RESEARCH REVIEW JOURNALS
Volume-04
Issue-04
April-2019

ISSN: 2455-3085 (Online)
RESEARCH REVIEW International Journal of Multidisciplinary
www.rrjournals.com[UGC Listed Journal]

Digital Identity Creation and E-Commerce for FPO's

[1]Prof. Anand Bhosle,[2]Hardik Bhanderi,[3]Saloni Agrawal, [4]Monika Dhamale
[1]*Professor, Department of Information Technology, International Institute of Information Technology, Pune (India)*
[2,3,4]*Student, Department of Information Technology, International Institute of Information Technology, Pune (India)*

ARTICLE DETAILS

Article History
Published Online: 15April2019

Keywords
FPO, E-Commerce, Online-shopping

***Corresponding Author**
Email:bhanderi77[at]gmail.com

ABSTRACT

We all know that some of the farmers suffer a lot in terms of money as they don't have proper land or get unexpected profits from their corresponding land holders. In this project we are trying to bridge the gap between farmers and their expected profits. Internet plays an important role in our daily life. We use internet daily almost for every single work. Before e-commerce buying and selling were done without internet physically in the markets but after the arrival of e-commerce in India our life has become more convenient because of its number of advantages. Online shopping is a part of e-commerce which is done mostly by the users due to e-commerce websites in India which allows us to buy and sell the products according to our choice at affordable price. E-commerce website has a lot of impacts on different markets and retailers. In this project we are introducing e-commerce in terms of seeds and crops which will be beneficial for customers and farmers both as well as for investors who are willing to sell their goods.

The other thing our project focuses on is creating farmer's unique identity that is his digital identity which will become easy for investors to find the respective farmers and hire them for work.

1. Introduction

Our project basically focuses on is every farmer should get opportunity to earn and grow goods. The question arises is how? We know that there are some investors who are ready to invest money on farmers for growing goods on their land. Our main aim is to connect such needy farmers and willing investors together. For that we will make a web application in which each farmer will have their own digital identities in which all records of farmer including personal details, past activities, work records, etc. will be stored. A contract PDF will get automatically generated by our application if both investor and farmers agree. In spite of this we will make a hybrid application which means a website and app both so that user can use it according to their convenience. In this project we are using E-Commerce for selling farmer's crop online. Customer can directly buy goods from farmer using our website/mobile application. Customer can order beforehand specifying the requirements.

1.1 Proposed System

We all know that in today's era of internet, e-commerce is growing by leaps and bounds keeping the growth of internet very rapidly. In many cases, e-commerce businesses are resorting to having a counterpart which is internet or e-commerce driven. People in the developed world and a growing number of people in the developing world now use ecommerce websites on a daily basis to make their everyday purchases. People now a day's prefer buying items online hence saving their time of shopping and availing great cash back offers online. But no such facilities for selling crops and seeds are available online. Moreover no such facility is available where farmers get their digital identity online. We are trying to overcome that which will lead to increase in farmer's income and will help farmer's to sell their goods to online.

1.2 Motivation

Now a days, number of farmers are doing suicide because they don't get enough income according to their work. So that we decided to do this project which will help in increase their income and make them happy. Other than this online marketing is very common concept now a days which motivated us to bring that concept in seeds and crops which will make farmers and customers earn more profits on goods happily.

1.3 Problem Definition

To develop an android application which will help farmers to earn more profit and have an online contract with investors about the money and time too. Information about farmers will be collected from various FPOs based on the location. And then Digital Identity will be created of that farmers.

Likewise investors and farmers both will be able to sell the crops which will be dealing with E-commerce concept.

2. The Planning Process

Our goal was to develop a web application that would be attractive enough, have a professional look and user friendly. So that people of all age groups would be its end users. Our job started with subdividing the entire task and setting milestones. The milestones would be a marker of percentage of the work actually accomplished and success story. The entire planning process took the following steps.

2.1. Defining Use Case Model

Investors can search available farmers registered through FPOs online for making work on his land. A contract PDF will be generated between farmers and investors for the duration of time they are working together through our application. The PDF includes time periods for which they are working for investor and the profit they will make for the specified duration.

Farmers and investors both can sell their goods online through the e-commerce application developed. Customer can even pre-order us their requirements if not sufficient quantity of what he requires is available. Customers Adds items into the shopping cart and finally orders the products online when the electronic copy of the bill is automatically generated. So, from the stated use case model we found out the following to be the primary requirements:

- A registration page
- Search option
- Shopping cart
- Billing system
- Contract PDF
- Pre-order of goods

2.2. Domain Modeling

As with most of the web applications developed using the Object Oriented Programming (OOP) we followed the same. So we moved forward for Object Oriented (OO) analysis. We are developing hybrid application which follows OO analysis using Ionic Framework.

2.3. Architectural Pattern

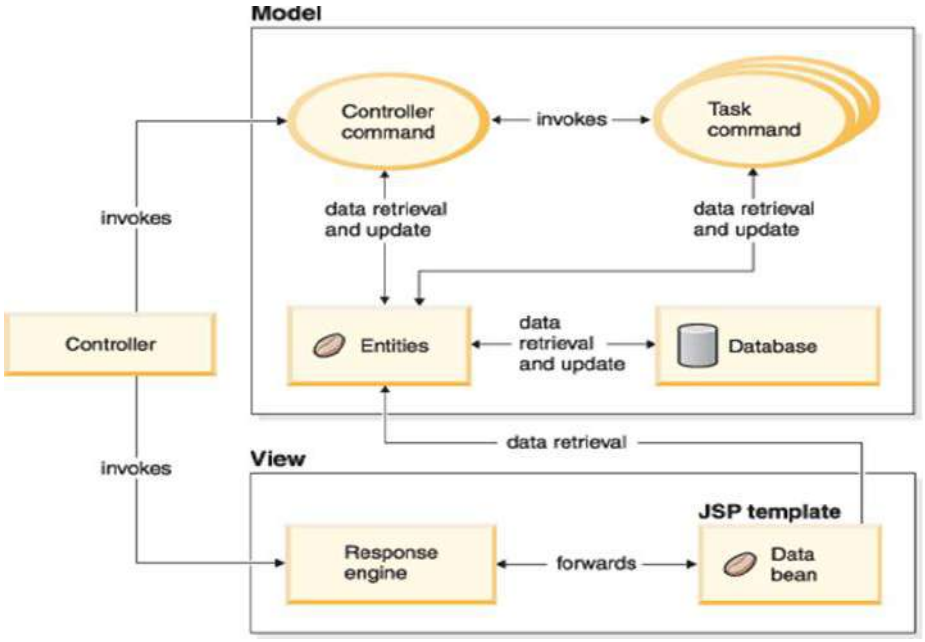


Fig 1: MVC model

Our application has been developed using the standard "Model-View-Controller" pattern. The model-view-controller (MVC) design pattern specifies that an application consist of a data model, presentation information, and control information. The pattern requires that each of these be separated into different objects.

The model (for example, the data information) contains only the pure application data; it contains no logic describing how to present the data to a user. The *view* (for example, the presentation information) presents the model's data to the user. The view knows how to access the model's data, but it does not know what this data means or what the user can do to manipulate it.Finally, the controller (for example, the control information) exists between the view and the model. It listens to events triggered by the view (or another external source) and executes the appropriate reaction to these events. In most cases, the reaction is to call a method on the model. Since the view and the model are connected through a notification mechanism, the result of this action is then automatically reflected in the view.

3. Development tools

The entire development process has been subdivided into two: the front end development and the backend development. The front end comprises of the visually visible parts such as the home page, admin panel, shopping cart page, register FPO page, generate PDF page. The back end contains the database and its interaction with the front-end.

3.1. Front End Development

The front end was initially raw coded using Type Script. Type Script is a client side scripting language which is a dedicated language for web development. Type Script code was simply mixed with the Hypertext Mark-up Language (HTML) code. Hypertext mark-up language is the language used to design the web pages of an application. A static page is an HTML document that is stored on the web server and does not change. This was performed by Cascading Style Sheet (CSS). CSS is a style sheet language used for describing the look and formatting a document written in a mark-up language. These CSS files are linked with the class files to put the panels in order, the text with correct font, size and color. We introduced Type Script in our application. Type Script is a client side scripting language most commonly used as part of web browsers and its implementations allow client side scripts to interact with the user, control the browser, communicate asynchronously and alter the document content that is displayed.

For example, in our website the clients while registering are asked to provide their specifications which contain their name, email address, age, etc. If they miss any of the criteria then immediately the browser asks him for filling the particular field. This is implemented by a Type Script.

3.2. Back End Development

In our application we have chosen the Cloud Fire store to hold the database. It is a cloud database management system. The main reason of using this is it can scale effortlessly as your app grows. Cloud Fire store allows you to run sophisticated queries against your data. This gives you more flexibility in the way you structure your data and can often mean that you have to do less filtering on the client, which keeps your network calls and the data usage more efficient.

3.3. Database Design

One of the most important and challenging task is the database design. The information passed by the customer while registering and registered FPOs with all their details are stored in the database. The products with their identification, description and image are stored in the database. Moreover, if we update any of the featured products the update takes place in the database. So the program has a lot to do with the database.

In our database there are different columns such as add FPO, add farmers, add items, product details, sell details, PDF generated, etc. Each table has its one or more attributes which included its detailed information like name, id, address, etc. Accessing all the queries becomes easy through cloud database. Other than this more tables are added in the database according to the requirement.

4. Coding Process

Our application has been developed using Model-View-Controller format as standard. We have separate files containing several functions to implement the proposed functionalities. View is a separate folder holding all the required files with .ts extension containing HTML tags and codes that creates the actual WebPages. These files have links to other files in the file holder entitled “Controller” which has necessary functions to correspond to the database.

Code for adding FPOs in the database is

```
createFpo(fpo: FpoDetails){
letfpoID = this.afs.createId();
console.log(fpo);
letdocRef: DocumentReference = this.afs.doc(`/FpoList/${fpoID}`).ref;
console.log(docRef);
```

Code for adding farmers in their specified FPOs is

```
createFarmer(farmer: FarmerDetails, fpoID: string){
letfarmerID = this.afs.createId();
console.log(farmer);
letdocRef: DocumentReference = this.afs.doc(`/FpoList/${fpoID}/FarmerList/${farmerID}`).ref;
console.log(farmer);
```

Code for adding items to sell is

```
createItem(item: ItemDetails){
letitemID = this.afs.createId();
console.log(item);
letdocRef: DocumentReference = this.afs.doc(`/ItemList/${itemID}`).ref;
console.log(docRef);
```

Code for accepting orders is

```
asyncaddOrder(salesOrder : SalesOrder){
letorderID = this.afs.createId();
console.log(salesOrder);
letdocRef: DocumentReference = this.afs.doc(`/SalesOrderList/${orderID}`).ref;
console.log(docRef);
```

5. Project Architecture and Design

5.1. Architecture

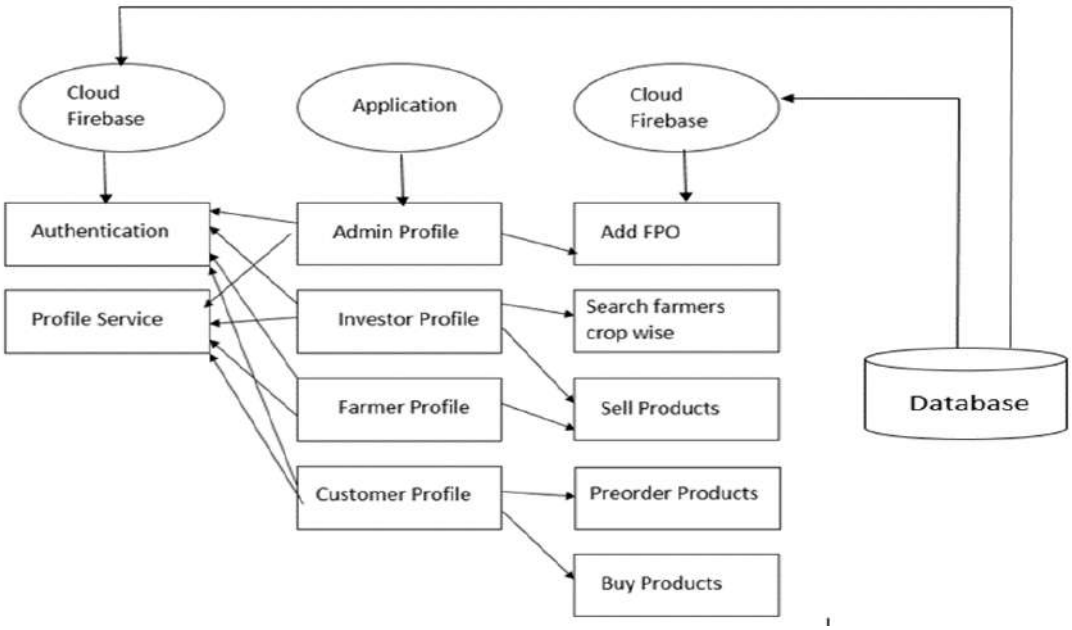


Fig 2: Architecture Diagram

1. FPO contains list of farmers. Admin registers the FPO.
2. FPO add farmers.
3. Investors can access the list of farmers.
4. Investors can also sell the crops.
5. Customers can purchase crops from Investors as well as Farmers.

5.2 Design

The design of project consists 4 modules:

A. Admin Module:

In this module admin can add FPO's and add farmers in the FPO's.

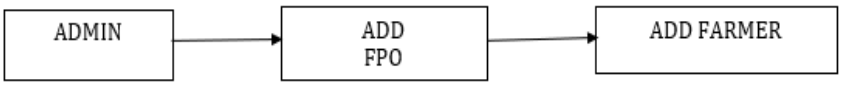


Fig 3: Admin Module

B. Investor Module:

In this module investor who wants to invest can search FPO's based on interested crops to grown and hire farmers willing to work in their land. Investor can sell crops grown by farmers to the customer.

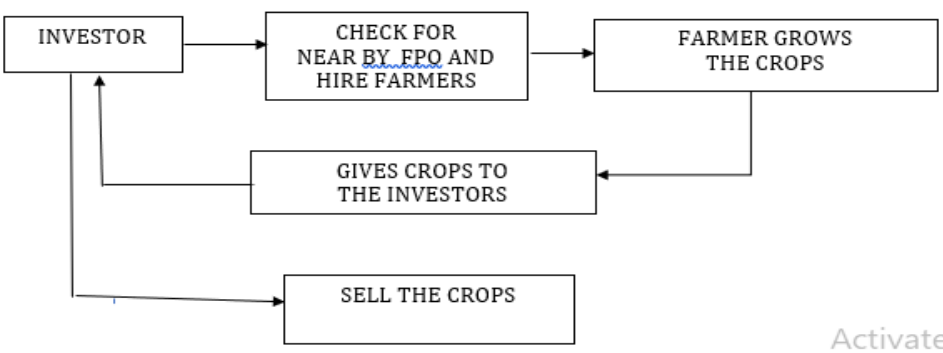


Fig 4: Investor Module

C. Farmer Module:

In this module if farmer have their own land they can grow their own crops or they can work for investors also they can sell their crops.

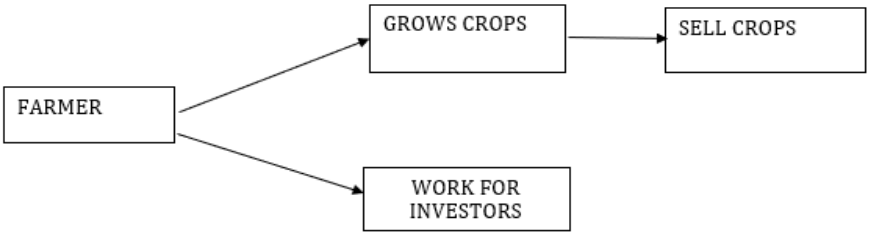


Fig 5: Farmer Module

D. Customer Module:

In this module customer can purchase crops from investors.

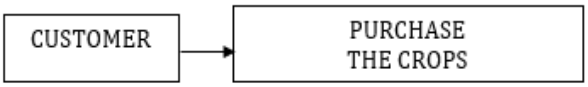


Fig 6: Customer Module

6. Conclusion

This project is design in order to increase the small farmer's income by connecting them with investors who are ready to invest.Due to this project buying and selling of crops became easy and possible which will give double profit to the farmers.This is a good step toward cashless transaction and to avoid the third party involvement which is unnecessarily increases the cost of product. The cashless transaction avoids the utilization of black money and makes system transparent in terms of taxpaying. There are no time barriers in selling the products. One can log on the internet even at midnight and can sell products at a single click of mouse. An interactive user friendly and focused Web App in the form of online shop can generate good business.

Acknowledgements

Working on the project "Digital Identity creation and E-commerce for FPO's" is the great experience. It was a source of immense knowledge. We would like to express my sincere gratitude to project guide Prof. AnandBhosle for his guidance and support throughout the semester. We acknowledge his encouragement and moral support. His guidance was very helpful in different matters regarding the project.

We would also like to express my gratitude to Project Co-coordinator Prof. Vishal Chaudhari for a support and the guidance for various issues regarding project. He was very kind and patient while explaining us the outlines of project. We thank him for his overall support. We would also like to thank all teachers and my fellow friends who all supported us.

References

1. Ullah, S. E., Alauddin, T., &Zaman, H. U. (2016). Developing an E-commerce website. 2016 International Conference on Microelectronics, Computing and Communications (MicroCom). https://ieeexplore.ieee.org/document/7522526
2. AnandUpadhyay, Ambrish Pathak, Nirbhay Singh, International Journal of Commerce and Management Research ISSN: 2455-1627, Impact Factor: RJIF 5.22 June 2017 http://www.managejournal.com/download/510/3-6-18-450.pdf
3. Camp, L. J. (2004). Digital identity. IEEE Technology and Society Magazine, 23(3), 34–41. doi:10.1109/mtas.2004.1337889 https://ieeexplore.ieee.org/document/1337889
4. Feng Liu, Wenjuan Tang, Yi Zhang, & Makoto, H. (2011). Construction of the agricultural products e-commerce mode linked by rural economic cooperation organization — through two Japanese cases study. 2011 International Conference on Business Management and Electronic Information. https://ieeexplore.ieee.org/document/5916898
5. http://www.sourcetrace.com/digital-technologies-fpos-achieved-together-case-study-india/
6. Anand A. Bhosle, Secure Data Distribution using Secret Splitting Over Cloud, Global Journal of Computer Science and Technology, Online ISSN: 0975-4172 & Print ISSN: 0975-4350 ,Volume 17 Issue 1 Version 1.0 Year 2017
7. AnandBhosle, Efficient Estimation of Complex Queries, International Journal of Advanced Research in Computer Science and Software Engineering, ISSN: 2277 128X, Volume 6, Issue 4, April 2016.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 7 Issue IV, Apr 2019- Available at www.ijraset.com

Plant Species Classification and Disease Detection using Convolutional Neural Network

Himanshu Joshi[1], Reeya Saxena[2], Snehal Bhise[3], Diksha Sawant[4], Prof.Anand Bhosale[5]
[1, 2, 3, 4, 5]*Information Technology, Savitribai Phule Pune University*

Abstract: ***This paper presents a survey on detection and classification of leaf spices. It is di□cult for human eyes to identify the exact type of leaf spices. Thus, in order to identify the leaf spices accurately, the use of image processing and machine learning techniques can be helpful. The images used for this work were acquired from the digital camera. In pre-processing step, background removal technique is applied on the image in order to remove background from the image. Then, the background removed images are further processed for image segmentation. Di□erent segmented images will be used for extracting the features such as color, shape and texture from the images. At last, these extracted features will be used as inputs of classifier. The goal of proposed work is to diagnose the spices using image processing of plant leaf. In the proposed system, leaf image with complex background is taken as input. Threshold is deployed to mask green pixels and image is processed to remove noise using an isotropic di□usion. Then leaf spices segmentation and then classification using classifier will be done.***
Keywords: 1. Preprocessing,2. Feature Extraction, 3. Classifier, 4. Threshold,5. Isotrofic-Difiusion.

I. INTRODUCTION

Plants are an integral part of ecosystem. Due to deforesting many plant species are under the risk of annihilation. Plants are useful for human being and other living things. They are useful as foodstu□, as medicine and also in many industries. Identifying plants helps ensure the protection and survival of all-natural life. Plant identification can be performed using many di□erent techniques using the planta's leaves. Leaves are useful to classify plants since they are more readily available than the other bio-metric components like flowers which are available for a short period. Plant classification by using leaves requires different bio-metric features of leaf like colour, shape, texture. This identification manually is time consuming and expensive. Leaves can be classified based on colour that include similarity between two images with the help of colour histogram, but the colour based classification is depend on seasonal e□ect of sunlight. So, by using classifier we can classify the Species of leaf.

A. Problem Statement

One of the important sectors of Indian Economy is Agriculture. Employment to almost 50 percent of the countries workforce is provided by Indian agriculture sector. India is known to be the world's largest producer of pulses, rice, wheat, spices and spice products. Farmers' economic growth depends on the quality of the products that they produce, which relies on the planta's growth and the yield they get. Therefore, in field of agriculture, detection of spices in plants plays an instrumental role. In order to detect a plant spices at very initial stage, use of automatic spices detection technique is advantageous. Manual detection of plant spices using leaf images is a tedious job. Hence, it is required to develop computational methods which will make the process of leaf species detection and classification using leaf images automatic.

B. Background Study

Below figure shows the fundamental steps of an automated plant classification system, that is nothing but a system architecture. Initially, the leaf images would be acquired using digital camera, scanner or some other equipment's. The images were then pre-processed to remove noise and improve the quality. Noise occurs as pixel values which do not represent the true intensities of an image during the image acquisition. Image enhancement is a process that is used to emphasise the features of an image. It is a necessary step to remove the image noises in order to highlight or enhance the important features of an image. Subsequently, the region of interest (ROI) was segmented from the images, followed by feature extraction[1]. Finally, the extracted features are fed to classification or recognition system. Leaves are commonly used in plant species recognition due to their availability throughout the year, especially, in the tropical areas. [1,2,5] Shape is the most common feature that have been used to develop plant identification systems. [1,2,5] Texture is one of the important features of the plant identification system, which can be used to characterize the leaves based on the surface structure of the leaves. [1,2,5]

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 7 Issue IV, Apr 2019- Available at www.ijraset.com

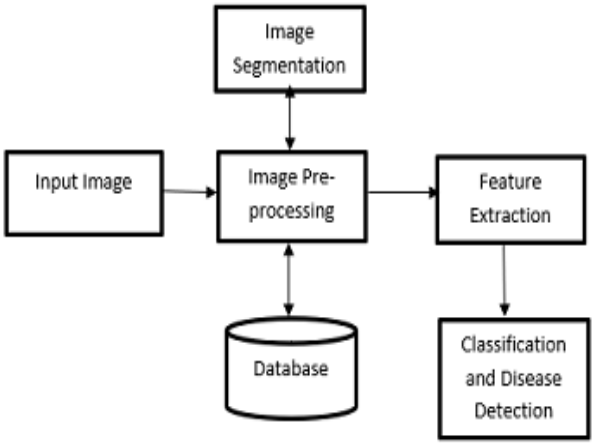


Fig -1: Proposed System Architecture [5]

II. PREPROCESSING

It is common practice to have the pre-processing of Cotton leaf images before it has been extracted and classified.
There are five main steps used for the detection of plant leaf diseases as shown in fig. The processing scheme consists of image acquisition through digital camera or web, image pre-processing includes image enhancement and image segmentation where the affected and useful area are segmented, feature extraction and classification. Finally, the presence of diseases on the plant leaf will be identified. [8]

In the initial step, RGB images of leaf samples were picked up. The step-by-step procedure as shown below:

1) RGB image acquisition;
2) Pre-processing of image using Histogram equalization;
3) Resize the image;
4) CNN for Image Segmentation;
5) Computing features extraction;
6) Classification & Recognition using neural networks.
7) Statistical analysis.

A. Pre-Processing of Leaf Image

The input image has to be pre-processed because images are corrupted by a type of multiplicative noise like light intensity and shadow on a cotton leaf images that may contain useful information about the leaf spot that can be used in the diagnosis. The pre-processing is done with the contrast enhancement using Histogram equalization.

B. Leaf Segmentation

The leaf spot in the capture image generally contains reflection from source, which forms some intense spot in the cotton leaf, but pixel value within the cotton leaf is over a particular threshold (20) then it is replaced by pixel value of some neighbourhood pixel. This operation fills all intense leaf spot present in cotton leaf area as shown in the figure below.

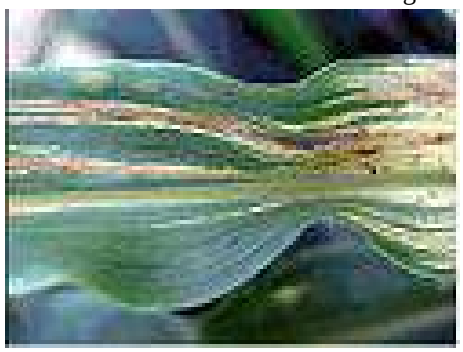


Fig -2 a) Captured Image

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 7 Issue IV, Apr 2019- Available at www.ijraset.com

Fig -2 b) Image After Segmentation.

C. Feature Extraction

Feature extraction is the process of transforming the raw pixel values from an image, to a more meaningful and useful information that can be used in other techniques, such as point matching or machine learning. In the proposed system we are using the Scale Variant Feature Transform (SIFT) algorithm. Features of the leaf like colour, texture, shape, vein. Numerous methods can be employed for feature extraction such as Histogram of Oriented Gradient (HOG), Zernike Moments, Hu's Moment and others.[1] In this research we used SIFT & Canny Edge Detection Algorithm (CEDA). Where CEDA is used for feature extraction of leaf.

The following information can be extracted from the Image.

1) *Convex hull Information:* It is formed using boundary point of leaf. Convex hull is approximated and number of vertices is extracted. The area of Hull and Perimeter is also Calculated.[7]
2) *Morphological Information:* The length and width of leaf is calculated by finding minimum and maximum x and y coordinates. The perimeter and area of leaf is also calculated using boundary points.[7]
3) *Distance Maps:*

a) *Vertical and Horizontal maps*: Lines are drawn on segmented image and each line consist of minimum and maximum image which intersect with leaf and its respected axis. The distance of each line is estimated and stored. This is above in Fig 3a.[7]
b) *Centroid radial map:* The centroid of leaf is found by intersecting the diagonal axes though the bounding box along the leaf. Sixteen point radiating from centroid are taken on boundary box. The Euclidian distance between this point and centroid are calculated as well as that between the centroid and the point which intersect the leaf boundary. This is shown in Fig 3.b[7]

The Information extracted are then used to create ratio which will be used in pattern matcher.

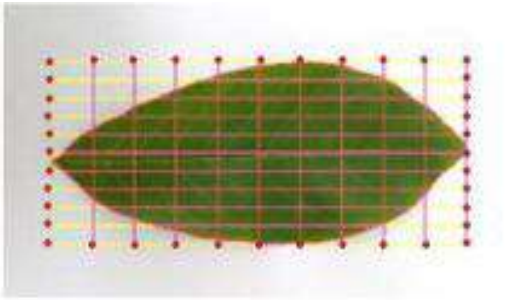


Fig -3 a) Distance Map X and Y[7]

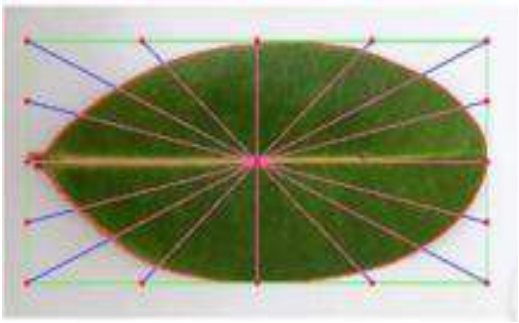


Fig -3 b) Centroid Radical Distance Map[7]

4) *Color Histogram:* It is a computed for a cropped part of image since if the whole is used, white spaces surrounding the leaf would affect the histogram. To Crop the image, the length and width of bounding box are used as markers to crop the central part of leaf image.[7]

5) *Matcher:* It consisted of 2 stages of CNN algorithm. All ratio is normalized to 0 or 1 before any comparison are made.

a) *Stage 1*: The leaf to be recognized undergoes the same process as the ones in database. The new leaf is then compared to training set one by one. The sum of all the Euclidian distance between the leaf and those in database are calculated. The 3 closest distance are returned. Each Ratio then feature in CNN classifier. For the distance and centroid map all the distance in set are considered as individual feature. [1,7]

b) *Stage II*: For instance, where the result set from stage 1 consist of different plants, color histogram of the new leaf compared those from result set. The correlation is calculated. Its value lies between -1 to 1. Value close to 1 means that very high correlation which means that the image is very similar. The closest match is calculated using CNN algorithm. [1,7]

More are the images greater is the efficiency of CNN algorithm.

Table below shows the ratio that have been calculated from previous measurement.[7]

Aspect Ratio	White area ratio	Perimeter to area	Perimeter to hull	Hull area ratio	Distance map x	Distance map y	Centroid radical distance
Width/Length	Area of Leaf/ (Length*width)	Perimeter of Leaf/ Area of Leaf	Perimeter of hull/ Perimeter of leaf	Area of leaf/Area of hull	Distance of lines parallel to x-axis/Length	Distance of the lines parallel to y-axis/Width	Distance from centroid to intersecting points/distance from centroid to boundary point

Table1: Ratio

III. CLASSIFICATION OF PLANT SPECIES AND DISEASE DETECTION

The processing scheme consists of test RGB image acquisition from database or web. Image pre-processing includes image enhancement and image segmentation where the affected and useful area are segmented each filter having size of 512 X 512 pixels. Here the size of feature vector is the size of image 512 X 512 pixels. Fig -4b shows that Enhance Test image using histogram equalization. Preprocessing the test image using histogram equalization is applied to increase the contrast in low contrast image where, leaf spot is highlight in Fig -4b. [2,4,5,7]

Fig -4 a) Test RGB Image

Fig -4 b) Histogram Equitization of Image

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 7 Issue IV, Apr 2019- Available at www.ijraset.com

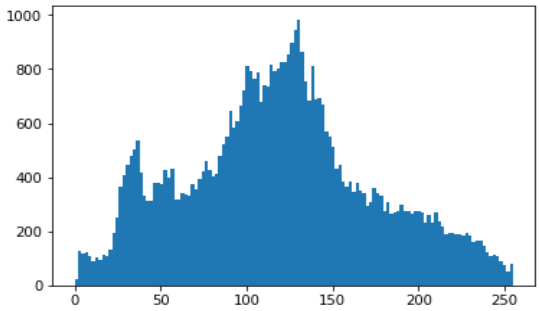


Fig -4 c) Histogram of Concentrated Pixels of image

A. Leaves Disease Detection Parameter

Recognition Accuracy Comparisons, Execution Time Comparisons, False Accept and False Reject Rates for D-Leaf datasets are compared below in respective table. Table -2 shows that recognition accuracy for detecting diseases on D-Leaf dataset. It shows that CNN algorithm has highest accuracy up to 97%.

Sr. No	Feature Extraction method/No of dataset	CNN Classifier Accuracy
1	100	85.6
2	200	86.9
3	300	87.9
4	400	90
5	500	91.3
6	600	93.6
7	700	96.8

Table 2: Recognition accuracy comparison

B. Execution Time

Table -3 shows that execution time in second for detecting diseases on D-Leaf dataset. Out of which CNN Classifier algorithm takes less execution time.

Sr. No	Feature Extraction method/No of dataset	CNN Classifier Execution or Compiling time in sec
1	100	45.6
2	200	60.6
3	300	70.5
4	400	80.6
5	500	100.4
6	600	120.6
7	700	144.5

Table 3: Execution Time

IV. RESULTS

The method of testing used was to use every photo of leaves in the database as an input image to the system, compare it to all other leaves and calculate the percentage accuracy of the system. This technique has an advantage of testing all images in database rather than just a small percentage of it. Every time the system applies the matcher to a leaf, it will create in a CSV file with the actual plant name and the predicted plant name.[5]

Testing was done on 740 leaves coming from different species of plants. In Particular we noted a 100% accuracy for seven different type of leaves. Only few plant species recognitions were below 85%.

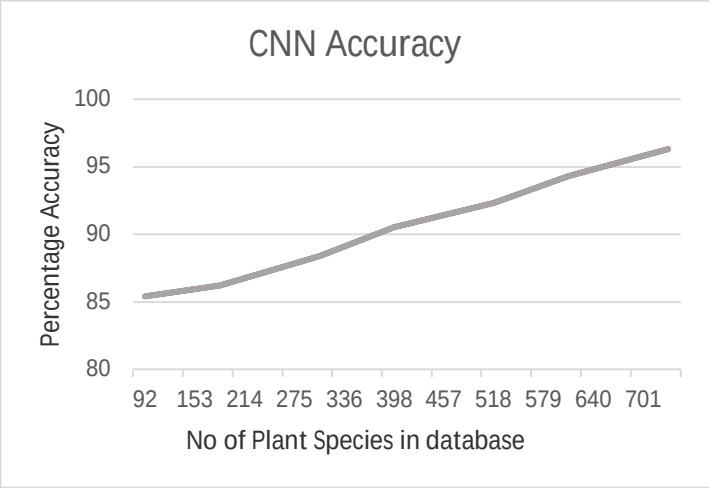


Fig. 5 Effect of increasing in number of plants species on classifier.

From above fig 5 it can be noted that with only eight species, we have a very low classifier accuracy of 85.4% which rapidly increases with the increase in number of species and touches to 96.3% in our system. We usually expect that the recognition accuracy to go down for more variety of dataset. However, in this case the accuracy is going up rapidly and it is still very less for 740 species. As the no of leaves of the plant increases the efficiency of classifier also increases in proposed algorithm. However, there is no increase in the accuracy of the classifier once the saturation limit is been reached. Overall accuracy follows similar trend but there is increase in almost (+-1%) overall accuracy using CNN classifier, after adding each additional leaf considering the starting amount of leaves. Thus, it is possible to obtain high number of accuracies by using relatively large number of species but with small no of leaves per species. We also demonstrated how accuracy varies with number of plant species and number of leaves. The accuracy obtained are comparable with existing work. However, as our approach is based on CNN classifier it is expected to run faster than PNN and KNN classifier. The effect of varying number of species and number of leaves had been successfully tackled in this literature. The Overall accuracy at the first phase was close to 90%. As the dataset increased the accuracy for the classifier also got increased significantly. The Color histogram matching operation was then applied on result from the first phase and accuracy rose. Moreover, the same test was carries on entire D leaf dataset the much higher accuracy close to 95% where obtained in first and second stages respectively. Thus, we can see this approach is very effective in classification of plant and leaves. Further experiment was conducted in order to assess the effect of increasing the number of plant species and the number of leaves on classification accuracy. We used CEDA and SIFT as we have greater amount of dataset and the other algorithm like SEDA and PCA are applicable only for short amount of dataset. We found that by using above algorithm our efficiency on out has significantly increased and does not vary time to time. Thus, we Successfully tested various species of plant in the database where we came to conclusion that as the number of species increases the efficiency of the Convolutional Neural network classifier also increases.

V. CONCLUSIONS

Plants play an important role in our lives, without plants there will not be the existence of the ecology of the earth. The large amount of leaf types now makes the human being in a front of some problems in the specification of the use of plants, the first need to know the use of a plant is the identification of the plant. The proposed system is an automated client-server system and is capable of classifying plants by extracting all morphological features of leaf from binary masks of leaf and predicting plant species using a SIFT feature detection algorithm and by canny edge detection algorithm giving the highest accuracy of 90 and 97.5 percent respectively. Hence saving the loss of biodiversity and reducing the dependency on the expert to a certain extent is possible. It can provide the help for a person having less knowledge about the plant species. So, we can conclude that these algorithms are the most suitable for this task. The proposed system is web-based system with user interface. In future, we can design an application for plant species classification with some improvement in technology support.

VI. DATASET

The leaf images dataset used in this research and a graphical user interface of D-Leaf is available for download from: JW Tan & Siow-Wee Chang (2017), D-Leaf dataset. fig share. https://figshare.com/articles/D-Leaf_Dataset/5732955.[1]

VII. ACKNOWLEDGMENT

We welcome this opportunity to express my heartfelt gratitude and regards to Prof Anand Bhosale, Department of Information Technology, International Institute of Information Technology, Pune, Maharashtra, India, for his unconditional guidance.

REFERENCES

[1] Jing Wei Tan, Siow-Wee Chang, Sameem Abdul-Kareem, Hwa Jen Yap, Kien-Thai Yong, "Deep Learning for Plant Species Classification using Leaf Vein Morphometric", : DOI 10.1109/TCBB.2018.2848653, IEEE/ACM Transactions on Computational Biology and Bioinformatics
[2] Pawan P. Warne1, Dr. S. R. Ganorkar2 , "Detection of Diseases on Cotton Leaves Using K-Mean Clustering Method", Volume 2 April 2015,IRJET.
[3] Krithika, N., Selvarani, "An Individual Grape Leaf Disease Identification Using Leaf Skeletons and KNN Classification," ICIIECS March 2017.
[4] Dias, R. O. Q., Borges, D. L., "Recognizing Plant Species in the Wild: deep learning results and a new database," 2016 IEEE ISM.
[5] Vishnu S and A. Ranjith Ram, "Plant Disease Detection Using Leaf Pattern: A Review," IJISET Volume 2 Issue 6 June 2015.
[6] Elhariri, E., El-Bendary, N., Hassanien A. E.," Plant Classification System based on Leaf Features",ICCES December 2014
[7] Trishen Munisami, Mahesh Ramsuran, Somveer Krishnah, SameerchandPudaruth," Plant leaf recognition using shape feature and color histogram with k nearest classifier".August 2015,Procedia Science direct.
[8] J. Fan, N. Zhou, J. Peng, L. Gao, "Hierarchical Learning of Tree Classifiers for Large-Scale Plant Species Identification," IEEE TRANSACTIONS ON IMAGE PROCESSING, 2015.
[9] Prof.Anand Bhosale ," Efficient Estimation of Complex Queries",International Journal of Advance Research in Computer Science and Software Engineering,Volume 4 April 2016.
[10] Sagar P. Jaikar, Ramkrushna C. Maheshwar, Sameer P. Mamadapure Anand A. Bhosle ," Secure Data Distribution using Secret Splitting Over Cloud",Global Journal of Computer Science and Technology :B Cloud and Distributed.,Volume 17 Issue 1,2017.

RESEARCH REVIEW JOURNALS
Volume-03
Issue-12
December -2018

ISSN: 2455-3085 (Online)
RESEARCH REVIEW International Journal of Multidisciplinary
www.rrjournals.com [UGC Listed Journal]

Amigo: A Smart Helper

[1]Ashwini Gote, [2]Maithilee Inge, [3]Nikita Jadhav & [4]Prof. Prashant Gadakh

[1,2,3]Student, Department of Computer Engineering, International Institute of Information Technology, Pune (India)
[4]Professor, Department of Computer Engineering, International Institute of Information Technology, Pune (India)

ARTICLE DETAILS

Article History
Published Online: 10 December 2018

Keywords
Personal Assistant, Natural Language processing, Speech recognition, Speech Processing, DialogFlowApi, Smart cleaner, Intrusion detection system, Raspberry Pi3,Dc motor

***Corresponding Author**
Email: ashwinigote144[at]gmail.com

ABSTRACT

Robots have become an integral part of 21st century due to their excessive use in industries, household, hotels and offices. A new concept of using a robot as a friend and personal assistant along with cleaning facility is been introduced here. There are many personal assistants available in the market like Amazon's Alexa, Apple's Siri and for cleaning purpose iRobot's Roomba is there but when we club these two things together in a very cost effective manner so versatility is increased and no need of buying two separate things. It can also act as a friend for alone people as it is intelligent robot to generate the reply from the voice based request received using some AI algorithms and NLP. Basic Objective is to enable machine to perform such a intellectual tasks as decision making, problem solving, perception, understanding, also human communication. It aims to act as a friend who can think rationally and help when needed.

Proposed system will be capable of activities like acting as chat-bot, reading books, playing music, making a call, capturing images. Cleaning is also an important factor for healthy living and hygiene but due to lack of time it is being neglected. Hence we can have this objective as automatic and intelligent automatic floor cleaner having two facilities like vacuum cleaner and floor cleaner with a mop attached to it. There are options available but those are costly to buy so the main aim to build a robot who can act as a friend and cleaner to the user. Also for extensive cleaning we are using intrusion detection system to avoid the robot failure and cover more floor area to clean.

1. Introduction

In the era of technology, humans are planning and inventing such new technologies that would reduce human efforts and improve the living standard. Even the manual work is also possible with the help of robotics and many of the industries are extensively using robot appliances too. The work of personal assistant is also possible with the help of artificial intelligence and robotics called as virtual personal assistant. Physically challenged people and the elderly persons find it difficult to handle new objects with new technology therefore they require assistance for the same. Hence the idea of virtual personal assistant came from here.

Now the aim is to develop a user friendly robotic assistant which is operated with the help of voice commands. Personal assistants, known by various names such as intelligent personal assistants, digital personal assistants, voice assistants or mobile assistants,. This assistant robot can be extensively applied in many fields like chemical industry, homes, manufacturing and also in medical field. This voice based system is intelligent enough which is able to help users to complete their tasks in a more efficient way and remember their to-do list. Voice based assistant is being applied in various devices and applications for providing guidance to users like in smart phones and car for navigation purpose. It has a wide range of applications like in business, education, government, entertainment and healthcare.

Available products in the market for virtual personal assistant are Microsoft's Cortana, Apple's Siri, Amazon Alexa, and Google Assistant. These big giants have used different methodologies for developing the product like Google used Deep Neural Networking concepts to design Google Assistant and Microsoft applied machine learning for developing Microsoft Azure. But the cost is high for people to buy.

Our Proposed system is the great combination of virtual personal assistant and smart floor cleaner. As the available personal assistant have the problem of mobility. Now a days, people are more attracted towards robotic cleaners due to their effectiveness in floor cleaning applications at homes, industries and hospitals. Robotic cleaners are distinguished on their method of cleaning like mopping or vacuum cleaning. Also some products have obstacle detection and avoidance mechanism using the intrusion detection sensors and ultrasonic sensors for the precise measurement of distance between system and intrusion or obstacle. The available products are like iRobot's Roomba which is a smart floor cleaner have advantages like intrusion detection mechanism in it and covers more floor area to clean but not very cost effective so it may appear as a luxury item to buy for common people. The main objective of this work is to provide a substantial solution to the problem of manufacturing robotic cleaner utilizing local resources while keeping it low costs with the help of Raspberry pi 3.

As discussed above, Amigo is the solution for the above two problems for personal assistant and smart cleaner. Amigo in Spanish means a friend whereas it also acts as a helper for mankind. It means a friend will help you when you will require to do so. Amigo can talk with you as a friend, can answer your questions, can read a book for you, also it can follow your commands, it recognizes your voice. It also gives information about the current traffic and current affairs too. Additional

feature here is it can also clean the floor reducing the work. Here the user will totally get a helping friend in case of knowledge, chit-chatting and floor cleaning.

