherical formula. For the validity of the ANN output few samples were tested with the real time data. Again the error was linear for normal cases and exponential for overloading or under loading cases.

Since we are mainly concentrating on the linear operating region, these samples can be neglected. It

was observed that the system is effectively able to produce the required damping controller parameters and hence the initial negative damping have been changed to positive damping. Also the synchronizing torque component has been increased to almost 1.7 times the uncontrolled system's output. This is an indication of the real power output increment. Again when the ANN is suggested for larger values of controller parameters (for overloaded conditions) the accuracy of prediction has been reduced. It is evident that for few values of ANN outputs, the system starts producing negative damping, which is undesirable.

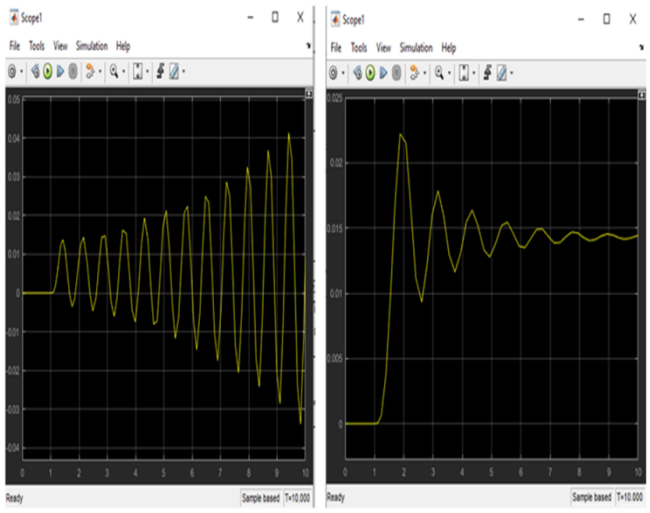


Fig: 4. FROC output for K_5= -0.1103

It was observed from the graph (4) , even though the FROC is able to suppess the oscillations, the settiling time for the machine is comparitively high. The other important factor is the change in the power angle when there is a change in change in load. The role of any controller is to ensure this change very negligible so that the magnitue of oscillations and the oscillatory period will get reduced. The following table has developed by varying the value of DC_5 from positive to negative and the corresponding change in rotor angle tabulated. The rotor angle can be directly measured from the output terminals of the simulink model.

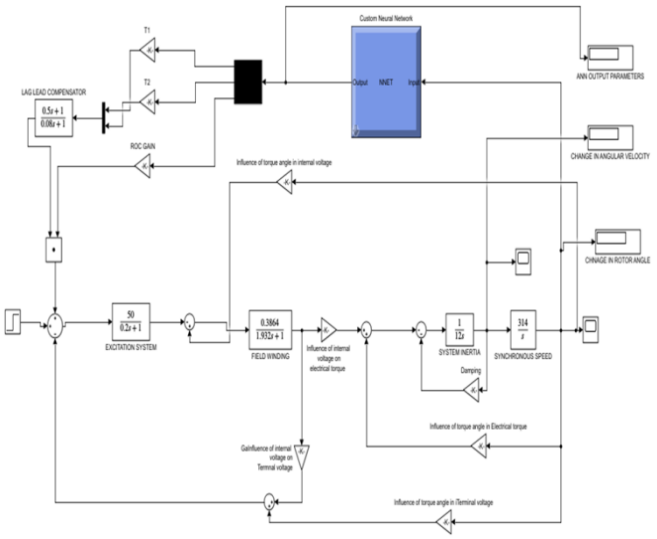


Fig 5. Simulink model of ANN –ROC-SGG system

The Fig (5) shows the Simulink model of the ANN-ROC designed for SGG system. The ANN is connected across the output terminnals of the system and fetch the real power , reactive power and terminal voltage as its input. For every change in DC_5, there is a corresponding change in these paramters.The ANN is able to produce the corresponding ROC parametrs which in turn guide the FROC to adjust itself. For this sake, the ROC is connected in cascade with the ANN block developed using SIMULINK. A demultiplexer at the output of ANN block is deviding the single ANN signal into three so that the gain and the time constant values for the lag-lead networks being given to their appropriate elements. The output of the FROC guided by ANN is given to the input summation block.

The effect due to the DC_5 also reaches the point, where the impacts are trying to nullify each other. Still due to the un availability of exact value of compensation, the system continue to produce some oscillations but within the prescribed limits. The sutained existance or change in the rotor angle may leads the system to unstable mode. And hence the variation of this angle going beyond certain extend, the ANN is able to deliver the suitable compensation. There are some conditiones where keeping the gain constant and varying only the time constants

produces the satisfatory output. To achieve this , the link between the gain block and the dot product block has been disconnected and simulated. This can be reversed by disable the phase compensation block also. This flexiblity is not available in the FROC system. In the above fig 7 ,We have tried with 4 different sets of training patterns within the operating limits, the testing process repeated. The system output was compared with the actual data available and the error calculated with respect to each parameter.The tabulations are given below.

It was found that the ANN was able to produce the output with error of approximately 8%. The table in fig (6) shows the comparion between ANN output and the actual database values for the given loading condition. It was found that for all the three parameters, the ANN response is satisfactory.

Real Power	ANN output			Actual Data		
(pu)	Gain	T_{lead}	T_{lag}	Gain	T_{lead}	T_{lag}
0.67	37.31	0.67	0.07	33.68	0.61	0.06
0.84	57.63	0.96	0.13	55.61	0.93	0.12
1.01	62.31	1.02	0.08	66.51	0.99	0.07
0.82	51.13	0.97	0.18	48.98	0.98	0.18
0.76	47.01	0.86	0.13	44.0	0.85	0.11
0.74	43.0	0.77	0.12	45.23	0.74	0.13
0.66	36.44	0.74	0.13	33.31	0.72	0.12
0.79	49.66	1.10	0.13	47.1	1.07	0.11

Fig.6.Comparion between ANN ROC and FROC

The fig (7) shows a sample respose for K_5= -0.1725. It is to be noted that the peak overshoot was very high for the ANN-ROC which is due to the sub synchronous oscillations and the system is able to settle within comapritively lesser time duration.

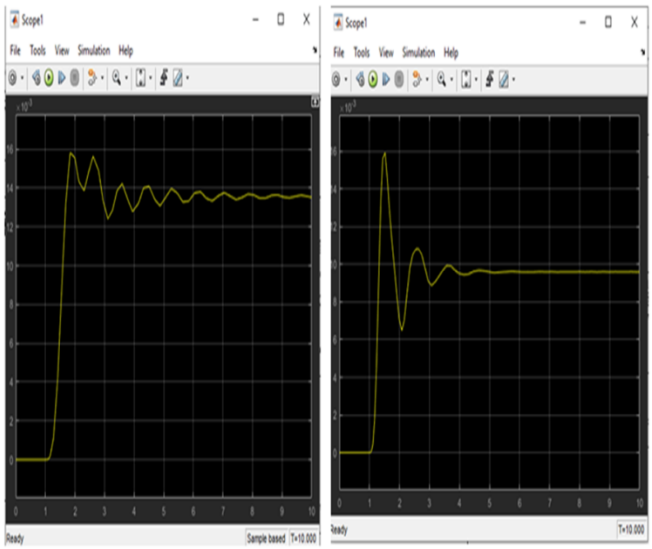


Fig.7.ANN-ROC response for K_5= -0.1725

V. CONCLUSION

In this paper, an Artificial Neural Network based rotor oscillations Controller have been designed on the MATLAB platform. The ANN training process, fixing the number of processing layers, processing neurons, activation function type are discussed in brief. The desirable property of the ANN –ROC is its ability to ensure the improvement of the dynamic stability and at the same time not permitting the machine to produce negative synchronizing torque for any value of exciter gain. Also a random tuning may push the system to even more un stable mode. Not only that, again, the system may start producing negative damping torque.

Hence, the ANN has been trained with suitable training parameters (particularly the ROC gain) such that any of the considered training set is not producing the negative T_s and only produce sufficient positive T_d. The ultimate limitation of the developed ANN-ROC model is the larger processing time. Slower processor and lack of parallelism makes the overall response delayed.Hence we are planning to design a hardware Neural Network controller which may be doing the same task with comparitively lesser time duration.

VI. REFERENCES

[1]. Prabha Kundur , John Paserbay, Venkat Ajjarapu , Göran Andersson , "Definition and Classification of Power System Stability", IEEE Transactions On Power Systems, Vol. 19, No. 2, May 2004,1387-1401.

[2]. R. E. Bourguet, P. J. Antsaklis, "Artificial Neural Networks in Electric Power Industry," Technical Report of the ISISGroup, No. ISIS-94-007, Univ ofDame, April 1994, pp:1-37.

[3]. G. Shahgholian , "Review Of Power System Stabilizer: Application, Modeling, Analysis And Control Strategy" International Journal on Technical and Physical Problems of Engineering September 2013, Issue 16 Volume 5 Number 3 Pages 41-52.

[4]. Byung Ha Lee and Kwang Y. Lee, "Dynamic And Static Voltage Stability Enhancement Of Power Systems" IEEE Transactions on Power Systems, Vol. 8, No. 1, Feb,1993,pp:231-238.

[5]. Zhihua Cui , Xiaozhi Gao, "Theory and applications of swarm intelligence" Neural Computations & Applications (2012) 21:205–206.

[6]. Izni Nadhirah Sam'on , Zuhaila Mat Yasin , Zuhaina Zakaria, "Ant Lion Optimizer for Solving Unit Commitment Problem in Smart Grid System" Indonesian Journal of Electrical Engineering and Computer Science Vol. 8, No. 1, October 2017, pp. 129 - 136 .

[7]. K. Gnana Sheela and S. N. Deep, "Review on Methods to Fix Number of Hidden Neurons in Neural Networks"Hindawi Publishing Journal of Mathematical Problems in Engineering Volume 2013,pp:1-12.

[8]. Janardhan Misra and Indranil saha, "Artificial Neural Networks in hardware: a survey of two decades ofprogress" Elsevier proceedings of neuro Computing74 (2010),pp239-255.

[9]. Philip H.W. Leong , "Recent Trends in FPGA Architectures and Applications" ,Proceedings of the 4th IEEE International Symposium on Electronic Design, Test & Applications, September, 2008, pp:137-143.

[10]. Parthasarathy, Muralidhara B, "Design of Rotor oscillations Controller using State Space modeling" proceedings of 1st IEEE International Conference on Advanced Technologies in Intelligent Control, Environment, Computing & Communication Engineering (ICATIECE)Dayananda Sagar Academy of Technology, Bengaluru, Feb, 2019, pp 234-239.

The International Conference on Research Perspectives : IoT in Hybrid Grid Integrated Renewable Energy Sources
In association with International Journal of Scientific Research in Computer Science, Engineering and Information Technology
ISSN : 2456-3307 (www.ijsrcseit.com)

Experimental Investigation of Cryogenically Treated Tool in Turning

D. M. Dalwe[1], R. G. Tated[2]
[1]Mechanical Engineering Department, S. T. B. College of Engineering, Tuljapur, Maharashtra, India
[2]Mechanical Engineering Department, Matoshree College of Engineering, Nashik, Maharashtra, India

ABSTRACT

In this research work experimental investigations on the wear resistance of Powder Metallurgy tools in turning operation are carried out. The tool steel is exposed to various cryogenic actions beside with the conservative heat treatment. For this a Taguchi's orthogonal array method is applied to finalize experimental investigations. The effect of cryogenic treatment on hardness and micro hardness of tool steel is found. Scanning electronic microscopy is used for analysis of effect of cryogenic treatment on the microstructure of tools. Electronic Dispersive Spectroscopy (EDS) method is used to obtain composition of specimen as a whole and the composition of individual components.

Keywords : Powder Metallurgy-steel tools, cryogenic treatment, Taguchi O. A., hardness, wear resistance.

I. INTRODUCTION

Lot of publications inspected DCT of usually made tool steels. However insufficient considerations have been given to materials made by powder metallurgical tool steels. The PM - powder metallurgical technique can accomplish a further homogeneous microstructure equated to unadventurously made steels. The powder metallurgical method was selected as segregation-free and additional homogeneous microstructures through an extra clearness can be accomplished in contrast to a conservative metallurgical (IM) method. Dissimilar mechanical properties may be reached by PM technique which is determined by on whether models were taken parallel or perpendicular to the route of hot deformation. Then PM tool steels offer constant spacing among single carbides in all directions [1].

Tool life is nothing but a time period in minute among two continuous grindings. Tool life is principally linked to the tool wear. Performance of a tool is based on material of tool, work material and cutting conditions or environment. Cutting tools have a limited life because of wear [2, 3].

Sub-zero behavior (furthermore considered to as cryogenic treatment) is a method of treatment at low temperatures (lower than -80°C) further to the Conventional Heat Treatment (CHT). Fundamentally it is the lee way of regular heat treatment processes. In event of tool steels, cryogenic treatment at temperatures fluctuating from -140°C to -196°C, denoted to as Deep Cryogenic Treatment (DCT),

counted in among the quenching and tempering stays the utmost frequently used [7].

A. Oppenkowski, et al. [1], the maximum noteworthy features which influenced the material goods i.e. properties of tool steels were the austenitizing and tempering temperatures during conservative heat treatment. The factors which were concerning to deep cryogenic treatment, the allotment time and the heating rate had a important effect on the material properties. It was seen for a lengthier holding time of 36 Hr, the wear level ranges minimum and rise again on auxiliary holding.

G. Straffelini et al. [4], the cryogenic behaviors convinced an development in the wear resistance. T.V. Sreerama Reddy et al. [5], the upgrading in flank wear resistance was 21.2% and upgrading in tool life up to 11.1%. The foremost cutting force was smaller as linked to unprocessed tungsten carbide enclosures and upgrading in surface roughness. B. Podgornika et al. [6], the deep-cryogenic behavior amended microstructure of PM high-speed S390 Micro-clean from Boehler steel as advanced needle-like martensitic structure. It revealed in higher exterior hardness and enhanced tribological properties, mainly in terms of friction and infuriating resistance alongside stainless steel. Sobotova J et al. [7], for Vanadis6 steel later Deep Cryogenic Treatments there was a slight rise in the wear rate, while Vanadis30 did not display somewhat effect of the cryogenic treatment on wear rate. Martin Kurik et al. [8], the bending-strength was closely matching after the conservative and cryogenic treatment of commercial manufacturing method tool steel 1.2379 and powder metallurgy HSS Vanadis23. The strong point of steel Vanadis23 was found 1000 MPa greater than that of steel 1.2379 in entirely the categories of heat treatment. Then again the hardness standards for the two considered materials were rather reduced afterwards the cryogenic treating. T. V. Sreerama Reddy et al. [9], cryogenically exposed tungsten carbide tool tips were found to reduced tool wear and here was rise in the tool life, decreased cutting force and provided well surface finish related to unprocessed tools. A.Y.L. Young, et al. [9], the cryogenic behavior enhanced the life length of cutting tools would be contingent a lot on the cutting circumstances. Mild cutting conditions of tools had increase due to cryogenic treatment, and increased cutting processes with long times of heating of the cutting tool would not advantageous from it. I. Gunes et al. [10], the cryogenic method augmented the wear resistance of the Vanadis4 steel and it was not considerable in the DCT-24 Hr samples tempered at 525ºC. Sanja Solic et al. [11], observed the effect of deep cryogenic treatment on the microstructure in high speed steel grade PM S390 MC Form made by powder metallurgy. They noted that appropriately accompanied heat treatment could significantly affect the finite characteristics of high speed steels; henceforth the heat treatment factors were selected depending on the definite characteristics of the particular tool.

II. EXPERIMENTAL WORK

A. Material

For the experimentation purpose tool steel Vanadis23 material was procured from Bohler-Uddeholm India Pvt. Ltd. The Table 1 shows the chemical composition of material provided by source Company

Table 1- PM Vanadis23 Tool Steel-Chemical Composition

C	Si	Mn	P	S	Cr	Mo	W	V
1.30	0.57	0.33	0.022	0.008	4.02	4.85	6.12	3

B. Cryogenic Treatment

First 48 Vanadis23 PM steel tool pieces are conventionally heat treated in vacuum oven. Fig. 1

shows the heat treatment method applied. The material is exposed to austenitizing at temperature 1040°C and tempered twice at a temperature of 500°C and 530°C respectively.

The objective of this research work is to study the consequence of cryogenic temperature on improvement in wear resistance of cryogenically treated Vanadis23 powder metallurgy steel tool. To study the effect on the tool wear resistance four factors, cryogenic temperature, cutting speed, feed and depth of cut are considered. Therefore Taguchi Orthogonal Array method is used for experiments. Taguchi L_{16} (2^{15}) array is decided for experimentation. The main factor is considered is cryogenic temperature hence it is used at four levels. From the rest of three parameters cutting speed has highest effect hence for this also four levels are decided. Feed and depth of cut are each used at two levels. Consequently L_{16} modified orthogonal array with two factors at four levels and two factors at two levels is finalized for cryogenic treated tools experimentations.

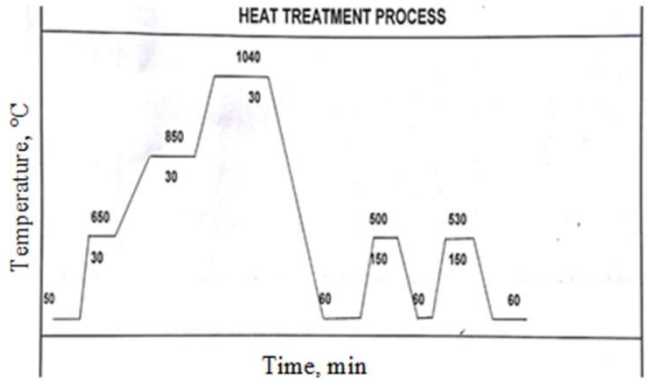


Figure 1: Heat Treatment Process

36 tools, after heat treatment later on subjected to cryogenic treatment. Three different cryogenic treatments at temperatures as -70ºC, -125ºC and -190ºC respectively followed by four tempering cycles at temperature 150ºC, 150ºC, 180ºC and 525ºC respectively are applied to the samples. The soaking time for each cryogenic temperature used is 4 h and tempering period used is 1.5 h for first three tempering and 2 h for fourth tempering. From 36 pieces, each set of 12 samples are given cryogenic treatment temperatures at: -70ºC, -125ºC and -190ºC respectively after conventional heat treatment and then to four tempering cycles as mentioned above. Later on all samples are ground to single point cutting tools as per standard geometry.

C. Testing

For testing, these tools and mild steel - Fe 410 as work material is used. The turning operations are conducted on the CNC machine ACE DESIGNERS, APPOLLO. In the experimentations, modified L_{16} array is used for three times keeping all other factors constant in dry conditions. To measure the flank wear of tools Mitutoyo make tool maker's microscope is used. Table 2 shows the average tool wear.

Table 2- Average tool wear

Cryogenic temperature, T_c (ºC)	Cutting speed, V (m/min)	Feed, f (mm/rev)	Depth of cut, d (mm)	Tool Wear, w_{cryo} (mm)
28	12.00	0.100	0.500	0.048
28	13.80	0.100	0.575	0.056
28	15.87	0.115	0.500	0.055
28	18.25	0.115	0.575	0.072
-70	12.00	0.115	0.500	0.049
-70	13.80	0.115	0.575	0.061
-70	15.87	0.100	0.500	0.056
-70	18.25	0.100	0.575	0.056
-125	12.00	0.115	0.575	0.060
-125	13.80	0.115	0.500	0.061
-125	15.87	0.100	0.575	0.064
-125	18.25	0.100	0.500	0.055
-190	12.00	0.100	0.575	0.043
-190	13.80	0.100	0.500	0.050
-190	15.87	0.115	0.575	0.055
-190	18.25	0.115	0.500	0.069

III. RESULTS DISCUSSION

From the Table2 is seen that cryogenic temperature affected on tool wear. For -70°C treatment less tool wear is obtained for a cutting speed of 12 m/min, feed of 0.115 mm/rev and depth of cut 0.5 mm. But in case of tools subjected to -125°C tool wear obtained is more for all combinations of parameters. While the tools exposed to -190°C show reduced tool wear and is least among all temperatures for a cutting speed of 12 m/min, feed 0.1 mm/rev and depth of cut 0.575 mm. Hence as cryogenic treatment reduces the tool wear. It means that wear resistance increases and hence tool life increases. This improvement in wear resistance may be due to decreased carbide size after cryogenic treatment. Figures 2 to 5 show graphs got for the factors cryogenic treatment temperatures, cutting speed, feed and depth of cut versus tool wear respectively. Note: NC = for conventionally treated samples atmospheric temperature of 28° c is considered while plotting the graph of cryogenic temperature versus tool wear.

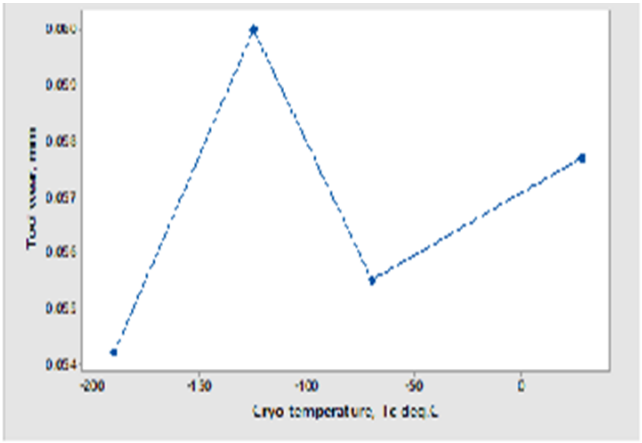


Figure 2: Tool wear versus Cryogenic Temperature

The graph of cryogenic temperature versus tool wear indicates that as the tool wear reduces the cryogenic process temperature is reduces up to -70ºC. It is increased for next level of cryogenic temperature of -125ºC. The tool wear is decreased after reduction in cryogenic temperature of -190ºC. For lower process temperature, lower is the tool wear. This shows that there is improvement in wear resistance of Vanadis23 PM steel tool when there is reduction in cryogenic temperature. It is in accordance with the earlier research done by researchers on different types of tool, die and PM steel [1, 4, 5, 6, 7, 9, 10]. From the graph it is observed that reduction in wear rate for the zone of room temperature to -70ºC is slow. In the zone of - 70ºC to -125ºC it is rapidly increased. The reduction in wear rate from -125ºC to -190ºC is rapid.

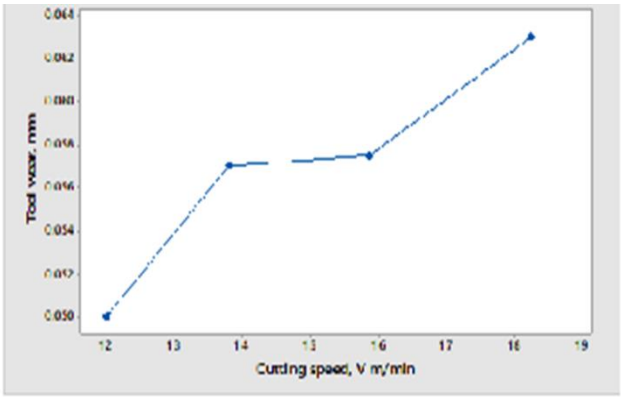


Figure 3: Tool wear versus Cutting Speed

From the figure 3, it shows that the effect of cutting speed on tool wear is rapid in first zone, 12 to 13.8 m/min. For the zone of cutting speed 13.8 to 15.87 m/min it is nearly constant. For cutting speed of 15.87 to 18.25 m/min it is increased. The result of feed and feed on tool wear is similar as seen from Figures 4 and 5 respectively. There is increase in tool wear when feed and depth of cut are increased.

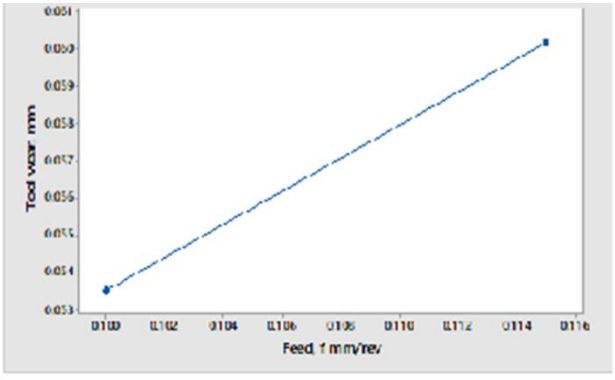


Figure 4: Tool wear versus Feed

There is less wear in cryogenic treated tools it means that wear resistance is improved of the Vanadis23 PM steel tool owing to cryogenic treatment for -70ºC and -190ºC. The improvement of wear resistance in tools results in increased tool life.

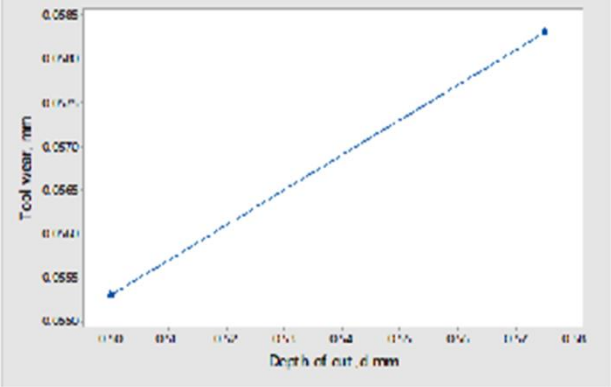


Figure 5: Tool wear versus depth of cut

Figure 6 shows the average tool wear by factor level for cryogenic treated tools.

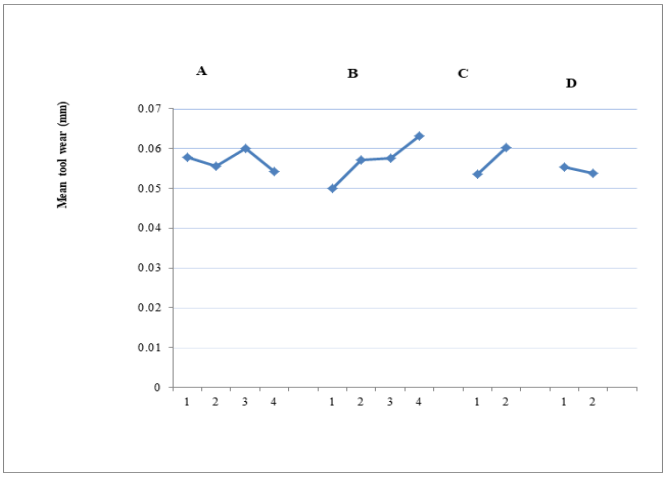


Figure 6: Mean Tool Wear by Factor Level

A – Cryogenic treatment temperature, Tc (ºC), B – cutting speed (m/min), C- Feed (mm/rev), D – Depth of cut (mm)

As per the literature review the hardness usually heat treated PM steels there is decrease in hardness value after cryogenic treatment [6]. In this present work this is not in agreement with the results obtained by other authors. The hardness values of usually heat treated Vanadis23 PM steel tool samples are lower (56 HRC) than the cryogenically treated. It is highest (62.2HRC) for cryogenically treated samples at -70ºC and lowest (59.5 HRC) for cryogenically treated samples at -125ºC.

As well as micro hardness values of one sample each of Vanadis23 PM tools from conventionally treated and cryogenically treated at -70ºC, -125ºC and -190ºC are obtained. The average micro hardness is lowest (667.33 Hv1.0 kg) for conventionally heat treated samples. It is highest (781Hv1.0 kg) for cryogenically treated samples at -70ºC and lowest (771 Hv1.0 kg) for cryogenically treated samples at -125ºC.

Figures 7 and 8 show the graphs of hardness and microhardness.

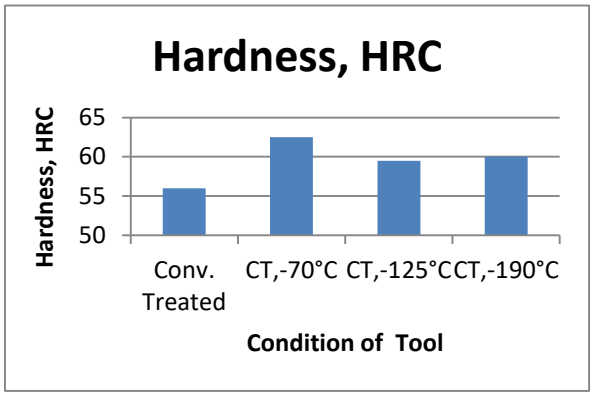


Figure 7: Hardness of conventionally and cryogenically treated samples

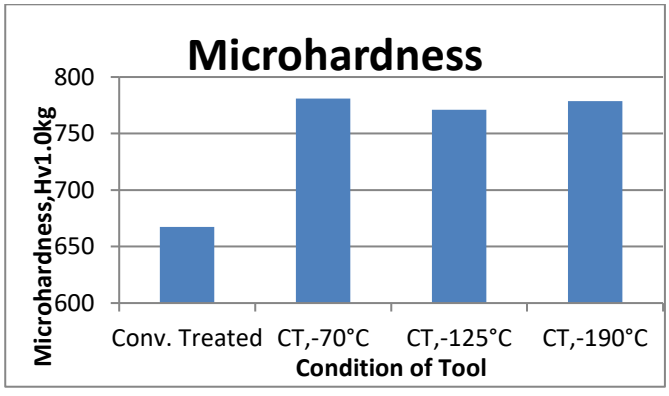


Figure 8: Micro hardness of conventionally and cryogenically treated samples

SEM and EDS analysis of usually heat treated and cryogenically treated samples are carried out. Figure

8 shows the microstructure of conventionally treated sample obtained by SEM (Scanning Electronic Microscopy). However in heat treated sample no martensite is seen but spherical carbides are uniformly distributed among them principal carbides of chromium and secondary carbides of vanadium and tungsten.

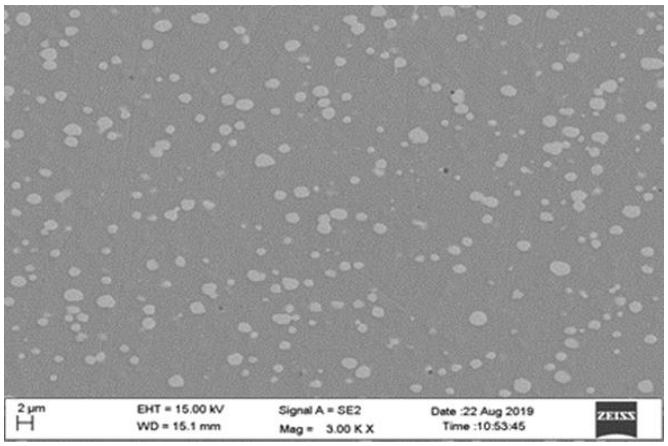


Figure 9: Microstructure of Conventionally TreatedVanadis23 PM Specimen

The Figures 9 to 11 show the microstructures on SEM for specimens cryogenically treated at -70ºC, and -190ºC. The SEM analysis does not show any significant difference in the structure between the usually heat treated and cryogenically treated samples. In case of cryogenically treated samples same pattern is seen but carbide size is decreased.

