IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Analysis of Emotion: A Technical Approach.

1Chandrakant N. Patil, 2Pratik S. Jaiswal, 3Mukta G. Dhopeshwarkar

1P.hD. Scholar, 2M.Phil. Scholar , 3Assistant Professor

1Department of Computer Science & I.T, Dr. Babasaheb Ambedkar Marathwada University, Aurangabad, ,

2Department of Computer Science & I.T, Dr. Babasaheb Ambedkar Marathwada University, Aurangabad, ,

3Department of Computer Science & I.T, Dr. Babasaheb Ambedkar Marathwada University, Aurangabad,

Abstract: This paper discusses the use of EEG waves to identify emotions. The approach used, the dataset used for simulation, the results obtained, as well as the limitations and future work/gaps are all summarized in this review paper. This allows future researchers to focus on the problems to be addressed and the methods to be provided, which might be entirely new or a combination of existing approaches or algorithms, as well as the dataset to be utilized.

Keywords: EEG, mRMR, IMF, KNN, Neural Network.

1. INTRODUCTION:

Emotion is prevalent in all people and combines intelligence and consciousness. Emotions are important in displaying effectiveness. As a result, the study of human emotions has expanded in fields such as computer science, neuroscience, artificial intelligence, cognitive science, and psychology. Because human emotions are linked to the brain, EGG may be used to classify them. The aggregate behavior of human emotions may be adequately described using EEG data.

Human emotion is exceedingly complicated, encompassing not only the psychological response to the external environment, but also the physiological reactions to the psychological reactions. The valence-arousal scale by James A. Russell, shown in Figure 1, is a 2D model that has been used to characterize human emotion.

The four main ideas may be split using the scale's quarters, with the remainder of the sentiments falling into one of three categories: good, poor, or neutral. The emotion characterized by the valence-arousal scale may be seen as a compass, with the horizontal (east to west) dimension representing the pleasant to unpleasant, or negative to positive. The vertical (south to north) component, on the other hand, represents sleep to alertness or low to high engagement. Based on the activation degree and pleasantness of the emotion, the emotion will be coordinated into the valence arousal scale.

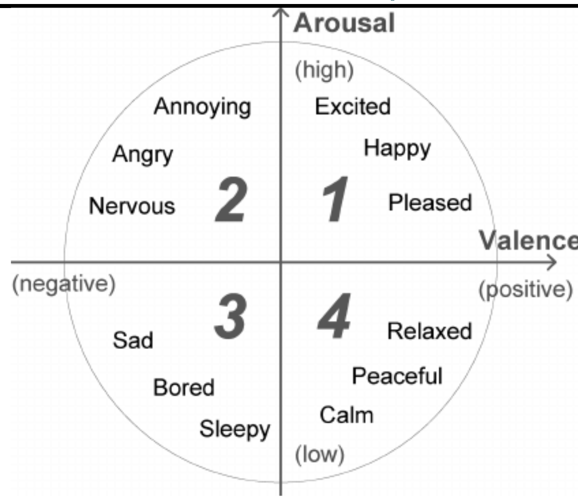


Figure 1: 2D Valance-arousal emotion space

2. LITERATURE REVIEW

Suitability in EEG-based emotion recognition. Various feature extraction algorithms have been applied, mRMR was used for feature choice and classification was done the usage of SVM and Random Forests [1]. In the research, three records are used to advocate an emotion cognizance gadget based on EMD. To decide the efficiency of the traits for emotion categorization, a thorough investigation used to be conducted. The findings demonstrate that the three characteristics are suitable for emotion recognition. The influence of every IMF element is
Sub-sequent investigated. Finally, the new strategy is tested against a number of traditional strategies & proven to be the most accurate [2]. They suggest an emotion identification device for human talent indicators based totally on EEG records in this study. They take EEG alerts that are related to emotion, split them into five frequency bands based totally on power spectrum density, then do away with low frequencies from zero to 4 Hz to dispose of EEG artefacts [3].

They advise a feature extraction approach based totally on multivariate empirical mode decomposition (MEMD) for emotion detection as high/low arousal and high/low valence states in this research. Benchmarking is performed the use of multichannel EEG recordings from the publicly handy DEAP emotional EEG data set, and the findings of past investigations are compared to the new MEMD-based technique [4].

The literature evaluation exhibits two opposing viewpoints. Using completely the EEG output, one set of researchers done negative emotion categorization accuracy. This crew believes that physiological signals such as galvanic skin resistance, heart rate, and breathing rate, among others, should be exploited to obtain excessive accuracy. Another crew believes that the EEG signal is adequate for emotion categorization, however with a lesser degree of accuracy [5]. The 3D-CNN emotion identification approach is developed in this lookup to extract spatiotemporal characteristics in order to characterize the temporal connections between EEG data. Because the 3D-CNN requires 3D inputs, a new strategy for changing multi-channel EEG facts to a 3D format has been created [6].

The frequency bands and channels employed for EEG-based emotion recognition were enhanced in this study [7]. Recognizing a person's feelings and comprehending his or her mental condition has been regarded critical. The specified future scope is boundless and has many good effects in various industries, including medical, security and surveillance, education, product marketing, and so on, with more futuristic uses on the horizon, notably in the field of security and surveillance [8].

Recognizing and comprehending a person's feelings and mental condition has been regarded critical. The future scope outlined is boundless and has several beneficial implications in a variety of industries, including medical, security and surveillance, education, product marketing, and so on, with more futuristic uses on the horizon, notably in the field of security and surveillance[9]. The emotions were classified as Low/High Arousal, Low/High Valence, and Accuracy [10]. In this work, Matlab software is used to programme and test the system. The approach does not require radiologists to identify the cancerous region of the image. Following the execution of the programme, the geometry and intensity characteristics are received as output. If the nodule's area and perimeter are both big, the nodule is confirmed to be malignant [11].

Table 1. Authors, techniques and result obtained in comparative analysis

Sr. No.	Author	Title of paper	Techniques	Conclusion
1.	Pascal Ackermann	EEG-based Automatic Emotion Recognition: Feature Extraction, Selection and Classification Methods	IMFs, mRMR,	Suitability in EEG-based emotion recognition. Various feature extraction algorithms were applied, mRMR was used for feature selection and classification was done using SVM an Random Forests.
22.	Ning Zhuang	Emotion Recognition from EEG Signals Using Multidimensional Information in EMD Domain.	empirical mode decomposition (EMD), Intrinsic Mode Functions (IMFs), fractal dimension (FD)	In this research, three statistics are used to suggest an emotion recognition system based on EMD. To determine the efficiency of the characteristics for emotion categorization, a thorough investigation was conducted. The findings demonstrate that the three characteristics are appropriate for emotion recognition. The impact of each IMF component is next investigated. Finally, the new approach is tested against various traditional methods and shown to be the most accurate.
33.	Kwang-Eun Ko	Emotion Recognition using EEG Signals with Relative Power Values and Bayesian Network.	Bayesian network, QEEG-8, FFT	They propose an emotion identification system for human brain signals based on EEG data in this study. They take EEG signals that are related to emotion, split them into five frequency bands based on power spectrum density, then remove low frequencies from 0 to 4 Hz to remove EEG artefacts.
44.	Ahmet Mert	Emotion recognition from EEG signals by using multivariate empirical mode decomposition.	multichannel IMFs, empirical mode decomposition (EMD)	They propose a feature extraction technique based on multivariate empirical mode decomposition (MEMD) for emotion detection as high/low arousal and high/low valence states in this research. Benchmarking is done using multichannel EEG recordings from the publicly available DEAP emotional EEG data set, and the findings of past investigations are compared to the new MEMD-based technique.
55.	Mandeep Singh	Emotion Recognition Using Electroencephalography (EEG): A Review	Functional Magnetic Resonance Imaging(FMRI),Positron Emission Tomography (PET)	The literature review reveals two opposing viewpoints. Using solely the EEG output, one set of researchers achieved poor emotion categorization accuracy. This group believes that physiological signals such as galvanic skin resistance, heart rate, and breathing rate, among others, should be exploited to achieve high accuracy. Another group believes that the EEG signal is sufficient for emotion categorization, but with a lesser degree of accuracy.
66.	Elham S. Salama	EEG-Based Emotion Recognition using 3D Convolutional Neural Networks	3D-CNN, Fast Fourier Transform (FFT), SLC, MLC	The 3D-CNN emotion identification technique is developed in this research to extract spatiotemporal characteristics in order to characterize the temporal connections between EEG data. Because the 3D-CNN requires 3D inputs, a new approach for converting multi-channel EEG data to a 3D format has been created.
77.	Mi Li	Emotion recognition from multichannel EEG signals using K-nearest neighbor classification.	KNN	This study improved the frequency bands and channels used for EEG-based emotion identification.
88.	Raheena Bagwan	Facial Emotion Recognition using	CNN, Image Processing, Facial Emotion	Recognizing a person's feeling and understanding his or her state of mind has been

		Convolution Neural Network	Recognition, Deep Learning	deemed crucial. The future scope described is limitless and has many positive outcomes in many industries such as medical, security and surveillance, education, product marketing, and so on, with more futuristic applications on the horizon, particularly in the realm of security and surveillance.
9.	V. Ramachandran	Facial Expression Classification System with Emotional Back Propagation Neural Network	PCA, LDA, LPP	The findings are compared to those of a standard feed forward neural network with back propagation. The new approach outperforms the existing algorithm.
110.	Nitin Kumar	Bispectral Analysis of EEG for Emotion Recognition.	bi-spectral analysis,	The emotion classification was done into Low/High Arousal and Low/High Valence and an accuracy of 64.84% and 61.17%.
111.	N. Malligeswari	A Novel Approach for Lung Pattern Analysis using Neural Networks and Fuzzy Interface System	ANN, ANFIS, MDC	Matlab software is utilized in this study to programme and test the system. The technique does not need radiologists to identify the malignant area of the picture. The geometry and intensity features are obtained as output once the programme has been executed. If the area and perimeter of the nodule are both large, the nodule is verified to be malignant.

3. CONCLUSION

As can be seen, several journal and conference studies on emotion detection using EEG data have been reviewed. In today's world, human emotion is a cutting-edge signal analysis, and determining validity is always a difficulty. The current state of the art techniques and strategies for identifying emotion using EEG waves have been investigated. Various datasets used for testing were studied, and it should be noted that the DEAP statistics set was formerly used as a popular reference by the majority of the researchers. Furthermore, the findings of categorization accuracy in terms of arousal and valence have been established for overall performance evaluation. As a consequence, future researchers will be able to employ or combine a variety of existing approaches to improve classification accuracy in the near future.

4. Acknowledgment

I would like to thank the department of computer science & I.T, Dr.Babasaheb Ambedkar Marathwada University, Aurangabad for providing me the infrastructure.

5. REFERENCES

[1] P. Ackermann, C. Kohlschein, J. Á. Bitsch, K. Wehrle and S. Jeschke, "EEG-based automatic emotion recognition: Feature extraction, selection and classification methods," *2016 IEEE 18th International Conference on e-Health Networking, Applications and Services (Healthcom)*, 2016, pp. 1-6, doi: 10.1109/HealthCom.2016.7749447.

[2] Ning Zhuang,1 Ying Zeng,1,2 Li Tong,1 Chi Zhang,1 Hanming Zhang,1 and Bin Yan1 , 1China National" Emotion Recognition from EEG Signals Using Multidimensional Information in EMD Domain" Hindawi BioMed Research International Volume 2017, Article ID 8317357, 9 pages https://doi.org/10.1155/2017/8317357

[3] Kwang-Eun Ko, Hyun-Chang Yang, and Kwee-Bo Sim* "Emotion Recognition using EEG Signals with Relative Power Values and Bayesian Network" International Journal of Control, Automation, and Systems (2009) 7(5):865-870 DOI 10.1007/s12555-009-0521-0 http://www.springer.com/12555

[4] Ahmet Mert• Aydin Akan." Emotion recognition from EEG signals by using multivariate empirical mode decomposition " Pattern Anal Applic (2018) 21:81–89https://doi.org/10.1007/s10044-016-0567-6 Received: 16 June 2015 / Accepted: 18 June 2016 / Published online: 29 June 2016_ Springer-Verlag London 2016

[5] Mandeep Singh[1], Mooninder Singh[2], Surabhi Gangwar[3] "Emotion Recognition Using Electroencephalography (EEG): A Review" IJITKMI Volume 7 • Number 1 • December 2013 pp. 1-5.
[6] Elham S.Salama, Reda A.El-Khoribi, Mahmoud E.Shoman, Mohamed A.Wahby Shalaby " EEG-Based Emotion Recognition using 3D Convolutional Neural Networks" (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 9, No. 8, 2018 www.ijacsa.thesai.org
[7] Mi Lia;c, Hongpei Xua;c, Xingwang Liua;c and Shengfu Lua, "Emotion recognition from multichannel EEG signals using K-nearest neighbor classification" Technology and Health Care 26 (2018) S509–S519 DOI 10.3233/THC-174836 IOS Press.
[8] Raheena Bagwan1, Sakshi Chintawar1, Komal Dhapudkar1, Alisha Balamwar1, Prof. Sandeep Gore2 "Facial Emotion Recognition using Convolution Neural Network". International Journal of Trend in Scientific Research and Development (IJTSRD) Volume 5 Issue 3, March-April 2021 Available Online: www.ijtsrd.com e-ISSN: 2456 – 6470.
[9] V.Ramachandran B. Sivaiah Dr. E. Srinivasa Reddy "Facial Expression Classification System with Emotional Back Propagation Neural Network" International Journal of Scientific & Engineering Research, Volume 4, Issue 10, October-2013 ISSN 2229-5518 http://www.ijser.org
[10] Nitin Kumara,, Kaushikee Khaunda, Shyamanta M. Hazarikaa "Bispectral Analysis of EEG for Emotion Recognition" 7th International conference on Intelligent Human Computer Interaction, IHCI 2015 Available online at www.sciencedirect.com
[11] N.Malligeswari ,2C.Rajani , 3 Dr.G.Kavya "A Novel Approach for Lung Pattern Analysis using Neural Networks and Fuzzy Interface System" Advances in Engineering Research (AER), volume 142. International Conference for Phoenixes on Emerging Current Trends in Engineering and Management (PECTEAM 2018) license (http://creativecommons.org/licenses/by-nc/4.0/).

IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Analyzing Political Opinions and Prediction of Election Result of The Indian Election Using Data Mining Approach

[1]Ashweeni A. Kulkarni, [2]Mukta G. Dhopeshwarkar, [3]Pratik S. Jaiswal

[1]M.Tech Research Scholar, [2]Assistant Professor & Research Guide, [3]M.Phil Research Scholar

[1]Department of Computer Science and IT,

[1]Dr. Babasaheb Ambedkar Marathwada University, Aurangabad 431004, India

Abstract: Sentiment Analysis is considered in a category of natural language processing and machine learning. Sentiment analysis has been a popular field for records and scientists. It is a technique of calculating sentiment of a particular statement or sentence for Political review, Movie review, social media like twitter review, hotel review and categorizes them as positive, negative and neutral. Election is conducted to find the public opinion, where candidate choose by group of people using votes and many methods to predict result, Although many agencies and media companies conduct pre poll survey and expert views to predict result of election. We use twitter data to predict outcome of election by collecting twitter data and analyze it to predict the outcome of the election by analyzing sentiment of twitter data about the candidates. We used Machine learning and lexicon based approach to find emotions in tweets and predict sentiment score. We performed data (text) mining on Political and Election based generated tweets. We utilized Dictionary based approach, Naïve Bayes, Support Vector Machine and Decision Tree algorithm to build our classifier and classified the test data as positive, negative and neutral. We also utilized comparative study between classifier for better accuracy result. We identified the sentiment of Twitter users towards each of the considered Indian political leaders and national political parties. We begin to use the case study by selecting 3 National Parties attend in 17th Lok Sabha Election of Indian General Election. We indicate that Naive Bayes can perform anticipation and classification processes with high accuracy in compared with two other algorithms to anticipate participation.