Problem Statement:

Amigo which is the proposed system comprising of two facilities like virtual personal assistant and smart cleaner and designed with the help of system on chip Raspberry pi 3. Hence it reduces the need for buying two gadgets separately so it can termed as the cost effective solution. Amigo is a robot who can think rationally and operates on voice based communication i.e. request and reply. This system uses NLP and speech recognition algorithms from artificial intelligence and generates the reply from DialogflowApi provided by Google.

Basic Idea/Survey:

• Basic idea behind developing a robot is to enable machine to perform such intellectual tasks as decision making, problem solving approach with easy method, perception which consider benefits, understanding the problem, also human – machine communication which allows humans to interact with machines.

- Amigo aims to act as a friend who can think rationally, suggest good things for you, help you whenever you need, act as guide for solving your problems and a good entertainer for you. .
- The voice based communication is needed where one can ask amigo for anything he/she wants to get performed. As its communication so amigo also answers the query through a voice. Here this communication requires text-to-speech and speech-to-text technologies.
- If amigo doesn't have solution or answer for your query it can search it on web for the correct answer, the network connectivity can be provided with the help of Wi-fi or GSM module.
- Even amigo has a battery support so that one can carry it anywhere and amigo will be ready to help you every time. It can also clean the floor when required.

a) Personal Assistant Robot

1. **Technologies Used:** Face Recognition, Autonomous Navigation, Digital Surveillance, Remote Control
2. **Description:** This paper presents the idea of design and implementation prototype of personal assistant robot. It aims to investigate how an intelligent machine interact with a human in a smart way. Technologies like Face Recognition and tracking is used for interaction with master in smart and secure manner.Environment information is essential for robot obstacle avoidance. The robot is equipped with an ultrasonic sensor module to estimate the distance between Personal Assistant Robot and its immediate environment. Autonomous Navigation helps to navigate in unstructured environment.

b) A Voice-Controlled Personal Assistant Robot

1. **Technologies Used:** control over voice, Bluetooth, Android based smart devices
2. **Description:** In this paper control over voice is used for communication between the human and robot. Personal robotic assistants help reducing the manual efforts being put by humans in their day-to-day work. The voice commands are given to the robot using a smart mobile phone which is based on an Android OS based platform. The voice signal is then converted to the text form using an online cloud server, in real time. This text command is sent via Bluetooth network of the smart-phone to the Bluetooth module on the robot. The Bluetooth module on-board the robot receives the text signals and forwards it to the microcontroller on-board the robot's body.

c) Next-Generation of Virtual Personal Assistants (Microsoft Cortana, Apple Siri, Amazon Alexa and Google Home)

1. **Technologies Used:** Multi-modal Dialogue Systems; Gesture Recognition; Image Recognition; Image Recognition.
2. **Description :** In this paper have the multi-modal dialogue systems is used which process two or more combined user input modes, such as speech, image, video, touch, manual gestures, gaze, and head and body movement in order to design the Next Generation of Virtual Personal Assistant model. The new model of Virtual personal Assistant will be used to increase the interaction between humans and the machines by using different technologies, such as gesture recognition, image/video recognition, speech recognition, the vast dialogue and conversational knowledge base, and the general knowledge base. Virtual Personal Assistant system can be used in other different areas of applications, including education assistance, medical assistance, robotics and vehicles, disabilities systems, home automation, and security access control

d) Design and Development of Floor Cleaner Robot (Automatic and Manual)

1. **Technologies Used:** RF Module, LCD, IR sensor
2. **Description:** This paper proposed the system where floor cleaner robot can work in any of two modes "Automatic and Manual". All hardware and software operations are controlled by AT89S52 microcontroller. The developed robot can perform sweeping and mopping task.

RF modules have been used for wireless communication between remote (manual mode) and robot and having range 50m.IR sensors are used to obstacle detection so that robot can avoid the obstacle and navigate to other direction properly . Here robot is with IR sensor for obstacle detection and automatic water sprayer pump. Four motors are used, two for cleaning, one for water pump and one for wheels. Dual relay circuit used to drive the motors one for water pump and another for cleaner. In the manual mode, RF module is used

for transmitting and receiving the information between remote and robot and display the information related to the obstacle detection on LCD.

2. Project Architecture

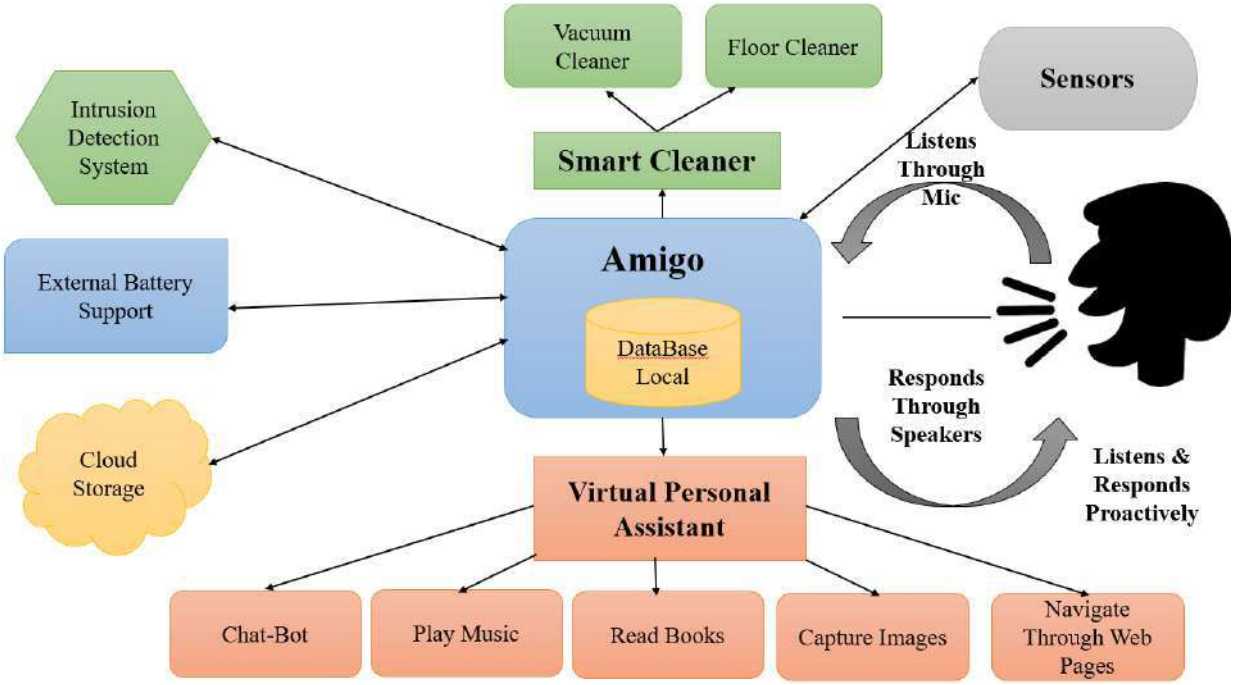


Fig 1: Architecture of Amigo

1. For processing of the whole system Raspberry Pi is used and interfaced with sensors.
2. Transducer mic will be always in the state of sensing the voice input by user then it will listen to the input and send it to speech recognition module and speech recognition module will convert the voice into text form and then analyze whether it belongs to smart cleaning module or personal assistant module.
3. If it belongs to cleaning module or it will compare with the command then start the motor and vacuum cleaner for cleaning purpose also the intrusion detection system will work here to avoid the system hardware loss.
4. If the user input is the request for task related to personal assistant then accordingly it will generate the response using DialogflowApi and will respond using speaker.
5. It will firstly search in its local database for the query if not found then will look for the cloud storage or will navigate the user through web pages.
6. Also the DHT11 sensor will be in active state always for sensing the humidity and temperature of the atmosphere and it will be displayed on the 2 x 16 LCD screen attached to it.

3. Project Approach

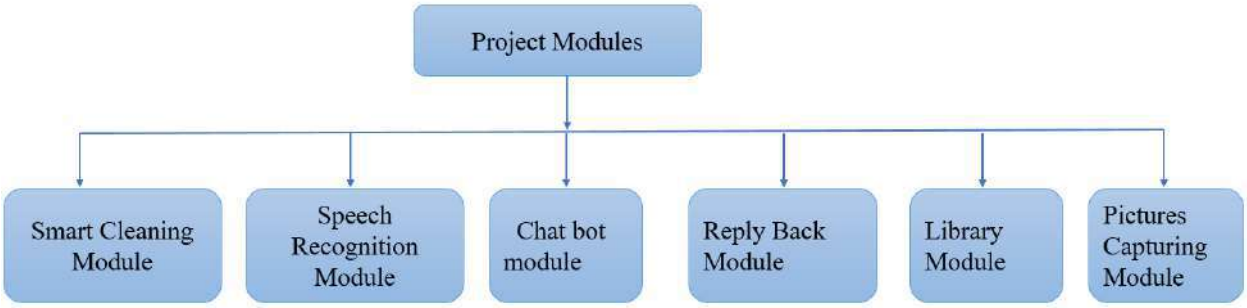


Fig 2: Project Modules

1) **Smart Cleaning Module:** Includes both vacuum cleaner and floor cleaner also detects the intrusion for machine safety. Also turns on the dc motor to clean the floor.

2) **Speech Recognition Module:** Includes the speech recognition that keeps the microphone continuously on for receiving the sound signal at any moment.

3) **Reply Back Module:** Includes the processing and analysis of the received command and matching it with the most appropriate data set and answering back with the help of speakers.
4) **Pictures capturing Module:** Includes the capturing of pictures with the help of camera attached to it.
5) **Library Module:** Includes reading of books which are there on the local storage or on the internet or may be cloud.
6) **Chat-Bot Module:** It includes the communication between Amigo and human owner with the help of dialogflow or Jarvis Amigo will be able to understand what the user wants to say and reply accordingly

Algorithm:

Algorithm for Floor Cleaner:

- Map can be interpreted as 2-d grid, with obstacles as obstructed points in the map.
- Obstacles can be arbitrary, both in quantity and position.
- The algorithm knows nothing about the surrounding environment.
- The robot only provides 3 API: turn_left, turn_right (rotate the looking direction) and move (move ahead 1 point)

Algorithm for Virtual Personal Assistant:

- DialogFlow uses speech recognition algorithm and then for processing of the given command it uses the google speech to text converter to make the command machine understandable.
- Now it analyses the text written command and firstly checks it with the local dataset available & if not then it uses web pages to get the data from navigation.
- Here the interpreted command is being processed and then the most appropriate reply is generated and processed with GTTS and finally the machine starts communicating.

4. Conclusion

Hence this paper describes about voice controlled system comprising an idea of combining a virtual personal assistant and concept of smart cleaner using Raspberry Pi and interfacing it with different sensors like ultrasonic sensor, infrared sensor for developing an intrusion detection system to avoid the system loss which will work in smart cleaner module and for the development of personal assistant a transducer mic is used for sensing the voice input from user and reply generation is done using DialgflowApi which is then given to the user through speakers attached to the system. The available systems in the market are having high costs hence our motive is to provide additional features in low cost and with improved efficiency. So no need to buy two gadgets separately. The paper shows a better and simple approach to provide an overview of design of a simple personal assistant &robotic cleaners using RPI3, sensors and available Apis. Functionalities covered in personal assistant are chat-bot, reading books, playing music, capturing images and navigate through web page to tell information about current affairs. In future we can extend the ability of Amigo clubbing it with emerging trend IOT for converting this personal assistant into home assistant. To develop self-learning ability and self-correction in its functions to avoid human intervention. In future work we can also add healthcare facilities like Heart-beat sensor, BP measurement, simply it will act as a caretaker.

References

1. Chia-How Lin, HenrryAndrian, Yao-Qing Wang, and Kai-Tai Song, “Personal Assistant Robot ”, Proceedings of the 2005 IEEE International Conference on Mcchafronics July 10·12, 2005, Taipei, Taiwan.
2. Anurag Mishra, Pooja Makula, Akshay Kumar, Krit Karan and V. K. Mittal, ”A Voice-Controlled Personal Assistant Robot”,2015 International Conference on Industrial Instrumentation and Control (ICIC) College of Engineering Pune, India. May 28-30, 2015.
3. VetonKëpuska,GamalBohouta, ”Next-Generation of Virtual Personal Assistants (Microsoft Cortana, Apple Siri, Amazon Alexa and Google Home) “,Proceedings of the 2005 IEEE International Conference on Mcchafronics July 10·12, 2005, Taipei, Taiwan.
4. Shripad Malavadikar, Swapnil Mungale, ToshikaJohri, HarshadLokhande, ”Automatic Cleaner Robot”,International Engineering Research Journal (IERJ), Volume 2 Issue 8 Page 2617-2620, 2017 ISSN 2395-1621.
5. Manreet Kaur, PreetiAbrol, ”Design and Development of Floor Cleaner Robot (Automatic and Manual)”, International Journal of Computer Applications (0975 – 8887) Volume 97– No.19, July 2014.
6. PaPeRo, NEC Personal Robot Center, http://www.incx.nec.co.jp/roboUrobotccnter.html.
7. Development Overview, htlp:llwww.sdia.or.jp/mhikobec/produets/ete/robot.html.
8. P. Hall, D. Marshall, and R. Martin, "Incremental eigenanalysis for classification," in Proe. British Machine Vision Conference, 1998, pp. 286-295.
9. J. Borenstein, Y Koren, "Error Eliminating Rapid Ultrasonic Firing for Mobile Robot Obstacle Avoidance," IEEE Transactions an Robotics and Automation, vol.l!. no. I, pp.132-138, 1995.
10. Wen-Pin Chen, "Development of an Environment Exploration system for a Home Robot," Master Thesis, National ChiaoTung University, July, 2002,
11. Kai-Tai Song and Wen-Jun Chen, "Face Recognition and Tracking for Human-Robot Interaction," in Proc. of IEEE 2004 International Conjerence on Systems, Man and Cybernetics, the Hauge, the Netherlands, 2004, pp.2877-2882.
12. P. Hall, D. Marshall, and R. Martin, "Incremental eigcnana1ysis for classification," in Proc. British Machine Vision Conjerence, 1998, pp. 286-295.

RESEARCH REVIEW JOURNALS
Volume-03
Issue-12
December -2018

ISSN: 2455-3085 (Online)
RESEARCH REVIEW International Journal of Multidisciplinary
www.rrjournals.com [UGC Listed Journal]

Personalized Outfit Recommendation Mobile Application

[1]Aravind Renjan, [2]Bhumika Kaushik, [3]Vaishnavi Gaikwad, [4]Yash Pathak, [5]Prof. Bailappa Bhovi & [6]Prof. Prashant Gadakh

[1,2,3,4] Student, Department of Computer Engineering, International Institute of Information Technology, Pune (India)
[5, 6] Professor, Department of Computer Engineering, International Institute of Information Technology, Pune (India)

ARTICLE DETAILS

Article History
Published Online: 10 December 2018

Keywords
Machine Learning, Apriori algorithm, QR Code Technology, Association Rules, Data Mining

***Corresponding Author**

Email: aravindrenjan[at]gmail.com

ABSTRACT

Organizing and managing a wardrobe is a tedious and time-consuming affair on a daily basis, so is keeping up with the current trends in fashion and finding the perfect outfit that suits the occasion. This mobile application envisions automating the entire process of choosing an outfit according to the weather and occasion with as minimal effort from the user as possible. The application uses Apriori algorithm to find the association rules of clothing items that can be worn together based on certain properties like fabric, color, texture, pattern, etc. These rules will be used to model a recommendation system. The properties are recognized via QR code which will be attached to the label of the clothing item generated by the vendor from the application website. The application would primarily create a virtual closet for the user and also suggest outfits according to the user preference, weather and occasion.

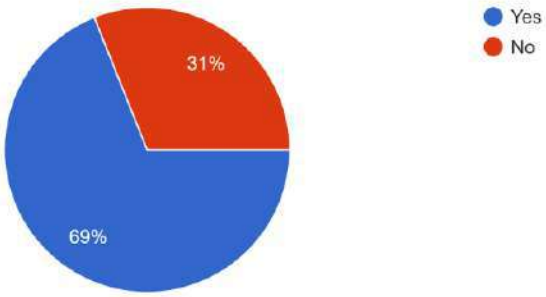
1. Introduction

Every day a great amount of time is spent in finding the outfit for the day and occasion. Choosing a set of clothes that match and also compliment the weather is a tedious task. It is almost impossible for a person to keep up with the rapid changes in trends in fashion unless they are from the fashion industry or have a keen interest in it. Physically taking out each piece of clothing that a person owns and seeing if two of them match and repeating the same to see what other choices are available is a difficult thing to do and also takes up a good amount of time on a daily basis. Moreover, while purchasing new clothes a person has to keep in mind all the clothes that he/she already has available in his/her wardrobe and envisage if the clothes available to buy match with any of those in their wardrobe.

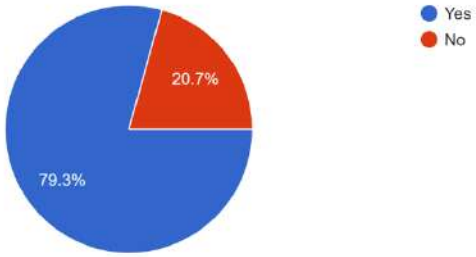
All the solutions to these difficulties and efficient use of time can be achieved by automating these everyday actions. The proposed application aims to create a virtual closet for each user which will have a repository of all the clothes owned by a person available online all the time. The application will suggest an outfit for the user from the available clothes in the closet by taking into consideration the occasion and the weather.

2. Market Survey

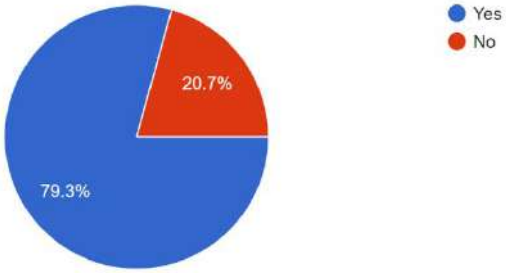
Is wardrobe management a hassle?



Is choosing an outfit for a day difficult?

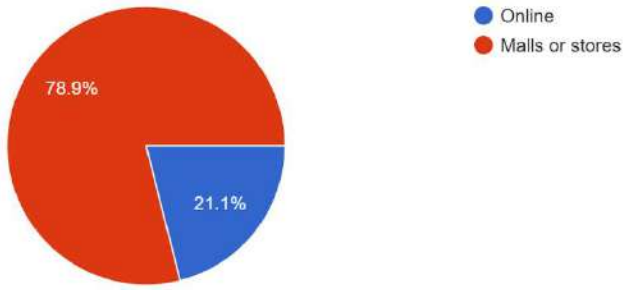


Would you like to have an app that can choose an outfit out of your own closet?

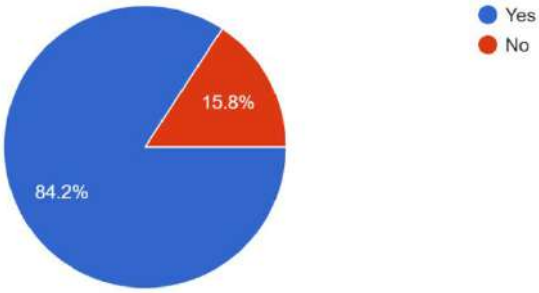


The surveys conducted shows that wardrobe management is a big hassle for many. Choosing what to wear every day when the wardrobe is not properly managed is a time-consuming activity. Having an app that would help choose an outfit from already available clothes in the wardrobe is also preferred highly. It can be fairly assumed app that would create and manage wardrobe virtually all on its own and help recommend outfits for everyday wear and also depending on the occasion and weather would be highly appreciated by the survey demographic.

Where do you shop usually?



Would you prefer buying from malls/store if you got recommendations based on what you bought previously and based on your personal choice?



And it is also seen in the result of the surveys that most consumers still prefer going to malls and stores to buy clothes. Even though going to malls and stores have many advantages a major disadvantage that can be seen is that while buying clothes from a mall or a store a consumer has to imagine whether the clothing item would make a complete outfit with the already available clothes in their wardrobe. This can be solved by our app as it can put together the outfit using both the new and the available clothing units in the wardrobe. A consumer can see an image of the selected clothes as a whole outfit and decide on buying the clothes.

3. Existing System

Initially a survey to find similar existing systems was conducted where in two different already existing apps were found and studied upon. One was a styling application that gives outfit ideas for a user. The application provides the user with an outfit that goes well together but does not take into consideration the preferences of the user. Also, the application suggests only outfits that the user can buy, not an outfit constructed out of the user's closet. Basically, this application suggests a user outfits to buy rather than wear.

The second system is another styling application that recommends outfits from user's clothes which the user has to add to the repository by taking pictures and entering necessary information along with it. The suggestions are made from the outfit ideas created by the user which has to be done manually. The user need not have the enough knowledge or time to create outfit ideas. And there are no suggestions for buying clothes that help create an outfit.

4. Proposed System

Our application would suggest a user outfits from the set of clothes available with the user to reduce the everyday hassle of finding an outfit. This suggestion model would take into consideration the preference of the user, weather and also the occasion. In the initial stage of use by any user the clothes that are already available with the user need to be manually entered into the virtual closet by the user. But with time the need for user entry of data will not be necessary as vendors can print the QR code with the tag on the cloth. And a user would simply need to scan the QR code to add the cloth to the private closet. As the application is a virtual closet organizer the user will be able to delete a cloth from the closet too. The application will also suggest users with clothes that can be bought that make an outfit with the clothes that are already present in the closet of the user.

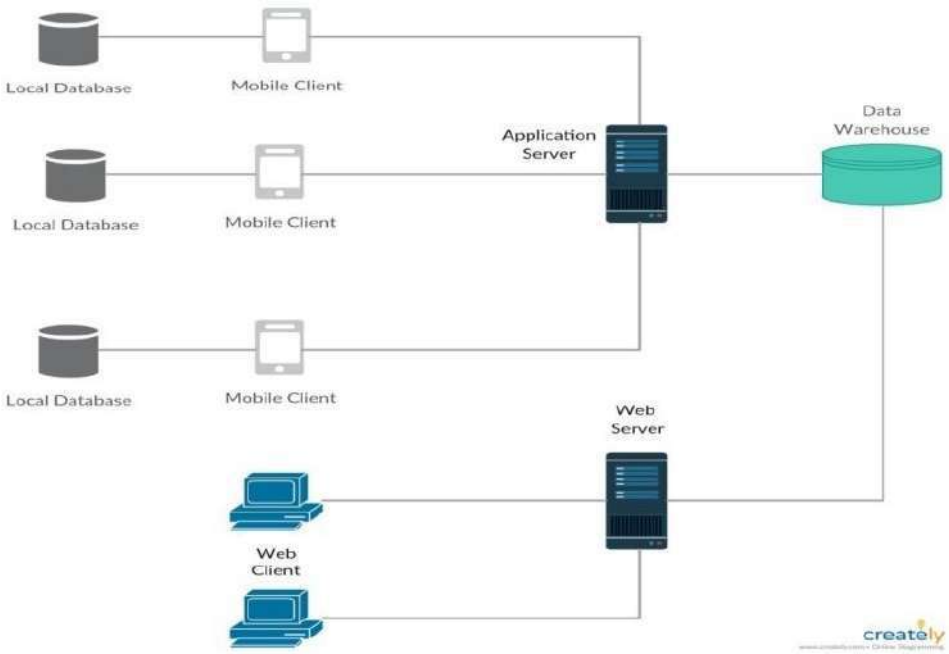


Fig.1 Architecture Diagram

The proposed system will be making use of the following:

Apriori Algorithm:

Association rule mining is the technique of finding frequent rules that define relations between seemingly unrelated items in a dataset. The two main measurements used for association rule mining are confidence and support. Association rule mining generally works in two steps. The first is to find all item sets with pre-decided minimum support and the second is generating association rules by combining these frequent items.

Apriori algorithm is one of the most popular data mining approach to find frequent patterns in a dataset and to mine association rules from it. These rules can be used to find relations from user preferences and can be used to recommend outfits to a user[1]. For example a relation like:

{(cotton, white,shirt) → (blue,jeans) → (sunny)}

Tells us that a cotton white shirt is preferred with blue jeans on a sunny day.

Similarly, many relations like the one above can be used to recommend an outfit to a user. Most of the rules will be general and can be used for many users and then can be bettered according to their personal preferences.

Apriori uses breadth first search and a tree structure to count the candidate item sets efficiently. The algorithm generates candidate item sets of length n from item sets of length n-1. The candidate item set is pruned to remove item sets that are not frequent. According to the downward closure lemma, the candidate set contains all frequent n-length term sets. It is an iterative process that repeats until the most frequent item set that are above the minimum support threshold are mined. The result is then used to generate association rules which are above the minimum confidence threshold.

Limitations of Apriori algorithm:

1. Slow and candidate generation bottleneck.
2. Repeated scans of the database increasing time complexity.
3. Large number of non-frequent item set generation increasing the space complexity.

There are multiple solutions to solve these limitations of apriori algorithm like FP Growth (Frequent Pattern Growth) method which generates an FP tree using with frequent patterns are generated thus reducing the database scans to only two times. And also, methods like partitioning which partitions the data into various partitions to find local frequent items sets which can then be used to find the global frequent item sets and sampling can implemented to increase the efficiency of the apriorialgorithm[2].

QR code:

SumitTiwari[3] gives a brief about QR code technology in his paper. A QR code is a type of matrix bar code or two-dimensional code that can store data information and designed to be read by smart phones. QR stands for “Quick Response” and hence implying that the contents are to be decoded rapidly. The code consists of black modules arranged in a square pattern on a white background. It has many advantages which include omnidirectional and fast scanning, small size, more storage space compared to a barcode, error correction, varied types of data support and so on. Increased use of smart phone in everyday life and QR scanners in smart phones have made this technique more accessible to public in recent times. The clothes will be added by scanning a QR code attached to the tag of the clothing item, which a manufacturer can obtain via the website of the application. It can be used to add clothes or match up with other clothes available in the virtual wardrobe of the application.

5. Conclusion and Future Works

In this paper, we develop a novel application to help people with the everyday hassle of choosing an outfit; Creating and easy organization of a virtual closet. Providing personalised suggestions to users with regards to various parameters like occasion, weather and preference. The system also provides the manufacturers with a website from which they can procure a QR code to be attached to the cloth tag that the user can use to easily add the item to their closet or be used by the system to suggest the clothing item to any user that has another that goes with it and completes the outfit. I future the we plan to add existing clothes to the closet with only the click of an image of the clothing and recognizing the necessary data (tags associated with the item) from the image. Also expand the closet horizon by adding accessories to the closet; And also providing a platform for purchase of clothes and other items by connecting the manufacturers and customers directly to each other.

References

1. Jugendra Dongre, Gend Lal Prajapati, S.V. Tokekar, “The Role of Apriori Algorithm for Finding the Association Rules in Data Mining”,2014 IEEE.
2. Chanchal Yadav, Shuliang Wang, Manoj Kumar, “An Approach to Improve Apriori Algorithm Based in Association Mining”, 2013 IEEE.
3. Sumit Tiwari, “An Introduction to QR Technology”, 2016 IEEE.
4. Berry, M. J. A. and Linoff, G. Data mining techniques for marketing,
5. sales and customer support, USA: John Wiley and Sons,1997
6. Andre Bergmann, "Data Mining for Manufacturing: Preventive
7. Maintenance, Failure Prediction, and Quality Control".
8. Fayyad, U. M; Piatetsky-Shapiro, G.; Smyth, P.; and Uthurusamy, R. 1 996. Advances in Knowledge Discovery and Data Mining. Menlo Park, Calif.: AAAI Press.
9. Dr. Gary Parker, vol 7, 2004, Data Mining: Modules in emerging fields, CD-ROM.
10. Jiawei Han and Micheline Kamber (2006), Data Mining

Concepts and Techniques, published by Morgan Kauffman, 2nd ed.
11. Literature Review: Data mining, http://nccur.lib.nccu.edu.twlbitstream/ 1 40.1 I 9/3523 I/S/35603 I OS.pdf, retrieved on June 2012.
12. H. Mahgoub,"Mining association rules from unstructured documents" in Proc. 3rd Int. Conf. on Knowledge Mining, ICKM, Prague, Czech Republic, Aug. 25- 27, 2006, pp. 1 67-1 72.
13. S. Kannan, and R. Bhaskaran "Association rule pruning based on interestingness meas ures with clustering". International Journal of Computer Science Issues, IJCSI, 6(1), 2009, pp. 35-43 .
14. M. Ashrafi, D. Taniar, and K. Smith "A New Approach of Eliminating Redundant Associa tion Rules". Lecture Notes in Computer Science, Volume 31 S0, 2004, pp. 465 -474.
15. P. Tang, M. Turkia "Para llelizing frequent itemset mining with FPtrees".
16. Technical Report titus.compsci.ualr.edu/-ptang/papers/par-fi.pdf, Department of Computer Science, University of Arkansas at Little Rock, 2005.
17. M. Ashrafi, D. Taniar, and K. Smith "Redundant Association Rules Reduction Techniques". L ecture Notes in Computer Science, Volume 3S09, 2005, pp. 254 -263 .
18. M. Dimitrijevic, and Z. Bosnjak "Discovering interesting association rules in the web log usage data". Interdisciplinary Journal of Information, Knowledge, and Management, 5, 20 I 0, pp.191 -207.
19. R. Agrawal, H. Mannila, R. Srikant, H. Toivonen, and A. I. Verkamo.: Fast discovery of association rules- In Advances in Knowledge Discovery and Data Mining (1 996).
20. Z. HONG-ZHEN, C. DIAN-HUI, and, Z. DE-CHEN "Association Rule Algorithm Based on Bitmap and Granular Computing". AIML Journal, Volume (5), Issue (3), September, 2005.
21. K. Yun Sing "Mining Non-coincidental Rules without a User Defined Support Threshold". 2009.
22. C. Yin-Ling and F. Ada Wai-Chee "Mining Frequent Itemsets without Support Thres hold: With and without Item Constraints". 2004.
23. Agrawal R, Srikant R (1 994) Fast algorithms for mining association rules. In: Proceedings of the 20thVLDB conference, pp 4S7-499.

Agribot

Smit Agrawal[1], Amit Choudhary[2], Raquib Siddiqui[3], Prof. Prashant Gadakh[4] Dr. Sashikala Mishra[5]

International Institute of Information Technology, Pune[1, 2, 3, 4, 5]

Department of Computer Engineering[1, 2, 3, 4, 5]

smitanilagrawal@gmail.com prashantg@isquareit.edu.in

***Abstract*—In today's world, due to technical advancements in the field of automation and robotics, it is being applied in a various number of areas such as defense, medical field, surveillance systems, industrial and various others. IoT is making use of actuators and sensors and integrating them together along with Internet. The existing system includes making use of man made efforts to keep a check of the field and applying various pesticides. But in case of large fields, this process becomes difficult to manage. Large fields cannot be easily managed by human efforts. The objective of this paper is to reduce human efforts, improve crop yield and assist the farmers by making use of the technological advancements. "Agri-bot" is a robot used for agricultural purposes, this system uses the powerful Raspberry pi for the computational work. By making use of image processing techniques it identifies the dissimilarity between the healthy leaf and the infected leaf. It does this with by making use of KNN algorithm and CNN algorithms. By making qualitative and quantitative analysis we can also predict the total crop yield and find appropriate prices to set for maximizing the profit.**

Keywords— **KNN, CNN, Cropping, Gaussian Filtering, Histogram, Segmentation, R-pi, Variance, Kurtosis.**

I. INTRODUCTION

Man's most vital need is Supplies. For the ever growing population throughout the years, people have struggled to produce large amounts of food. Agriculture is one of the most important and vital factors in the ecnomy of the country. There has been increase in land, improved techniques and technological advances, however, there is also substantial decrees in the number of population engaging in agriculture practices. Agriculture Automation and Robotics provides a solution for these problems. The applications of Automation and Robotics are getting outpaced every day.

To distribute supplies among large population needs proper value of production. Major percentage of population in India lives in rural areas, and is largely dependent on agriculture or agricultural related activities. Keeping a track of health of the plants and crops from the early stage is very necessary. If proper care of the crop is taken even from the early stages, a lot of damage can be controlled if proper care is taken. Plants can be saved from damage if they are well managed like, proper care of soil is taken, good quality seeds are used, appropriate irrigation and harvesting techniques are applied. Another important task is making proper use of Pesticides. A very large number of farmers are unaware about making the appropriate use of pesticides for the crops. Some of the specific crops are prone to certain types of pesticides while some are not, this knowledge is not accessible to all the farmers. This can be harmful not only to crops that are getting sprayed with pesticides, but also the people who consume them.

Disease in plant is usually identified by the farmers by keeping a check on the crops manually. One of the most common practice around the globe is by making use of Sprayers. In the conventional method of threshing, mostly mechanical sprayer are used. Furthermore, farmers are directly exposed to these harmful chemicals without any protective measures. Exposure to these pesticides can cause irritation to nose and skin, further leading to critical diseases.

The proposed system, works on Raspberry pi and automated system which works in several phases such as image acquisition, pre-processing, data analysis and feature extraction. It tries to tackle all the problems occurring in the traditional farming practices. The disease detection can take place on the basis of the colour of the leaf. But for the diseases causing deformity in plant leaf other algorithm needs to be applied.

II. BACKGROUND

- RPi Camera is used to take the digital image of the leaf. The captured images are then processed by applying different techniques, to get useful and necessary features for the purpose of analyzing.
- Capturing the image of the leaf is the first and foremost step in analysis.
- The acquired image is pre-processed to improve the quality of image and remove any noise or distortions from the image. Only the interested part of the image is clipped and image smoothing is done.
- The green pixels are masked from the image and the threshold value is computed. If the computed

value is less than that of the pre computed value, then zero value is assigned to the red, blue or yellow pixels.

- The infected cluster is masked out and the masked cells are removed.
- We then obtain the useful segment to classify the leaf disease.
- Ever image is made up of a number of pixels, and every pixel has red, blue and green components. Every chromosome in the leaf shows a sequence of K cluster center.
- In the first step of computation, the pixels are clustered according to the nearest cluster center.
- In the further computations, new cluster centers are formed based on the mean of each pixel from that Custer center.
- The function can be computed by calculating the Euclidean distance between the pixels and their respective cluster center.

$$dist((x, y), (a, b)) = \sqrt{(x - a)^2 + (y - b)^2}$$

- Next we compute the features using color co-occurrence method. In this method, both the texture and the color of the image is taken into consideration.
- The co-occurrence features are first stored, and then SVM is used to do the classification process.
- Success can be measured by using the classification gain.

$$Gain\ (\%) = No.\ of\ correct\ classification\ /\ Total\ no.\ of\ test\ images\ *\ 100$$

- It gives us the details of the problem that may exist and the possible solution to it like:
  - Inappropriate use of pesticide problems.
  - Low availability of light.
  - Less fertilization.
- In this module, we calculate the approximate quantity and the approximate price of a particular crop.
- By doing this, we can predict future profit and yield.

III. TECHNICAL ASPECTS

- Image Acquisition;

The images of the plant leaf area unit captured through the camera. This image is in RGB (Red, inexperienced and Blue) kind. Semblance transformation structure for the RGB leaf image is made, and then, a device-independent semblance area transformation for the semblance transformation structure is applied.

- Image Pre-processing:

To remove noise in image or alternative object removal, varicolored pre-processing techniques is taken into account. Image clipping i.e. cropping of the leaf image to induce the interested image region. Image smoothing is sunk exploitation the smoothing filter. Image improvement is distributed for increasing the distinction. The RGB pictures area unit changed into the gray pictures exploitation semblance conversion by exploitation the equation,

$$f(x)=0.2989*R + 0.5870*G + 0.114.*B$$

Here, we choose a higher value of the Green pixels so as to capture the greener part of the leaf and it ignores the other aspects. So a smaller or lesser value of Red and Blue pixels is used.

- Image Segmentation:

Image segmentation is a technique of classifying an image into various segments based on the basis of some similarity or dissimilarity. There are various image segmentation techniques like neural network, edge detecting technique, dual clustering, histogram based methods, etc. One of the most widely used techniques is K-Means clustering technique. K-Means is one of the most useful and widely used techniques for a number of reasons, mainly because it is simple to implement and has high computational speed. It also has a large number of applications. It also is unsupervised learning technique. It segments the image into K different segments. The value of K is always positive. The segmentation is done on the basis of similarities between the features. The grouping is done on the basis of distance between the segment steroid and the pixels. In our case, the infected leaf clusters will be formed separately. We can also choose the value of K on runtime. Spot detection will help us to figure out he disease present oin the plant.

- Feature Extraction:

A lot of future extraction techniques are available. In our system, we are taking into consideration a lot of feature such as Skewers, Standard Deviation, Variance, Kurtosis for feature extraction. KNN algorithm will be based on the cluster center only. It might not be very useful for us, in case of Artificial Neural Network, we can train the algorithm much faster and it will be beneficial as it uses back propagation. SVM on the other hand is binary classification algorithm. By making use of SVM we can classify the leaf into the disease it falls.

1) Semblance Co-occurrence Method:

In this step, the captured image is applied some filtering to remove the noise from the image and only the leaf part is extracted. This image is converted from the RGB image to HIS image for feature extraction process.

$$H = \begin{cases} Theta, & if\ B < G \\ 360 - Theta, & if\ B > G \end{cases}$$

$$S = 1 - \frac{3}{(R+B+G)}[\min(R+B+G)]$$

2) Feature extraction from Geometric features:

In traditional feature extraction, only the features are extracted on the basis of leaf color. Digital morphological feature extraction also takes into account any geometric and shape based feature extractions. These features are derived from a variety of features like radius, parameter of the leaf, perimeter and the size of the leaf. With so large number of features, it will be difficult to manage the calculations based on these features. So to counter this, we will make use of Principal Component Analysis. PCA makes simpler calculations and creates a more meaningful dataset. This dataset is now eligible to make simpler feature extractions.

- Classification:

When the process of feature extraction is completed, the trained dataset is classified by making use of artificial neural network. In artificial neural network, the framework will try to learn from the results produced by feature extraction. The classifier will classify the results to each image and give us the desired output.

IV. EXISTING SYSTEM

- The current existing system keeps the track of yield manually or by laborers. They generally might fail to discover issues caused within the crops at terribly early stages, which could injury the yield inflicting major losses to the farmer. Lack of resources is another major issue because of improper usage of pesticides or water.

V. PLANNED SYSTEM

- Setting a platform to remotely management the system that keeps a track of all the crops and yield. The system can gather information and acquire pictures. The gathered pictures are processed and analyzed with reference to the already hold on pictures, to predict an efficient resolution for the malady found within the crop. It'll work on four totally different phases.

1) Information Gathering.

2) Information Analysis.

3) Method information.

4) Predict Effective Solutions

VI. CONCLUSION

Thus we will conclude that, our project on Agriculture exploitation Drones will be extremely economical in terms of Human Laborer and Resource Management.

- Better yield and good quality crops will be achieved by exploitation image process and information analysis.
- We can make early predictions on the quality of crops and detect diseases, if any.
- By making use of information analysis, we will predict the harvest and calculate the costs for the crops for higher profits and sales.

VII. REFERENCE

- Jayaprakash Sethupathy, Veni S, OpenCV Based Identification of Diseases in Plants, International Journal of Engineering and Technology (IJET)
- Gulam Amer, S.M.M. Mudassir, M.A Malik, Design and Operation of Wi-Fi Agribot Integrated System, 2015 International Conference on Industrial Instrumentation and Control (ICIC)
- Akhila Gollakota, M. B. Srinivas, Agribot — A multipurpose agricultural robot, 2011 Annual IEEE India Conference.
- Vijai Singh, A. K. Mishra, Detecton of plant leaf disease using segmentation.
- Dept. of Bioproduction & Machinery, Mie Univ., Japan, Agricultural robots in Japan, EEE International Workshop on Intelligent Robots and Systems, Towards a New Frontier of Applications.
- Shuma Ramesh, P V Vinod, Ramchandra Hebbar, Plant Diseaese Detection Using Machine Learning, 2018 International Conference on Design Innovations for 3Cs Compute Communicate Control

International Journal of Technical Innovation in Modern Engineering & Science (IJTIMES)
Impact Factor: 5.22 (SJIF-2017), e-ISSN: 2455-2585
Volume 4, Issue 12, December-2018

DRIVER BEHAVIOUR ANALYSIS USING ANDROID APP

Preeti Kamble[1], Anant Dargude[2], Shital Chadar[3], Shrutika Mane[4], Prof. Sandeep Patil[5]

[1]*Computer Engineering, Pune university,*
[2]*Computer Engineering, Pune university,*
[3]*Computer Engineering, Pune university,*
[4]*Computer Engineering, Pune university,*
[5]*Computer Engineering, Pune university,*

Abstract— Texting-while-driving (TD) is one of the fatal accident resulting behaviours of drivers. Till date many interesting systems and mobile phone applications have been designed to help to detect or combat TD. However, for a practical TD detection system, one of the key property is its capability to distinguish between driver's mobile phone usage pattern and behaviour from passengers. As of now all the existing solutions are dependent on user's manual input. Also they utilize specific GPS devices to determine whether a mobile phone is at driver's location and in a particular position. In this paper, we propose a method which is able to detect TD automatically without using any extra devices thus a cost efficient method. The idea is very simple: when a user; (specifically Driver) is composing messages or misusing Smartphone while driving apart from regular usage, the Smartphone embedded sensors (i.e. gyroscopes, accelerometers, and GPS) collect the associated information such as touch strokes, holding orientation and vehicle speed. This information will b recorded and then be analyzed and categorised to see whether there exist some specific patterns similar to current recorded one. Extensive experiments were conducted by different persons and in different driving scenarios. The results show that our approach can achieve good detection accuracy with low false positive rate and thus focusing more on passenger's safety on road. This system is infrastructure free and provides high accuracy. Also this method does not access the content of messages and therefore is privacy-preserving.

Keywords— Drunk Driver Detection, Texting while Driving, Road safety and measures.

INTRODUCTION

As per survey performed by the U.S. Department of Transportation, in 2011, at least 23 percent of auto collisions involved cell phones, that equals 1.3 million crashes. Out of which, distracted driving activities associated with cell phones, texting while-driving (TD) has become the top one killer. Hence to cope up with such challenging issue in day to day life, we have proposed a new notion for the benefit of mankind and reduce such mishaps and accidents.The aim of our proposed system is that to build such a software application which automatically detects such TD without any extra device implantation or resource installation. This project is an attempt to focus on a security system that is designed merely to serve the purpose of providing security to Passengers so that they never feel helpless while facing Misbehavior of Driver.

In this system, we are using users Smartphone to collect relevant data when messages are being composed to conclude if there are TD pattern which will make us come to conclusion that vehicle speed is reduced or increased or diverted because of this behavior.

Literature survey

Title - Driver Driver Cell Phone Usage Detection from HOV/HOT NIR Images
Authors: Y. Artan, O. Bulan, R. P. Loce, and P. Paul,

Description:
Distracted driving due to cell phone usage is an increasingly problem in terms of lost lives and property damage. Thus for public safety and security, several state and federal governments have enforced regulations that prohibit driver from using mobile phone usage while driving. Here we have proposed a computer vision based method for determining driver cell phone usage using a near infrared (NIR) camera system directed at the vehicle's front windshield. This method consists of two stages; firstly we localize the driver's face region within the front windshield image using the deformable part model (DPM). Secondly we utilize a local aggregation based image classification technique to classify a region of interest (ROI) around the drivers face to detect the cell phone usage.