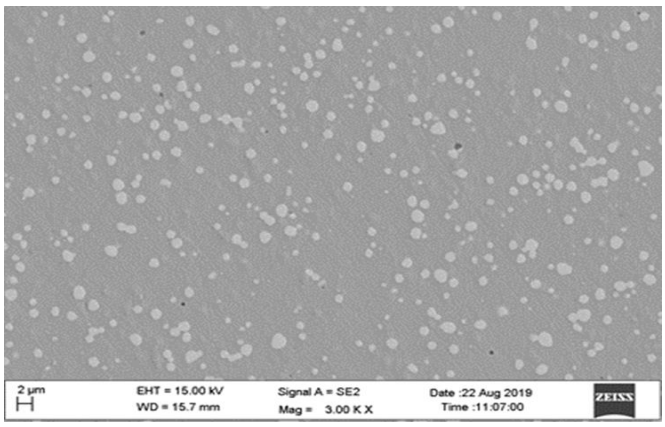


Figure 10: Microstructure of Cryo Treated Vanadis23 PM Specimen at -70ºC

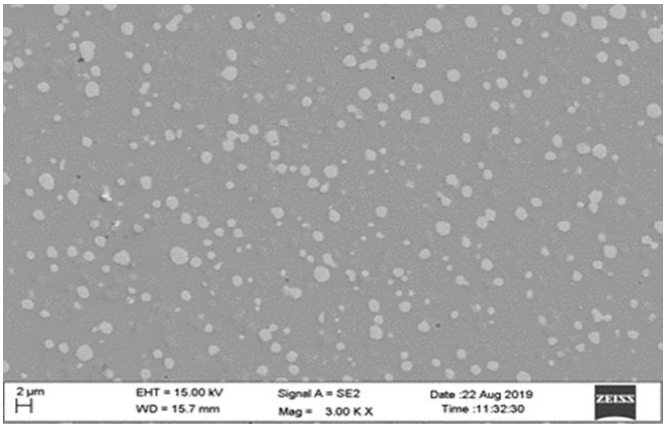


Figure 11: Microstructure of Cryo Treated Vanadis23 PM Specimen at -190ºC.

Electronic Dispersive Spectroscopy (EDS) method used to obtain composition of specimen as a whole and the composition of individual components. It is a non-destructive characterization technique. EDS spectrums of conventionally and cryogenically treated samples at -190ºC is obtained as shown in Fig. 12. The results revealed that Mn, V, Cr, W, Si, S, Mo and Fe the principal elements present in the inspection field, with Mn being the most abundant.

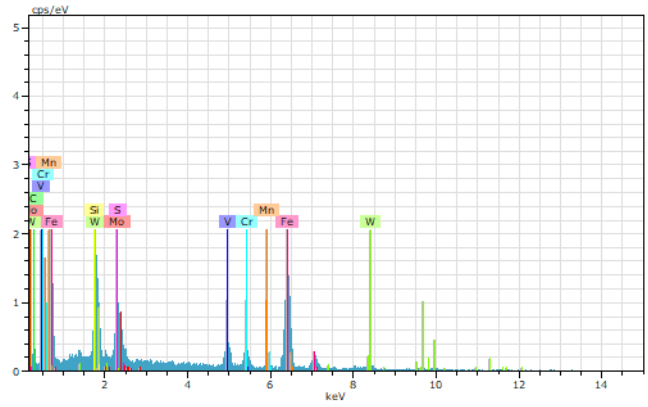


Figure 12: EDS image of -190ºC Cryogenic Treated Tool Sample

IV.CONCLUSION

In this study, experimental design with Taguchi method is applied for investigation of cryogenically and conventionally treated tool in turning operation with various cutting tools and machining conditions. Following conclusions are made from the experiment:

Cryogenic treatment affected on the hardness of tools. Both hardness and microhardness values of cryogenically treated tools are increased. Under the dry turning condition there is notable improvement in wear resistance of tools treated at -70°C and -190°C in comparison with the conventionally treated one. This improvement in wear resistance may be due to decreased carbide size after cryogenic treatment. Based on these laboratory results, the wear-rate values are much lower in cryogenically treated tool than conventionally treated one

V. REFERENCES

[1]. A. Oppenkowski, S Weber, W. Theisen, Evaluation of factors influencing deep cryogenic treatment that affect the properties of tool steels, Journal of Materials Processing Technology 210, (2010) 1949- 1955

[2]. Production Technology hmt, Banglore, Sixteenth reprint 1999, Tata McGraw – Hill Publishing Company Limited, New Delhi

[3]. David A. Stephenson, J. A. Agapiou, Metal Cutting Theory and Practices, Marcel Dekker, 1997, pp 577-601.

[4]. G. Straffelini, G. Bizzotto, V. Zanon, Improving the wear resistance of tools for stamping, Wear 269 (2010) 693-697.

[5]. T.V.Sreerama Reddy, T. Sornakumar, M. Venkatarama Reddy, R. Venkatram, Machining performance of low temperature treated P-30 tungsten carbide cutting tool inserts, Cryogenics 48 (2008) 458-461

[6]. B. Podgornik, F. Majdic, V. Leskovsek, J. Vizintin, Improving tribological properties of tool steels through combination ofdeep-cryogenic treatment and plasma nitriding, Wear 288(2012) 88– 93

[7]. Sobotova J, Kurik M, Krum S and lacza J., Effect of sub-zero treatment on the wear resistance of P/M tool steels, Journal of Applied Mechanical Engineering 2016, 5:6

[8]. Martin Kurik, Jakub Lacza, Tomas Vlach, Jana Sobotova, Study of the properties and structure of selected tool steels for cold work depending on the parameters of heat treatment, Materiali in Technologij /Materials and Technology (1967–2017) – 50 LET/50 YEARS, MTAEC9, 51(4)585(2017) ISSN 1580-2949

[9]. A.Y.L. Young, K.H.W. Seah, M. Rahman, Performance evaluation of cryogenically treated tungsten carbide tools in turning, International Journal of Machine Tools and Manufacture 46 (2006) 2051 –2056

[10]. I. Gunes, M. Uzun, A. Cetin, K. Aslantas, A. Cicek, Evaluation of wear performance of cryogenically treated Vanadis 4extra tool steel, Kovove Mater. 54 2016 1-10.

[11]. SanjaSOLIC, ZdravkoSCHAUPERL, Matjaz GODEC, VladoTROPSA,Microstructural changes in heat treatment of PM high-speed steels. Technical, 11, 4(2017), 166-170

[12]. Mr. Dalwe D. M., Tated R. G., Study of Cryogenic Treatments on PM Tool Steels, International Journal in Engineering and Techniques, Vol. 5, Issue 3, May-June 2019, pp101-108 ISSN: 2395-1303

[13]. Dalwe D. M., Dr. Tated R. G. Performance Studies of Vanadis23 Tool in Turning using Taguchi Method,International Research Journal of Engineering and Technology, Vol.: 7 Issue: 01 Ian 2020, pp279-287 e- ISSN: 2395-0056, p-ISSN: 2395-0072

[14]. Dalwe D. M., Dr. Tated R. G., Performance Analysis of Cryogenically Treated Vanadis23 Tool under Dry Turning Conditions, International Research Journal of Engineering and Technology, Vol.: 7 Issue: 01 Ian 2020,

pp288-294 e- ISSN: 2395-0056, p-ISSN: 2395-0072

[15]. D. M. Dalwe, R. G. Tated, Investigations on Wear Resistance of PM Steel Tool after Heat Treatment, Journal of Interdisciplinary Cycle Research, Vol. XII, Issue III, March 2020, pp200 -204.

The International Conference on Research Perspectives : IoT in Hybrid Grid Integrated Renewable Energy Sources
In association with International Journal of Scientific Research in Computer Science, Engineering and Information Technology
ISSN : 2456-3307 (www.ijsrcseit.com)

Transient Stability Prediction using Artificial Neural Networks and Synchronized Measurements

Chandrashekhar P K[1], Dr. Srivani S G[2]

[1]Research Scholar, VTU Research Centre, Central Power Research Institute, Bengaluru, Karnataka, India

[2]Associate Professor, Department of E & EE, RV College of Engineering, Bengaluru, Karnataka, India

ABSTRACT

In this paper an post-fault transient stability assessment (TSA) method in terms of transient stability margin (TSM) prediction using artificial neural networks (ANN) and synchronized or PMU measurements is proposed. A post-fault multi-machine system is converted into a suitable OMIB using Single Machine Equivalent (SIME) concept. By using SIME P_a-δ trajectory, a normalized transient stability margin is calculated. By using pre and during fault synchrophasor measurements as input ANN model is trained to predict normalized stability margin. Using the synchronized measurements available at generator buses and the trained ANN model, post-fault TSA is carried out in terms of TSM prediction. If the predicted margin is negative then the post-fault system is declared unstable and if the predicted margin is positive then the system is declared stable. The proposed assessment method is implemented using New England 39 bus test system. The results are compared with time domain simulations.

Keywords- Artificial Neural Networks (ANN), Single Machine Equivalent (SIME), Transient Stability Assessment (TSA), Transient Stability Margin (TSM)

I. INTRODUCTION

Post fault transient stability assessment after the occurrence of a severe contingency in present day modern interconnected power system plays an important role. The prediction of post fault system status well in advance plays an important role from the operator point of view. If the operator has the information of the post fault power system status immediately after the fault clearance then there will be sufficient amount of time left for corrective action. Thus the prediction of post fault power system plays an important role from the power system operation point of view. By knowing the post fault system status, suitable corrective actions can be adopted to bring the system back into stable operating condition. Transient stability of a power system is related with the ability of that system to remain in synchronism after subjected to a large disturbance as stated above [1-3]. TSA methods are widely classified as time domain, energy function and artificial intelligence (AI) methods. Time domain simulation (TDS) method is the practical method used for accurate assessment of the complicated power systems with detail modeling of its components. It requires more time for solving large number of equations. Energy function (EF) methods also known as direct methods [4-6] overcome this heavy computation burden of TDS

method, but are also have problem of limited scalability and their conservativeness.

Hybrid methods which are derived by combining both TDS and EF methods [7-12] improved the performance of TEF methods. These hybrid methods also improved the performance of TEF methods but the errors present due to different models and contingencies still present. Another category of TSA methods are AI or machine learning approaches such as artificial neural networks (ANN), decision trees (DT), support vector machines (SVM), fuzzy based systems, extreme learning machines (ELM) etc. With good number of offline training cases these methods can assess the stability of the system with good accuracy. With the help of synchronized measurements these machine learning methods are becoming more useful in online applications. These methods are faster and more accurate compared to other methods.

In the proposed approach single machine equivalent (SIME) method is used to transform a multi machine system into its equivalent OMIB system. Then using the accelerating and decelerating area of the equivalent OMIB P_a-δ trajectory, a normalized transient stability margin (TSM) is defined. The stability margin not only gives the stability status but also gives the information about severity of the contingency. ANN models are developed and trained to predict the TSM. The simulations are carried out using Matlab based simulation packages [13-15].

II. SINGLE MACHINE EQUIVALENT (SIME) METHOD

SIME method [7] is a hybrid method which is derived by combining time domain simulation and equal area criterion concept. Irrespective of the size and complexity the power system, any multi-machine system can be reduced into its equivalent OMIB with the help of suitable equations. Using this method we can also assesses the system in terms of CCT, stability margin and contingency ranking. Different approaches are used to convert multi machine system into two machine groups, critical and non-critical machine groups. Maximum post-fault rotor angle difference criteria is the most used one. Critical machines (CM) are the one which swing together and are likely to lead the post-fault power system into unstable condition. Non-critical machines (NM) are the one which swing together and remain in stable condition even after the power system becomes unstable.

The expressions for OMIB parameters are given below.

The rotor angles and speed for critical machines, C:

$$\delta_C(t) = \frac{1}{M_C}\sum_{k\in C} M_k\, \delta_k(t$$

(1)

$$\omega_C(t) = \frac{1}{M_C}\sum_{k\in C} M_k\, \omega_k(t)$$

(2)

For non-critical machines, N:

$$\delta_N(t) = \frac{1}{M_N}\sum_{j\in N} M_j\, \delta_j(t)$$

(3)

$$\omega_N(t) = \frac{1}{M_N}\sum_{j\in N} M_j\, \omega_j(t)$$

(4)

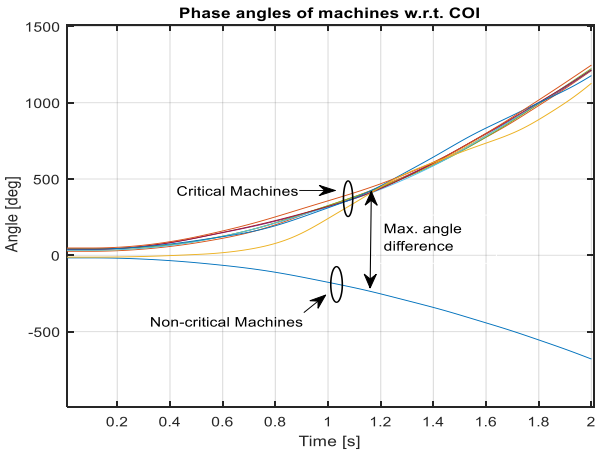


Figure 1. Grouping of machines

The equivalent OMIB parameters:

Rotor angle:

$$\delta_{OMIB} = \delta_C(t) - \delta_N(t)$$

(5)

Rotor speed:

$$\omega_{OMIB} = \omega_C(t) - \omega_N(t)$$

(6)

Mechanical power:

$$P_m(t) = M\left(\frac{1}{M_C}\sum_{k\in C} P_{mk}(t) - \frac{1}{M_N}\sum_{j\in N} P_{mj}(t)\right)$$

(7)

Electrical power:

$$P_e(t) = M\left(\frac{1}{M_C}\sum_{k\in C} P_{ek}(t) - \frac{1}{M_N}\sum_{j\in N} P_{ej}(t)\right)$$

(8)

Accelerating power:

$$P_a(t) = P_m(t) - P_e(t)$$

(9)

$$M_C = \sum_{k\in C} M_k\,; \quad M_N = \sum_{j\in N} M_j\,; \quad M = \frac{M_C M_N}{M_C + M_N}$$

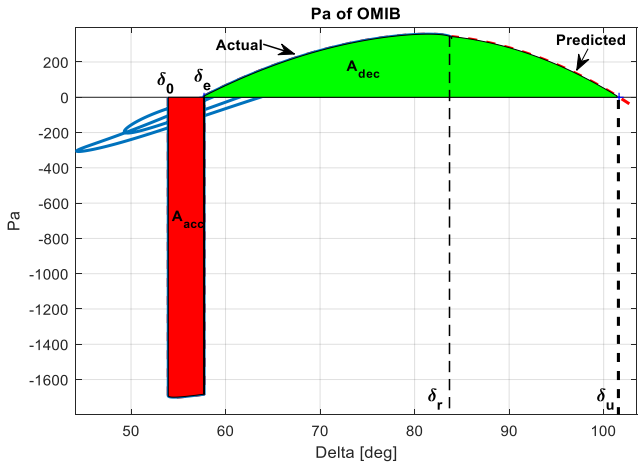


Figure 2. SIME OMIB P_a-δ trajectory for stable case

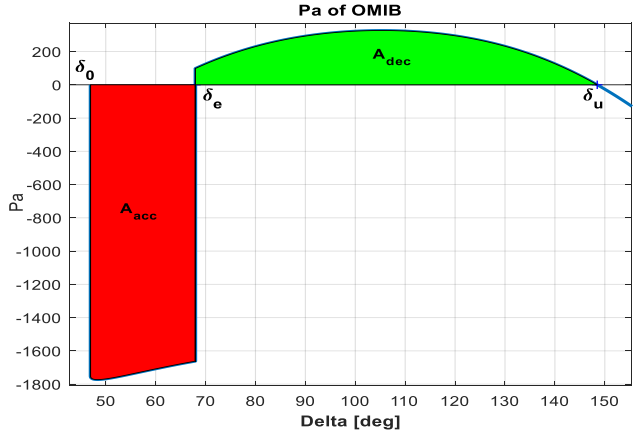


Figure 3. SIME OMIB P_a-δ trajectory for unstable case

III. TRANSIENT STABILITY MARGIN (TSM)

In the proposed approach the it is assumed that PMUs are placed on all the generator buses and their synchronized measurements are available[7]. Using the OMIB P_a-δ trajectory of SIME, a normalized transient stability margin (TSM) is defined as given below.

$$A_{acc} = \int_{\delta_0}^{\delta_e} P_a \, d\delta$$

(10)

$$A_{dec} = \int_{\delta_e}^{\delta_u} P_a \, d\delta$$

(11)

where, δ_0 – fault starting instant

δ_e – fault clearing instant

δ_u – end of observation

Normalized accelerating power based transient stability margin (TSM),

$$\eta = \begin{cases} \frac{A_{dec}-A_{acc}}{A_{dec}} & \text{If } A_{dec} > A_{acc} \text{ (Stable)} \\ \frac{A_{dec}-A_{acc}}{A_{acc}} & \text{If } A_{acc} > A_{dec} \text{ (Unstable)} \end{cases}$$

(12)

The above defined normalized stability margin is positive and lies between 0 to 1 for stable cases where decelerating area is more than the accelerating area. It is negative and lies between -1 to 0 for unstable cases. Thus the stability margin lies between -1 to 1.

IV. DEVELOPMENT OF ANN FOR TRANSIENT STABILITY PREDICTION

With the advent of synchronized measurements (PMU) and availability of online measurements it is easy to predict the post-fault transient stability. A two layer feed forward neural network model used for prediction of TSM is as shown in figure 4.

A. Inputs

The generator connected to the bus with highest difference between pre and during fault voltage

magnitude is termed as severely disturbed generator (SDG). The bus voltage magnitudes at four different instances and rotor angles at two instances of this SDG are chosen as inputs to the ANN model. The instances considered for synchronized measurements are (i) just before fault starting (ii) fault starting (iii) fault clearing (iv) immediately after fault clearing. The bus voltage magnitudes at all four instances and rotor angles at instants (i) and (iii) are chosen as inputs. The input vector is X = {V_1 V_2 V_3 V_4 δ_1 δ_3}.

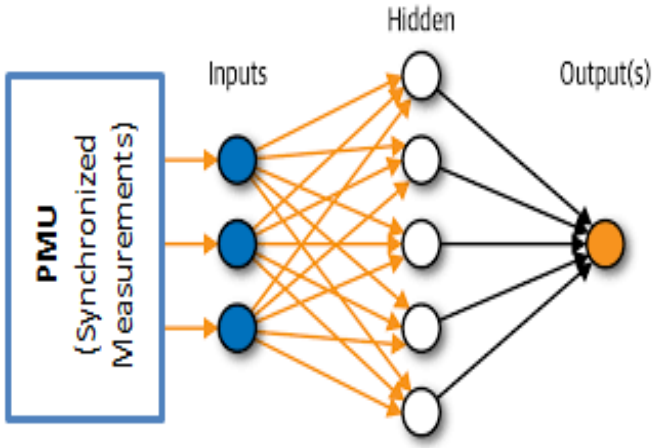


Figure 4. Proposed Artifitial Neural Network

B. Output

The output of the ANN consists of one neuron representing the post-fault system status in terms of TSM. The value of TSM and its indication is as follows.

$$TSM(\eta) = \begin{cases} 0 \text{ to } 1 \rightarrow \text{Stable} \\ -1 \text{ to } 0 \rightarrow \text{Unstable} \end{cases}$$

V. SIMULATION RESULTS

The proposed approach is tested by applying it to the New England 39 Bus system shown in figure 5. The power flow simulations were carried out using MATPOWER [13]. The transient simulations were carried out using MatDyn [14] package. The ANN simulations were done using Neural Networks toolbox from Matlab [15]. The base case load was varied from 80 to 120% and for each scenario, a three phase fault at a bus is considered. The fault was cleared by removing the connected line. For each fault, the fault clearing time was varied from 5 to 10 cycles. Thus totally 2446 valid cases were generated for training the ANN model. Out of the total 2446 cases 1545 were stable and remaining 901 cases were unstable.

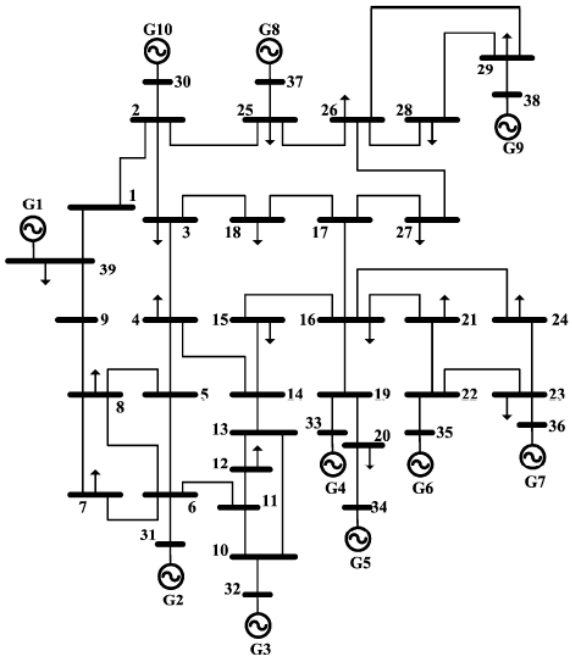


Figure 5. New England 39 Bus Test System

The structure of the ANN model with six inputs and one put is as shown in figure 6. The details of the trained Multi-Layer Perceptron Neural Network (MLPNN) are as given in table I. The input data of the ANN is divided into three parts, training, validation and testing data. 70% data is chosen for training, 15% for validation and remaining 15% is used for testing the ANN model. Different regression plots of the ANN model are as shown in figures 7 to 10. The results obtained for different contingencies by the ANN model are listed in table II. From the results it can be seen that the trained ANN model able to predict the TSM with good accuracy. From the predicted TSM value the post-fault transient stability status can be identified accurately.

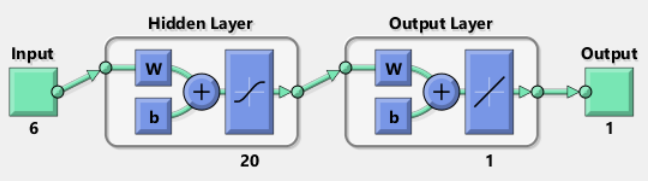


Figure 6. Structure of ANN model

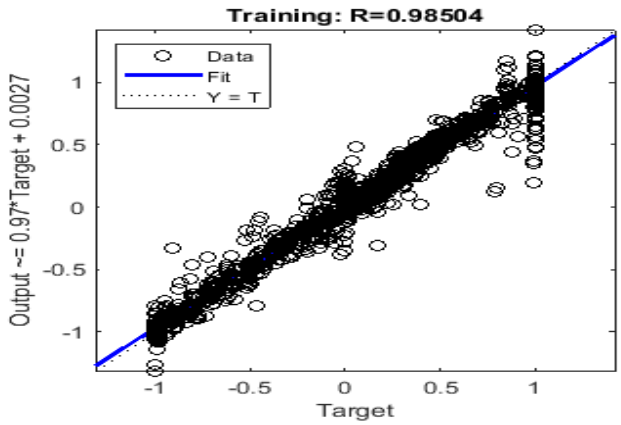


Figure 7. Training regression plot

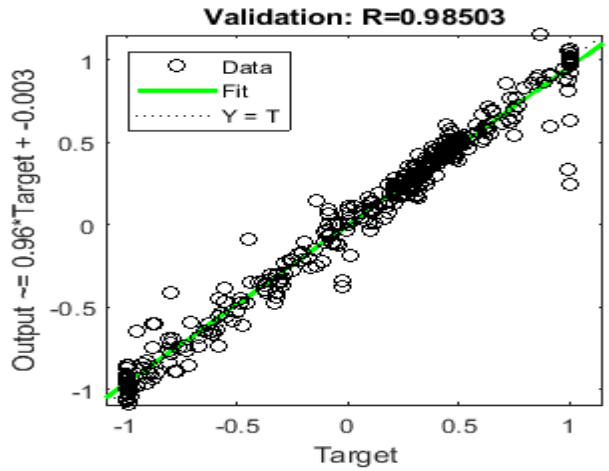


Figure 8. Validation regression plot

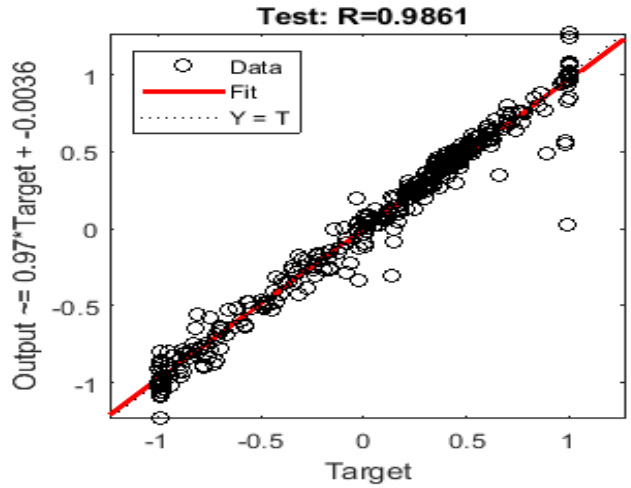


Figure 9. Test regression plot

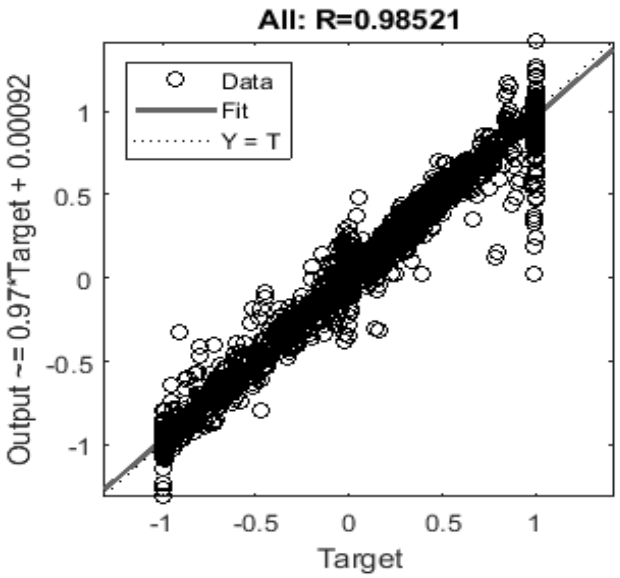


Figure 10. Overall regression plot of the ANN model

TABLE I. STRUCTURE OF MLPNN

ANN Model	Details
MLPNN	No. hidden layers = 1 No. neurons in hidden layers = 20 Learning rate = 0.02 Learning Algorithm: Levenberg–Marquardt (LM)

VI. CONCLUSION

In this paper an approach for transient stability assessment in terms of transient stability margin prediction using ANN and synchronized measurements is presented. The proposed approach is tested on New England 39 bus test system and shows good performance. From the results it is observed that with the availability of synchronized measurements from devices such as PMUs, the ANN based approach is fast, accurate and can also be used for online transient stability assessment.

TABLE II. PREDICTION RESULTS

Faulte	Remov	Cleari	TSM(η)	Assessme

d Bus	ed Line	ng Time (cycles)	Actual	ANN	nt
1	1-2	8	0.7610	0.7633	Stable
1	1-39	10	0.6166	0.6042	Stable
6	6-7	5	0.2897	0.2820	Stable
10	10-13	9	-0.8207	-0.8548	Stable
16	16-17	7	-0.9886	-0.9669	Unstable
17	17-27	6	-0.4718	-0.4395	Unstable
26	26-28	5	-0.2059	-0.2104	Unstable
28	28-29	10	0.9996	0.9880	Stable

VII. REFERENCES

[1]. IEEE/CIGRE Joint Task Force. Definition and classification of power system stability. IEEE Trans Power Syst 2004: 1(2).

[2]. Kundur P. Power system stability and control. McGraw-Hill; 1994.

[3]. Sauer P W, Pai M A. Power system dynamics and stability. New Jersey: Prentice-Hall; 1998.

[4]. Pai M A. Energy function analysis for power system stability. Kluwer Academic; 1989.

[5]. Fouad A A, Vittal V. Power system transient stability analysis using the transient energy function method. Prentice-Hall; 1992

[6]. Xue, Yusheng, Thierry Van Custem, and Mania Ribbens-Pavella. "Extended equal area criterion justifications, generalizations, applications." IEEE Transactions on Power Systems 4.1 (1989): 44-52.

[7]. Mania Pavella, Damien Ernst, and Daniel Ruiz-Vega. Transient stability of power systems: a unified approach to assessment and control. Vol. 581. Springer, 2000.

[8]. Wadduwage, D. Prasad, Udaya D. Annakkage, and Christine Qiong Wu. "Hybrid algorithm for rotor angle security assessment in power systems." The Journal of Engineering 2015.8 (2015): 241-251.

[9]. Zhang, Ywee, et al. "SIME: A hybrid approach to fast transient stability assessment and contingency selection." International Journal of Electrical Power & Energy Systems 19.3 (1997): 195-208.

[10]. Fang, D. Z., T. S. Chung, and A. K. David. "Improved techniques for hybrid method in fast-transient stability assessment." IEE Proceedings-Generation, Transmission and Distribution 144.2 (1997): 107-112.

[11]. Fang, D. Z., et al. "Improved hybrid approach to transient stability assessment." IEE Proceedings-Generation, Transmission and Distribution 152.2 (2005): 201-207.

[12]. Maria, G. A., C. Tang, and J. Kim. "Hybrid transient stability analysis (power systems)." IEEE Transactions on Power Systems 5.2 (1990): 384-393.

[13]. R. D. Zimmerman, C. E. Murillo-Sanchez (2015) MATPOWER [Software]. Available: https://matpower.org

[14]. MatDyn website. [Online]. Available: http://www.esat.kuleuven.be/electa/teaching/matdyn/

[15]. MATLAB. (2019). version 9.6.0 (R2019a). Natick, Massachusetts: The MathWorks Inc.

The International Conference on Research Perspectives : IoT in Hybrid Grid Integrated Renewable Energy Sources
In association with International Journal of Scientific Research in Computer Science, Engineering and Information Technology
ISSN : 2456-3307 (www.ijsrcseit.com)

Prediction of opening and closing of Company Stocks Using Machine Learning

Ashlesha Bhagat, Prof. Parnal P. Pawade
Department of Information Science and Engineering P.R. Patil College of Engineering & Technology, Amravati, India

ABSTRACT
In the finance world stock trading is one of the most important activities. Stock market prediction is an act of trying to determine the future value of a stock other financial instrument traded on a financial exchange . The technical and fundamental or the time series analysis is used by the most of the stockbrokers while making the stock predictions. In Prediction of stocks, the aim is to predict the future opening and closing of company stocks value of the financial stocks of a company. The recent trend in stock market prediction technologies is the use of machine learning which makes predictions based on the values of current stock market indices by training on their previous values. a Machine Learning ML approach that will be trained from the available stocks data and gain intelligence and then uses the acquired knowledge for an accurate prediction. In this context this study uses a machine learning technique called Support Vector Machine SVM to predict stock prices for the large and small capitalizations and in the three different markets, employing prices with both daily and up to the minute frequencies. Machine learning itself employs different models to make prediction easier and authentic. The paper focuses on the use of Regression and LSTM based Machine learning to predict stock values. Factors considered are open, close, low, high and volume.

Keywords - Close, high, low, SVM, LSTM model, open, regression, and volume.

I. INTRODUCTION

A correct prediction of stocks can lead to huge profits for the seller and the broker. Frequently, it is brought out that prediction is chaotic rather than random, which means it can be predicted by carefully analyzing the history of respective stock market. Machine learning can be defined as the data which is obtained by knowledge extraction. Machines don"t have to be programmed explicitly instead they are trained to make decisions that are driven by data. Instead of writing a code for every specific problem, data is provided to the generic algorithms and logic is developed on the basis of that data. When a machine improves its performance based on its past experiences it can be said that machine has truly learnt.

The technique for most accurate prediction is by learning from past instances, and to make a program to do this is best possible with machine learning techniques. Machine learning is an efficient way to represent such processes. It predicts a market value close to the tangible value, thereby increasing the accuracy. Introduction of machine learning to the area of

stock prediction has appealed to many researches because of its efficient and accurate measurements [1] [2].

The vital part of machine learning is the dataset used. The dataset should be as concrete as possible because a little change in the data can perpetuate massive changes in short
outcome. In this project, supervised machine learning is employed on a dataset obtained from Yahoo Finance. This dataset comprises of following five variables: open, close, low, high and volume. Open, close, low and high are different bid prices for the stock at separate times with nearly direct names. The volume is the number of shares that passed from one owner to another during the time period. The model is then tested on the test data.