Index Terms **- Sentiment Analysis, Naïve Bayes, Support Vector Machine, Decision Tree, Sentiment Analysis with Python, SentiWordNet dictionary, Indian Election.**

I. INTRODUCTION

Elections are of utmost importance in every Democratic country. As we all know, “DEMOCRACY IS FOR THE PEOPLE,BY THE PEOPLE ,OF THE PEOPLE”, democracy is defined as a government of the people, for the people and by the people. All the powers in democracy are in the hand of people. Election is the process of voting to choose someone to be their political leader or the representative in government. People gives equal right for every citizen of the nation to select his leader. For every citizen, Vote is prominently considered in election and people are very concerned towards the winner of their choice, so they express their views on exit polls or opinion poll on results before election.

An opinion poll has existed since the early 19th century[1]. Currently, there are many scientifically proven statistical models to forecast an election[2]. But sometimes, even in the developed countries, the polls failed to accurately predict the election outcomes. Past research shows that several failed polls result such as in the 2004 European elections in Portugal, the 1992 British General Elections, 2007 French presidential elections, the 1998 Quebec Elections the 2006 Italian General Elections, the 2002 and the 2008 Primary Elections in the States.[3]

Latterly, it is observed that traditional polls may fail to make an accurate prediction. Then the scientific community has turned its interest in analyzing web data, such as blog posts or social networks users activity as an alternative way to predict election outcomes, hopefully more accurate.

Social media is a part of our daily lives from some years now. People increasingly tend to express their opinions via social media platforms. On a daily basis, data generated from social media is large volume of data. The question arrives for the future discussion is that can we use these data in order to detect trends, preferences, patterns and predict outcomes of future event? Social media is used for research and more specifically Twitter is used.. Twitter is considered one of the most successful social media. The community of the popular platform counts more than 328.000.000 active users at the moment [4]. Twitter is a microblogging web service that was launched in 2006. Now, it has more than 200 million visitors on a monthly basis and 500 million messages daily. The user of twitter can post a message (tweet) up to 140 characters. The message is then displayed at his/her personal page (timeline). Originally, tweets were intended to post status updates of the user, but these days, tweets can be about every imaginable topic. Based on the research in [5], rather than posting about the user’s current status, conversation and endorsement of content are more popular. [5] The advantages of using tweets as a data source are as follows; first, the number of tweets is very huge and they are available to the public. Second, tweets contain the opinion of people including their political view.

II. DATABASE INFORMATION

A. Sentiment Analysis.

Sentiment analysis is a text mining technique that uses machine learning and natural language processing (NLP) to automatically analyze text for the sentiment of the writer in positive, negative and neutral [6]. Sentiment analysis allows to organizing text like customer feedback or product reviews or political leaders review, first by category (Features, Shipping, Customer Service, Politics etc.), and then mining text for sentiment so you can see which categories are positive or negative or neutral.

It allows you to analyze thousands of online reviews or social media comments in just minutes. However, before performing any kind of sentiment analysis, you'll need to break down comments, paragraphs, or documents, into smaller fragments of text. Suppose Political Party Opinion, for example, 'I like BJP' sentiment towards this statement is positive, 'But it seems really slow as comparing to other parties' sentiment towards this statement is negative, 'I am ok with BJP, lets see what they can do in future' sentiment towards this statement is neutral. Like this it also contains multiple ideas or opinions, to analyzing the overall sentiment of reviews, tweets, documents, and so on.

B. Sentiment Analysis on Twitter

The goal of Twitter sentiment analysis is to classify tweets into three categories: positive sentiment class, negative sentiment class, and neutral sentiment class. Similarly, further calculations and functions, such as the most commonly used phrases, commonplace phrases, most commonly used emoticon, and frequent sentiment in the midst of the data, may be computed.

As previously stated, performing sentiment analysis on Twitter data is difficult.

The following are the reasons behind this:[7]

1. Limited Tweet size: With just 280 characters to work with, succinct statements are constructed, resulting in a limited number of possibilities. It's also tough to examine the numbers because of the use of slang, acronyms, and emoticons in tweets.
2. Slang: These words are distinct from English terminology, and their use might age a method owing to the growth of slangs.
3. Features of Twitter: Hashtags, customer references, emojis, and URLs are all supported. These need to be processed differently than other phrases.
4. User Variety: Customers express their opinions through a variety of tactics, including the employment of unique language in between, as well as the use of repeated words or symbols to convey an emotion. Some people choose to use a sequence of emoticons to represent a visual snapshot, while others prefer to make sarcastic words that appear to be something else but are highly recommended.

In our project, we have used the Twitter dataset. In that, we are fetching tweets regarding Loksabha Election 2019 and the parties which are included in the Loksabha Election 2019. We have used the dataset of 60000 tweets. In that, we used three different datasets of 20000 tweets for each party like datasets for BJP, NCP, and AAP. Tweets are fetching on monthly basis from December 2018 to March 2019. After fetching tweets for every month then merge that data to our existing dataset. Every month we have created a newly updated dataset for each party for getting the more accurate result.

We are fetching tweets using the following keywrds :

- **BJP**

'BJP', 'BhartiyaJanataParty', 'modi', 'NarendraModi

- **NCP**

'congress', 'NCP', 'Gandhi, 'RahulGandhi

- **AAP**

'AAP', 'ArvindKejriwal', 'AamAadamiParty'

Following are the SearchQueries which are used for fetching tweets :

#Congresswins, #BJPwins, #AAPwins, #CongressvsBJP, #RahulvsModi, #Loksabha, #LoksabhaElection, # LoksabhaElection2019, #BJPforIndia, #ModivsGrandAllowance, #BJP2019, #ModiforPM2019, #GeneralElection2019, # GeneralElection, #RahulforPM2019, #AAPwins2019, #kejariwal, #ArvindKejariwal, #BJP, #AAP, #NCP

We are using tweets per query is 100.

While fetching tweets, we are also fetching the location of the user who tweets, the count of total hashtags used in that tweet, the count of users mentioned in that tweet, and also the count of URLs and symbols used in that tweet.

III. METHODOLOGY

When we create dataset it was not in line way , so we need to clean database in an efficient manner. When we done entire database cleaning process, database setup is prepare to perform particular experiment on it. Then we calculate sentiment score for all the three perties and apply classfier on that database. We used Naïve Bayes classifier, Support Vector Machine classifier, Decision Tree Classifier. Above figure shows the entire execution process from row database collection to the final output of the classifier.

Following is the flow chart of the experiment

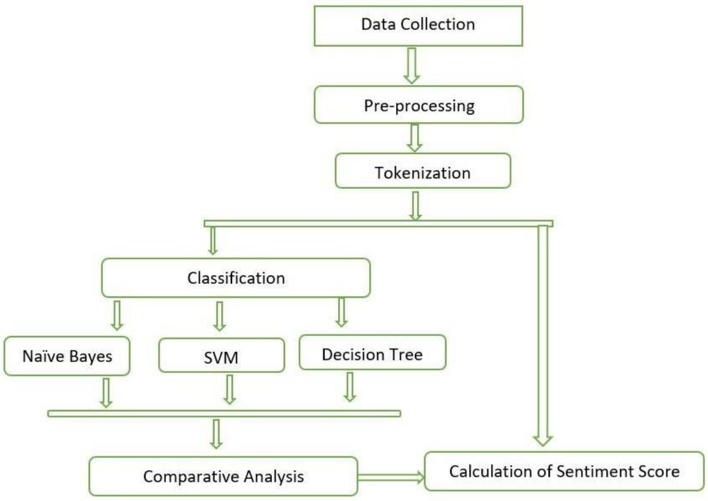


Figure : Flow Chart

IV. EXPERIMENT

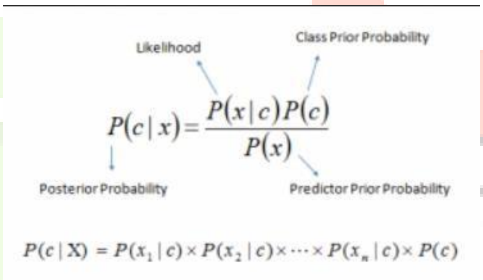
A. Naive Bayes

Naive Bayes is the easiest and fastest classification algorithm for a large chunk of data. In various applications such as junk mail filtering, textual content classification, sentiment analysis, and recommendation systems, Naive Bayes classifier is used successfully. It makes use of the Bayes chance theorem for unknown classification prediction.

The Naive Bayes classification approach is a simple and effective classification task in computer learning. The use of Bayes' theorem with a robust independence assumption between the facets is the foundation for naive Bayes classification. When used for textual information analysis, such as Natural Language Processing, the Naive Bayes classification yields true results.

Naive Bayes model is easy to construct and mainly beneficial for very giant records sets. Along with simplicity, Naive Bayes is recognised tooutperform even rather state-of-the-art classification methods.

Bayes theorem offers a way of calculating posterior probability P(c|x) from P(c), P(x) and P(x|c). Look at the equation under :



Above,

- P(c|x) is the posterior probability of class (c, target) given predictor (x, attributes).
- P(c) is the prior probability of class.
- P(x|c) is the likelihood which is the probability of predictor given class.
- P(x) is the prior probability of predictor.

Naïve Bayes Algorithm flow :

i. first, we have created a training and testing dataset for each party.
ii. Creating pickle file of that dataset
iii. Data Preprocessing
iv. Sentiment Analysis using Lexicon-based Method
v. Naive Bayes with TF-IDF on original text data
vi. Naive Bayes with TF-IDF on pre-processed text data achieved very minimal accuracy improvement
vii. Finally, we got precision, Recall, F1Score, Support and Accuracy of our dataset

B. Support Vector Machine

A universal learner is the Support Vector Machine. Both the input and output formats for the Support Vector Machine have been established. The input is vector space, and the output is either positive or negative. The document's text is unsuitable for learning. These texts are formatted in a structured manner. The text is converted into a format that the machine learning system can understand. The texts' scores are computed, and the results are then fed into the Support Vector Machine. The Support Vector Machine has been proven to be one of the most powerful text categorization learning systems.

However, text classification can occasionally result in an error. A text classifier comparison is required to determine which is superior across texts. In this scenario, the performance metric is employed.

Support vector machine Algorithm flow :

i. first, we have created a training and testing dataset for each party.
ii. Creating pickle file of that dataset
iii. Data Preprocessing
iv. Sentiment Analysis using Lexicon-based Method
v. SVM with TF-IDF on original text data
vi. SVM with TF-IDF on pre-processed text data - achieved very minimal accuracy improvement
vii. Finally, we got precision, Recall, F1Score, Support, and Accuracy of our dataset

C. Decision Tree

The most powerful and widely used tool for categorization and prediction is the decision tree. A decision tree is a flowchart-like tree structure in which each internal node represents an attribute test, each branch reflects the test's conclusion, and each leaf node (terminal node) stores a class label. As decision trees are supervised algorithms, they must be trained using annotated data.

As a result, the main notion is the same as for any text classification: given a set of documents (for example, TFIDF vectors) and their labels, the algorithm will determine how strongly each word connects with each label.

Decision Tree Algorithm flow :

i. first, we have created a training and testing dataset for each party.
ii. Creating pickle file of that dataset
iii. Data Preprocessing
iv. Sentiment Analysis using Lexicon-based Method
v. Decision Tree with TF-IDF on original text data
vi. Decision Tree with TF-IDF on pre-processed text data achieved very minimal accuracy improvement
vii. Finally, we got precision, Recall, F1Score, Support, and Accuracy of our dataset.

We consider the following evaluation measures in order to compute the overall performance of the system.

	Positives	negatives
positives	True Positive(tp)	False Positive(fp)
negatives	False Positive(fp)	False Negative(fn)

1. **Precision:** Precision is defined as portion of true positive predicted instances among all positive predicted instances.
Precision = $tp/tp + fp$
2. **Recall:** Recall is calculated as portion of true positive predicted instances against all actual positive instances.
Recall = tp /tp + fn
3. **Accuracy:** Accuracy basically is the portion of true predicted instances against all predicted instances.
Accuracy = $tp{+}tn/tp{+}tn{+}fp{+}fn$
4. **F-measure:** F-measure is the combination of Presicion and Recall and is calculated as:
F-Measure = 2∗Precision∗Recall/Recall + Precision

Preprocessing data :

- Convert every tweets to lower case
- Remove Twitter username
- Remove punctuations, numbers and special characters
- Convert more than 2 letter repetitions to 2 letter (example (wooooooow --> woow))
- Remove extra spaces
- Remove URLs
- Emoji analysis
- Handle contractions words " can't " >> " can not "
" won't " >> " will not " " should't " >> " should not "
- Remove Stop word

The following pie chart shows the locations of the tweets that we downloaded from Twitter.

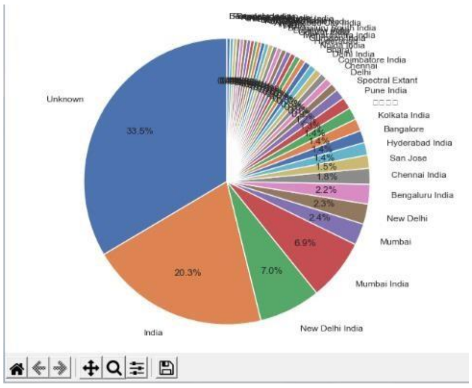


Figure: Tweets fetch for BJP from different location

4.1 BJP

The following bar chart shows the sentiment score and polarity of tweets for BJP. In that bar chart indicate nearly about 35% of Positive tweets, 21% of Negative tweets, 24% of Neutral tweets. The neutral intensity of the tweets indicates some of the users were neutral about the BJP they are neither positive nor negative.

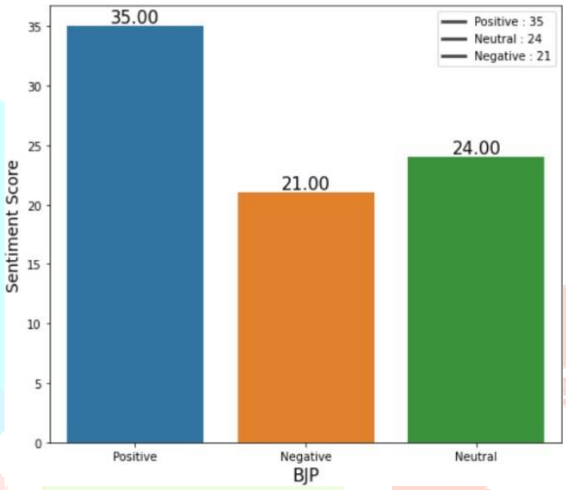


Figure: Sentiment Score for BJP

This is the word cloud for the positive tweets used for BJP. This word cloud we showing from our dataset.

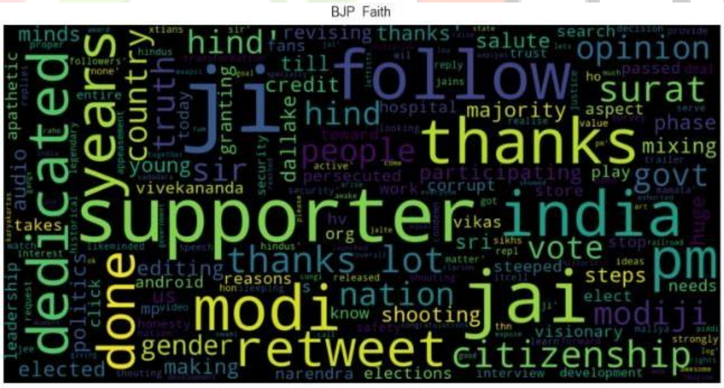


Figure : Positive word cloud for BJP

This is the word cloud for the Negative tweets used for BJP. This is the word cloud we showing from our dataset which indicates negative tweets.