Title : Detection of Drivers Mobile Phone Usage
Authors: Hayrullah Yasar, Phd Student, De La Salle.

Description:
Mobile phone usage while driving is dangerous. It may cause traffic accident. Detection and proof of usage should be done by a system. Anti-Distracted Driving Act that became a law last August 1, 2016 will now be enforced and put into

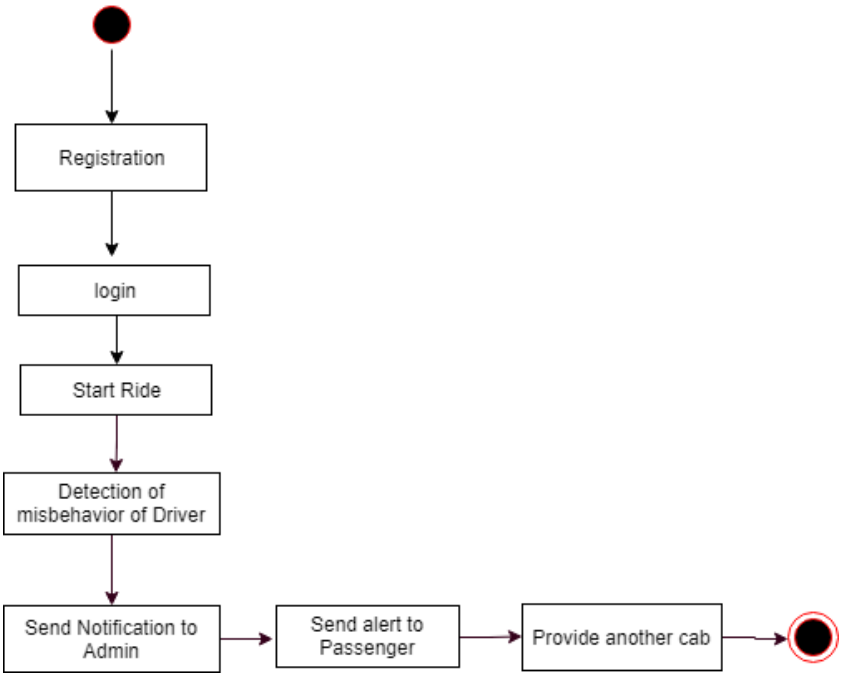
practice by many countries. So drivers may get penalized if they use mobile phone while driving. This paper is intended to develop a neural network application that can detect mobile phone usage. For the system training and testing, sample pictures would be used. Based on this pictures we will train the Cascade Object Detector on MATLAB.

PROPOSED SYSTEM

In proposed work is first of all we are going to developed a Smartphone application. This application will collect data from user's mobile phone and will do a comparative analysis from database to identify similar trends as defined in system. Here we would be recording driver's behaviours and its physical state such as drunk or normal. Secondly this system proposes an algorithm using sensor to check whether the user, particularly driver is using his Smartphone for general navigation purpose as defined by commercial Cab services such as OLA and Uber.

Our application will detect the behavioural usage pattern of drivers mobile. It will check for its state i.e whether it's in landscape mode or portrait. If yes for how much time duration and what purpose its serving while diving. Apart from this it will check, if there is continuous activity in mobile position i.e can be an example of user playing game while driving. This again can lead to distraction from driving and result into mishaps. Thirdly, it would check if driver is texting while driving. This can be recorded if user is using continuous key strokes for texting while driving. Thus our system covers all the drawbacks of existing system and proposes a new and more secure reliable system. This is not only cost effective but also saves human life and improves road driving quality.

FLOW OF THE PROJECT



ADVANTAGES OF PROPOSED SYSTEM

- Reduce road accidents and mishaps which happen due to TD behaviour.
- Alerts the passengers as well as user about Smartphone usage is exceed its maximum limit.
- Detecting behavioural drunk driver patterns thus alerting the passenger and system.
- Cost and resource effective.

MATHEMATICAL MODEL

Let W be the set of whole system which consists of the input, process and output of the system.

W = input, process, output.

Where,
Input = is the set of inputs given to the system to achieve the problem statement.
Process = is the procedure or the algorithm applied to the system which gives the expected output.
Output = is the output of the system.

input = S, U, A, R, P, N, Avg.
Let,

1. S = Drivers activity.

2. U = be the set of users/Drivers.
U = u1, u2, u3, ...un.

3. A = be the set of Miss behavior activities.
A = a1, a2, a3, ...an.

5. P be the process which monitors users behaviour.

Process:

Step 1: User 'Ui' will registered while booking and bording the cab.

Step 2: User 'Si' activities will be starting to monitor once the ride ahs started.

Step 3: User will give get an alert if the drivers behaviour or smart phoer usages is as per predefined pattern in database.

Depending on the user history of particular driver 'Si' system will apply the efficient algorithm to detect the drunk driver pattern and Smartphone usage and the categorize it accordingly to send alert by the system.

Avg = (sum of S) / total number of that U users.

if avg Avg is greater than threshold average value then that ad post is considered as positive category else
it is negative.

Step 4: As per Negative Rating System will Notify to passenger i.e. alert about drivers behaviour.

Step 5: System will record and report against that driver for safety issues.

SYSTEM REQUIREMENTS

HARDWARE REQUIREMENTS:

- System Processors : CoreI7 an compatible
- Speed : 2.4 GHz and above
- Hard Disk : 3 GB and above

SOFTWARE REQUIREMENTS:

- Operating system : 32bit Windows 7 and on words
- Coding Language: : Java J2EE\Android
- IDE : Eclipse Kepler
- Database : XAMP Server

CONCLUSIONS

In this paper, we have explained how the influence of utilization of Smartphone and its adverse effects which may have severe impacts such as loss of human life's. We have proposed a novel and cost effective method which can be easily implemented to to detect TD(Texting-while-Driving) without using any extra devices. The system is designed in such way that CAB Organization automatically get drivers activity while ride is going on. Using inbuilt sensors Detection system get current running application status with respective sensors. In this project system generate alert message with device details to Organization. And every week CAB Department generates weekly report for drivers behaviors. If driver's activity always gives negative feedback then Service providers will take action against him.

REFERENCES

[1] Y. Artan, O. Bulan, R. P. Loce, and P. Paul, Driver cell phone usage detection from hov/hot nir images, in 2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops, June 2014, pp. 225230.

[2] D. Wang, M. Pei, and L. Zhu, Detecting driver use of mobile phone based on incar camera, in 2014 Tenth International Conference on Computational Intelligence and Security, Nov 2014, pp. 148151.

[3] P. Viola and M. Jones, Rapid object detection using a boosted cascade of simple features, in Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001, vol. 1, 2001, pp. I511I518 vol.1.

[4] Y. Wang, J. Yang, H. Liu, Y. Chen, M. Gruteser, and R. P. Martin. Sensing vehicle dynamics for determining driver phone use. 2013.

International Journal of Research in Advent Technology, Vol.7, No.2, February 2019
E-ISSN: 2321-9637
Available online at www.ijrat.org

Diabetic Fighter Diet Plan Android App

Yogesh Ghule, Swapnil Bobe, Pratik Agrawal, Himanshu Deshpande, Prof. Ashwini Jarali
International Institute of Information Technology, Pune
Yogeshghule2971@gmail.com

Abstract: Intake of a good diet is very important to fight any disease and the most common diseases now a day is diabetes. We observed in our environment and we found out many people don't know what that they should follow in their diet to fight their diabetes and how to maintain healthy living with the healthy diet. App will take information from the user about food habits and their current Health information. After taking necessary information from the user app will process the data and will create a diet plan accordingly for a diabetic patient which can fight diabetes.

Keywords: diabetes, Decision Tree.

1. INTRODUCTION

The disease diabetes increasing rapidly and many people are not able to maintain a healthy diet to fight with diabetes. Nowadays everybody is using a mobile phone and thus we decided to make a app which can provide a healthy diet for a diabetic patient. Most of the people and having habit of using mobile app to track each and every activity and set reminders this app for that will be of great used for the people who are having this app for that will be of great use for the people who are having diabetes. Current users health data will be asked by the app for the first time of login according to the input diet plan will be created by the app. This app will smartly manage users diabetes with the help of diet plan if users sugar level is high then this app
will provide you such a diet which will help you in controlling users sugar level. App will ask input from user for his current food habits and we are not going to drastically changed the diet because its impossible for a person to adopt a truly new diet from the current diet which he is following, so we have decided to customize the diet slowly in the way such that the person can adopt to the diet slowly.

2. EXISTING SYSTEM

You might have seen many bad apps for healthy living and fitness on App Store and there are many apps for diabetes as well but every app is a static app. Have you ever imagine that they are providing us same diet plan to each and every person who is using that app how its possible every persons body is different and everybody should be provided with a personalized that according to his requirement and here comes the solution of the diabetic app which we are creating app is a dynamic app which will create that according to the persons input data which is his health information and his current food habit. We are not claiming that our app is totally different from the apps which are available in the market but our app is unique when you compare it to some other apps in the market. Because our app will provide you dynamic diet plan and other apps are providing users static diet plan, a static diet plan is same for each and every user

3. LITERATURE SURVEY

The diabetic data sets which are available online and not fulfilled data sets and are having many impurities. with the help of data smoothing techniques we are removing the impurities in the data set which can be used for prediction and analysis of diabetes. Thus the raw data will be turned into informative data with the help of data smoothing technique.[1].

In this study, they uses F-score feature to identify the accuracy of predicted values. F-score gives us the minimal clustering error. The correctly classified instance finds the pattern for diagnosis and are used for further classification process. The improved performance of the in terms of Accuracy of the classifier, Sensitivity proves that the proposed feature approach indeed improves the performance of classification. [2].

In this study, decision tree algorithm is used to predict diabetes. They implement algorithm in two phases. In the first phase they performed data preprocessing including attribute identification and selection, handling missing values, and numerical discretization and In the second phase they performed a diabetes prediction model construction using the decision tree method.[3].

Diabetes self-management app for android smartphones to manage the diabetic dietary. Also manage weight, blood glucose self monitoring and

promoting behavior changes. It will suggest the diet plan for diabetic patient and also maintain the dietary habit of diabetic patients. [4].

The use of mobile phone is increase exponentially in the 21st century. For each and every aspect of tracking mobile phone is used in 21st century. For various health monitoring and fitness tracking many application have been introduced for android and ios platform. As diabetic is a majorconcerned with thought of developing app which can track your diabetic condition and provide result in percentage. Various inputs like current food intake, blood sugar and BMI. The minimum the percentage better is the health. It develops a competition within to fight and overcome the previous result.[5].

4. PROPOSE SYSTEM

Our app will take users data input for analysis and prediction purpose. The most important uses data like health information will be asked and secondly the current diet plan According to the input given by the user and by understanding the current diet habits of the user, app will be preparing a custom diet plan for the user We understand that every individual has different body type and different food habits the user to user that plan will be different and it will be customize only for particular user none of the diet plan of two person will be same

Mathematical Model:

- INPUT :

I : { I1, I2, I3 }
I1 : User Personal Information.
I2 : User Health Information.
I3 : ToDoList.

- FUNCTION :

F : { F1, F2, F3, F4, F5, F6, F7 }
F1 : UserSignup (Email, Password, CPassword, Mob_No) / UserLogin (Email, Password)
F2 : UserHealthInfo (Age, BMI, Insulin, BP)
F3 : ToDoList ()
F4: StoreData ()

F5 : CheckSugerLevel ()
F6 : MakeDietPlan ()
F7 : DisplayDiet ()

- OUTPUT :

O : { O1, O2 }
O1 : Low / Medium / High Sugar Level
O2 : Diet Plan For Low / High Sugar Level.

- CONDITION :

C : { C1, C2, C3}

C1(O1= 'LOW') : F7(LOW).
C2(O1= 'MEDIUM') : F7(MEDIUM).

C3(O1= 'HIGH') : F7(HIGH).

Architecture Diagram:

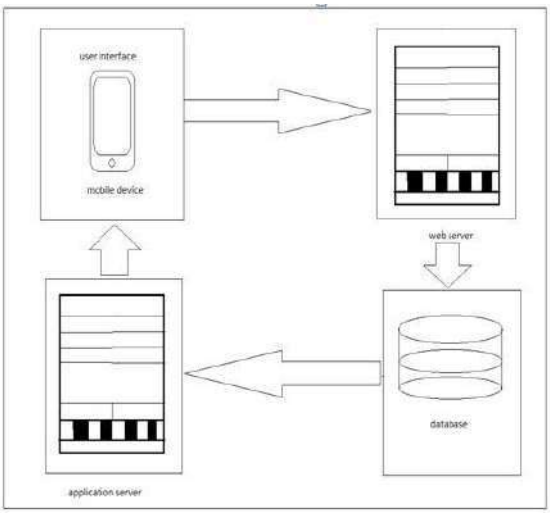


Figure 2 - Architecture Diagram

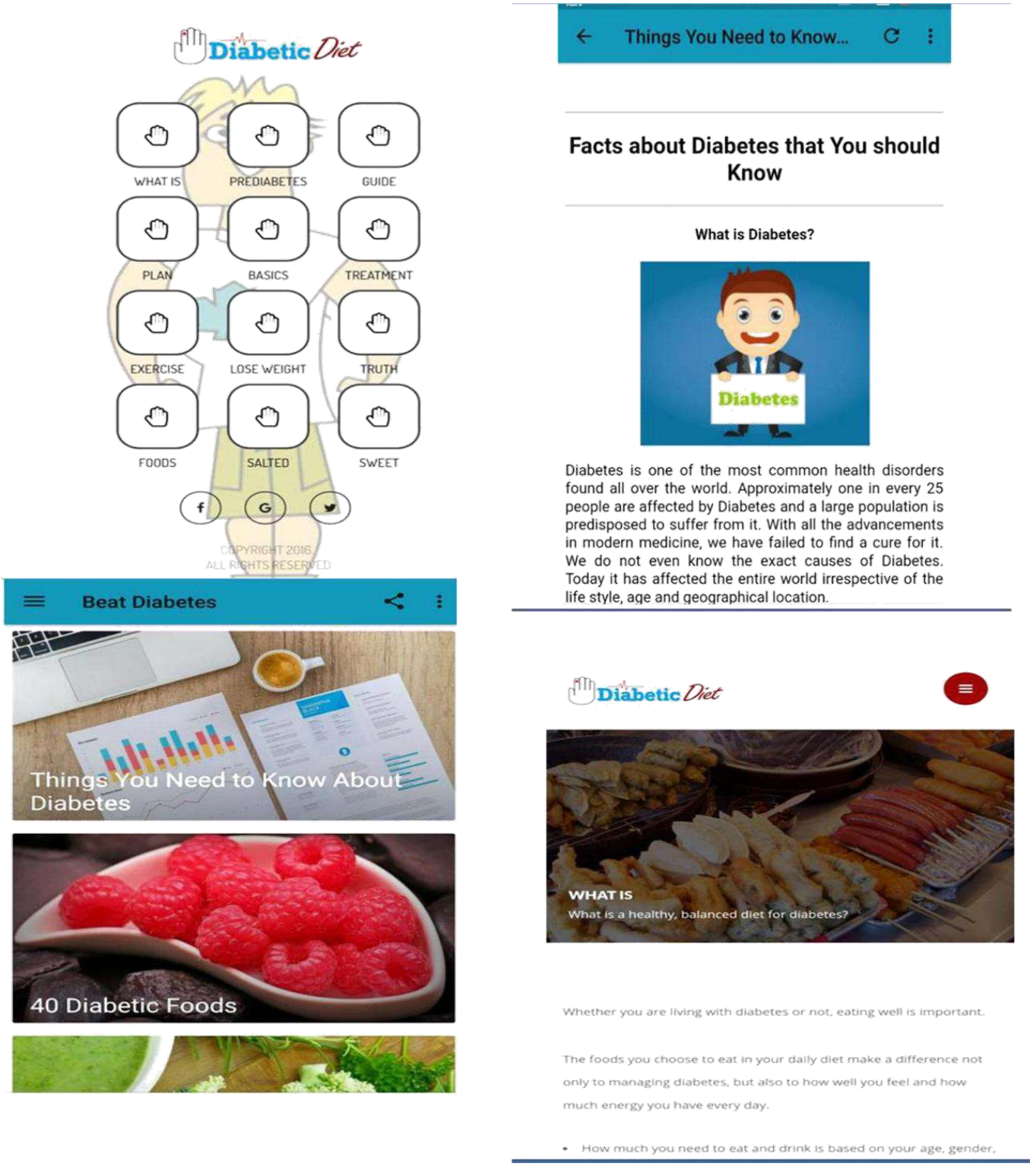


Figure 1 - Existing system

5. CONCLUSION

Our app will be to help for senior citizens for home managing a diet is difficult. Hence, our Android app will Help users smartly organize their daily diet to fight the diabetes and our main aim is to fight diabetes with the help of healthy diet and not with the help of medicines or any external

6. FUTURE SCOPE

In future we will be adding home delivery of good diet food. Which is recommended to the users by the app most of the times we have observed that our senior citizens are not capable of going to the market and buying good diet food and best to make it easier we will be adding a basic online card feature from which our users can order food recommended by our diet chart online and it will be delivered to their homes.

REFERENCES

[1] "A Computational Approach of Data Smoothening and Prediction of Diabetes Dataset"https://dl.acm.org/citation.cfm?id=2791572August10 - 13, 2015
[2] " A Hybrid Prediction Model with F-score Feature Selection for Type II Diabetes Databases"https://dl.acm.org/citation.cfm?id=1858391September 16 - 17, 2010 -
[3] "Decision Tree Discovery for the Diagnosis of Type II Diabetes"https://ieeexplore.ieee.org/abstract/document/5893838 25-27 April 2011-
[4] "An evaluation of diabetes targeted apps for Android smartphone in relation to behavior change techniques"http://eprints.whiterose.ac.uk/104077/3/Full%20manuscript_second%20revision.pdfMarch, 2007
[5] "Content Analysis of Mobile Health Application on Diabetes Mellitus"https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5711785/
[6] "Clinical prediction models for diabetes prediction"https://link.springer.com/article/10.1186/s40537-017-0082-7
[7] " Mobile Health Applications on Diabetes Mellitus"https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5711785/
[8]" Blood glucose level prediction"https://ieeexplore.ieee.org/abstract/document/7591358/

International Journal of Research in Advent Technology, Vol.7, No.2, February 2019
E-ISSN: 2321-9637
Available online at www.ijrat.org

Novel Methodology for Crop Disease Detection

Asit Dhaware[1], Akash Fulmali[2], Sahil Hatankar[3], Dr. Sashikala Mishra[4], Prof. Prashant Gadakh[5]
International Institute of Information Technology, Pune[1, 2, 3, 4, 5]
Department of Computer Engineering[1, 2, 3, 4, 5]

Abstract- Our country's economy is something, which is highly depends on the agricultural productivity. Due to this reason the disease detection in agricultural field has become one of the most important factors as it plays an important role for successful cultivation .Even though the technology has emerged periodically, due to some reasons traditional methodologies are being used. Due to usage of traditional methodologies ,the plant diseases are wrongly identified which leads to huge loss of time, money, productivity, quality, and quantity of the cultivated crops .Now a days, expertise use manual prediction methodology ,but for this large number of experts are required and it also requires continuous monitoring of plants, due to which its cost is very high for a middle class farmer .Generally, by observing the symptoms on the leaves of a plant, we can predict the diseases of a plant.

Keywords: Neural networks, CNN, Machine learning, Crop diseases, Disease detection, Diagnosis.

1. INTRODUCTION

Worldwide, crop diseases cause major production loss and economical loss in agricultural industry. For agriculture to sustain, it is impossible for a farmer to monitor the health of plants or trees and to detect their diseases. In the current time, there are no sensors available for real-time assessment of health situation in trees on commercial basis. Diagnosis of the diseases based on the detection of symptoms occurred earlier, has been a usual threshold taken into account for integrated pest management strategies. Earlier, sanitary treatment with regard to pests and pathogens has been a great procedure to minimize crops yield losses and to increase the efficiency and efficacy of the treatments. However, there are complications in early identification, which delays the application of the appropriate corrective actions due to the possibility of new diseases associated with new resistant crop variants. Earlier, detection of diseases among farmers and technicians can be leveraged by the use of image-based automated identification systems, but under real field conditions they perform poorly using smart phone devices [7].

By observable patterns of specific plant, the studies of plant or fruit can be determined. But, it is very difficult to detect disease and monitor health of the plant. Pesticides, fungicides and chemical applications, using such proper management strategies, one can facilitates the control of diseases which interns improved quality [8].In recent years, with the development of computational systems, and in particular Machine Learning-related Artificial Intelligence applications, Graphical Processing Units (GPUs) embedded processors, IOTs, Cloud Computing have achieved exponential growth in technology, leading to the development of application models and novel methodologies [9].

2. LITERATURE SURVEY

In March 2011, Authors Heba Al-Hiary, Mohamad Hashem experimentally proposed and provided a software solution for fast and more accurate detection and classification of plant diseases. After the segmentation, successively two steps were added. The identification of mostly green pixels are done in the first step. Based on the specific threshold values, these pixels are masked in second step. By using Otsu's method these threshold values are computed, then those mostly green pixels are masked. Complete removal of pixels with zeros red, green and blue values and the pixels on the boundaries of infected object was the another addition step. The proposed technique is a one of the robust technique for the detection of plant leaves diseases, was demonstrated by the results of this experiment. The developed algorithm's efficiency would successfully detect and classify the examined diseases with a precision between 83% and 94%[1].

In 30 may 2014, Author Emanuel Cortes proposed a paper "Plant Disease Classification Using Convolution Networks and Generative Adversarial Networks". As per author, due to the growth of Smartphone technology throughout the globe, it has now become technically feasible to leverage image processing techniques to identity type of plant disease by capturing a simple photo. To reduce the effects of crop diseases on food supply, quicker interventions are implemented that can be achieve by identifying disease. A public dataset of 86,147 images of diseased and healthy plants are used, a deep convolution network and semi supervised methods

are trained for classification of crop species and disease status of 57 different classes[2].

In 29 January 2016, Authors Jayme Garcia, Arnal Barbedo proposed a review on main challenges in automatic plant diseases identification based on visible range images. When system tried to identify disease using simple image, many problems had occurred during this. In this paper, the problems or challenges that faced by system are given. Extrinsic factors and Intrinsic factors are specified. Image background and image captured conditions were the extrinsic factors addressed. Intrinsic factor contains system segmentation, multiple simultaneous disorders and different disorders with similar symptoms. Real time operation, computational power etc. are the other challenges that were identified. The problems and their possible solutions were also given[3].

In 29 May 2016, Authors Srdjan Sladojevic, Marko Arsenovic, Andras Anderla, Dubravko Culibrk, and Darko Stefanovic proposed a software solution for leaf image classification using deep neural networks based recognition of plant diseases. Impressive results were achieved in the field of image classification by the convolution neural networks (CNNs). Based on leaf image classification, there has been development of plant disease recognition model, by the use of deep convolution networks which is concerned in this paper which has been a new approach. This developed model proposed by author can recognize different types of plant diseases out of healthy leaves, and system can easily remove noise. This is first time that CNN algorithm is used for plant disease identification. All necessary step which are required for developing model such as gathering images, creating database, training and testing models, etc. are described in this paper. "Caffe" (a deep learning framework) used for the deep CNN training. The developed model achieved accuracy between 91% and 98%, when separate class tests carried then model gave average 96.3% accuracy [5].

A model using deep learning for image-based plant disease detection was proposed by the Authors Sharada P. Mohanty, David P. Hughes and Marcel Salathe in 22 September 2016. Plant disease detection using deep learning was the primary motive. Plant village dataset is used for training purpose. A dataset contained 54,306 images 14 crop species and 26 diseases were specified. There are two ways for building a model, transfer learning and training from scratch. For training purpose, AlexNet and GoogleNet deep learning architecture choosen. 5 different training and testing set distribution used; wherever Train: 80% and Test:20% give high accuracy. According to result, Transfer learning way and GoogleNet architecture is highly efficient. An accuracy of 99.359% on a test set was achieved by the trained model. On field testing model accuracy is reduced to just above 31%[4].

In March 2017, Author Suja radha proposed a solution for detection of leaf disease using image processing. In agricultural field, disease identification of plants has always been difficult. In this paper, they used image processing in MATLAB for identification of the disease. It included the following steps: 1) Loading the images, 2) Contrast enhancement, 3) Converting RGB to HSV, 4) Extracting of features and SVM. In this project, for feature extraction they used RGB, HSV, YIQ and Dithered Images. They used cotton leaf for experiment. For segmentation K-means clustering algorithm is used and SVM is used for classification[6].

In 21 april 2017, Authors Alexander Johannes, Artzai Picon, Aitor Alvarez-Gila, Jone Echazarra, Sergio Rodriguez-Vaamonde, Ana Díez Navajas proposed a paper "Automatic plant disease diagnosis using mobile capture devices applied on a wheat use case". Unlike other papers, this paper tries to focus on mainly early disease detection and multiple disease detection. Dataset were created using 7 mobile devices in field condition crop. The wheat is considered and mainly 3 disease septoria, rust and tan_spot, where more than 3500 images captured. In algorithm system extracted hotspot candidate, then analysed each hotspot in detail by local descriptors that extract and categorised each region to check against different disease using classifier. Obtained results in field condition higher than 0.30 for all the analysed diseases[7].

In Jan-2018, Authors Dixit Ekta Gajanan, Gavit Gayatri Shankar, Gode Vidya Keshav proposed a technique for plant disease identification in android based system using "Feature Extraction". For the recognition of plant diseases in the early stage, we can use intelligent systems instead of assigning professional agriculture engineers. The purpose of system is Supervising the diseases on crop and suggesting solution forgood productivity. The recognition of a disease can often be based on symptoms like lesions or spots in various parts of a plant. The color, area and the number of these spots on different parts of plant can determine to a great extent, the disease by which plant has been affected. If necessary higher cost molecular analyses and tests can be followed. This application can easily be extended for different plant diseases and different smart phone platforms [8].

In 12 Feb 2018, Authors Konstantinos P. ferentinos proposed a solution for plant disease detection and diagnosis using deep learning models. For this purpose, they used healthy and disease leaf images where total

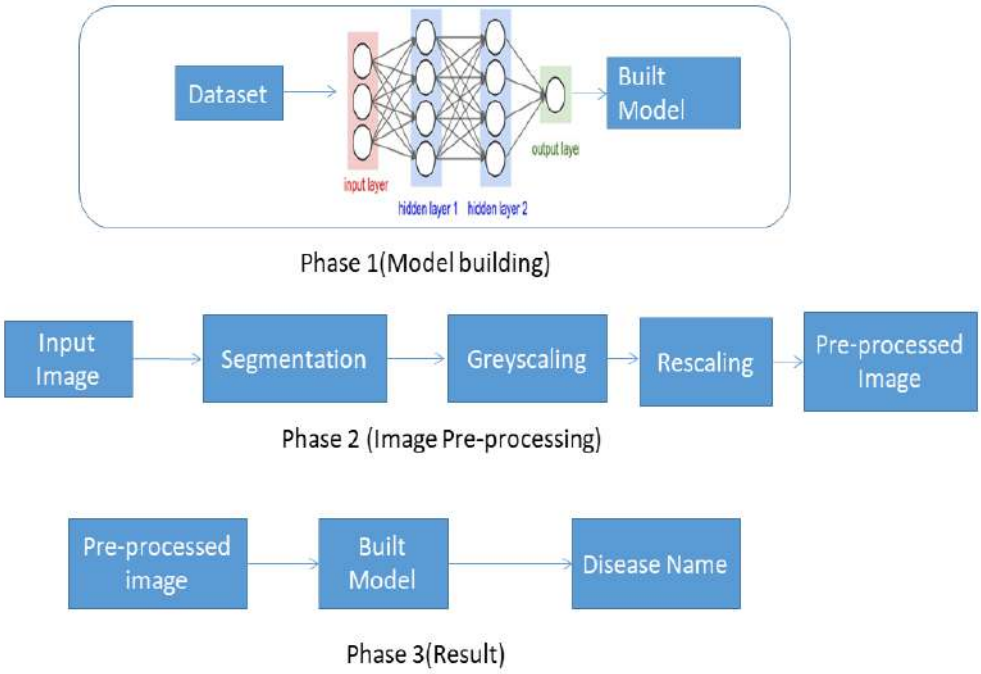
87,848 images are used for 25 different plants disease identification. Best result is 99.53% with CNN (VGG Model). Based on the edges and lines in image, CNN algorithm can easily found disease. Plant diseases were identified using plantvillage database of leaves images of 14 different plants. As per paper, results were quite accurate using CNN. But there has been one drawback, that the images were solely from datasets, and no real photos were taken from cultivation field. They used 80% of images for training and 20% images for testing. Among all CNN models, VGG Model and AlexnetQWTBn architecture gave best success rate. Training time: 5.5 days [9].

In 22 April 2018, Authors, Andreas Kamilaris, Francesc X. Prenafeta-Boldú surveyed on different technique and proposed a paper on it.With promising results and large potential, deep learning constitutes a recent modern technique for image processing and data analysis. After successfully applying deep learning in various domains, now it has entered into the domain of agriculture. A survey of 40 researches has given. The different data pre-processing techniques were examined. Comparison has been made between "Deep learning" and other existing popular techniques (with respect to differences in classification or regression). Results of the comparison clearly indicated that, deep learning provides high accuracy, outperforming existing commonly used image processing technique. Disadvantages and limitations of deep learning methods are specified in the paper [10].

Table 1: Conclusion Table

Sr. No.	Published Year	Techniques/Algorithm used	Accuracy Percentage
1	March 2011	Otsu's Method & Neural Network	83-93%
2	May 2014	Generative Adversarial Networks	91.42%
3	Jan 2016	Challenges in automatic plant diseases identification	-
4	May 2016	Caffe Framework & CNN algorithm.	96.3%
5	Sept 2016	Googlenet	99.357%
6	March 2017	K-means clustering & SVM classifier	98.45%
7	April 2017	Hotspot extraction(k means clustering) & Navie Bayes classifier	86.67%
8	Jan 2018	Image Processing	90-95%
9	Feb 2018	VGG & AlexnetQWTBn	99.53%
10	April 2018	Deep Learning Techniques	-

3. PROPOSED SYSTEM



Steps in Proposed System:
Step 1: Importing dataset. Dataset is taken from Plantvillage Dataset.
Step 2: Training the dataset to build a model. Approximately 80% of dataset is provided for testing.
Step 3: Taking the input image from testing dataset which is approximately 20%.
Step 4: Algorithms:
AlexNet, resNet
Image processing:
a. Segmentation
b. Greyscale
c. Rescale
Step 5: Pre-processed image.
Step 6: Feeding pre-processed image to "Built Model".
Step 7: We will get results in terms of disease name.

Proposed system in brief:
First of all the most important thing was the data source, based on which we were going to implement the application. The dataset has been taken from an open source website called "PlantVillage.com", which provides different datasets of different crops. Among all other datasets, tomato crop image dataset has been selected as it is one of the common crops .
Secondly, conventional neural network (ConvNets or CNNs) was selected for image recognition and image classification. CNN is a class of deep neural network, which is most commonly applied for the usage of visual imagery. CNN uses different multilayer for feature selection. CNN takes input as an image, processes it and classify it according to the given class labels.
Inside CNN, each input image of training and testing dataset will go through a series of convolution layer with filter, pooling, etc. We are using "Softmax" as an activation function in CNN. Inside the model different image pre-processing processes are happening which will minimize the further processing, will improve the image data and eliminate the unwanted noise and distortions. We are using Segmentation, grey-scaling and rescaling for the image pre-processing.
In segmentation, we partition the image into different regions, the main part of an image which is to be considered is taken into account, and the unwanted background is eliminated using techniques such as edge detection, region based detection. In grey-scaling, the

segmented image is converted into grayscaled image, by converting into shades between white and black. Computers don't understand the image, but it understands in the terms of pixels. In rescaling, the greyscaled image is converted into arrays of pixels, and algorithms processes on such arrays of pixels. Image pre-processing is done in both training and testing phase.

After the model is built, we take the real-life image of a tomato crop and again on the leaf, several processes are done and we get output in the form of text format i.e. the disease name and the primitive steps to avoid/prevent the infected crop from disease.

4. CONCLUSION

Hence we concluded that among all techniques proposed by the authors in above papers, deep neural network gives very high accuracy. We will be implementing AlexNet and resNet algorithms which give more accurate results. On lab condition cultivated plant, but fail on real condition cultivated crop. So dataset should need more field images. For image preprocessing, we will be doing some basic steps like segmentation, greyscale and rescaling. In a result, we provide disease name and diagonsis of disease.

REFERENCES

[1] Heba AlHiary, Mohamad Hashem, "Fast and Accurate Detection and Classification of Plant Diseases", March 2011.

[2] Emanuel Cortes, "Plant Disease Classification using Convolutional Networks and Generative Adversarial Networks", 30 May 2014.

[3] Jayme Garcia, Arnal Barbedo, "A review on the main challenges in automatic plant disease identification based on visible range images", 29 Jan 2016.

[4] Srdjan Sladojevic, Marko Arsenovic, Andras Anderla, Dubravko Culibrk, Darko Stefanovic, "Deep Neural Networks based Recognition of Plant Diseases by Leaf Image Classification", 29 May 2016.

[5] Sharada P. Mohanty, David P. Hughes, Marcel Salathe, "Using Deep Learning for Image based Plant Disease Detection", 22 Sept 2016.

[6] Suja Radha, "Leaf Disease Detection Using Image Processing", March 2017.

[7] Alexander Johannes, Artzai Picon, Aitor Alvarez-Gilla, Jone Echazarra, Sergio Rodriguez-Vaamonde, Ana Diez Navajas, "Automatic Plant Disease Diagnosis using Mobile Capture Devices Applied on a Wheat Use Case", 21 April 2017.

[8] Dixit Ekata Gajanan, Gavit Gayatri Shankar, Gode Vidya Keshav, "Android Based Plant Disease Identification System Using Feature Extraction Technique", Jan 2018.

[9] Konstantinos P. Ferentinos, "Deep Learning Medels for Plant Disease Detection and Diagnosis", 12 Feb 2018.

[10] Andeas Kamilaris,Francesc X. Prenafeta-Boldu, "Deep Learning in Agriculture: A Survey" 22 April 2018.

[11] H.M.Deshmukh, Jadhav Sanjivani, Lohar Utkarsha, Bhagat Madhuri, Salunke Shubhangi, "Plant Leaf Disease Identification System for Android", March 2016.

[12] M. Akila, P. Deepan, "Detection and Classification of Plant Leaf Diseases by using Deep Learning Algorithm", Jan 2018.

[13] Pramod S. Landge, Suhil A. Patil, Dhanashree S. Khot, Omkar D. Otari, Utkarsha G. Malavkar, "Automatic Detection and Classification of Plant Disease through Image Processing", July 2013.

[14] Jihen Amara, Bassem Bouaziz, Alsayed Algergawy, "A Deep Learning-based Approach for Banana Leaf Diseases Classification", March 2017.

[15] Kamlesh Golhani, Siva K. Balasundram, ganesan Vadamalai, Biswajeet Pradhan, "A review of neural networks in plant disease detection using hyperspectral data", April 2007.

[16] Savita N. Ghaiwat, Parul Arora, "Detection and Classification of Plant Leaf Diseases Using Image processing Techniques: A review", Feb 2014.

[17] Geng Ying, Li Miao, Yuan Yuan, Hu Zelin, "A Study on the Method of Image Pre-Processing for Recognition of Crop Diseases", March 2009.

[18] Jiazhi Pan, Min Huang, Yong He, "Crop and Weed Image Recognition by Morphological Operations and ANN Model", July 2007.

[19] G.Anthonys, N. Wickramarachchi, "An Image Recognition System for Crop Disease Identification of Paddy fields in Sri Lanka", Sept 2009.

[20] Jose Bermudez Castro, Raul Queiroz, Laura Cue La Rosa, Pedro Achanccarray Diaz, Ieda Sanches, "A Comparative Analysis of Deep Learning Techniques for Sub-tropical Crop Types Recognition from Multitemporal Optical/SAR Image Sequences", June 2017.

GPS Based Encapsulation of Car Keyprint using On Board Diagnostics

Saif Lakhani

Department of Computer Engineering International Institute of Information Technology Pune, India

Shubham Kokate

Department of Computer Engineering International Institute of Information Technology Pune, India

Niharika Sachdev

Department of Computer Engineering International Institute of Information Technology Pune, India

Pranav Patil

Department of Computer Engineering International Institute of Information Technology Pune, India

Prof. Deptii Chaudhari Department of Computer Engineering International Institute of Information Technology Pune, India

Abstract—The ever rising fuel prices is a huge concern for motor vehicle owners.. In the context of Indian market, driving the vehicles with increased fuel efficiency can help the vehicle owners adapt to these fuel price variations. In this work we propose a novel approach to help drivers monitor the fuel efficiency of their vehicles and provide driver finger printing.. In the context of Indian market, driving the vehicles with increased fuel efficiency can help the vehicle owners adapt to these fuel price variations. In this work we propose a novel approach to help drivers monitor the fuel efficiency of their vehicles and provide driver finger printing.All cars 2010 and later contain the OBD port and it presents a valid opportunity for car manufacturers, insurance companies and users alike. Notwithstanding its advan- tages already outlined, this also provides analysis of fault codes from the car, for additional information and diagnosis.

I. Introduction

OBD is a computer-based system originally designed to reduce emissions by monitoring the performance of major engine components.The data from the on-board sensors from each car is aggregated using ELM327.A basic OBD system consists of an ECU (Electronic Control Unit), which uses input from various sensors to control the actuators to get the desired performance. A modern vehicle can support hundreds of parameters .The ELM327 is a programmed microcontroller produced by ELM Electronics for translating theon-board diagnostics(OBD) interface. Its command protocol is one of the most popular PC-to-OBD interface standards and is also implemented by other vendors. The original module is imple- mented on the PIC18F2480 microcontroller from Microchip Technology. It is one of a family of OBD translators from ELM Electronics. Other variants implement only a subset of the OBD protocols. To collect this data torque app is used which can use the GPS to provide tracker logs with OBD engine logging so you can see what you were doing at any point in time. The paper presents opportunities to apply various algorithms on this data to provide an accurate representation of a driver, his style and his efficiency with the end goal of improving these parameters. This review paper analyses the different classification and regression algorithms used for this, their effectiveness and their feasibility.

II. Objectives

The main objective of this work is to allow drivers to drive more efficiently by providing them information about their driving patterns.To develop a method to rate the drivers and provide feedback for increasing fuel efficiency we use Kalman filters and Adaboost algorithms. For driver identification and finger printing we use classification and regression techniques with Hidden Markov model. Also, to rate the drivers behavior we use continuous regression to identify distraction of the driver. The aims and objectives of this work are :

A. To allow drivers to drive more efficiently

B. To fingerprint the driver

C. To rate driver behaviour

III. Related Work

Today, the traditional approach to reduce accidents due to driver distractions majorly consists of introducing laws to prohibit use of cell phones and text messaging while driving.In addition, National Highway Authority of India (NHAI) estimates that 1.35 lakh of all motor vehicle crashes are caused by driver distraction. However, as the number and complexity of on-board information and entertainment systems increase, these restriction laws cannot stop drivers from performing secondary tasks.

Data analysis showed that the average vehicle speed is lower under distracted driving, compared to neutral driving. Also, distracted driving has a wider, counterbalanced short-term variance than non-distracted (neutral) driving. There are six long-term driving behaviors: Turn Left, Turn Right, Lane Change Right, Lane Change Left, Stop, and neutral driving. In the paper published by SangJo Choi and et. al [3], driver action classification, driver distraction detection, and driver identification were studied using on-board CAN-Bus signals with both GMM and HMM frameworks. Here, the authors employee two statistical modeling frameworks to model driver behavior. GMMs are used to model driver characteristics. HMM using 3 states yields the best performance for both driving durations.

The figure below compares vehicle speed of neutral driving and driving while interacting with a voice portal (distracted driving) using the same vehicle route twice.

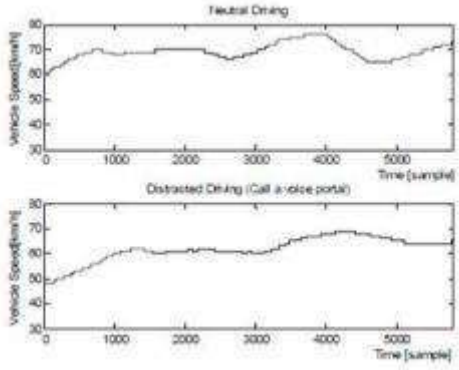


Fig. 1. Comparison of vehicle speed for neutral driving and under distraction [3].

The second plot shows the accuracy of the most efficient HMM (with 3 states).

The section is devoted to driver modeling based on GMM and HMM with applications on action classification, distraction detection, and driver identification. The neutralized short-term variance of steering degree under the neutral driving condition is 1.21, compared to 1.69 for the distracted driving condition. The next experiment is distraction detection. Total six driver models were trained. The average speed varies among drivers, with an average speed of 41.08 km/h during neutral driving.

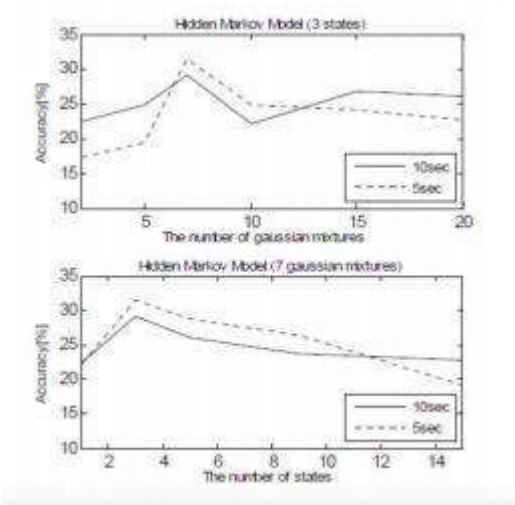


Fig. 2. Accuracy of driver identification [3].

There are five activities considered as distraction tasks amid driving: calling a voice portal, controlling radio, controlling the window, talking with an assistant, and performing some common Tasks. On another route, the average vehicle speed of distracted driving, calling a voice portal was 64.55 km/h, compared to the average vehicle speed of 68.52 km/h under neutral driving. Under the neutral driving condition, the average vehicle speed was 48 km/h and under distraction talk (driver controlled the windows while driving), the average vehicle speed was 41 km/h. That is, the longer driving data test shows more consistent characteristics of distraction. The average speed is slower when he/she drives with some distraction. The average accuracy is 69%. Gaussian mixtures are known for their ability to generate arbitrarily shaped densities. For driver identification, driving signals are also used to generate models to identify drivers.

Figure below plots the steering degree of neutral driving and distracted driving (having conversation with passenger).