Regression and LSTM models are engaged for this conjecture separately. Regression involves minimizing error and LSTM [3] [4] contributes to remembering the data and results for the long run. Finally, the graphs for the fluctuation of prices with the dates (in case of Regression based model) and between actual and predicted price (forthe LSTM based model) are plotted. The rest of the paper consists of following: Section II discusses the related work. Section III puts forward the two models used and the methods used in them in detail. Section IV discusses the results produced with different plots for both the models in detail. While Section V consists of conclusion and the last section involves the references.

II. RELATED WORK

From the literature survey, it was observed that the application of machine learning techniques to stock market prediction is being undertaken thoroughly throughout the world. Machine Learning techniques are proving to be much more accurate and faster as compared to contemporary prediction techniques.

Significant work has been done throughout the world in this field. A testament to which is the work of M. Usmani, S. H. Adil, K. Raza and S. S. A. Ali [1] and that of K. Raza
[2] who have surveyed the application of machine learning techniques and presented the current advancements in this field.

H. Gunduz, Z. Cataltepe and Y. Yaslan [3] predicted stock prices using deep neural network techniques. Similarly, M. Billah, S. Waheed and A. Hanifa [4] suggested further improvements to stock prediction using

neural networks through the use of a training algorithm which they designed on their own. K. V. Sujatha and S. M. Sundaram [6] suggested insightful techniques on handling non-normal situations which may often arise during the working of the system and cause disruptions or lead to inaccurate predictions.

Liu, G. Liao and Y. Ding [7] conducted similar work and designed a model for applying LSTM to stock prediction with lots of scope for improvements to prediction accuracy. K. A. Althelaya, E. M. El-Alfy and S. Mohammed [9] further contributed to the field by staging experiments and simulations to assess the feasibility of applying deep learning techniques to prediction of stock prices.

III. METHODOLOGY

Stock market prediction seems a complex problem because there are many factors that have yet to be addressed and it doesn't seem statistical at first. But by proper use of machine learning techniques, one

can relate previous data to the current data and train the machine to learn from it and make appropriate assumptions.

The dataset being utilized for analysis was picked up from Yahoo Finance. The dataset consisted of approximately 9 lakh records of the required stock prices and other relevant values. The data reflected the stock prices at certain time intervals for each day of the year. It consisted of various sections namely date, symbol, open, close, low, high and volume. For the purpose of simulation and analysis, the data for only one company was considered. All the data was available in a file of csv format which was first read and transformed into a data-frame using the Pandaslibrary in Python. From this, the data for one particular company was extracted by segregating data on the basis of the symbol field. Following this normalization of the data was performed through usage of the sklearn library in Python and the data was divided into training and testing sets. The test set was kept as 20% of the available dataset.

Although machine learning as such has many models but this paper focuses on two of the most important amongst them and made the predictions using these.

A. Regression Based Model

In general, the Regression based Model is used for predicting continuous values through some given independent values [5]. Regression uses a given linear function for predicting continuous values:

൭·൯ (1)

Where, V is a continuous value; K represents known independent values; and, a, b are coefficients.

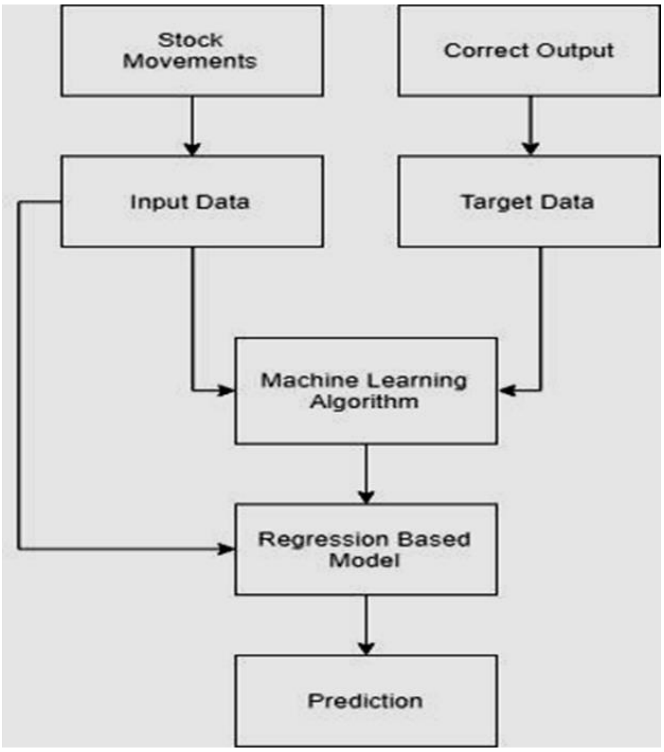


Fig. 1 Flow Chart for Regression Based Model

corresponds to the problem of Vanishing Gradient. LSTM prevents this from happening. The LSTM consists of a remembering cell, input gate, output gate and a forget gate. The cell remembers the value for long term propagation and the gates regulate them.

The paperutilizes the gradient descent linear regression algorithm for predicting correct values by minimizing the error function as given in Fig.1.

Linear Regression [6]as governed by the above equation is performed on the data and then the relevant predictions are made. The factors considered for the regression were low, open, high, close and volume. The R-square confidence test was used to determine the confidence score and the predictions were plotted to show the results of the stock market prices vs time.

B. Long Short Term Memory (LSTM) Network Based Model

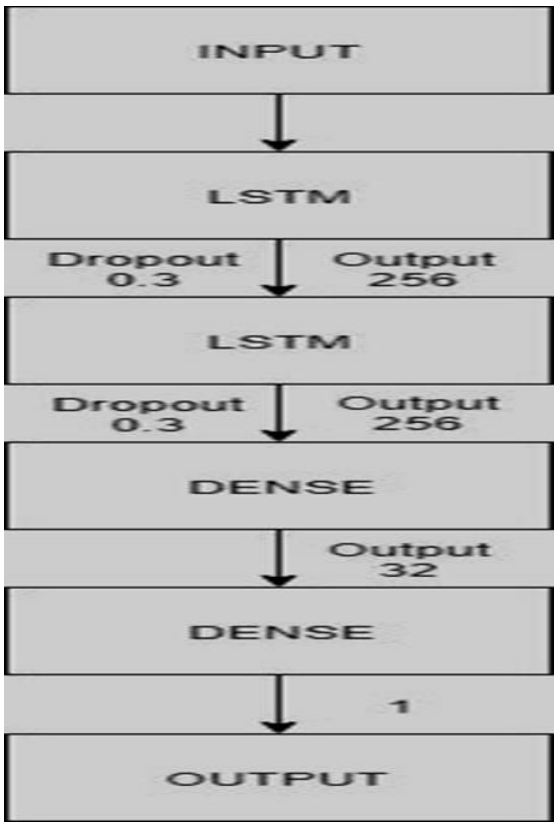


Fig. 2 LSTM Layers

LSTM is the advanced version of Recurrent-Neural-Networks (RNN) where the information belonging to previous state persists. These are different from RNNs as they involve long term dependencies and RNNs works on finding the relationship between the recent and the current information. This indicates that the interval of information is relatively smaller than that to LSTM.

The main purpose behind using this model in stock market prediction is that the predictions depends on large amounts of data and are generally dependent on the long term history of the market [6]. So LSTM regulates error by giving an aid to the RNNs through retaining information for older stages making the prediction more accurate [7]. Thus proving itself as much more reliable compared to other methods.

Since stock market involves processing of huge data, the gradients with respect to the weight matrix may become very small and may degrade the learning rate.[8].This

In this paper, a sequential model has been made which involves stacking two LSTM layers on top of each other with the output value of 256. The input to the layer is in the form of two layer [0] and layer[1]. A dropout value of 0.3 has been fixed which means that 0.3 out of total nodes will be frozen during the training process to avoid over-fitting of data and increase the speed of the training process. At last, the core dense layer where each neuron is connected to every other in the next layer is added providing input of 32 parameters to the next core layer which gives output as 1. The model is compiled with a mean square cost function to

maintain the error throughout the process and accuracy is chosen as a metric for the prediction [9].

IV. EXPERIMENTAL RESULTS

The proposed system is trained and tested over the dataset taken from Yahoo Finance. It is split into training and testing sets respectively and yields the following results upon passing through the different models:

A. Regression Based Model Results

The plot in figure3 is the result of application of linear regression algorithm on the dataset to predict varying prices with respect to the time.

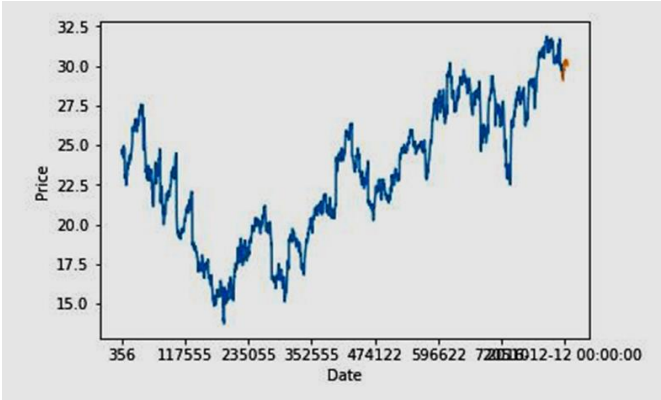


Fig.3 Plot between Price and Date Using Regression

The above graph Fig. 3 is plot over the data having batch size 512 and 90 epochs. The R-square confidence test resulted in a confidence score of 0.86625.

B. LSTM Based Model Results

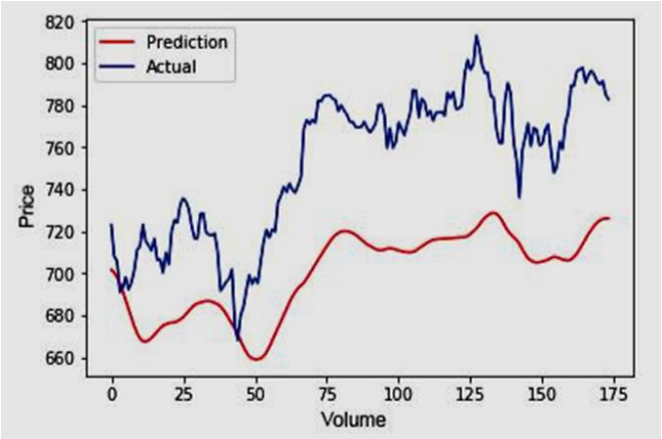


Fig. 4 Plot between Actual & Predicted Trend of LSTM

The prediction is shown by red line and the actual trend is shown by blue. The proximity of these two lines tells,
how efficient the LSTM based model is. The prediction approximates real trend when a considerable amount of time has passed. The model resulted in a Train Score of 0.00106 MSE (0.03 RMSE) and a Test Score of 0.00875 MSE (0.09 RMSE). The more thesystem is trained and the greater the size of the dataset utilized the greater the accuracy which will be attained. The LSTM Model offered more accuracy than the Regression based Model.

V. CONCLUSION

This paper was an attempt to determine the future prices of the stocks of a company with greater accuracy and reliability using machine learning techniques. The primary contribution of the researchers being the application of the novel LSTM Model as a means of determining the stock prices.

Both the techniques have shown an improvement in the accuracy of predictions, thereby yielding positive results with the LSTM model proving to be more efficient. The results are quite promising and has led to the conclusion that it is possible to predict stock market with more accuracy and efficiency using machine learning techniques.

In the future, the accuracy of the stock market prediction system can be further improved by utilizing a much bigger dataset than the one being utilized currently. Furthermore, other emerging models of Machine Learning could also be studied to check for the accuracy rate resulted by them. Sentiment analysis though Machine Learning on how news affects the stock prices of a company is also a very promising area. Other deep learning based models can also be used for prediction purposes.

VI. REFERENCES

[1]. M. Usmani, S. H. Adil, K. Raza and S. S. A. Ali, "Stock market prediction using machine learning techniques," 2016 3rd International Conference on Computer and Information Sciences (ICCOINS), Kuala Lumpur, 2016, pp. 322-327.

[2]. K. Raza, "Prediction of Stock Market performance by using machine learning techniques," 2017 International Conference on Innovations in Electrical Engineering and Computational Technologies (ICIEECT), Karachi, 2017, pp. 1-1.

[3]. H. Gunduz, Z. Cataltepe and Y. Yaslan, "Stock market direction prediction using deep neural networks," 2017 25th Signal Processing and Communications Applications Conference (SIU), Antalya, 2017, pp. 1-4.

[4]. M. Billah, S. Waheed and A. Hanifa, "Stock market prediction using an improved training algorithm of neural network," 2016 2nd International Conference on Electrical,

Computer & Telecommunication Engineering (ICECTE), Rajshahi, 2016, pp. 1-4.

[5]. H. L. Siew and M. J. Nordin, "Regression techniques for the prediction of stock price trend," 2012 International Conference on Statistics in Science, Business and Engineering (ICSSBE), Langkawi, 2012, pp. 1-5.

[6]. K. V. Sujatha and S. M. Sundaram, "Stock index prediction using regression and neural network models under non normal conditions," INTERACT-2010, Chennai, 2010, pp. 59-63.

[7]. S. Liu, G. Liao and Y. Ding, "Stock transaction prediction modelling and analysis based on LSTM," 2018 13th IEEE Conference on Industrial Electronics and Applications (ICIEA), Wuhan, 2018, pp. 2787-2790.

[8]. T. Gao, Y. Chai and Y. Liu, "Applying long short term memory neural networks for predicting stock closing price," 2017 8th IEEE International Conference on Software Engineering and Service Science (ICSESS), Beijing, 2017, pp. 575-578.

[9]. K. A. Althelaya, E. M. El-Alfy and S. Mohammed, "Evaluation of bidirectional LSTM for short-and long-term stock market prediction," 2018 9th International Conference on Information and Communication Systems (ICICS), Irbid, 2018, pp. 151-156

The International Conference on Research Perspectives : IoT in Hybrid Grid Integrated Renewable Energy Sources
In association with International Journal of Scientific Research in Computer Science, Engineering and Information Technology
ISSN : 2456-3307 (www.ijsrcseit.com)

An Enhanced ANTSEC Framework with Clustering based Cooperative Caching in the Mobile AdHoc Networks

P. A. Gaikwad, Dr. S. S. Sherekar, Dr. V.M. Thakre
SGBAU, Amravati, Maharashtra, India

ABSTRACT

The gadgets in versatile mobile Ad Hoc Network (MANETs) are for the most part fueled by battery.Battery limit is fixed and a few procedures needed to save energy at the gadget level or at the convention stack ought to be applied to improve the MANETs life time. In climate, a portability hub is capricious; this is considered as a trait of remote organizations. Due to the broken or malignant hubs, the organization is helpless against steering mis behavior. The asset compelled attributes of MANETs prompts expanded inquiry delay at season of information access. In this paper, AntHocNet+Security (ANTSEC) structure is recommended that incorporates an upgraded collaborating storing plan implanted with counterfeit resistant framework. In this system improves security by infusing insusceptibility into data packets, to improves the parcel conveyance proportion and lessens start to finish defer utilizing cross layer plan.

Keywords: mobile Ad Hoc Network, Ant Hoc Net + Security

I. INTRODUCTION

MANET are an unstructured there is no focal organization to the oversee hub portability, recognition and avoidance of irregularities. MANETs are the exceptionally powerless towards security and the organization flaws. The security incorporates not just giving assurance against known and obscure an assaults however includes infusing invulnerability into the information parcels or portable nodes. The group of correspondence has been essential for many applications in MANETs. The normal number of the user support by the organization have become progressively assets serious. These thus, has increased to the significance of data transmission effectiveness in the MANETs. It is urgent for medium access control convention of MANET adjust to the powerful climate as well as in the effectively oversaw transfer speed use. In this paper strategy proposed is Ant Hoc Net+ Security (ANTSEC) outline work is proposed in that incorporates an upgraded helpful storing plan inserted with counterfeit insusceptible framework. In this outline work improves security by infusing invulnerability into information parcels, improves the bundle conveyance proportion and to diminish start to finish postpone utilizing the cross layer plan. The issue hub disappointment and the hub glitch are tended to in the reserve management. As portable hubs in impromptu organizations may have comparative assignments and offer the regular interest, agreeable storing, which allows the sharing and coordination of stored information among different hubs, can be used to diminish the exchange speed and a force use. The Mobile gadgets are not

needed to send solicitations to the data source as they are share data with the adjoining hubs.

II. PREVIOUS WORK DONE

G. Radhamaniet al.,[2012] in this paper an improved ACO based calculation for upgrading execution in remote specially appointed organization. ACO is presented counterfeit resistant framework (AIS)has demonstrated to give adequate execution against anamolies anyway hub disappointment cause overhead which can be decreased by actualizing cross layer designing. B. Sateesh, etal.,[2012]in this paper novel deviated helpful reserve approach utilizing a powerful source directing convention was suggested that presents an agreeable store component in shared remote organizations. V. M. Thakare ,et al.,[2012]in this paper investigation of interruption location methods in MANET an interruption recognition framework (IDS)is needed in a MANET by both legitimate organization hubs and malevolent hubs to screen the arrange and distinguish the miss conduct and oddities. S.M. Iyengar et al.,[2010] in this paper is calculation has the property of flexibility and responsiveness to adjusting ecological conditions in different issue domains, it can be applied MANET steering problem.N. Chauhanand et al.,[2012]in this paper Prefetching dependent on agreeable storing in portable impromptu networks. N Shrivastava Ant province advancement with order calculations utilized for intrusion location International J. Computational Engineering and Management The pheromone table update is performed concurring an Artificial Immune System. The idea of AIS fused in the proposed structure is like the human safe framework that shields to the human from microorganisms, for example, infections and microbes. In this paper, Ant Hoc Net + Security (ANTSEC) structure is proposed with cross-layer plan that improves security by infusing invulnerability into the information bundles, expands information openness and decreases question delay. The primary segments of the proposed structure are the improved co-operative reserving plan (ECOCA) middleware and the stack profile (cross layer).

III. LITERATURE SURVEY

M.K Denko et al., [2009] proposed Cluster-based cross-layer plan for helpful reserving in portable impromptu organization. The primary segment proposed outline work are the upgraded agreeable reserving plan (ECOCA) center product and stack profile (cross layer). H. Artial et al., [2008] proposed and intriguing element of subterranean insect settlements with regards to the searching conduct and, specifically, their capacity to find the most brief course between their nest and a food source understanding they are practically visually impaired. H. Artial et al., [2008] COACS An agreeable versatile storing framework for MANET. Hubs in insect hoc net forward information stochastically When a subterranean insect moves from the I to j hubs pheromone is stored. To one another curve (I,j)of a chart the measure of counterfeit pheromone saved ij This data can be perused and composed by the ants to administer their development to the following hub.

IV. PROPOSED METHODOLOGY

In this structure consolidates the highlights of the Ant Hoc Net convention, AIS, cross layer configuration bunching. The primary parts of the ANTSEC structure are ECOCA middleware and the stack profile (cross layer).Ant Hoc Net directing convention is assessed with an AIS and the cross layer plan ECOCA. The framework a design is portrayed in the Fig. 1. ECOCA is a middleware that executes under an Application Layer and to gives functionalities, for example, reserving, information the executives, and security. The useful component of the ECOCA specialist incorporates data search,

caching/pre fetching, grouping, stack profile, and the client interface that capacities in a manners middleware. The ACO calculation utilizes control parcel called the subterranean insect agentto test conceivable a way towards an objective. In the ECOCA, the directing data got by the subterranean insect specialists is refreshed in the stack profile. The steering table holds a pheromone variable that are constantly refreshed by path quality values determined by the insect agents. When the insect specialists proceed onward a similar way consistently and simultaneously, the way testing age an outcome in the accessibility of a heap of ways at every hub with an expected proportion of quality. Middleware.

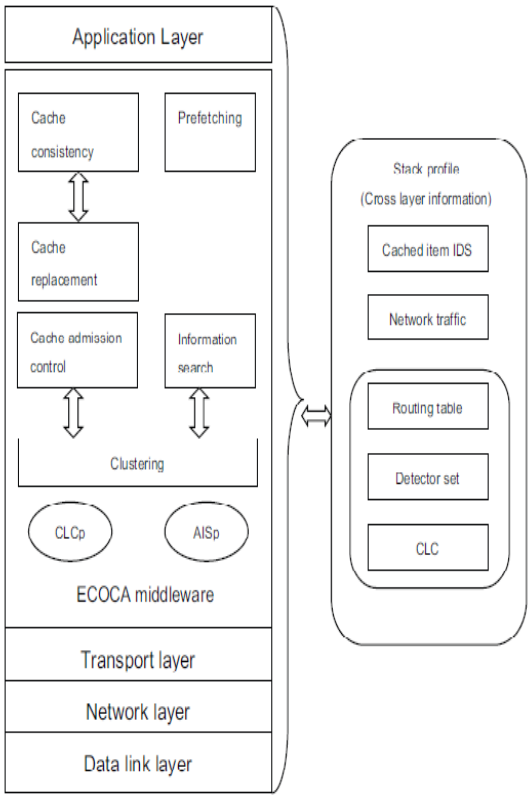


Fig. 1. System architecture of ANTSEC framework.

V. TOOLS USED

Network Simulator (NS2)

VI. Data Set Used:

Kaggel kiva

VII. MATHEMATICAL MODEL

Procedure AIS (Cb, **detector set)**
Begin
If the Cb == Db
then
Discard current behaviour
Else
Propagate
End
Procedure CLC (datagram)
Begin
Get pheromone value from Routing Table
then
If *ph value* < 30% then
S is high

1. Get BP from source cache
2. Send BP directly to intermediate node where path distraction occurred
3. Intermediate node reroute to destination through BP

Else if *ph value* >30% and *ph value* <60% then
S is moderate
Propagate until time T
If T is expired then
Get the Correct Path from CH
Reroute to the Correct Path
End if
End if
End

8:IMPLIMENTATION OF PROPOSED METHOD

ECOA work

Step 1 : To receives DP from application layer.
Step 2 : Anembed AISp and CLCp alongside DP.
Step 3 : During course disclosure stage, Ant Hoc Net convention proliferates and the finds briefest way. Subterranean insect chooses the best way dependent on the nature of pheromone esteem between hubs. After spread of mode the ants move to progress mode. When first

DP arrives at the objective, BP is put away in source reserve.

Step 4 : Detector set is acquired from the negative selection algorithm. It contains unusual conduct.

Step 5 : This Severity depends on the amount of pheromone value.

Step 6 : AISp gets enacted when a halfway hub attempts to open the information.

Step 7 : CLCp gets initiated when there is less amount of pheromone esteem in a specific hub. It is pass the control to CLC.

VIII. EVALUTION FURTHERMORE, RESULTS

The proposed an ANTSEC system was assessed in the NS-2 recreation climate [20]. Apathetic reenactment situations were performed. The exhibition measurements used to assess the proposed structure were PDR and start to finish delay. In recreation, every portable host moves in the reenactment region following the irregular waypoint portability model. The arbitrary waypoint model is utilized for the reproducing the development example of versatile hosts in a MANET. For 100 and 150 hubs reproduction was acted in a 1,500 m × 1,000 m region. The reenactment parameters are recorded in Table 2. An Ant Hoc Net convention was utilized in the recreation. The accompanying presentation measurements were utilized in the simulation tests.

- PDR: The proportion between the quantity of information bundles got and the quantity of an information parcels sent.
- End-to-end delay: The normal time delay for getting an information bundle. To assess the exhibition of the proposed ANTSEC system, two situations dependent on some of the versatile hubs were thought of.

Table 4. Statistical analysis for 100 nodes.

100 Nodes	Mean	SD	t-value	df	p-value
PDR-ANTSEC vs. PDR-AODV+COCA	4.55	1.33	7.633	4	0.002
DELAY-ANTSEC vs. DELAY-AODV+COCA	-35.82	19.81	-4.043	4	0.016

Table 5. Performance on 150 nodes.

	PDR		Delay	
150 Nodes Sce-nario 2	ANTSEC	AODV+ COCA	ANTSEC	AODV+ COCA
1	93.99	90.18	153.31	171.29
2	91.13	87.61	269.93	304.04
3	84.76	81.45	304.68	337.82
4	96.20	92.60	205.63	239.97
5	95.22	92.12	170.69	207.74
Mean	92.26	88.79	220.85	252.17
SD	4.60	4.55	64.70	68.36

IX. PROS

1. This can improve the self-resistance of information bundles and can diminish delay brought about by hub disappointment.
2. The fruitful conveyance of the information parcels with the capacity to shield themselves from a gentle assault makes the edge work more effective.

X. CONS

1. AIS is activated when a hub played out an unforeseen occasion that would be unfavorably impact the typical conduct.
2. The normal association lifetime under various portability models a practically the equivalent, which are demonstrates that the presentation grouping plan isn't tremendously influenced by the particular versatility models.

XI. APPLICATIONS

In this paper, the design of the ANTSEC framework was formulated. This can improve self-immunity of the data packets and decrease delay caused by node failure.

XII. CONCLUSION

In this paper, the plan of the ANTSEC system was defined. This can develop self-invulnerability the information bundles and decline delay brought about by hub disappointment. The effective conveyance of information parcels with the capacity to a shield themselves from gentle assaults makes the casing work more productive when contrasted with customary secure information transmission components utilizing AODV. The ANTSEC system executes AIS and ECOCA, which can a remarkable mix of information security and store the executives. The proposed grouping plan depends on the assessed versatility data and work can autonomously of outside frameworks (e.g., GPS). The proposed grouping plan can generously an improve network steadiness and the adaptability

FUTURE WORK

In future pre fetching scheme can be improved to pre fetch the data items when a network traffic is high, the clustering algorithm can be improved to accommodate a large network. We should like to study the clustering by taking into consideration the energy availability of anodes, the traffic load distribution over the network and a requirement of different types of traffic.

XIII. REFERENCES

[1]. Minming Ni , Zhangdui Zhong , and Dongmei Zhao, “MPBC: A Mobility Prediction-Based Clustering Scheme for Ad Hoc Networks”, IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, Vol.60.,No.9,November 2011,PP.4549-4559

[2]. BoraKaraoglu, Member , IEEE and Wendi Heinzelman , Senior Member,”Cooperative load balancing and dynamic channel allocation for cluster-based mobile Ad Hoc networks.”, IEEE TRANSACTIONS ON MOBILE COMPUTING.,Vol.14,No.5,May 2015,PP.951-963

[3]. M. Kaliappan , E. Mariappan , M. Viju Prakash ,and B. Paramasivan ,”Load balanced clustering technique in MANET using Genetic Algorithms “,.Defence science journal, Vol.66, No.3,May 2016,pp.251-258

[4]. V.S.Anitha , M.P. Sebastian ,”Dominating set-based , weighted and adaptive clustering algorithms for mobile ad hoc networks.”, .IET Communications, Vol.66, No.3,May 2016Proc.ietdl, 20oct 2009.,pp.1836-1853

[5]. B.Karaoglu , t. Numanoglu, and W. Heinzelman , “Analytical formance of soft clustering protocols” , Ad Hoc Networks.,vol.9.no.4, Jun.2011, pp.635-651

The International Conference on Research Perspectives : IoT in Hybrid Grid Integrated Renewable Energy Sources
In association with International Journal of Scientific Research in Computer Science, Engineering and Information Technology
ISSN : 2456-3307 (www.ijsrcseit.com)

IoT Data Link Layer Communication Protocols Frame Format in Controller Area Networks

Dr. Sreekanth R
Department of MCA, New Horizon College of Engineering, Bengaluru, Karnataka, India

ABSTRACT
The crucial part of in Internet of Things (IoT) is its protocols. Protocols specify the communication for each layer of the network where the data is exchanged as per the specified frame formats. The Open System Interconnection model has specified and well-defined frame format of data link layer on how the data bits are taken from physical layer and formed as frames. This frame has a header field, payload field and trailer fields. Data link layer (DLL) provides an error free data to the above layer i.e Network layer. Data link protocols provide framing, error detection, correction and flow control. IoT has various standards mentioned for data link protocol. These standards for protocols are provided by few international organizations which are known as Internet Engineering Task Force (IETF), Institute of Electrical and Electronics Engineers (IEEE) and International Telecommunication Union (ITU). In this paper we first study about the DLL protocols, frame format and then propose an additional payload field required for extended protocols frame format in controller area networks (CAN).

Keywords - Error detection, Error Control, Flow control, Framing, Protocols.

I. INTRODUCTION

Data Link Layer provides services to network layer by utilizing various communication protocols that are used in IoT. Smart devices which are connected to the network the data communication happens with the assistance of protocols. There are several standard protocols provided by the telecommunication organizations. Bluetooth is a short-range wireless communication which is present in most of smart devices.

Data transfer rate for Bluetooth provides 3 Mbps for a short range of meters. Zigbee is similar to Bluetooth technology and supports Mesh topology. In this technology the physical layer link with data link layer protocol standards are provided by IEEE 802.15.4.

Blue Tooth Low Energy (BLE) is used in Personal Area Networks. Z-wave provides wireless communication for longer range IoT. Similarly, many other protocols like Sigfox, RFID also can be used in many smart devices for IoT. CAN is a bus standard which is designed so that the devices and microcontrollers communicate with each other with host computer.

II. CAN DATA LINK LAYER

The data link layer protocols for controller area networks have certain unique features. Bit errors are quite common in the communication. All Single bit errors are detected in this layer. There is high probability that multi but errors are also detected. The data frame for the CAN is shown in the fig1.

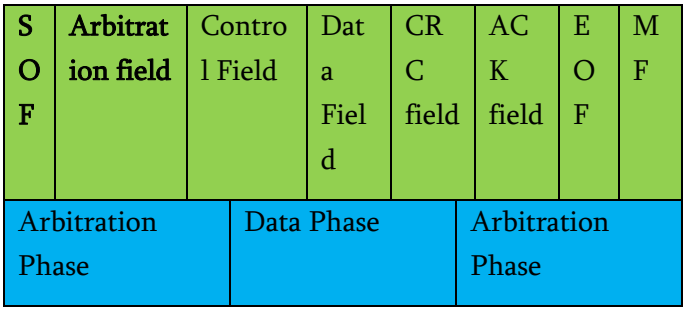


Fig 1: A CAN Data Frame

The transmission speed limit for CAN usually for short networks is in the range of 1 Mbps. Payload field in this frame is 8 bytes. CAN with Flexible Data Rate, CAN FD can have a payload filed till 64 bytes. The standardized frame format of CAN link layer is mentioned in ISO 11898. The implementation services of this layer is mentioned in Logical Link Control and Medium Access Control layers of CAN. Fig 2 shows the Data Link layer services.

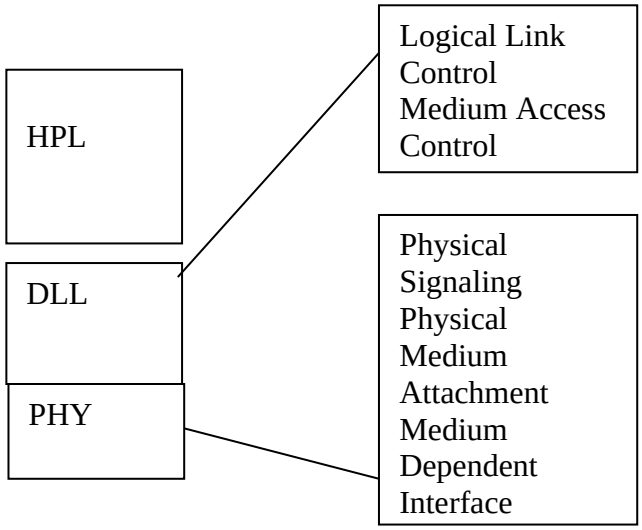


Fig 2. Data Link Layer Services

Some of the services like acceptance filtering, overload notification and recovery management are provided by the logical link control layer(LLC). Bit timing and Data Encapsulation is provided by medium access control layer .

MAC also responsible for the error detection, error signaling and acknowledgement. The data link layer should provide better communication services in order to provide an error free data to Network layer. there are two such services here in this layer. Write object service and Read Object service.