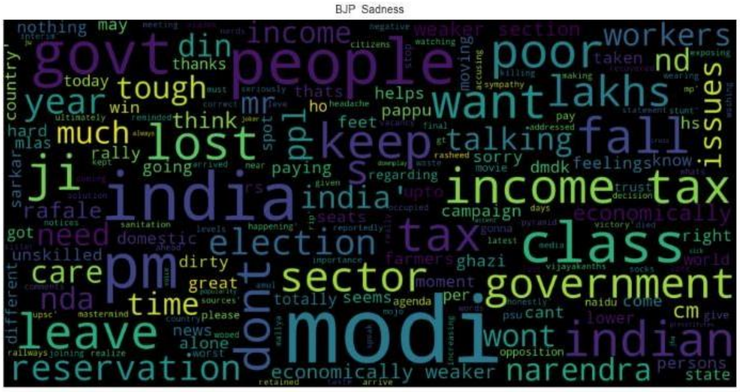


Figure : Negative word cloud for BJP

The following graph shows retweets detection, we are detecting tweets that are already used in our experiment and then removing that from our dataset. This will help to reduce the redundancy of the tweets and get a more accurate result.

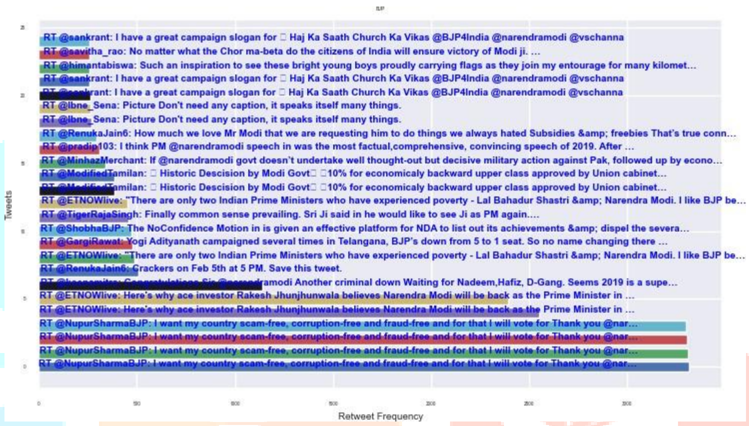


Figure : Retweets detection for BJP

Table 1 : Results using Naïve Bayes for BJP

	Precision	Recall	F1-Score	Support
positive	0.53	0.77	0.63	4
negative	0.56	0.75	0.64	9
Accuracy			0.87	100
Weighted avg	0.55	0.76	0.64	9

Table 2 : Results using SVM for BJP

	Precision	Recall	F1-Score	Support
positive	0.76	0.77	0.63	3
negative	0.75	0.75	0.64	8
Accuracy			0.79	100
Weighted avg	0.75	0.75	0.75	8

Table 3 : Results using Decision Tree for BJP

	Precision	Recall	F1-Score	Support
positive	0.42	0.53	0.47	3
negative	0.40	0.22	0.27	4
Accuracy			0.64	100
Weighted avg	0.41	0.33	0.39	3

4.2 NCP

The following bar chart shows the sentiment score and polarity of tweets for NCP. In that bar chart indicate nearly about 26% of Positive tweets, 35% of Negative tweets, 31% of Neutral tweets. The neutral intensity of the tweets indicates some of the users were neutral about the NCP they are neither positive nor negative.

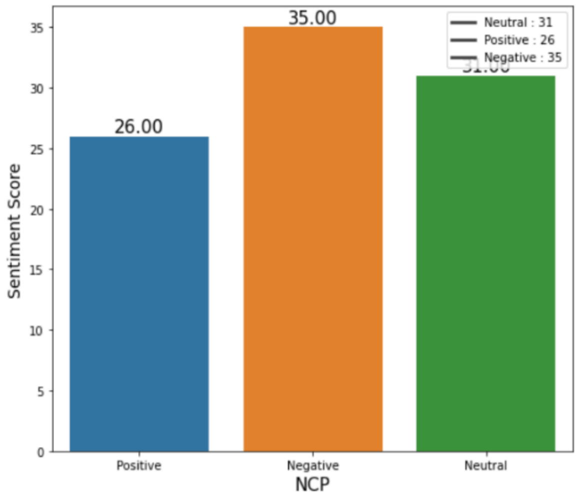


Figure: Sentiment Score for NCP

Table 4 : Results using Naïve Bayes for NCP

	Precision	Recall	F1-Score	Support
positive	0.42	0.55	0.48	5
negative	0.50	0.43	0.46	8
Accuracy			0.81	100
Weighted avg	0.46	0.51	0.47	6

Table 5 : Results using SVM for NCP

	Precision	Recall	F1-Score	Support
positive	0.41	0.76	0.53	5
negative	0.62	0.60	0.61	8
Accuracy			0.70	100
Weighted avg	0.52	0.65	0.55	7

Table 6 : Results using Decision Tree for NCP

	Precision	Recall	F1-Score	Support
positive	0.34	0.66	0.45	5
negative	0.33	0.12	0.19	1
Accuracy			0.64	100
Weighted avg	0.30	0.41	0.34	3

4.3 AAP

The following bar chart shows the sentiment score and polarity of tweets for AAP. In that bar chart indicate nearly about 24% of Positive tweets, 28% of Negative tweets, 31% of Neutral tweets. The neutral intensity of the tweets indicates some of the users were neutral about the AAP they are neither positive nor negative.

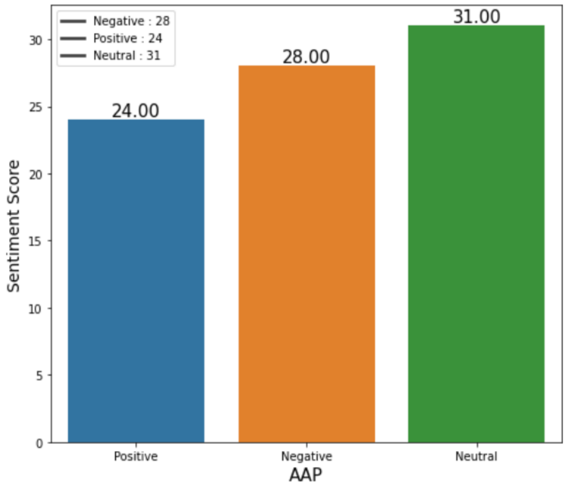


Figure: Sentiment Score for AAP

Table 7 : Results using Naïve Bayes for AAP

	Precision	Recall	F1-Score	Support
positive	0.47	0.72	0.57	3
negative	0.50	0.42	0.46	5
Accuracy			0.72	100
Weighted avg	0.45	0.63	0.50	5

Table 8 : Results using SVM for AAP

	Precision	Recall	F1-Score	Support
positive	0.43	0.72	0.54	6
negative	0.49	0.55	0.52	5
Accuracy			0.70	100
Weighted avg	0.46	0.66	0.53	6

Table 9 : Results using Decision Tree for AAP

	Precision	Recall	F1-Score	Support
positive	0.48	0.41	0.44	6
negative	0.32	0.67	0.42	3
Accuracy			0.51	100
Weighted avg	0.41	0.52	0.40	4

- **COMPARATIVE STUDY OF DATA MINING CLASSIFIER :**

We are using Naïve Bayes Classifier, Support Vector Machine, and Decision Tree classifier to calculate the result accuracy of Tweets regarding BJP, NCP, AAP.
We got following results:
Using Naïve Bayes Classifier for the tweets for BJP accuracy is 87%, for NCP accuracy is 81%,for AAP accuracy is 72%.

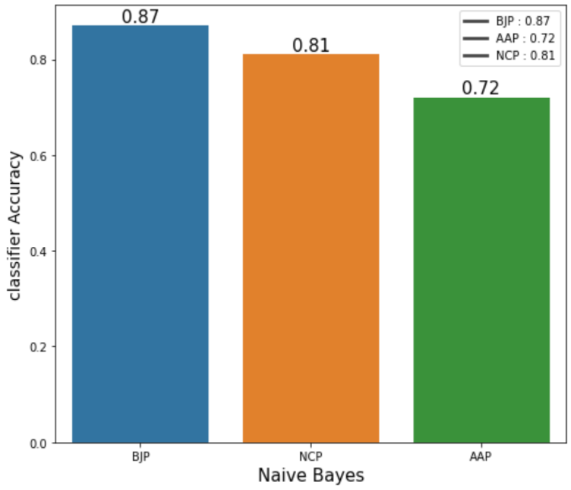


Figure : Naïve Bayes classifier result for BJP, NCP, and AAP

Using Support Vector Machine Classifier for the tweets for BJP accuracy is 79%, for NCP accuracy is 47%,for AAP accuracy is 61%.

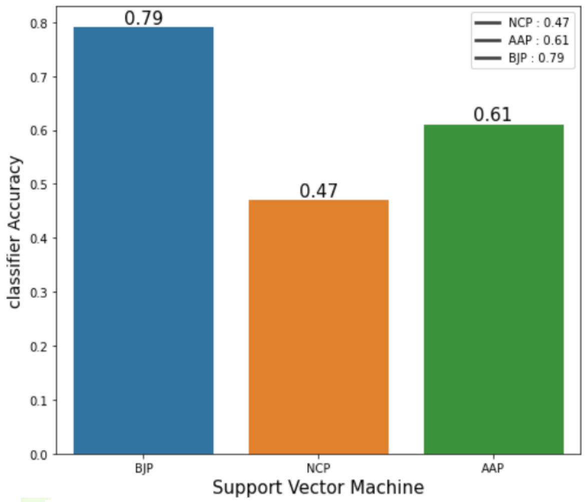


Figure: SVM classifier result for BJP, NCP, and AAP

Using Decision Tree Classifier for the tweets for BJP accuracy is 64%, for NCP accuracy is 65%, for AAP accuracy is 51%.

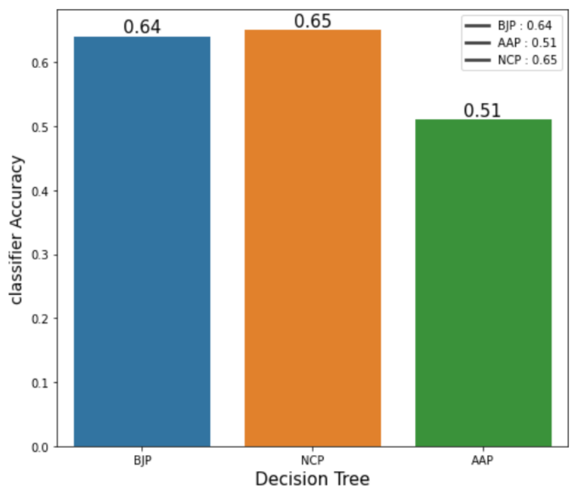


Figure: Decision Tree classifier result for BJP, NCP, and AAP

V. RESULTS AND DISCUSSION

On the basis of the results of our experiment we got the highest accuracy for the Naïve Bayes classifier, Using Naïve Bayes Classifier for the tweets for BJP accuracy is 87%, for NCP accuracy is 81%,for AAP accuracy is 72%.
Using Support Vector Machine Classifier for the tweets for BJP accuracy is 79%, for NCP accuracy is 47%,for AAP accuracy is 61%.
Using Decision Tree Classifier for the tweets for BJP accuracy is 64%, for NCP accuracy is 65%, for AAP accuracy is 51%.
Also we got highest positive sentiment for BJP is 35% , for NCP is 26% , and for AAP is 24%.
So, according to this experiment Naïve Bayes Classifier is better classifier to got accurate result.

VI. CONCLUSION

In our project, we have mainly focused on 2 approaches The first approach involved is the highest positive sentiment score for the party participating in the loksabha election 2019. The second approach is using a data mining algorithm we calculate sentiment score and accuracy from our tweet dataset of three parties and also show comparative about classifiers for which classifier is the better classifier.

Also, among the three classifiers i.e Naïve Bayes classifier, SVM classifier, and Decision Tree classifier, the Naïve Bayes classifier proves to generate a better result than the others.

According to our experiment, we got a total number of positive tweets for BJP to be 36%, congress to be 27%, and AAP to be 25%. As we are calculating the winning in the Lok Sabha election and the major parties contesting are BJP and congress we can clearly see that BJP is winning over the people's hearts and according to that BJP wins the Lok Sabha election.

On the basis of these classifiers, BJP has the highest accuracy in all three classifiers. Hence according to that BJP won the Lok Sabha election.

VII. LIMITATIONS

When we consider the total population we are not considering the people eligible to vote. There are many people whose views might change during the election. There are states with so less population that their vote affecting is negligible in this method. Because the long retweets couldn't be fully recovered, they were represented with "...", which the computer interprets as neutral sentiment. Due to a misuse of semantics, sarcasm was not detected in several statements. The Twitter search API could only obtain data from the last seven days. In other circumstances, using hashtags as a shorthand depiction of the party produced unclear outcomes. It's impossible to attain 100 percent accuracy while analysing tweets.

REFERENCES

[1] Hillygus, D. S. (2011). The evolution of election polling in the United States. Public opinion quarterly, 75(5),, 962-981.

[2] Lewis Beck, M. S. (2005). Election forecasting: principles and practice. The British Journal of Politics & International Relations,7(2), 145-164.

[3] Fumagalli, L. &. (2011). The total survey error paradigm and pre-election polls: The case of the 2006 Italian general elections. ISER Working Paper Series. 2011-29.

[4] "Number of monthly active Twitter users worldwide from 1st quarter 2010 to 1st quarter 2017 (in millions)" Link: https://www.statista.com/statistics/282087/number-of- monthly-active-twitterusers/ [Accessed 2/2/2018].

[5] Pratik Jaiswal, Mukta Dhopeshwarkar, Mangesh Patil, Anupriya Kamble, Gajanand Boywar, Ramesh R. Manza, and Surekha B. Jaiswal, "Identification of Educationally Backward Countries in Primary, Secondary and Tertiary Level Students by Using Different Classification Techniques", Springer Nature Singapore Pte Ltd. 2021 V. S. Rathore et al. (eds.), Rising Threats in Expert Applications and Solutions, Advances in Intelligent Systems and Computing 1187, https://doi.org/10.1007/978-981-15-6014-9_91

[6] Dann, S. (2010). Twitter content classification. First Monday, 15(12).

[7] Bahrainian S.-A., Dengel A., "Sentiment Analysis and Summarization of Twitter Data", 16th IEEE International Conference on Computational Science and Engineering, pp. 227-234, Sydney, Australia, December 2013.

iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 9 **Issue:** VII **Month of publication:** July 2021

DOI: https://doi.org/10.22214/ijraset.2021.36874

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue VII July 2021- Available at www.ijraset.com

Digit Recognition from EEG Signals on Smart Devices a Novel Approach

Subodh Kumar Jha[1], Chandrakant Patil[2], Sushil Pandharinath Bedre[3], Pratik Jaiswal[4], Pravin Yannawar[5], Mukta Dhopeshwarkar[6]
[1, 3, 5]*Vision and Intelligent system laboratory, Dr. Babasaheb Ambedkar Univarsity, Aurangabad, Maharashtra - 431004 India.*
[2, 4, 6]*Department of Computer Science & I.T, Dr. Babasaheb Ambedkar Univarsity, Aurangabad, Maharashtra - 431004 India.*

Abstract: The Advancement of communication system has given us the freedom to think beyond traditional communication system and stage is set for thought oriented communication system. There are thousands of thoughts generated and vanished in a timeframe but out of these some prominent thoughts persist and we proceed with the same in our day to day activities. The advancement in Electroencephalogram has provided a chance to see the activity in the human brain in non-invasive manner. The proposed research work presents the method for Digit recognition using the EEG signals acquired and processed on smart devices. The results show the implementation of Computation neural network for the recognition of digits from EEG signals. It was seen that, the 90.64% correct classification was achieved.
Keywords: EEG, FAST Independent Component Analysis, Computational Neural Network

I. INTRODUCTION

An electroencephalography (EEG) based Brain Computer Interface (BCI) enables people to communicate with the outside world by interpreting the EEG signals of their brains to interact with devices such as wheelchairs and intelligent robots described by Zhang et.al. 2018[1]. Recent advancement in Electroencephalography (EEG) allows user to measure recording of the electrical signals (*voltage potential developed across the human scalp due to iconic transfer between the neurons of brain cells during its activity*) by using arrays of sensor placed across the across the scalp of subject. These potentials are recorded from different regions of the scalp for which each region of scalp has got its own importance and it depicts the neuron activity activated on different locations of the brain. Although there is plentiful research have been made in past several years and dedicated towards exploiting EEG technology in the fields of neuro and cognitive science. The researchers have explored the possibility of controlling electronic devices based on thought signals based on EEG measurements and gives rise to the potential area of Brain Computer Interface.