The neutralized short-term variance of the steering degree under neutral driving is 0.27, and 0.82 under distracted driv- ing (with an analyzed window length of 300 samples). By analyzing the signals of each driver, the authors observed that both neutral driving and distracted driving of each driver have its own characteristics, they did not consider the differences between neutral driving and distracted driving to classify driver identity. Therefore, the authors designed a multi-modal data transcription framework and transcribed data based on driver activities (e.g., making a left turn, talking with an assistant) according to the reference labels in Table.

Identifying driving styles of drivers using in-vehicle sensor data is an interesting research problem and an must have real-world requirement for automotive industries. A good

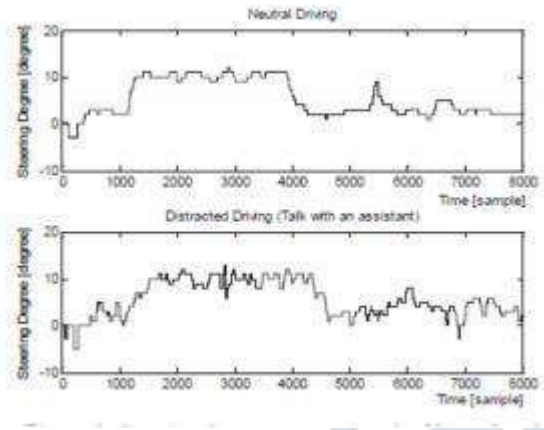


Fig. 3. Steering degree versus time for Neutral and Distracted driving (conversation with passenger) along same route conditions. [3]

Tasks		Start point		End Point	
Code	Description	Data type	Action	Data type	Action
TL	Turn Left	S	W	D	D>0
TR	Turn Right	D	W	D	D>0
LR	Lane Change Right	Video	A	Video	B
LL	Lane Change Left	Video	A	Video	B
ST	Stop	S	S=0	S	S>0
CT	Call a voice portal	Audio	X	Video	Y
CR	Control Radio	Audio	X	Video	Y
CW	Control window	Audio	X	Video	Y
TA	Talk with an assistant	Audio	X	Audio	Y
CM	Common Task	Audio	X	Video	Y
FR	Free-style Driving	-	O.W	-	O.W

A: When driver starts to glance at the rear mirror
B: When vehicle becomes parallel to the lane
S: Vehicle speed
D: Steering Degree
W: When steering degree starts to increase
X: When instruction is given
Y: When driver finishes action
O.W: Otherwise

representation of driving features can be extremely valuable for anti-theft, auto insurance, autonomous driving, and many other application scenarios. We address the problem of driver identification using real driving datasets consisting of measure- ments taken from in-vehicle sensors. The paper unfolds the minimum learning and classification times that are required to achieve a desired identification performance. Further, feature selection is carried out to extract the most relevant features for driver identification. Finally, in addition to driving pattern related features, driver related features (e.g., heart-rate) are shown to further improve the identification performance.

Meng and et.al. [7] proposed a Hidden Markov Models (HMM) method, coupled with an HMM-based similarity mea- sure, using mainly three features: acceleration, brake, and steering wheel data. The authors have shown that driver identification can be accomplished at rates ranging from 30 to 70

The authors conducted experiments comparing brake and turning signals from two different drivers using K-means and Support vector machine (SVM) algorithms. Another effort in drivers differentiation was performed by Zhang and et.al. who used HMM to analyze the data of the accelerator and steering wheel of each driver, and achieved an accuracy of 85

Analysis and methodology followed by Meng and et.al. [7] for driver fingerprinting is also studied in the proposed paper. A higher accuracy of driver identification will likely require multiple driving parameters and a larger learning time. In this paper, the author investigated the relationship between accuracy, the number of features and the learning time with the objective to optimize the driver fingerprinting task. The process of driver fingerprinting consists of preprocessing the driving datasets, selecting the most relevant features thereafter developing appropriate classification models using machine learning. The data collection was carried out in South Korea using a recent model of KIA Motors Corporation. Ten drivers participated in the experiments setting which consists of four paths of three types, city main roads, highway and parking area, with a total length of 23km.

The city way has signal lamps and crosswalks, but the highway had none. In the parking area, the drivers were required to drive slowly and cautiously. The drivers completed two round trips for a reliable classification. The driving data per driver were labeled from A to J. A total of 94,401 records every second were captured leading to a 16.7MB dataset. The data were collected from the vehicles CAN bus through the On Board Diagnostics 2 (OBD-II) and CarbigsP (OBD-II scanner). The used vehicle has many measurement sensors and control sensors which are managed by the Electronic Control Unit (ECU).

IV. Proposed Theory

A. Using Kalman Filter model for pre-processing data for Fuel Economy

Kalman filters are ideal for systems which are continuously changing. They have the advantage that they are light on memory (they dont need to keep any history other than the previous state), and they are very fast, making them well suited for real time problems and embedded systems. This is suitable for our sensor readings of KMPL and mileage

The Kalman filter model assumes the true state at time k is evolved from the state at (k − 1) according to :

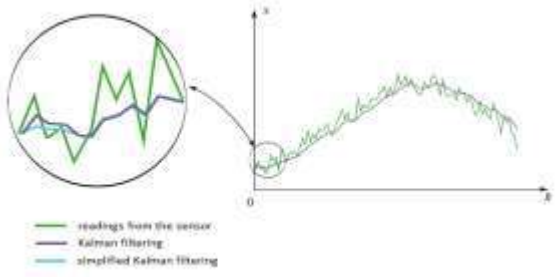


Fig. 4. An illustration of Kalman Filter.

$$\mathbf{x}_k = \mathbf{F}_k\mathbf{x}_{k-1} + \mathbf{B}_k\mathbf{u}_k + \mathbf{w}_k$$

This is where each value is filtered out of its predecessor.

x_k : is the one dimensional vector consisting of the PID associated with Kilometers per Liter (instant)
F_k, the state-transition model;
H_k, the observation model;
Q_k, the covariance of the process noise;
R_k, the covariance of the observation noise;

This is to remove outliers like 0 and 90 and bring the rated fuel efficiency to a large average

At time k an observation (or measurement) zk of the true state xk is made according to

$$\mathbf{z}_k = \mathbf{H}_k\mathbf{x}_k + \mathbf{v}_k \mathbf{z}_k = \mathbf{H}_k\mathbf{x}_k + \mathbf{v}_k$$

When a new value of fuel KMPL is received, it is now applied in the updation formula,

$$\mathbf{P}_{k|k} = (\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)\mathbf{P}_{k|k-1}$$

Where K_k is the optimal Kalman Gain

When required, the fuel efficiency can now be predicted by simply applying the estimation or prediction model

$$\hat{\mathbf{x}}_{k|k-1} = \mathbf{F}_k\hat{\mathbf{x}}_{k-1|k-1} + \mathbf{B}_k\mathbf{u}_k$$

B. Driver Identification

- Data Preprocessing

The objective of this task is to transform the collected data for the subsequent analysis and classification algo- rithms. This task consists of the following subtasks.

– Data preparation and cleaning

Constant and identical columns are removed. For example, the engine torque value is identical to the correction of engine torque value. After deleting redundant features, we replace the missing values or wrong ones using the K Nearest Neighbor (KNN) method.

– Feature selection

After data preparation, we select the most con- tributing features and exclude those that are highly correlated with them in order to improve the driver identification performance in terms of accuracy and speed.

– Feature transformation

First, the time is transformed from date-time for- mat to timestamp format in order to easily include this feature in the learning algorithms. Then, the dataset is split into multiple segments to be used in the subsequent optimization process. Further, as the features have different scales, they are normalized prior to their use in the machine learning algorithms. Indeed, the normalization process is necessary for algorithms that are based on the distance between data elements, such as the KNN algorithm. This normalization is performed using Eq.(1), where Xi is the normalized version of feature xi; the resulting normalized features lie between 0 and 1.

- Classification algorithms

The machine learning algorithms considered in the classification task are Decision Tree, Random Forest, Extra Trees, KNN, SVM, Gradient Boosting, AdaBoost based on Decision Tree, and multi-layer perception (MLP).

For the Security dataset, we used the fifteen most important features, along with the normalized timestamp. For the HciLab dataset, all features were used in the classification task.

For the UAH-DriveSet dataset(training dataset) , only the GPS-related features were used in the classification, since the sensors have different sampling frequencies and are not synchronized. Each of the above-mentioned classification algorithms generates a driver identification model.

Using the classification model trained by authorized users, driver verification consists of testing whether or not the user is classified into one of the pre-defined classes, e.g., authorized drivers.

The testing process is based on the computation of the probability of occurrence of each the pre-defined classes given the new data samples.

For the Random Forest algorithm, these probabilities are computed using the frequencies of each class, given a new driving pattern, among the large number of generated trees.

If all computed probabilities fall below a pre-defined threshold, the driver is declared not to be one of the authorized drivers, and thus an alert may be sent to the owner of the car or the vehicle control center. To minimize the probability of false alert, this threshold must be chosen judiciously, according to the minimum accuracy obtained in the training phase. The proposed time-optimized driver fingerprinting method is based on the driving patterns. It is shown that in-vehicle network data, such as fuel trim, brake pedal and steering wheel

data, are relevant in accurately identifying drivers. It is also shown that it is possible to identify drivers with a very high accuracy within the first 3min of driving, using a limited amount of sensor data collected from a restricted but judiciously chosen set of sensors.

C. Model for Driver Fingerprinting

Input:
Observations $y = y_t, t = 1, 2, 3$
1. Vehicle speed
2. Steering Degree
3. When steering degree starts to increase

Output:
Observation Probability:
$P(y_t|x_t)$
x : true state(hidden state)
y : observation

Transition Probability:
$P(x_t|x_{t-1})$
where t: state

Joint probability:
$P(X, Y) = P(x_1) \sum P(x_{t+1}|x_t)$
from $t = 1$ to $t - 1$

newline Joint probability of current state to all historical observations:
$t(x_t) = P(x_t, y_1.y_t)$

Conditional probability of all future observations:
$t(x_t) = P(y_{t+1}, ...y_T|x_t)$

Marginal probability:
$P(x_t|y) = P(x_t, y_1.y_t)P(y_{t+1}, ...y_T|x_t)$

D. Rating the Driver

The proposed method collects vehicle operation informa- tion, including vehicle speed, engine RPM, throttle position, and calculated engine load, via OBD interface. Then the proposed method makes use of AdaBoost algorithms to create a driving behavior classification model, and finally could determine whether the current driving behavior belongs to safe driving or not. Experimental results show the correctness of the proposed driving behavior analysis method can achieve average of 99.8% accuracy rate in various driving simulations. The proposed method has the potential of applying to real- world driver assistance system.

Currently, with the economy developing, the amount of the vehicles increases every year. As the same time, the amount of non-professional drivers increases rapidly.

Since most novice drivers are unskilled, unfamiliar with the vehicle condition and in weak awareness of traffic safety, the drivers' personal factors have become the main reasons of traffic accidents.

The driving auxiliary equipment is urgently needed to remind drivers the vehicle information in time and correct improper driving behavior.

Current research methods of driving behavior analysis include the driving data collection, driving modeling algorithms, and applications. Driving data collection includes automotive video capture car-mounted sensors, and the on board diagnostic (OBD).

In terms of driving behavior modeling algorithms, there are HMM, support vector machine (SVM), decision trees, and other principles.

The main application of the driving behaviour analysis is the identification of driving lethargy or the driver's actions forecast.

To judge the driver based on his driving skills, a method is proposed that includes the data from the OBD port and AdaBoost algorithm.

This proposed method collects the vehicle operation information, including vehicle speed, engine RPM, throttle position, and calculated engine load, from the OBD interface. Then the proposed model makes use of AdaBoost Algorithms to create a driver behaviour classification model, and finally could determine whether the current driving behaviour belongs to safe driving or not. Experimental results show the correctness of the proposed driving behaviour analysis method can achieve average of 99% accuracy rate in various driving simulations .

The method consists of splitting of datasets , pre-processing

and then the classification algorithm which is the AdaBoost algorithm .

Firstly, the dataset is split appropriately into training and testing dataset.

After splitting , the preprocessing is applied to the datasets simultaneously.

Classification makes the test samples into the driving model based on AdaBoost algorithms to classify and determine the sample category.

The classifiers are then divided into right and wrong classifiers based on the error rate. The strong classifier obtained as the result then finally provides if the driver has a normal condition or a bad vehicle condition data

- The normal condition data

  The relative ratio of the vehicle speed and engine speed is maintained at between 0.9 and 1.3 (test in the same gear); the relative ratio of the engine speed and throttle valve is maintained at between 0.9 and 1.3; the engine load is maintained between 20% and 50%.

- The bad vehicle condition data

  The relative ratio of the vehicle speed and engine speed is maintained more than 1.3 or less than 0.9; the relative

ratio of the engine speed and throttle valve is maintained more than 1.3 or less than 0.9; the engine load is maintained greater than 50% or less than 20%.

E. Model for Rating the Driver

Training data set T= $\{(x_1, y_1), (x_2, , y_2), ., (x_n, y_n)\}$
where x = p_1, p_2, p_3
p_1 = vehicle speed
p_2 = engine speed
p_3 = throttle position

- Initialization of the weight value distribution of the training data
  $D_1 = (w_{11}, ..w_{1i}, ..., w_{1N})$,

  $w_{1i} = 1/N, i = 1, 2, ..N$
  m = 1,2...M (m is the times of train.)

- (a) Using training data set has the weight distribution D_m to learn, get the basic classification

  $G_m(x) : X \in \{+1, -1\}$

- (b) The classification error rate of $G_m(x)$ is calculated on the training data

  $e_m = P(G_m(x) \neq y_i) = \sum_{i=1}^{N} w_{mi} I(G_m(x) \neq y_i)$

- (c) Calculation the coefficient of $G_m(x)$, the logarithm is the natural logarithm

  $\alpha_m = \frac{1}{2} log(\frac{1-e_m}{e_m})$

- (d) Update the weight value distribution of the training data $D_m = (W_{m+1,1}, ..., W_{m+1,i}, ., W_{m+1,N})$
  $W_{m+1,i} = W_{mi}(exp(-\alpha_m y_i G_m(x_i)))$

  Where Z_m is normalization factor which could make the D_m become a probability distribution .

  $Z_m = \sum_{i=1}^{N} W_{mi} exp(-\alpha_m y_i G_m(x_i))$ [where i = 1...N]

Build a linear combination of basic classifiers :
$f(x) = \sum_{m=1}^{M} \alpha_m G_m(x)$ [Where m=1..M]

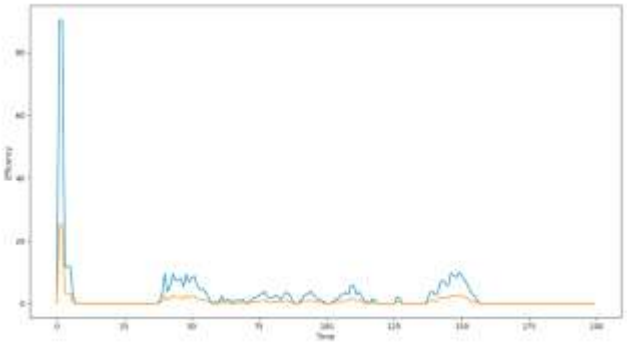


Fig. 5. Result after applying Kalman Filter on the data.

q = 0.262

r = 0.017

As shown in the above graph, we have used the Kalman filter to reduce the noise in the system, where the upper line represents instant values, and the lower line represents kalman noise. Now these Kalman values are averaged out over 2000 readings to give an average fuel efficiency, in our case, of 14.52 kilometers per liter, while using a 1.4 liter engine, diesel

VI. Conclusion

The papers evaluated show that there exist various meth- ods for fingerprinting, fuel efficiency as well as distraction detection, solely through data gathered by OBD-II ports. This further goes to suggest that a project to build a system with these elements to provide a framework for drivers can be built and improved upon. Our approach uses Kalman Filters to reduce the noise in the data and by applying various classification algorithms our model will be able to fingerprint as well as rate the driver.

V. Discussion and Results

Our work shows that using the values received by logging OBD values, we have successfully applied filters to reduce noise and fine tuned these values to show an acceptable fuel efficiency. Below are the results of the application of the kalman filter, with noise values

VII. Acknowledgment

We would like to thank our project sponsors QuicSolv

Technologies Pvt. Ltd. for their financial support and backing.

REFERENCES

[1] Saad Ezzini, Ismail Berrada and Mounir Ghogho. Who is behind the wheel? Driver identification and fingerprinting. Journal of Big Data.

[2] Fabio Martinelli, Francesco Mercaldo, Vittoria Nardone, Albina Or- lando, Antonella Santone : Whos Driving My Car? A machine learning based approach to driver identification.

[3] Choi, S., Kim, J., Kwak, D., Angkititrakul, P., and Hansen, J. H. (2007).

Analysis and classification of driver behavior using in-vehicle can-bus information. In Biennial Workshop on DSP for In-Vehicle and Mobile Systems, pages 1719.

[4] Nishiwaki, Y., Ozawa, K., Wakita, T., Miyajima, C., Itou, K., and Takeda, K. (2007). Driver identification based on spectral analysis of driving behavioral signals. In Advances for In-Vehicle and Mobile Systems, pages 2534. Springer.

[5] Wakita T, Ozawa K, Miyajima C, Igarashi K, Katunobu I, Takeda K, Itakura F. IEICE Driver identification using driving behavior signals. TRANSACTIONS on Information and Systems. 2006;89(3):118894.

[6] Miyajima C, Nishiwaki Y, Itou K, Takeda K, Ozawa K, Itakura F, Wakita T. Driver modeling based on driving behavior and its evaluation in driver identification. Proc IEEE. 2007;95(2):42737.

[7] Meng X, Lee KK, Xu Y. Human driving behavior recognition based on hidden Markov models. In: IEEE international conference on robotics and biomimetics. ROBIO06. IEEE; 2006. p. 2749.

[8] Choi S, Kim J, Kwak D, Angkititrakul P, Hansen JH. Analysis and classification of driver behavior using in-vehicle can-bus information. In: Biennial workshop on DSP for in-vehicle and mobile systems. 2007. p. 1719.

[9] Wahab A, Quek C, Tan CK, Takeda K. Driving profle modeling and recognition based on soft computing approach. IEEE Trans Neural Netw.

2009;20(4):56382.

International Journal of Pure and Applied Mathematics
Volume 120 No. 6 2018, 547-556
ISSN: 1314-3395 (on-line version)
url: http://www.acadpubl.eu/hub/
Special Issue

http://www.acadpubl.eu/hub/

Food Nutritional Value Evaluation System

Dr. Sashikala Mishra [1], Prof. Prashant Gadakh[2],
Mr. Chinmay Gosavi[3], Mr. Shrinit Padwal[4], Mr. Vishal Kasar[5]
[1] HoD, Department of Computer Engineering
[2] Faculty, Department of Computer Engineering
[3,4,5] Final year Students, Department of Computer Engineering
[1,2,3,4,5] I2IT, Pune

July 8, 2018

Abstract

Food plays a vital role in human existence and is our source of energy. With the demand for food increasing daily, the quality of food is suffering due to malpractices such as adulteration and heavy use of various pesticides. The traditional approach to analyze food nutritional values involves the use of various sensors and laboratory procedures to detect the quality, but such sensors and methods are not readily available. There is a need of a system which we can use to quickly evaluate the quality of food by methods which are ubiquitous. The number of handheld devices and their processing capabilities has increased manifolds over the last few years. Thus, the aim to the project is to help detect the nutritional quality of the food by utilizing the various sensors which are present in smartphones such as cameras and microphone. The main benefit of such a project is obvious, no extra investment is required as the only requirement is to own a Smartphone which are extremely common.

Keywords: Statistics, Malpractices, NLP, Machine Learning, DietCam.

1 Introduction

Food plays a vital role in human existence and is our source of energy. It is basic necessity as well as luxury depending on circumstances. With the ever increasing population the demand for food has also increased exponentially. Due to food being a vital commodity the competition between producers has increased to such extent that the originality and the natural value of food have been diminishing day by day. This has led to many malpractices by food producers and vendors. We have all become a little skeptical of the purity and quality standards of the food we consume these days. Food colors, chemicals and additives often creep up not just in our local produce, but even in packaged products. Such practice where the quality of food is altered by chemical means or procedures is referred to as food adulteration. The side effects of such adulteration can lead to a variety of diseases ranging from some common diseases like stomach disorders ranging to complex diseases like liver disorders or cancer. Deaths due to poor quality of food are

increasing yearly and statistics point to about 2.2 million deaths globally. Thus, it has become necessary to be check of the nutritional value and quality of the food we are consuming on daily basis. However, the procedures for checking are not always readily available to us as they require laboratory apparatus and expertise in the field. This project addresses this problem by trying to provide a method which can be available to all mobile phones. The formatter will need to create these components, incorporating the applicable criteria that follow.

2 SURVEY WORK

The topic of evaluation of Food Nutrition value has been researched heavily by health and nutrition related areas. The traditional method involves detection of color and texture analysis and optical sensing techniques described in [1], [2] and [6]. Pouladzadeh et al. highlights the various factors which may cause discrepancies in the food and nutrition recognition process when done by using just Smartphone cameras in [1]. Some of them include not properly understanding the amount of food present in the dish, the condition where the food isnt detected in the camera due to it being out of range

of visibility, etc. The paper offers a solution to these problems by utilizing both color and texture analysis. The data is prepared in the above factors and the food is properly segmented and important features are extracted before classifying it using a SVM which predicts target value of data instances in the testing set, which are given only by their attributes.

Kong et al. uses a different approach for the food recognition problem by proposing a method which involves two major factors - ingredient detection and food classification. Food ingredients are detected through part-based model and texture verification model while food is classified from detected ingredients using multi-view multi-kernel SVM. The results obtained from experiments on developed database shows that DietCam is most accurate among classifiers with accuracy of about 85% for food items [2]. Similarly, Anthimopoulos et al. have described a method which involves a bag-of-features (BoF)-based system for food image classification, as a first step towards the development of a portable application. This provides a method for dietary advice to diabetic patients through automatic CHO(carbohydrate) counting [6].

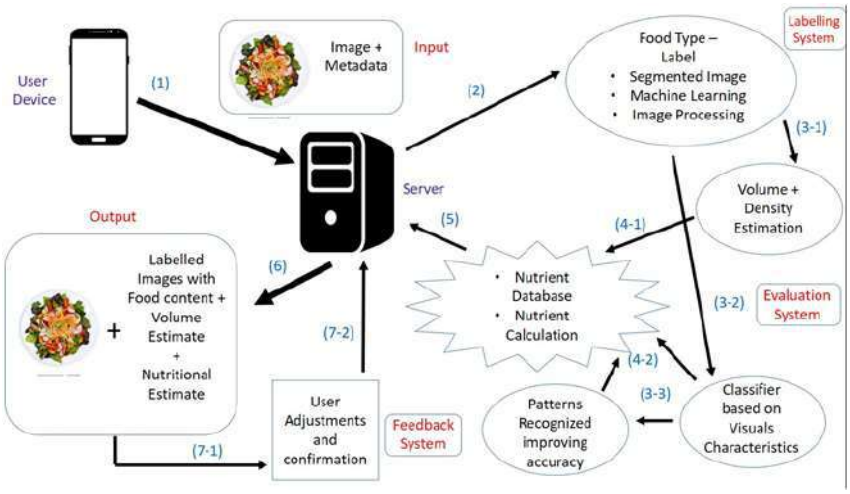
As described above detection of adulteration is also a problem. Ginesu et al. deals with the problem of detection of foreign bodies. Here well contrasted thermographic images are obtained using thermographic camera. Plancks law is used radiation emission from black body. Gray levels are measured and gray levels vs time graph is obtained. Foreign bodies and main food have different heat conductivities and capacities and thus different cooling rates. Difference is calculated later, absolute levels between gray levels of food and foreign bodies by using various pre-processing methods such as dead pixel correction, enhancement filter applications, shading correction and histogram stretching [5]. Another approach to adulteration detection has been described by Karuppuswami et al. [7] demonstrates a simple, inexpensive, and reusable hybrid sensor for real time liquid sample measurements (in this paper extra virgin olive oil). Authors has designed a hybrid passive wireless sensor that measures both electrical and mechanical properties of liquid, providing a two factor quality control. The hybrid sensor is based on an inductor-capacitor (LC) resonant tank coupled with a magnetoelastic strip. The mechanical and electrical resonances change as a function of viscosity and dielectric constant, respectively. Hence,

detection of adulteration in extra virgin olive oil is achieved by measuring the change in viscosity and dielectric constant for different adulteration levels.

The problem of detecting the quality or grade of fruits is also an important problem both commercially and for health reasons. Nandi et al. proposed a scheme for automated grading of mango according to maturity level in terms of actual-days-to-rot and quality attributes like size, shape and surface defect using hardware and software. The hardware includes the conveyor belt speed control mechanism, feed control, camera control, light intensity control and switching control. The software system analyses the still frame extraction, pre-processing of image, features extraction and finally gradation. The performance accuracy achieved using this proposed system for grading of mango fruit is nearly 87% [9].

Milk is another common candidate for adulteration practices with adulterants like unhygienic water, chalk powder, urea, starch, etc. Banwari et al. describe a method proposes a completely non-contact type method for detection of milk adulteration, preserving the consistency and quality of milk sample and making it reusable for testing again. Authors built an embedded system consisting of AVR microcontroller integrated with the optical sensor. They used refractive index as principle parameter for detecting the adulteration. As the amount of water changed, the refractive index changed [8].

3 PROPOSED DESIGN AND ARCHITECTURE



4 METHODOLOGY

1. Android app developed in Java. Usage of android.hardware API to access camera to take images of food.

2. The image data processed on device using Image Processing, converting it into required format for prediction.

3. The data (image data along with model name, version name and input data file written in JSON file) send to cloud via internet (TCP/IP) for online prediction. Communication with TensorFlow Application with help of Cloud ML Engine APIs.

4. The Cloud ML service runs the model with sent data and returns the predictions as response message for the call. Unpack the message and display the output (the food category with highest predicted percentage).Back-End

5. Gathering and Preparing training data Collecting data with labels (food category and nutrition value); Cleaning data by

looking for anomalies such as instances with missing features, multiple methods representing same feature; Transforming data to format compatible with Cloud ML Engine. Uploading the dataset to cloud (Google Cloud Platforms storage) so it can be accessible by Cloud ML Engine. Splitting the data into training data, evaluation data and test data.

6. Training the model Create a TensorFlow application that defines the computation graph and trains the model(classifier); Package and transfer the application to Google Cloud Storage bucket to access the project(automated using the gcloud command-line tool to run a training job). Cloud ML Engine training service sets up resources for your job by allocating one or more virtual machines. Training service runs your trainer, passing the command-line arguments specified when creating the training job. After training of model export it as TensorFlow SavedModel saved on Google Cloud Storage.

7. Deploying the model Create Cloud ML Engine model (gcloud ml-engine models create food_idf); Create the model version (gcloud ml-engine versions create version_name –model food_idf –origin bucket_path).

5 CONCLUSION

The problem of increasing food adulteration is posing a serious harm to the society. While laboratory procedures are the ultimate test for checking quality its not feasible to use them every time. The final system, if properly implemented, will be a huge benefit in day to day life as well be able to quickly decide whether the food is fit for consumption and what its nutritional value is. The main advantage of such a system being that no extra investment is required for this procedure as smartphones are widely available. The project is divided into 4 modules input, communication, processing and feedback. Input - The input module involves capturing the image using the end-users camera. Basic checking of the image will be done on the phone in order to ensure that the image quality is acceptable. Communication - The communication module consists of every communication task both input and output involved in

the project. The initial task is to send the captured Image over the cloud for further processing. The module also includes services like confirmation of whether the Image received or not. Other type of conformation includes the output confirmation whether the output was received by the user device. Processing - The image received from the communication module will be transformed to appropriate format and segmented in order to identify food item(s) present in the image. The image will then be calibrated in in order to establish a scale so that the quantity of food items can be calculated. Visual pattern recognition and metrics are used in order to verify if the items are properly identified. Using the food type and quantity obtained in the previous steps, the nutritional value can be calculated by referring to the database. Feedback - The results are displayed to the user after the processing step using the communication module and then the user is asked for feedback to confirm the accuracy of identified food items. The ratings from the user are supplied to the processing module in order to improve the models accuracy.

References

[1] P. Pouladzadeh, S. Shirmohammadi and R. Al-Maghrabi, "Measuring Calorie and Nutrition From Food Image", IEEE Transactions on Instrumentation and Measurement, vol. 63, no. 8, pp. 1947-1956, 2014.

[2] H. He, F. Kong and J. Tan, "DietCam: Multiview Food Recognition Using a Multikernel SVM", IEEE Journal of Biomedical and Health Informatics, vol. 20, no. 3, pp. 848-855, 2016.

[3] G. Villalobos, R. Almaghrabi, P. Pouladzadeh, and S. Shirmohammadi, An image processing approach for calorie intake measurement, in Proc. IEEE Symp. Med. Meas. Appl., Budapest, Hungary, May 2012, pp. 15.

[4] G. Villalobos, R. Almaghrabi, B. Hariri, and S. Shirmohammadi, A personal assistive system for nutrient intake monitoring, in Proc. ACM Workshop Ubiquitous Meta User Inter., Scottsdale, AZ, USA, 2011, pp. 1722.

[5] G. Ginesu, D. Giusto, V. Margner and P. Meinlschmidt, "Detection of Foreign Bodies in Food by Thermal Image Processing", IEEE Transactions on Industrial Electronics, vol. 51, no. 2, pp. 480-490, 2004.

[6] M. Anthimopoulos, L. Gianola, L. Scarnato, P. Diem and S. Mougiakakou, "A Food Recognition System for Diabetic Patients Based on an Optimized Bag-of-Features Model", IEEE Journal of Biomedical and Health Informatics, vol. 18, no. 4, pp. 1261-1271, 2014.

[7] S. Karuppuswami, A. Kaur, H. Arangali and P. Chahal, "A Hybrid Magnetoelastic Wireless Sensor for Detection of Food Adulteration", IEEE Sensors Journal, vol. 17, no. 6, pp. 1706-1714, 2017.

[8] A. Dave, D. Banwari, S. Srivastava and S. Sadistap, "Optical sensing system for detecting water adulteration in milk", 2016 IEEE Global Humanitarian Technology Conference (GHTC), 2016.

[9] C. Nandi, B. Tudu and C. Koley, "A Machine Vision Technique for Grading of Harvested Mangoes Based on Maturity and Quality", IEEE Sensors Journal, vol. 16, no. 16, pp. 6387-6396, 2016.

[10] F. Kong and J. Tan, Dietcam: Regular shape food recognition with a camera phone, Body Sensor Networks (BSN), 2011 International Conference on, pp. 127 132, 2011.

[11] D. Casasent, M. A. Sipe, T. F. Schatzki, P. M. Keagy, and L. L. Lee, Neural net classification of X-ray pistachio nut data, LebensmittelWissenschaft Und-Technol., vol. 31, pp. 122128, 1998.

[12] V. Davidson, J. Ryks and T. Chu, "Fuzzy models to predict consumer ratings for biscuits based on digital image features", IEEE Transactions on Fuzzy Systems, vol. 9, no. 1, pp. 62-67, 2001.

[13] P. Neelamegam, S. Abirami, K. Vishnu Priya and S. Valantina, "Analysis of rice granules using image processing and neural

network", 2013 IEEE CONFERENCE ON INFORMATION AND COMMUNICATION TECHNOLOGIES, 2013.

[14] B. S. Anami, V. Burkpalli, S. A. Angadi, and N. M. Patil, Neural network approach for grain classification and gradation, Proceedings of the second national conference on document analysis and recognition, pp. 394408, July 2003.

[15] M. O'Farrell, C. Sheridan, E. Lewis, C. Flanagan, J. Kerry and N. Jackman, "Online Optical Fiber Sensor for Detecting Premature Browning in Ground Beef Using Pattern Recognition Techniques and Reflection Spectroscopy", IEEE Sensors Journal, vol. 7, no. 12, pp. 1685-1692, 2007.

[16] I. Khosa and E. Pasero, "Artificial neural network classifier for quality inspection of nuts", 2014 International Conference on Robotics and Emerging Allied Technologies in Engineering (iCREATE), 2014.

[17] S. Sasano, X. Han and Y. Chen, "Food recognition by combined bags of color features and texture features", 2016 9th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), 2016.

[18] A. Mueen, M. S. Baba and R. Zainuddin, Multilevel feature extraction and x-ray image classification Applied Sciences, 7.8 (2007): 1224-1229.

[19] K. Lee, Q. Li and W. Daley, "Effects of Classification Methods on Color-Based Feature Detection With Food Processing Applications", IEEE Transactions on Automation Science and Engineering, vol. 4, no. 1, pp. 40-51, 2007.

[20] V. Singh and S. Dwivedi, "Ultrasonic detection of adulteration in fluid foods", Proceedings of the First Regional Conference, IEEE Engineering in Medicine and Biology Society and 14th Conference of the Biomedical Engineering Society of India. An International Meet.

Medical Data Security Using Distributed Server

[1]Ananya Sinha, [2]Kimi Ramteke, [3]Nikita Praveen Kumar, [4]Kritika Asreth

[1]Student, [2]Assistant Professor, [3]Student, [4]Student

[1]Computer Engineering,

[1]Hope's Foundation Isquareit, Pune, Maharashtra, India

Abstract : Medicinal services applications unit are thought-about as promising fields for remote devices systems, wherever patient is also checked utilizing remote restorative device systems (WMSNs). WMSN welfare investigate patterns that focus on patient solid correspondence, quiet state and vitality mean guiding, as one or two of precedents. Be that as a result of it might, inflicting new advancements in human services applications whereas not considering security makes persistent protection helpless. In addition, the physiological information of a private unit is deeply sensitive. Hence, security is also a foremost necessity of welfare applications, notably on account of patient protection, if the patient choices a mortifying unhealthiness. We have an inclination to feature some notable human services ventures utilizing remote healthful device organizes, and mention their security frameworks arrangements will primarily produce bound the patient information amid transmission but still cannot guarantee at intervals assault wherever the supervisor of the patient information uncovers the sensitive patient information. Thus we have an inclination to face while proposing to keep at intervals assault by utilizing completely utterly different information servers to store quiet information. The principle commitment of this paper is to disperse patient's information safely in multiple information servers and working out the Paillier cryptosystems to perform measurable investigation on the patient information and not commerce off the patient's protection.

Index Terms **- Wireless medical sensor network, Patient data privacy, Paillier encryption.**

I. INTRODUCTION

There are lots of applications that require the protection of susceptible user information from hackers and unauthorized users. It's any organization's biggest responsibility to protect the information of their clients/users from being hacked and/or misused. If these very delicate data is revealed to any unauthorized person, it could result in a major loss for the client thus also reducing the organization's reputation. There are many security issues that an organization needs to take care of, like stealing of data, misuse of data, altering of data and incorrect values. Suppose if the unwanted person is making an attempt to hack the coed details, there are several possibilities that the data may be misused, leading to grave impact. The hackers may even change the info due to dearth of security. Severe consequences, like death of a patient is possible if the analysis provided by the physician is hacked. To stop these problems, an invasion detection system is projected. An invasion detection system could be a system accustomed to check the malignant actions and produce computerized statements to an administrative site. It constitutes of Paillier key crypto-systems. The formula is employed to cypher the coded particulars prior to storing it within the knowledge-base and perform coding once required by the medical practitioner. In hardware we tend to square measure scanning patient finger print for the protection purpose and providing the patient information to hospital.

II. RELATED WORK

Accumulating and dealing with susceptible data may be a troublesome duty. Veritably, there's no accepted formula for constructing the required data systems. SHAREMIND—proposed in [1], may be a virtual mechanism for privacy-conserving processes that depends on shared computing approach, which can be a regular manner for firmly assessing methods during a multiparty computing setting. The uniqueness of this answer is within the alternative of the key sharing theme. SHAREMIND protocols measure information-hypothetically safe within the honest-but-curious design with 3 computing members. Though the honest-but-curious model doesn't allow malignant members, yet caters considerably enhanced privacy preservation in comparison to straightforward centralized databases. Though it cannot guarantee ideal secure model and privacy conserving model absolutely, however it will to some extent [1].

A novel procedure downside, specifically the Composite Residuosity category obstacle including the operations to public-key cryptography are measured in [2]. It offers a replacement trapdoor mechanism and extracts against this method 3 encoding designs: a trapdoor permutation and 2 homomorphic probabilistic encoding designs. The cryptosystems, supported by standard arithmetic, incontrovertibly protect below applicable assumptions within the normal model. However, public secret is slower than a secret key encoding, one key provides that solely thanks to write and rewrite, simplifying and rushing up the method [2], [8].

[3] proposes secure and light-weighted systems for WMSNs. It utilizes hash-chain based key updating mechanism and proxy-protected signature technique in order for achieving capable safe communication and data access control all through the system to provide rearward confidence and confidentiality conservation. However, it is only applicable for low-power sensor nodes and not for large sensor nodes [3], [5], [6], [7].

Depending on secretiveness of information, the requirement of security may rise. It is of utmost importance to prevent the information from non-legitimate disclosure or misuse of information.These days, because of the changing technology, large amount of multimedia is produced and communicated, which leaves our individual information to be misused, i.e., edit, modify and duplicate. Cryptography is a skill of obscure writing. It validates information and data of high value. It safeguards the system from

authentic intrusion. Advanced Encryption Standard (AES) [4] is one such safety algorithm, which provides data security. It includes encryption and decryption, that are correlated with a key. The key has to be obscure [4],[12],[13],[14],[15].

III. PROPOSED SYSTEM

For safe-guarding the delicate information of patients from intrusion, we have proposed a brand new information assortment protocol. The susceptible patient data is divided into 3 parts in line with a random value generator supported hash function and are then sent to three servers using secured mediums. To keep the susceptible patient data in informational access, we have come up with a fresh info access protocol on the thought of the Paillier cryptosystem. This covenant permits the user (e.g. doctor) to access the patient info whereas not disclosing it to any other server. For conserving the confidentiality of the patient data in arithmetic study, we tend to propose some new privacy-preserving arithmetic study protocol on the thought of the Paillier cryptosystems. It allows the user (e.g., medicinal analyst) to perform arithmetic study on the patient info whereas not compromising the patient information privacy.

IV. PROPOSED SYSTEM ARCHITECTURE

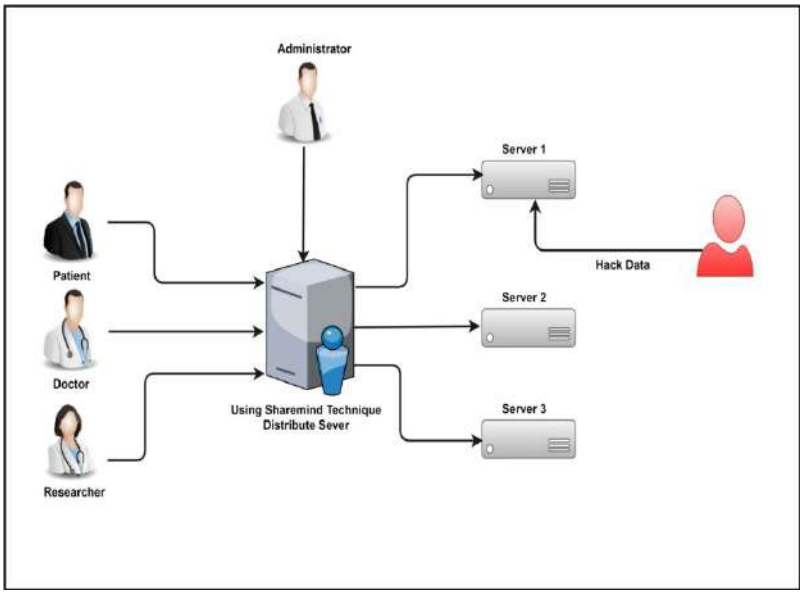


Fig 1. Proposed System Architecture

Our design contains a random knowledge generator that sends the patient's information to a patient's information system. At the side of a random generator, it has a patient information directory. The patient information directory stores the patient information from the random-generator. It also administers querying assistance to users (e.g., doctors and medicinal specialists). It additionally contains a patient's knowledge access system that is employed by the user (e.g., doctor) to access the patient's information and monitor the same system that is employed by the user (e.g., medicinal analyst) to question the patient's information directory system and analyse the patient's knowledge analytically.

V. FUTURE SCOPE

Health care operations are rising fields for WMSNs. It is where patients can be observed. Instead of randomly generating data, the data can be taken via wireless devices that patients can be. These wireless devices may or may not continually monitor the patient and send information to patient knowledge database. Retrieval of compromised data through proxy server, for data security purpose, system gives permission to user regarding compromised data. This data can now be subjected to privacy protection using distributed servers.

VI. CONCLUSION

We have proposed the security and privacy issues among the medical detector information assortment storage and queries and given a complete resolution for privacy-preserving medical detector network through the ad-hoc network. To retain the privacy of the patient data, we may tend to project a latest data assortment protocol that divides the patient's information into 3 parts and keeps them in 3 data servers. Until the joined data server is not jeopardized, the confidentiality of the patient data are retained. In case of genuine users (i.e., doctor) to have an access to the patient information, we may tend to project associate access management protocol, where three data is distributed in numerous servers.

VII. ACKNOWLEDGEMENT

Working on the project "MEDICAL DATA SECURITY USING DISTRIBUTED SERVER" was a great experience. It was a source of immense knowledge. We would like to express our sincere gratitude to project guide Prof. Kimi Ramteke for her guidance and support through out the semester. We acknowledge her encouragement and moral support. Her guidance was very helpful in different matters regarding the project. We would also like to express our gratitude to Project Coordinator Prof. Prashant Gadakh for support and guidance for various issues regarding the project. He was very kind and patient while explaining us the outlines of the project. We thank him for his overall support. We would also like to thank all teachers and our fellow friends who all supported us.

REFERENCES

[1] Dan Bogdanov, Sven Laur, Jan Willemson, "Sharemind: A Framework for Fast Privacy-Preserving Computation," Proceedings of the 13th European Symposium on Research in Computer Security: Computer Security, 2008

[2] Pascal Paillier, "Public-Key Cryptosystems Based on Composite Degree Residuosity Classes," In Proc. EUROCRYPT: International Conference on Theory and Applications of Cryptographic Techniques pages 223-238, 1999

[3] .Yi, Xun, et al. "Privacy Protection for Wireless Medical Sensor Data," IEEE Transactions on Dependable and Secure Computing 13.3 (2016): 369-380.

[4] Radhika D.Bajaj , Dr. U.M. Gokhale, "Design and Simulation of AES Algorithm for Cryptography", IJESC (Volume 6, Issue no. 6)

[5] S. Dağtaş, G. Pekhteryev, Z. Şahinoğlu, H. Çam, and N. Challa, "Real-Time and Secure Wireless Health Monitoring," International Journal of Telemedicine and Applications, Volume 2008, Article ID 135808, 10 pages.

[6] Daojing He, Sammy Chanand, Shaohua Tang, "A Novel and Lightweight System to Secure Wireless Medical Sensor Networks", IEEE Journal of Biomedical and Health Informatics (Volume: 18 , Issue: 1 , Jan. 2014)

[7] P. Belsis and G. Pantziou, "A k-anonymity privacy-preserving approach in wireless medical".