III. LITERATURE REVIEW

In [1] the authors proposed a common layered model which are used in standard models like OSI [2]. The model is also compared with the Transmission Control and Internet Protocol (TCP/IP) model [3]. The OSI model has seven layers while TCP/IP model has four layers. The authors proposed a protocol stack which require five layers for the IoT stack. In this layer the physical and data link layer provides the Radio Frequency capabilities.

In [4] Routing, security of data is provided by the Network layer in the IoT stack. Transport and application layers provide commands for the protocol. It was proposed to use few protocols like Open Stack for IoT, Blue Tooth Low Energy, Zigbee, Z-wave and Sigfox.

Each protocol mentioned has range, interoperability, security practices and topology type. In [5] the definition of Physical layer and data link layer are provided. In order to provide interoperability with non-IoT networks which are using IPV6, a layer called 6LoWPAN is introduced between data link layer and network layer. this layer ensures to provide interoperability in Non-IoT devices which will use the IPV6 address.

The Medium Access Layer is defined for two nodes which provides Reduce Function and Full Function

device. The major challenge here for IoT is to transpose the IP network to networks which are using Low power wireless personal Networks.

In order to achieve this in [6] Internet Engineering Task Force provide a protocol named 6LoWPAN. Transport layer in IoT also relies on the basic protocols. Two such protocols are Transmission control protocol and User Datagram protocol. In [7] the authors provided control area networks communication services as write and read object service. Write object service will usually transmit the data frames in a way like producer to consumer.

IoT can be viewed as an integration of various networks which include wired networks, wireless networks, mobile networks, adhoc networks, and mobile ad-hoc networks[8]. The protocols used in IoT communication layer can be known from the below fig3.

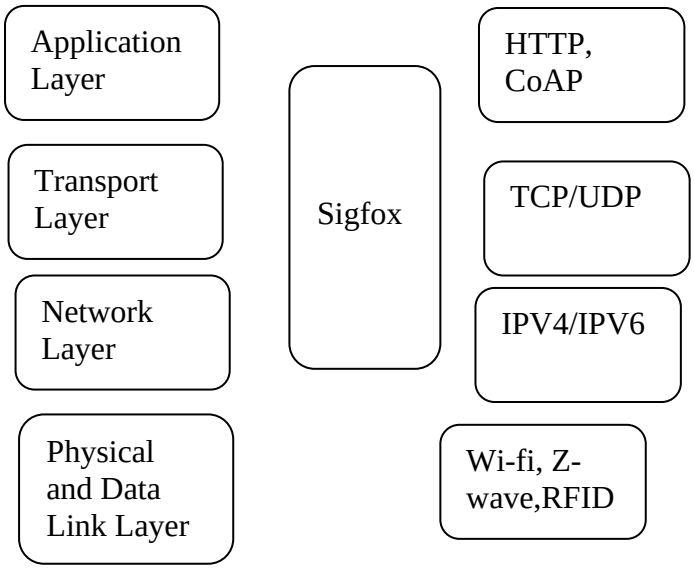


Fig 3. Protocols in IOT Communication

IV. IOT DATA LINK LAYER

MAC Layer is the sub layer of the Data link layer. IEEE 802.15.4 is the standard defined for MAC layer [9]. This standard defines the frame format which includes the header, trailer and payload fields. Node communication is also mentioned in this layer [10]. These formats can be used for traditional networks but when we use the low power communication protocols in IoT, these formats are not suitable due to their overhead. A new defined protocol is needed for low power communication protocols. These are defined in IEEE 802.15.4e[11] which supports low power in IoT devices and meet low-cost communication for IoT. The features that defined the standard for the frame format as shown in Table 1.

Feature	Description
Slot frame Structure	Scheduler for each node
Scheduling	Handles Mobility Scenarios
Synchronization	30 second Prespecified intervals
Channel Hopping	Changing Frequency channel
Network formation	Joining components

Table 1 Features for frame format

IEEE 802.11ah is another low power wireless standard access medium [12]. The features of this standard are given in below Table 2.

Feature	Description
Synchronization Frame	Probe delay can be configured
Efficient Bi directional packet exchange	Power saver for uplink and downlink
Short MAC Frame	Provides short MAC frame for 12 bytes
Null Data packet	Have ACK frames
Increase Sleep Time	Exchanging data

Table 2 Feature of 802.11 ah Frame format

V. EXISTING FRAME FORMAT OF DATA LINK LAYER OF CAN

Data exchange in any serial communication system usually will be in the form of frames. CAN networks has two basic formats [13]. One is the standard base

format which supports bits for identifier, whereas extended frame format supports 29 bits for the identifier. In an extended frame format bit are assigned to the base frame and remaining 18 bits are assigned to extension which can be termed as identifier extension[14]. The frame format for standard base frame is shown in fig 4.

SOF	11bit	RT	Control	Data

Fig4. Standard base format

CAN controllers which support the extended base format can also able to send and receive the messages through standard base format [16] the extended base format for CAN is shown in Fig 5.

SOF	11bit	SRR	IDE	18 bit	RTR

Fig 5. Extended frame format

The control field in the extended format contains r1,r0 and DLC fields [17]. The bit filed of the standard and extended CAN can be summarized as follows in table 3

Field	Description
SOF (1 Bit)	Start of Frame – synchronize the nodes
Identifier (11 bit)	Indicates priority
Remote Transmission Request (1 bit)	Dominant in data frames
Reserver (2 bits)	Dominant
Data length code (4 bits)	Number of data bytes
Data field (0-8 bytes)	Determined by DLC
CRC (15 bits)	Cyclic Redundancy check
CRC delimiter (1 bit)	Recessive
ACK (1 bit)	Acknowledgement field
ACK delimiter (1 bit)	ACK delimiter field
EOF(7 bit)	End of Frame

Table 3. Bit field of Standard and Extended CAN

In the CAN data frame, a 15-bit Cyclic Redundancy Check Sequence need to be derived for the frames being transmitted [18]. This process is done by calculating the CRC from starting of the frame till the data field is encountered.

VI. CRC GENERATION FOR DATA FRAME

The CRC can be generated by using a generator polynomial. In this paper we use the LFS (Linear Feedback Shift) to generate the CRC code and implement it. A polynomial is generated and can be used here. This generator polynomial is shown as follows.

$$X^{15}+X^{14}+X^{10}+X^{8}+X^{7}+X^{4}+X^{3}+1$$

The transmitter of the frame is treated as polynomial and it will be divided by the polynomial generator. Modulo 2 division is performed on the sequence and then this is transmitted along with the message.

The receiver will also perform the same check as done by the transmitter. If the result found is not as same as the bits send by receiver the frames are discarded. It then transmits an error frame and then request the sender to retransmit the frame.

CAN sends the message from source to destination. Usually, CAN bus has four different frame types. Data frame is used to transmit the data over the network. Remote frame is used to request the data in an network from another node. Error frame is used to get the message again in case of error found.

VII. PROPOSED ADDITIONAL PAYLOAD FOR AN EXTENDED FRAME FORMAT

In this section an extended frame format with additional data field bits is proposed in order to accommodate more data. An error free frames can be sent in short time if the data in the payload field can be accommodated more than the currently available frame formats. Low power for IoT is taken care to accommodate more data with other fields designed to

accommodate less bits. This will decrease the power consumption for low cost IoT devices.

The proposed frame format is shown in the following figure.

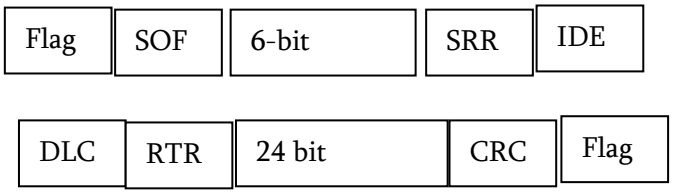


Fig 6. 24 bit payload proposed frame format

In the above fig we propose the frame format for IoT data link layer for 24 bit payload field. The other fields remain same as the previous format. We also propose a Flag field for beginning and ending of the frame format. This frame format better fits for current IoT devices which also require low power. The cost involved for these frame formats are also minimized.

VIII. RESULTS

From the above study and CRC generation for a data frame it can be observed that CAN has two standard formats one is base identifier and other is extended frame format for data link layer in IoT devices. But these frame formats also have some limitations for the communication and data transfer. This frame formats have the starting and ending of the frame which has 1-bit SOF and 7-bit EOF. In the proposed frame format, we can have a 1 bit Flag both at beginning and end which can reduce the bit size. As IoT devices require less power and minimize the cost design this proposed frame format may control the overhead.

IX. ACKNOWLEDGMENT

I thank New Horizon College of Engineering for providing me opportunity to write this paper. Further I thank my Department Head, Master of Computer Applications and all Colleagues for their continued support in completing this paper on time.

X. REFERENCES

[1]. H. Li, Z. Weishi, and Z. Weifu, "A Novel Vulnerability Detection Method for ZigBee MAC Layer," in IEEE 12th International Conference on Dependable, Autonomic and Secure Computing, 2014.

[2]. Shyan, S. Y. Wei, and S. C. Chou, "A Comparative Study of Wireless Protocols: Bluetooth, UWB, ZigBee, and Wi-Fi," in The 33rd Annual Conference of the IEEE Industrial Electronics Society(IECON),2013pp5-8.

[3]. Z. Zhou, Kista, W. Zhiqang, W. Fan, and W. Yuexin, "A middleware of Internet of Things (IoT) based on Zigbee and RFID," in IET International Conference on Communication Technology and Application(ICCTA),2011,pp.34-41.

[4]. Y. Xiolin, J. Zhiqang, Z. Wenjun, and W. Zhingning, "The Research and Implementation of ZigBee Protocol-Based Internet of Things Embedded System," in International Symposium on Information Engineering and Electronic Commerce (IEEC), 2010, pp. 7-10.

[5]. C. H. Kuor, "Bluetooth : A Viable Solution for IoT?," in IEEE Wireless Communications, vol. 21, no. 6, 2011, pp. 6-7.

[6]. J. H. Christophher, "Internetworking with Bluetooth Low Energy," in Mobile Computing and Communications," 2015 pp 34-38.

[7]. Niemen, C. Gomez, M. Isomaki, and T. Savolainen, "Networking Solutions for Connecting Bluetooth Low Energy Enabled Machines to the Internet of Things," in IEEE Network, vol. 5, no. 4, 2014 pp. 83-90.

[8]. C. J. Hansen, "Internetworking with Bluetooth Low Energy," in Mobile Computing and

Communications, vol. 19, no. 2, 2015, pp. 34-38.

[9]. Z-Wave "Understanding Z-Wave Networks, Nodes & Devices"

[10]. P. Amaro, R. Corteso, R. Landeck, and J. Santos, "Implementing an Advanced Meter Reading infrastructure using a Z-Wave Compliant Wireless Sensor Network," in Proceedings of the 3rd International YouthConferenceonEnergetics,2011.

[11]. C. S. Zhengguo, Z. Chunseng, and C. M. Victor "Surfing the Internet-of-Things: Lightweight Access and Control of Wireless Sensor Networks Using Industrial Low Power Protocols," in EAI Endorsed Transactions on Industrial Networks and Intelligent Systems,vol.1,no.1,2014.

[12]. D. Baldo, G. Beneelli, and A. Pozzebon, "The Siesta Project : Near Field Communication," in Proceedings of 7th International Symposium on Communication Systems Networks and Digital Signal Processing, 2010, pp. 721-725.

[13]. V. Coskun, K. Ozdinici, Near Field Communicatioin (NFC), Wiley, 2012.

[14]. P. Urien, "LLCPS: A new security framework based on TLS for NFC P2P applications in the Internet of Things," in IEEE Consumer Communications and Networking Conference (CCNC), 2013, pp. 845-854.

[15]. C. Vedat, O. Busra, and K. Karim, A Survey on Near Field Communication (NFC) Technology. Springler, 2013.

[16]. HomePlug Powerline Alliance, HomePlug Green PHY 1.1, the standard for in-house smart grid powerline communications: an application and technology overview White paper, October 2012, 17p.

[17]. H. Jun, Qingdao, M. Zheng, S. Chungseng, "Energy-efficient MAC Protocol Designed for Wireless Sensor Network for IoT," in Seventh International Conference on Computational Intelligence and Security (CIS), 2011.

[18]. S. Eckehard, K. Matthias, A.N. Anas, D. Stefan, A. Moller, R. Luis, S. Florian, "Advances in Media Technology Internet of Things," technical report, Technische Universitat Munchen Institute for Media Technology, 2013. IoT "New Weightless 2-Way Communication IoT Standard launches"

[19]. A. Sikora and V. F. Groza, "Coexistance of IEEE 802.15.4 with other systems in the 2.4 Ghz-ISM-Band,' in Proc. IEEE Instrumentation & Measurement Technology Conference, May 2005,

[20]. K. Shuaib, M. Boulmalf, F. Sallabi, and A. Lakas, "Coexistance of ZigBee and WLAN: A performance study," in Proc. IEEE Int. Conf. Wireless & Optical Communication Network, April 2006.

[21]. P. S. Neelakanta and H. Dighe, "Robust Factory Wireless Communication: A performance appraisal of the Bluetooth and the ZigBee collocated on an industrial floor," in Proc. IEEE Int. Conf. ind. Electron. (IECON'03), November 2003, pp. 2831-2386.

The International Conference on Research Perspectives : IoT in Hybrid Grid Integrated Renewable Energy Sources
In association with International Journal of Scientific Research in Computer Science, Engineering and Information Technology
ISSN : 2456-3307 (www.ijsrcseit.com)

Predicting Accuracy of Loan Using Machine Learning Techniques

Latha R[1], Subham Sharma[2], Naganandan P[2], Tiyasha Dasgupta[2], Pavan D[2]

[1]Assoc. prof., Department of MCA Sir M Visvesvaraya Institute of Technology, Bangalore, Karnataka, India

[2]PG Student, Department of MCA Sir M Visvesvaraya Institute of Technology, Bangalore, Karnataka, India

ABSTRACT

The number of customers is increasing daily for the loan, their must be a specific mechanism to identify the right person to sanction the loan. In the banking system, various approaches are used to identify the correct person to sanction the loan. In this paper, we are discussing various machine learning methodology that exhibits the best performance for the given set of data .Loan Prediction is very helpful to employees of banks as well as to the applicant too. The Loan Prediction System can automatically calculate the minimum balance in the customer's account for taking loan processing.

Keywords -Loan Prediction, Machine Learning, KNN (K- Nearest Neighbors), Logistic Regression, Random Forest Decision Tree.

I. INTRODUCTION

In our country, most of them are depending on loans to improve their economical growth, market reputation at national and international level. So banks supply the long term and medium term loans to their customers to fulfill there needs. The large amount of data is used by every organization nowadays. Since, it is very important to analyze the data in a correct order to extract some useful information and to implement with the existing algorithm based on analysis. So it can be achieved through data mining and machine learning. Loans is the core business part of almost every banks. The most profit earned by bank's is came from the disbursed loan to the account holder's. Nowadays all the bank approves loan after verifying customer details to check whether the applicant is eligible to get the loan or not through the machine learning techniques .So ,whole process of validation of features is automated by implementing machine learning supervised technique which are Linear Regression, kNN(k-Nearest Neighbors), Logistic Regression, Random Forest. This paper has the following sections (II) Data Set, (III) Methodologies (IV) Implementation (V) Conclusion.

II. DATA SET DESCRIPTION

The input data contains 5581*18 of instances each instances as many as attributes and one quality attribute is called class (such as low, medium and high credit). In this table, it describing the attributes used in the bank_marketing.csv file[8]. The main motive of the machine learning

methodologies is to maintaining the Integrity of the specified attributes.

Different Machine learning techniques are used to analyze the given input data and to use the existing model to classify required output data. The below table (2.1) describes the variable names, description and its types of the given data set. Data classification technology are used by machine learning classifiers to approval for client loan, positioning the target market, medical diagnosis, fault detection, effectiveness aanalysis, graphics processing and for the fraud analysis in the field of insurance.

Table 2.1 Description of data set

Variable	Description	Type
Age	Age of all applicants	Integer
Job	Employed or unemployed	Object
marital	Married or unmaried	Object
education	Graduate/Under Graduate	Object
balanace	Balance of customer	Integer
housing	Yes/no	Object
Loan	Yes/no	Object
contact	Contact communication type	Object
Day	Last contacted days	Integer
month	Last contacted month	Object
duration	Last contacted duration	Integer
campaign	No of campaign	Integer
pdays	Number of days that passed after the client was ontacted from a previous campaign	integer
previous	No of loans taken	integer
poutcome	Success or failure	object
deposit	Yes or no	object

III. METHODOLOGIES

Three machine learning classification models have been used for prediction of loan approval.

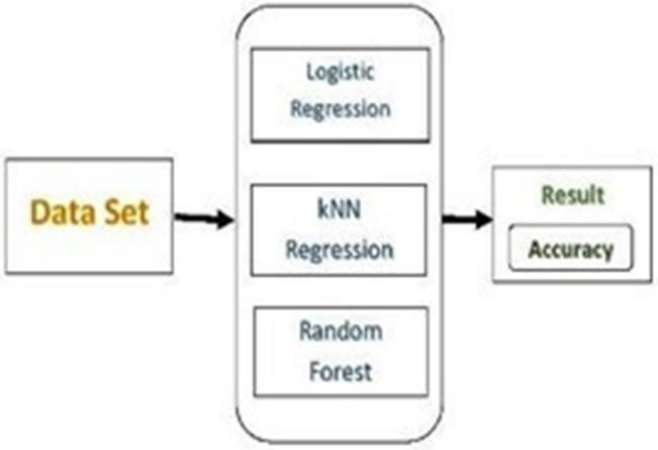


Fig 1 Architecture

The above Architecture shows the brief details of each model is described below:

A. kNN (k- Nearest Neighbors):

It can be widely used for classification and regression problems. It is a simplest way of storing all available cases and classification of new cases by having a majority of votes of its k neighbors. kNN algorithm specifies for a given graph it takes a nearest point of k neighbors that distance will be measured by a distance function called Euclidean and we are considering the value of k=7

B. Logistic Regression

Logistic regression [7] is used for predicting the output of a categorical dependent variable(loan) and to solve the regression problem. Since, the result of output will be the discrete value or categorical values. The values produced by the logistic regression which lies between 0 and 1. Logistic regression is a kind of regression that predicts the probability of an by fitting test data to a logistic function. Logistic regression uses several predictor variables that may be categorical or numerical variable

C. Random Forest

Random Forest is a classifier technique which contains the various subsets for decision tress for the given data set and takes the average to improve the predictive accuracy of that dataset.

This algorithm is used for both classification and regression applications. The accuracy of these models is higher than other above mentioned techniques. The below figure shows the decision tree for random forest

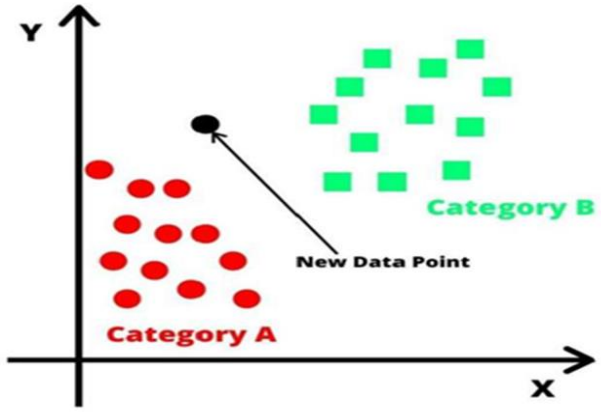


Fig 2 kNN Graph

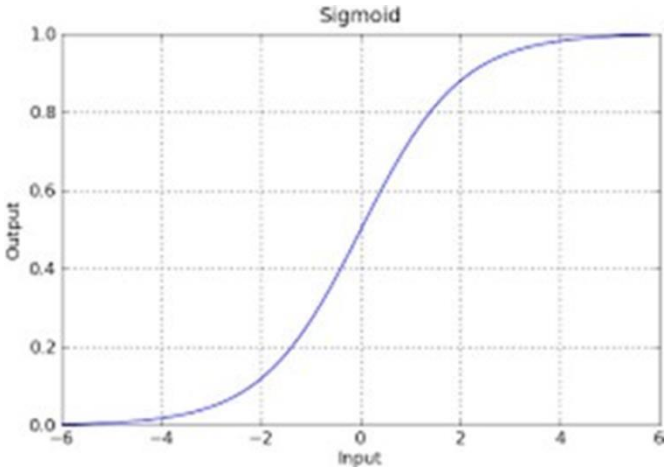


Fig 3 Sigmoid curve for Logistic Regression

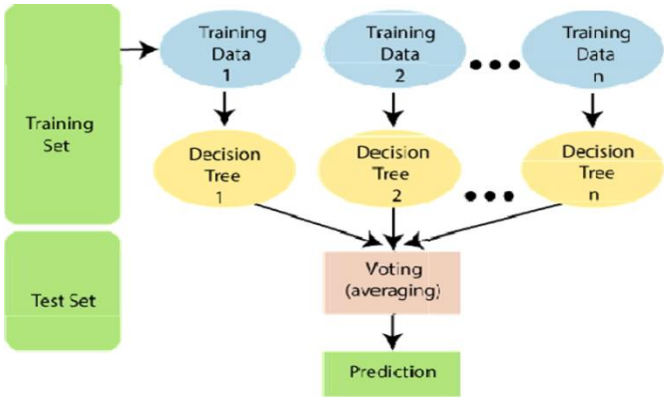


Fig 4 Decision tree for Random Forest.

IV. IMPLEMENTATION

Here are implementing different ML techniques to find the accuracy for loan prediction by using bank_marketing.csv file[8], fig 5 represents the total number of tuples and attributes in the file as shown below.

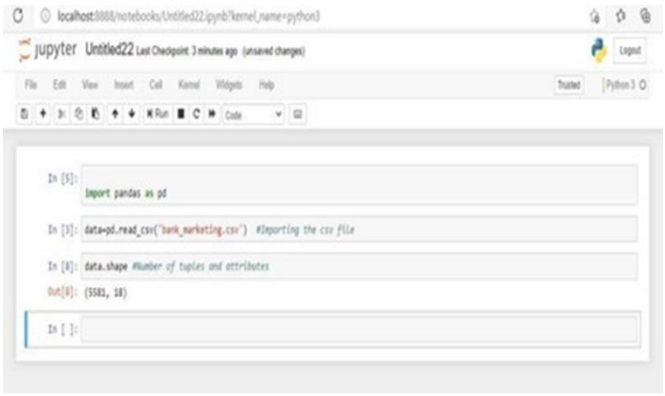


Fig 5. Size of Data Set

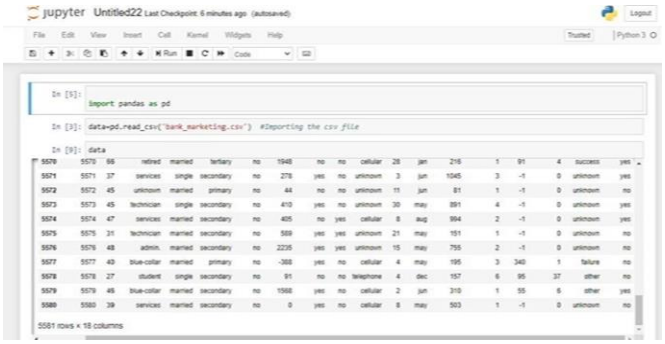


Fig 6 bank_marketing.csv file

A. Logistic regression

Logistic regression [7] shows the logistic relationship, Which means it finds how the value of the dependent variable(loan) changing according to the value of then independent variable(deposit).

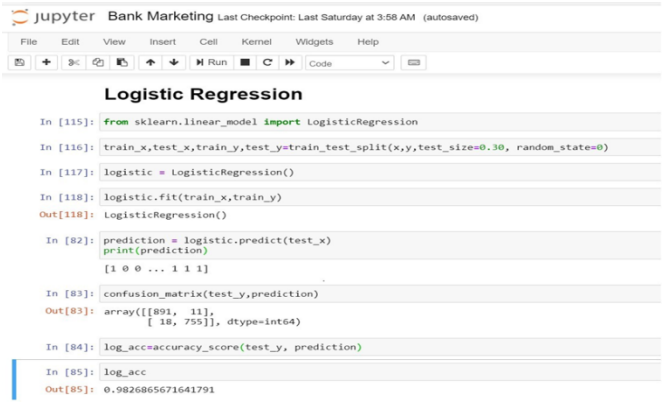


Fig 7 Implementation of logistic regression

B. Random Forest

Random forest algorithm can be used for both classifications and regression task. It provides higher accuracy through cross validation. Random forest classifier will handle the missing values and maintain the accuracy of a large proportion of data.

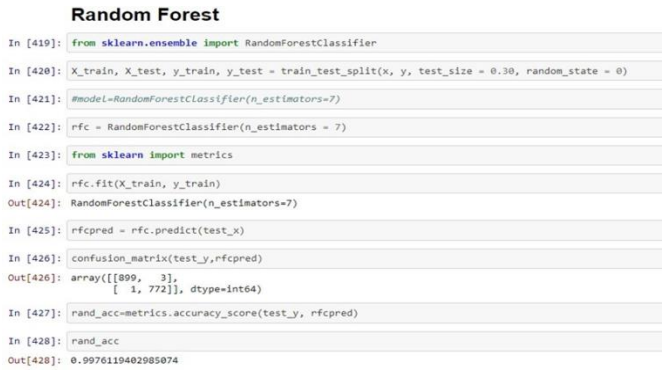


Fig 8 implementation of random forest

C. KNN

A k-nearest-neighbor algorithm, often abbreviated k-nn [4], is an approach to data classification that es imates how likely a data point is to be a member of one group or the other depending on what group the data points nearest to it are in

Fig 9 Implementation of kNN Neighbor

Number	Classifiers	Accuracy
0	Logistic Regression	0.982687
1	Random Forest	0.997612
2	KNN Model	0.754627

Table 2 Accuracy rate

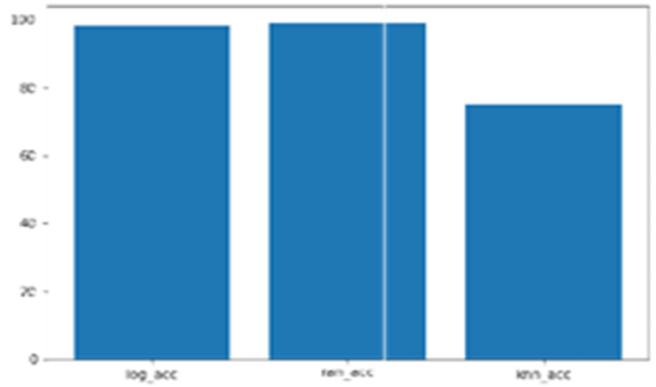


Fig 10. bar graph for the accuracy of all three techniques.

V. CONCLUSION

The existing work implemented by using supervised machine learning techniques certainly proves that the random forest technique performs well with the accuracy when compared to kNN and Logistic Regression techniques to predict loan for the customer for the given large dataset.

VI. REFERENCES

[1]. Pang-Ning Tan, Michael Steinbach, Vipin Kumar: "Introduction to Data Mining, Addison- Wesley, 2005".

[2]. Jiawei han and Micheline Kamber: Data Mining – "Concepts and Techniques, 2nd Edition,Morgan Kaufmann Publisher, 2006".

[3]. "Introduction to linear algebra" - by Gilbert Strang

[4]. "Applied statistics and probability for engineers" – by Douglas Montgomery

[5]. "Mastering python for data science", Samir Madhavan

[6]. Shen Rong, Zhang Bao-wen;" The research of regression model in machine learning" field,School of information engineering of Ningxia University, YIN Chuan, China,750021,School of mathematics and statistics,YIN Chuan,China,750021(2019)

[7]. Vladimir Nasteski, Univeristy "St. Kliment Ohridski"– Bitola "An overview of the supervised machine learning methods".

[8]. https://onlinecourses.nptel.ac.in/noc20_cs80/course

The International Conference on Research Perspectives : IoT in Hybrid Grid Integrated Renewable Energy Sources
In association withInternational Journal of Scientific Research in Computer Science, Engineering and Information Technology
ISSN : 2456-3307 (www.ijsrcseit.com)

Intrusion Detection System using Data Mining Techniques

Miss R. K. Borikar, Dr. S. S. Sherekar, Dr. V. M. Thakre
SGBAU, Amravati, Maharashtra, India

ABSTRACT

Intrusion detection system (IDS) is a process or system that monitors events occurring on a network and analysing it to detect any kind of activity that threat computer security policies.This paper focusedon analysis of five different IDS techniques likePCA and Fuzzy PCA Techniques, Improving K-Means Clustering Using Discretization Technique, Intrusion Detection System Using an Optimized Framework, Anomaly Detection Approach using Hybrid Algorithm, Combination of the DM techniques, etc. But some pros and cons are observed in these various techniques.These approaches have been analysedto address the limitations of intrusion detection system suchas low accuracy, high false alarm rate, and time consuming, etc.So, to overcome these problems the proposed method 'the hybrid algorithm' is used to reduce the rate of false positive alarm, false negative alarm rate, detect zero-day attackers,to improve the detection rate, etc.

Keywords: Intrusion detection, hybrid algorithm, PCA, feature selection SMO, K-means, Clustering.

I. INTRODUCTION

Principal Component Analysis (PCA)is a common statistical technique for data analysis and pre-processing that has been widely applied in many fields of research. PCA is designed to transform the data in a reduced form and keep most of the original variance present in the initial data.Complementing preventative technologies like firewalls, sturdy authentication, and user privilege. IDSs became an importantpart of enterprise IT security management. They're generally classified as misuse based or anomaly-based systems [1]. Network Intrusion Detection Systems (NIDSs) have always been designed to boost and improve the network security issue by agency of investigation, distinguishing, assessing and reportage any unauthorized and illegitimate network connections and activities [2].Intrusiondetection systems (IDSs) have presently attracted the attention of a major portion of world, specified their development and sweetening represent a high priority for organization and analysts and science centers [3].Intrusion is one among the most threats to the net. Therefore, security problems had been huge drawback.So numerous techniques and approaches are given to deal with the constraints of intrusion detection system like low accuracy, high warning rate, and time overwhelming [4]. Associate in the nursing of intrusion detection system (IDS) it monitors either networks or alternative systems for malicious or abnormal behaviours. Complementing preventative technologies like firewalls, sturdy authentication, and user privilege can be used for IDS. [5].

This paper is focused on analysis of techniques such as PCA and Fuzzy PCA Techniques, Improving K-Means Clustering Using Discretization Technique, Intrusion

Detection System Using an Optimized Framework, Anomaly Detection Approach using Hybrid Algorithm, Combination of the DM techniques. This paper proposes a hybrid machine learning techniquefor network intrusion detection based on combination of PCA feature selection, K-means clustering and Sequential Minimal Optimization (SMO) classification. It introduces hybrid approach that is able to reduce the rate of false positive alarm, false negative alarm rate, to enhancethe detection rate and detect zero-day attackers. The classification is performed by using Sequential Minimal Optimization.

II. BACKGROUND

Many studies on data mining models have been done to develop the mobility scheme in recent past years such schemes are:

K nearest Neighbour algorithm is used to classify the test samples of connections into a normal or attack category. K nearest neighbour formula so as to classify samples of connections in to a traditional or attack class,to make anomaly-based IDS more efficient.These techniques are used to reduce the high dimensional data obtained from network traffic, before applying any anomaly-based algorithms [1]. Two sources of dataset being used for testing and evaluating the method which are the generated data and real traffic data with clustering algorithm which has improved the detection quality [2].The genetic algorithm is used to derive a set of classification rules from network audit information, and the support-confidence framework utilized as fitness function to judge the quality of each rule [3].Author employed to derive a group of classification rules from network audit information, and therefore the support confidence framework is employed as fitness perform to evaluate the standard of every rule.The new anomaly detector approach is based on k-means clustering algorithm and Sequential Minimal Optimization (SMO) to detect online network anomaly detection.The proposed approach aims to generate a suitable number of detectors with high detection accuracy [4]. Feature selection is an important step to discard irrelevant information, which in turnincreases the detection accuracy and computational efficiency of the proposed models [5].