The Brain Computer Interface is upcoming potential area of research. The researchers working in this domain are trying to make system more robust and scalable. The major challenges faced by the researchers are like *computational inaccuracies*, *delays*, *false positive detections*, *inter people variances*, *high costs*, and *constraints on invasive technologies* that needs further research in this area. The primary research that utilizes EEG technology is based on the fact that this rhythmic activity is dependent upon mental state and can be influenced by level of alertness or various mental diseases. One of the most common cause of *artifacts* is *eye movement and blinking*, however other causes can include the use of *scalp*, *neck*, or other muscles or even poor contact between the scalp and the electrodes by Millett, D. 2001[2], Rampil, I. J.,1998[3] and Fetz, E. E, 1969[4]. Kennedy, P et.al 1997[5] has describe in his work in Electroencephalography (EEG) is the process of picking up the electrical activity from the cortex. Hans Berger et.al, suggested that periodic fluctuations of the EEG might be related in humans to cognitive processes. Carey, B. 2008 used the electrical activity recorded of the scalp with surface electrodes constitute a non-invasive approach to gathering EEG data, while semi-invasive or invasive approaches implant electrodes under the skull or on the brain, respectively [6].

Hochberg, L. R et.al., 2006 worked on the recent technological advances were helpful for the researchers to design of some of the state of art prototype system based on thoughts [7]. The group at Carnegie Mellon University and the University of Pittsburgh allowed a monkey to feed itself via a prosthetic arm using only its thoughts. The work described by Gotman, J. 1982 was extremely promising for the disabled, and indeed by 2006 a system was developed for a *tetraplegia* that enabled subject to use prosthetic devices such as a mouse cursor, and a television via a 96-micro-electrode array implanted into primary motor cortex of subject [8].

Similarly, many automated EEG signal classification and seizure detection systems were also in-place and designed by using different approaches.

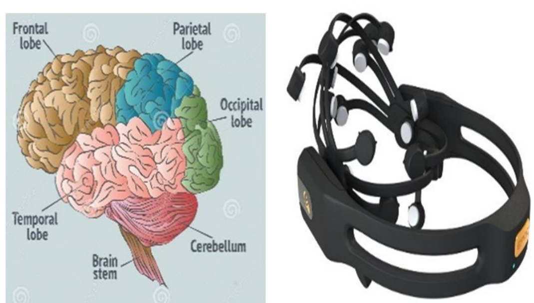
In-line with reported studies, Gotman et.al 1982 presented a computerized system for detecting a variety of seizures, while Qu and Gotman 1997[9], proposed the use of the nearest-neighbor classifier on EEG features extracted in both time and frequency domains to detect the onset of epileptic seizures. Gigola et al. 2004 applied a method based on the evolution of accumulated energy using wavelet analysis for the prediction of epileptic seizure onset from intracranial epileptic EEG recordings [10], while Adeli et al. 2007[11] and Ubeyli et al. 2006[12] & 2010 discussed the potential of nonlinear time series analysis in seizure detection[13].

Artificial neural network-based detection systems for diagnosis of epilepsy have been proposed by several researchers Tzallas, A. T., 2007[14] and Ghosh-Dastidar, S., 2008[15]. The method put forward by Weng and Khorasani 1996[16] uses the features proposed by Gotman and Wang 1991[17], namely, average EEG amplitude, average EEG duration, variation coefficient, dominant frequency and average power spectrum, as inputs to an adaptive structured neural network[18]. The method proposed by Pradhan et al. 1996 exploits raw EEG signal input to a learning vector quantization network. Nigam and Graupe 2004 proposed a new neural network model called LAMSTAR (Large Memory Storage and Retrieval) network and two time-domain attributes of EEG; namely, r*elative spike amplitude* and *spike rhythmicity* have been used as inputs for the purpose of detecting seizures [19].

In-line with the method presented by the researchers in the literature is particularly towards study of EEG Signals in the context to neuroscience and cognitive science. This paper presents an entirely new method of processing EEG signals on to the Smart Devices such as smart phones, tablets, notebook based on the preprocessing, features extraction methods and classifiers. The content of this paper is organized in following section. *Section I* has provided the background, *Section II* addresses database specification, *Section III* presents detailed methodology adopted and performance analysis of the method, *Section IV* presents conclusion of the work followed acknowledgement and references.

II. DATABASE

In order to design robust EEG recognition system, the researchers have designed the database as per requirement that meets to their research problem in general. The EEG signals, for each mode, was captured by EmoEngine as shown in figure 1



(a) (b)

Figure 1. (a) Brain lobes and (b) Emotive EPOC device for brain wave data acquisition

Data is read from the headset and sent to an output file for later analysis. The subject is asked to wear the Emotive head set which sends the data about the activity performed by the subject to the remote smart device through the available communication mechanism. The data is stored on the smart device and further be used for training and testing of the samples over Smart Devices. Traditionally, the data received from the subject is seen for five broad spectral sub-bands of the EEG signal which are generally of clinical interest they are *delta* (0 - 4 Hz), *theta* (4 - 8 Hz), *alpha* (8 - 16 Hz), *beta* (16 - 32 Hz) and *gamma waves* (32 - 64 Hz). These five frequency sub-bands provide more accurate information about neuronal activities underlying the problem and, consequently, some changes in the EEG signal, which are not so obvious in the original full-spectrum signal, can be amplified when each sub band is considered independently. Each EEG segment was considered as a separate EEG signal resulting in a total of 125 EEG data segments.

The brain is formed using five lobes, these lobes perform all the critical neurological activities such as *frontal lobe* controls the activity of Speech, Thought, Emotions, Problem solving and skilled movements, *Parietal lobe* identifies and interprets sensations such as touch, pain etc. *Occipital lobe* collects and interprets visual images that is sight, *Temporal lobe* controls the activities related to hearing and storing memory and *Cerebellum* controls the coordinate's familiar movements. Similarly, the relationship between brain lobes that the excreted (energy) frequency of signal are as below

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue VII July 2021- Available at www.ijraset.com

Type	Frequency Range	Origin
Delta	0Hz – 4Hz	Cortex
Theta	4Hz – 8Hz	Parietal and Temporal
Alpha	8Hz – 13Hz	Occipital
Beta	13Hz – 20Hz	Parietal and Frontal
Gamma	20Hz – 40Hz	Parietal and Frontal

Table 1. Signal Type, Frequency and its origin

The data set designed in this research work is basically developed on two modalities that is KEYWORD and DIGITS. The dataset of DIGITS have been utilized for all experiments. This set contains digits from (0-9) and EEG Signal recordings of 10 subjects (i.e. 7 Male and 3 Female) in the age group of (20-25) were taken. The cumulative size of the database is 10x10x10 = 1000 samples. All participating subjects in the process of data collection/acquisition were normal and away from any physical and mental disorder. The data acquisition set up was developed at Vision and Intelligent System Laboratory of Department of Computer Science and Information Technology, Dr. Babasaheb Ambedkar Marathwada University, Aurangabad. All subjects were made to sit comfortably on an arm chair facing the screen in electromagnetically shielded room. The subjects had given their written consent for recording EEG signals before participating. All subject has good knowledge of digits. All subjects were instructed that this experiment has been designed to be used for Brain Computer Interface applications. A simple display system in power point is prepared for the data collection under proposed research work. This system generates digit and keyword signal with interval of 2 sec. After every 2 sec a next number is displayed on the screen. Demonstration of display system was shown to each subject before experiment start so that he was more familiar to the task and we will get proper signals. This process was repeated for five times. So the total volume of digit dataset is 1000 samples and keyword is 1000 samples. After extracting the EEG Signal from all subject it will be processed as per methodology

III. METHODOLOGY

To implement the above digit recognition system on smart devices, the critical aspect of consideration is the accuracy of the EEG based thought recognition algorithm. This paper presents a method of acquiring, preprocessing, feature extraction, normalization and classification model from raw EEG signal. In this section, we provides an overview of proposed approach and ANTARANG framework for interpretation of EEG data.

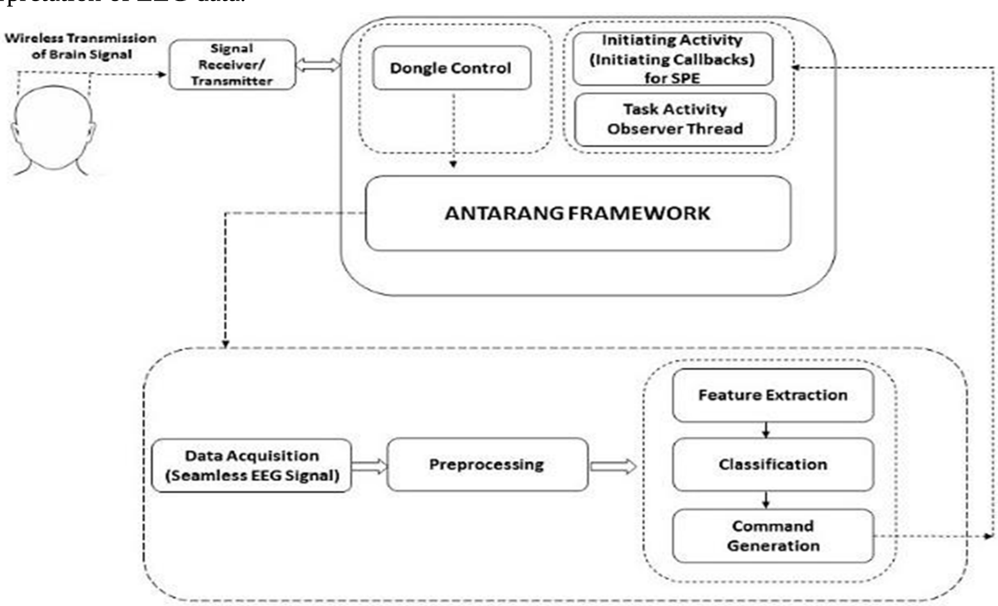


Figure 2 (a)

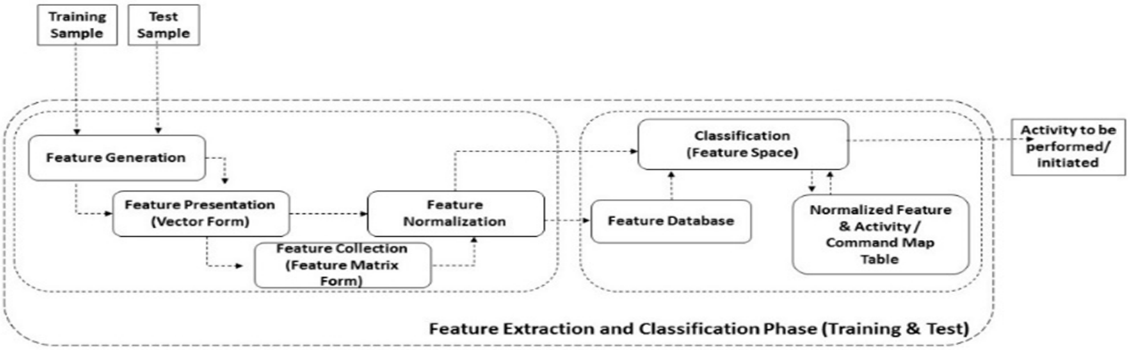


Figure 2 (b)

Figure 2: (a) and (b) Block Diagram of thought processing system 'ANTARANG'

A. Overview

The above Figure 2 (a) and 2(b) illustrate the organization of various steps involved in the recognition of digits. The functionalities of the systems are as discussed below,

1) *Data Acquisition*: the method of data acquisition is as discussed in section II is followed. The subject is required to place EMOTIVE EPOC head gear in order to acquire EEG signals corresponding to the activity. The data is acquired by EMOTIVE head set is transferred to the wireless dongle connected to the device seamlessly via Bluetooth mechanism. The dongle control mechanism connected to the smart device acts as receiver. This will store the test sample primarily in the storage of Smart device and handed over to ANTARANG framework.
2) *Preprocessing*: the captured test sample was preprocessed using, FAST *Independent component analysis* (ICA) was performed on EEG sample data to removing artifact and resulting ICs were passed for feature extraction.
3) *Feature Extraction*: the objective of this phase is to generate unique set of features such that the overall performance of classification is improved. In this research work stack of feature extraction methods were used which contains methods like *Short Time Fourier Transform (STFT), Discrete Cosine Transform (DCT) and discrete wavelet transform (DWT)* were utilized towards computation of features.
4) *Feature Normalization*: the computed features are normalized. This is required to reduce the size of feature space and speedup the classification of the system. The Linear Discriminant analysis were utilized towards reducing the feature space. This is feature normalization is performed with all training vectors as well the test sample is also normalized before classification.
5) *Classification:* the classification phase has immense potential in the design of any automated system. The proposed system is developed with the stack of classifiers such as *Support vector machine, k-Nearest Neighbor, Random forest, Naïve Bias classifier [20], Multi-Layer perceptron, and Convolution neural network*. The result of classifier will be handed over to the native command translation mechanism which initiate the activity in the smart processing elements (Smart Devices).
6) *Command Map Table and Task observer thread:* The command map table contains information about the mapped callback corresponding the thought. The task observer thread observers the activity and invoke/dispatch the task for execution on the smart devices.
7) *Tools and Software:* As part of this work, the preprocessing and freature extraction was implemented in the SciPy and Numpy library of Python language. Convolutional neural network models were designed using the Keras library and run using Tensorflow in an attempt to classify the time-frequency representations. The matplotlib library was used to create plots the figures and data visualization.

B. Working

The subject is required to gear with Emotive EEG set at the time of data acquisition as well as during testing samples. The electrode or subset of electrodes in an EEG device may move during data acquisition this may leads into bad contact with the scalp and therefore a poor quality signal may be received. More rarely, electrodes may also have mechanical faults, for example frayed wiring, which can partially or completely degrade the signal received. Such electrodes can produce artifacts into the signals. So in a preprocessing step, FAST Independent component analysis (ICA) was performed on EEG sample data to removing artifact and resulting ICs were pass for feature extraction. Fundamentally ICA in biomedicine involves the extraction and separation of statistically independent sources underlying multiple measurements of biomedical signals.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue VII July 2021- Available at www.ijraset.com

1) *Feature Extraction using DCT:* The Discrete Cosine Transform is a transformation method for converting a time series signal into basic frequency components. Low frequency components are concentrated in first coefficients and high frequency in last ones. The one-dimensional DCT for a list of N real numbers is expressed by eq (1) as,

$$Y(u) = \sqrt{\frac{2}{N}}\alpha(u)\sum_{x=0}^{N-1} f(x)cos\left(\frac{\pi(2x+1)}{2N}\right) \quad (1)$$

Where u=0, 1, 2, 3… N-1;

$$\alpha(0) = \frac{1}{\sqrt{2}}$$

a(j) = 1, j≠0;

An acquired input EEG sample from training set is a set of ‘N’ data values and the output is a set of N-DCT transform coefficients Y(u). The first coefficient Y(0) is called the DC coefficient and holds average signal value. The rest coefficients are referred to as the AC coefficients. DCT exhibits good energy compaction for highly correlated signals. If the input data consists of correlated quantities, then most of the N transform coefficients produced by the DCT are zeros or small numbers, and only a few are large. Compressing data with the DCT is therefore accomplished by quantizing the coefficients. The small ones are quantized coarsely and the large ones can be quantized finely to the nearest integer. Applying this feature for EEG signals allow compressing useful data to the first few coefficients. Therefore, only these coefficients can be used for classification using machine learning algorithms. This kind of data compression may dramatically reduce input vector size and decrease time required for training and classification. These feature were calculated for all the samples of ‘Number set’ and ‘Keyword set’. The ‘DCT Feature Matrix’ for the samples of ‘Number set’, are as shown in table 2.