[8] T. ElGamal, "A Public-Key Cryptosystem and a Signature Scheme Based on Discrete Logarithms. IEEE Transactions on Information Theory," 31 (4): 469-472, 1985

[9] D. Angluin and D. Lichtenstein, "Provable Security of Cryptosystems: A Survey", Computer Science Department, Yale University, TR-288, 1983.

[10] D. M. Alghazzawi, S. H. Hasan and M. S. Trigui, "Advanced Encryption Standard - Cryptanalysis research," 2014 International Conference on Computing for Sustainable Global Development (INDIACom), New Delhi, 2014.

[11] J. Ko, C. Lu, M. B. Srivastava, J. A. Stankovic, A. Terzis and M. Welsh, "Wireless Sensor Networks for Healthcare," in Proceedings of the IEEE, vol. 98, no. 11, pp. 1947-1960, Nov. 2010.

[12] Hassan Anwar, Masoud Daneshtalab, Masoumeh Ebrahimi, Juha Plosila, Hannu Tenhunen "FPGA Implementation of AES-based Crypto Processor", IEEE 2013.

[13] Abhijith.P.S, Mallika Srivastava, Aparna Mishra, Manish Goswami, B.R.Singh "High Performance Hardware Implementation of AES Using Minimal Resources", International Conference on Intelligent Systems and Signal Processing (ISSP), IEEE 2013.

[14] Bin Liu, Bevan M. Bass,"Parallel AES Encryption Engines for Many-Core Processor Arrays", IEEE TRANSACTIONS ON COMPUTERS, VOL. 62, NO. 3 MARCH 2013.

[15] Ohyoung Song, Jiho Kim "Compact Design of the Advanced Encryption Standard Algorithm for IEEE 802.15.4 Devices", Journal of Electrical Engineering & Technology, Vol. 6, No. 3, pp. 418-422,2011.

TPR, PPV and ROC based Performance Measurement and Optimization of Human Face Recognition of IoT Enabled Physical Location Monitoring

Ajitkumar S. Shitole, Manoj H. Devare

Abstract: ***This paper describes the construction of Internet of Things (IoT) enabled system which not only captures the sensors data in textual and numeric form but also performs live human face recognition to monitor physical location effectively. The dataset used in order to apply supervised machine learning algorithms is the combination of automatically captured live sensor data along with name of the human face recognized or unknown and additional manually introduced class label. Performance measurement of face recognition is done with the help of Decision Tree (DT), K-Nearest Neighbors (KNN), Naïve Bayes (NB) and Logistic Regression (LR). The results show that DT gives the best performance with respect to classifier's accuracy; True Positive Rate, Positive Predictive Value and area under curve of Receiver Operating Characteristics (ROC) for face recognition prediction whether the recognized face is true or false.***

Index Terms: ***Machine Learning, Physical Location Monitoring, Confusion Matrix, ROC, Decision Tree, Naive Bayes, Logistic Regression, K-Nearest Neighbors.***

I. INTRODUCTION

Internet of Things (IoT) is the one of the emerging and rapidly developing technology in the field of Information Technology and Communication Engineering. Lots of devices can be connected to each other with the help of IoT to communicate and exchange their information and data. In today's life, it is necessary to monitor the physical location with the help of IoT where numbers of different sensors are connected to single board computer. Analysis of physical location is required in order to identify any abnormal conditions in the environments like home locations, sensitive laboratories, hospitals, educational institute, industries etc. Abnormal conditions can be sudden increase or decrease in temperature and humidity, increase in intensity of light, increase in gas sensor values, unknown person's detection in the premises which in turn can cause severe damage to the location and surroundings. So it is essential task to capture sensor data continuously on regular intervals and perform statistical as well as systematic analysis of the same to create decision support system which is required to avoid further loss in the environment. IoT enabled system with multimedia data such as digital images of human faces are useful for face detection and recognition. Face recognition is useful in various scenarios such as intrusion detection, identifying the several actions such as switch ON/OFF various devices, identifying user's routine in the environment to know when user is at home and interacting with the devices and so on. Development of IoT enabled system with face recognition makes significant change in safety and security of premises. More robust and powerful system can be achieved with the help of IoT and face recognition. The objective of this paper is to present prescient scientific models for IoT enabled with face recognition system for monitoring physical location. Location considered here is the living room of a home and data is captured for one month continuously. The system employs four supervised machine learning predictive models with DT, KNN, NB and LR for analysis of human face recognition to find accuracies of applied classifiers, precision, recall and ROC curve and compare them.

II. RELATED WORK

Sankar Mukherjee et al. addressed an issue of meeting sensor connect with the Mobile Adhoc Network (MANET) organizes on the grounds that hubs have distinctive power levels, heterogeneous conventions and have odds of co-channel obstructions another design of IoT systems, where sensor systems and MANET are joined together for proficient correspondence with the Internet Gateways [1].

Neelesh Mishra et al. presented an overview of different congestion control calculations utilized at transport layer. IoT requires a vehicle layer convention which offers blockage control, adaptability and dependability as indicated by necessity of gadgets [2]. Dragos Mocrii et al. presented a survey of real advancements of IoT-based smart homes and current difficulties of brilliant home advances and their scattering, and indicate some interesting arrangements and future patterns [3]. Adel Alkhalil et al. recommended the usage of information provenance as an imperative instrument that can improve the security and protection of IoT frameworks and reviewed the most difficult issues in IoT information provenance. Seven issues have been talked about including provenance security, monstrous measure of information, ordering, different customers, change, question, and interoperability [4]. Nallapaneni Manoj Kumar et al. expounded the conceivable security and protection issues considering the segment cooperation in IoT and concentrates how the Distributed Ledger based Block Chain (DL-BC) innovation add to it [5].

Revised Manuscript Received on July 20, 2019.

Ajitkumar S. Shitole, Research Scholar, Amity University Mumbai, India, Asso. Prof, I²IT, Hinjawadi, India.
Dr. Manoj H. Devare, HoI, AIIT, Amity University Mumbai, India.

Mustafa Alper Akkaş et al. displayed a Wireless Sensor Network model comprising of MicaZ hubs which are utilized to quantify nurseries' temperature, light, weight and stickiness. With this framework farmers can control their nursery from their cell phones or PCs which have web association [6].

Partha Pratim Ray reviewed mainstream IoT cloud stages in light of explaining a few administration areas such as application advancement, gadget the executives, framework the board, heterogeneity the executives, information the board, devices for investigation, arrangement, checking, perception, and research. An examination is displayed for in general spread of IoT mists as per their appropriateness [7].

Nawaz Mohamudally et al. featured the difficulties significant to center components engaged with the advancement of an Anomaly Detection Engine (ADE). It was discovered that an exact and dependable ADE depends on three fundamental determination factors to be specific, the nature of the information focuses, the time arrangement change, and where investigation are executed [8].

Mohammad Saeid Mahdavinejad et al. evaluates the different machine learning strategies that bargain with the difficulties exhibited by IoT information by considering shrewd urban areas as the fundamental use case. The key commitment of this investigation is the introduction of a scientific classification of machine learning calculations clarifying how unique strategies are connected to the information so as to remove larger amount data [9].

Bill Karakostas proposed an engineering that utilizes a Bayesian occasion expectation display that utilizes chronicled occasion information produced by the IoT cloud to ascertain the likelihood of future occasions. Framework anticipated outbound flight defer occasions, in view of inbound flight delays, in light of authentic information gathered from avionics measurements databases [10].

Huseyin Yildirim et al. concentrated to break down the variables that impact representatives' aim to utilize wearable gadgets at the work environment. In this examination, an audit of the writing with respect to acknowledgment of innovations and affecting elements, for example, hazard and trust is utilized to build up an applied model [11].

Ajitkumar Shitole et al. clarified about proposed showing of relevant adjustment approach and the executives customization that abuses distinctive identification methodology to give a proactive control advantage at home is conceivable and attainable. Principle based administration customization technique that utilizes a standard happenstance strategy dependent on semantic separation to settle on choices about the unique situation and a set hypothesis strategy dependent on set hypothesis to screen benefit customization [12].

Manoj Devare explained about the huge amount of statistics values captured from the sensor want to be analyzed because one cannot forget the ideal values. The hassle may additionally arise at some stage in the handshaking of the sensors with the libraries established inside the SBCs. The sensor information gathered within the SBC, and pushing it either at the internet-server or Cloud is likewise having a few synchronization issues. The handshaking and synchronization troubles can be detected and appropriately analyzed the usage of the statistical gear and techniques which applied to the accrued sample data within the preliminary checking out [13].

Ajitkumar Shitole and Manoj Devare depicted about the observing of a physical area isn't only a basic action yet suggests vital restorative measures after efficient investigation, to stay away from the further misfortune in the materials and also dangers in nature. The sensor information caught as time arrangement is helpful for examination of the anomalous conditions in the environment. The content based and numerical qualities from the sensor are valuable for the examination utilizing the factual instruments and procedures [14].

Alexandra Moraru et al. presented vertical framework mix of a sensor hub and a toolbox of machine learning calculations for anticipating the quantity of people situated in a shut space. The dataset utilized as a contribution for the learning calculations is made out of consequently gathered sensor information and extra physically presented information. The framework broke down the dataset and assessed the execution of two kinds of machine learning calculations on this dataset. The investigations demonstrated that enlarging sensor information with appropriate data can enhance forecast results and furthermore the arrangement calculation performed better [15].

Joseph Siryani et al. depicted machine learning Decision-Support System (DSS) which enhances the IoT Smart Meter Operations. The model is observationally assessed utilizing informational indexes from a business organize. The framework shows the effectiveness of methodology with a total Bayesian Network forecast model and contrast and three machine learning expectation demonstrate classifiers: Naïve Bayes, Random Forest and Decision Tree. Results show that approach creates factually critical estimations and that the DSS will enhance the cost effectiveness of Electric Smart Meter (ESM) arrange tasks and support [16].

Purnendu Shekhar Pandey et al. acquainted significance with advice the individual about his undesirable way of life and even alert him/her before any intense condition happens. To distinguish the pressure heretofore framework have utilized heart beat rate as one of the parameters. IoT alongside ML is utilized to alert the circumstance when the individual is in genuine hazard [17].

Go Takami et al. outlined the ML techniques and described a sensor identification experiment and the results of a deterioration determination experiment that suggests the possibility of understanding the sensor deterioration process. System believed that there is a great possibility that analysis of sensor data using the ML techniques can be used for the preventive maintenance such as sensor deterioration estimation [18].

Rui Madeira et al. clarified ML Approach for Indirect Human Presence Detection Using IoT Devices. The gave data was anonymized at the source. The initial step was to extricate satisfactory highlights for this issue. A naming advance is presented utilizing a blend of heuristics to affirm the probability of anybody being home at a given time, in light of all data accessible, including, yet not constrained to, coordinate nearness indicators.

The arrangement lays chiefly on the utilization of regulated learning calculations to prepare models that recognize the nearness with no data dependent on direct nearness finders [19]. Che-Min Chung et al. examined about new methodology that utilized ML strategies to determine the gigantic information issue in the quickly business of the IoT. This arrangement is represented considerable authority in IoT information and connected to a genuine case of a keen working with more than 100 associated sensors and its execution is contrasted with industry benchmarks [20].

III. EXPERIMENTATION

IoT enabled device is created to reveal the physical area in real time the usage of sensors connected the usage of the jumper wires. Various sensors together with digital temperature and humidity sensor, light intensity, physical presence, and gasoline detection sensors are connected to Raspberry Pi3 to optimize the physical location tracking. The device is evolved to fetch actual-Time facts from the sensors. The web camera is likewise linked to Raspberry Pi3 to capture snap shots of human face for recognition. In order to monitor physical location, the sensor data is captured regularly in real time fashion and stored onto local server. Whenever human face is detected and recognized as either known or unknown person, same data is also stored onto Go Daddy Cloud Service for further use and analysis. Subset of original dataset is pushed onto cloud to create labeled dataset and then apply supervised machine learning algorithms to measure the performance of human face recognition. Fig. 1 shows the IoT Enabled System with Face Recognition for live face recognition and sensor data capturing. Fig. 2, Fig. 3, and Fig. 4 show Face Recognition of Known Persons. Face recognition and sensor readings are captured simultaneously using multithreading programming in python. The system is used to monitor the home location continuously for one month. To create labeled dataset, cloud data was downloaded daily to add the class label manually. Whenever the person was recognized incorrectly, false entry was registered manually in register to create labeled dataset to apply supervised machine learning algorithms.

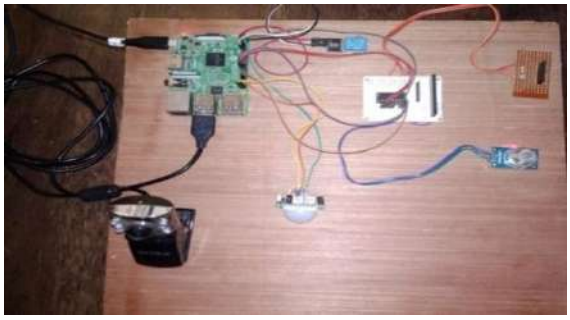


Fig. 1 IoT Enabled System with Face Recognition

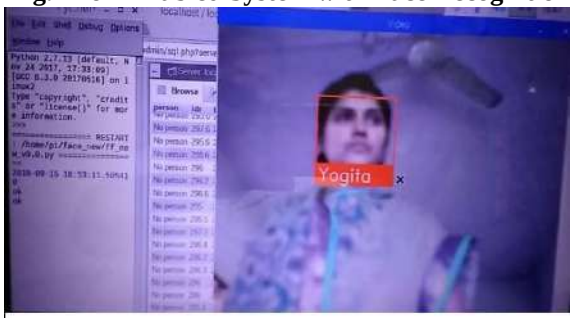


Fig. 2 Face Recognition of Known Person: Yogita

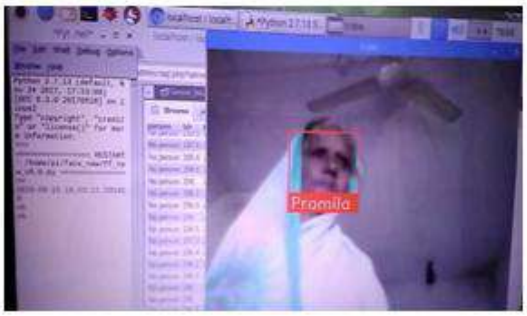


Fig. 3 Face Recognition of Known Person: Pramila

Fig. 4 Face Recognition of Known Person: Swaroop

In order to capture sensor values and to recognize human face in real time, multithreading programming in python is applied as Raspberry Pi 3 supports a quad core processor. Main thread along with two additional threads is created to achieve simultaneous processing which in turn to get maximum throughput. Main thread is used to capture the live image frame by frame and to perform processing on that captured image for face recognition. First thread out of two additional threads is used to read temperature and humidity sensor values. Second thread is used to read LDR, Gas, and PIR sensor values. To perform face recognition activity effectively in IoT enabled environment, face recognition library which recognizes and manipulates faces from Python is installed onto system. Local database of known faces is created to compare with live captured images frame by frame. Face recognition library consists of various in built methods to perform tasks such as to load image file, to get face locations, to get face encodings, to compare faces etc. Every known image is loaded into temporary variable for encoding of facial characteristics that can be contrasted with some other picture of a face. Two arrays are initialized to represent known face encoding and known face names. Live image is captured and processed to get areas and frameworks of every individual's eyes, nose, mouth, and jaw. Face location is applied to get face encodings. Captured image's face encodings are compared with known face encodings and if match is found known face name is displayed on screen otherwise unknown string is displayed. System consists of heterogeneous data as it combines numeric, string, and image data. Although image data is combined with sensor data, captured images are not stored either on local database or cloud. Whenever face is recognized, the names of known persons along with other sensor values are stored onto local server as well as cloud.

Irrespective of face detection and recognition, all entries with sampling rate of 2 to 4 seconds are maintained onto local database. Cloud database is a subset of local database as it contains entries when face is recognized either as known or unknown. Local database contains dataset in csv file with 5, 86,506 entries of size 40.2 MB where as cloud database contains dataset in csv file with 3025 entries of size 213 KB.

IV. MACHINE LEARNING MODELS

Classification is utilized to discover in which gather every datum example is connected inside a given dataset. It is utilized for characterizing information into various classes as indicated by some obliges. A few remarkable sorts of arrangement calculations including Decision Tree, K-Nearest Neighbors, Naive Bayes, and Logistic Regression are utilized for it. Arrangement is a two stage process. During initial step the model is made by applying arrangement calculation on preparing informational collection at that point, in second step the extricated model is tried against a predefined test informational collection to gauge the model prepared execution and precision. So grouping is the procedure to allot class mark from informational index whose class name is unclear.

A. Decision Tree

Decision tree fabricates regression or classification models as a tree structure. The last outcome is a tree with choice nodes and leaf nodes. A choice node has at least two branches, each speaking to values for the quality tried. Decision trees utilized in information mining are of two principle types: Classification tree investigation is the point at which the anticipated result is the class to which the information has a place. Regression tree examination is the point at which the anticipated result can be viewed as a genuine number. The Gini coefficient is a factual proportion of conveyance. The coefficient ranges from 0 (or 0%) to 1 (or 100%), with 0 speaking to consummate fairness and 1 speaking to consummate disparity. The impurity measure used in building decision tree in Classification and Regression Trees (CART) is Gini Index. The decision tree built by CART algorithm is always a binary decision tree.
Gini index for a given node *t*:

$$GINI(t) = 1 - \sum_j [p(j|t)]^2 \qquad (1)$$

$p(j|t)$ is the relative frequency of class *j* at node *t.*
When a node t is split into k partitions (child nodes), the quality of split is computed as,

$$GINI_{split} = \sum_{i=1}^{k} \frac{n_i}{n} GINI(i) \qquad (2)$$

Where n_i = number of records at child node i
n = number of records at parent node t
Trait that boosts the decrease in impurity or having minimum gini index is chosen as dividing attribute.

B. Naïve Bayes

The Naive Bayes Classifier system depends on Bayesian hypothesis and is especially utilized when the dimensionality of the information sources is high. Bayes hypothesis gives a method for computing the posterior probability P(A|B), from P(A), P(B), and P(B|A). Naive Bayes classifier thinks about that the impact of the estimation of an indicator (B) on a given class (A) is autonomous of the estimations of different indicators.

$$P(A|B) = \frac{P(A)P(B|A)}{P(B)} \qquad (3)$$

Where, P (A|B) is the posterior probability of class (target) given predictor of a class, P (A) is called the prior probability of a class, P (B|A) is the likelihood which is the probability of predictor of given class, and P (B) is the prior probability of predictor of a class.

C. K Nearest Neighbors (KNN)

KNN recognizes the order of unclear information point based on its nearest neighbor whose class is as of now known. It makes usage of the more than one nearest neighbor to decide the class in which the given information point has a place with and subsequently it is called as KNN.
The Euclidean distance between the points x and u is

$$d(x,u) = \sqrt{\sum_{i=1}^{n}(x_i - u_i)^2} \qquad (4)$$

D. Logistic Regression (LR)

Logistic Regression is utilized to portray information and to clarify the connection between one dependent binary variable and at least one nominal, ordinal, interim or proportion level autonomous factors. LR is a factual strategy for breaking down a dataset in which there is at least one autonomous factor that decide a result. The result is estimated with a dichotomous variable.
The "logit" function is given below

$$ln\left[\frac{p}{(1-p)}\right] = \alpha + \beta X + e \qquad (5)$$

Where, p is the probability that the event Y occurs, p(Y=1)
p/(1-p) is the "odds ratio"
ln[p/(1-p)] is the log odds ratio, or "logit"
The logistic distribution constrains the estimated probabilities to lie between 0 and 1.
The estimated probability is:

$$p = \frac{1}{[1 + \exp(-\alpha - \beta X)]} \qquad (6)$$

Where if $\alpha + \beta X = 0$, then p = .50
as $\alpha + \beta X$ gets really big, p approaches 1 , as $\alpha + \beta X$ gets really small, p approaches 0.

E. Open Source Distribution for ML Predictive Models

For ML predictive models, Anaconda Jupyter is used. Anaconda is open supply circulation of the Python and R programming languages for information technology and device studying related applications. The Jupyter notebook is open-source web software that lets in you to create and percentage files that include stay code, equations, visualizations and all. Four supervised machine learning algorithms: DT, NB, KNN and LR are applied.

First dummy variables are created for categorical attributes using pandas in python. Input and output variables are created using pandas. Input variable consists of values of all features except class label. Output variable with class label is created. Data set is divided into training and testing dataset. To create the models machine learning algorithms are applied on training data set. Classifier's performance is measured with the help of testing dataset. Graph of all confusion matrices and Receiver Operating Characteristics (ROC) are plotted for interpretations of accuracies.

V. EXPERIMENTAL RESULTS

To monitor physical location i.e. home, sensor values are collected for one complete month and to detect outliers, if any, box whisker plots are created for temperature, humidity, LDR, and gas sensor values.

A. Box Whisker Plots

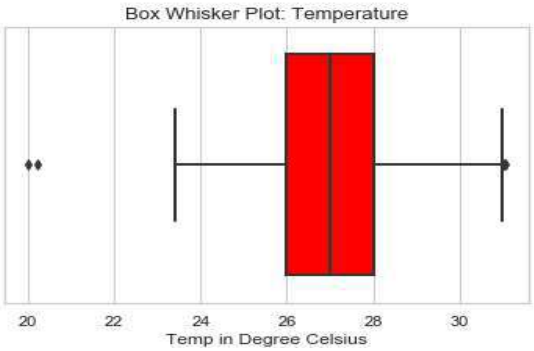


Fig. 5 Box Whisker Plot for Temperature

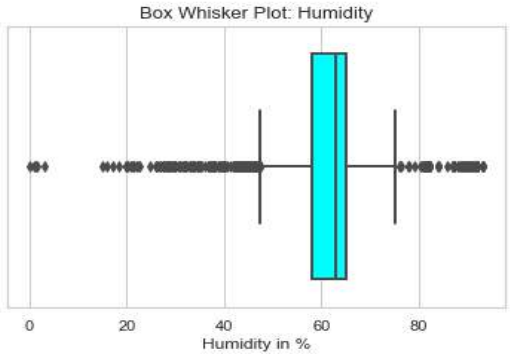


Fig. 6 Box Whisker Plot for Humidity

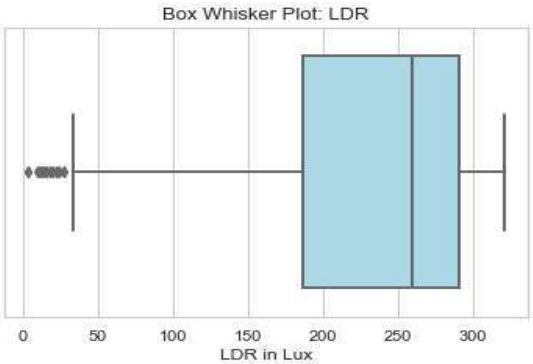


Fig. 7 Box Whisker Plot for LDR

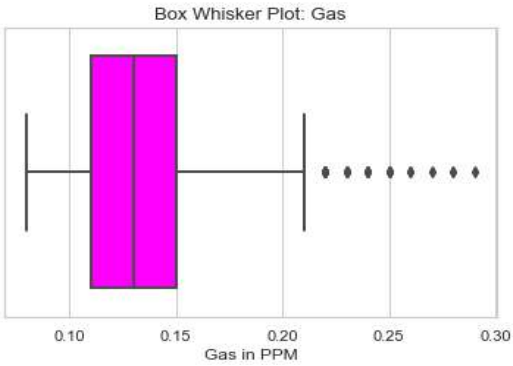


Fig. 8 Box Whisker Plot for Gas

Fig. 5, Fig. 6, Fig. 7, and Fig. 8 show box-whisker plots of temperature, humidity, LDR, and gas sensors respectively. Box whisker plot divides entire data into four regions and every region consists of 25% of total data. The first region is from minimum value to first quartile (Q1), second region is from first quartile to median (Q2), the third region is from median to third quartile (Q3) and fourth region is from third quartile to maximum value. Difference between Q3 and Q1 is called as Inter Quartile Range (IQR).

$$IQR = Q_3 - Q_1 \qquad (7)$$

The data points having values less than 1.5 times IQR and values greater than 1.5 times IQR are called as outliers. Outliers indicate unusual happenings in the environment where the system is located.

$$Outliers < 1.5 * IQR \qquad (8)$$

$$Outliers > 1.5 * IQR \qquad (9)$$

Fig. 5 and Fig. 8 show that very few outliers exist for temperature and gas sensor values. Fig. 6 shows too many outliers exist for humidity sensor and Fig. 7 shows outliers for LDR more than temperature and gas sensor values but less than humidity sensor values.
Various experiments are also carried out to assess classification accuracy, classification report, ROC curves, evaluation and the analytical model selection based on ML classifiers.

B. Classification Accuracy

A confusion matrix is an abstract of forecast results on a classification problem. It is a two dimensional matrix of order 2*2 for binary classification problem. Row is reserved to indicate actual values of negative and positive samples. Column is reserved to indicate predicted values of negative and positive samples. Matrix is divided into four cells such as True Negative (TN), False Positive (FP), False Negative (FN), and True Positive (TP) respectively. Entries along the diagonal from left most upper corner to right most bottommost corner represent true entries representing either TN or TP otherwise remaining entries are false. FPs are called as type-I error and FNs are called as type-II error.

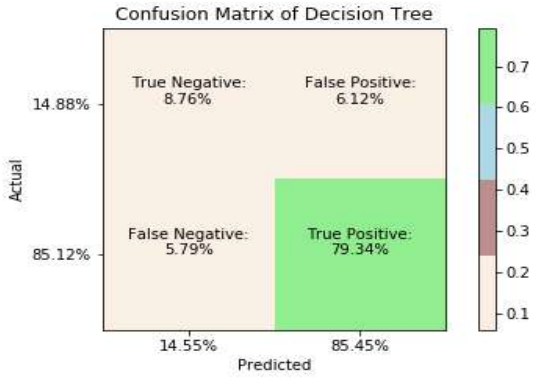


Fig. 9 Confusion Matrix of DT

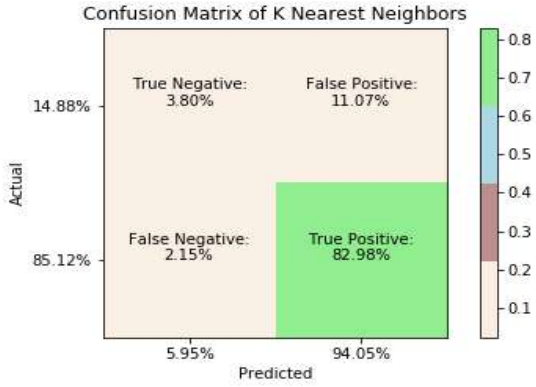


Fig. 10 Confusion Matrix of KNN

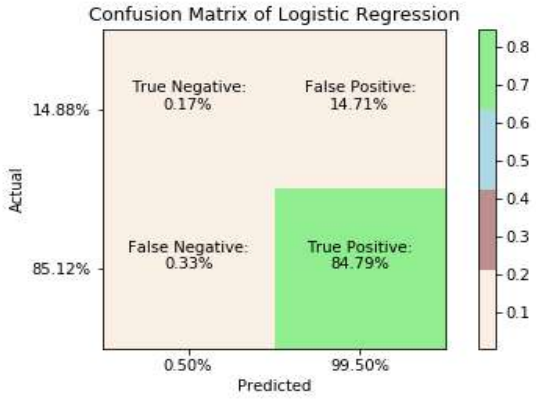


Fig. 11 Confusion Matrix of LR

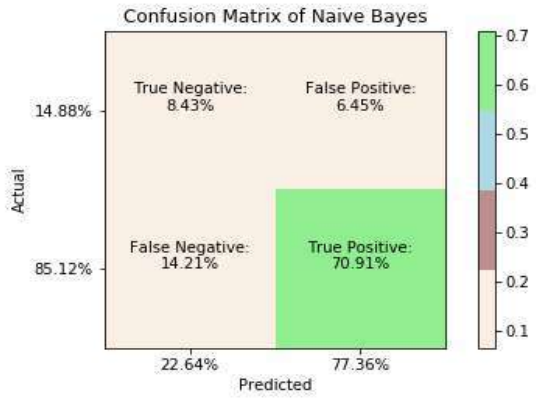


Fig. 12 Confusion Matrix of NB

$$Accuracy = \left(\frac{TP + TN}{TP + TN + FP + FN}\right) \qquad (10)$$

Accuracy of a classifier is the ratio of summation of TP and TN to total number of samples or instances. Accuracy is also called as recognition rate which specifies the proportion of total samples that are correctly identified. Misclassification rate is the difference between 1 and recognition rate.

Fig. 9, Fig. 10, Fig. 11 and Fig. 12 show confusion matrices of ML classifiers: DT, KNN, LR and NB respectively. Out of total instances, there are 85.12 % instances are true instances and 14.88 % instances are false instances. Fig. 9 shows that 8.76 % instances are correctly predicted as false instances and 79.34 % instances are correctly predicted as true instances for DT. Fig. 10 shows that 3.80 % instances are correctly predicted as false instances and 82.98 % instances are correctly predicted as true instances for KNN. Fig. 11 shows that 0.17 % instances are correctly predicted as false instances and 84.79 % instances are correctly predicted as true instances for LR. Fig. 12 shows that 8.43 % instances are correctly predicted as false instances and 70.91% instances are correctly predicted as true instances for NB.

C. Classification Report

The classification report summarizes, gives the precision, recall, f1-score, and support for the model. Precision is a classifier's ability not to label a positive instance which is in fact negative. It is the percentage of predicted positive instances that are correctly predicted as true positives. It is the ratio of true positive values to the summation of true and false positive values. Precision is also called as Positive Predictive Value (PPV).

$$Precision = PPV = \frac{TP}{TP + FP} \qquad (11)$$

Recall is a classifier's ability to find all positive events. It is the ratio of true positive values to the summation of true positive and false negative values. Sensitivity also called the True Positive Rate (TPR), the recall, measures the proportion of actual positives that are correctly classified as true positives. In binary classification, recall of the positive category is also recognized as sensitivity and recall of the negative category is recognized specificity.

$$Sensitivity = TPR = \frac{TP}{TP + FN} \qquad (12)$$

$$Specificity = TNR = \frac{TN}{TN + FP} \qquad (13)$$

Specificity also called the True Negative Rate (TNR) measures the proportion of actual negatives that are correctly classified as true negatives. Another appraise is F1-score which is the harmonic mean of precision and recall such that greatest score is 1.0 and bad score is 0.0.

$$F1 - score = \frac{2 * Precision * Recall}{Precision + Recall} \qquad (14)$$

In general, F1-scores are lower than accuracy measures as they include precision and recall into their calculations but they can be used to evaluate classifier models, not universal accuracy. Support is the number of real class occurrences in the data set specified. The support does not alter between models, but diagnoses the evaluation procedure as an alternative.

Table 1: Classification Report of DT

	Precision	Recall	F1-Score	Support
FALSE	0.60	0.59	0.60	90
TRUE	0.93	0.93	0.93	515
Avg / Total	0.88	0.88	0.88	605

Table 2: Classification Report of KNN

	Precision	Recall	F1-Score	Support
FALSE	0.64	0.26	0.37	90
TRUE	0.88	0.97	0.93	515
Avg / Total	0.85	0.87	0.84	605

Table 3: Classification Report of LR

	Precision	Recall	F1-Score	Support
FALSE	0.33	0.01	0.02	90
TRUE	0.85	1.00	0.92	515
Avg / Total	0.77	0.85	0.79	605

Table 4: Classification Report of NB

	Precision	Recall	F1-Score	Support
FALSE	0.37	0.57	0.45	90
TRUE	0.92	0.83	0.87	515
Avg / Total	0.84	0.79	0.81	605

Table 1, Table 2, Table 3 and Table 4 show the classification report of DT, KNN, LR and NB respectively. Table 1 shows that TPR and PPV of a DT is 93 % each. Table 2 shows that TPR of a KNN is 97 % and PPV is 88%. Table 3 shows that TPR of a LR is 100 % and PPV is 85%. Table 4 shows that TPR of a NB is 83 % and PPV is 92%. System's major class of interest is TRUE class and task is to minimize False Positives as well as False Negatives as much as possible to get good performance of a model. There is a trade-off between PPV and TPR. Among four predictive models, DT gives good results for TPR as well as PPV.

D. ROC Curve

The ROC curve is formed by plotting the True Positive Rate (TPR) along Y axis and the False Positive Rate (FPR) along X axis at various threshold settings. The ROC curve is thus the sensitivity as a function of FPR. The model with TPR=1 and FPR=0 is called as perfect model. Area Under Curve (AUC) is applied in classification examination in order to find out which of the used models predicts the best results. Fig. 13, Fig. 14, Fig. 15 and Fig. 16 show ROC Curves of ML classifiers: DT, KNN, LR and NB respectively where dotted line indicates the random guessing with AUC=0.5. Curve below the dotted line indicates bad performance of a model and curve above the random guessing shows good performance of a model. Fig. 13 shows that AUC of a DT is 0.76. Fig. 14 shows that AUC of a KNN is 0.62. Fig. 15 shows that AUC of a LR is 0.50. Fig. 16 shows that AUC of a NB is 0.70. Among four predictive models, Decision Tree gives the maximum AUC.

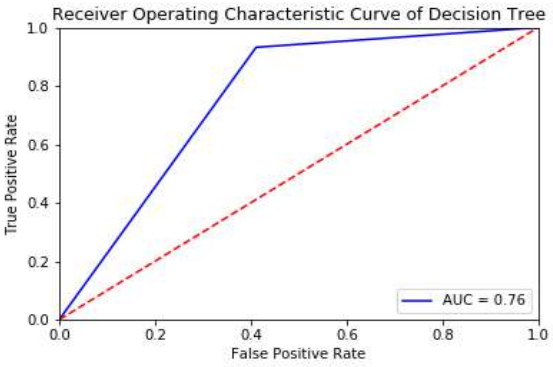


Fig. 13 ROC Curve of DT

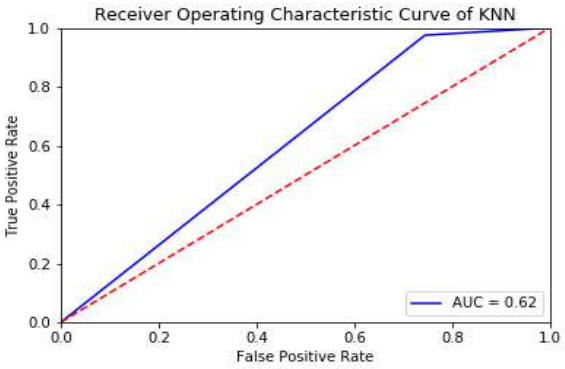


Fig. 14 ROC Curve of KNN

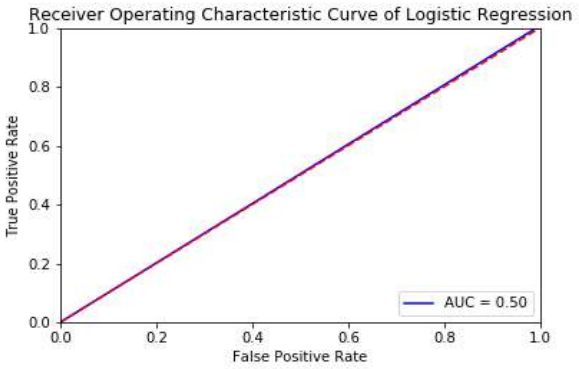


Fig. 15 ROC Curve of LR

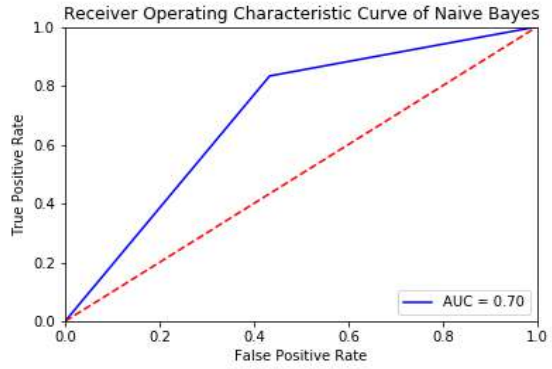


Fig. 16 ROC Curve of NB

E. Comparison and Selection of the Best ML Model

Fig. 17 shows that accuracy performance comparison of four ML models. In holdout method, the dataset is distributed into separate training and testing dataset - the former is used for creation of model and later is used to estimate its performance. As dataset applied belongs to imbalanced binary class, stratified k-fold cross validations is also used to compare and select the model effectively. In stratified k fold cross validation, k indicates number of folds which is equal to 10 and the type proportions are maintained in every fold to make sure that each fold is consultant of the category proportions in the training dataset. Among these four models, DT to be the best model for prediction of human face recognition with the highest accuracy of 87.77 % using hold out method and 89.81 % using stratified k-fold cross validation. Precision and recall of DT model is also very good as compared to other ML models. It is also observed that AUC of DT is the highest with 0.76 units. The model having curve nearest to the uppermost left area indicates the best performance. So DT is proved to be the best model with the highest classification accuracy, very good recall, precision, and highest AUC.

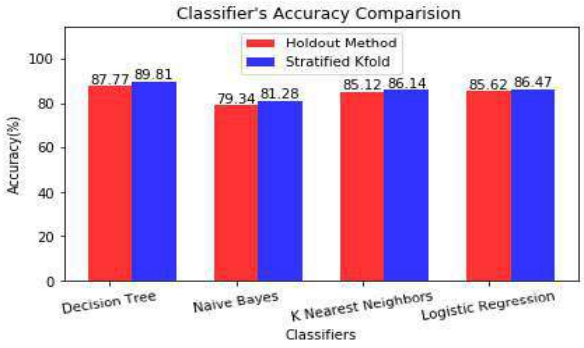


Fig. 17 Accuracy Comparison Chart of Four Classifiers

VI. CONCLUSION

Live sensor captured data along with multimedia data is useful for analysis of abnormal conditions in the environment of various physical locations. Sensor data analysis with the support of digital images of human face for human presence detection and recognition is useful for confirmation of abnormal conditions in the surroundings. As only sensitive information is pushed on to the cloud whenever the human presence is recognized either as known or unknown face, optimization of IoT enabled physical location monitoring is achieved because only the subset of original dataset is stored onto the cloud. The projected scheme is new and competent because it offers proper accuracies of classifiers and effectiveness of the approach. The Decision Trees, amongst the quite a number of analytical models, is a remarkable method for the evaluation of multimedia sensor records, with the maximum correctness of 87.77 % using hold out method and 89.81 % using stratified 10-fold cross validation, very good TPR and PPV of 0.93 each and ROC with AUC of 0.76, followed by NB, KNN and LR respectively. The prediction of person who is either known or unknown using sensor data analysis in the physical location, sending notifications and alert messages to mobiles and email accounts will be extended work of this system to enhance the robustness.

REFERENCES

1. Sankar Mukherjee , G.P. Biswas, "Networking for IoT and applications using existing communication technology", Egyptian Informatics Journal 19 (2018) 107–127.
2. Neelesh Mioshra, Lal Pratap Varma, Prabhat Kumar Srivastava, Ajay Gupta, "An Analysis of IoT Congestion Control Policies", Procedia Computer Science 132 (2018) 444–450
3. Dragos Mocrii, Yuxiang Chen, Petr Musilek, "IoT-based smart homes: A review of system architecture, software, communications, privacy and security", Internet of Things 1–2 (2018) 81–98
4. Adel Alkhalil, Rabie A. Ramadan, "IoT Data Provenance Implementation Challenges", Procedia Computer Science 109C (2017) 1134–1139.
5. Nallapaneni Manoj Kumar,Pradeep Kumar Mallick, " Blockchain technology for security issues and challenges in IoT"International Conference on Computational Intelligence and Data Science (ICCIDS 2018), Procedia Computer Science 132 (2018) 1815–1823.
6. Mustafa Alper Akkaş, Radosveta Sokullu, "An IoT-based greenhouse monitoring system with Micaz motes", International Workshop on IoT, M2M and Healthcare (IMH 2017), Procedia Computer Science 113 (2017) 603–608
7. Partha Pratim Ray, "A survey of IoT cloud platforms", Future Computing and Informatics Journal 1 (2016) 35-46.
8. Nawaz Mohamudally, Mahejabeen Peermamode-Mohaboob, " Building An Anomaly Detection Engine (ADE) For IoT Smart", The 15th International Conference on Mobile Systems and Pervasive Computing, Procedia Computer Science 134 (2018) 10–17.
9. Mohammad Saeid Mahdavinejad , Mohammadreza Rezvan, Mohammadamin Barekatain , Peyman Adibi, PayamBarnaghi, Amit P. Sheth, "Machine learning for internet of things data analysis: a survey", Digital Communications and Networks 4 (2018) 161–175.
10. Bill Karakostas, "Event prediction in an IoT environment using naïve Bayesian models", The 7th International Conference on Ambient Systems, Networks and Technologies (ANT 2016), Procedia Computer Science 83 (2016) 11 – 17.
11. Huseyin Yildirim, Amr M.T. Ali-Eldin, "A model for predicting user intention to use wearable IoT devices at the workplace", Journal of King Saud University-Computer and Information Sciences (2018)
12. Ajitkumar S. Shitole, Dr. Kamatchi R. Iyer, "SMART HOME CONTEXT-AWARE AUTOMATION BY CUSTOMIZATION STRATEGY", International Journal of Advanced in Management, Technology and Engineering Sciences(pp. 277-283), April 2018, ISSN NO: 2249-7455.
13. Devare M. (2018). Analysis and Design of IoT Based Physical Location Monitoring System. In Lucio Grandinetti, Seyedeh Leili Mirtaheri, Reza Shahbazian, Thomas Sterling, Vladimir Voevodin (Eds.), Advances in Parallel Computing,Volume 33: Big Data and HPC: Ecosystem and Convergence (pp. 120 - 136). IOS Press. doi: 10.3233/978-1-61499-882-2-120.
14. Ajitkumar S. Shitole, Manoj Devare, "Machine Learning Supported Statistical Analysis of IoT Enabled Physical Location Monitoring Data", International Conference On Computational Vision and Bio Inspired Computing, Nov 2018.

15. Alexandra Moraru, Marko Pesko, Maria Porcius, Carolina Fortuna and DunjaMladenic,"Using Machine Learning on Sensor Data", Journal of Computing and Information Technology - CIT 18, 2010, 4, 341–347 doi:10.2498/cit.1001913, 341.
16. Joseph Siryani, Bereket Tanju, and Timothy Eveleigh,"A Machine Learning Decision-Support System Improves the Internet of Things', Smart Meter Operations", IEEE 2017.
17. Purnendu Shekhar Pandey, "Machine Learning and IoT for Prediction and Detection of Stress", 978-1-5386-3893-4/17/$31.00 ©2017 IEEE.
18. Go Takami, Moe Tokuoka, Hirotsugu Goto, Yuuichi Nozaka, "Machine Learning Applied to Sensor Data Analysis ",Yokogawa Technical Report English Edition Vol. 59 No. 1(2016), pp. 27-30.
19. Rui Madeira, Luís Nunes, "A Machine Learning Approach for Indirect Human Presence Detection Using IOT Devices", The Eleventh International Conference on Digital Information Management 2016, pp. 145-150.
20. Che-Min Chung, Cai-Cing Chen, Wei-Ping Shih, "Automated Machine Learning for Internet of Things", 2017 IEEE International Conference on Consumer Electronics - Taiwan (ICCE-TW) pp. 295-296.