This paper introduces some data mining technique i.e.,PCA and Fuzzy PCA Techniques, Improving K-Means Clustering Using Discretization Technique, Intrusion Detection System Using an Optimized Framework, Anomaly Detection Approach using Hybrid Algorithm, Combination of the DM techniques, etc. The paper is organised as follows.

Section I Introduction. **Section II** discusses Background. **Section III** discusses previous work. **Section IV** discusses existing methodologies. **Section V** discusses attributes and parameters and how these are affected on data mining techniques. **Section VI** gives proposed method **Section VII** gives outcome and possible result. **Section VIII** Conclude this paper. Finally, **Section IX** gives future Scope.

III. PREVIOUS WORK DONE

In research literature, many mobility models have been studied to provide various data mining techniques schemes and improve the performance in terms of accuracy and detection rate.

Amal Hadri et al. (2016) [1] have applied K nearest neighbour algorithm in order to classify the test samples of connections into a normal or attack category. The conducted experiments were created by using KDDcup99 dataset. The results obtained reveal that Fuzzy PCA methodology outperforms PCA in detecting U2R and DoS (Denial of Service) attacks. Author purposeto make anomaly-based IDS more efficient, by using some techniques to reduce the high dimensional data obtained from network traffic, before applying any anomaly-based algorithms.

Hatim Mohamad Tahir et al. (2016) [2] proposed a novel flow-based anomaly detection scheme based on the K-Means clustering algorithm. Author deployed the two distance-based methods for classification and outlier detection using K-Means clustering results. Two sources of dataset being used for testing and evaluating the method which are the generated data and real traffic data with clustering algorithm which has improved the detection quality.

Elham Ariafar etal. (2017) [3] has proposed the optimized framework in order to improve the new ensemble clustering method (NEC) in terms of detection rate and false alarm rate. The genetic algorithm is used to derive a set of classification rules from network audit information, and the support-confidence framework is utilized as fitness function to judge the quality of each rule.

Saad Mohamed Ali Mohamed Gadal et al. (2017) [4] has proposed the new anomaly detector approach based on k-means clustering algorithm and Sequential Minimal Optimization (SMO) to detect online network anomaly detection, the proposed approach aims to come up with a suitable number of detectors with high detection accuracy.

Fadi Salo et al. (2018) [5] have proposed an ensembled learning approach that combines many of the discussed DM techniques. Feature selection/extraction was broadly found in the literature, and is an important step to discard irrelevant information, which, in turn, increases the detection accuracy and computational efficiency of the proposed models. This work focused on the base feature selection techniques used as part of the hybrid feature selection approaches proposed by many of the discussed works.

IV. EXISTING METHODOLOGY

Many datamining schemes have been implemented over the last several decades. There are different methodologies that are implemented for different data mining models i.e. PCA and Fuzzy PCA Techniques, Improving K-Means Clustering Using Discretization Technique, Intrusion Detection System Using an Optimized Framework, Anomaly Detection Approach using Hybrid Algorithm, Combination of the DM techniques.

A) PCA and Fuzzy PCA Techniques:

The main goal of this approach is to have a standard format attributes before applying any dimensionality reduction.For ths the author converted the discrete attributes values of dataset into continuous values following the idea used.Suppose there are m possible values for a discrete attribute i. For each discrete attribute correspond m coordinates, and can associate one coordinate for every possible value of the attribute. The coordinate corresponding to the attribute value has a value of 1, and the remaining coordinates has a value of 0. The two dimensionality features reduction techniques PCA and FPCA used to reduce the high dimensionality of data (for the training and testing data) at the same time keeping the maximum variances present in the original dataset. ClassificationFinally, KNN classifier is used for classification in order to check whether these sample test network connections are normal or abnormal.The fuzzy covariance matrix is $C_{fpca} = \frac{1}{M}\sum_{i=1}^{M} V_i V_i^T$ [1].

B) Improving K-Means Clustering Using Discretization Technique:

In this technique the dataset has been tested using two methods which are native Bayes Classifier (NBC) and a hybrid method of K-Means Clustering together with Naive Bayes Classifier (KMC+NBC). Measuring the accuracy, detection rate and false alarm is computed. The performance results produced using this testing data is evaluated and compared to the previous hybrid method of KMC+NBC, so as to

confirm the performance of this proposed method with discretization technique of KMCD+ NBC.

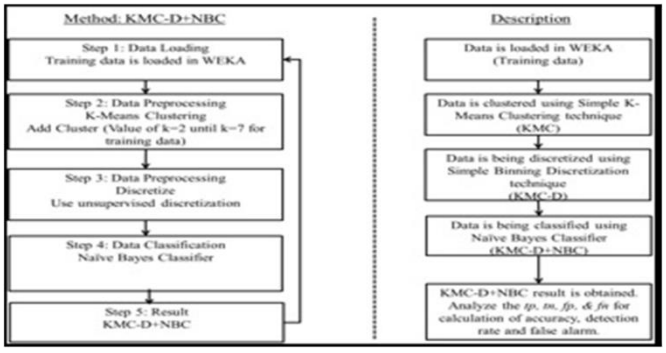


Fig. 1. The Process Flow of KMC-D+NBC Method [2].

C) Intrusion Detection System Using an Optimized Framework:

Here, step 1 isthe various K and max-run values were evaluated following the implementation of the suggested framework in order to determine their optimal values. Clusters were evaluated based on the mean centroid distance. The intended dataset should be uploaded to be used in datamining process. The intended features from the dataset uploaded in the step 1 are selected. Theselection basis may include a specific feature, a subset of features or all features. Determining feature roles are specified. Converting nominal data to numerical data are needed to be converted to numerical ones in the clustering technique. Then feature values are normalized based on Z-transformation method.

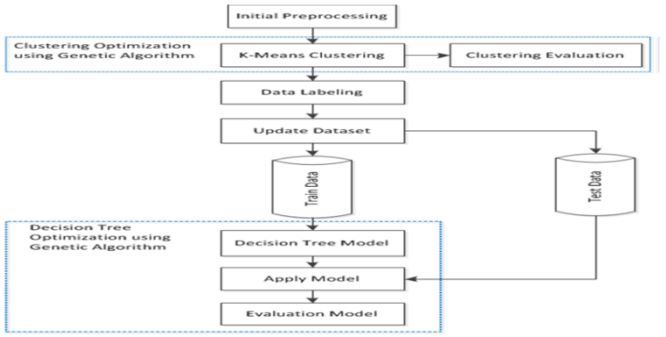


Fig. 2. The Process Flow of genetic algorithm.

The genetic algorithm (GA) was used for clusteringoptimization following the initial pre-processing of theintended dataset. Clusters were used as labels for clustermembers. Thus, the dataset was updated and subsequentlyused in the classification process. The updated dataset which incorporated cluster labels was then used for dataclassification. The GA was used to optimize the confidenceparameter, hence improving the quality of the DTclassification process. Finally, the output of the suggestframework was presented. Given the application of clusteringand classification techniques in this framework, it can be usedfor both labelled and unlabelled datasets, i.e. supervised andunsupervised methods [3].

D) Anomaly Detection Approach using Hybrid Algorithm :

Here,the main idea is based on using feature selection in pre-processing phase to reduce the number of features in dataset, The ConsistencySebsetEvel and Genetic search, algorithms have been applied to select specific features, from the dataset and remove those features which are irrelevant before clustering and classification phases, after attribute selection k-means clustering algorithm selected to reduced training dataset in order to decrease time and processing complexity. In classification phase supervised algorithm Sequential Minimal Optimization (SMO) selected to improve the quality of detection.

$$\textbf{Decoction rate} = \text{DTR} = \frac{TP}{TP+FN} * 100$$

$$\textbf{False positive rate alarm} = \text{FPR} = \frac{FP}{TN+FP} * 100$$

$$\textbf{Accuracy} = \text{AC} = \frac{TP+TN}{TP+TN+FP+FN} * 100$$ [4].

E) Combination of many of the DM techniques:

The methodology is consisting of three primary phases: planning, conducting, and reporting. Each phase has specific and distinct step. A crucial step for the first phase is to create a review protocol. All the selected data is filtered and stored in the table format without containing duplication of any data sets. Datasets considered here are pages containing research questions to simplify the tracing procedure of the data extraction process results, a binary model was used to show whether each dataset answers the research questions or not i.e.to synthesize the data

extracted from the selected articles andto answer the research questions (RQs). Narrative synthesis is the method of tabulating the results with respect to the research question and visualizing them using techniques like bar charts and pie charts to enhance the result presentation and applied it to IDS.

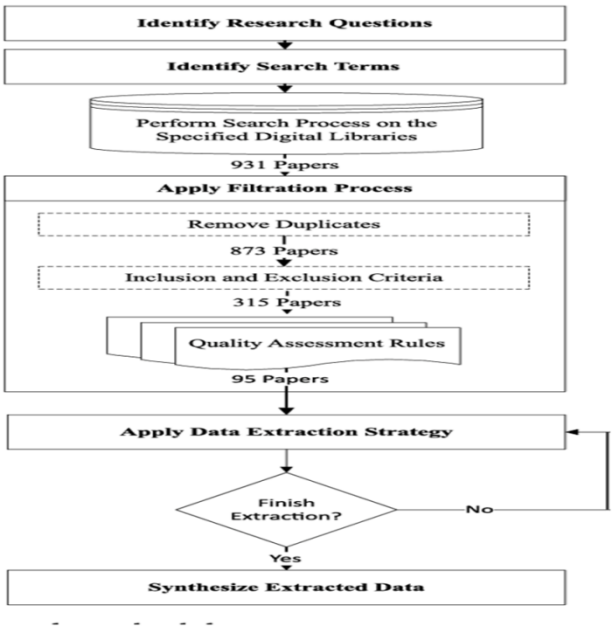


Fig.3. Research Methodology[5].

V. ANALYSIS AND DISCUSSION

The results obtained reveal that Fuzzy PCA method outperforms PCA in detecting U2R and DoS (Denial of Service) attacks [1]. The outcome depicts thatthe proposed method generates better detection rate and accuracyup to 99.3% and 99.5% respectively and reduces the false alarmto 1.2% with better efficiency of 0.03 seconds time taken to buildmodel[2]. The NSL-KDD 2009 dataset have revealed that the method achieved a 99.1% of detection rate (DR) and 1.8% of false alarm rate (FAR), shows an improvement compared with the new ensemble clustering (NEC) method[3].The proposedtechnique (K-mean + SMO) has achieved a positivedetection rate of (94.48%) and reduce the false alarm rate to(1.2%) and achieved accuracy of (97.3695%)[4]. A research gap in establishing the effectiveness of classifiers to identify intrusions in modern network traffic when trained with aging datasets [5].

Data mining Techniques	Advantages	Disadvantages
PCA and Fuzzy PCA Techniques	Most used and overall best classifier, Flexibility when selecting parameters.	Higher false positive than other algorithms make it difficult to use, Training takes more time in average.
Improving K-Means Clustering Using Discretization Technique	Relatively simple to implement. Usedfor large data sets,	Being passionate about initial values, bunch information of varied sizes and density, bunch outliers, Scaling with range of dimensions.
Intrusion Detection System Using an Optimized Framework	It monitors and analysis the user and system activities.	More maintenance, False positives, False negatives, Staff requirements.
Anomaly Detection Approach using Hybrid Algorithm	Most used and overall best classifier. Flexibility when selecting parameters,	Size and dimensionality of dataset affects training complexity considerably, Higher false positive than other algorithms make it difficult to use.

Combines many of the DM techniques	Most used and overall bestclassifier, Flexibility when selecting parameters, High accuracy and simple,	Higher false positive than otheralgorithms make it difficult to use, Training takes more time in average, Lower detection precision for low frequent attacks.

VI. PROPOSED METHODOLOGY

The proposedhybrid algorithm isbased on PCA feature selection, k-means clustering algorithm andSequential Minimal Optimization (SMO) to detect onlinenetwork anomaly detection, the proposed approach aims togenerate a suitable number of detectors with high detectionaccuracy rate. The main idea is based on using feature selectionin pre-processing phase to reduce the number of features indataset, to select specific featuresfrom the dataset and remove those features which areirrelevant before clustering and classification phases, afterattribute selection k-means clustering algorithm selected toreduce training dataset in order to decrease time andprocessing complexity. In classification phase supervisedalgorithm Sequential Minimal Optimization (SMO)is used to improve the quality of detection.

Basic steps of algorithm:

Step1: Training datasets are loaded

Step2: Pre-processing

A. Features Selection:The features reduction techniques PCA and FPCA in order to reduce the high dimensionality of data (for the training and testing data) at the same time keeping the maximum variances present in the original dataset.

B. Clustering phase: The clustering phase done by applying the K-meansclustering algorithm, two clusters were specified andcreated. As the algorithm iterates through the training data,each cluster's architecture transferred to another. Theupdating of clusters causes the modification of the values of the centroids. This change is a reflects in the current clusterelements. When there are no changes to any cluster, theclustering of the K-Means algorithm becomes completed.

Step3: In Last phase classification is supervisedalgorithm i.e. Sequential Minimal Optimization (SMO) wasused to classify dataset to normal or anomaly.

Step 4: Result is obtained.

Diagrammatic representation of proposed method is shown as follows:

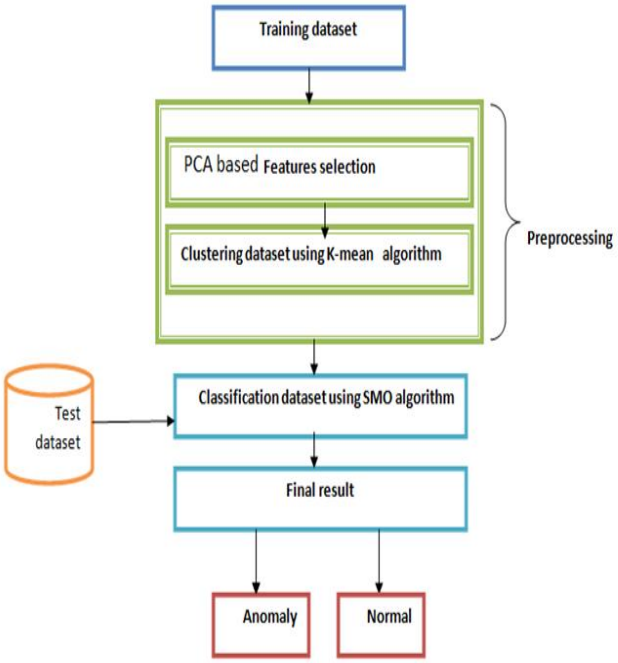


Figure4: Flowchart of the hybrid algorithm using data mining techniques

VII. OUTCOME AND POSSIBLE RESULT

The results derived from this proposed method and approach outperforms other by a positive detection rateand reduce the false alarm rate and giveshighaccuracy.

VIII. CONCLUSION

In recent years, by spread of using the Internet, need ofinformation security has been felt more than ever to preventpersonal and confidential information from unauthorizedintrusion. The different approaches introduced for intrusion detection. This paper presents a hybrid approach toanomaly detection using PCA feature selection, K-means clustering andSequential Minimal Optimization (SMO) classification.It uses feature selection in pre-processing phase to reduce the number of datasets, and remove those features which are irrelevantbefore clustering and classification phases and it usesk-means clustering to reduce the size of the training datasetwhile maintaining datasets. After that in classification phasesupervised algorithm called Sequential MinimalOptimization (SMO) selected to improve the quality ofdetection.

IX. FUTURE SCOPE

From observations of the proposed method the future work willinclude high accuracy, high positive detection rate, reduce the false alarm rate.

X. REFERENCES

[1]. Amal Hadri, Khalid Chougdali and Rajae Touahni , "Intrusion Detection System using PCA and Fuzzy PCA Techniques" 2016 International Conference on Advanced Communication Systems and Information Security (ACOSIS), 10.1109/ACOSIS.2016.7843930

[2]. Hatim Mohamad Tahir , Abas Md Said,Nor Hayani Osman, Nur Haryani Zakaria, Puteri Nurul 'Ain M Sabri and Norliza Katuk., "Improving K-Means Clustering Using Discretization Technique In Network Intrusion Detection System"2016 3rd International Conference on Computer and Information Sciences (ICCOINS), 10.1109/ICCOINS.2016.7783222

[3]. Elham Ariafar and Rasoul Kiani," Intrusion Detection System Using an Optimized Framework Based on Datamining Techniques" 2017 IEEE 4th International Conference on Knowledge-Based Engineering and Innovation (KBEI). 10.1109/KBEI.2017.8324903

[4]. Saad Mohamed Ali Mohamed Gadal and Rania A. Mokhtar,"Anomaly Detection Approach using Hybrid Algorithm of Data Mining Technique"2017 International Conference on Communication, Control, Computing and Electronics Engineering (ICCCCEE), 10.1109/ICCCCEE.2017.7867661

[5]. Fadi Salo, Mohammadnoor Injadat, Ali Bou Nassif, Abdallah Shami, And Aleksander Essex," Data Mining Techniques in Intrusion Detection Systems: A Systematic Literature Review"2018IEEE Access (Volume:6) 10.1109/ACCESS.2018.2872784

	Rakshanda K. Borikar has completed B.E. Degree in Computer Science and Engineering from Sant Gadge Baba Amravati University, Amravati, Maharashtra. She is pursuing Master's Degree in Computer Science and Information Technology from P.G. Department of Computer Science and Engineering, S.G.B.A.U. Amravati. (e-mail id: rakshandaborikar@gmail.com)
	Dr. Swati S. Sherekar Presentlyworking as Professor in the P. G.Department of Computer Science and Engg and having 25 years of

	teachingand research experience. Her area ofresearch is Network security, MobileComputing, Cloud Computing andworking as supervisor for Ph.D.guidance. Completed one MRP ofUGC one MRP of AICTE. She haspublished more than 70 papers inInternational & National levelJournals and also Invited for Talks in many International / National Conferences and National level Conferences.(e-mail id: ss_sherekar@rediffmail.com)
	**Dr. Vilas M. Thakare** is Professor and Head in Post Graduate department of Computer Science and Engg, Faculty of Engineering & Technology, SGB Amravati university, Amravati. He is also working as a coordinator on UGC sponsored scheme of e-learning and m-learning specially designed for teaching and research. He is Ph.D. in Computer Science Engineering and completed M.E. in year 1989. He has done his PhD in area of robotics, AI and computer architecture. His area of research is Computer Architectures, AI and IT. He has published more than 150 papers in International & National level Journals and also International Conferences and National level Conferences. (e-mail id: vilthakare@yahoo.co.in)

The International Conference on Research Perspectives : IoT in Hybrid Grid Integrated Renewable Energy Sources
In association withInternational Journal of Scientific Research in Computer Science, Engineering and Information Technology
ISSN : 2456-3307 (www.ijsrcseit.com)

Development of Efficient Clustering Approach for Analysis of Real Data

Miss.R. J. Wadnare, Dr. S. S. Sherekar, Dr.V. M. Thakare
SGBAU, Amravati, Maharashtra, India

ABSTRACT

Clustering is very important techniques in various fields like machine learning, data mining, image processing, knowledge discovery etc. But techniques developed for clustering i.e. k-means, partition based clustering, DBSCAN suffer from different problem such as noise data, parameter selection, cluster shape, overlapping of clusters. The partition based clustering is simple and effective but more fragile to noise. This paper analyzesdifferent existing clustering techniques namely KMDD, symmetric neighborhood graph with density based clustering, initial cluster center selection, density clustering framework, and spectral clustering via massage passing and density similarity.Butthese techniques have some limitations like density partition method is time consuming for large data sets, KMDD fail to identify core and non-core points, parameter selection etc.To improve theseapproaches thispaper proposed the development of efficient clustering approach for analysis of real data.

Keywords :- Clustering, Data Mining, Spatial Clustering, Diverse Shape And Densities, Revere-K-Algorithm, Neighbourhood Density Estimation, Density Peak Clustering.

I. INTRODUCTION

Clustering is a very effective unsupervised learning approach and it is widely used in various fields including data mining pattern recognition and image analysis. Due to the evidence importance of clustering various, clustering algorithms have been proposed in the past decades, which include partitioning methods, density based clustering and hierarchical clustering. The partition based clustering is simple and effective but more fragile to noise. A density-based algorithm- DBSCAN which is robust to noise, shape and densities but it's performance totally depends upon the parameter cut off distance, which it need to select in advanced. In Density based clustering framework (DCF)Density partition needs to be repeated for large data sets and cluster with more overlap so it may be time consuming [1]. Recently Density and Distance based clustering algorithm is invented in which time consumption is high but core idea is novel. Motivated by their work a new KMDD method for large spatial data sets and highdimensional datasets is invented.AS the k-means are order sensitive to the data, result of KMDD are not stable. In some case core and noise are difficult to identify [2]. Density Peak Clustering With Symmetric Neighbourhood Graph (DPC-SNR) issuitable for larger data sets.Clustering of data sets can be done correctly.Cluster centre can successfully identify cluster centre regardless of their distributions and dimensionality.The efficiency of algorithm is depending upon the selection of

parameter k [3]. ICCK-K-a means novel method for initial cluster centre selection which automatically determines the number of clusters. It is robust, more stable and gives high quality results but generation of a distance matrix causes large amount of time consumption [4]. Clustering algorithm based on message passing.(MPSC) effectively deal with multi-scale data sets, but it consumes a lot of time when dealing with large-scale data, just like traditional spectral clustering [5].

To overcome this problems paper, propose a new method which combine the advantages of different clustering algorithm. By using the k- nearest neighbour algorithm and Reversek- nearest neighbour it identification is done for the core points and non-core points.initial cluster centre for partition clustering is core points identified byk- nearest neighbour and rknn. By using this initial centre KMDD perform clustering on data which can give more stable and accurate clustering results as compared to other algorithm.

II. BACKGROUND

Many studies have been done to develop an effective density cluster method are: The method proposed is density-based clustering framework (DCF) which is simple and effective. In DCF, data is partition into core points and non-core points by using neighbourhood density estimation model. Clusters are formed using the core point which represents the cluster center. Finally, in sequence non-core points are classified in initial clusters. Experiment show that DCF results are comparable with the DBSCAN and Dp algorithm. It also shows that core points are better to represents cluster structure [1]. Author proposed [2] advance clustering method called KDMM, which discover fast cluster with diverse shape and densities in special dataset. It performs partition and merges strategy to form cluster. Dissimilar of traditional hierarchical clustering it apply both density and distance approach to accumulate sub-clusters. It effectively identifies clusters with density and diversities. Its results are vigorous with respect to choice of parameter. In otherscheme [3], a symmetric neighbourhood graph over all data points is establish, based on the *k*-nearest neighbours and reverse *k*-nearest neighbours of each point called density peak clustering using symmetric neighbourhood relationship. Local densities of each point are calculated using the reverse *k*-nearest neighbours, which enables more efficient identification of initial cluster. Similar clusters are aggregated on the symmetric neighbourhood graph-based on the initial center of clusters. According to distance from higher density, all points are sorted by applying Breadth first search on symmetric neighbourhood graph. It form very successful cluster. To determine the initial cluster center this method[4] uses MNN(M nearest neighbour), density, distance which find the neighbourhood radius r. The average of density, distance and MNN is r. The density of each data point lie within the range of r is calculated. The probability factor f is used to find the clustercenter.The data point with bigger f is set as a cluster center. On comparing with different clustering algorithm the proposed algorithm obtained results with quality cluster center identification [4]. The proposed[5] method uses the adaptive density similarity measure, which describe the relation between cluster centers and data points with the help of message passing mechanism in AP clustering. According to density-sensitive similarity distance between pairs similarity matrix is constructed. The problems in clustering multi-scale data set is effectively deal by proposed method. The performance of proposed algorithm is better than traditional spectral clustering algorithm and very stable.

These are organizes as follows. **Section I** Introduction. **Section II** discusses Background. **Section III** discusses previous work. **Section IV** discusses existing methodologies. **Section V**discusses attributes and parameters and how these are affected on clustering techniques, **Section VI** proposed method**Section VII**for outcome result**section VIII** Conclude this. Finally**Section IX** gives future scope.

III. PREVIOUS WORK DONE

In research literature, many clustering models have been studied to provide various clustering algorithm and to improve the performance in terms ofcluster identification, extract knowledgeable information.

Jian yun Lu et al (2017) [1] proposed method uses the adaptive density similarity measure, which describe the relation between cluster centers and data points with the help of message passing mechanism in AP clustering. According to density-sensitive similarity distance between pairs similarity matrix is constructed. The experiments show that the proposed algorithm can effectively deal with the clustering problem of multi-scale datasets. The performance of proposed algorithm is better than traditional spectral clustering algorithm and very stable.

JIANG WANG et al(2017) [2]] had proposed advance clustering method called KDMM, which discover fast cluster with diverse shape and densities in special dataset. It performs partition and merges strategy to form cluster. Dissimilar of traditional hierarchical clustering it apply both density and distance approach to accumulate sub-clusters. It effectively identifies clusters with density and diversities. Its results are vigorous with respect to choice of parameter.

Chunrongwu et al (2019) [3] had worked on a new clustering algorithm that is robust to outliers. Reverse k-nearest neighbour algorithm is used to find local density of data points on symmetric neighbourhood graph. Based on the mutual connectivity on the graph tiny cluster and outliers are merged into larger cluster. This scheme identify cluster irrespective of data distribution and dimensionality.

YatinLi et. al (2019)[4] worked on to solve the difficulty in initial cluster center selection, to find number of cluster automatically. It finds the data objects that reflect the actual distribution of data. It removes the dependency of cluster result on cluster center selection and number of clusters.

Lijuan Wang et al (2019) [5] has proposed, a special clustering algorithm based on massage passing technique[MPSC]. Density sensitive similarity is used to improve the clustering performance. It describes more accurately complex relation between data points. It also maintains global consistency of data set by better distribution of data. It gives high quality cluster as compared to the K-means, AP, MKAP & NJW algorithm.

IV. EXISTING METHODOLOGIES

Many clustering techniques been implemented over the last several decades. There are different technique that are implemented for different purpose such asPartitioning methods, Hierarchical clustering. Fuzzy clustering.

Density-based clustering, Model-based clustering.

A) Density based clustering framework (DCF):

The DCF mainly contains four modules: density partition, initial clustering, ordering and partition clustering.

Density Partition: In this module, dataset D is divided into core pointsand non-core points. This is achieved using the neighbourhood density method.Initial clustering: In this module, dataset D is divided into sets. One set contain only core points and other sets contain non-core points. It also apply different clustering algorithm to form cluster s. the set with only core points form better cluster. Ordering: It

classifiesnon-core points into initial clusters in a priority sequence. To achieve this, it used priority sequence of non-core points with neighbourhood density.Partition Clustering: In thefinal step non-core points are added into initial clusters. This module needs to design a function to measure the similarity between non-core points and the initial cluster[1].

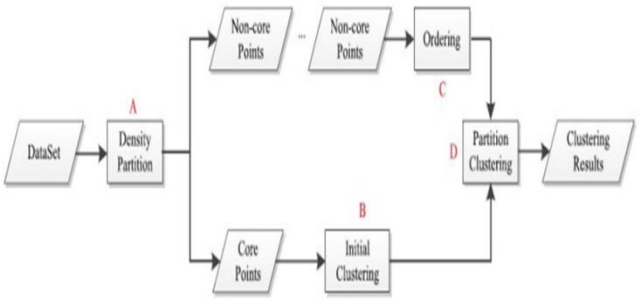


Fig1.The density-based clustering framework

B) KMDD:

KMDD combines two clustering methods, k-means and DD. Firstly k-means clustering partitions datasets into sub-clusters and calculate minimum distance between sub-clusters, also local density for each sub-cluster. Decision graph is used to choose core-cluster. It is assumed that genuine cluster has only one core, on which each non-core sub-cluster is assigned to the nearest neighbour with high density and Noise point is identified with lower density. The local density of sub-cluster is defined as:[2].

$$\rho_i = \sum_{j=1}^{K} |s_k| \exp(\frac{d_{ij}^2}{d_c^2})$$

C) Density Peak Clustering With Symmetric Neighbourhood Graph(DPC-SNR):

It isclustering method based on Density Peak with symmetric Neighbourhood Relationship is introduced. To improve the (DPC), it uses the reverse k- nearest neighbour ((kNN) and symmetric Neighbourhood Relationship for computing the local density of a point and change the assignment method. In data mining technology knn is simplest method. Euclidean distance method is used to calculating the distance between two points. After evaluating the distance between the points, sort the distance in ascending order to find the k nearest distance. All points contained in the k nearest distance correspond to kNN, and also get reverse kNN during the process. The graph constructed by linking the symmetric neighbourhood of each point is called as symmetric neighbourhood graph(SNG).The outlier is detected as point with less than two neighbours in symmetric graph. Some rules must be followed while building a cluster. These rules are:Suppose X and Y are two point in symmetric neighbourhood graph. X is already extended in graph and Y has not been extended yet. Y will be extended if it the following rules.

1. Y is not a outlier.
2. Y is not visited yet.
3. The distance between X and Y are less than the mean distance.

The algorithm choose cluster from a point with the largest product value of local density and distance by undertaking a breadth first search of SNG. After traversing whole SNG, it found some initial clusters. The main cluster is found in the process of clustering. It also combines the small cluster with large cluster whose connected edges are larger in the SNG. This algorithm avoids erroneous assignment of data and improves the efficiency of clustering. Therefore, it is suitable for lager data set. It is robust to outlier. The local density can be defined as[3].:

$$\rho_i = \sum_{j \in RkNN(i)} exp(-dist^2(i,j))$$

D) ICCK-K-a means novel method for initial cluster center selection:

This method used distance cost function F(S, K) to find the number of cluster K. For determining initial cluster center following three steps are used. Firstly, the distance between two objects is calculated. The average value of distance, density and MNN is used to calculate neighbourhood radius r. the density p of data point lie within the range of r is calculated. The probability f of data point to become cluster center is finding by dividing data set n with corresponding distance. By sorting f in descending order top k output are k initial clusters. The number of clusters identified by minK{F(S,K)} K= 1,2 ,3….n. where F(S,K) is distance cost function. Initial cluster center selected by applying f=p/r.Where p is the density of data point and r is the distance corresponding to MNN [4].

E) Clustering algorithm based on message passing.(MPSC):

This paper proposed the new clustering technique based on density similarity measurement, which is insensitive to parameter selection and can deal with the multi-scale clustering problem. Like Spectral clustering algorithm MPSC algorithm is also consider the data points as a vertex and construct the undirected weighted graph. This algorithm construct similarity matrix based on the density sensitive distance. The density sensitive distance is adjustable with the line segment length, which can adjust the distance in region with different densities. Density sensitive similarity is measured on the basis of density sensitivity distance. As compared to the Guass kernel which is used in spectral density clustering method here the similarity measure does not require introducing kernels and directly calculate the similarity in distance measure. Similarity in the same matrix data points is higher and data point with different matrix is lower. The distance measure is depends on the data and adjustable with the density sensitivity. The manifold similarity can be reduced the distance between two data points on the same manifold and vice-versa. Purpose of building the MPSC algorithm is to reduce influence of randomly selected parameter and to obtain more stable clustering results [5].