Zero (0)	One (1)	Two (2)	Three (3)	Four (4)	Five (5)	..	Nine (9)
10.62964	12.48839	10.48546	9.819233	11.02994	16.9704	..	27.57632
5.413632	7.304988	6.416848	5.132164	5.326457	8.545044	..	5.128143
1.423464	2.693033	3.74308	2.304048	0.652103	-1.00238	..	2.486984
0.543664	0.927368	3.214933	0.981311	-0.04452	-0.52716	..	-1.96868
-0.16563	-1.0361	1.963108	-0.1841	-0.70961	-0.04245	..	4.004859
-0.50003	-1.7836	1.128882	-0.54477	-1.36299	1.116709	..	-0.70907
0.288363	-1.68334	0.503528	-0.85261	-1.13287	0.399226	..	3.714574
1.684606	-0.99388	-0.03555	-0.89706	-1.03169	-0.58246	..	-8.12342
2.571268	1.571514	-0.46821	0.533503	-0.52996	1.070756	..	-2.49936
..	..	..	..	..	..	..	..
0.921427	2.309691	-0.5832	0.605643	1.048911	1.025788	..	4.177428

2) *Feature Selection Using LDA:* After signal analysis as well as feature extraction using DCT, the feature vector, $Y = [y_1, y_2, y_3, \ldots, y_n]$ is derived. Its dimension should be reduced since the dimension *n* is often too large and the design of classifiers for a large dimension suffers from various difficulties. Those are mostly numerical problems that involve operation with high-order matrices. At the same time, a classifier in *n* -dimensional space is very difficult to analyze and almost impossible to imagine. Thus Linear Discernment Analysis (LDA) was applied on feature vector to deduce the feature and selecting most prominent features for classification. The aim of LDA is to use hyper planes to separate the data representing the different classes proposed by Duda, R. O.,et,at. 2012[21]. The separating hyper plane is obtained by seeking the projection that maximize the distance between the two classes means and minimize the inter classes variance by Fukunaga, K. 2013[22]. To solve an N-class problem (N > 2) several hyper planes are used. This technique has a very low computational requirement which makes it suitable for BCI system. So all the sample of ‘Digit database’ and ‘Keyword dataset’ normalized using LDA and selected 100 features of each sample for classification.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue VII July 2021- Available at www.ijraset.com

IV. RESULTS AND DISCUSSION

The recognition of EEG Signal sample was carried out by DCT and LDA. These features were calculated for all sample of training set and stored for recognition purpose. The entire preprocessed features data set of EEG Digits were divided into 70-30 ratio that is 70% (Training samples), 30% (Test samples) and evaluated using Convolution Neural Network (CNN). This artificial neural network is improved in both parameters that is shift and translational invariance describe in Fukushima, K. 1980[23]. CNN is a subset of deep learning which has attracted a lot of attention in recent year and used in image recognition such as analysis of x-ray medical images by Kallenberg, M et. al., 2016[24], magnetic resonance images by Pereira, S et. al. 2016[25], histopathological images by Hatipoglu, N. et. Al, 2017[26] fundus images by Tan, J. H. et. al., 2017[27], and computed tomography images describe by Setio, A. et. al., 2016[28]. But, very little research has been done on the use of CNN using physiological signals. The CNN architecture consists of three different types of layer i.e. convolutional layer, pooling layer, and a fully connected layer. CNNs are very effective models for Image classification tasks [29].

For the proposed work CNN model was designed, where EEG digit dataset data first Convolution layer takes this 1-dimensional array as input and the Convolution operation uses 10 initial convolution filters and a convolutional kernel of size 11. Where, the first convolution layer uses 'relu' as the Activation ('Relu' or Rectilinear units as Activation for Arousal model). In these proposed experiments, the choice of activation functions for this first layer are of cardinal importance, as some functions like *sigmoid or softmax* might not be able to activate neurons of later layers consistently this improper activation function contributed towards making model defective. The next layer is another Convolutional Neural Layer which again with 100 filters and 3*3 size kernel. This layer uses '*relu*' as the Activation function for both Valence and Arousal classification. Thus, dropout on the outputs of *MaxPooling* layer, with a dropout probability of 0. 5, to form a flat 1 dimensional layer. The final dense layer uses '*softmax*' as its activation function. The model uses the Categorical Cross Entropy as the loss function and '*rmsprop*' as the optimizer used. The experiment were carried out upto 500 epochs and train the model using batches of 32 experiments each. Following is confusion matrix of CNN classification of EEG digit signal.

Digit	Total Test Sample	Training Samples										Correct Classified	Miss-classified	Accuracy
		0	1	2	3	4	5	6	7	8	9			
0	15	13	0	1	0	1	0	0	0	0	0	13	02	86.66
1	15	0	13	0	1	1	0	0	0	0	0	13	02	86.66
2	11	0	0	9	0	1	0	1	0	0	0	9	02	81.81
3	14	0	0	0	14	0	0	0	0	0	0	14	00	100.00
4	15	0	0	0	0	13	0	2	0	0	0	13	02	86.66
5	16	0	0	1	0	0	13	2	0	0	0	13	02	81.25
6	22	1	0	0	0	0	1	20	0	0	0	20	02	90.90
7	12	0	0	0	0	0	0	0	12	0	0	12	00	100.00
8	7	0	0	0	0	0	0	0	1	6	0	6	01	85.71
9	13	0	0	0	0	0	0	0	0	0	13	13	00	100.00
	139	Classification Result										126	13	90.64%

Table 3: Confusion Matrix for Digit Classification using CNN

The confusion matrix as shown in Table 3, the total 139 test samples of tested on the training data set. It was see that out of 15 samples of 'zero', 13 were classified correctly and 2 samples were misclassified so that the found in class of two and four. Similarly, all test samples of 'three', 'seven' and 'nine' were completely classified in 'three', 'seven' and 'nine' classes correctly, there were no misclassification seen in this.

Out of these all 139 test samples, the 16 test samples of number 'five', 13 were classified correctly and 03 were misclassified into 'two' and 'six'. The average classification resulting into 90.64% classification accuracy that is out of 139 samples, 126 samples were classified into correct classes and only 13 samples were misclassified.

V. CONCLUSION

The proposed research work presents the system for automatic classification of EEG signal of digits for smart devices. The proposed work evaluates the performance of CNN classifier evaluated over normalized features of Discrete Cosine Transform. The work also signifies method of feature minimization using linear discriminant analysis. The overall accuracy was observed to be 90.64% and the work will be also extended towards automatic classification of 'keywords'. The proposed work also is extended towards design of EEG operated smart devices.

VI. ACKNOWLEDGMENT

This piece of work is registered under Indian **Patent application number** 201621005217 **Titled** SYSTEM FOR THOUGHT BASED COMMUNICATION **Year -** 2016.

The Authors would like to thanks Dept. of Computer Science and IT, Dr. Babasaheb Ambedkar Marathwada University, Aurangabad for providing necessary infrastructure.

REFERENCES

[1] Zhang, X., Yao, L., Sheng, Q. Z., Kanhere, S. S., Gu, T., & Zhang, D. (2018, March). Converting your thoughts to texts: Enabling brain typing via deep feature learning of eeg signals. In 2018 IEEE International Conference on Pervasive Computing and Communications (PerCom) (pp. 1-10). IEEE.
[2] Millett, D. (2001). Hans Berger: From psychic energy to the EEG. Perspectives in biology and medicine, 44(4), 522-542.
[3] Rampil, I. J. (1998). A primer for EEG signal processing in anesthesia. Anesthesiology: The Journal of the American Society of Anesthesiologists, 89(4), 980-1002.
[4] Fetz, E. E. (1969). Operant conditioning of cortical unit activity. Science, 163(3870), 955-958.
[5] Kennedy, P. R., & Bakay, R. A. (1997). Activity of single action potentials in monkey motor cortex during long-term task learning. Brain research, 760(1-2), 251-254.
[6] Carey, B. (2008). Monkeys think, moving artificial arm as own. New York Times.
[7] Hochberg, L. R., Serruya, M. D., Friehs, G. M., Mukand, J. A., Saleh, M., Caplan, A. H., & Donoghue, J. P. (2006). Neuronal ensemble control of prosthetic devices by a human with tetraplegia. Nature, 442(7099), 164.
[8] Gotman, J. (1982). Automatic recognition of epileptic seizures in the EEG. Electroencephalography and clinical Neurophysiology, 54(5), 530-540.
[9] Qu, H., & Gotman, J. (1997). A patient-specific algorithm for the detection of seizure onset in long-term EEG monitoring: possible use as a warning device. IEEE transactions on biomedical engineering, 44(2), 115-122.
[10] Gigola, S., Ortiz, F., D'attellis, C. E., Silva, W., & Kochen, S. (2004). Prediction of epileptic seizures using accumulated energy in a multiresolution framework. Journal of neuroscience methods, 138(1-2), 107-111.
[11] Adeli, H., Ghosh-Dastidar, S., & Dadmehr, N. (2007). A wavelet-chaos methodology for analysis of EEGs and EEG subbands to detect seizure and epilepsy. IEEE Transactions on Biomedical Engineering, 54(2), 205-211.
[12] Ubeyli, E. D. (2006). Analysis of EEG signals using Lyapunov exponents. Neural Network World, 16(3), 257.
[13] Ubeyli, E. D. (2010). Recurrent neural networks employing Lyapunov exponents for analysis of ECG signals. Expert systems with applications, 37(2), 1192-1199.
[14] Tzallas, A. T., Tsipouras, M. G., & Fotiadis, D. I. (2007). Automatic seizure detection based on time-frequency analysis and artificial neural networks. Computational Intelligence and Neuroscience, 2007.
[15] Ghosh-Dastidar, S., Adeli, H., & Dadmehr, N. (2008). Principal component analysis-enhanced cosine radial basis function neural network for robust epilepsy and seizure detection. IEEE Transactions on Biomedical Engineering, 55(2), 512-518.
[16] Weng, W., & Khorasani, K. (1996). An adaptive structure neural networks with application to EEG automatic seizure detection. Neural Networks, 9(7), 1223-1240.
[17] Gotman, J., & Wang, L. Y. (1991). State-dependent spike detection: concepts and preliminary results. Electroencephalography and clinical Neurophysiology, 79(1), 11-19.
[18] Pradhan, N., Sadasivan, P. K., & Arunodaya, G. R. (1996). Detection of seizure activity in EEG by an artificial neural network: A preliminary study. Computers and Biomedical Research, 29(4), 303-313.
[19] Nigam, V. P., & Graupe, D. (2004). A neural-network-based detection of epilepsy. Neurological research, 26(1), 55-60.
[20] P. Jaiswal, M. Dhopeshwarkar, M. Patil, A. Kamble, G. Boywar, R. R. Manza. "Identification of Educationally Backward Countries in Primary, Secondary and Tertiary Level Students by Using Different Classification Techniques." Department of Computer Science and Information Technology, Dr. Babasaheb, Ambedkar Marathwad University, Aurangabad, Maharashtra, India. © Springer Nature Singapore Pte Ltd. (2021) Rising Threats in Expert Applications and Solutions, Advances in Intelligent Systems and Computing 1187, https://doi.org/10.1007/978-981-15-6014-9_91 , (757-763)
[21] Duda, R. O., Hart, P. E., & Stork, D. G. (2012). Pattern classification. John Wiley & Sons.
[22] Fukunaga, K. (2013). Introduction to statistical pattern recognition. Elsevier.
[23] Fukushima, K. (1980). Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. Biological cybernetics, 36(4), 193-202.

[24] Kallenberg, M., Petersen, K., Nielsen, M., Ng, A. Y., Diao, P., Igel, C. & Lillholm, M. (2016). Unsupervised deep learning applied to breast density segmentation and mammographic risk scoring. IEEE transactions on medical imaging, 35(5), 1322-1331.

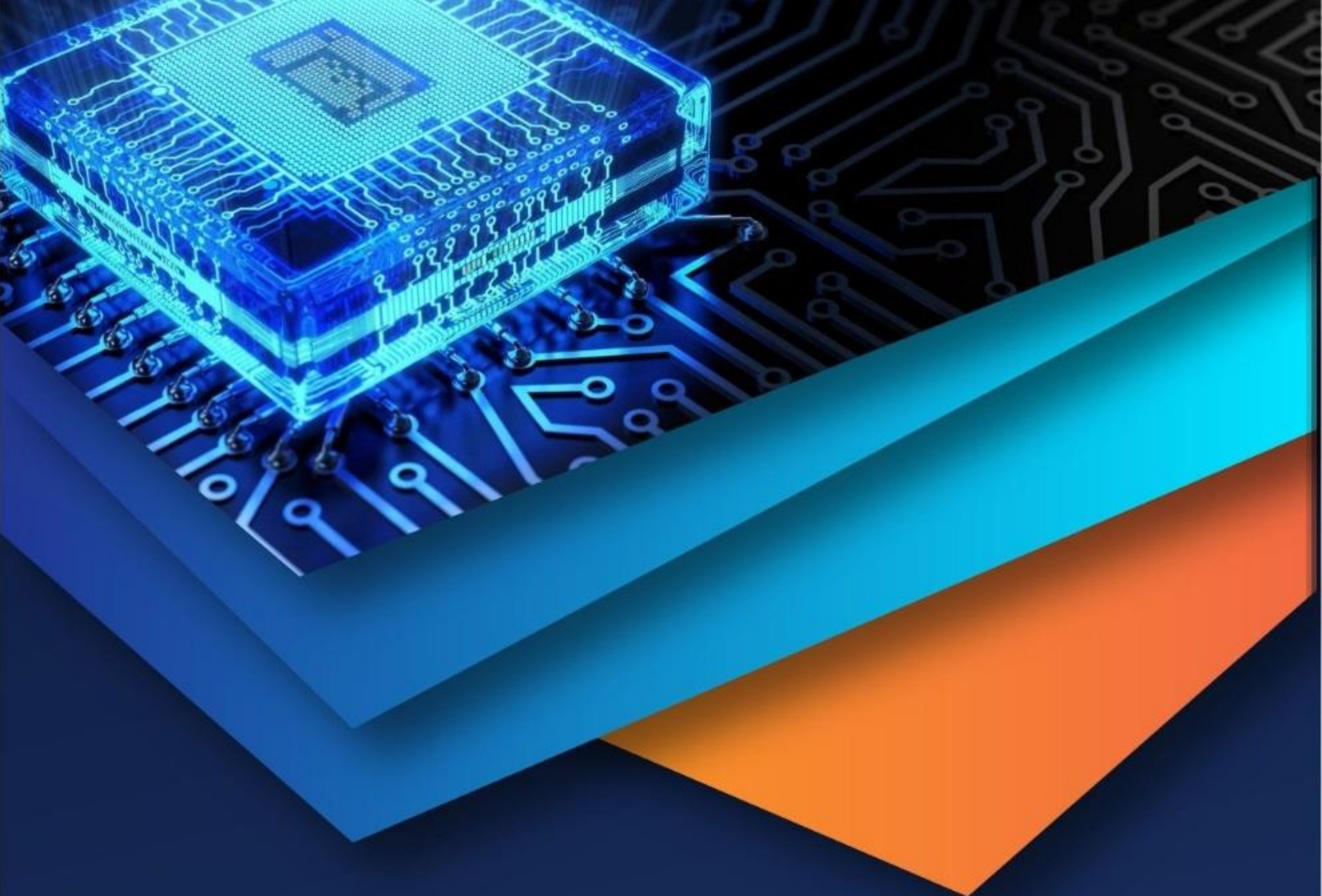
[25] Pereira, S., Pinto, A., Alves, V., & Silva, C. A. (2016). Brain tumor segmentation using convolutional neural networks in MRI images. IEEE transactions on medical imaging, 35(5), 1240-1251.

[26] Hatipoglu, N., & Bilgin, G. (2017). Cell segmentation in histopathological images with deep learning algorithms by utilizing spatial relationships. Medical & biological engineering & computing, 55(10), 1829-1848.