AUTHORS PROFILE

Ajitkumar S. Shitole is pursuing Ph.D. in CSE from Amity University Mumbai. He has published more than 21 research papers in various International Journals/ Conferences. His area of interest is Data Mining, Machine Learning, and Algorithms. Currently he is working as an Associate Professor in Computer Engineering Department at I^2IT Hinjawadi, Pune.

Dr. Manoj Devare currently holding position of Associate Professor at Amity Institute of Information Technology, Amity University Mumbai. He has been served as Post Doctorate Fellow at Centre of Excellence on HPC, University of Calabria, Italy. His research deals with Multi-Scale High Performance Computing in Grids, Virtualization, and Clouds. He is also involved in the review process of the International Journals papers, edited chapters from the reputed journals like Elsevier's Future Generation Computer Systems (FGCS), Springer's review system, International Journal of Computing, IEEE Transactions on Parallel and Distributed Systems (TPDS), and SCIT International Journal.

Astrophys Space Sci (2018) 363:182
https://doi.org/10.1007/s10509-018-3403-1

ORIGINAL ARTICLE

Modeling of local ionospheric time varying characteristics based on singular value decomposition over low-latitude GPS stations

Raghavendra Neeli[1] · J R K Kumar Dabbakuti[2,3] · V. Rajesh Chowdhary[1] · Nitin K. Tripathi[1] · Venkata Ratnam Devanaboyina[3,4]

Received: 1 February 2018 / Accepted: 27 July 2018 / Published online: 8 August 2018

Abstract Singular Value Decomposition (SVD) model is implemented to recognize the Total Electron Content (TEC) time series of daily, temporal as well as seasonal characteristics throughout the 24th solar cycle period of the year 2015 in the study. The Vertical (vTEC) analysis has been carried out with Global Positioning System (GPS) data sets collected from five stations from India namely GNT, Guntur (16.44° N, 80.62° E), and IISC, Bangalore (12.97° N, 77.59° E), LCK2, Lucknow (26.76° N, 80.88° E), one station from Thailand namely AITB, Bangkok (14.07° N, 100.61° E), and one station from South Andaman Island namely PBR, Port Blair (11.43° N, 92.43° E), located in low latitude region. The first five singular value modes constitute about 98% of the total variance, which are linearly transformed from the observed TEC data sets. So it is viable to decrease the number of modeling parameters. The Fourier Series Analysis (FSA) is carried out to characterize the solar-cycle, annual and semi-annual dependences through modulating the first three singular values by the solar (F10.7) and geomagnetic (Ap) indices. The positive correlation coefficient (0.75) of daily averaged GPS–TEC with daily averaged F10.7 strongly supports the temporal variations of the ionospheric features depends on the solar activity. Further, the significance and reliability of the SVD model is evaluated by comparing it with GPS–TEC data and the standard global model (Standard Plasma-Spherical Ionospheric Model, SPIM and International Reference Ionosphere, IRI 2016).

Keywords Ionosphere · GPS · SVD · FSA · TEC · IRI · SPIM

✉ V.R. Devanaboyina
dvratnam@kluniversity.in

[1] School of Engineering and Technology, RS&GIS Field of Study, Asian Institute of Technology, P.O. Box 4, Klong Laung, Pathumthani 12120, Thailand

[2] Department of ECM, KLEF, K L University, Vaddeswaram, Guntur Dt, 522502, Andhra Pradesh, India

[3] Department of ECE, KLEF, K L University, Vaddeswaram, Guntur Dt, 522502, Andhra Pradesh, India

[4] Department of Atmospheric Science, KLEF, K L University, Vaddeswaram, Guntur Dt, 522502, Andhra Pradesh, India

1 Introduction

Ionosphere is an essential domain of near-Earth's space environment, where adequate ionization can exist and influence the trans-ionospheric radio-wave propagation (Dabas 2000). The complex variations in the ionospheric structure in time and space lead to degradation of satellite-based navigation, positioning, and communication system performances (Jakowski et al. 2001). Global Positioning System (GPS) signals propagate through the ionospheric medium and carry the signatures regarding Total Electron Content (TEC). The propagation speed of the GPS signal depends on the number of free electrons. TEC is defined as the number of electrons in cube of 1 m^2 cross section extending from the receiver to the satellite (Kaplan and Hegarty 2006). The day to day variations in TEC depend on solar ionizing radiation, solar radio bursts and the interaction between the ionosphere, thermosphere and magnetosphere, interplanetary conditions, geomagnetic activity, and some events like solar eclipses (Kumar et al. 2015). In particular, the estimates of TEC across the low/equatorial latitude region is a crucial task because of the large spatio-temporal gradients associated with the associated local electrodynamic parameters, like Equatorial Ionization Anomaly (EIA) and Equatorial Electrojet (EEJ) strength, etc. (Appleton 1946). EIA

is characterized as two ionization crests at about 15° magnetic latitudes on both sides of the geomagnetic equator and a trough over the geomagnetic equator. Therefore, a method is essential to model the ionospheric TEC variations for a reliable GPS positioning performance at low-latitudes.

In grid based regional/global ionosphere modelling various interpolation techniques are used, for example linear/multiquadric interpolation as suggested by Skone (1998). Another examples are statistically based spatial interpolation techniques like kriging which can be used to develop a global ionospheric model (Ansari et al. 2018). The researchers have been improve the ionospheric delay predictions from several empirical ionospheric models like Kriging, Auto Regressive Moving Average (ARMA) and Holt–Winter methods (Sivavaraprasad and Ratnam 2017; Ansari et al. 2017). Based on the GPS–TEC observations, earlier researchers have proposed empirical models to fulfill the scientific and technological application needs (Bilitza et al. 2008). Several ionospheric regional models have been suggested by researchers over various geographical regions, such as, Spherical Cap Harmonic Analysis (SCHA) model over Australia (Bouya et al. 2010), Empirical Orthogonal Function (EOF) model over China (Mao et al. 2008; Yu et al. 2015), Singular Value Decomposition (SVD) model over Denmark (Jakobsen et al. 2010), EOF model over Japan (A et al. 2011), Neural Network (NN) method over Southern Africa (Habarulema et al. 2011), EOF over North America (Chen et al. 2015), Mathematical model over Turkish (Ansari et al. 2017).

Concerning low latitude and equatorial region over Indian sub-continent, several studies have been carried out to attribute the ionospheric TEC variations to the solar-terrestrial effects, geomagnetic activity and equatorial ionization anomaly (EIA) phenomena (Bagiya et al. 2009; Kumar et al. 2012; Dashora and Suresh 2015). Further, numerous researcher studies has been carried investigated GP-based ionospheric TEC/Electron Density Profiles (EDP) variations and compared with the IRI model estimations over low latitude region, Thailand (Kenpankho et al. 2011; Arunpold et al. 2014; Chowdhary et al. 2015; Jamjareegulgarn et al. 2017). These studies complement/supplement the climatological improvements in the global empirical International Reference Ionosphere (IRI) model. Most of these papers from Indian context describe that TEC analysis by ground-based GPS TEC observations. It is necessary to model TEC response over the low latitude regions.

Hence, the present study an attempt to characterize and model the ionospheric variations of GPS–TEC over the low latitude EIA regions ranging from 13.02° N to 16.37° N latitude (Bangalore, AITB, and KLU-Guntur, Lucknow, Port Blair) during the solar maximum phase (2015). In this paper, we implemented an approach with SVD decomposition technique. The decomposed SVD modes consist of Singular values U and V those reproduce the daily, seasonal, and

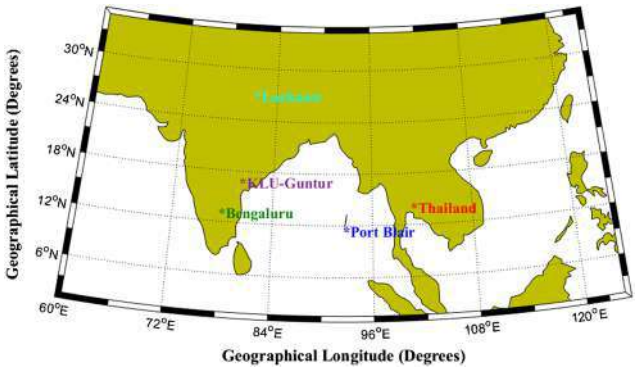


Fig. 1 GPS stations locations for low-latitude ionospheric variability TEC data analysis

yearly ionospheric variations. Additionally, the Fourier Series Analysis (FSA) has been performed to modulate the first three SVD modes of GPS–TEC by the solar (F10.7) and geomagnetic (Ap) indices to replicate the solar-cycle, semi-annual and annual variations. The study also includes a comparative analysis of the observed TEC with the model TEC values (SVD, IRI2016, and SPIM) to demonstrate the reliability of the SVD approach for portraying the low latitude ionospheric variations.

2 Data analysis

The experimental TEC data from dual-frequency GPS receivers is considered to carry out the investigation of ionospheric variability and modeling using SVD method. The geomagnetic and geographic coordinates of the GPS stations under the investigation are listed in Table 1. The GPS–TEC study has been conducted of the year 2015, which is the solar decreasing peak of the 24th solar cycle.

The study area covers five GPS stations in the low latitude ionosphere with the geographical extent from 13° N to 26° N latitudes and 77° E to 100° E longitudes as shown in the Fig. 1. The analysis of ionospheric TEC data is performed during 24th solar-cycle for the year 2015. The GPS station at Asian Institute of Technology, Thailand (AITB 14.07° N, 100.61° E) belongs to SCINDA network consists of Novatel GPS receiver (GSV40004B). Another permanent GPS station is at KLU-Guntur (16.37° N, 80.37° E) comprises of a Novatel GPStation-6 receiver directly providing the slant TEC (sTEC) data with a sampling rate of 1 sec. The other GPS station at Bangalore (13.02° N, 77.58° E), Lucknow (26.76° N, 80.88° E) and Port Blair (11.43° N, 92.43° E) is under IGS network, observables from which freely accessible at the archives of Scripps Orbit and Permanent Array Center (SOPAC) and website is noted as http://sopac.ucsd.edu. The sTEC from KLU Guntur is transformed to its vertical equivalent (vTEC) through a mapping function, whereas the

Table 1 Detail of GPS stations

S. No.	GPS stations	Geographical		Geomagnetic	
		Latitude (N)	Longitude (E)	Latitude (N)	Longitude (E)
1	Port Blair	11.43°	92.43°	2.03°	165.25°
2	Bangalore	13.02°	77.57°	4.40°	150.77°
3	AITB	14.07°	100.61°	6.49°	172.54°
4	KLU-Guntur	16.37°	80.37°	7.50°	153.76°
5	Lucknow	26.76°	80.88°	17.98°	155.22°

vTEC at other stations are estimated from the GPS observables by using the GPS–TEC analysis program (Seemala and Valladares 2011). Further, the geomagnetic activity (Ap) and solar activity (F10.7) indices have been taken from (http://omniweb.gsfc.nasa.gov/form/dx1.html). The IRI model latest version (IRI 2016) can be accessed online through the website (https://omniweb.gsfc.nasa.gov/vitmo/iri2016_vitmo.html) whereas the subroutines of SPIM model for the modeled vTEC data is retrieved at ftp://ftp.izmiran.rssi.ru/pub/izmiran/SPIM.

3 Methodology

SVD is a method of decomposing a matrix into a product of three-elements matrices. The elements represents the recognizable patterns when the data set resembles a samples of continuous variables (Preisendorfer and Mobley 1988).

For a given matrix $m \times n$, $m \geq n$, the matrix X can be written as (Parker 2004).

$$X = USV^T \tag{1}$$

where U represents the $m \times m$ matrix, V represents the $n \times n$ matrix, S represents the diagonal matrix of size $m \times n$. U and V are orthogonal matrices. S is a nonnegative values diagonal matrix, called the singular values of X.

The GPS–TEC measurements of X are ordered into 24 hours $\times$ 365 days. The rows of X matrix correspond to a day of the year (DOY) ($d = 1, 2, \ldots, 365$) and columns correspond to Hour of the Day (HPD) ($h = 1, 2, \ldots, 24$). The U matrix will correspond to the daily variation and the V matrix will correspond to the yearly variation. The SVD is done using the LAPACK routine (Anderson et al. 1999).

Further, singular values of V are modeled using the FSA, to figure out the impact of solar and geomagnetic activities on GPS–TEC, and their relative contributions is expressed as (A et al. 2012).

$$V_k(d) = Sol_{k1}(d) + Annual_{k2}(d) + Semiannual_{k3}(d) + \varepsilon \tag{2}$$

where $Sol_{k1}(d)$ indicate the solar activity influence, $Annual_{k2}(d)$ and $Semiannual_{k3}(d)$ are the periodic component at two different frequencies.

The solar activity component can be determined by daily averaged solar and geomagnetic activity indices (F10.7 and Ap).

$$Sol_{k1}(d) = u_1 + u_2 F_{10.7p}(d) + u_3 Ap(d) \tag{3}$$

The amplitude and phase of each periodic component can be attained via the sine and cosine terms. The periodic oscillations at different frequencies 182.6 and 365.25 represent the Semi-Annual Oscillations (SAO) and Annual Oscillations (AO). It can be determined by daily averaged (F10.7 and Ap) indices via the sine and cosine terms.

$$\begin{aligned} &Annual_{k2}(d) \\ &= \left(u_4 + u_5 F_{10.7}(d) + u_6 Ap(d)\right) \cos 2\pi d/365.25 \\ &\quad + \left(u_7 + u_8 F_{10.7p}(d) + u_9 Ap(d)\right) \sin 2\pi d/365.25 \end{aligned} \tag{4}$$

$$\begin{aligned} &Semiannual_{k3}(d) \\ &= \left(u_{10} + u_{11} F_{10.7p}(d) + u_{12} Ap(d)\right) \cos 2\pi d/183.6 \\ &\quad + \left(u_{13} + u_{14} F_{10.7p}(d) + u_{15} Ap(d)\right) \sin 2\pi d/182.6 \end{aligned} \tag{5}$$

The unknown coefficients $u_1, u_2, \ldots, u_{15}$ are determined by the least squares method (Menke 1984). The deviations among the observed GPSTEC and the predicted model TEC values are determined by its residuals. The deviations are analyzed by correlation coefficients, Root Mean Square Deviations (RMSD) and the percentage RMSD.

4 Results and discussions

The contour maps of diurnal ionospheric variations of TEC over Bangalore, AITB, KLU-Guntur stations are shown in Fig. 2. It is delineated from the Fig. 2 that the diurnal level of TEC exhibits a steady increase from morning to reach the peak value during the local afternoon and dropped thereafter to attain the day-minimum value in the post-midnight period. During the maximum solar period, we observe that the peaks of diurnal vTEC exists in-between 6:00 to 14:00 UT; and reaches a minimum value at night around 20:00 UT.

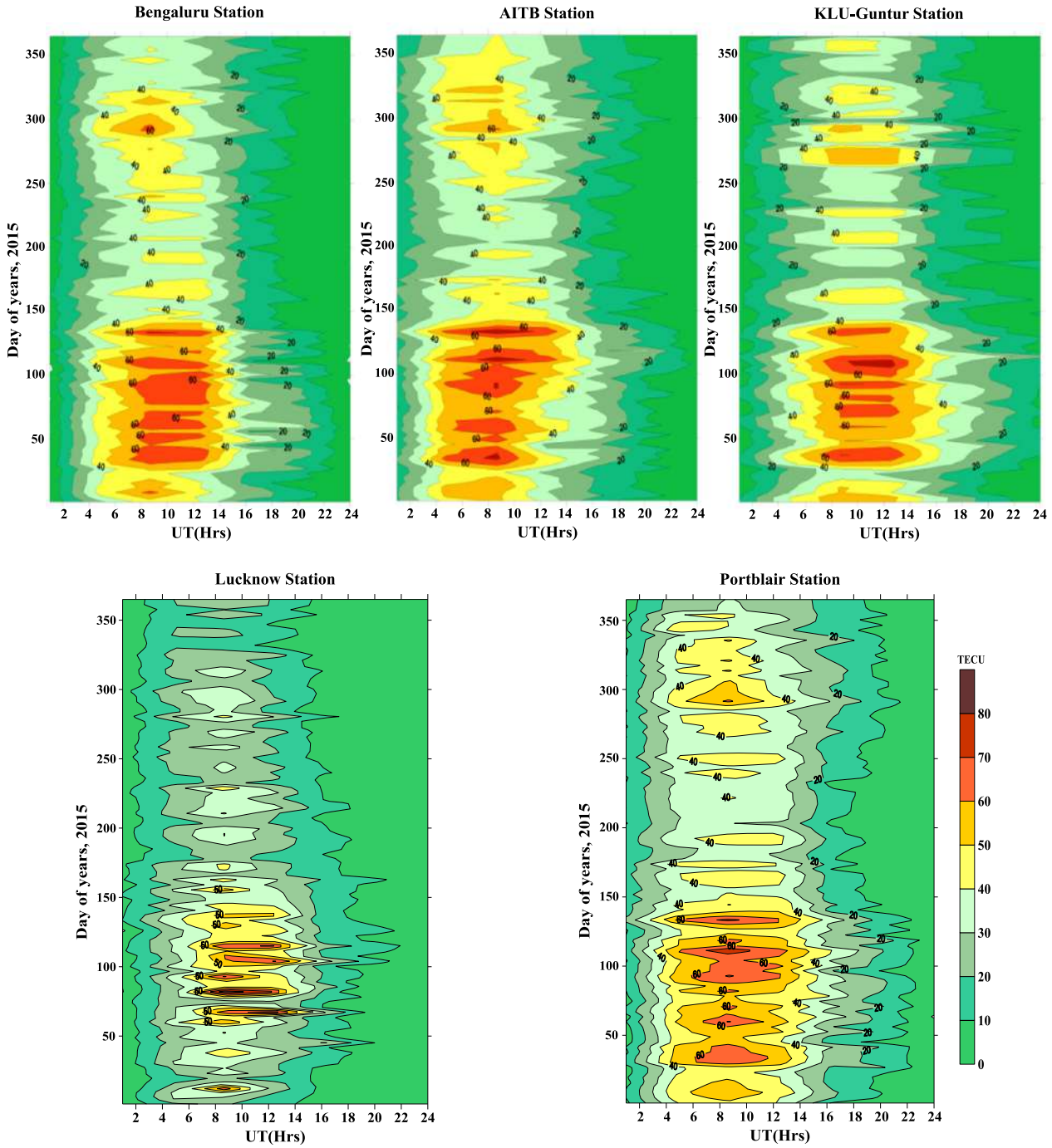


Fig. 2 The vTEC diurnal variations for five stations during the year 2015

From Fig. 2 we detect that ionospheric vTEC variations are maximum for the period 8:00 to 13:00 UT over Bangalore, 5:00 to 12:00 UT over AITB, 8:00 to 14:00 UT over KLU-Guntur, which shows the TEC non-linear spreading (Fig. 2). During winter vTEC reaches it's peak value (8:00 UT) in prior to the peak in summer (13:00 UT) and equinox (14:00 UT) months. vTEC exhibits a normal trend of the day to day deviation of a minimum in the early hours (05:00 to 07:00 UT) and maximum during the afternoon (13:00 to 14:00 UT). The reason may be the rotation of the Earth about the sun and its rotation in an elliptical orbit that affects the ionospheric variations. The axis of the earth is inclined at 66.50 to the Earth's orbit surface, and the length of the day and night changes (Panda et al. 2015). The position of the Earth around the sun determines the season and amount of solar energy that the two hemispheres

Table 2 Outline of the variances through SVD-decomposition for the five GPS stations by the first five singular values

Singular values	Bangalore (13.02° N, 77.57° E)		AITB (14.07° N, 100.61° E)		KLU-Guntur (16.44° N, 80.62° E)		Lucknow (16.44° N, 80.62° E)		Port Blair (16.44° N, 80.62° E)	
	Variances (%)	CV (%)	Variances (%)	CV (%)	Variances (%)	CV (%)	Variances (%)	CV (%)	Variances (%)	CV (%)
1	94.32	94.32	95.02	95.02	95.14	95.1	91.21	91.2	95.08	95.08
2	2.64	96.96	1.85	96.86	1.93	97.1	4.70	95.9	1.98	97.06
3	1.17	98.14	1.30	98.16	1.01	98.1	1.36	97.3	0.99	98.05
4	0.67	98.81	0.57	98.73	0.57	98.7	0.94	98.2	0.72	98.76
5	0.30	99.11	0.39	99.12	0.35	99	0.44	98.7	0.33	99.09

received. The effects of sunrise and sunset will cause the generation (photoionization) and disappearance (recombination) of electrons in the ionosphere, this can be observed from contour plot (Fig. 2).

The production rate of ionospheric electrons during winter reaches its peak at earlier hours due to the shorter duration of the daytime compared to summer (Fig. 2). The annual ionospheric TEC pattern variations over Bengaluru, AITB, and KLU-Guntur. From Fig. 2, it is also observed that the magnitude of absolute TEC at the five stations are comparable with fewer inconsistencies during the pre-sunrise and post-midnight hours and more deviations during early afternoon in all seasons. The early afternoon discrepancies are attributed to the apparent increase in solar radiation from morning to noon and related equatorial electrodynamics associated with $E \times B$ drifts, Fountain effect and development of EIA over the low latitude to consequence non-linear distribution of electrons from dip equator to farther latitudes (Panda et al. 2015). The maximum TEC value has been found in March and September equinoxes and the minimum values are observed during June and December solstices. It could be associated with the pre-reversal enhancement (PRE) over the low-latitude regions (Appleton 1946). The fraction of the variance and its cumulative-variances for the first six modes are listed in Table 2. The variance for the first five singular values at Bangalore (94.32%, 2.64%, and 1.17%), AITB (95.02%, 1.85% and 1.30%), KLU-Guntur (95.14%, 1.93%, 1.01%), Lucknow (91.21%, 4.70% and 1.36%) and Port Blair (95.08%, 1.98%, 0.99%), confirm that first five modes signify more than 98% of total variance in TEC variations.

In this present study, 5 SVD modes are considered, which account for more than 98% of the total variance in the original data set were retained to develop the SVD model and the contribution of each component is shown in Table 1. The variations in the outputs are found to be imitating the variation pattern of original data sets shown in Fig. 3 and it is decomposed into Hour of the Day (HOD) denoted as U and Day of the Year (DOY) variations denoted as V. The first five singular values of (U_1 to U_5) in the left panel and their corresponding coefficients (V_1 to V_5) in the right panel for all the five stations (Fig. 3). The first singular value U_1 clearly represents the diurnal variation pattern of TEC whereas the V_1 is analogs to the variation of solar activity level during the period which is illustrated from Fig. 4. This agrees well that the TEC variation is precisely following the 24th solar cycle phases (2015). Ionospheric TEC is also affected by other factors including interhemispheric flow, neutral wind and diffusion. In this way, it is expected that small disturbances and irregularities due to the above-mentioned influences, which are often represented by the change of U_2 and U_3 on the left panel, overlap with diurnal variation. U_2 shows a semi-annual variation pattern, which is related to the annual variation from summer to winter. This is clearly evident from the second relevant coefficient V_2 (Fig. 3), which mainly includes the annual component. It can be noticed that the characteristic of the important equatorial ionospheric anomaly may be related to the pre reversal enhancement (PRE) effect in U_3. The third singular value V_3 represents semiannual variations and solar activity dependence (Fig. 3). The PRE phenomenon is mainly due to the generation of the dynamo electric field in the F region under post sunset decay of the electric conductivity (Abdu 2005; Fejer 2011).

Figure 4 shows the comparison between solar-index (F10.7) and the first singular value V_1 at five GPS stations. It is clearly seen from Fig. 4 that the first Singular value V_1 is well-correlated with the F10.7 indicating the ionospheric variations due to the solar activity. It is verified that V_1 variation is related to the solar activity which is the leading factor that restraints the TEC variability. Moreover, the contribution of the first order Singular value is more than 91% for five GPS stations. This clearly reveals that the solar activity is the leading factor that restraints the TEC variability.

After TEC decomposition singular value V were then modeled in terms of Ap and F10.7 indices using linear regression analysis, in order to introduce the influence of solar and geomagnetic activities on TEC into the model (A et al. 2012). By using the formal Fourier series according to the Eq. (2) to (5), the solar cycle, annual and semi-annual variations of each order coefficients from the five

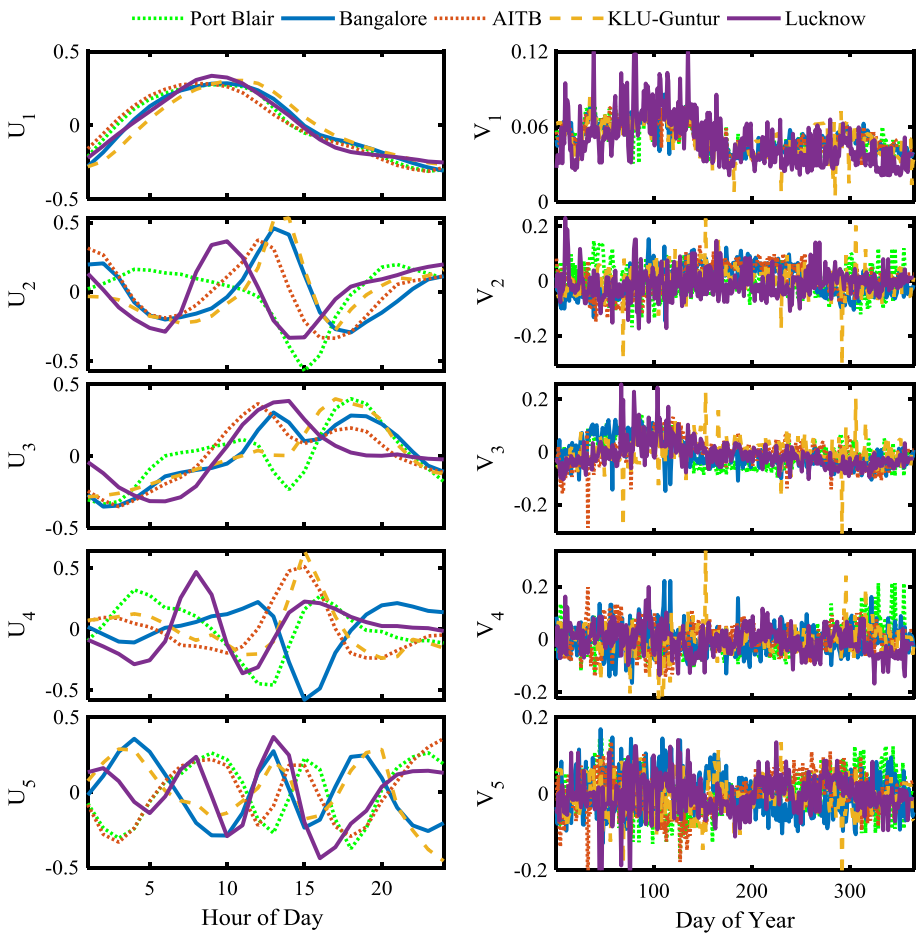


Fig. 3 First five singular values of U (left panel) and V (right panel) matrices of five GPS stations for the year 2015

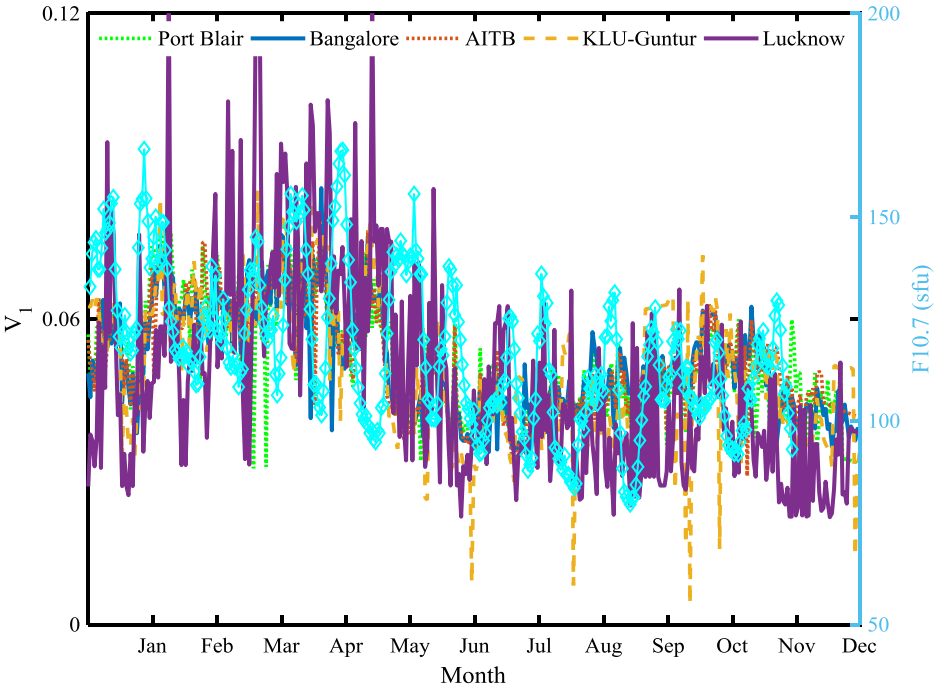


Fig. 4 Comparison of solar index (F10.7) and GPS–TEC at five GPS stations for the first singular value V_1

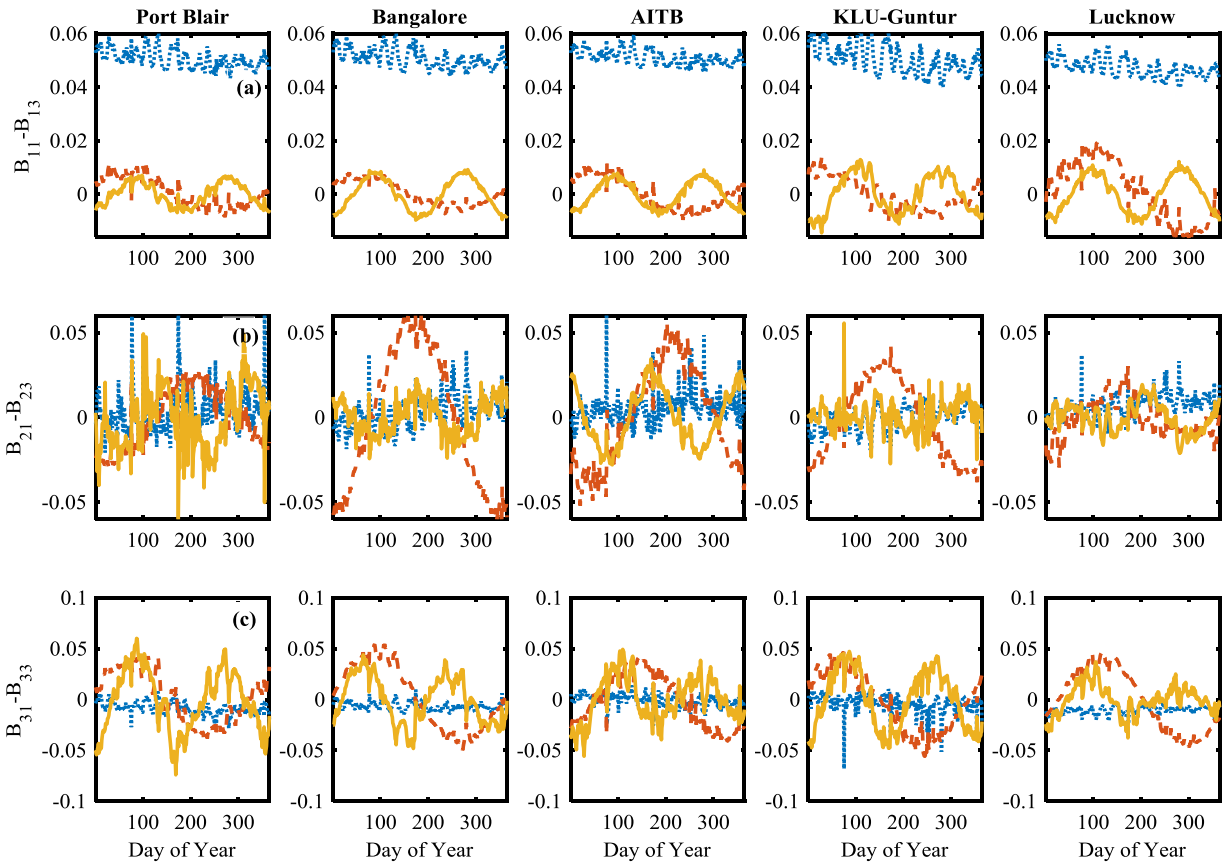


Fig. 5 Modeled mode coefficients at five GPS based on FSA during the year 2015

GPS stations are plotted in Fig. 5. From Fig. 5a, top panel (blue colour) $B_{11} - B_{13}$, observed that the variation to the solar cycle contribution is primarily obtained by the first-order component than that of other components. The contribution of second order component middle panel (red colour) $B_{21} - B_{23}$, represents the annual variation (Fig. 5b). The semi-annual variation is more prominent from third components ($B_{31} - B_{33}$) and less prominent in the first component ($B_{11} - B_{13}$) (Fig. 5c).

Figures 6a–6e show comparison of observed GPS–TEC with SVD, IRI-2016 and SPIM models at five GPS stations respectively. The monthly mean GPS–TEC and SVD model are having quiet similar trends throughout the year for five GPS stations. From Fig. 6a, it is noticed that the SPIM model overestimates SVD model and GPS–TEC values during all months expect the solictices months. The IRI model follows the patterns of GPS–TEC during summer solictices months and overestimates GPS–TEC and SVD model values during all the seasons. The IRI2016 model underestimates the SVD model and GPS–TEC values throughout the equinox months (March and April) and solstices months (January, February and May). The SPIM model overestimates GPS–TEC and SVD model values during the equinoctial months (September and October) and solstice months (June, July, August, November and December) at Bangalore GPS station. The residual error is less for SVD model, whereas IRI2016 model shows higher error when compared to SVD and SPIM models (Fig. 6b).

From Fig. 6c, it is noticed that the IRI2016 model underestimates SVD model and GPS–TEC values during the equational months (March and April) and solictices months (January, February and May). The SPIM model overestimates GPS–TEC and SVD model values during the equational months (September and October) and solictices months (June, July and August) at AITB GPS station. The residual error is less for SVD model, whereas IRI2016 model shows higher error when compared to SVD and SPIM models (Fig. 6c). From Fig. 6d it is noticed that the IRI2016 and SPIM model overestimates SVD model and GPS–TEC values during the all months in the 2015 year except in February. The residual error is less for SVD model, whereas IRI2016 and SPIM models show higher error when compared to SVD model (Fig. 6d). One significant characteristic is noticed that standard models of IRI and SPIM shows higher residuals at the KLU-Guntur GPS station which is located at the inner edge of EIA crest region and lesser residuals at AITB GPS station. Subsequently, the correlation coefficient, Root Mean Square Deviation (RMSD) and

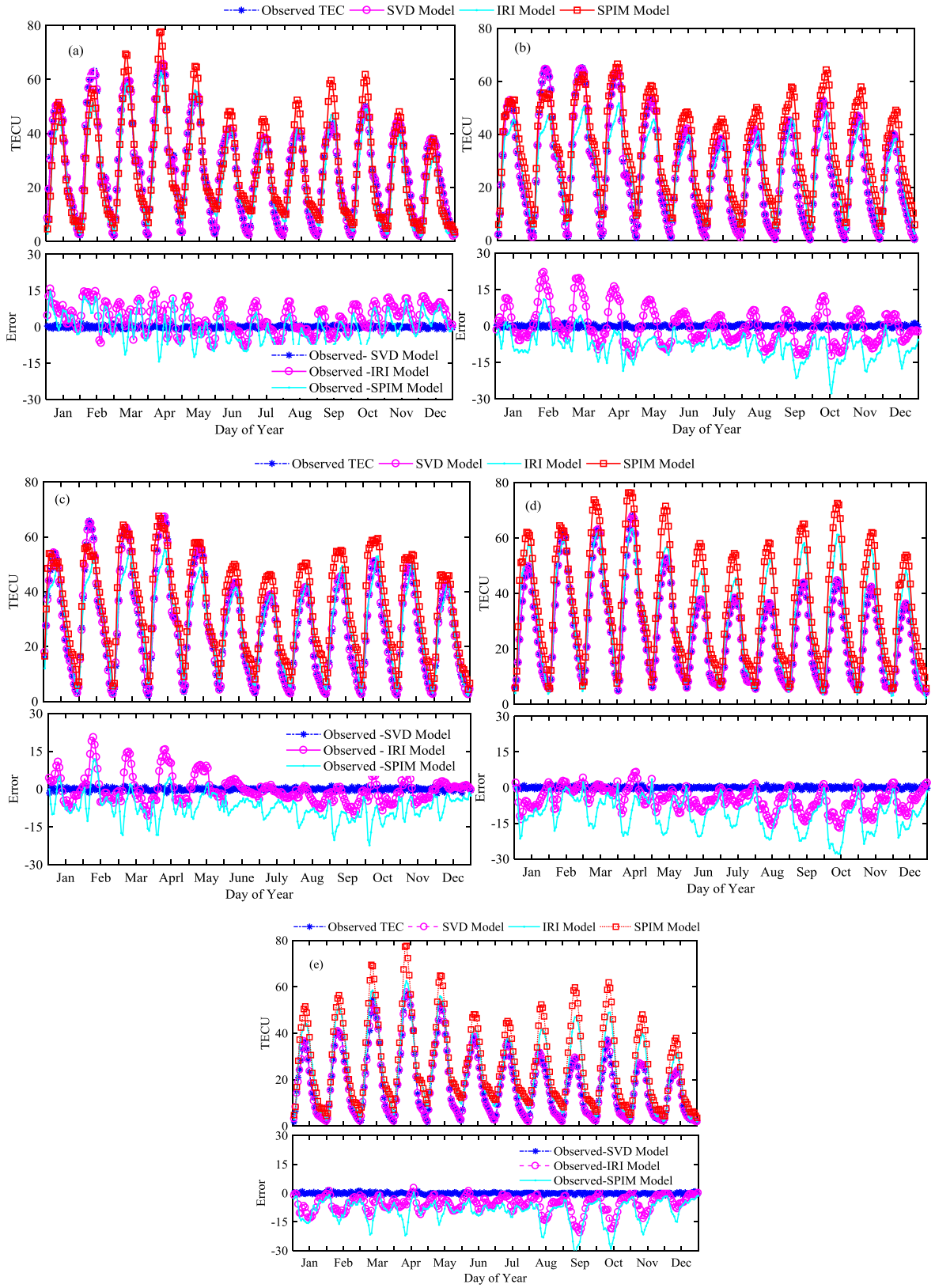


Fig. 6 Upper panel (**a**), (**b**), (**c**), (**d**) and (**e**): comparison of measured GPS–TEC with SVD and IRI2016, SPIM models at. Lower panel: errors between measured GPS–TEC and modeled TEC values

Table 3 Correlation coefficients and average root mean square deviation (RMSD) between models and observed TECS for five GPS stations

Station	Corr. coefficient			RMSD (TECU)			RMSD (%)		
	SVD model	IRI model	SPIM model	SVD model	IRI model	SPIM model	SVD model	IRI model	SPIM model
Bangalore	0.99	0.94	0.95	0.54	7.40	9.61	1.5	22.9	29.8
AITB	0.99	0.96	0.96	0.57	5.44	8.58	1.72	16.43	25.89
KLU-Guntur	0.99	0.96	0.95	0.71	6.91	12.78	2.00	22.63	41.87
Lucknow	0.99	0.96	0.95	0.63	7.21	11.31	2.78	31.78	49.05
Port Blair	0.99	0.94	0.93	0.47	6.73	6.34	1.48	20.97	19.74

percentage of RMSD of observed TEC and model TEC values are determined to check the validity of the SVD model.

From Fig. 6e, it is noticed that the IRI2016 model overestimates SVD model and GPS–TEC values during all months expect the solictices months (June and July). The SPIM model overestimates GPS–TEC and SVD model values during all the months at Lucknow GPS station. The residual error is less for SVD model, whereas SPIM model shows higher error when compared to SVD and IRI models (Fig. 6e).

Table 3 illustrates the Correlation coefficients and Root Mean Square Deviation (RMSD) between GPS–TEC and models for five GPS Stations. RMSD and RMSD percentage are given in the Table 3. It was noticed that RMSD between the observed TEC versus SPIM TEC model has shown higher values than observed versus SVD model and Observed versus IRI models. The lower linearity (higher RMSE) of with standard models (IRI and SPIM models) with SVD model is probably due to the discrepancy in estimated TEC over the low Correlation coefficient between observed versus SVD model shows higher linearity value of 0.99 whereas IRI2016 and SPIM model shows relatively lower linearity (0.94 and 0.95). From Table 3, it can be seen that the RMSD values increases from Bangalore towards Lucknow station just above the anomaly crest region. The latitude of crest development varies with EEJ strength and the season of the year. As the Lucknow station is on above the EIA crest, the TEC abnormal variation is noticeable and the day maximum also occurs slightly later than near equator stations, i.e., Bangalore and KLU-Guntur. This is probably due to the time lag between EEJ strength and maximum intensity of fountain effect.

The seasonal variations are categorized into four types: March and April are grouped to be Autumn season, September and October are grouped as Spring season; November, December, January and February are grouped to be winter season; May, June, July and August are combined to represent the summer season. From Fig. 7a, it is noticed that the IRI2016 model underestimates SVD model and GPS–TEC values at noon hours and overestimates at 1500 to 2400 hrs during the all the seasons, whereas SPIM model overestimates at 1500 to 2400 hrs during the all the seasons and follows the trends of GSP–TEC and SVD model during the winter and autumn seasons. From Fig. 7b it is noticed that the IRI2016 model underestimates SVD model and GPS–TEC values at noon hours during the autumn and winter seasons and overestimates at 1500 to 2400 hrs during the autumn and spring seasons. Whereas SPIM model overestimates at $\sim$ (0100 to 0500) hrs during the all the seasons and follows the trends of GSP–TEC and SVD model during all the seasons except spring season. From Fig. 7c it is noticed that the IRI2016 and SPIM model overestimates SVD model and GPS–TEC values at all the hours during all seasons except in autumn season at 0100 to 2400 hrs. From Fig. 7d it is noticed that the IRI2016 model overestimates SVD model and GPS–TEC values during all seasons and follows at 1800 to 2400 hrs during the winter and spring seasons. Whereas SPIM model overestimates during the all the seasons. From Fig. 7e it is noticed that the IRI2016 underestimates SVD model and GPS–TEC values at all the hours during all seasons except at 0100 to 0400 hrs. Whereas SPIM model overestimates, TEC values during 0100 to 0400 hrs.