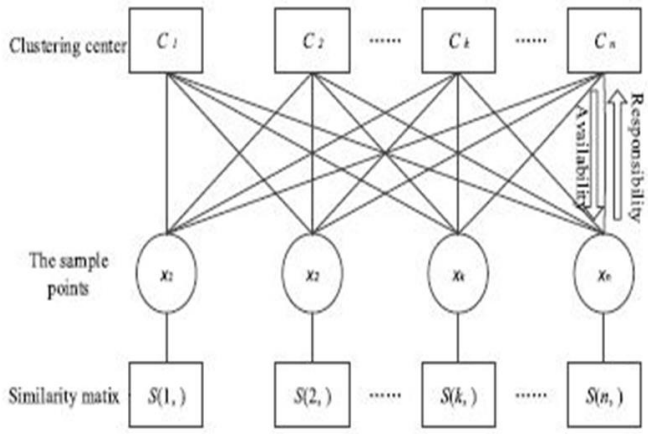


Fig1 Description of the clustering process in AP algorithm

V. ANALYSIS AND DISCUSSION

Density-based clustering framework (DCF) overcome problems such as noise data, parameter selection, cluster shape, overlapping of clusters problems by using density-based clustering framework (DCF). It proved that utilization of core point is more reliable to represent cluster characteristics than use one point to represent a cluster always [1]. KMDD is method for large spatial data sets and high dimensional datasets, is the first method to use density and distance-based concept to aggregate the sub-clusters. Earlier agglomerative algorithm used distance and probability to measure connectivity or dis-connectivity between different sub-clusters [2]. Density Peak Clustering With Symmetric Neighbourhood Graph(DPC-SNR) is method suitable for large datasets which establishes a symmetric neighbourhood graph over all data points, which enable more efficient identification of initial cluster[3].In traditional clustering methods results of clustering depend on the cluster center selection and

numbers of clusters which leads to lack of stability. ICCK-K-Means determine number of cluster using MNN, distance and density to determine initial cluster center. High quality cluster centers are identified by ICCK-K-Means as compared to anotheralgorithm [4].

Clustering algorithm based on message passing. (MPSC) remove dependency of initial cluster selection and give more stable result in multi-set dataset as compared to the spectral density algorithm[5].

Clustering Method	Advantages	Disadvantages
Density based clustering framework (DCF)	The noise, border and dense points are effectively identified using RKNN.Core points are better to represent cluster structure. Cluster overlap and density variation are eliminated by using the core points.Final cluster shape is not influence by the classifying non-core point because it uses priority sequence. It combines the advantages of different clustering algorithm.	Density partition method needs to be repeat for large data sets and cluster with more overlap so it may be time consuming.
KMDD	The cluster result is robust with respect to the choice of parameter. The efficiency of KMDD is not influence by data dimension. It has capability to identify cluster with different shape and size	AS the k-means are order sensitive to the data result of KMDD is not stable. In some case core and noise are difficult to identify.
Density Peak Clustering With Symmetric Neighbour hood Graph(DP C-SNR)	It is suitable for larger data sets. Clustering of data sets can be done correctly. Cluster center can successfully identify cluster center regardless of their distributions and dimensionality.	The efficiency of algorithm is depending upon the selection of parameter k.
ICCK-K-a means novel method for initial cluster countersele ction	Automatically determines the number of clusters. Automatically determines the initial clusters. Highrobustness. More stable and high-quality results .	Generation of a distance matrix causes large amount of time consumption. It had randomness in selecting the distance of the MNN.
Clustering algorithm based on message passing.(M PSC)	Effectively deal with multi-scale data sets. A performance is very stable. It can better reveal the distribution of data and maintain the global consistency of datasets. The density sensitive similarity measure can correctly describe the	Consume a lot of time when dealing with large-scale data, just like traditional spectral clustering. .

	complex relationships among data points.	

TABLE 1: Comparisons between different clustering techniques.

VI. PROPOSED METHODOLOGY

KMDD algorithm can identify cluster with diverse shape and density and robust with respect to parameter but it fail some time to differentiate core and non-core points and k-means is order sensitive with respect to input data. The new neighbourhood density estimation with KMDD algorithm proposed to overcome this problem. It use reverse k-nearest neighbour as density estimation model because it simple and effective to calculate local density. Dataset is partitioned into number of sub-cluster k, by using k-means clustering algorithm. The distance matrix is obtained by finding each sub-clusters distance. The minimum distance between the sub-clusters is calculated. The sub-cluster which is non-core and noise free are merged into core sub-cluster. Final cluster is found and labelled according to the sub-cluster.

Basic steps of algorithm:

Step 1 : Calculate the distance between two points using Euclidean distance.

Step 2 : Identify the core points by using reverse k-nearest neighbour

Step 3 : Apply k-means algorithm to partition the data.

Step 4 : Merge the sub-cluster by applying decision graph.

Diagrammatic representation of proposed method is shown as follows:

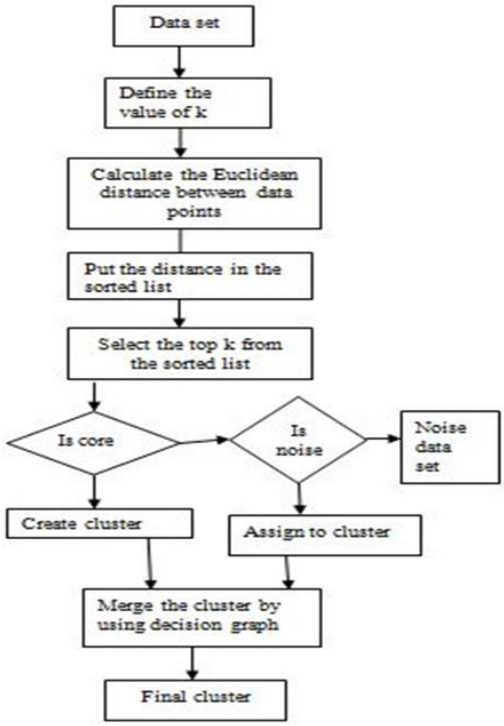


Fig1. Flowchart of neighborhood density estimation with KMDD algorithm

VII. OUTCOME AND POSSIBLE RESULT

In this way the proposed method is performing clustering for data with varies size and shape. This gives more accurate result than KMDD algorithm.

VIII. CONCLUSION

This paper focused on the study of various clustering techniqueKMDD, symmetric neighborhood graph with density-based clustering, initial cluster center selection, density clustering framework, and spectral clustering via massage passing and density similarity, But KMDD sometimes fails to differentiate between core points and non-core points which leads to wrong result. To overcome this problem this paper proposed development of efficient clustering approach for analysis of real data. It gives proper

initial centre for clustering and improves the performance of KMDD algorithm.

IX. FUTURE SCOPE

From observations of the proposed method the future work will include to reduce time complexity of method.

X. REFERENCES

[1]. Jianyun Lu, And Qingsheng Zhu, "An Effective Algorithm Based on Density Clustering Framework" IEEE Access vol.5 February 9, 2017.

[2]. Jiang Wang , Cheng Zhu, Yun Zhou , Xianqiang Zhu, Yilin Wang, And Weiming Zhang" From Partition-Based Clustering to Density-BasedClustering: Fast Find Clusters With DiverseShapes and Densities in Spatial Databases" IEEE Acess on SPECIAL SECTION ON ADVANCED DATA ANALYTICS FOR LARGE-SCALE COMPLEXDATA ENVIRONMENTS vol 6 September 4, 2017,

[3]. Chunrong Wu 1, Jia Lee TeijiroIsokawa, Jun Yao1,And Yunni Xia "Efficient Clustering Method Based on Density Peaks With Symmetric Neighbourhood Relationship" vol 7 April 4, 2019,

[4]. Yating Li, Jianghui Cai , Haifeng Yang , Jifu Zhang , And Xujun"A Novel Algorithm for Initial Cluster Center Selection" IEEE Acess vol.7 April 30, 2019,

[5]. Lijuan Wang, Shifei Ding, And Hongjie Jia "An Improvement of Spectral Clusteringvia Message Passing and Density Sensitive Similarity" IEEE Acess vol 7June 28, 2019

Rupali Wadnarehas completed B.E. Degree in computer engineering from RastrsantTukadojimaharaj Nagpur university Maharashtra. She is persuing Master's Degree in Computer Science and Information Technology from P.G. Department of Computer Science and Engineering, S.G.B.A.U. Amravati. (rupaliwadnare@gmail.com)

Dr. Swati S. Sherekar Presentlyworking as Professor in the P. G.Department of Computer Science andEngg and having 25 years of teachingand research experience. Her area ofresearch is Network security, MobileComputing, Cloud Computing and working as supervisor for Ph.D.guidance. Completed one MRP ofUGC one MRP of AICTE. She haspublished more than 70 papers inInternational & National level Journals and also Invited for Talks in many International / NationalConferences and National levelConferences.
(e-mail id:ss_sherekar@rediffmail.com)

Dr. Vilas M. Thakare is Professor and Head in Post Graduate department of Computer Science and Engg, Faculty of Engineering & Technology, SGB Amravati university, Amravati. He is also working as a coordinator on UGC sponsored scheme of e-learning and m-learning specially designed for teaching and research. He is Ph.D. in Computer Science Engineering and completed M.E. in year 1989. He has done his PhD in area of robotics, AI and computer architecture. His area of research is Computer Architectures, AI and IT. He has published more than 150 papers in International & National level Journals and also International Conferences and National level Conferences.
(e-mail id: vilthakare@yahoo.co.in)

The International Conference on Research Perspectives : IoT in Hybrid Grid Integrated Renewable Energy Sources
In association withInternational Journal of Scientific Research in Computer Science, Engineering and Information Technology
ISSN : 2456-3307 (www.ijsrcseit.com)

Study of Various Models for Popularity Prediction in Online Social Networks and Designing Appropriate Model

Mr. S. N. Ugale, Dr. S.S. Sherekar, Dr. V. M. Thakare
SGBAU, Amravati, Maharashtra, India

ABSTRACT

Popularity prediction models of online social networking (OSN) have been still research area in OSN with lots of models which helps to designing the efficient popularity prediction. This paper is focused on analysis of five differentexisting popularity prediction techniques such as Social-Transfer, Two-stage prediction method, social-driven propagation dynamics-based popularity prediction model (SPDPM),popularity stage problem (PSP) algorithm, and Factor prior weighting (FPW) algorithm etc.But some pros and cons are observed in these techniques These approaches have been presented to address the limitations of popularity prediction system such aslow accuracy, less efficiency, high false rate, time consuming etc. So, to overcome the problems the proposed methodcalled “the online streaming with popularity prediction" is used to reduce the false rate, time, overhead, delay and to increase the accuracy and performance. Use of error rate as a metric, throughput, capacity to improve the popularity prediction rate.Leads topresenting a new method called “the online streaming with popularity prediction". This method is a combination of the Topic space, Social Transfer and the clustering techniques which is useful for predicting the bursty or sudden rise in the popularity of the posts or videos.

Keywords : Mobile ad hoc network, Capacity, delay, throughput, Opportunistic Routing, interference prediction and location prediction.

I. INTRODUCTION

Predicting the popularity of online contents means simply to understand the bursty video concept which exhibits the sudden rise in online contents. Such videos provide a unique opportunity for advertising and caching. An approach is to simply calculate the social prominence of a video by using the trends learned from Twitter as social sensors of video popularity [1]. Predicting the future popularity of online video content is measured by using the view count. Such meaningful knowledge can help the service providers to design more effective information services, such as video ranking, recommendation schemes and video searching. Hence,proposed a two-stage prediction method which is used for predicting the future popularity of online videos [2]. The author proposed a social-driven propagation dynamics-based prediction model which needs neither training phases nor prior

knowledge. Then, a discrete-time markov chain approach is proposed to predict the viewing probability of certain contents from the perspective of individuals [3]. The main aim is to model the dynamic evolution and focus on predicting multiple popularity stages of online contents in social media. The prediction of popularity suffers from one of 4 stages, the four stages are mainly- burst, tall, rise and valley [4]. The task of predicting the popularity is achieved by images shared on social media over time. Such kinds of task are known as "Popularity Dynamic Prediction" [5].

This paper is focused on evolution of five different popularity prediction technique such as Social-Transfer,Two-stage prediction method, Social-driven propagation dynamics-based popularity prediction model (SPDPM),popularity stage problem (PSP) algorithm, Factor prior weighting (FPW) algorithm.This paper proposes the online streaming with popularity prediction model for popularity prediction which is the combination of Social-Transfer and Factor prior weighting (FPW) algorithm along with topic space and clustering. It introduces the online streaming with popularity prediction model that is able to reduce the false rate, time, overhead, delay to enhance the prediction rate. The previous methods have some limitation and to overcome such problems, it improves the version of popularity prediction scheme that is "the online streaming with popularity prediction" model is proposed here that depend upon the bursty rise as well as the online streaming.

II. BACKGROUND

Many studies on popularity prediction of OSN have been done to develop the popularity prediction schemein recent past years.Such schemes are:

The transferlearning algorithm. It can learn topics from social streams and improve popularity predictions in the video domain. Social-Transfer is used to improve video popularity prediction as well as it can also be used to identify the fake videos and its fake popularity[1].Two-stage prediction methods are used for predictingthe future popularity of online videos. The proposed method slightly reduces the predictionerrors [2].Machinelearninglog-linear correlation model and evolutionary predicationmodel has proposed to improve the popularity prediction accuracy[3].To extract the pattern of popularity stage, apply the K-means clustering algorithm and also a one more scheme is used called popularity stage problem algorithm which simply combine the clustering results and the predictions results[4].The factor prior weighting (FPW) algorithm considers the contributions of each dynamic factor to the PSP problemandmethod compares the popularity achieved by different contents up to the current time step and performs a clustering over the set of defined behaviour patterns [5].

This paper analysed the various popularity prediction schemesi.e.Social-Transfer,Two-stage prediction method, social-driven propagation dynamics-based popularity prediction model (SPDPM),popularity stage problem (PSP) algorithm, Factor prior weighting (FPW) algorithm.The paper is organised as follows:

Section I Introduction. **Section II** discusses Background. **Section III** discusses previous work. **Section IV** discusses existing methodologies. **Section V**discusses and analyses attributes and parameters and how these are affected on popularity prediction models. **Section VI** gives proposed method **Section VII** gives outcome and possible result. **Section VIII** Conclude this review paper. Finally, **Section IX** gives future scope.

III. PREVIOUS WORK DONE

In research literature, many popularity predictions models have been studied to provide various popularity prediction schemes and to improve the performance,reduces the predictionerrors,identify

the fake videos and its fake popularity,and improve the popularity prediction accuracy.S. D. Roy et. al. (2013) [1] have worked on the transferlearning algorithm that can learn topics from social streams and improve popularity predictions in the video domain. Transfer learning aims to transfer knowledge to improve the learninginrelatedtargetdomain Tag enrichment by comment extraction or visual object understanding. Itimproves the prediction power of the model.

Shuxin Ouyang et al.(2016) [2] has proposed Two stage prediction model gainsignificant improvements of the predictionperformance.The proposed method has reduced the predictionerrors.Thelargerthepopularitylevel transition gives the more improvements in method gains.

Shuo He et al.(2017)[3]has worked on the Machinelearninglog-linear correlation model and evolutionary predicationmodel and proposed to improve the popularity prediction accuracy.

Qingchao Konget al.(2018) [4] has presentedBoth the FP-based method ad IPW algorithm, the classification algorithms, and both algorithmscombine the different classifiers to improve the prediction results. Clustering-based method is used to improve the performance ofthePSP-clustering method.

Alessandro Ortis et. al.(2019) [5]has presentedthe factor prior weighting (FPW) algorithm which considers the contributions of each dynamic factor to the PSP problem and also extracts the patterns of popularity evolution to facilitate the prediction of popularity stages.

IV. EXISTING METHODOLOGIES

Many popularity predictions schemeshave been implemented over the last several decades. There are different methodologies that are implemented for different popularity prediction models i.e. Social-Transfer,Two-stage prediction method, social-driven propagation dynamics-based popularity prediction model SPDPM,popularity stage problem (PSP) algorithm, Factor prior weighting (FPW) algorithm.

A) Social-Transfer:

Used to calculate the social prominence of a video by using trends learned from Twitter streams as social sensors of video popularity. The social prominence of a video is substantially responsible for its bursty popularity in the video domaindepends on the topic that a video belongs to. The real-time to image updating allows calculating the social prominence of a video in real-time, which makes prediction of bursty video popularity possible. Social influence is capable of improving performance and the results show that the online video content which will gain sudden bursty popularity. The framework of the proposed method is:

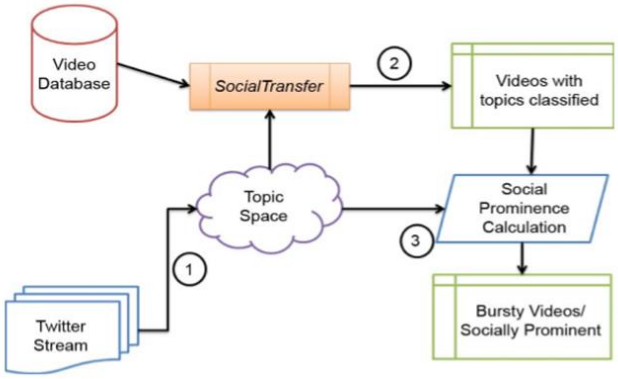


Fig 1.1: Interaction among different components of the system

(1) The OSLDA algorithm extracts topics from Twitter stream in real-time and populates the topic space, (2) The Social Transfer algorithm allows for classifying videos with social media topics by using this topic space and the transfer graph [1].

B) Two-stage prediction method:

It is used for predicting the future popularity of online videos. First, it estimates the future popularity level of a video based on a rich set of features and effective classification technique. Then, constructs regression model to predicts the future view count

for the video. Compare the prediction performance with two state-of-the-art baseline models of online video popularity prediction and tests such method on a real-world dataset.The multi-class classification model is employed in the popularity level prediction, and the regression model is used in the future view count prediction and it compare these two-stage prediction methods with two state-of-the-art baseline models for predicting the popularity of online videos.

$$\frac{(\text{Daily Increase in View Count})}{(\text{Total View Count})}$$

Historical Popularity Features calculates the ratio of daily increment score over total view count score[2].

C) Social-Driven Propagation Dynamics-Based Popularity Prediction Model (SPDPM):

The encoding method was proposed and optimized to release the burden of caching videos and also studied the content delivery strategies with coordinated multipoint (CoMP) technique. However, the mentioned works are all based on a common assumption: the popularity distribution of the contents known as the prior information. Machinelearninglog-linearcorrelation model and evolutionary predicationmodel has proposed to improve the popularity prediction accuracy. Contents are shared by users via social networks. Once content goes online, it will be propagated through online social networks over a period of time, e.g., minutes, hours and days. In order to describe the propagation mechanism of contents and the dynamic interactions between users, author adopts the SIR epidemic model.

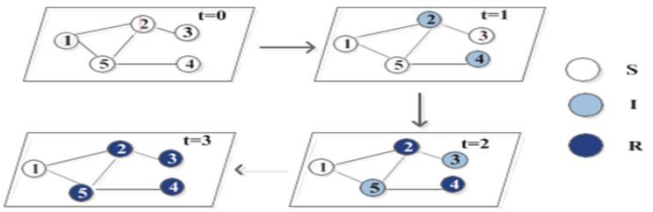


Fig 1.2: The diagram of spreading process of a content.

Users in different coverage areas of SBSs can interact with each other through online social networks, e.g., Facebook and Twitter. To model the propagation of contents among users, first define three states and approximate corresponding parameters under SIR model. Then, an undirected graph is constructed to describe the social reaction[3].

D) Popularity stage problem (PSP) algorithm:

A multiclass classification task is treated as a popularity stage prediction problem, i.e., the prediction result will be one of them, namely burst, fall, rise and valley. Itidentifies the evolution of the entire patterns of popularity stage and its effects on the prediction problem. Basically to extract the pattern of popularity stage apply the K-means clustering algorithm and propose a scheme that is called PSP algorithm which simply mixing the prediction results and the clustering results viz; (1) Dynamic Factors: A factor that continually modify or change when multiple users re-tweet the original tweet is called as dynamic factor. (2) Popularity Stage Clustering: Apply the K-means clustering technique to create a group or to create a cluster in which the sequences of tweets are collected. (3) Popularity Stage Prediction Algorithm: Propose a scheme that is called PSP algorithm which simply mixing the prediction results and the clustering results. The popularity prediction suffers from one of 4 stages of burst, valley, rise and valley. So, there is need to identify the patterns of stages of popularity with its effects. This paper proposed a technique called k-means clustering technique and also proposed a one more scheme called popularity stage problem algorithm which simply combining the clustering results and the prediction results. The dynamic factors are continually modifying or change when multiple users re-tweet or reposts the original tweet along with k-means popularity stage clustering takes place.

Index	Popularity Stage Evolutions										Cluster Size
1	F	F	R	F	F	F	R	F	F	F	1639
2	B	B	B	B	B	B	B	B	B	B	1312
3	F	V	V	V	R	F	V	V	V	V	1067
4	B	B	B	F	F	F	F	F	R	R	982

k-means clustering scheme, simply creates the group of the various classification, in which the posts sequence or tweets is collected. Suppose consider the 2000 posts or tweets and then apply the popularity stage clustering[4].

E) Factor prior weighting (FPW) algorithm:

The author defined three different experimental settings. Each experimental setting provides a certain quantity of knowledge about the shape and the scale of the sequences. The achieved error rates resulting from these experiments can be considered as lower bounds for any approach and that aims is to predict the popularity dynamic without any prior knowledge on the output sequence. The output sequence is obtained by combining the shape prototype assigned by the clustering procedure (*sshape*∗) and the ground truth scale value (*sscale*). This method achieves the minimum possible error. In this step the scale value is taken from the Ground Truth (*sscale*), whereas the sequence shape isclassifier (RNDF). This case exploits the shape prototype assigned as Ground Truth and combines it with the scale value inferred by the trained Support Vector Regressor(*s*ˆscale). Let *p*ˆ be the popularity score estimated by the SVR and the number of views is computed by applying the following formula: [5].

$$\hat{s}_{scale} = (e^{\hat{p}} - 1) \times n$$

V. ANALYSIS AND DISCUSSION

The chunks of tweets are provided to the social transfer algorithm in time-sequential batches based on the time slots [1].The results of twostagepredictiongives that the method leads to significant reductions in the prediction errors reaching to the baseline models respectively. Then based on the result of the first stageprediction forecast the precise value of the future video popularity[2]. The mRse is widely used to measure the mean prediction accuracy for a set of contents. The model outperforms BP and S-H with 86% and 94% on average, respectively[3]. The methods also predict the popularity growth, i.e., it analyzes there will be any burst in future regards with other stages. The results show that the methods achieve better performance against the baseline methods and thus verify the effectiveness of PSP. [4].The approaches are combined the results obtained by two different algorithms and aimed to estimate the maximum number of the views reached by the photo in the period of observation (scale) and the shape of the sequence[5].

Mobility scheme	Advantages	Limitations
Social-Transfer	Captured a great market and increase the graph of view count. Mapping the traditional videos to the video portal. Dealing with the great profit. Provides good opportu -nity for advertising, Caching and great marketing. Works on social prominence.	More chances of fake popularity. More chances of fake videos.

Two-stage prediction method	Delivering the network resources(e.g.bandwid -th and cache servers) can avoid the potential bottlenecks. Easily identifying the next rising star of online video. Design effectiveinfor -mation services, such as video-ranking, video-searching and recommendation schemes.	• More chances of fake popularity. More chances of fake videos.
Machinelearninglo g-linear correlation model and evolutionary predicationmodel	It allows for the consol -edation of customer data and the basis for deep insights. It speeds-up the sales conversion process. It increases staff productivity, lowering time-cost. Improves customer experience by allowing personalization and query resolution.	Security and data protectionwith centralized data. Time and initial productivity cost of implementation.
popularity stage problem (PSP) algorithm,	Used in Decision making for many applications. The proposed scheme can be applied on clustered data. Better performance efficiency.	Without cluster data such technique is difficult to implement. Costlier technique.
Factor prior weighting (FPW) algorithm.	To support the publication and effectiv -e diffusion of contents through social media. The proposed scheme can be applied on clustered data. Betterperformance efficiency.	Without cluster data such technique is difficult to implement. Costlier technique.

TABLE: Comparisons between different popularity Predictionschemes.

VI. PROPOSED METHODOLOGY

The model for Predicting the popularity of postsis important and crucial, the model isto analyse and discuss about various tasks based on different parameters i.e. accuracy, performance, use error rate as a metric, time, overhead, throughput, delay, capacity etc. for different mobility models.This new popularity prediction method called“the online streaming with popularity prediction” model is proposing here to overcome the problems of this model. As this model is dependingupon the online streaming along with the sudden rise in the popularity

As shown in Fig 1.3, the different components of this model are connected in the following way: (1) the system extracts topics from online social network stream in real-time and populates the topic space, (2) the *SocialTransfer* block allows for clustering videos and post with social media topics by using this topic

space and the transfer graph and label a video or posts with a topic learned from social domain. The block also allows for continuous updating of the transfer graph and seamless integration of fresh topics as newer social media data is encountered.(3) The system calculates the social prominence of each topic and make a prediction that postor videos with social prominence will demonstrate bursty behaviour i.e. sudden rise in popularity. This would be empirical evidence that popularity signal of social media traverses across domains to affect video popularity.

Basic steps of algorithm:

Step 1 : The system extracts topics from online social network stream in real-time with topic space.

Step 2 : *Social-Transfer* algorithm allows for classifying videos and posts with social media topics by using this topic space and the transfer graph.

Step 3 : continuous updating of the transfer graph and seamless integration of fresh topics as newer tweet data is encountered.

Step 4 : calculate the social prominence of each topic.

Step 5 : predicting the videos with social prominence will demonstrate bursty rise in popularity.

Diagrammatic representation of proposed method is shown as follows:

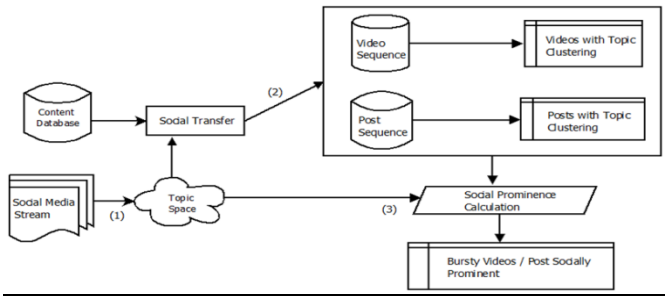


Fig 1.3: Block Diagram for Predicting Popularity of Posts/ Video.

VII. OUTCOME AND POSSIBLE RESULT

In this paper, the chunks of tweets are provided to the social transfer algorithm in time-sequential batches based on the time slots in which they are generated. The topic modelingrequires less time than the speed of incoming tweets. In this paper, the aim of the proposed approach is to calculate the maximum number of the views reached to the posts or videos in the period of observation. Individually, the two clustering are inferred, then the two results are combined to predict the popularity dynamics over *n* days.

VIII. CONCLUSION

This paper focused on the study of various popularity prediction models i.e.Social-Transfer,Two-stage prediction method, social-driven propagation dynamics-based popularity prediction model SPDPM,popularity stage problem (PSP) algorithm, Factor prior weighting (FPW) algorithm. But in previous methods there are some problems in predicting the popularity of posts so this proposed method improvesthis problem-solving technique by presenting topic space, social transfer algorithm along with the separation of posts and videos by doing clustering.

IX. FUTURE SCOPE

From observations of the proposed method the future work will include the exact accuracy and false news prediction of popular posts with the help of more close form of mathematical expression.

X. REFERENCES

[1]. Suman Deb Roy, Tao Mei, Wenjun Zeng, Fellow and Shipeng Li,"Towards Cross-Domain Learning for Social video popularity prediction",IEEE Transactions on multimedia (Volume: 15 , Issue: 6 , Oct. 2013).

[2]. Shuxin Ouyang, Chenyu Li and Xueming Li,"A Peek Into the Future: Predicting the Popularity of Online Videos",IEEE Access (Volume: 4, Issue: 06, JUNE 2016).

[3]. Shuo He, Hui Tian, and Xinchen Lyu "Edge Popularity Prediction Based on Social-Driven Propagation Dynamics", IEEE Communications Letters (Volume: 21 , Issue: 5 , May 2017).

[4]. Qingchao Kong, Wenji Mao, Guandan Chen and Daniel Zeng, "Exploring Trends and Patterns of Popularity Stage Evolution in Social Media",IEEE Transactions on systems, man, and cybernetics: System Systems, Man, and Cybernetics: Systems, 10.1109/TSMC.2018.2855806.

[5]. Alessandro Ortis, Giovanni Maria Farinella and Sebastiano Battito,"Predicting Social Image Popularity Dynamics at Time Zero", IEEE Access, 10.1109/ACCESS.2019.2953856.

Shubham N. Ugale has completed B.Tech Degree in Information technology from Dr. Babasaheb Ambedkar Technological University, Lonere, Raigad, Maharashtra. He is perusing Master's Degree in Computer Science and Information Technology from P.G. Department of Computer Science and Engineering, S.G.B.A.U. Amravati. (E-mail id: shubhamugale0218@gmail.com)

Dr. Swati S. Sherekar Presentlyworking as Professor in the P. G.Department of Computer Science andEngg and having 25 years of teachingand research experience. Her area ofresearch is Network security, MobileComputing, Cloud Computing and working as supervisor for Ph.D.guidance. Completed one MRP ofUGC one MRP of AICTE. She haspublished more than 70 papers inInternational & National level Journals and also Invited for Talks inmany International / NationalConferences and National LevelConferences. (e-mail id:ss_sherekar@rediffmail.com)

Dr. Vilas M. Thakare is Professor and Head in Post Graduate department of Computer Science and Engg, Faculty of Engineering & Technology, SGB Amravati university, Amravati. He is also working as a coordinator on UGC sponsored scheme of e-learning and m-learning specially designed for teaching and research. He is Ph.D. in Computer Science Engineering and completed M.E. in year 1989. He has done his PhD in area of robotics, AI and computer architecture. His area of research is Computer Architectures, AI and IT. He has published more than 150 papers in International & National level Journals and also International Conferences and National Level Conferences. (e-mail id: vilthakare@yahoo.co.in)

The International Conference on Research Perspectives : IoT in Hybrid Grid Integrated Renewable Energy Sources
In association with International Journal of Scientific Research in Computer Science, Engineering and Information Technology
ISSN : 2456-3307 (www.ijsrcseit.com)

Prediction of Crop Yield and Cost by Finding Best Accuracy using Machine Learning Approach

Swathi, Mrs. Soja Rani
Department of CSE, New Horizon College of Engineering, Bangalore, Karnataka, India

ABSTRACT
Among around the world, agribusiness has the significant duty regarding improving the financial commitment of the country. Still the most agrarian fields are immature because of the absence of arrangement of biological system control advances. Because of these issues, the yield creation isn't improved which influences the farming economy. Subsequently an improvement of rural profitability is upgraded dependent on the plant yield expectation. To forestall this issue, Agricultural areas need to anticipate the yield from given data set utilizing AI procedures. The outcomes show that the viability of the proposed AI calculation strategy can be contrasted and best exactness with accuracy.

Keywords : Dataset, Crop Yield, Machine Learning- Classification Method.

I. INTRODUCTION

In agricultural nations, cultivating is considered as the significant wellspring of income for some individuals. In current years, the rural development is locked in by a few advancements, conditions, methods and civic establishments. Moreover the usage of data innovation may change the state of dynamic and in this manner ranchers may yield the most ideal way. For dynamic cycle, information mining procedures identified with the horticulture are utilized. Information mining is a cycle of separating the most huge and helpful data from the gigantic measure of data sets. These days, we utilized AI approach with created in harvest or plant yield forecast since agribusiness has distinctive informationlike soil information, crop information, and climate information. Plant development expectation is proposed for checking the plant yield adequately through the AI strategies. It is additionally relevant for the computerized cycle of cultivating is the start of another time that will be reasonable for the ranchers who look for specialists to take proposal about the fitting yield on explicit area of their property and don't have any desire to fail to remember any progression of the development all through the cycle.

II. CROP YIELDPREDICTION

Crop yield prediction that is a fundamental endeavor for the pioneers at public and typical levels for enthusiastic dynamic. Careful gathering of yields understand model can help farmers with picking what to make and when to make. There are different approaches to manage administer oversee crop yield

assumption. Different advances that can be associated with crop yield forecast are information obtaining, information pre-handling, highlight choice, order and forecast. The harvests that were considered in the model for forecast incorporate coriander, beats, cotton, paddy, sorghum, groundnut, sugarcane, banana and vegetables. Various credits of the dirt were considered in request to anticipate the harvest, which included pH, profundity, disintegration, surface, waste, dater holding also, soil tone. The arranged work presents effective degree affordable crop proposal framework. Utilization of guileless mathematician makes the model horrendously practical as far as calculation.