[27] Tan, J. H., Acharya, U. R., Bhandary, S. V., Chua, K. C., & Sivaprasad, S. (2017). Segmentation of optic disc, fovea and retinal vasculature using a single convolutional neural network. Journal of Computational Science, 20, 70-79.

[28] Setio, A. A. A., Ciompi, F., Litjens, G., Gerke, P., Jacobs, C., Van Riel, S. J. & van Ginneken, B. (2016). Pulmonary nodule detection in CT images: false positive reduction using multi-view convolutional networks. IEEE transactions on medical imaging, 35(5), 1160-1169.

[29] Mouad MH Ali, Pravin L Yannawar, AT Gaikwad "Multi-algorithm of palmprint recognition system based on fusion of local binary pattern and two-dimensional locality preserving projection" 2017/1/1 Procedia computer science Volume115 Publisher Elsevier.



IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661,p-ISSN: 2278-8727, Volume 24, Issue 1, Ser. I (Jan. –Feb. 2022), PP 50-56
www.iosrjournals.org

Emotion Recognition using Smart Devices

[1]Chandrakant Patil, [1]Mukta Dhopeshwarkar, [1]Pratik Sunil Jaiswal, [2]Sushil Pandharinath Bedre, [2]Subodh Kumar Jha
[1]Department of Computer Science & I.T, Dr. Babasaheb Ambedkar Univarsity, Aurangabad, Maharashtra - 431004 India.
[2]Vision and Intelligent system laboratory, Dr. Babasaheb Ambedkar Univarsity, Aurangabad, Maharashtra - 431004 India.

Abstract: *The progress of communication systems has allowed us to think beyond traditional communication systems, and the scene has been set for thought-oriented communication systems. Thousands of thoughts are formed and then evaporate in a short period of time, yet certain notable concepts remain and we carry on with our daily routines. EEG has advanced to the point that it is now possible to see the activity in the human brain in a non-invasive manner. The approach for emotion identification utilizing EEG data recorded and processed on smart devices is presented in this study. The results demonstrate the use of a computational neural network to recognize emotions from EEG data. It was discovered that the correct categorization rate was 90.17 percent.*

Keywords: *Emotions, EEG, FAST Independent Component Analysis, Computational Neural Network*

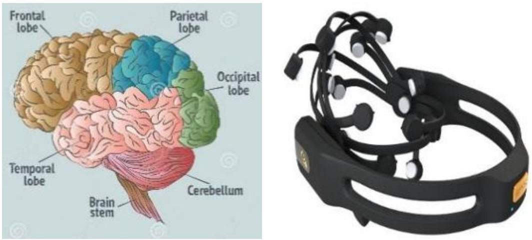
Date of Submission: 06-01-2022 Date of Acceptance: 18-01-2022

I. Introduction

We are approaching the next communication assessment, which will see a significant change in communication channels and mechanisms. When it comes to human-machine communication, we've come a long way from punch cards to keyboards and mice to touch screens and gesture recognition to BCI devices. We can now analyze enormous real-time data in various complicated algorithms to derive essential information thanks to developments in computer and network capability. A brain-computer interface (BCI) is a type of neuro-technology that decodes a user's central nervous system signals [1]. The BCI enables direct thought–based communication with other users or operation of different appliances (e.g., a direct brain–robot interface) without efferent peripheral nervous system fibers or muscles being involved [2]. In the case of locked-in-syndrome (LIS) users who cannot focus or regulate their eye movements, BCIs offer feasible alternatives, and in certain circumstances they are the more appropriate communication augmentation solutions [3]. EEG has a lot of untapped potential since it has a lot of untapped information that is directly related to real-time brain activity. This data is being processed using advanced approaches, resulting in novel applications for EEG interpreted data. Controlling smart devices with brain signals, performance assessment of persons using EEG analysis on generated cognitive work load, control of prostate organs using EEG, and so on are only a few instances [4, 5]. BCIs (Brain Computer Interfaces) are an emerging prospective study field. Researchers in this field are attempting to make the system more resilient and scalable. The researchers encounter considerable hurdles such as computational errors, delays, false positive detections, inter-person variations, high prices, and restrictions on intrusive technologies, all of which necessitate more study in this field. The fundamental study that makes use of EEG technology is based on the concept that this rhythmic activity is impacted by mental state and can be altered by alertness or other mental illnesses. Eye movement and blinking are one of the most prominent sources of artefacts, but additional causes include the usage of scalp, neck, or other muscles, or even insufficient contact between the scalp and the electrodes [6]. In a Human-in-the-Loop Cyber-Real-Systems scenario, users' intentions are inferred from brain and body sensor networks linked to them and sent directly into sensors and actuators in the physical environment, allowing for autonomous system adaption to their demands. Humans must be instrumented and integrated into the system in this scenario. This is a future that is still a long way off and far from ideal, because the desire of people and the general public to engage is still a huge issue, and it is not addressed by [7]. Similarly, several automated EEG signal categorization and seizure detection systems existed and were built using various methods. Gotman et al. [8] presented a computerized system for detecting a variety of seizures, while Qu and Gotman [9] proposed using the nearest-neighbor classifier on EEG features extracted in both the time and frequency domains to detect the onset of epileptic seizures, which is in line with previous research. While Adeli et al. [11], Guler et al. [12], and Ubeyli et al. [13] discussed the potential of nonlinear time series analysis in seizure detection, Gigola et al. [10] used a method based on the evolution of accumulated energy using wavelet analysis for the prediction of epileptic seizure onset from intracranial epileptic EEG recordings. Several studies have suggested artificial neural network-based detection methods for epilepsy diagnosis [14-15]. Weng and Khorasani [16] present an adaptable structured neural network that incorporates the characteristics suggested by Gotman and Wang [17], namely average EEG amplitude, average EEG length, variation coefficient, dominant frequency, and average power spectrum, as inputs. Pradhan et al. [18] offer a technique that uses raw EEG data input to a learning vector quantization network. Nigam and Graupe [19] introduced a novel neural network model called LAMSTAR (Large Memory Storage and Retrieval), which uses two time-domain EEG properties, namely relative spike amplitude and spike rhythmicity, as inputs to identify seizures. In line with the methodologies described in the literature, the study of EEG Signals in the framework of cognitive science. Based on data extraction, preprocessing, feature extraction algorithms, and classifiers, this literature proposes a completely new approach of processing EEG signals utilizing Smart Computing/Communication Devices such as smart phones, tablets, and notebooks. Using the aforementioned approaches, the team conducted research to recognize emotions from real-time EEG data, which may then be utilized to operate smart devices or to offer as input to external systems, which can then use it to carry out user activities in the context of the apps. The following sections organize the material of this work. The backdrop is presented in Section I, followed by database specifications in Section II, the methodology employed and performance analysis of the technique in Section III, and the work's conclusion in Section IV, which is followed by acknowledgements and references.

II. Database:

In order to create a reliable EEG recognition system, the researchers created a database that fits the requirements of their study topic in general. EmoEngine collected the EEG signals for each mode, as seen in figure 1.



(a) (b)

Figure 1. (a) Brain lobes and (b) Emotive EPOC device for brain wave data acquisition

The data from the headset is read and saved to an output file for further processing. The subject is required to wear an Emotive head set, which transmits data about the subject's activities to a distant smart device via the available communication method. The data is saved on mobile phones and may then be utilized for sample training and testing through mobile phones. The data acquired from the subject is traditionally analyzed for five wide spectral sub-bands of the EEG signal that are often of therapeutic interest: delta (0 - 4 Hz), theta (4 - 8 Hz), alpha (8 - 16 Hz), beta (16 - 32 Hz), and gamma waves (32- 64 Hz). These five frequency sub-bands give more precise information about the underlying neural activity, and as a result, some changes in the EEG signal that aren't visible in the full-spectrum signal can be enhanced when each sub-band is analyzed separately. There were 125 EEG data segments since each EEG segment was treated as an unique EEG signal. The brain is divided into five lobes, each of which is responsible for a different set of neurological functions. For example, the frontal lobe is in charge of speech, thought, emotions, problem solving, and skillful movements. The parietal lobe is responsible for identifying and interpreting experiences such as touch, discomfort, and so on. The Occipital lobe receives and analyses visual images, while the Temporal lobe is in charge of hearing and memory storage, and the Cerebellum is in charge of the coordinates' familiar motions. Similarly, the association between the output (energy) frequency of signal and the brain lobes is as follows:

Table 1. Signal Type, Frequency and its origin

Type	Frequency Range	Origin
Delta	0Hz – 4Hz	Cortex
Theta	4Hz – 8Hz	Parietal and Temporal
Alpha	8Hz – 13Hz	Occipital
Beta	13Hz – 20Hz	Parietal and Frontal
Gamma	20Hz – 40Hz	Parietal and Frontal

The data set used in this study is essentially Emotions that have been analyzed. For all of the studies, the Emotions dataset was used. This collection comprises seven emotions: 'Happy', 'Excited', 'Content', 'Calm', 'Angry', 'Afraid', 'Sad' and EEG Signal recordings of ten participants (seven males and three females) in the age range of 20-25. The database has a total size of 12x10x10 = 1200 samples. All of the participants in the data collection/acquisition procedure were healthy and free of any physical or mental disorders. In an electromagnetically insulated room, all patients were told to sit comfortably in an arm chair facing the screen. Before taking part in the study, the participants had provided their written agreement to have their EEG signals recorded. Every subject has an excellent understanding of emotions. All of the participants were told that this experiment was created for use in Brain Computer Interface applications. For the data gathering under the proposed research project, a basic power point display system has been built. This device provides a 2-second interval emotion signal. A new emotion was flashed on the screen every 2 seconds. Before the experiment began, each participant was provided a demonstration of the display system so that they would be more comfortable with the job and we would receive correct signals. This procedure was carried out five times. As a result, the total number of samples in the emotion dataset is 1000. The EEG signals from all individuals will be extracted and analyzed according to the technique.

III. Methodology / ANTARANG framework

We created the ANTARANG framework, which is utilized to design a mobile emotion identification system. This framework or approach of obtaining, preprocessing, feature extraction, normalization, and classification model from raw EEG data is presented in this research. We give an overview of the suggested technique and the ANTARANG framework for EEG data interpretation.

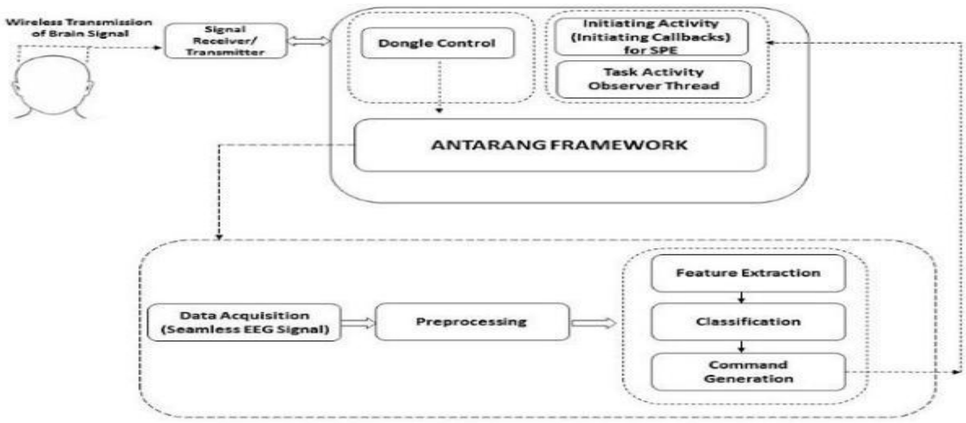


Figure 2 (a)

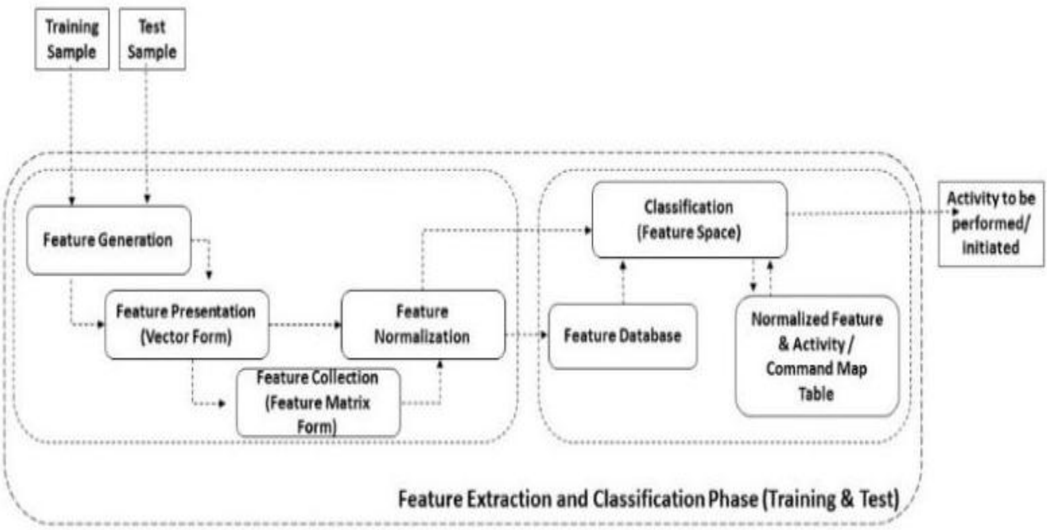


Figure 2 (b)

Figure 2: (a) and (b) Block Diagram of thought processing system 'ANTARANG'

a. Overview

Figures 2(a) and 2(b) above demonstrate how the several phases involved in emotion recognition are ordered. The systems' functions are described in the following sections,

1. **Data acquisition:** The data collection approach is the same as that described in section II. In order to get EEG signals relating to the activity, the individual must wear EMOTIVE EPOC head gear. The data collected by the EMOTIVE head set is easily sent to the wireless dongle attached to the smartphone via Bluetooth. The smart device's dongle control mechanism serves as a receiver. This will save the test sample primarily in the Smart device's storage and pass it on to the ANTARANG framework.
2. Preprocessing: FAST Independent component analysis (ICA) of EEG sample data was done to remove artefact, and the resultant ICs were passed for feature extraction.
3. Feature Extraction: The goal of this step is to develop a unique collection of characteristics that will increase classification performance overall. In this study, a stack of feature extraction methods was employed to compute features, including Short Time Fourier Transform (STFT), Discrete Cosine Transform (DCT), and discrete wavelet transform (DWT).
4. Feature Normalization: The characteristics that have been calculated were normalized. This is necessary to decrease the dimensionality of the feature space and speed up the system's categorization. The feature space was reduced by the use of Linear Discriminant Analysis. This feature normalization is done on all training vectors, as well as the test sample, before to classification.
5. Classification: In the design of any automated system, the categorization step has enormous potential. Support vector machine, k-Nearest Neighbor, Random forest, Nave Bias classifier, Multi-Layer perceptron, and Convolution neural network are among the classifiers used in the suggested system. The classifier's output will be sent on to the native command translation mechanism, which will start the smart processing elements working (Smart Devices).
6. Command Map Table and Task observer thread: The mapped callback corresponding to the idea is stored in the command map table. The task watcher thread watches for activity and then invokes/dispatches the task for smart device execution.
7. Tools and Software: Preprocessing and feature extraction were implemented in Python's SciPy and Numpy libraries as part of this project. In order to categories the time-frequency representations, convolutional neural network models were created with the Keras toolkit and executed with Tensorflow. To plot the figures and visualize the data, the matplotlib software was utilized.

b. Working

The participant must wear an Emotive EEG set during data collection as well as during the testing samples. During data gathering, an EEG device's electrode or subset of electrodes may shift, resulting in poor contact with the scalp and, as a result, a low-quality signal. Electrodes can also have mechanical problems, such as frayed wire, which can weaken the signal received partially or totally. Artifacts in the signals can be caused by such electrodes. FAST Independent component analysis (ICA) was conducted on EEG sample data as a preprocessing step to remove artefact, and the resultant ICs were passed for feature extraction. In biomedicine, ICA entails the extraction and separation of statistically independent sources underlying various biological signal measurements.