Table 4 illustrates the seasonal deviations of the observed GPS–TEC values with Model values during the period. It can be seen that the GPS–TEC versus modeled-SVD shows a lower RMSD and percentage of RMSD when compared to IRI2012 and SPIM models at five GPS stations. The deviations of IRI and SPIM are lower at AITB station and higher at KLU-Guntur station. One more significant feature is noticed during HSA period (2015) that the error deviation of SVD model is higher in winter season as compared to the summer season. The phenomenon is similar to the seasonal ionospheric anomaly and could be attributed to the strong summer to winter neutral spread in solar maximum.

5 Conclusions

In the present study we present a TEC model based on the SVD analysis using the GPS–TEC during the year 2015 with corresponding IRI and SPIM model values at low latitude stations. The performance of SVD model is compared with IRI and SPIM models.

The following are the key observations are.

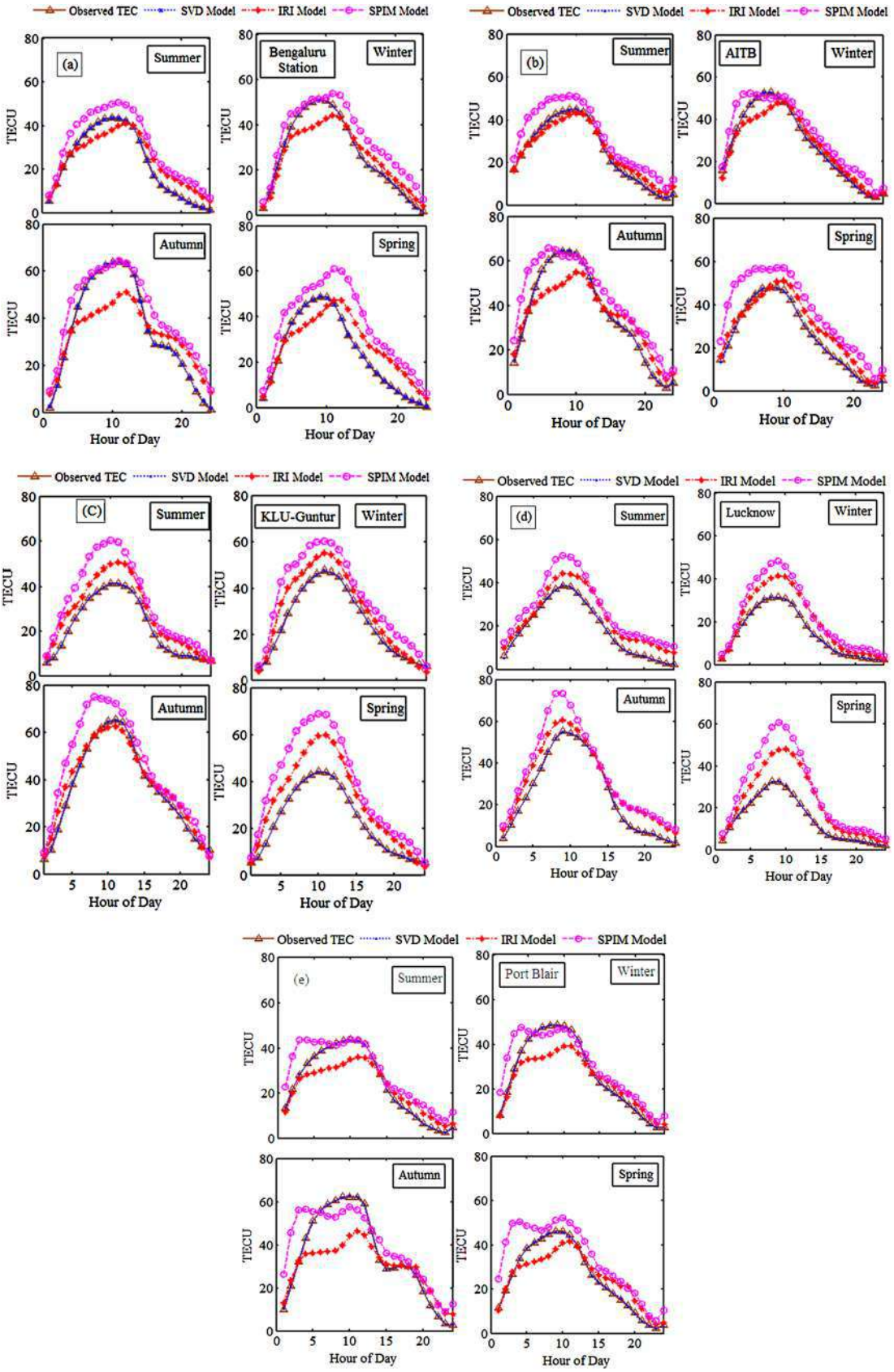


Fig. 7 Comparison of seasonal variations among the observed GPS–TEC with SVD Model, IRI2016 Model and SPIM Model during the year 2015 over (**a**) Bangalore station, (**b**) AITB Thailand station, (**c**) KLU-Guntur station, (**d**) Lucknow station, and (**e**) Port Blair station

Table 4 Seasonal Root Mean Square Deviation (RMSD) between observed (GPS–TEC) and model values

Station	Season	RMSD (TECU)			RMSD (%)		
		SVD model	IRI2016 model	SPIM model	SVD model	IRI2016 model	SPIM model
Bangalore	Autumn	0.41	10.12	8.76	1.01	24.66	21.36
	Spring	0.25	7.75	13.30	0.86	26.55	45.54
	Summer	0.18	5.14	7.98	0.65	18.64	28.94
	Winter	0.27	5.32	7.54	0.85	16.71	23.68
AITB	Autumn	0.60	7.67	8.91	1.45	18.38	21.35
	Spring	0.40	4.91	11.93	1.34	16.38	39.79
	Summer	0.27	2.61	7.38	0.95	9.00	25.96
	Winter	0.31	4.18	6.02	0.96	12.75	18.33
KLU-Guntur	Autumn	0.50	3.64	10.10	1.05	8.80	24.39
	Spring	0.30	9.45	16.77	0.90	35.20	62.46
	Summer	0.19	5.84	11.36	0.62	19.97	38.86
	Winter	0.30	6.24	12.26	0.70	24.29	47.57
Lucknow	Autumn	0.25	6.34	10.83	0.79	19.93	34.01
	Spring	0.45	10.32	16.0	2.55	57.37	60.88
	Summer	0.28	5.73	9.42	1.26	25.64	42.12
	Winter	0.38	6.25	9.62	2.11	34.18	50.62
Port Blair	Autumn	0.34	11.48	9.84	0.84	28.38	24.32
	Spring	0.44	5.38	10.15	1.50	18.34	34.54
	Summer	0.13	5.75	6.85	0.48	20.17	24.00
	Winter	0.18	6.53	6.50	0.60	20.88	20.78

The SVD model can reproduce the original data sets of SVD adequately by utilizing the first six singular values, which constitute about 98% of the total variance, which are linearly transformed from the observed TEC data sets. The singular values (U) represent the diurnal variation and small-scale irregularities, while the resultant singular values of (V) can show the long-term variations (solar cycle, annual, semi-annual, and seasonal).

The temporal GPS–TEC variations in the first three singular values of V have been modulated by the geomagnetic (Ap) and solar (F10.7) indices through the FSA. It reveals that singular values of V are associated with the geomagnetic and solar activity and the solar activity is the dominant factor to control the TEC variability.

The RMSD values increases from Bangalore towards Lucknow station, as the Lucknow station is on above the EIA crest, the TEC abnormal variation is noticeable and the day maximum also occurs slightly later than near equator stations, i.e., Bangalore. This is probably due to the time lag between EEJ strength and maximum intensity of fountain effect. The error deviation of SVD model is higher in winter season as compared to the summer season. The phenomenon could be attributed to the strong summer to winter neutral spread in solar maximum.

A comparative study has been conducted on the observed GPS–TEC data with SVD model and the standard global models (IRI2016 and SPIM). SVD model agrees quite well with the observational data and more accurately predicts the ionospheric irregularities than the IRI2016 and SPIM models. The performance of standard models is needed to be relatively improved.

Acknowledgements The present work has been carried out under the project titled 'Development of Single Frequency Ionospheric correction & plasma bubble detection algorithms using GPS Aided GEO Augmented Navigation (GAGAN) & Navigation Indian Constellation (NavIC) TEC observations' sponsored by NavIC–GAGAN Utilization Programme at Space Applications Centre, Ahmedabad, India, Project ID: NGP-10. The contribution is also supported by Department of Science and Technology (DST), New Delhi, India, SR/FST/ESI-130/2013(C) under DST–FIST Program. The authors thank the reviewers for their helpful comments.

References

A, E., Zhang, D.H., Xiao, Z., Hao, Y.Q., Ridley, A.J., Moldwin, M.: Modeling ionospheric for F2 by using empirical orthogonal function analysis. Ann. Geophys. **29**(8), 1501–1515 (2011)

A, E., Zhang, D., Ridley, A.J., Xiao, Z., Hao, Y.: A global model: empirical orthogonal function analysis of total electron content

1999–2009 data. J. Geophys. Res. Space Phys. **117**, A03328 (2012)
Abdu, M.: Equatorial ionosphere–thermosphere system: electrodynamics and irregularities. Adv. Space Res. **35**(5), 771–787 (2005)
Anderson, E., Bai, Z., Bischof, C., Blackford, S., Demmel, J., Dongarra, J., Croz, J.D., Greenbaum, A., Hammarling, S., McKenneyDS, A.: LAPACK Users Guide, vol.3. Society for Industrial and Applied Mathematics, Philadelphia (1999)
Ansari, K., Panda, S.K., Althuwaynee, O.F., Corumluoglu, O.: Ionospheric TEC from the Turkish Permanent GNSS Network (TPGN) and comparison with ARMA and IRI models. Astrophys. Space Sci. **362**(9), 178 (2017)
Ansari, K., Panda, S.K., &Corumluoglu, O.: Mathematical modelling of ionospheric TEC from Turkish permanent GNSS Network (TPGN) observables during 2009–2017 and predictability of NeQuick and Kriging models. Astrophys. Space Sci. **363**(3), 42 (2018)
Appleton, E.V.: Two anomalies in the ionosphere. Nature **157**(3995), 691 (1946)
Arunpold, S., Tripathi, N.K., Chowdhary, V.R., Raju, D.K.: Comparison of GPS–TEC measurements with IRI-2007 and IRI-2012 modeled TEC at an equatorial latitude station, Bangkok, Thailand. J. Atmos. Sol.-Terr. Phys. **117**, 88–94 (2014)
Bagiya, M.S., Joshi, H., Iyer, K., Aggarwal, M., Ravindran, S., Pathan, B.: TEC variations during low solar activity period (2005–2007) near the equatorial ionospheric anomaly crest region in India. Ann. Geophys. **27**(3), 1047–1057 (2009)
Bilitza, D., Reinisch, B., Lastovicka, J.: Progress in observation-based ionospheric modeling. Space Weather **6**, S02002 (2008)
Bouya, Z., Terkildsen, M., Neudegg, D.: Regional GPS-based ionospheric TEC model over Australia using spherical cap harmonic analysis. In: Proceedings of 38th COSPAR Scientific Assembly, Bremen, Germany, vol. 38, p. 4 (2010)
Chen, Z., Zhang, S.R., Coster, A.J., Fang, G.: EOF analysis and modeling of GPS TEC climatology over North America. J. Geophys. Res. Space Phys. **120**(4), 3118–3129 (2015)
Chowdhary, V.R., Tripathi, N.K., Arunpold, S., Raju, D.K.: Characterization of GPS–TEC in a low-latitude region over Thailand during 2010–2012. Ann. Geophys. **58**(5), 0553 (2015)
Dabas, R.S.: Ionosphere and its influence on radio communications. Resonance **5**(7), 28–43 (2000)
Dashora, N., Suresh, S.: Characteristics of low-latitude TEC during solar cycles 23 and 24 using global ionospheric maps (GIMs) over Indian sector. J. Geophys. Res. Space Phys. **120**(6), 5176–5193 (2015)
Fejer, B.G.: Low latitude ionospheric electrodynamics. Space Sci. Rev. **158**(1), 145–166 (2011)
Habarulema, J.B., McKinnell, L.A., Opperman, B.D.: Regional GPS TEC modeling; Attempted spatial and temporal extrapolation of TEC using neural networks. J. Geophys. Res., Space Phys. **116**(A4) (2011)
Jakobsen, J., Knudsen, P., Jensen, A.B.: Analysis of local ionospheric time varying characteristics with singular value decomposition. J. Geod. **84**(7), 449–456 (2010)
Jakowski, N., Wehrenpfennig, A., Heise, S., Schlueter, S., Noack, T.: Space weather effects in the ionosphere and their impact on positioning. In: ESA Workshop Paper, December (2001)
Jamjareegulgarn, P., Supnithi, P., Hozumi, K., Tsugawa, T.: Study of ionospheric topside variations based on NeQuick topside formulation and comparisons with the IRI-2012 model at equatorial latitude station, Chumphon, Thailand. Adv. Space Res. **60**(2), 206–221 (2017)
Kaplan, E., Hegarty, C.: Understanding GPS, Principles and Applications. 2. Artech House, INC, Norwood (2006)
Kenpankho, P., Watthanasangmechai, K., Supnithi, P., Tsugawa, T., Maruyama, T.: Comparison of GPS TEC measurements with IRI TEC prediction at the equatorial latitude station, Chumphon, Thailand. Earth Planets Space **63**(4), 365–370 (2011)
Kumar, S., Priyadarshi, S., Krishna, S.G., Singh, A.: GPS–TEC variations during low solar activity period (2007–2009) at Indian low latitude stations. Astrophys. Space Sci. **339**(1), 165–178 (2012)
Kumar, S., Tan, E.L., Murti, D.S.: Impacts of solar activity on performance of the IRI-2012 model predictions from low to mid latitudes. Earth Planets Space **67**(1), 42–58 (2015)
Mao, T., Wan, W., Yue, X., Sun, L., Zhao, B., Guo, J.: An empirical orthogonal function model of total electron content over China. Radio Sci. **43**(2) (2008)
Menke, W.: Geophysical Data Analysis: Discrete Inverse Theory. Academic, San Diego (1984)
Panda, S.K., Gedam, S.S., Rajaram, G.: Study of ionospheric TEC from GPS observations and comparisons with IRI and SPIM model predictions in the low latitude anomaly Indian subcontinental region. Adv. Space Res. **55**(8), 1948–1964 (2015)
Parker, R.L.: Geophysical Inverse Theory. Princeton University Press, Princeton (2004)
Preisendorfer, R., Mobley, C.D.: Principal Component Analysis in Meteorology and Oceanography. Elsevier, Amsterdam (1988)
Seemala, G.K., Valladares, C.E.: Statistics of total electron content depletions observed over the South American continent for the year 2008. Radio Sci. **46**, 1–14 (2011)
Sivavaraprasad, G., Ratnam, D.V.: Performance evaluation of ionospheric time delay forecasting models using GPS observations at a low-latitude station. Adv. Space Res. **60**(2), 475–490 (2017)
Skone, S.: Wide area ionosphere grid modelling in the auroral region. UCGE Reports, Number 20123, Department of Geomatics Engineering, The University of Calgary (1998)
Yu, Y., Wan, W., Xiong, B., Ren, Z., Zhao, B., Zhang, Y., Ning, B., Liu, L.: Modeling Chinese ionospheric layer parameters based on EOF analysis. Space Weather **13**(5), 339–355 (2015)

Journal of Engineering Technology
Volume 6, Issue 2, July. 2018, PP. 564-576
RESEARCH ARTICLE -ENGINEERNG TECHNOLOGY ASEE

Automatic raga identification in Indian classical music using the Convolutional Neural Network

Varsha N. Degaonkar[1], Anju V. Kulkarni [2]

[1]Research Scholar, Department of Electronics and Telecommunication, JSPM's Rajrshi Shahu College of Engineering, Pune, SPPU, Maharashtra, India

[2]Professor, Department of Electronics and Telecommunication, Padmashree Dr. D.Y. Patil Institute of Technology, Pune, SPPU, Maharashtra, India

Abstract: Automatic Raga Identification plays a vital role in Automatic retrieval of music file. Many researchers have used various combinations of multiple feature extraction methods and classifiers for identification of Raga till date. A number of problems associated with the above said methods are complexity of using many techniques in collaboration and processing time in general and a priori knowledge of the raga is a must for feature extraction in specific. In the proposed work, a different approach, namely Convolutional Neural Network (CNN) is used to extract high level features and also classify them without the necessity of a priori knowledge of the raga. The study shows that reduction in error rate is achieved by using CNN. To further improve the results, a novel technique is incorporated wherein the features obtained from the machine based and human based extractions are combined together in the CNN before further processing. This has resulted in another 5% reduction in the error rate. Local weight sharing characteristics of CNN appears to be of great advantage for raga identification and extraction since the features available at a particular part of classical music file may also be available in another part of the file and pooling avoids the need for decision making as regards to the overfitting parameter.

Keywords: Automatic Raga Identification, Convolutional Neural Network, Machine Based Feature, Human Based Features

1 Introduction

Automatic Retrieval of the Indian Classical music file plays a very important role in the development of automatic indexing and retrieval of the file from the huge database. Now a day, huge digital audio files are available, and automatic retrieval will save user's time from tedious and time consuming searches.

The tonic is the base of all the melodies present in Indian classical music. It depends upon the base pitch of the singer and it is always carefully chosen to decide the range of the pitch while singing. All the instruments (e.g. Tabla, Violin, Tanpura etc.) are tuned to the tonic of the lead singer of Raga. The sound of the Drone created by playing Tanpura is used to add a harmonic element to the performance of the Raga.

In Indian classical music, the Raga is the basic melodic framework upon which the music is built [1], [2], [3] and the Taal provides the rhythmic framework [4], [5]. Raga is made by the combination

of different Swaras or musical notes in a particular sequence. Taal is the repetitive form of the rhythmic and cyclic pattern by which creativeness is brought in.

In all the performances of Indian classical music, the base tonic is the "Sa" swara (Shadja) and the complete Raga is built on this Swara. All other swaras are derived in relation to this "Sa" Swara. In the classification of Indian Classical Music, Tonic identification plays an important role, but for many classical music types, Tonic identification is a complicated task. So there is a need for development of new algorithms or an approach for automatically identifying the Indian Classical Music [6].

A bandish in the Indian classical music is characterized by its mukhda, and it is mostly repeated at regular intervals. The automatic detection of the Bandish from the complete Classical music signal would contribute to important dataset. The mukhda can be detected by three ways, i.e. the lyrics, position in the cycle and its melodic shape. The main challenge in detection of Bandish is the nature of genre, as the grammar of the raga allows significant deviation in the shape of the melody of the phrase. [7]

In [8], the instrumental music analysis and classification is done using their spectral and temporal features. For extracting the features, a spectrum, chromagram, centroid, lower energy, roll off, and histogram are being used. Four ragas (i.e. Bhairav, Bhairavi, Todi and Yaman) have been classified using KNN and SVM classifier.

Chromagram patterns and Swara based features have been used for Scale-independent raga identification. GMM based Hidden Markov Models have been used for extracting the features consisting of chromagram patterns [9], Mel-cepstrum coefficients [10] and timbre features [11] on the specific dataset including 4 ragas- Sohini, Malhar, Khamaj and Darbari.

Raga (melody) and Tala (rhythm) are main foundational elements of Indian classical music. Both these are open frameworks for creativity and a very large number of possibilities are permissible.

The study shows that, all these systems need the primary knowledge base of classical music to start with selection of feature extraction algorithms and classifiers. Without the basic knowledge, one cannot select the correct algorithm.

CNN is more competent in finding the hidden features in input data as compared with the methods where the features are extracted manually by different methods or by combination of various methods. The main characteristics of the Convolutional Neural Network are local receptors with local connections, sharing of weights, and operation of pooling and dropout techniques

Local connections between the neurons of adjacent layers in CNN take advantage of spatially correlated data in classical music; as each neuron is only connected to a tiny section of the input data. As the features available at a particular part of classical music file, may be available in another part of the classical music file, weight sharing is very beneficial. The exact location of the feature is not of much importance as compared to other features. Reduction of spatial size is required in classical music data to avoid overfitting which is achieved by reducing the number of parameters and the amount of computation required in the network made possible by "Pooling" a major characteristic of CNN. Dropout reduces the overfitting of classical music data by avoiding the number of training nodes of training data. By doing so, it reduces the interaction among the nodes, so as to guide them to learn more robust features and generate a new set of data.

This research paper is arranged as follows: Section 2 describes the proposed methodologies used with details of implementation. Section 3 describes the various experiments done with the results and discussion. Finally the section 4 describes the conclusion of the research work.

2 Methodology

In this research work, the Convolutional Neural Network is used as a building block to extract the features and to classify the music files. The next section gives a brief introduction of the Convolutional Neural Network.

2.1 Convolutional Neural Network (CNN)

CNN is a multilayer feed-forward neural network, which is trained with the back-propagation algorithm. Nodes in CNN perform a scalar dot product (convolution) on the previous layer, but with only a small portion (receptive field) of the nodes in the previous layer.

For the Regular Neural Nets, the first layer consists of neurons and receives the input in vector format. And this is transformed through a series of hidden layers. A set of neurons is present in each hidden layer and each neuron in these hidden layer is fully connected to all the neurons in the previous layer, and the functioning of all these neurons in a single layer is completely independent and do not share any connections. The last layer (normally called Output layer) is fully-connected.

The Convolutional Neural Network is motivated by visual neuroscience. By applying a two dimensional input data, CNN can automatically find out the hidden features and creates a high level abstraction feature set, which is applied to either simpler classifier such as a Fully Connected Neural Network (NN) or a Support Vector Machine (SVM) for classification purpose. All the hidden patterns in the input signal, without human intervention, are learned from CNN and accumulated in the parameters of connections of the network, thus CNN needs very less labor-intensive identification of the parameters.

First, the Preprocessing is done on the input signal. This preprocessed data is applied to the Convolutional Neural Network. It has two layers. First is Convolutional layer and second is pooling layer. Multiple feature maps are used in each of the Convolutional layers so as to extract the higher level features from the previous layer. Each feature map is having multiple units, each of which is connected to receptive field in the previous layer.

In this research work CNN is used to automatically identify the Raga in Indian Classical Music file in three ways: In the first method, CNN is directly applied to the features of Indian Classical Music extracted by the Machine computation method. In the second method, CNN is applied to the features of Indian Classical Music extracted by the Human computation method. And in the third method, CNN is applied to the features of Indian Classical Music extracted from the hybrid combination of Machine computation and Human computation method. The following section describes the detailed methodology with steps.

2.2 Automatic Raga Identification in Indian Classical Music using Machine computation and Convolutional Neural Network (CNN)

Following steps are used for Automatic Raga identification in Indian Classical Music using Machine computation and Convolutional Neural Network:

- The input Indian Classical Music signal passed through a 25ms Hamming window with a fixed frame rate of 10ms.
- Fourier transform based filter bank analysis is used to generate the feature vectors, which includes forty log energy coefficients distributed on a mel scale.
- The log-energy is calculated directly from the mel-frequency spectral coefficients (i.e., without calculating DCT of the signal), which are denoted as Mel Frequency Spectral Coefficients (MFSC).
- These MFSC features are used to characterize each audio frame, along with their first and second derivatives. This portrays the acoustic energy distribution in numerous different frequency bands.
- Input music signal is divided into total 15 frames and for each frame 40 MFSC features along with their first and second derivatives are calculated, i.e. total 45 feature maps with 40 frequency bands are calculated.
- This is directly applied to the first Convolutional layer, where six feature maps are used and this is followed by the pooling layer.
- In second Convolutional layer, twelve feature maps are used followed by pooling layer.
- This output is then applied to the fully connected layer which consists of the output layer as classifier which gives direct Raga identification.

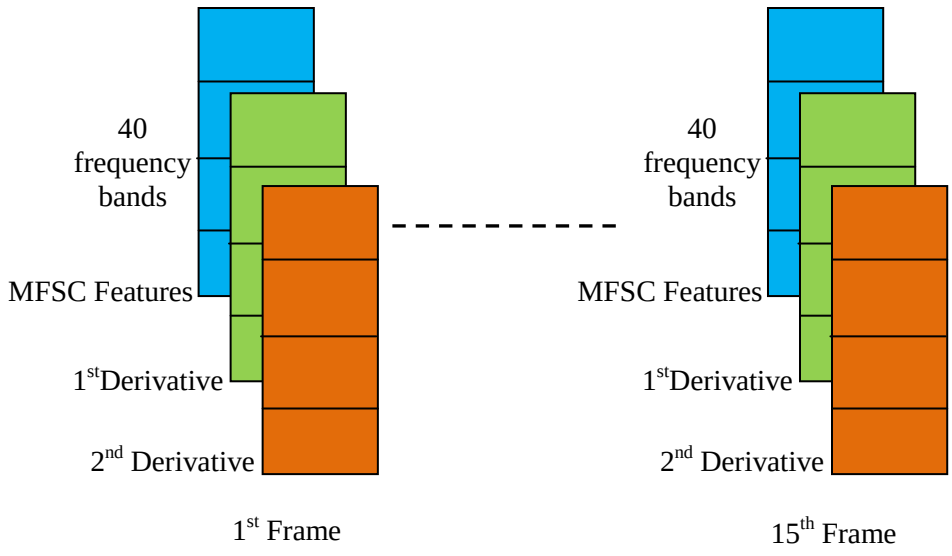


Figure 1. Organization of music input features

2.3 Automatic Raga Identification in Indian Classical Music using Human computation and Convolutional Neural Network (CNN)

By using a human computation and involving human in an activity, attributes are collected directly to the different music input [14]. In [14], Assisted and Unassisted activities are developed to collect different attributes from the players. In assisted activity, players have selected the correct option given for the particular music input. Here the players are assisted in the selection process. In unassisted activity, players have written the relevant attributes in the text boxes provided for the particular music

input. All these attributes, which are in the form of sentences/words, are directly applied to the Convolutional Neural Network.

One dimensional Convolutional Neural Network is used and the filter map is slide in only one dimension, as shown in the figure 2.

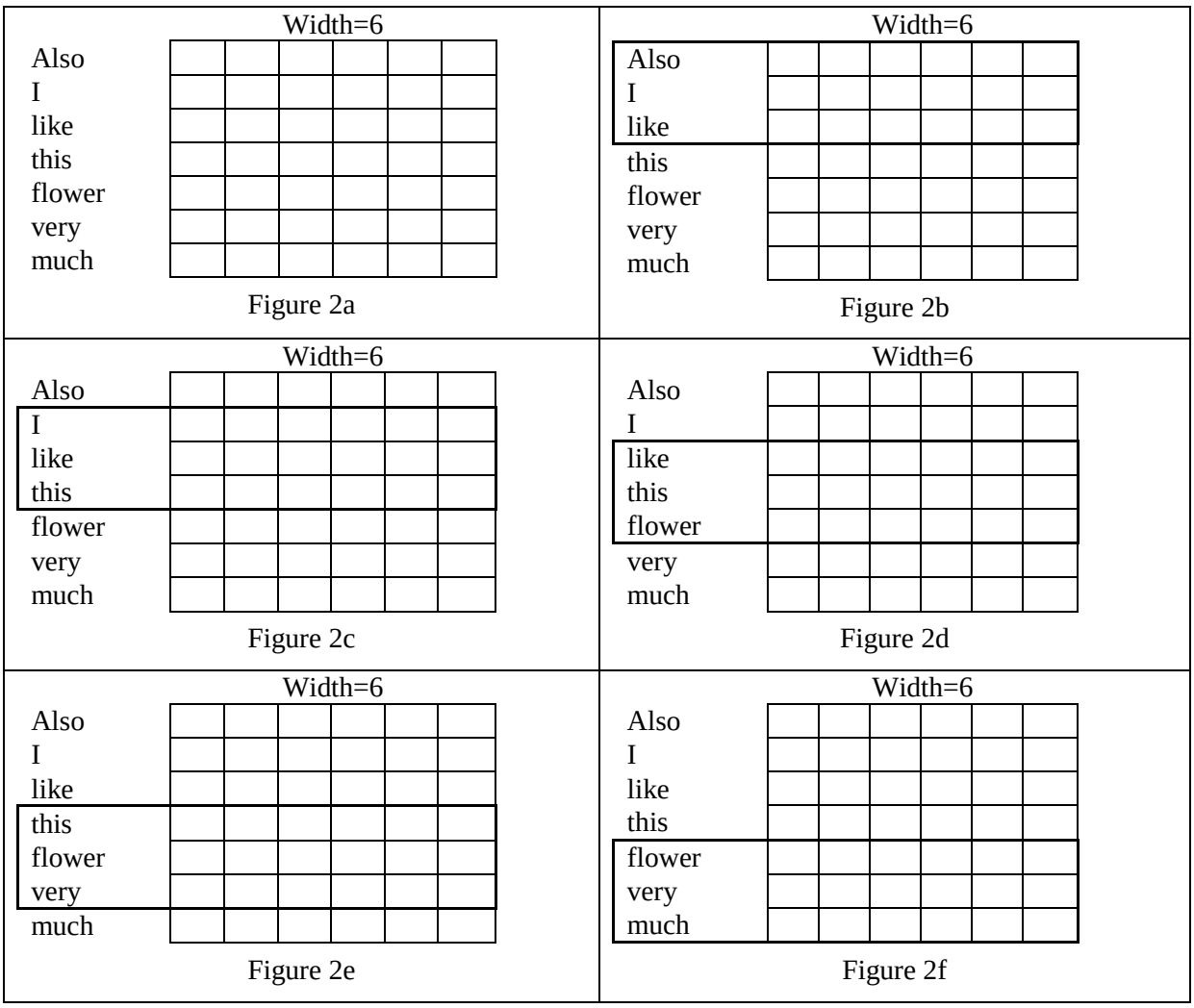


Figure 2a- 2f. Representation of sentence / words in a matrix and shifting the window

As shown in figure 2a-2f, after convolving one filter with the input music signal, one feature vector is generated. For 6 such feature maps, six different filters are convolved with input in the first convolution layer.

2.4 Automatic Raga Identification in Indian Classical Music using collaboration of Human computation, Machine computation and Convolutional Neural Network (CNN)

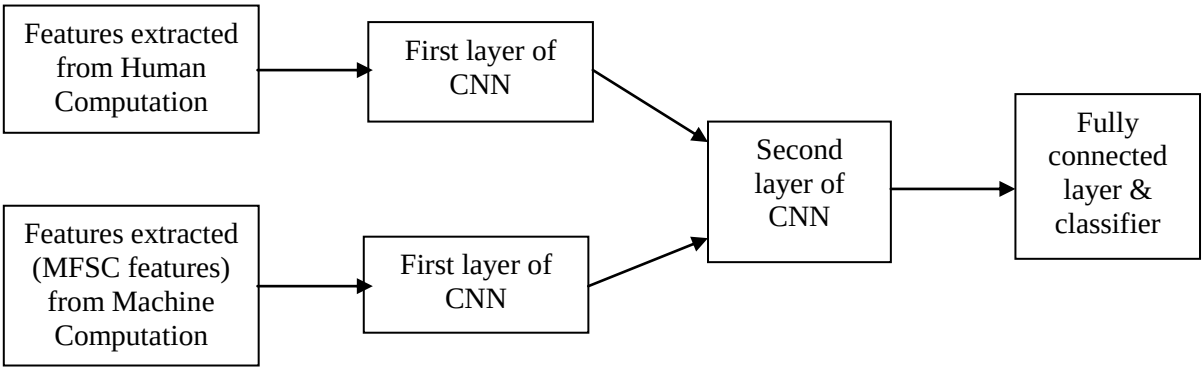


Figure 3. Structure of CNN for collaborative Human Based and Machine Based Techniques

As shown in figure 3, first layer of CNN is different for features extracted from Human Computation and features extracted from Machine Computation. With second layer, all the features are combined and given to classifier through the fully connected layer.

3. Results

Effect of variation in CNN parameters like Sub-sampling factor (Shift size), pooling size, the size of the filter, and a number of feature maps is checked for all the proposed methods.

3.1 For Automatic Raga Identification in Indian Classical Music using Machine computation and Convolutional Neural Network (CNN):

The Effect of varying Sub-sampling factor (CNN Shift sizes)

As per the figure 4, when the shift size is smaller, better results are achieved. This is achieved because, with the smaller shift sizes locality of the data is maintained.

Journal of Engineering Technology
Volume 6, Issue 2, July. 2018, PP. 564-576
RESEARCH ARTICLE -ENGINEERNG TECHNOLOGY

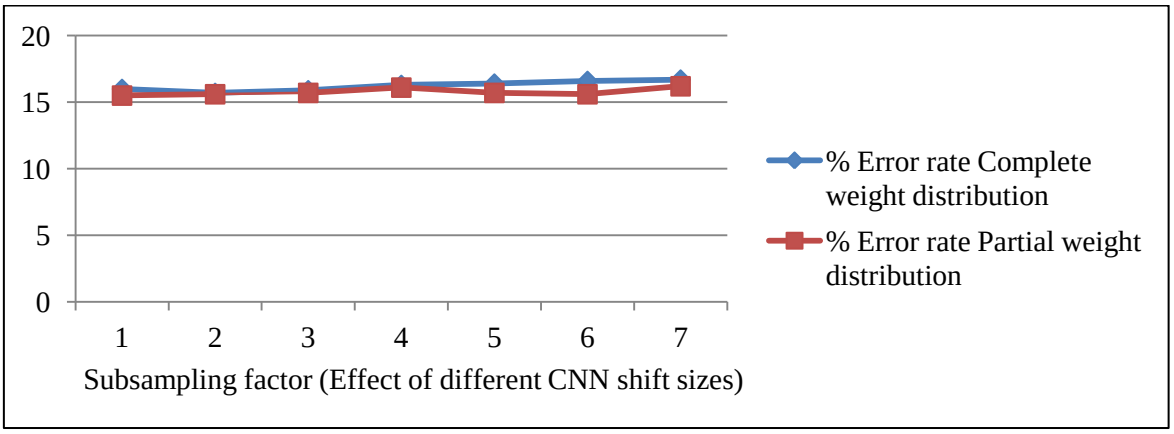


Figure 4. Effects of CNN shift size variation for Music input on % Error Rate for Complete Weight Distribution and Partial Weight Distribution

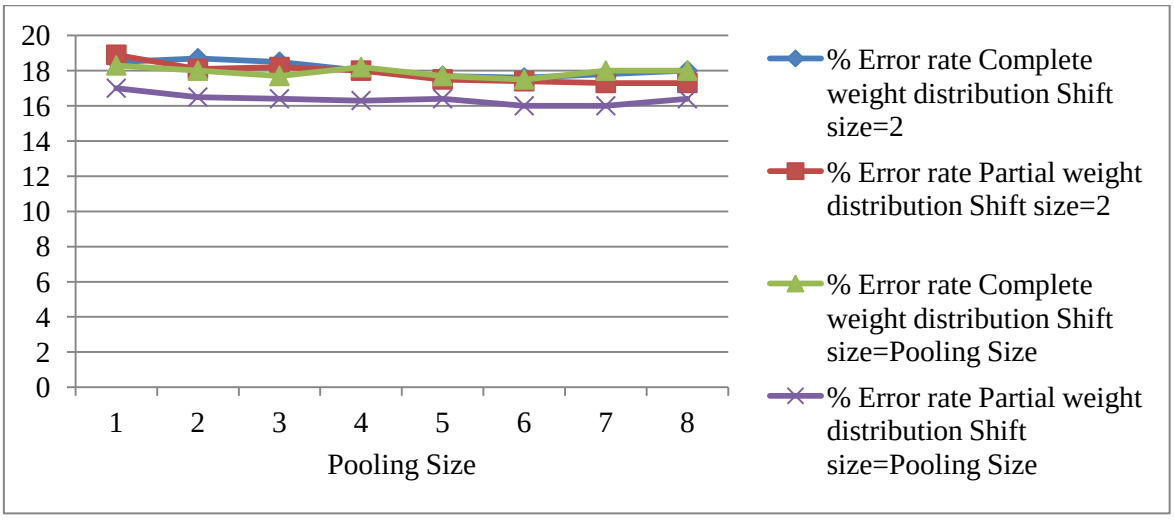


Figure 5: Effect of pooling size variation on % Error Rate for Music input for both Partial Weight Distribution and Complete Weight Distribution.

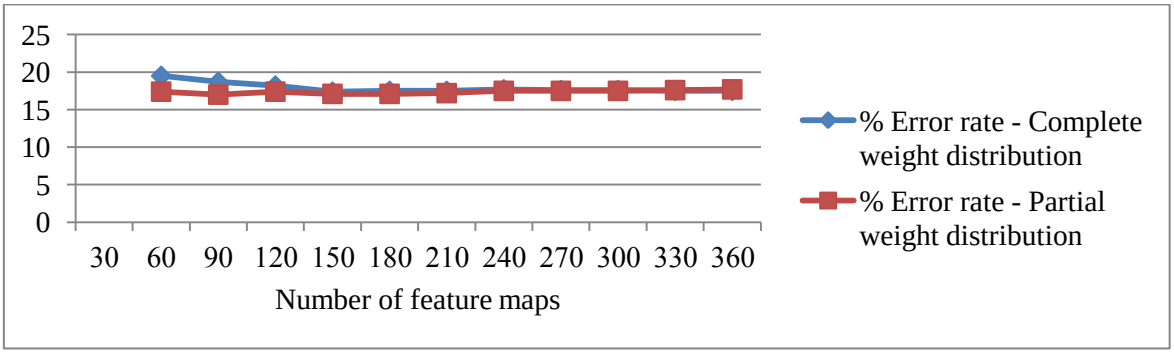


Figure 6. Effect of variation in number of feature maps for Music input on % Error rate for both Partial Weight Distribution and Complete Weight Distribution.

Effect of varying Pooling sizes

The figure 5 shows that there is no clear performance gain when the overlapping pooling window is used. But when both the pooling size and the shift size have the same value, reduction in the percentage error rate is found and decreases the complexity of the model.

Effect **of varying number of feature maps**

From figure 6, it is observed that with very less and very high number of feature maps, it does not produce any clear performance gain. For Partial Weight Distribution, 80 feature maps and for Complete Weight Distribution, 150 feature maps are giving best results and good retrieval efficiency.

Effect of varying size of the filter

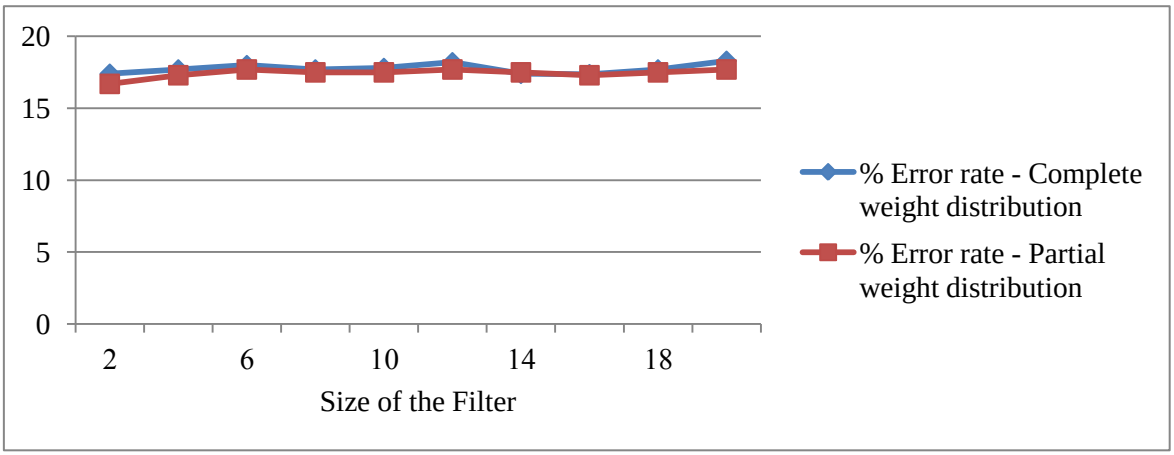


Figure 7. Effects of variation in filter size for Music input on % Error Rate for both Partial Weight Distribution and Complete Weight Distribution.

As per figure 7, when the size of the filter is smaller, better results are achieved, as with smaller shift sizes locality of the data is maintained. In the convolution layer and pooling layer pooling size equal to 4, shift size equal to 2, 150 feature maps for Complete Weight Distribution, and 80 feature maps per frequency band for Partial Weight Distribution is used.

Table 1. Effect of different CNN parameters on the average percentage error rate for Automatic Raga Identification using Machine computation and Convolutional Neural Network (CNN)

The Effect of	Network Structure	Average % Error rate
Pooling size	Complete Weight Distribution Shift size=2	18.1
	Partial Weight Distribution Shift size=2	**17.8**
	Complete Weight Distribution Shift size=Pooling Size	17.9
	Partial Weight Distribution Shift size=Pooling Size	**16.4**
Subsampling factor	Complete Weight Distribution	16.2
	Partial Weight Distribution	**15.8**
A Number of feature maps	Complete Weight Distribution	17.9
	Partial Weight Distribution	**17.4**
Size of the Filter	Complete Weight Distribution	17.8
	Partial Weight Distribution	**17.4**

Partial Weight Distribution: As the properties of the input music signal varies over diverse frequency bands. By using a different set of weights for different frequency bands are more suitable. As by doing so it gives the flexibility for detection of distinctive feature patterns in different filter bands along the frequency axis.
Complete Weight Distribution: The same type of patterns may present in an image at different location, thus the distribution of complete weight may be good for image input.

3.2 For Automatic Raga Identification in Indian Classical Music using Human computation and Convolutional Neural Network (CNN)

The effect of varying Sub-sampling factor (CNN Shift sizes)
As per the figure 8, when the shift size is smaller, better results are achieved, as with smaller shift sizes locality of the data is maintained.

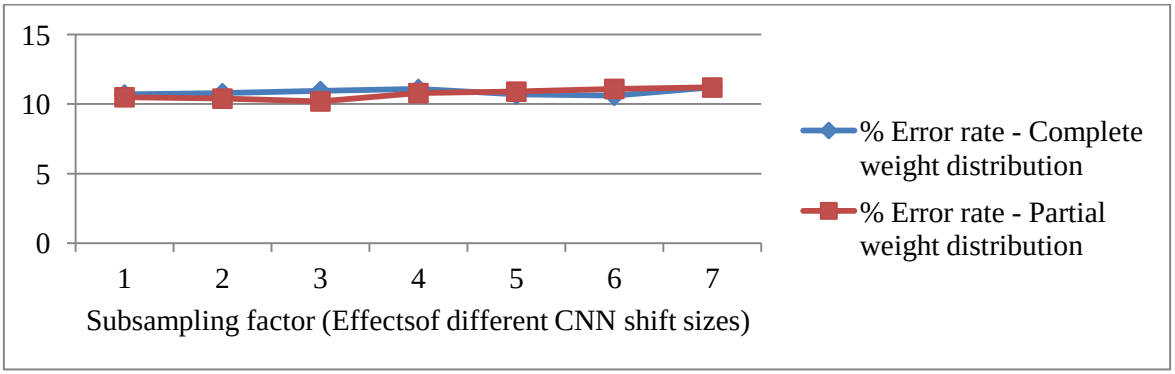


Figure 8. Effects of variation in shift size on for hybrid Music input % Error Rate for both Partial Weight Distribution and Complete Weight Distribution

Effect of varying Pooling sizes

Figure 9 shows, that there is no clear performance gain when the overlapping pooling window is used. But when the same value for both the pooling size and the shift size is used, it reduces the percentage error rate and decreases the complexity of the model. A shift size equal to 2 and a pooling size equal to shift size is used to check the effect.