III. DATASETS

A mix of models is a dataset and when working with AI procedures we normally need a couple datasets for various purposes. Preparing Dataset: A dataset that we feed into our AI check to set up our model. It could be known as the underwriting dataset.

A. Preparing the Dataset

The demo dataset is right now gave to AI model dependent on this educational assortment the model is readied. Each new detail involved at the hour of design goes probably as a test instructive assortment. After the movement of testing, model estimate reliant on the inferring it closes dependent on the readiness educational assortments. Satellite Imagery (Remote Sensing Data), has been comprehensively used for predicting crop yield. This dataset is accumulated using the sensors mounted on satellites or planes, which recognize the energy (electromagnetic waves), reflected or diffracted from surface of the earth. Inaccessible distinguishing data has a lot of energy gatherings to bring to the table, anyway basically relatively few of them have been used for crop yield conjecture. In any case, there are a couple of gathering who have made a pass at creating material features using the gatherings which are typically dismissed, and they have been productive with improving results with that. In case of this dataset, by far most only sometimes research the high- demand depictions of the features. Considering these datasets people have used computations like Regression models, Random Forest and Nearest Neighbor, etc

Variable	Description
Crop	Crop name
State Name	Indian state name
District Name	District name list of each state
Cost of Cultivation	Cultivation amount for C2 Scheme
Cost of Production	Production amount for Scheme
Yield (Quintal/ Hectare)	Yield of crop
Crop year	Crop year list
District Name	District name for each state
Area	Total area of each place
Rainfall	Water availability of each crop
Average Moistness	Straightforwardly impacts the water relations of plant and by implication influencesleaf development
Mean Temperature	Climate of r each particular crop

Fig.1. Table shows details of the datasets

B. Exploratory Data Analysis

In this piece of the information, you will stack in the data, check for cleanliness and a short time later trim and clean your dataset for examination. Guarantee that you fileyour methods carefully and legitimize your cleaningdecisions.

C. Training Dataset:

The first line imports iris instructive assortment which is as of now predefined in sklearn module. Iris enlightening assortment is basically a table which contains information about various combinations of iris blooms.

- For model, to import any computation and train_test_split class from sk-learn and numpy module for use in this program.
- This procedure secludes dataset into getting ready and test data discretionarily in extent of point we encapsulate any computation.
- In the accompanying line, we fit our planning data into this estimation so PC can get readied using this data. By and by the readiness part is done.

D. Testing dataset:

- Now we have measurements of another bloom in a numpy cluster called 'n' and we need to foresee the types of this blossom.
- We do this utilizing the foresee technology which accepts exhibit info which lets active objective incentive as yield.
- Objective worth recovers out to be 0. At last discovery grade which is the respective no. of expectations discovered right complete forecasts made.
- This utilizing the technique that fundamentally analyzes an real estimations of the test set with the anticipated qualities.

IV. MACHINE LEARNING

Machine learning is concerned with learning the computer programs and improve automatically with the experience. However there is no clear idea about how to make computers learn as we humans do but there are several algorithms that are used for certain type of learning task. From age old year's humans worked under every different categories as time passed machines came into use which was trained with algorithms and spoon fed how to work then complete a given task and now it is the decades of the concept ML where in machines are not trained how to work in real but allowed itself to learn from experience using the input, analyze the data by itself and give the results. The input can be divided into 2 sector training data and testing data. Training data is used to guide the system; the system will learn from these data and will produce the output. Later comes the role of testing data where the trained system is checked for its correctness. Machine learning algorithms also performed in the field called data mining to determine the knowledge from the databases, loan transactions, and financial transactions.

A. Types of Machine Learning

There are three sector of ML techniques which is based on the input type, input length, time duration given to solve etc.

❖ Supervised Learning

In this technique set of data in the form of examples need to be provided. Each example will be associated with labels; a learning algorithm is to be given to example. It is possible for this technique to predict the label for the examples that never seen before when it is fully trained. Since supervised learning focuses on singular task, so we can feed more and more examples until it performs the task accurately.

Supervised learning can further be divided into two kinds of problems regression problems and classification problems. Regression problems have to find or draw a linear line (in case of linear regression) which classifies that given data maximally correct. As the name suggests classification problem has to classify the given input correctly based on the other inputs whose classification is previously known. A simple example is given input data of which is the suitable day to play tennis in the form of attributes such as temperature, wind, outlook etc., and giving the testing input to check if tennis can be played on a particular given day.

❖ Unsupervised Learning

It is too sensitive converse to the supervised learning because we need to provide lot of data to understand its properties, as the data are unlabeled Unsupervised learning is said to be data driven because it is mainly based upon data and its properties. The outcomes are based on data and the way they are formatted. As in supervised learning the input data consists of inputs and also its possible classification (in terms of target attribute) the unsupervised learning does not consist of target attribute, only input data is given and based on the common things/ similarities the system should group the data and provide the output. Example is given the input of number of patients who are having diabetes, attributes such as their age, height, sex, blood pressure etc., are also given, now the system has to learn by itself from these data and when a testing input is given it should predict whether that person has diabetes or not.

❖ Reinforcement Learning

Let us consider a simple example of playing checkers game, in this game we make several moves and at the end the result is either win, loss or a draw. Let us assume that in the first game you lost the game so while playing second round you will think which are the moves you took in the first game which led to failure and change your some of the moves accordingly (we cannot say that all the moves made in the first game is wrong, some moves made may be the reason for the failure) and eventually win the game. So Reinforcement Learning is somewhat different from supervised and unsupervised learning as it will learn from mistakes. For support learning, we need a specialist and a climate. To associate the climate and the specialist we need to give two signs: refreshed state and award.

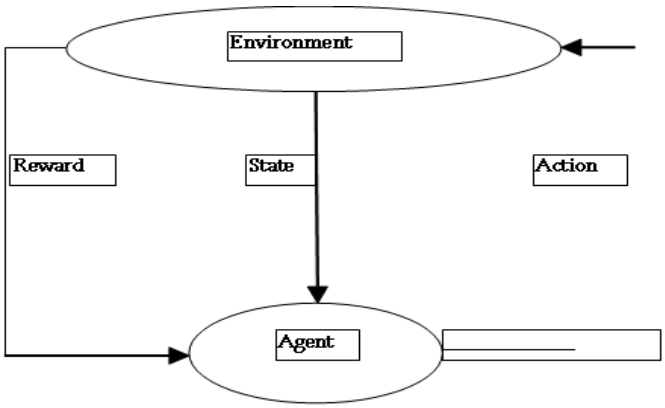


Figure 2: Reinforcement Learning

In the above diagram it is clearly explained how a reinforcement learning works, the environment gives the changes in the environment as input to the learning system called agent, the agent using its previous experience takes an action and depending on its decision the environment either provides rewards or blames the system. This is represented in equation form.

B. Designing a learning system

- Choosing the preparation experience: this is initial step included where we will pick the preparation information. This preparation information will have an impact on progress or disappointment of the student.

There are several key attributes. The first key attribute is providing a direct or indirect feedback about the choices made. Second is how much the student controls succession of the preparation models.

The third characteristic is the means by which well the circulation of the models over the last framework execution is addressed. That is if the preparation models follow the conveyance like the future test models then the learning is more solid.

- Choosing the target function: here we need to think about what kind of knowledge is gained and used by the performance system. There will be several optimization problems encountered such as scheduling and controlling manufacturing processes where the steps are understood but the strategy for sequencing is not.
- Choosing a Representation for the target function: We need to select a representation that will be crucial and very expressive such that it is almost approximate to the ideal objective capacity. Assuming the portrayal is expressive, we can pick among the elective speculation it can address.
- Choosing the capacity guess calculation: We require a bunch of preparing guides to gain proficiency with the objective capacity to do this we need to initially appraise the preparation esteems and afterward change the loads appropriately so it best fits the preparation models and diminishes the squared mistake between the preparation esteems and the qualities that are anticipated.
- Final Design: It contains four distinct component
  - The performance system: It takes new problem instance as input and output as a trace of solution. It is mainly used to solve given performance task .The performance should increase as the evaluation function becomes more accurate.
  - Critic: takes the input as history and output as set of training examples. Critic is nothing but it corresponds to the training rule.
  - Generalizer: The input is a training examples obtained from the critic and the hypothesis is the output. The hypothesis will be generalized such that it covers all the training examples and the cases beyond the training examples.
  - Experiment Generator: input will be a current hypothesis and the output will be new problem for the performance system in this manner it increases the learning rate of the system

C. Commonly used Machine Learning algorithms

Some of the most commonly used ML algorithms are

1. **Decision Trees**
2. **Support Vector Machines(SVM)**
3. **Naïve Bayes**
4. **K-Nearest Neighbors (KNN)**
5. **K-Means**
6. **Random Forest**

1. Decision Trees

The most frequently used supervised learning algorithm is decision tree. It works well for both categorical (discrete) as well as for continuous dependent variable. The various distinct groups are made. In simple words the structure is similar to a tree present in data structures. The tree consists of a main node *Root* internal nodes which is nothing but attributes given in input labeled based on some given conditions(test on the attribute) and finally the last nodes are the leaves which is the target concept / possible outputs for a given problem statement, branches are the values obtained after the attribute is tested.

2. Support Vector Machines

It is a support learning algorithm. In this modular the points will be represented in N dimensional space so that examples with different features will be clearly separated by a large gap. The new instances will be then classified to the category in which there is small gap. SVM can easily perform both linear and non-linear classification using kernel trick that is it maps

the inputs to high dimensional output. Nonlinear SVM is nothing but the boundary that the algorithm calculates is not a straight line. SVM is very much suitable in the following cases:

- When there is huge number of training data.
- When the number of zero values are more.
- To solve problems like image classification, genes classification etc.

There are several drawbacks as well:

- Input data has to be labeled.
- Probabilities of estimated data has finite data will not be estimated.
- It is difficult to interpret the parameters of solved model.
- SVM is applicable only for class two tasks. If there is multi class task binary problems have to be applied.

3. Naïve Bayes

It is one of the highly sophisticated classification methods very much useful for larged at a set. It is mainly based on Bayes' Theorem. Naive Bayes classifier assumes that each of the feature is independent of the other. The Naive Bayes classifier will consider all properties contribute independently to the probability even if the features are related on each other.

4. K-Nearest Neighbors

It is commonly used to solve the classification and regression problems. KNN algorithm will store all the available cases and when a new case is encountered the case will be assigned for class which is pure common among the K nearest neighbor calculated by Distance function. There are various distance functions such as Euclidian, Manhattan and Murkowski used for similarity of function and hamming distance used for particular sector variables.

5. K-Means

It is unsupervised centroid based algorithm used for clustering. It mainly aims at forming a k clusters from observations. Clustering is done based on the nearest mean value. In order to pause the clustering in K-means algorithm

- Do not change the centroid of newly formed cluster.
- All the points should remain in the same cluster.
- Maximum iterations has to be reached.

6. Random Forest

Random forest is a collection of decision tree mainly used for classification and regression. It is supervised algorithm that doesn't over fit the model, handles the missing value and mainly modeled for categorical value. It is better than decision tree because it contains only the subset of features

D. Disciplines of machine learning

❖ Artificial intelligence

Symbolic representation of concepts are learnt using artificial intelligent. It is also used as approach to improve problem solving. Using prior knowledge along with training details used as guide for learning.

❖ Bayesian method

The best hypothesis which is most probable can be determined by using Bayes' theorem. It is important in machine learning because it provides quantitative approach in identifying the evidence supporting alternating hypothesis. It provides basis for learning algorithm as it will directly manipulate probability and provides framework for analyzing the operation of algorithm.

❖ Computational complexity theory

The multifaceted design of different learning task are assessed by the computational effort, bungles made, planning models, etc

❖ **Control theory**

To predict the next state of the process and also to optimize the predefined objectives we need to learn the control process Information theory learning to provide minimum description length. For encoding hypothesis with optimal codes and their relationship to optimal training sequences.

❖ **Philosophy**

To determine the best hypothesis that is simple. To generalize the observed data analyzing the justification.

❖ **Psychology and neurobiology**

According to the power law for a very wide range of learning problem response time of people increases with practice. Artificial neural network models of learning are motivated by Neurobiological studies.

❖ **Statistics**

When estimating the accuracy of the hypothesis based on the data given characterization of errors may occur.

E. Applications of machine learning

❖ **Recognition of spoken words:**

SPHINX system is one such that recognizes the primitive sounds and words from the speech signal and learns the speaker specific strategies. There are certain methods that are effective for customizing individual speakers, vocabularies, background noise such as methods for hidden Markov models, Neural network learning methods.

❖ **Driving an autonomous vehicle:**

The PC controlled vehicles are prepared to guide accurately when driving on assortment of street types. ALVINN framework is trained to mimic the steering commands of human driving the vehicle and it was successful in driving on public highways at speed up to 70 miles per hour and for a distance of 90 miles.

❖ **Classifying the astronomical structure**

In order to classify the celestial object decision tree learning algorithm have been used by NASA. It will automatically classify objects in the sky survey.

V. REFERENCES

[1]. Guanghui1,DongryeolRYU2,∗,JIAOCaixia1 and HONG Changqiao1 Estimation of Organic Matter Content in Coastal Soil UsingReflectance Spectroscop Research 2015

[2]. ManashProtimGoswami , BabakMontazer, and UtpalSarma, Member, IEEE Design and Characterization of a Fringing Field CapacitiveSoil MoistureSensor 2018

[3]. Zhiqiang Cheng, JihuaMeng * and Yiming Wang Improving Spring Maize Yield Estimation at Field Scale by Assimilating Time-Series HJ-1 CCD Data into the WOFOST Model Using a New Method with Fast Algorithms2016

[4]. Becker-Reshef, E. Vermote, M. Lindeman A generalized regression-based model for forecasting winter wheat yields in Kansas and Ukraine using MODIS data2010

[5]. Kim, NariLee,Yang-Won Machine Learning Approaches to Corn Yield Estimation Using Satellite Images and Climate Data: A Case of Iowa State 2016

[6]. SudhanshuSekharPanda, GerritHoogenboom, and Joel O.Paz Remote Sensing and Geospatial Technological Applications for Site-specific Management of Fruit and Nut Crops: A Review 2010

[7]. K. Sathishkannan; G.ThilagavathiOnline farming based on embedded systems andwireless sensor networks2013

[8]. B. Abishek; R. Priyatharshini; M. AkashEswar; P. DeepikaPrediction of effective rainfall and crop water needs using data miningtechniques 2017

[9]. Crop Yield Estimation Using Time-Series MODIS Data and the Effects of Cropland Masks inOntario, Canada 2019

[10]. Sheena Angra; SachinAhujaMachine learning and its application2017

[11]. FilippoSciarroneMachine Learning and Learning Analytics: Integrating Data with Learning2018

[12]. Pavan Patil1, Virendra Panpatil2, Prof. ShrikantKokate Crop Prediction System using Machine Learning Algorithms2020

[13]. Md. TahmidShakoor, KarishmaRahman, Chakrabarty.2017."Agricultural Production Output Prediction Using Supervised Machine LearningTechniques".978-1-5386-3831-6/17/$31.00 ©2017 IEEE

[14]. I. Ahmad, U. Saeed, M. Fahad, A. Ullah, M. Habib-ur-Rahman, A. Ahmad, J.Judge Yield forecasting of spring maize using remote sensing and crop modeling in Faisalabad-Punjab

[15]. S. Pudumalar, E. Ramanujam, R. H. Rajashree, C. Kavya, T. Kiruthika and J. Nisha, "Crop recommendation system fo r prpecision agriculture," 2016 Eighth International Conference on Advanced Computing (ICoAC), Chennai, 2017, pp. 32-36. doi: 10.1109/ICoAC.2017.7951740.

[16]. Naive Bayes classifier available at https://en.wikipedia.org/wiki/Naive_Bayes_classifier

[17]. S Brunda1, Nimish L2, Chiranthan S2, Arbaaz Khan2 Cro p Price prediction using Random Forest and Decision Tree Regression2020

[18]. s://medium.com/swlh/random-forest- and-its-implementation-71824ced454f

[19]. Y. Peng, C. Hsu and P. Huang, "Developing crop price forecasting service using open data from Taiwan markets," 2015 Conference on Technologies and Applications of Artificial Intelligence (TAAI)

[20]. Ananya Roy; Prodipto Das; Rajib DasTemperature and humidity monitoring system for storage rooms of industries2017

The International Conference on Research Perspectives : IoT in Hybrid Grid Integrated Renewable Energy Sources
In association withInternational Journal of Scientific Research in Computer Science, Engineering and Information Technology
ISSN : 2456-3307 (www.ijsrcseit.com)

An Overview of Machine Learning on Heart Disease

Dr. Clara Kanmani*, Harshitha M
Department of CSE, NHCE, Maharashtra, India

ABSTRACT
As the use and the advancement in technology is increasing in today's world, this leads to the exposure of new technologies to detect heart disease, which in turn leads to the increase in detection of disease as early as possible. Hence, there are many technologies to detect the heart disease but some may be time consuming and few may be costly. Hence the Machine Learning (ML) algorithm will help to detect the disease at early and with low cost. In the paper we are understanding about the machine learning algorithm that are used in detection of heart disease.

Keywords - Heart Disease, Machine Learning, Diagnosis.

I. INTRODUCTION

The major challenge of the Healthcare industry is that it faces now-a-days is superiority of facility. The diagnosis of disease correctly & providing effective treatment to patients will define the quality of service. The Poor diagnosis will cause a disastrous consequence that will not be accepted. Records or data of medical history is very large, but these are from many dissimilar foundations. The interpretations which are done by physicians are the most essential components of these data. The data in real world might be incomplete and inconsistent, noisy, so data preprocessing will be required to fill the omitted values in the database. In ancient years cardiovascular diseases is found as the important source of death in world, these have been announced as the most avoidable and manageable diseases. The complete accurate management of a disease will rest the well-timed judgment of that disease. A correct and methodical tool for recognizing high-risk patients. The mining data for timely analysis of heart infection looks a serious want.

Even though heart disease is acknowledged as the supreme chronic sort of disease in the world, it can be most avoidable one also at the same time. A timely analysis (inferior prevention) and healthy way of life (main prevention) are the two major origins of heart disease director. Conducting steady check-ups (inferior prevention) will shows a outstanding role in the judgment and also in the early prevention of the heart disease. Several tests comprising of echocardiography, chest X-rays, angiography, and exercise tolerance test support to this significant issue. These tests are expensive and that may involve the availability of accurate medical equipment. Researchers make use of several data mining techniques that are available to help the specialists or physicians identify the heart disease. Commonly used procedures used are k-nearest, decision tree, and Naïve Bayes. There are many other different

classification-based techniques used are bagging algorithm, kernel density, straight Kernel self-organizing map, and sequential minimal optimization and neural networks, and SVM (Support Vector Machine).

Machine learning algorithms are programs that will not require human intervention and can learn itself from data and help to improve from the experience. Learning tasks which will contain learning the function and also will map the output from the input, learning from the concept of hidden structure in the data of unlabelled or from the instance-based learning, where the class label would be produced to a new instance. That will be stored by the help of comparing the new instance (row) value to the instances from the training data. Instance-based learning will not create an abstraction from specific instances. Machine learning will allow doctor to speed up the process of heart detection and it will provide the accurate results and also to put preventative measures in place.

These algorithms will combine the beliefs of three classifiers: Naïve Bayesian, k-Nearest Neighbor (kNN), and Decision Tree. Then these three algorithms showed the effect to get the value than the individual algorithm used. That is, the way in which they combine all the three specified algorithms to get the effective result [1]. ECG is a test that will measure the heart electrical activity, and it will provide the valuable information regarding the heart's status. The paper tells how the classification method will be used for extracting multi-parametric features which will be obtained by analyzing HRV from ECG. Also tells about the heart disease pattern. The method used is classifier which will be based on the efficient FP-growth method. They conducted an experiment for associative classifier that will utilizes a set of multiple rules and pruning, and along with the biased confidence [4]. These papers will present a new heuristic for computing efficiently sparse kernel in SUPANOVA. This paper has applied a benchmark Boston housing market dataset. The non-invasive measurement of the heart activities based on magnetic field produced by the human heart. On this prediction of data, 83.7% predictions were correct. The exceeding of the results obtained using the equivalent kernels and standard Support Vector Machine. Equally good results were obtained by the spline kernel on a benchmark Boston housing market dataset [7].

Since the early 1980's, there has been a growth in hospital information systems (HIS) for storing a huge amount of data. Thus, data mining methods will help for obtaining the patterns which are interesting from the databases for the reuse of data that has been stored. This is necessary in the field of the medical research and practice, as the human beings will not be able to deal with such large amounts of data. The article focuses on the characteristics of medical data and how the data mining will help to store the large amount ofdata [8]. Hand and foot disease along with the tetanus is the serious infection that occurs in less income countries. This type of infection will often affect the larger number of infants along with the young children. The death will be due to the cause of the automatic nervous system dysfunction. Early detection of disease is much difficult [10]. At present the Heart failure disease has become one of the multifaceted diseases affecting huge amount of people worldwide. At initial stage Hospitals and cardiac centers would heavily depend on ECG, which is used as a general tool for evaluating and the diagnosing disease [12].

II. MACHINE LEARNING

Machine learning algorithms are those which contain programs that will not require the help of the human intervention and hence they can learn by itself from

the data and will improve from the experience. The concept of learning tasks will include the concept of learning, that will help to map the output from the input, learning structure which is hidden is unlabeled data or the instance-based learning, in which a class Iabe! will be produced to obtain a new class of instance that will be stored ay comparing values of the new instance (row) to the value of the instances from the set of the training data. 'Instance-based learning' will not create the abstraction for the specific instances.

Machine learning will allow the doctor to increase the process of the heart detection and will also provide the accurate results. To get to know how the machine learning will diagnosis one must have to know how the machines will be able to sense the large amount of data. Ontology is those where systems are made up of distinct things that will be known as entities and along with their relationships. Consider the following which tells a very simple ontology consisting of different types of coffee.

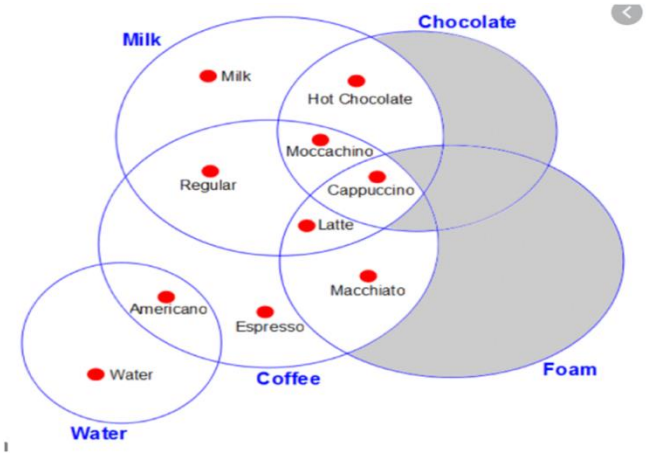


Fig. 4. Example of Venn Diagram.

Consider a simple Venn diagram, the individual ingredients which are the entities, but which will form an ontology that contains a set of relationships. From above ontology, we can obtain a combination of all ingredients milk, coffee and a foam that become a lattice.

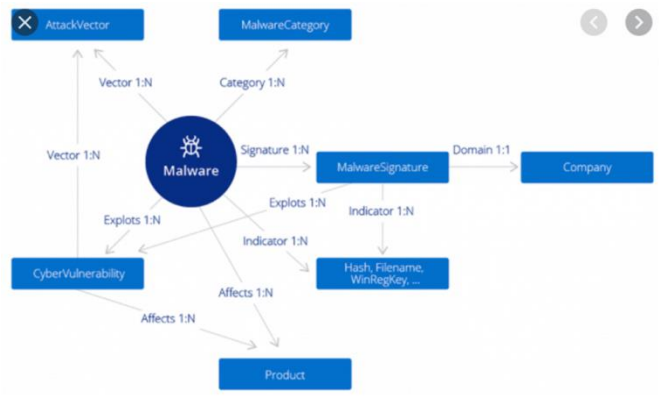


Fig. 5. Example of cyber security.

In the concept of cyber security shown in Fig, ontology is used to specify the real world inside the concept of the machine-learning environment. That means this ontology, see malware sitting at middle surrounded by various entities that could similar to that malware.

In this exist three types concept of machine learning (ML) algorithms as follows:

A. Supervised Learning Algorithms:

Supervised learning will use labelled training data value to learn the concept of the mapping function for the input variables value into the output variable (Y). This will help us to accurately obtain the outputs when an input of new is given.

In this, two types of the concept of supervised learning are there

Classification is a way that is used to identify the outcome from the given sample variable will be in the form of the categories.
The model of the classification looks at the value of the input data and will try to identify the labels like, "sick" or," healthy."

Regression will be used to identify the outcome of sample given when the output variable will be in form of the real values. Consider an example, a

regression - based mode! may process the values of input data to identify the height the person, amount from rainfall, etc.

Ensembling is also another type in the supervised. This will combine the multiple of predictions machines learning models which are the individually weak to obtain more accurate prediction on the newly generated sample.

B. Unsupervised Learning Algorithms:

Unsupervised learning models are used when there is the input variables (X) and no corresponding output variables. There are three types in the concept of the unsupervised learning:

The **Association** is a way that is used to find the probability of items that wi I co- occurrence in collection set.

Clustering method is a way that is used to combine samples which will oe that the objects that are in tne same cluster will be those ones that are those matching to each other than to the objects that are obtained from the different cluster.

Reinforcement learning is a way that allows to decide by an agent for the best next action that will be based on the state of current state after learning behaviours that will allows a reward to maximize. Algorithms of reinforcement will learn the optimal actions through the trial and error.

1) Random Forest

It is a classifier, which is ensembled classifier using many decision trees models. It is a collection of unpruned CARSTS that follows some designed rules for tree growing. A random seed is that will be chosen from the pulls out of the random collection of samples that is in the form of the training dataset. There is a huge advantage in the random forest where it can be used in the concept of both classification and concept of regression problems. The below figure shows how random forest would look like.

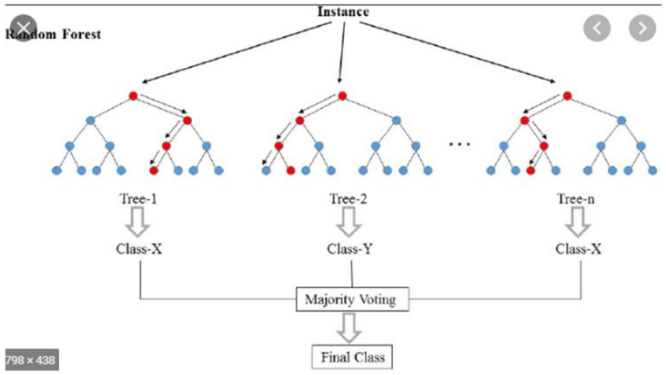


Fig.1. Random forest

2) Logistic Regression

Linear regression will help to predict values as the continuous values whereas the logistic regression will prediction will be in a discrete value after applying a transformation function.

Logistic regression is defined for binary classification: that is data sets that is where y = 0 or 1, 1 will represent the default class. For example, in the prediction of whether a seminar will happen or not, there are will be only two possibilities' are that it happen (which we denote as 1) or that it not happens (0).

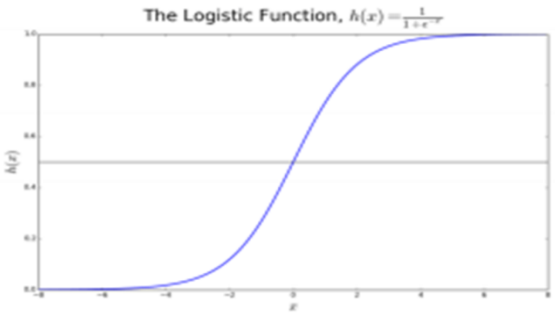


Fig.2. Logistic Regression

In the above Figure, to obtain whether a tumour will be malignant or not, the default variable would be y = 1 then it will be tumour = malignant. The x variable that will be a measure of the tumour, that is the size of the tumour. As shown from the figure, the logistic function will be able to transform the x-value of the

various instances of data set, into the limit of value 0 to 1. If the probability of value crosses the threshold of the value 0.5 (shown as horizontal line), the tumour would be classified as malignant.

The logistic regression of equation is of form $P(x) = e \wedge (b0 + b1x) / (1 + e(b0 + b1x))$ can transform into $ln(p(x) / 1-p(x)) = b0 + b1x$. The aim of the logistic regression will be to use the available training data and then to search the values of coefficients that are b0 and b1 in a way that it will minimize the error from the predicted and to the actual outcome.

3) Decision Trees.

Non-terminal nodes of Classification and Regression Trees will be the root node and other will be the the internal node. The terminal nodes is the leaf node, each nonterminal node will represent a input variable(X) and also a splitting point for that variable, leaf node represent the output variable, go through the tree to arrive at the leaf node and output the value arrived at the leaf node. The decision tree shown below will classifies the person whether he will buy a car or minivan depending on their age and marital status. If the person is above 30 and that person is not married, we walk over the tree as shown below follows 'over 30 years?' -> yes -> married?' -> no. Hence, this model will result as a sports car.

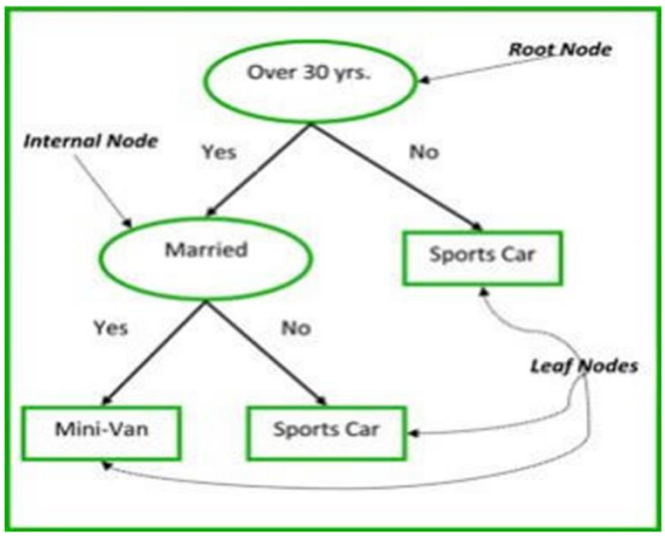


Fig.3. Some Parts of a decision tree.

4) K- Nearest Neighbour

Algorithm will use the whole data set as the concept of be training set, instead of the splitting of the data set into the concept of be training set and the test set. When the outcome is needed to obtain data instance of new data, thenthe concept of the KNN algorithm will help to go through the set of the entire data to obtain the k nearest instances for the fresh instance, or a k number of instances will be the most similar to the new record, and then outputs the mean of the outcomes that is a regression problem. The value of k will be specified by the user.

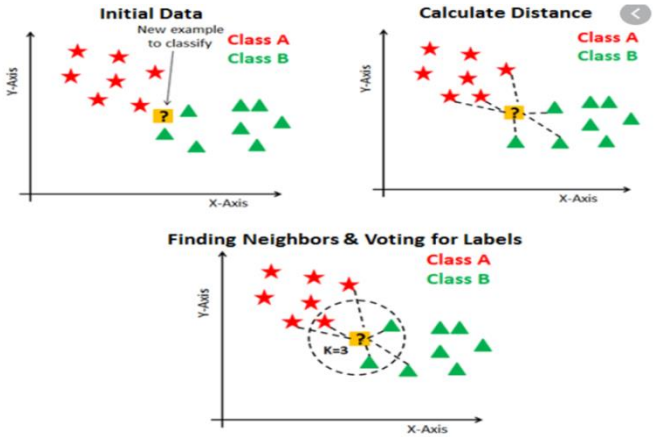


Fig.4. KNN algorithm.