1. Feature Extraction using DCT

The Discrete Cosine Transform is a technique for transforming a time series signal into its fundamental frequency components. Low frequency components are concentrated in the first coefficients, whereas high frequency components are concentrated in the final. Equation (1) expresses the one-dimensional DCT for a list of N real values as,

$$Y(u) = \sqrt{\frac{2}{N}}\alpha(u)\sum_{x=0}^{N-1} f(x)cos\left(\frac{\pi(2x+1)}{2N}\right)(1)$$

Where u=0, 1, 2, 3… N-1;

$$\alpha(0) = \frac{1}{\sqrt{2}}$$

a(j) = 1, j≠0;

The output of an acquired input EEG sample from the training set is a set of N-DCT transform coefficients Y, while the input is a set of 'N' data values (u). The first coefficient, Y(0), is known as the DC coefficient and is responsible for storing the average signal value. The AC coefficients stand for the rest coefficients [20]. For strongly correlated data, DCT shows good energy compaction. If the input data is correlated, the majority of the N transform coefficients produced by the DCT are zeros or tiny values, with only a few exceptions. As a result, quantizing the coefficients is used to compress data with the DCT. The little ones are coarsely quantized, whereas the large ones can be finely quantized to the closest integer. When applied to EEG signals, this characteristic allows meaningful data to be compressed to the first few coefficients. As a result, machine learning systems can only employ these coefficients for categorization. This type of data compression can significantly reduce the size of the input vector and the amount of time required for training and classification. These characteristics were estimated for all of the 'Emotion set' samples. Table 2 shows the 'DCT Feature Matrix' for the samples from the 'Emotion set.'

Happy	Excited	Content	Calm	Angry	Afraid	Sad
0.366515	1.081741	1.355358	4.177025	8.047682	1.687529	2.732452
0.639972	0.688207	1.752748	9.701809	3.125768	21.37542	31.93106
0.061715	0.542834	0.970531	2.899618	2.956171	7.894226	6.532051
0.031866	1.167931	0.613204	4.922474	6.982062	21.8817	1.435193
0.567666	0.526758	0.790951	7.802225	2.726657	22.01105	51.35523
1.073398	0.276145	0.029724	0.07606	0.027696	0.03968	0.680563
3.476802	2.356264	0.774597	0.754562	0.40443	0.505899	35.93576
0.194747	0.033649	0.033631	0.017214	0.009243	0.041233	0.107949
0.056582	0.016408	0.020714	0.024856	0.006497	0.016135	0.082541
0.402641	0.091293	0.008991	0.04552	0.04046	0.051385	13.32203

Table 2: Emotion

2. *Feature Selection using LDA*

The feature vector Y = [y1, y2, y3,....., yn] is produced after signal analysis and feature extraction using DCT. Its size should be lowered since the dimension n is frequently too huge, and designing classifiers for such a large dimension is challenging. The majority of these problems are numerical in nature and require the use of high-order matrices. At the same time, analyzing and imagining a classifier in n-dimensional space is quite challenging. As a result, Linear Discernment To determine the feature and pick the most significant characteristics for classification, LDA was used to the feature vector. The goal of LDA is to segregate data representing distinct classes using hyper planes [21]. The separation hyper plane is found by looking for a projection that maximizes the distance between the means of the two classes while minimizing the variance between them [22]. Several hyper planes are employed to solve an N-class issue (N > 2). This approach has a low processing need, making it ideal for use in a BCI system. As a result, all of the samples in the 'Emotion dataset' were normalized using LDA, and 100 features from each sample were chosen for classification.

IV. Results and Discussion

DCT and LDA were used to recognize the EEG signal samples. For the purpose of recognition, these characteristics were computed for each sample of the training set and saved. The full preprocessed features data set of EEG Emotions was split into a 70-30 ratio, with 70% (Training samples) and 30% (Test samples), then assessed using a Convolution Neural Network (CNN). The shift and translational invariance of this artificial neural network have been enhanced [23]. CNN is a subset of deep learning that has gained a lot of attention in recent years and is used in image recognition applications such as x-ray medical image analysis [24], magnetic resonance image analysis [25], histopathological image analysis, fundus image analysis, and computed tomography image analysis. However, there has been relatively little study on the application of CNN with physiological inputs. Convolutional layer, pooling layer [26], and fully connected layer are the three types of layers that make up the CNN architecture. For image classification applications, CNNs are particularly effective models.

Table 3: Confusion Matrix for emotion Classification using CNN

Emotions	Total Test Sample	Training Samples							Correct Classified	Miss-classified	Accuracy
		Happy	Excited	Content	Calm	Angry	Afraid	Sad			
Happy	**17**	15	0	0	0	0	1	0	**15**	**2**	**86.67**
Excited	**13**	0	13	0	0	0	0	0	**13**	**0**	**100.00**
Content	**13**	0	0	13	0	0	0	0	**13**	**0**	**100.00**

Calm	**10**	0	1	0	9	0	0	0	**9**	**1**	**88.89**
Angry	**12**	1	0	0	0	10	0	0	**10**	**2**	**80.00**
Afraid	**14**	0	0	0	0	0	12	0	**12**	**2**	**83.34**
Sad	**14**	0	0	1	0	0	0	13	**13**	**1**	**92.31**
	93	**Classification Result**							**85**	**8**	**90.17**

Table 3 shows the confusion matrix, with a total of 93 test samples being tested on the training data set. It was discovered that 15 of the 17 samples of the emotion 'Happy' were correctly categorized and two were incorrectly categorized, whereas 13 of the 13 test samples of the emotion 'Excited' were correctly classified and 0 were incorrectly labelled. Similarly, all 'Content', 'Calm', 'Angry', 'Afraid', and 'Sad' test samples were completely categorized. Only 8 test samples were misclassified out of a total of 93 test samples, resulting in a classification accuracy of 90.17 percent.

V. Conclusion

The method for automated categorization of EEG signal of Emotions for smart devices is presented in the suggested research effort. The proposed study assesses the effectiveness of a CNN classifier using normalized Discrete Cosine Transform information. The work also denotes a feature reduction strategy based on linear discriminant analysis. The total accuracy was found to be 90.17 percent, and the research will now be expanded to include automated categorization of 'emotions.' The suggested study is further extended to the construction of smart gadgets that are controlled by EEG.

References

[1]. Wolpaw, Jonathan, and Elizabeth Winter Wolpaw, eds. Brain-computer interfaces: principles and practice. OUP USA, 2012.

[2]. Rutkowski, Tomasz M., and Hiromu Mori. "Tactile and bone-conduction auditory brain computer interface for vision and hearing impaired users." Journal of Neuroscience Methods 244 (2015): 45-51.

[3]. H. Mori, Y. Matsumoto, V. Kryssanov, E. Cooper, H. Ogawa, S. Makino, Z. R. Struzik, and T.M. Rutkowski, "Multi-command tactile brain computer interface: A feasibility study," in Haptic and Audio Interaction Design. Springer, 2013, pp. 50–59.

[4]. H. Mori, Y. Matsumoto, S. Makino, V. Kryssanov, and T. M. Rutkowski, "Vibrotactile stimulus frequency optimization for the haptic BCI prototype," in Proceedings of The 6th International Conference on Soft Computing and Intelligent Systems, and The 13th International Symposium on Advanced Intelligent Systems, Kobe, Japan, November 20-24, 2012, pp. 2150- 2153.

[5]. Lebedev MA, Nicolelis MAL (2006) Brain-machine interfaces:past, present and future. TRENDS Neurosci 29(9):536–546.

[6]. Schirner, Gunar, Deniz Erdogmus, Kaushik Chowdhury, and Taskin Padir. "The future of human-in-the-loop cyber-physical systems." *Computer* 1 (2013): 36-45.

[7]. Gotman, J. "Automatic recognition of epileptic seizures in the EEG." Electroencephalography and clinical Neurophysiology 54, no. 5 (1982): 530-540.

[8]. Qu, Hao, and Jean Gotman. "A patient-specific algorithm for the detection of seizure onset in long-term EEG monitoring: possible use as a warning device." IEEE transactions on biomedical engineering 44, no. 2 (1997): 115-122.

[9]. Gigola, S., F. Ortiz, C. E. D'attellis, W. Silva, and S. Kochen. "Prediction of epileptic seizures using accumulated energy in a multiresolution framework." Journal of neuroscience methods 138, no. 1-2 (2004): 107-111.

[10]. Adeli, Hojjat, Samanwoy Ghosh-Dastidar, and Nahid Dadmehr. "A wavelet-chaos methodology for analysis of EEGs and EEG subbands to detect seizure and epilepsy." IEEE Transactions on Biomedical Engineering 54, no. 2 (2007): 205-211.

[11]. Übeyli, Elif Derya. "Recurrent neural networks employing Lyapunov exponents for analysis of ECG signals." Expert Systems with Applications 37, no. 2 (2010): 1192-1199.

[12]. Übeyli, Elif Derya. "Analysis of EEG signals using Lyapunov exponents." Neural Network World 16, no. 3 (2006): 257.

[13]. Tzallas, Alexandros T., Markos G. Tsipouras, and Dimitrios I. Fotiadis. "Automatic seizure detection based on time-frequency analysis and artificial neural networks." Computational Intelligence and Neuroscience 2007 (2007).

[14]. Ghosh-Dastidar, Samanwoy, Hojjat Adeli, and Nahid Dadmehr. "Principal component analysis-enhanced cosine radial basis function neural network for robust epilepsy and seizure detection." IEEE Transactions on Biomedical Engineering 55, no. 2 (2008): 512-518.

[15]. Weng, W. D., and Khashayar Khorasani. "An adaptive structure neural networks with application to EEG automatic seizure detection." Neural Networks 9, no. 7 (1996): 1223-1240.

[16]. Gotman, J., and L. Y. Wang. "State-dependent spike detection: concepts and preliminary results." Electroencephalography and clinical Neurophysiology 79, no. 1 (1991): 11-19.

[17]. Pradhan, N., P. K. Sadasivan, and G. R. Arunodaya. "Detection of seizure activity in EEG by an artificial neural network: A preliminary study." Computers and Biomedical Research 29, no. 4 (1996): 303-313.

[18]. Nigam, Vivek Prakash, and Daniel Graupe. "A neural-network-based detection of epilepsy." Neurological Research 26, no. 1 (2004): 55-60.

[19]. R. O. Duda, P. E. Hart, and D. G. Stork. Pattern Recognition, second edition. WILEYINTERSCIENCE, 2001.

[20]. K. Fukunaga, Statistical Pattern Recognition, second edition. ACADEMIC PRESS, INC, 1990.

[21]. Fukushima, K., 1980, Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position, Biological Cybernetics 36:193-202.

[22]. Kallenberg. M., Petersen. K., Nielsen. M., Ng. Y. A., Diao. P. F., Igel. C., Vachon. C. M., Holland. K., Winkel. R. R., Karssemeijer. N., Lillholm. M., 2016. Unsupervised deep learning applied to breast density segmentation and mammographic risk scoring, IEEE Transactions on Medical Imaging, 35(5):1322-1331.

[23]. Pereira. S., Pinto. A., Alves. V., Silva. C. A., 2016. Brain tumor segmentation using convolutional neural networks in MRI images, IEEE Transactions on Medical Imaging, 35(5):1240-1251.

[24]. Hatipoglu. N., Bilgin. G., 2017. Cell segmentation in histopathological images with deep learning algorithms by utilizing spatial relationships, Medical and Biological Engineering and Computing, 1-20, doi: 10.1007/s11517-017-1630-1.
[25]. Tan. J. H., Acharya. U. R., Bhandary. S. V., Chua. K. C., Sivaprasad. S., 2017. Segmentation of optic disc, fovea, and retinal vasculature using a single convolutional neural network, Journal of Computational Science, DOI: 10.1016/j.jocs.2017.02.006.
[26]. Setio. A. A. A., Ciompi. F., Litjens. G., Gerke. P., Jacobs. C., van Riel. S. J., Wille. M. M. W., Naqibullah. M., Sánchex. C. I., van Ginneken. B., 2016. Pulmonary nodule detection in CT images: False positive reduction using multi-view convolutional networks, IEEE Transactions on Medical Imaging, 35(5):1160-1169.

Chandrakant Patil, et. al. “Emotion Recognition using Smart Devices.” *IOSR Journal of Computer Engineering (IOSR-JCE)*, 24(1), 2022, pp. 50-56.

iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL
FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 Issue: IV Month of publication: April 2022

DOI: https://doi.org/10.22214/ijraset.2022.42049

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue IV Apr 2022- Available at www.ijraset.com

Human Follower Robot

Priyanka P. Vetal[1], Dr. Mukta Dhopeshwarkar[2], Pratik S Jaiswal[3]
[1, 2, 3]*Department of Computer Science & I.T, Dr.Babasaheb Ambedkar Marathwada University, Aurangabad (M.S)*

Abstract: In this digital and automotive day and age, robotics, and IoT produce an impact on human life. One can't just rely on the traditional mode of work in this era. One has to adapt the robotics and keep digging in it, as it's the near future for humans. So to do this there are many aspects to implement automotive in day-to-day life. One such event is to study a robot that follows humans that means which can detect human movement and react as per this movement. The study shows that there are many researchers, scientists, engineers who have worked and still working to improve this human movement detection in robotics. This paper has studied some of the previous work and gave a comparative analysis of the same.
Keywords: Human Detection Robot, Sensor, Machine, Automotive

I. INTRODUCTION

Robot technology has grown tremendously in recent times. The same establishment was just a dream for some people a few times back. But in this seaside world, there is now a need for robots like “A Human Follower Robot” that can communicate and communicate with them. (1) To accomplish this, a robot needs the ability to dream and to perform. (2) (3) The robot must be intelligent enough to follow a person in tight spaces, in an image, and inside or out (4). Photo processing done to get information about nature by appearance is really important. The following points should be considered carefully in practice. Living conditions should be truly stable and should not change. The width should be well placed in the requested area when blurring. The target should not be too far away from the visible detector as distance is very important. We should avoid using the same color next to the target robot. Otherwise, the robot would be confused. Usually the next dead robots are equipped with several different combinations of icons i.e. light detection and various icons. All detectors and modules operate in accordance with the definition and target tracking.

The robot's ability to track and trace a moving object can be used for a number of purposes.

1) Helping people.
2) Generating people easily.
3) Can be used for self-defense purposes.

In this paper, we have introduced the dying robot system based on label identification and detection using a camera. Intelligent target recording is done using a variety of tools and modules (5) namely ultrasonic detector, magnetometer, infrared detector, and camera. The wise decision is made by the robots' control unit based on the information found in the icons and modules below, which is why it changes and tracks something by avoiding obstacles and without contradicting the target.

II. RELATED WORK

Another experimental work was done in this regard, in-depth photography was used by Calisi and the target was developed by designing a special algorithm (1). Ess and Leibe did the same job. They do a lot of work on sewing and acquisition. The great advantage of their system is that their algorithm worked in a complex environment and (2) (3). The stereo view is also made by Y. Salih to make discovery (4). This system enabled him to pursue the desired goal effectively. A combination of different detectors was used by R. Munoz ku. get information about the target to be tracked. In addition to using various equipment, he also used the stereo view to obtain accurate information. (5). The protective data combined with the information from the camera proved to be very useful in performing the task (6). Different algorithms are developed by testers for detection purposes. The beam was used ina single test to determine the style of moving legs (7) (8) (9) and the camera was used to describe an object or person (10) (11). Really simple fashion was also used for experimentation. In this way, one is accustomed to finding the distance to the robot and the person. These detectors detect radio swelling and are detected by detectors in a person to be tracked. In this way the robot followed the target (12).