Effect of varying number of feature maps

Figure 10 shows, with very less and very high number of feature maps, network does not create a clear gain in the performance. For Limited Weight Sharing 80 feature maps and for Full Weight Sharing 150 feature maps are giving best results and good retrieval efficiency.

Effect of varying size of the filter

As per the figure11, when the size of the filter is smaller, better results are achieved, as with smaller shift sizes locality of the data is maintained. In the convolution layer and pooling layer pooling size equal to 4, shift size equal to 2, 150 feature maps for Complete Weight Distribution, and 80 feature maps per frequency band for Partial Weight Distribution is used.

Journal of Engineering Technology
Volume 6, Issue 2, July. 2018, PP. 564-576
RESEARCH ARTICLE -ENGINEERNG TECHNOLOGY

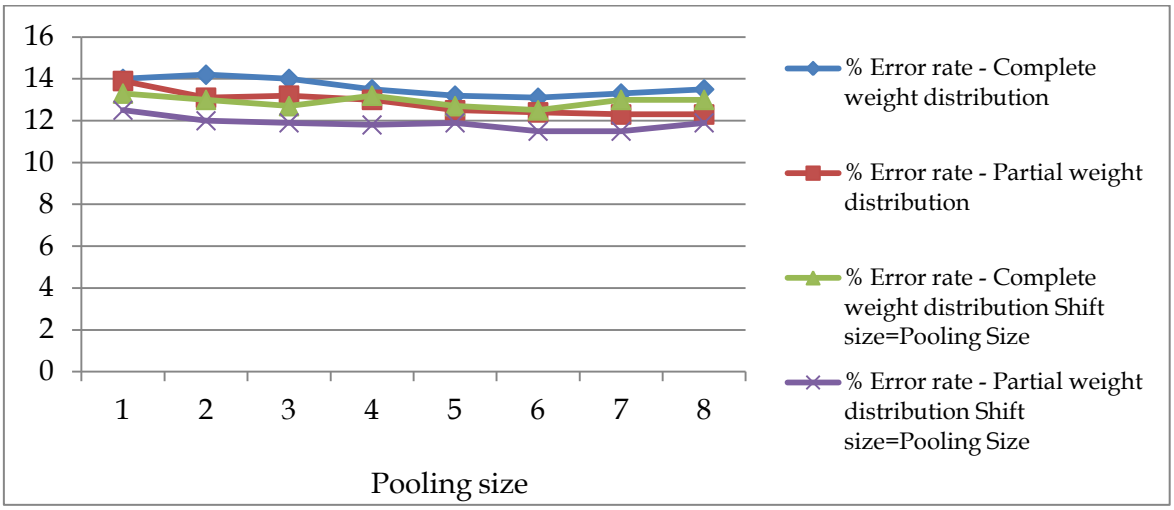


Figure 9. Effect of variation in pooling size on % Error Rate for hybrid music input for both Partial Weight Distribution and Complete Weight Distribution.

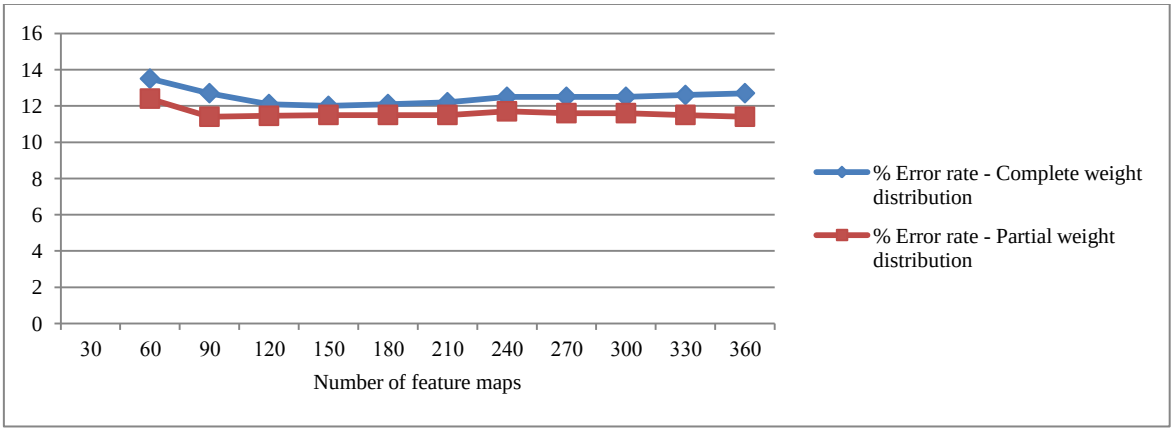


Figure 10. Effects of variation in numbers of feature maps for hybrid Music input on % Error for both Partial Weight Distribution and Complete Weight Distribution.

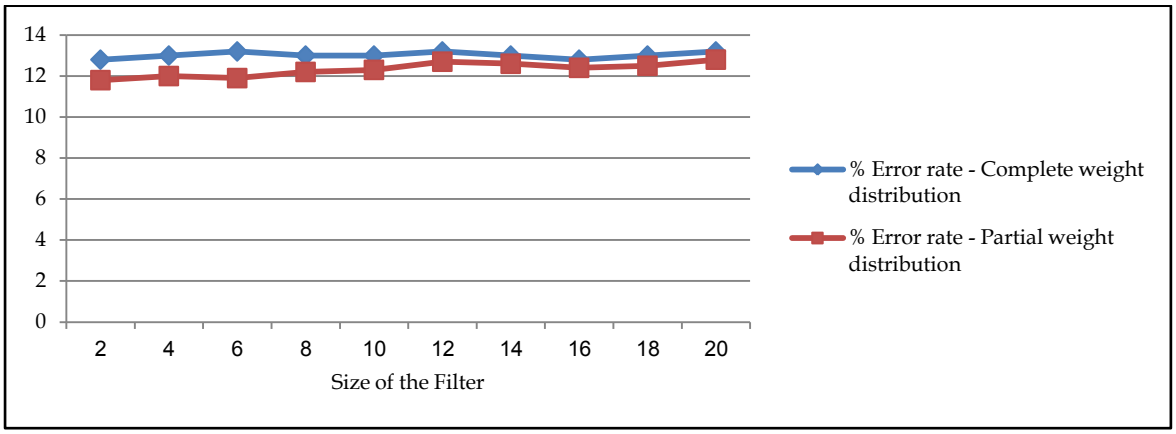


Figure 11. Effect of variation in filter size for hybrid Music input on % Error Rate for both Partial Weight Distribution and Complete Weight Distribution.

Table 2. Effect of different CNN parameters on the average percentage error rate for music & word / sentence dataset

The Effect of	Network Structure	Average % Error rate
Pooling size	Complete Weight Distribution Shift size=2	13.6
	Partial Weight Distribution Shift size=2	**12.8**
	Complete Weight Distribution Shift size=Pooling Size	12.9
	Partial Weight Distribution Shift size=Pooling Size	**11.9**
Subsampling factor	Complete Weight Distribution	10.9
	Partial Weight Distribution	**10.7**
A Number of feature maps	Complete Weight Distribution	12.5
	Partial Weight Distribution	**11.6**
Size of the Filter	Complete Weight Distribution	13.0
	Partial Weight Distribution	**12.3**

For the music dataset, Partial Weight Distribution gives 3% reduction of % error rate as compared to other methods. The properties of the music signal vary over diverse frequency bands. Using different sets of weights for separate frequency bands is more appropriate since it permits for the detection of distinctive feature patterns in separate filter bands along the frequency axis. For music dataset, using a collaborative approach, the percentage error rate is further reduced by an average 5%. For the music dataset, Partial Weight Distribution gives 3% reduction of % error rate as compared to other methods.

Table 3. Effect of different CNN parameters on the average percentage error rate for Automatic Raga Identification using a collaboration of Machine computation, Human Computation and Convolutional Neural Network (CNN)

The Effect of	Network Structure	Avg % Error Rate	
		Music Input	Hybrid Music Input
Pooling size	Complete Weight Distribution Shift size=2	18.1	13.6
	Partial Weight Distribution Shift size=2	**17.8**	**12.8**
	Complete Weight Distribution Shift size=Pooling Size	17.9	12.9
	Partial Weight Distribution Shift size=Pooling Size	**16.4**	**11.9**
Subsampling factor	Complete Weight Distribution	16.2	10.9
	Partial Weight Distribution	15.8	10.7
A Number of feature maps	Complete Weight Distribution	17.9	12.5
	Partial Weight Distribution	**17.4**	**11.6**
Size of the Filter	Complete Weight Distribution	17.7	13.0
	Partial Weight Distribution	**17.4**	**12.3**

The properties of the music signal vary over different frequency bands. Using different sets of weights for separate frequency bands is more appropriate since it permits for the detection of distinctive feature patterns in different filter bands along the frequency axis.

Result comparison for Classical Music input

Table 4. Result comparison of different techniques for music dataset

Techniques Used		% Error Rate
MFCC + FFNN [12]		28.4
MFCC + KNN [12]		22.4
MFCC + SVM [13]		22.3
MFCC +PCA + FFNN [13]		16.8
MFCC + PCA +KNN [12]		18.8
MFCC +PCA +SVM [13]		16.8
MFCC +ICA + FFNN [12]		28.3
MFCC + ICA +KNN [14]		20.7
MFCC +ICA +SVM [14]		20.3
MFCC Combined Features + SVM [13]		16.4
MFCC Combined Features+ PCA+ SVM [13]		18.3
Human Computation [14]		16.0
Machine computation & CNN	Pooling size variation	16.4
	Shift size variation	15.8
	Variation in number Feature maps	17.4
	Variation in Size of the Filter	17.4
Collaboration of Machine computation, Human computation & CNN	Pooling size variation	**11.9**
	Shift size variation	**10.7**
	Variation in number Feature maps	**11.6**
	Variation in Size of the Filter	**12.3**

As shown in table 4, as compared to other techniques, CNN gives performance improvement by reducing the % Error rate by 4%. And further performance improvement is achieved by CNN using the collaborative approach by further reducing the % Error rate by 5%.

Conclusion

In the proposed work, a different approach, namely Convolutional Neural Network (CNN) is used to extract high level features and also to classify them without the necessity of a priori knowledge of the raga. To further improve the results, a novel technique is incorporated wherein the features obtained from the machine based and human based extraction are combined together (hybrid combination) in the CNN before further processing. As the properties of input music signal vary over diverse frequency bands. By using different sets of weights in CNN for different frequency bands are more suitable since it gives the flexibility for detection of distinctive feature patterns in different filter bands along the frequency axis.

For music dataset, CNN gives performance improvement by reducing the % Error rate by 4% using machine computation. Further performance improvement is achieved by CNN using the collaborative approach by further reducing the % Error rate by 5%.

References

[1] Bagchee, S., "Nad: understanding raga music", Business Publications Inc, 1998.

[2] Danielou, A., "The ragas of Northern Indian music", New Delhi: Munshiram Manoharlal Publishers, 2010.

[3] Viswanathan, T., & Allen, M. H., "Music in South India",Oxford University Press, 2004.

[4] Clayton, Clayton, M. R.L., "Time in Indian music: rhythm, metre, and form in North Indian rag performance", Oxford University Press, 2000.

[5] Sen, Sen A. K., "Indian concept of rhythm (Second ed.)", New Delhi: Kanishka Publishers, Distributors, 2008.

[6] Sankalp Gulati, Ashwin Bellur, Justin Salamon, Ranjani H.G, Vignesh Ishwar, Hema A Murthy and Xavier Serra, "Automatic Tonic Identification in Indian Art Music: Approaches and Evaluation", Journal of New Music Research, Volume 43, Issue 1, 31 Mar 2014

[7] Kaustuv Kanti Ganguli, Abhinav Rastogi, Vedhas Pandit, Prithvi Kantan, Preeti Rao, "Efficient Melodic Query Based Audio Search For Hindustani Vocal Compositions", Proc. of Int. Soc. for Music Information Retrieval (ISMIR), Malaga (Spain), October 26-30, 2015

[8] A. C. Bickerstaffe and E. Makalic, "MML classification of music genres," in AI 2003: Advances in Artificial Intelligence, Springer Berlin Heidelberg, 2003.

[9] Chordia, P., Senturk, S., "Joint recognition of raag and tonic in north indian music", Computer Music Journal, 37(3), 82-98, 2013.

[10] Kartik Mahto, Abhilash Hotta, Sandeep Singh Solanki, Soubhik Chakraborty, "A Study on Artist Similarity Using Projection Pursuit and MFCC: Identification of Gharana from Raga Performance", International Conference on Computing for Sustainable Global Development (INDIACom), pp: 647:653, 2014

[11] Längkvist, M., Karlsson, L., Loutfi, A., "A review of unsupervised feature learning and deep learning for time-series modeling", Pattern Recognition Letters, 42(1): pp: 11-24, 2014

[12] Varsha N. Degaonkar, Anju Kulkarni, A Novel Hybrid Approach for Retrieval of the Music Information, International Journal of Applied Engineering Research, Vol. 12, No. 24, (2017), 15011-15017

[13] Varsha N. Degaonkar, Dr. Anju V. Kulkarni, "Classical Music Information Retrieval", International Journal of Pure and Applied Mathematics, Volume 118 No. 20 2018, 883-890

[14] Varsha N. Degaonkar, Dr. Anju V. Kulkarni, "Unassisted Crowd Sourcing Technique for Knowledge Generation", International Conference on Recent Trends in Engineering and Material Sciences (ICEMS-2016)

Developing Mathematical Models and Methodologies to Overcome Lean Strategy Selection and Leanness Assessment Problems

Rakhi Ashok Deshpande, P. Venkata Chalapathi

Abstract: This investigation inspects the impact of lean strategies among chose manufacturing organizations. In orderto remain focused; creating cheaper products at a quicker rate Lean Manufacturing would encourage the business. In this paper we discuss about adopting lean strategies by manufacturing industries with the intention of eliminating wastes. Here we create mathematical models and a systematic methodology for choosing fitting lean strategies. The main objectives of this paper are to assess the leanness levels in manufacturing organizations. In the work we will evaluate both quantitative and subjective info factors. Different tools like MATLAB, Microsoft Access and Excel, Visual Basic, DST tool will be used for developing mathematical models and methodologies to overcome lean strategy selection and leanness assessment problems. As a result we try to limit the conceivable disarray of executing different lean strategies which can be utilized by any manufacturer for surveying leanness and distinguishing their leanness benchmark Accordingly, we suggest that manufacturing firms nearby different types of organizations (benefit firms comprehensive) ought to receive lean reasoning, introduction or culture to dispense with wasteful practices in this way clearing a path for best practices that can attract main concern for the organizations.

Index Terms: Lean, strategies, manufacturing, organization, **industries, models, etc.**

I. INTRODUCTION

Manufacturing tasks are ceaselessly endeavoring to expand profitability and yield of their activities. They will probably fulfill the client with the correct item, quality, amount, and cost in the briefest measure of time. Lean manufacturing is in excess of a cost decrease program or a critical thinking approach. The principle thought is that a proficient creation can be accomplished by a far reaching way to deal with limit wastes. This implies taking out overabundance creation and stock, repetitive development of material, pausing and delays, over handling, abundance laborer movement, and the requirement for modify and remedies. Some portion of lean manufacturing is checking on activities for those parts, procedures or items that include cost as opposed to esteem. Each progression of the manufacturing procedure is observed to decide whether it increases the value ofthe item. In the event that it doesn't include Value, the procedure could be appointed to a subcontractor or redistributing organization with the end goal to concentrate the staff on esteem included tasks of its center business. Endless number of manufacturing organizations has made upgrades since the presentation of Lean management strategies and apparatuses. In any case, the advantages have not been as so great for service industries applying Lean management standards.

A. Concept of lean manufacturing and lean manufacturing strategies

Lean Manufacturing can be characterized as: Lean manufacturing or lean generation, which is regularly known just as" Lean", is the ideal method for producing goods through the evacuation of waste or we can say "Lean manufacturing is the system which points in end of the loss from the system with a systematic and constant methodology" OR Lean Manufacturing is an operational system arranged toward accomplishing the briefest conceivable process duration by disposing of waste. Lean manufacturing techniques depend on the utilization of five principles to manage management's activity toward progress.

A lean organization fathoms customer esteem and centers its key processes to constantly expand it. This strategy thinks about the regard for the laborers, the nature of the products and the steadiness of the process. This production method changes the focal point of management and guides it to the disposal of waste along whole value streams, rather than at individual focuses. To achieve this makes processes that need less human exertion, less space, less capital, and less time to make products and services at far less costs and with many less imperfections, contrasted and customary production systems. There are five fundamental ideas that characterize lean reasoning and empower lean generation: determine esteem; distinguish the esteem stream, stream, force, and error free. Clearly, lean reasoning underlines quality and incentive for every item from the point of view of the end client. Lean manufacturing, spearheaded by Toyota, includes stock and quality control, modern relations, work administration, and provider producer rehearses that vary on a very basic level from customary business rehearses. This attestation takes us to the five principle of lean thinking: Value, Value Stream, Flow, Pull and perfection.

Include Value: Value is characterized as a "capacity given to customer at the correct time and at a correct cost, as built up for each situation by the customer".

Revised Manuscript Received on August 05, 2019.
RAKHI ASHOK DESHPANDE , Research Scholar, Department of Mechanical Engineering, KoneruLakshmaiah Education Foundation, Vaddeswaram, Andhra Pradesh, 522502, India. Aadhaar Number - 5757 5645 9875.
DR. P. VENKATA CHALAPATHI , Professor, Department of Mechanical Engineering, KoneruLakshmaiah Education Foundation, Vaddeswaram, Andhra Pradesh, 522502, India.

RESEARCHERID
THOMSON REUTERS
[Varpe, 5(10): October 2018]
DOI- 10.5281/zenodo.1472142
ISSN 2348 – 8034
Impact Factor- 5.070

GLOBAL JOURNAL OF ENGINEERING SCIENCE AND RESEARCHES

TEMPORAL VARIABILITY OF AEROSOLS PARAMETERS OVER A RURAL AERONET OBSERVING SITE AT GANDHI COLLEGE

Sandeep Varpe[1]*, Mahesh Waghmare[1], Swati Kolet[1], Rupali Varpe[2], Mandar Datar[1] & Sachin Alhat[3]

[1]International Institute of Information Technology P-14, Rajiv Gandhi Infotech Park, Hinjawadi, Pune 411057, India
[2]Navmaharashtra Jr. College, Pimpri Waghere, Pune 411 014, India
[3]DG Walse Patil College, Paragaon, Pune 412 406, India

Abstract

A compressive measurement of aerosol optical depth (AOD), a particulate matter (PM) and black carbon mass concentration have been carried out over Gandhi College. The measured aerosol data was incorporated in an aerosol optical model to estimate various aerosol optical parameters which were subsequently used for radiative forcing estimation. The impact of aerosol on the global climate system is considered as a medium to low. The effect of atmospheric aerosol particle depends on the spatial and vertical distribution of aerosol. This paper aims on a metropolitan AERONET observing site at Gandhi College in India. The data analysis gives output the existence of large seasonal heterogeneity in the frequency distribution of aod_{500} mm, aod_{1020} mm and corresponding retrieved $ae_{440\text{-}870}$ mm during different season at Gandhi College.

Keyword: *AEROSOL OPTICAL DEPTH (AOD), ANGSTROM EXPONENT (AE), PRECIPITABLE WATER VAPOUR (PWV).*

I. INTRODUCTION

The atmosphere consists of tiny colloidal particles of different chemical composition, dispersed/embedded in a gas, smoke or fog, termed as aerosols [1]. These particles linger in the atmosphere for a long period of time and their lifetime is found to vary considerably depending on their size and composition. The coarse-mode volcanic dust particles injected into the atmosphere during large volcanic eruptions have a short lifetime of about 1–2 months due to their fast removal by gravitational settling [2]. On the other hand, tropospheric sulfate aerosols, mostly in the accumulation-mode size regime have a lifetime of 5–7 days [3]. The worldwide investigations of aerosol optical, microphysical, radiative and compositional properties have established that the aerosols disperse widely both on horizontal and vertical scales by means of circulation patterns in the atmosphere [4]. Aerosols perturb the Earth-atmosphere climate system at a local, regional and global scale through direct and indirect processes. The characterization of aerosols is further complex due to their large spatio-temporal uncertainty, heterogeneity in the nature of particles suspended in the atmosphere and the associated lifetime [5,6,7]. Aerosol mixtures composed of dust, sulfate, black carbon (BC), sea-salt particulates produce a potential challenge to satellite and sub-orbital remote sensing techniques. Hence, aerosols are at the focus of growing interest due to their impact on the air quality, human health and Earth-atmosphere climate system [8].

Investigations reveal that the mixtures of combustion aerosols and desert dust comprise of BC and iron oxides as the most important absorbing particulate species in fine- and coarse-mode dust particles respectively [9]. The iron oxides present in desert dust produce the significantly strong absorption in the ultraviolet and visible spectral regions [10]. The long-term AERosolROboticNETwork (AERONET) data sets have been used in various global studies to delineate spatio-temporal aerosol characteristics, their type discrimination, etc [11]. In the present study, the multiyear AERONET data on the direct Sun measurement and almucantar scan retrievals at Gandhi College have been analysed to investigate heterogeneity in aerosol optical and microphysical properties. For this, monthly/annual

Iran J Sci Technol Trans Sci
https://doi.org/10.1007/s40995-018-0619-4

RESEARCH PAPER

CrossMark

An Epidemiological Correlation of Oral Candidiasis Mice Model Study of an Isolate from *Mollugo pentaphylla* Linn. and In silico Docking Approach

Partha Niyogi[1] · Snigdha Pattnaik[1] · Laxmidhar Maharana[1] · Rajaram Mohapatra[1] · Swati P. Kolet[2]

Received: 24 October 2017 / Accepted: 20 July 2018

Abstract

The aim of the present study was to isolate the active compound against *Candida albicans* from the methanolic extract of aerial part of *Mollugo pentaphylla* Linn. and to peruse in vivo animal model of oral candidiasis. The study was designed to quantify the oral *candida* infection with different statistical approach and histopathological observation of infected tongue on treatment with 100 µg/mL of the isolated compound. The significance of age, body weight and tongue weight was also correlated with disease treatment on either sex of animal. The isolated active compound was characterized as stigmasterol (EX-1) based on FTIR, DSC, ^{1}H NMR, ^{13}C NMR and mass spectral analysis. An in silico docking approach predicted strong negative binding energy efficacy of stigmasterol with *C. albicans* PKh Kinase protein unveiled an in-visual perception for anticandidal efficacy.

Keywords Animal model · Statistical · Histopathological · Stigmasterol · DSC · NMR · In-visual

1 Introduction

Oral candidiasis, a superficial or deep infection, is caused by the colonization and proliferation of opportunistic *Candida albicans* in the oral mucosal layer and periodontal pocket resulting in mucocutaneous mycosis of the oral cavity (Haynes 2001; Järvensivu et al. 2004; Klis et al. 2009). Co-aggregation of fungi with the dental biofilm bacteria enhances the adherence of the *Candida* species in the oral epithelial layer and supports the progression of the infection along with the production of pathogenic collagens, phospholipase-C and protease enzyme, which may lead to degradation of extracellular matrix protein and immunoglobulin (Haynes 2001; Järvensivu et al. 2004; Bensadoun et al. 2011). AIDS (Farah et al. 2010; Margaix-Muñoz et al. 2009) and diabetes mellitus are found to be the prevalent predisposing factors for the oral candidiasis (Aguiree-Urízar 2002). It was observed that around 75% among the HIV patients, 50% among diabetics are associated with candidiasis (Wyk and Steenkamp 2011). Moreover, prolong administration of antibiotic and adrenal cortex steroid to treat candidiasis causes suppression of host defense system (Villar and Dongari-Bagtzoglou 2008); long use of polyene group or the azole and its derivatives to counter oral candidiasis (Hata et al. 1996) leads to hepatotoxicity and nephrotoxicity (Kauffman and Carver 1997) and also leads to improved resistance of *Candida* species to azole and its derivatives (Dupont et al. 1996).

Presently, the modern world has started utilizing the diversified medicinal areas to reduce side effect in acute or chronic clinical use. In this context, the plant *Mollugo pentaphylla* Linn., is well known to possess potent antioxidant and anti-diabetic activity (Valarmathi et al. 2013). However, till date the concrete component responsible solely or in amalgamation for anticandidal activity was not been identified.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s40995-018-0619-4) contains supplementary material, which is available to authorized users.

Snigdha Pattnaik
snigdhapattnaik@soa.ac.in

[1] School of Pharmaceutical Sciences, Siksha 'O' Anusandhan Deemed to be University, Kalinga Nagar, Ghatikia, Bhubaneswar, Odisha 751003, India

[2] International Institute of Information Technology, Pune, Maharastra 411057, India

Published online: 01 August 2018

Temporal Variability of Aerosol Parameters over a Metropolitan AERONET Observing Site at Pune.

Sandeep Varpe [1, 2*], Mahesh Waghmare [2], Swati Kolet , Amit Kasar[2], Rupali Varpe [1], Suvarna Bhagwat[2], Rupali Yeole[2], Mandar Datar[2]

Physics Department, Nowrosjee Wadia College, Pune 411 001, India

International Institute of Information Technology P-14, Rajiv Gandhi Infotech Park, Hinjawadi, Pune 411057, India.

E-mail :sandeepvarpe@rediffmail.com

Abstract: The effect of atmospheric aerosol particle depends on the spatial and vertical distribution of aerosol. This paper is based on the observations taken at the metropolitan AERONET observing station at Pune, India. In this paper the multiyear AERONET data on sun measurement and almucantar scan retrieval at Pune have been analyzed to investigate heterogeneity in aerosol optical and microphysical properties. The data analysis gives the existence of large seasonal heterogeneity in the frequency distribution of AOD_{500} mm, AOD_{1020} mm, Precipitable water vapour and correspondingly retrieves $AE_{440\text{-}870}$ mm during different seasons at Pune. The assessment of possible aerosol impact on the Indian monsoon circulation gives precise estimate of aerosol absorption.

Keywords: Aerosol optical depth, Angstrom exponent, Precipitable water vapour.

A. INTRODUCTION

The atmosphere consists of tiny colloidal particles of different chemical composition, dispersed/embedded in a gas, smoke or fog, termed as aerosols [1, 2]. These particles linger in the atmosphere for a long period of time and their lifetime is found to vary considerably depending on their size and composition. The coarse-mode volcanic dust particles injected into the atmosphere during large volcanic eruptions have a short lifetime of about 1–2 months due to their fast removal by gravitational settling [3]. On the other hand, tropospheric sulfate aerosols, mostly in the accumulation-mode size regime [4] have a lifetime of 5–7 days. A recent study [4, 5, 6, 7, and 2] has retrieved a significantly longer lifetime of tropospheric aerosols by diagnosing the e-folding of the deposited Cesium during a major nuclear power plant accident at Fukushima Dai-Ichi in Japan during March 2011. In the atmosphere, aerosols are produced from a mixture of natural (e.g., biogenic and volcanic) and anthropogenic activities viz., urban/industrial, biomass burning, fossil fuel combustion, etc. The worldwide investigations of aerosol optical, microphysical, radiative and compositional properties have established that the aerosols disperse widely both on horizontal and vertical scales by means of circulation patterns in the atmosphere [8, 9, 10, 11, and 12]. Aerosols perturb the Earth-atmosphere climate system at a local, regional and global scale through direct and indirect processes.

B. OBSERVATIONAL SITE AND METEOROLOGY

The present study is focused on a metropolitan AERONET observing site at Pune in Maharashtra. Pune is the eighth largest metropolitan city in India and has grown noticeably with a rapid increase in urbanization, industrial growth (automobile manufacturing, chemicals, foundries, food processing, etc.)

Artificial Intelligence: At A Glance

Amit Kasar[1]
Applied Sciences & Engineering Dept.
istant Professor, International Institute of Information Technology, Hinjewadi, Pune, India
amitk@isquareit.edu.in

Mayuri Dangare[2]
Computer Science Department
[2]Assistant Professor Sinhgad College of Science
Pune, india
danger.mayuri@gmail.com

stract – *The invention in Artificial Intelligence has based ools and techniques of many different disciplines, ding formal logic Linguistic intelligence, Musical elligence, Logical-mathematical intelligence, Intra-personal elligence Spatial intelligence, bodily kinesthetic ligence, probability theory, power systems management nce, decision theory, computer science, mathematics, ology, linguistics and philosophy. The application of these iplines in AI has necessitated invention of many ements and extensions. The evolution has some as .This paper focus on the* advancement, *current and constraint of Artificial Intelligence.*

Artificial Intelligence, AI evolution, AI , AI Disadvantages

1. INTRODUCTION

ld of artificial intelligence came into existence in the Its intention was that to get a computer able to do tasks human. If machine can do the things the human can do it can be said to be Intelligent Machine. This intelligent ne can mimic humans and then do it themselves or better human.

Invention of Artificial Intelligence

lan Turing is published the "Computing Machinery & Intelligence" in the year of 1950. He suggest "the imitation game" also known as the "Turing Test" in which a machine s to impersonate itself as a human being in an imitation by giving human-like responses to a series of questions. lieved that if a machine could make a human being e that he or she is communicating with another human then machine can be considered as intelligent as a being. On August 31, 1955, the concept of "artificial e" is produced in a proposal for a "2 month, 10 man artificial intelligence" submitted by John McCarthy outh College), Marvin Minsky (Harvard University), el Rochester (IBM), and Claude Shannon (Bell Laboratories). The workshop held in July and 956, is generally considered as new beginning.

During the 1980s and 1990s, new approach was made in artificial intelligence to build Artificial Intelligence to overcome problems human minds cannot calculate. Real world problems for example such as in medicine, finance and aeronautics. But this Artificial Intelligence has the ability to be faster and better than a human mind .Consequence of this was now the artificial system could potentially outperform a human brain. (Garrido, 2010)

December 1955	Herbert Simon and Allen Newell developed the Logic Theorist, the first artificial intelligence program
1957	Frank Rosenblatt developed the Perceptron. It is early artificial neural network enabling pattern recognition based on a two-layer computer learning network
1958	John McCarthy developed LISP programming language
1959	Arthur Samuel coins the term "machine learning," reporting on programming a computer "so that it will learn to play a better game of checkers than can be played by the person who wrote the program."
1959	Oliver Selfridge publishes "Pandemonium: A paradigm for learning" in the *Proceedings of the Symposium on Mechanization of Thought Processes*, in which he describes a model for a process by which computers could recognize patterns that have not been specified in advance
1961	Unimate, the first industrial robot, started working in New Jersey on an assembly line in a General Motors plant.
1961	James Slagle develops SAINT (Symbolic Automatic Integrator), a heuristic program that solved symbolic integration problems in freshman calculus.
1965	Joseph Weizenbaum developed ELIZA. It is an interactive program that carries on a dialogue in English language on any topic.
1965	Edward Feigenbaum, Bruce G. Buchanan, Joshua Lederberg, and Carl Djerassi started working on DENDRAL at Stanford University. DENDRAL is the first expert system. It automated the decision-making process and problem-solving behavior of organic chemists, with the general aim of studying hypothesis formation and constructing models of empirical induction in science.

Empirical analysis of Total Factor Productivity growth of Indian Steel Industry

Priyanka Narwade[1], Harshad Sonar[2*],

[1]Assisiant Professor, International Institute of Information Technology (I²IT), Pune, India- 411057

[2*]Research Scholar, National Institute of Industrial Engineering (NITIE), Mumbai, India- 400 087

Email id: narwadepriyanka3@gmail.com[1], harshadsonar@gmail.com[2*]

ABSTRACT

The steel industry is one of the basic or key industries in the national economy of any country. Keeping this as a backdrop, this paper evaluates productivity growth for three selected India's leading steel manufacturers which includes SAIL, Tata Steel and JSW. The study is based on firm-level panel data drawn from PROWESS, compiled by the Center for Monitoring Indian Economy (CMIE) for the period from FY 2007-08 to 2016-17. The estimation of productivity will be very useful to evaluate the variations in the performance of industry over a period. As far as steel industry is concerned, it has been observed that average TFP for selected industries is 1.21 during the study period. The results on partial factor productivity of factors show consistency in productivity of material, labour and capital. The result on the overall productivity shows declining total factor productivity growth over the study period. Very few literature is available which compares the productivity of India's leading steel industries. This study attempted to measure and compare productivity of three major steel producers in India.

Keywords: Total Factor Productivity, Steel Industry, Productivity Measurement, PROWESS

1 INTRODUCTION

The rapid development of steel production capacity is indeed a necessity for rapid industrialization of any developing country. The steel industry is one of the basic or key industries in the national economy of any country. The iron and steel industry comprises one of the main foundations on which the industrial structure of the country can be built. The Indian steel sector is booming, and now it occupies the fifth position globally. India's crude steel output increased by 5.4 percent to 61.8 million tonnes (MT) in the first seven months of the calendar year 2018, according to the World Steel Association. India is the world's third-largest producer of crude steel. The growth in the Indian steel sector has been driven by the domestic availability of raw materials such as iron ore and cost-effective labor. Steel consumption in India is growing. However, the pace of growth is much slower in contrast with world consumption.

In recent years, the burden of global competition has forced companies to focus on strategies for productivity improvements. In this context, total productivity measurement is needed in order to boost internal efficiency and thereby the competitiveness of a business unit (Hannula, 2002). Total factor productivity (TFP) growth is essential and therefore, TFP growth became synonymous with long-term growth as it reflects the potential for growth (Mahadevan, 2004; Rao, 1996). At the business unit level, productivity measures belong mainly to the group of non-financial measures. Any of the operational phases of a business unit, including purchasing, promoting, finance, sales, and support services contribute to total productivity. In the current scenario, the Indian steel industry plays a significant role in the country's economic growth. The country has also attained an imperative position on the global steel map due to its giant steel mills, acquisition of global scale capacities by players, continuous modernization & up gradation of old plants, improving energy efficiency, and backward integration into global raw material sources (Singh and Raina, 2015).

International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249 – 8958, Volume-8 Issue-3, February 2019

Investigation of Performance of Heat Transfer with the Help of Coiled Wire Inserts and Dimpled Tube

Deshmukh Yogiraj Ramakantrao, K. V. Narasimha Rao

Abstract: *Heat transfer enhancement is classified into active and passive methods. Active techniques require external power to type in the process; in comparison, passive methods don't require any extra power to enhance the thermo hydraulic overall performance of the product. Passive techniques are popular in each numerical and experimental uses when investigating heat transfer enhancement and friction losses to help save costs and power. The numerous passive ways for increasing heat transfer rate include different elements placed in the fluid flow path, like coiled wire, coiled and tangled cables, as well nozzle inhalator's. The existing paper belongs to an extensive review which centered on heat transfer enhancement techniques with coiled wire plus coiled wire inserts because the assembly of inserts is simpler and much more efficient. The study work involves experimentation and numerical evaluation of dimpled tube built with frequently spaced coiled wire inserts. In this paper, the explanation of the experimental set in place and instrumentation used, together with the experimental data and process reduction is presented. The primary goal of the testing is obtaining experimental data pertaining to heat transfer and fluid flow. Unique trials had been carried out with mixtures of dimpled tubes and coiled wire with different Reynolds number for warm fluid and by holding regular flow of cool fluid through annulus.*

Index Term : *Heat, existing, presented, Reynolds, Unique ,comparison.*

I. INTRODUCTION

Heat exchangers are utilized in various tasks ranging from conversion, recovery and utilization of thermal power in different industrial, domestic and commercial applications. A few typical examples include steam generation, condensation in power & cogeneration plants, smart heating & cooling in winter processing of substance, agricultural and pharmaceutical products, and fluid heating in producing & waste heat healing etc. Increase in heat exchanger's efficiency is able to result in less expensive look of heat exchanger that will assist with generate power, materials & cost savings associated with a heat exchange system. The requirement to boost the winter functionality of heat exchangers, therefore affecting power, materials & cost savings have resulted in use and growth of countless methods called as heat transfer augmentation. These methods can also be known as heat transfer enhancement and also intensification. Augmentation strategies grow convective heat transfer by decreasing the winter opposition in a heat exchanger. Use of Heat transfer enhancement methods lead increasing in heat transfer coefficient but in the price of increased pressure fall.

Revised Manuscript Received on 24 February 2019.

Deshmukh Yogiraj Ramakantrao, Department of Mechanical Engg, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur Dist, Andhra Pradesh, India.

K. V. Narasimha Rao Professor, Department of Mechanical Engg, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur Dist, Andhra Pradesh, India.

Thus, while developing a heat exchanger through these strategies, evaluation of heat transfer rate & pressure fall must be completed. Aside from this, problems like long term efficiency & detailed financial evaluation of heat exchanger must be studied. In order to attain higher heat transfer rate in a current or maybe new heat exchanger while caring for the greater pumping electricity, a few methods are recommended in recent years. Coiled wire a kind of passive heat transfer augmentation methods show drastically excellent outcomes in previous studies. For experimental work, various layouts of coiled wire utilized are normal coiled wire (TT). All these tapes are analyzed with 3 distinct twist proportions (y = 4.213, 5.337, 6.4606) as well as level of cut (d = six mm) for notched coiled wire.

II. REVIEW ON HEAT TRANSFER TECHNIQUES

2.1 Heat Transfer Enhancement with Coiled Wire Inserts

The enhancement of heat transfer rate by making use of coiled wire inserts is commonly studied both numerically and experimentally by a lot of scientists. Garcia et al. (2012) [1] analyzed the thermo hydraulic functionality of 3 kinds of passive heat transfer development dependent on synthetic roughness. Based on the end result, the usage of coiled wire in a reduced Reynolds number was much more useful compared to dimpled and corrugated tubes. Ozceyhan (2005) [2] numerically examined conjugate heat transfer as well as winter pressure in a tube with coiled wire inserts. Gunes et al. (2010) [3] investigated the consequences of coiled wire on heat transfer enhancement as well as pressure fall in a tube. It was discovered that top overall enhancement effectiveness was attained because of the configuration with P/D = one. In an additional study Gunes et al. (2010) [4] found on the characteristics of heat transfer as well as pressure fall in a tube with coiled wire inserts separated from the tube wall by 2 various distances. The results demonstrated that the Nusselt quantity as well as pressure fall expansion with minimizing distance between the coiled wire as well as the tube wall. Promvonge (2007) [5] given the experimental success because of the heat transfer as well as flow friction qualities in a circular tube where coiled wire with a square cross part was introduced.

The winter functionality of helically coiled wire was experimentally examined by Eiamsaard et al. (2012) [6]. They found the heat transfer rate as well as the friction component increase as the twist ratio as well as helical pitch ratio increase. In an additional study, Eiamsaard et al.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878, Volume-7, Issue-6C2, April 2019

An Analysis of Changing Economic Trends and Rapid Urbanisation Complicate Solid Waste Management

Madhuri Reddy, SS.Asadi

Abstract: Urbanization and population growth are exclusively in charge of high expanding rate of solid waste and its legitimate management is a noteworthy issue of Municipal Corporation. Solid waste damps are truly ruining the environmental conditions in creating nations. Negative environmental effects from inappropriate solid waste dumping can be effortlessly watched wherever in the creating scene. In this paper, A multicriteria decision making technique, named as Analytical Hierarchy Process (AHP), has been used to fulfill of all criterion objectives, criteria, sub criteria, and alternatives is applied in selection of an appropriate solid waste treatment technology. The obtained inputs with the help of experts have been used in pair wise comparison matrix in order to rank the technologies. These comparisons have been used to obtain the weights of importance of the decision criteria, and the relative performance measures of the alternatives. The finding of this study shows that the combination of recycling and composting technology is the most appropriate solid waste treatment technology and recommended to be implementing in Pune Municipal Council.

Index Terms: Population, solid waste, management, environment, etc.

I. INTRODUCTION

Humans have produced garbage from man's soonest time. Since the beginning garbage has been managed through dumping, burning, recycling and minimization. The gathering, treating and discarding solid waste came into human history not long after the beginning of industrialization and populace development of urban zones. Today ill-advised solid-waste management can cause negative impacts on our health through episodes of illnesses and our environment through toxins. Anyway "waste" is an emotional term since some consider waste to be a general health and environment hazard and others consider it to be a burden that is a wellspring of salary. Solid waste is a of/rejected material by the general public created various sources which is one of the real consuming on the planet. The quantity of solid waste is expanding from year to year as the populace and industry is at the rising end. Inappropriate

Dealing with and heaping up of solid waste is prompting unhygienic conditions in urban and major rural environments in India. The solid waste dumping yards in India are getting to be as a rearing reason for flies and stray creatures. Therefore, a few vector borne infections are spreading on the planet. Populace rise is the real explanation behind increment of solid waste due to escalated rural and industrial exercises. Utilization of normal assets is expanding quite a long time prompting weight on characteristic assets. As of late utilization of industrial goods are additionally expanding at worldwide dimension. This situation is prompting enormous solid waste age in all nations. Especially, the created nations are delivering more than the undeveloped nations. In any case, the kind of solid waste is fluctuating starting with one city then onto the next city. Essentially, concoction organization is likewise fluctuating from place to put.

A. Concept of Solid Waste and its types

Waste' in genuine sense undesirable or undesired material or might be any substance which is pointless in present setting. Contingent on the time and phrasing or the kind of material, it is otherwise called garbage, rubbish, trash or junk according to comfort. When it is with reference to living life forms, it alludes to undesirable substances or harmful materials which are to be removed out of their body. Waste management is the human control connected on the collection, treatment, and transfer of differed sorts of wastes. This is done to diminish the negative effects which are conceivable on the nearby or worldwide environment and society.

The rate of per capita waste generation in pune city is one and half kilogram and it is expanding with time. In the neighborhood town, in the city territory the city enterprise is the specialist to deal with the waste in the point of view zone. Pune City Corporation gathers 42 percent solid waste and dumps in the dumping destinations. Rest of the waste remains unmanaged in the source place and store locales. There was no logical management framework for the waste in Pune city and other town zone. In the Dhaka city territory 50 percent of the aggregate populace utilizes the dustbin for their waste. In basic word, solid waste can be characterized as any undesirable or disposed of materials. Human being and their exercises will produce waste. It isn't conceivable to maintain a strategic distance from waste in our life yet legitimate management is fundamental for waste to keep the environment healthy. One - fourth of the aggregate populace in the urban territory is confronting ill-advised management of the waste. The business and hospital waste are frightfully contaminating our environment. The household wastes likewise unmanaged and dumping all over consistently

Revised Manuscript Received on April 09, 2019.
Madhuri Reddy, Department of Civil Engineering., Koneru Lakshmaiah Education Foundation, Vaddeswaram ,Guntur District , Andhra Pradesh, India.
SS. Asadi, Department of Civil Engineering. Vignan's Foundation for Science, Technology and Research, Deemed to be University, A.P, India.