5) Support vector machines (SVMs)

An SVM mode! is a concept of the different classes that is represented in concept of hyperplane in multidimensional hyperplane in multidimensional space. In the hyperplane that generated will be in an iterative manner that will be formed by SVM and so that the error that is occurred can be minimized. The goa! of the SVM will be to divide the datasets into the concept of classes and then to find a maximum marginal hyperplane (MMH).It is a type of supervised algorithm and also can be used in the concept of both regression or classification challenges. This method will be used by in classification problems.

In the algorithm of SVM, first the plots of each of the data item will be as a point in n- dimensional space (which means n represent the number of features that you contain) along with the value of each feature

that will be the value of coordinate that will particular. Then, on that perform some classification and this can be done by finding the hyper- plane. This hyper plane will differentiate between the two classes very well.

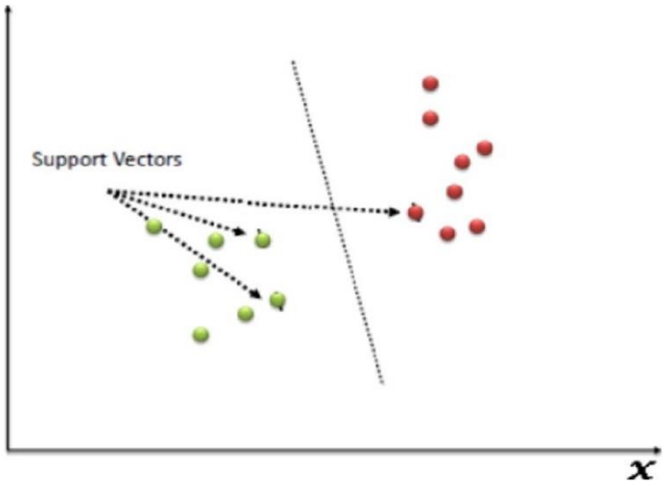


Fig. 5.1. SVM algorithm

The SVM classifier is one which has a 'frontier which best segregates the two classes (hyper-plane/ line). Identify the plane that is the right hyper-plane in the following condition:

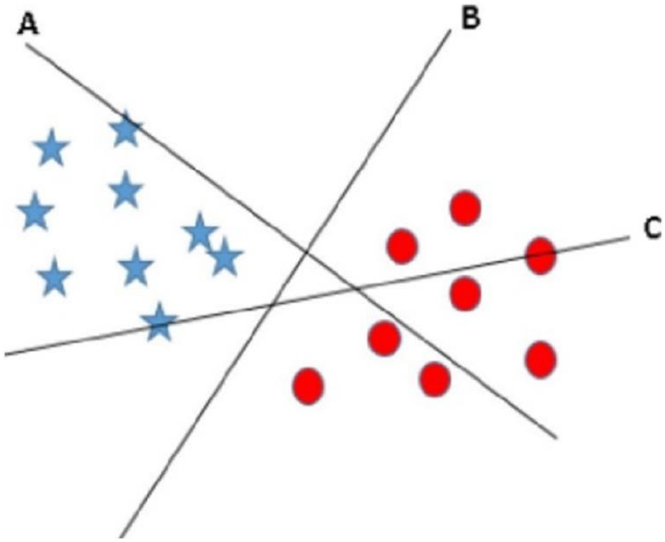


Fig.5.2. SVM algorithm Scenario

One must need to keep in mind about a thumb rule that will help to identify the plane that is the right hyper-plane. First select hyper-plane which segregates the two the two classes in a better. Here B has done the excellent job.

III. MACHINE LEARNING USES

Machine learning algorithm is not only used in detection of heart disease or other heart related problems but it can be used in many other fields like detection of spam emails where it can identify whether the email is spam or not, image recognition like when a photo is uploaded with friends in Facebook it identifies the and tags the name, speech recognition for identifying the voice, traffic prediction which helps to tell whether there is traffic or not, product recommendations which helps in ecommerce, self-driving cars ,virtual personal assistant like Alexa, online fraud detection, automatic language detection and in many other fields.

IV. CONCLUSION

This paper is presented with a literature review of ML methods for detection of heart disease and the types used. Heart disease is the most common problems which occurs suddenly in the whole world. These problems cannot be seen with a naked eye and comes instantly when its limitations are reached. Hence many algorithms have been used but some are costly and time-consuming hence Machine Learning Algorithm can be used at correct time to predict the heart disease early. The analytical process started from data cleaning and processing, and the missing value, and exploratory analysis and finally model building and till evaluation. Finally, the prediction of the heart disease using the machine learning algorithm with different results will bring some of the following insights about heart disease prediction. As maximum types of dataset will be covered in this system, doctor will get to know about the disease exactly using ML algorithms, it helps the doctor in decision making weather patient will get heart disease or not. The machine algorithm can be used in many areas of the applications.

V. REFERENCES

[1]. Y. Alp Aslandogan, Gauri A. Mahajani :"Evidence Combination in Medical Data Mining".

[2]. Zhao, W., Wang, C. and Nakahira, Y. (2011) Medical Application on Internet of Things. IET International Conference on Communication Technology and Application (ICCTA 2011), Beijing, 14-16 October 2011, 660-665.

[3]. William Grant Hatcher; Weixian Liao; Weichao Gao; Wei Yu," Machine Learning for Security and Internet of Things: The Good, the Bad, and the Ugly". In IEEE access v0l 7,2019.o

[4]. Kiyong Noh,Heon Gyu Lee,Ho-Sun Shon,Bum Ju Lee,Keun Ho Ryu"Associative Classification Approach for Diagnosing Cardiovascular Disease".(volume 345)

[5]. Nabil Alshurafa, Costas Sideris, Mohammad Pourhomayoun, Haik Kalantarian, Majid Sarrafzadeh "Remote Health Monitoring Out-come Success Prediction using Baseline and First Month Interven-tion Data" in IEEE Journal of Biomedical and Health Informatics

[6]. Soni, J., Ansari, U. and Sharma, D. (2011) Intelligent and Effective Heart Disease Prediction System Using Weighted Associative Classifiers. International Journal on Computer Science and Engineering (IJCSE), 3, 2385-2392.

[7]. Boleslaw Szymanski,arsten Sternickel,alexander Ross,lijuan Zhuu: "using Efficient Supernova Kernel for Heart Disease Diagnosis".

[8]. S. Tsumoto, "PROBLEMS WITH MINING MEDICAL DATA" Department of Medicine Informatics, School of Medicine, Shimane Medical University, Izumo, Shimane, Japan.IEEE Oct. 20000

[9]. Deeanna Kelley "Heart Disease: Causes, Prevention, and Current Research" in JCCC Honors Journal.

[10]. Girmaw Abebe Tadesse; Tingting Zhu; Nhan Le Nguyen Thanh; Nguyen Thanh Hung;"Severity detection tool for patients with infectious disease"IEEE 2020

[11]. Shouman, M., Turner, T. and Stocker, R. (2012) Using Data Mining Techniques in Heart Disease Diagnosis and Treatment. Electronics, Communications, and Computers, Alexandria, 173-177

[12]. Gamal G. N. GeweId; Mahmoud A. Abdallah"A New Automatic Identification Method of Heart Failure Using Improved Support Vector Machine Based on Duality Optimization Technique"IEEE2019

[13]. Ponrathi Athilingam, Bradlee Jenkins, Marcia Johansson, Miguel Labrador "A Mobile Health Intervention to Improve Self-Care in Patients With Heart Failure: Pilot Randomizd Control Trial" in JMIR Cardio 2017, vol. 1, issue 2,

[14]. Applying k-Nearest Neighbour in Diagnosing Heart Disease Pa-tients Mai Shouman, Tim Turner, and Rob Stocker International Journal of Information and Education Technology, Vol. 2, No. 3, June 2012

[15]. Amudhavel, J., Padmapriya, S., Nandhini, R., Kavipriya, G., Dha-vachelvan, P., Venkatachalapathy, V.S.K., "Recursive ant colony optimization routing in wireless mesh network", (2016) Advances in Intelligent Systems and Computing, 381, pp. 341-351.

[16]. Alapatt, B.P., Kavitha, A., Amudhavel, J., "A novel encryption al-gorithm for end to end secured fiber optic communication", (2017) International Journal of Pure and Applied Mathematics, 117 (19 Special Issue), pp. 269-275.

[17]. Amudhavel, J., Inbavalli, P., Bhuvaneswari, B., Anandaraj, B.,Vengattaraman, T., Premkumar, K., "An effective analysis on har-mony search optimization approaches", (2015) International

Journal of Applied Engineering Research, 10 (3), pp. 2035-2038.

[18]. Amudhavel, J., Kathavate, P., Reddy, L.S.S., Bhuvaneswari Aadharshini, A., "Assessment on authentication mechanisms in dis-tributed system: A case study", (2017) Journal of Advanced Re-search in Dynamical and Control Systems, 9 (Special Issue 12), pp. 1437-1448.

[19]. Amudhavel, J., Kodeeshwari, C., Premkumar, K., Jaiganesh, S., Rajaguru, D., Vengattatraman, T., Haripriya, R., "Comprehensive analysis on information dissemination protocols in vehicular ad hoc networks", (2015) International Journal of Applied Engineering Re-search, 10 (3), pp. 2058-2061.

[20]. Amudhavel, J., Kathavate, P., Reddy, L.S.S., Satyanarayana, K.V.V., "Effects, challenges, opportunities and analysis on security based cloud resource virtualization", (2017) Journal of Advanced Research in Dynamical and Control Systems, 9 (Special Issue 12), pp. 1458-1463.

[21]. Naïve Bayes * Shadab Adam Pattekari and Asma Parveen Department of Computer Science and Engineering Khaja Banda Nawaz College of Engineering .

The International Conference on Research Perspectives : IoT in Hybrid Grid Integrated Renewable Energy Sources
In association with International Journal of Scientific Research in Computer Science, Engineering and Information Technology
ISSN : 2456-3307 (www.ijsrcseit.com)

IoT Based Component for Smart Museum (I-Smart)

Santhosh Krishna B V[1], Sanjeev Sharma[2]
Senior Assistant Professor[1], Professor & Head[2]
Department of Electronics and Communication Engineering, New Horizon College of Engineering, Bengaluru, Karnataka, India

ABSTRACT

Today's world is becoming smarter due to the development of various technologies. To Communicate in this smart world, our environment needs to become smarter because environment speaks to people and people speak to the world. One such smart technology is IOT (Internet of Things). Our paper's ultimate aim is to enhance the user experience in a museum which relies on a wearable device that acts as a guide in the museum. This wearable device captures the image of an artifact or artwork and compare the captured image with the images in the cloud using image processing algorithms. An audio and textual description of the artifact's multimedia contents stored in the cloud is delivered to the user such that everyone can easily access the cultural profiles through this smart device without manual intervention.

Keywords : IOT, Wearable Device, Artifact, Museum

I. INTRODUCTION

Once Art and culture was the part and parcel of human's life. Museum act as sources of objects and associated information (sometimes known as "metadata" for virtual objects in the field of information technology) which are worth saving for future generations.Over the years, hundreds of museums and art galleries have preserved our diverse cultural custom and served as important sources of education and learning. A substantial number have established a Web presence for several years and are now considering how this should be developed for the future.Today, museums and galleries usually provide visitors either with booklets or with audio guides. Visits at museums are often considered boring, because it is hard for museums curators to catch the attention of tourists.Therefore, a smart museum needs to be created for unplanned and customized historical center visits which can be made possible by the smart technology called IOT.

The paradigm of the Internet of Things has become a core element of the Internet. Sensing with low cost and actuation is available to everyone. It allows unified information exchange and networked interactions of physical and digital objects. IoT expects to make a superior world for individuals, where smart articles around us understand what we like, what we need and act in like manner without explicit motions .Considering all these, the main objective of this paper is to realize a real smart environment, in which cultural things are able to speak to the visitors. Once an artwork is detected ,the multimedia story is delivered to the user by means of proper multimedia delivery in order to stimulate the

visits. When a user is in front of an artwork or artifact, several details such as name, historical context, critical review can be easily and automatically provided without the necessity of personal guides. All these contents can also be shared in the social networks by the application of cloud services.

The integration of image recognition capabilities with a specified digital scenario allows achieving the desired goals. Furthermore, the possibility to share the cultural experience can be a strong driving factor to approach young people to the cultural world.

II. RELATED WORKS

There are several works related to our system. One of the key features of the proposal is represented by image matching mechanism, is an important research topic. There are several works based on the localization mechanism which seems to be complex. In [1] the author proposed a design that can(i) provide cultural contents to the visitors, (ii) communicate useful information to external users, (iii) interact with diverse technologies that control the status of the environment. Another example of location-aware services in a smart environment is reported in [2]The museum wearable, a storytelling device: it is a museum guide which observes his/her path and length of stops and based on the location, selects content from a database of available movie clips and audio. The localization is based on an infrared positioning system; and hence allows the system to roughly estimate the position and therefore the content delivery may not be correct. Works [3,4] focuses on RFID technologies. In these System, the authors exploits the interaction between an RFID reader integrated on users' mobile devices and RFID tags placed near each artworks to provide users with cultural contents . These solutions require the use of mobile devices equipped with RFID readers, which are expensive and not so common. Another example is the system "Smart Museum" [5], in which visitors can gather information about what the museum displays and customize their visit based on specific interests. This system, that integrates PDAs and RFIDs, brought freshness when first released, but it has some limiting flaws. many authors use local descriptors to support detection and recognition of object in real world images and video. One such example is, In [6] the paper presents a method in which the edges of the given images are matched with the new entries using different algorithms and it determines its validity with the available pictures. This has a drawback that if image contains less edges, comparison becomes quite difficult. In fact, local visual descriptors like Maximally Stable Extremal Regions [7], Scale Invariant Feature Transform [8], Speeded Up Robust Features [9], have been proven to be able to capture sufficiently discriminative local elements with some invariant properties to statistical or photometric transformations and are vigorous to occlusions. However, they suffer of high blur, a typical characteristic of video acquired by a wearable camera, and in this scenario achieve a low recognition accuracy. In [10,11] the author proposed a better algorithm ORB which is faster than SURF and SIFT. The author in this paper [12] discuss a general reference framework for the design of an urban IOT by using the technical solutions and best-practice strategy adopted in the Padova Smart City project, a proof-of-concept deployment of an IoT island in the city of Padova, Italy, performed in collaboration with the city municipality.

III. SYSTEM DESCRIPTION

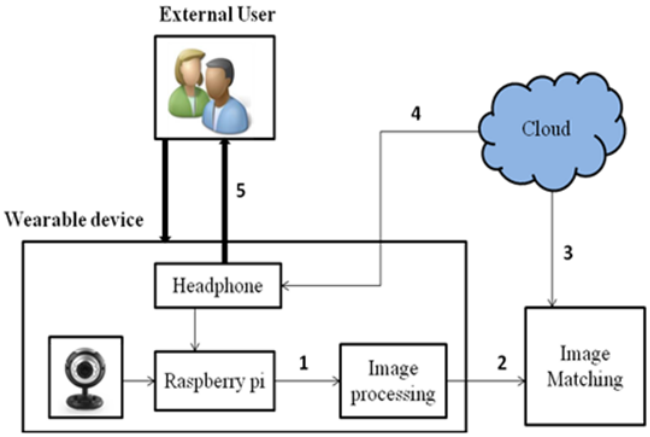


Fig 3.1. Architecture diagram

3.1.WEARABLE DEVICE

3.1.1.CAMERA

Webcam is a digital camera which can send live pictures from wherever it is positioned to another location by means of the internet and is also connected to a computer..A webcam is generally connected by a USB cable. Some webcams are connected to computers through USB ports, but others are wireless (wifi).

Fig 3.2 Camera

3.1.2. RASPBERRY PI

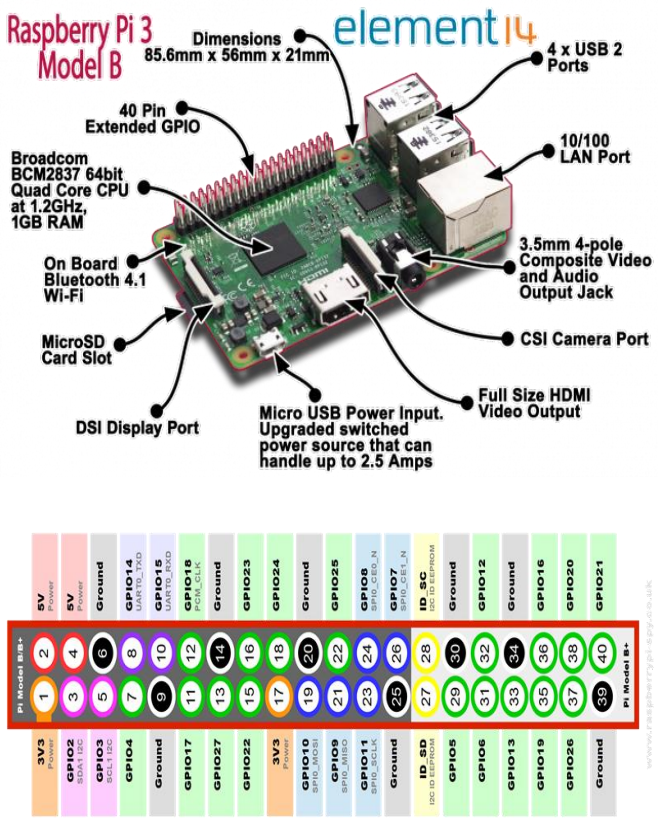


Fig 3.3 Raspberry pi 3 model B

The Raspberry Pi is a single-board computer in the size of a credit card with an open-source platform that has a promising community of its own, similar to that of the audio. There are a few varieties of the Raspberry Pi, but the latest version, has improved upon its antecedent in terms of both form and functionality. **Raspberry Pi 3 Model B**, a 5V USB power supply of 2 amps with a micro USB cable. We can also connect any standard USB keyboard and mouse, an HDMI cable and monitor for display, and a micro SD card. A black 3.5 mm audio jack located on the top edge of the Pi next to the yellow phone connector provides analogue audio. This is the same connector used for headphones and microphones on consumer audio equipment, and it's wired in exactly the same way. The CSI camera connector port is used to connect capturing element to the device. It is possible to add Wi-Fi support to any Pi using a USB wireless adapter.

3.1.3.PIR SENSOR

Passive Infrareds sensor (PIRs) is an electronic device which is used in security alarm systems to detect motion of an IR emitter, usually a human body. When the amount of radiation falling on the crystal changes, the amount of charge also changes and can then be measured with a FET device which is sensitive. The PIR325 sensor has two sensing elements connected in a voltage opposing configuration. This arrangement cancels signals caused by vibration, heat changes and light. A body passing in front of the sensor will activate first one and then the other element whereas other sources will affect both elements simultaneously and be cancelled.

3.1.4 IMAGE PROCESSING

ORB Algorithm

Oriented FAST and rotated BRIEF (ORB) is a binary algorithm proposed by Ethan Rublee[11]. ORB algorithm is based on FAST and BRIEF algorithm. It is a method to describe feature points by using the binary string. the feature point of ORB is detected by the FAST feature detection, and is described using an improved BRIEF feature descriptor. the speed of FAST and BRIEF are very fast, hence ORB has an absolute advantage in speed. The greatest feature of this algorithm is fast and having rotational invariance and reducing sensitivity to noise. Its aim is to provide a fast and efficient alternative to SIFT.

a.FAST

FAST (Features from Accelerated Segment Test) is a algorithm proposed by Edward Rosten and Tom Drummond in their paper[13]. FAST corner detector uses a circle of 16 pixels to classify whether a point p is actually a corner. Each pixel in the circle is labelled from number 1 to 16 clockwise. If a set of N adjoining pixels in the circle are all brighter than the intensity of pixel p (denoted by I_p) plus a threshold value t or all darker than the intensity of candidate pixel p minus threshold value t, then p is classified as corner.

Algorithm :

1. Select a pixel p in the image which is to be tested. Let its intensity be I_p.
2. Select threshold value t.3.
3. Consider a circle of 16 pixels around the pixel p.(picture taken from[13])
4. Now the pixel p is a corner if there exists a set of n adjoining pixels in the circle (of 16 pixels) which are all brighter than I_p+t, or all darker than I_p-t . n was chosen to be 12.

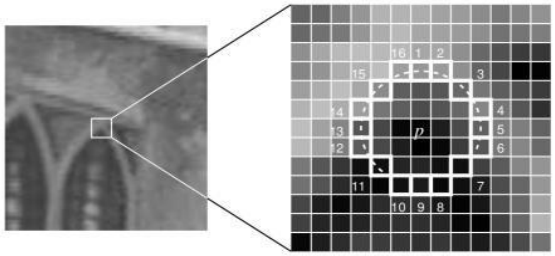


Fig.3.4 Corner Detection

b.BRIEF

We know SIFT and SURF takes a lot of memory which are not feasible for resource-constraint applications especially for embedded systems. Larger the memory, longer the time it takes for matching. BRIEF comes into picture at this moment. It provides a shortcut to find the binary strings directly without finding descriptors. It takes smoothened image patch and selects a set of nd (x,y) location pairs .Then some pixel intensity comparisons are done on these location pairs. For eg, let first location pairs be p and q. If I(p)<I(q), then its result is 1, else it is 0. This is applied for all the nd location pairs to get a nd-dimensional bit string.

3.2. SOFTWARE

Image processing algorithm runs on the wearable device and it is able to detect, in real-time, the artwork the user is observing. This algorithm is developed using Open CV (Open Source Computer Vision Library) which is an open source computer vision and machine learning software library. Being a BSD-licensed product, Open CV makes it easy to utilize and modify the code. The library contains more than 2500 enhanced algorithms, which includes a complete set of both classic and state-of-the-art computer vision and machine learning systems. These algorithms can be used to detect and recognize faces, identify objects, find similar images from an image database ,classify human actions in videos, track camera movements, track moving objects etc., It has C++, C, Python, Java and MATLAB interfaces and supports Windows, Linux, and Mac OS.

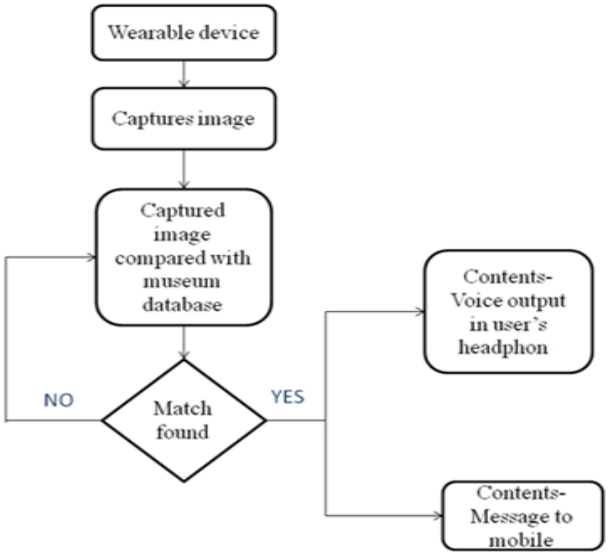
3.3 .CLOUD SERVER

Storing, organizing, and retrieving information and multimedia contents in a large amount is an expensive process from both the computational and memory point of view. For this reason, the Cloud seems to be the solution that suits this kind of requirements, as its storing and computing abilities allow to process data more resourcefully. In particular, in the proposed system, the Cloud accomplishes several tasks.

3.4 OPERATION

When the user wearing the device is in front of the article ,the PIR sensor senses the human body and changes the amount of radiation. This initiates the camera present in the device and triggers it to capture the image. The captured image is processed by the suitable algorithms[1] and compared with the images stored in the cloud[2,3] .If the match is found, then the multimedia contents about the article is accessed from the cloud and given as input to wearable device[4].These contents are then delivered to the user via headphones connected to the device.

3.5 FLOWCHART



IV. RESULTS AND DISCUSSION

In this section our proposed system has been tested with database images and the results were observed as follows.

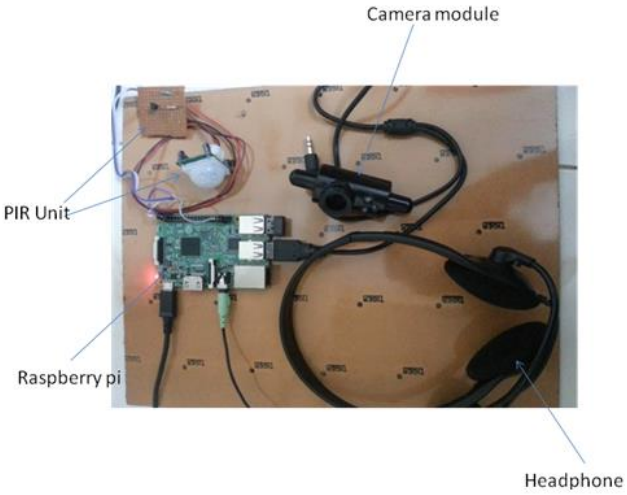


Fig 4.1 Experimental setup

The experimental arrangement is shown in the fig 4.1.The major componenents of the system are the camera and a headphone which can be implemented as a wearable device. A PIR sensor is connected to the Raspberry pi which can sense the human motion and can initiate the camera.

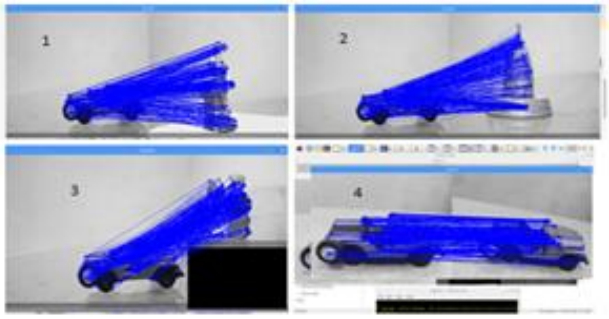


Fig 4.2 Image comparison

Here the captured image is compared with database images. In this images 1, 2,3 represents the unmatched condition. Image 4 represents matched condition which is shown clearly in fig 4.3.

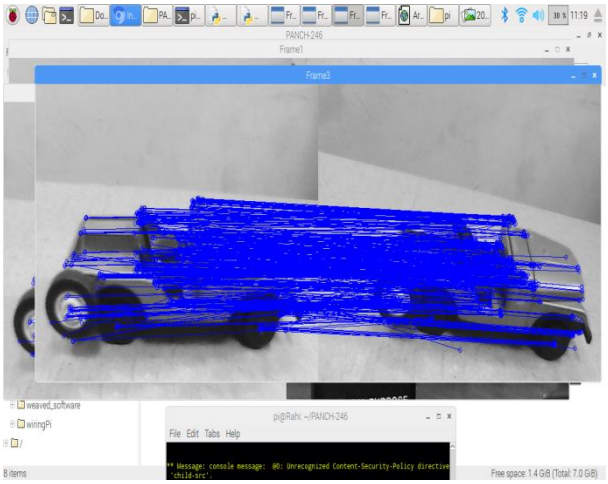


Fig 4.3 Matched image frame

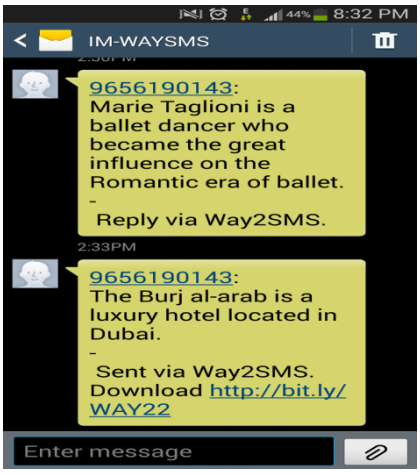


Fig 4.4 SMS sent to mobile

After image comparison and matching process the description about the artifact stored in the database is sent to the mobile as a SMS (way2SMS) as shown in the figure 4.4 .As said above the voice output is delivered using the headphone

Fig 4.2 Future view of our wearable device

V. CONCLUSION AND FUTURE WORK

In this paper we have presented the use of IOT technology, in the smart museum guide system. It describes how collaboration between museum staff and university researchers can lead to new and innovative uses of technology in the museum and how these technologies can support a variety of activities in addition to its traditional role in regular museum visitors. The proposed system relies on a wearable device equipped with image matching capability to automatically provide users with cultural contents related to the observed artwork. Even if the proposed system would suffer from serious maintenance issues caused by hardware damages due to vigorous use, it does not require the use of expensive devices, as well as it does not force the users to exploit their own devices. In future, the wearable device will be powered by the phone which clips into the front and the screen and shows specially written software to put on the show.

VI. REFERENCES

[1]. S.Alletto, R.Cucchiara, G.Del Fiore, L.Mainetti, V. Mighali , L. Patrono, and G. Serra, “An Indoor Location-aware System for an IOT based Smart Museum”, IEEE Internet of things journal,2015.

[2]. F.Sparacino, “The museum wearable: real-time sensor-driven understanding of visitors’ interests for personalized visually-augmented museum experiences,” in Museums and the Web, 2002.

[3]. Y.Wang, C.Yang, S.Liu, R.Wang, X.Meng, “A RFID & handheld device-based museum guide system", 2nd Int. Conference on Pervasive Computing and Applications (ICPCA07), pp. 308 - 313, July 2007.

[4]. L.Caviglione, M.Coccoli, A.Grosso, "A framework for the delivery of contents in RFID-driven smart environments," in 2011 IEEE International Conference on RFID-Technologies and Applications.

[5]. A.Kuusik, S. Roche, F. Weis et al., 2009. Smart museum: Cultural content recommendation system for mobile users. In Proc. Fourth Int. Conference on Computer Sciences and Convergence Information Technology.

[6]. SomiR.Kevin,”Image matching with edge detection”, published in International journal of electronics communication and computer engineering,2012.

[7]. J.Matas,O.Chum, M. Urban, and T. Pajdla, “Robust wide-baseline stereo from maximally stable extremal regions,” Image and vision computing”, 2004.

[8]. D.G.Lowe,"Object recognition from local scale-invariant features," in Proc. of the seventh IEEE international conference on computer vision. Kerkyra, 1999,

[9]. H.Bay, T.Tuytelaars, and L.V.Gool. "Surf: Speeded up robust features," in Proc. of European Conference on Computer vision, Springer, Berlin, Heidelberg, 2006.

[10]. Ethan Rublee, Vincent Rabaud,Kurt Konolige Gary Bradski,” ORB: an efficient alternative to SIFT or SURF” in IEEE International Conference on Computer Vision, 2001

[11]. Lei Yu, zhixin, Yan Gong,”An Improved ORB Algorithm of Extracting and Matching Features”, International Journal of Signal Processing, Image Processing and Pattern Recognition,2015.

[12]. A. Zanella, N. Bui, A. P. Castellani, L. Vangelista, M. Zorzi, "Internet of things for smart cities", IEEE Internet of Things Journal, 2014.

[13]. Edward Rosten and Tom Drummond,” Machine learning for high-speed corner detection”,2006.

[14]. W. Zhao, "Internet of things - what and how", IEEE Real-Time Systems Symposium, 2012.

[15]. Majed Rostamian, Maryam Parsa and VoicuGroza “Design and fabrication of a smart electronic guide for museums” in IEEE Transaction on Applied Computational Intelligence and Informatics (SACI), 30 July 2012.

[16]. Angelo Chianese, Francesco Picciali,”Designing a Smart Museum: When Cultural Heritage Joins IOT”, IEEE published in (NGMAST), 2014 Eighth International Conference.

IJSRCSEIT

www.ijsrcseit.com

Sponsored By

International Journal of Scientific Research in
Computer Science, Engineering and Information Technology
(International Journal Bimonthly Publication)
www.ijsrcseit.com

Published by :
TechnoScience Academy

www.technoscienceacademy.com

The International Conference on Research Perspectives : IoT in Hybrid Grid Integrated Renewable Energy Sources

NEW HORIZON
COLLEGE OF ENGINEERING

Organised by
Department of Electrical and Electronics Engineering,
New Horizon College of Engineering,
Bellandur Main Rd, Kaverappa Layout,
Bengaluru, Karnataka, India

Email: editor@ijsrcseit.com, ijsrcseit@gmail.com