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue IV Apr 2022- Available at www.ijraset.com

III. BLOCK DIAGRAM

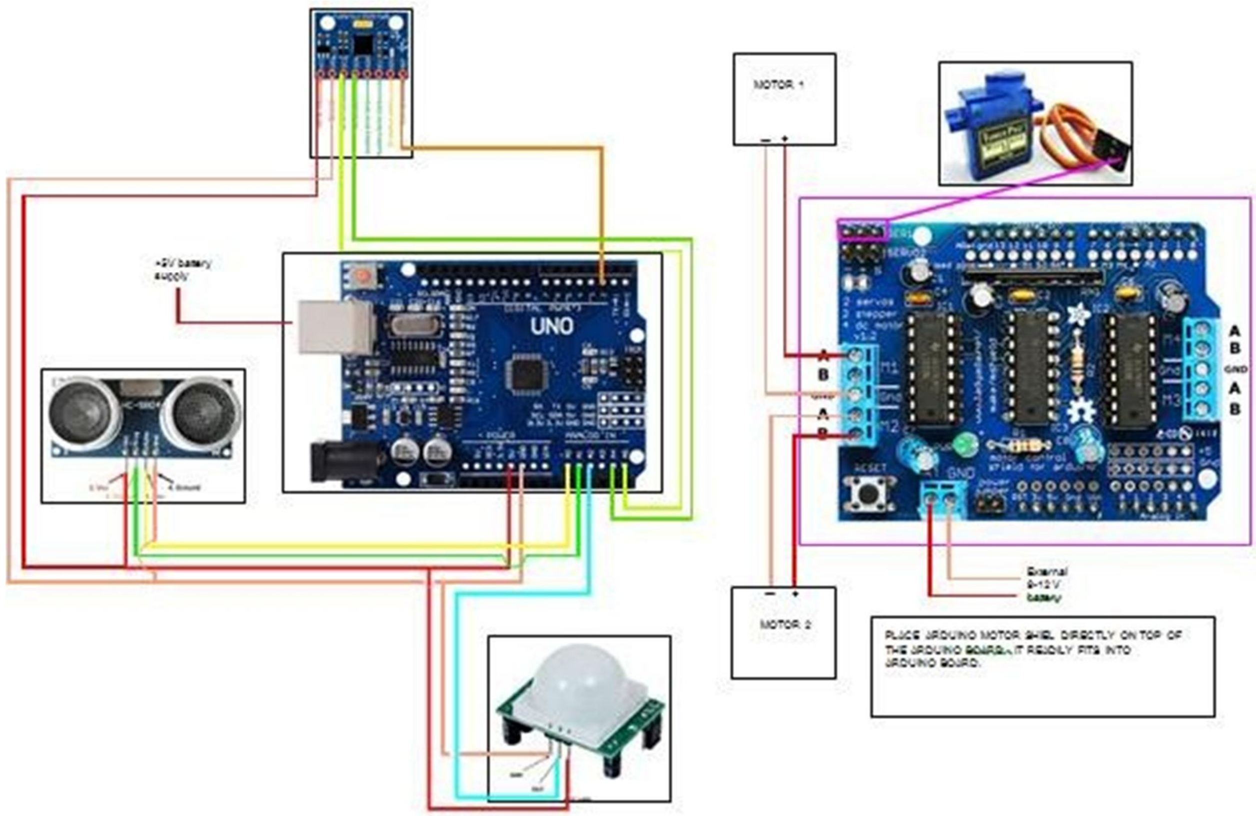


Fig.1: Block diagram

IV. PROPOSED SYSTEM

An Arduino-based human tracking robot is used to track a person or a person with a leg movement wherethe person is moving the robot will follow him or her forward, left, right the robot will detect human movement and properly follow its movements. Using an ultrasonic sensor robot to locate a person and the ultrasonic sensor calculates the distance between the human and the robot also analyzes it. In our project the ultrasonic sensor is mounted on a servo motor which is a direct actuator that allows precise control of angular or linear position.

The Servo motor has a shaft that goes up to 180 degrees .in the servo motor we have an ultrasonic sensor that also moves the servo motor to get the correct position and direction of the moving person.

We have also used the gyro sensor also known as the angular rate sensor which is the angular velocity sensor .the gyro sensor is used to detect human movement angles and detects moving body angles at 30, 45., 60, 90, 180 degree respectively. Analyzing and deciding and sense the person moving slowly left, right, forward take the step of following a particular person where he or she is moving in the right direction with his or her moving direction. The main function of the gyro sensor is an angle sensor and control mechanism.

Many of the earliest constructions of flower flowers are not used by the PIR sensor .we have used the PIR sensor in our project which is used to detect the difference between living and non-living things. The PIR is an infrared sensor used to detect when a person has entered or exited a range of sensors.

A. Component Used

1) *Arduino Uno:* It is the brain of our design. It can give all the instructions on their low items to be used by the mortal geste. It also provides feedback on other aspects and people. So that it can be used as a means of communication between humans and robots and vice versa. It has 8 bit CPU specification, 16 MHZ clock speed. Speed, 2 KB SRAM 32 KB flash Memory, 1 KB EEPROM.

2) *DC Gear Motors:* DC Motor is a device that converts any type of energy into mechanical power or transmits chaos. In building a robot, the engine often plays an important role in providing robotic movement. After that 4 DC motors are used to drive robot.
3) *Motor Shield:* Motor Shield is a motor driver module that allows you to use Arduino to control the speed and direction of the car. The Motor Shield may be powered by Arduino directly or 6V 15V external power with terminal input. The Motor Motorist Board was then designed to work with the L293D IC.
4) *Ultrasonic Detector:* Ultrasonic detector is a tool that measures the distance to an object using an ultrasonic irradiator. The operating principle of this module is simple, it transmits ultrasonic palpitation from 40kHz through the air, and if there is a defect or object, it will jump back into the detector. By calculating travel time and sound speed, distance can be calculated.
5) *Servo Motor:* Servo motor operates in PWM (Pulse width modulation) means that its gyration angle is controlled by the length of the heartbeat applied to its Control leg. Basically a servo motor is made with a DC motor controlled by a flexible resistor (potentiometer) and other gears.
6) *PIR Sensor:* PIR detectors allow you to inhale the chaos, almost always used to determine whether a person has entered or exited a machine. They are small, affordable, low power, easy to use and never wear out. For that reason they often invest in electrical appliances and widgets used in homes or businesses.
7) *Gyro Scope Sensor:* A gyroscope detector is a device that can measure and maintain exposure and angular velocity of an object. These are much more advanced than accelerometers. These can measure the exposure of the cock and the side of the object while the accelerometer can only measure direct movement. Gyroscope detectors are also called Angular Rate Detectors or Angular Haste Detectors. These detectors are activated when the exposure of an object is soft to human. Measured by degrees per second, angular acceleration is the change in rotation angle of an object per unit time.

When the upper part of the gyro rotates 90 degrees to the side, it continues its desire to move left. The same is true for the lower section-- it rotates 90 degrees to the side and continues its dream of moving right. This force rotates thewheel where it goes.

B. Algorithm

1) *Step 1:* Initially turn the ultrasonic sensor into 30 degrees using a servo motor. (Considered 90 degree as forwardultrasonic sensor, 0 degree vertical)
2) *Step 2:* Check the presence of the object within 600mm.
3) *Step 3:* If the object is B / W 600mm to 250mm. Turn the robot over to the object. (In this case at (90-30) = 60degree using the sensor to measure the angle).
4) *Step 4:* Rotate the ultrasonic sensor at 60 degrees to 90 degrees. Standing angle pointing to an object.
5) *Step 5:* Move the robot forward to the object until the robot is 250-300mm long or there is a delay.
6) *Step 6:* If the object in step 2 is 250-200mm make the robot stop.
7) *Step 7:* If the object in step 2 is less than 200mm, move the robot back to a distance of 250mm.
8) *Step 8:* When an item is found in Step 2. Increase the servo angle by level. (In the second multiplication the servoangle is 31 degrees.)
9) *Step 9:* Repeat steps 2 to step 3 until the servo angle = 1150 degree. (Total sweep angle 150-30 = 120 degree)
10) *Step 10:* Set the servo direction from 150 to 30 degrees. Keeping step 2 to 8 in the loop. (Back lowers the servoangle by 1 degree instead of magnification.)

C. Working

Our system consists of a four-wheeled robot car equipped with a separate microprocessor and control unit as well as various equipment and modules namely ultrasonic trackers, infrared detectors that help navigate the people and objects around you. The finders below work in tandem and assist the robot in its operation and movement in its path by avoiding obstacles and keeping a certain distance from the object. We have used an ultrasonic detector to avoid crashes and to maintain a certain distance of the object. Ultrasonic detector works directly within 4 steps.

At an angle set to 0^o, the finding area between the two receivers is small enough to make the robot difficult to make a decision to define the human leg. While, the angle set to 10^o, the large finding area between the finders indicates that it is suitable for the analysis of the death leg within the area.

At that point, the angle set to 20^o, the acquisition angle is considered to be slightly larger than the 10^o angle. It may make the robot unsuitable for carrying legs when the distance between the legs and the robot ascends. Therefore, thefun can be eliminated. The angle of discovery of angle 30^o is so complex that it can be confusing to make decisions there.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue IV Apr 2022- Available at www.ijraset.com

V. RESULT

Various tests were performed and the performance of the dying robot was tested. Testing is performed on ultrasonic and infrared detectors. It was noted that the detector operated directly within the 4-dimensional range. We also did experimenting to see if the robot kept a certain distance from the target object. We also tested intermittent interactions between Arduino, motor guard, and colored motors. Based on the results obtained from these tests and tests, we have made the necessary changes to the processing and control algorithm. After it was completed, we saw thatthe results produced truly satisfying the robot was following the wrong person wherever he went.

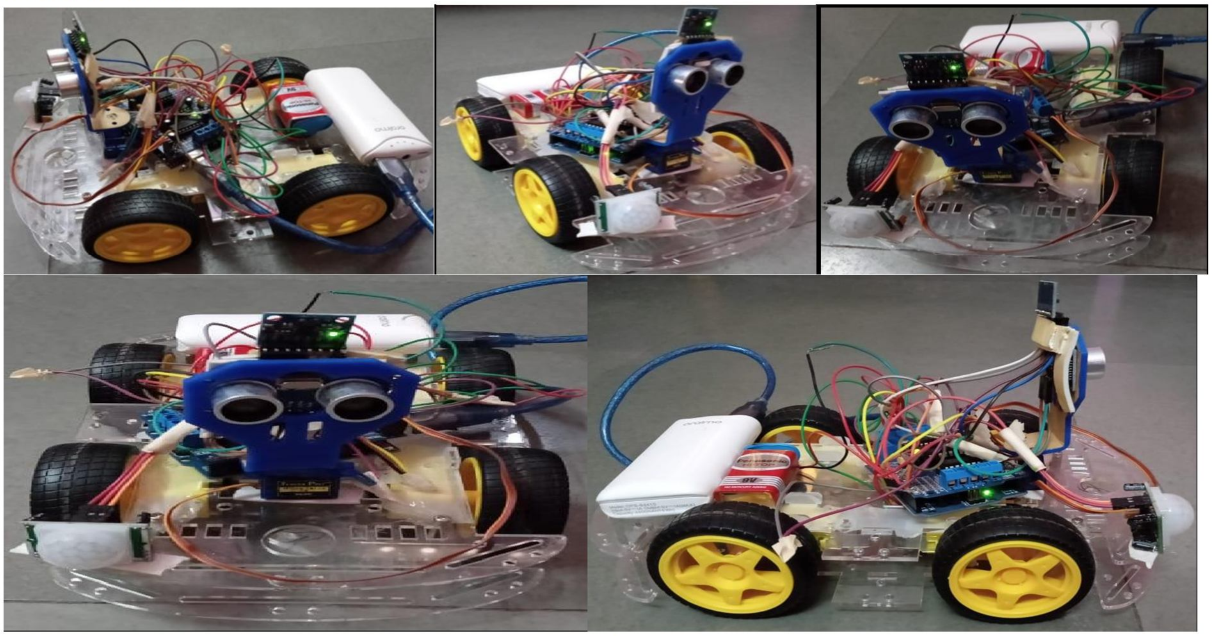


Fig.2: Project Result

Existing System Comparison Table

	Existing system	Actual System
Boat Weight	Moderate weight	Light weight
Chassis size	200m x 1x 49mm	150 x 25 x 18 mm
Motor	BO Motor	DC gear motor
Battery	9v	11v
Sensors	Ultrasonic, IR	Ultrasonic , PIR
Controller	Arduino	Arduino
Servo motor	MG90	MG90

VI. CONCLUSION

A successful implementation of a prototype of human follower robot is illustrated in this paper. This robot does not only have the detection capability but also the following ability as well. While making this prototype it was also kept in mind that the functioning of the robot should be as efficient as possible. Tests were performed on the different conditions to pin point the mistakes in the algorithm and to correct them. The different sensors that were integrated with the robot provided an additional advantage. The human following robot is an automobile system that has ability to recognize obstacle, move and change the robot's position toward the subject in the best way to remainon its track. This project uses Arduino, motors different types of sensors to achieve its goal. This project challenged the group to cooperate, communicate, and expand understanding of electronics, mechanical systems, and theirintegration with programming.

A. Application

1) In military area also in household, travel and in shopping where human follower robot can used as anautonomous cart or automated trolley which follows the person.
2) In industrial application a robot can help to carry heavy items for a long distance.

B. Future Scope

1) We can attach a camera on this robot to analyze and record the entire situation where human is going. There are numerous intriguing operations of this exploration in different fields whether service or medical.
2) Wireless communication functionality can be added to the robot to make it more protean and control it from alarge distance. This capability of a robot could also be used for military purposes. By mounting a real- time videotape archivist on top of the camera, we can cover the surroundings by just sitting in our apartments. We can also add some variations to the algorithm and the structure as well to fit it for any other purpose. Also, it can help the public in shopping promenades. So there it can act as a luggage carrier, hence no need to carryup the weights or to pull that. Also, an ample quantum of variations could be done to this prototype for far and wide operations.

REFERENCES

[1] K. Morioka, J.-H. Lee, and H. Hashimoto, "Human-following mobile robot in a distributed intelligent sensor network," IEEE Trans. Ind. Electron., vol. 51, no. 1, pp. 229–237, Feb. 2004.
[2] Y. Matsumoto and A. Zelinsky, "Real-time face tracking system for human-robot interaction," in 1999 IEEE International Conference on Systems, Man, and Cybernetics, 1999. IEEE SMC '99 Conference Proceedings, 1999, vol. 2, pp. 830– 835 vol.2.
[3] T. Yoshimi, M. Nishiyama, T. Sonoura, H. Nakamoto, S. Tokura, H. Sato, F. Ozaki, N. Matsuhira, and H. Mizoguchi, "Development of a Person Following Robot with Vision Based Target Detection," in 2006 IEEE/RSJ International Conference on Intelligent Robots and Systems, 2006, pp. 5286–5291.
[4] H. Takemura, N. Zentaro, and H. Mizoguchi, "Development of vision based person following module for mobile robots in/out door environment," in 2009 IEEE International Conference on Robotics and Biomimetics (ROBIO), 2009, pp.
[5] Muhammad Sarmad Hassan, MafazWali Khan, Ali Fahim Khan,"Design and Development of Human Following Robot", 2015,Student Research Paper Conference,Vol-2, No-15.
[6] N. Bellotto and H. Hu, "Multisensor integration for human-robot interaction," IEEE J. Intell. Cybern.Syst., vol. 1, no. 1, p. 1, 2005.
[7] K. Morioka, J.-H. Lee, and H. Hashimoto, "Human-following mobile robot in a distributed intelligent sensor network," IEEE Trans. Ind. Electron., vol. 51, no. 1, pp. 229–237, Feb. 2004.
[8] Y. Matsumoto and A. Zelinsky, "Real-time face tracking system for human-robot interaction," in 1999 IEEE International Conference on Systems, Man, and Cybernetics, 1999. IEEE SMC '99 Conference Proceedings, 1999, vol. 2, pp. 830–835 vol.2.
[9] T. Yoshimi, M. Nishiyama, T. Sonoura, H. Nakamoto, S. Tokura, H. Sato, F. Ozaki, N. Matsuhira, and H. Mizoguchi, "Development of a Person Following Robot with Vision Based Target Detection," in 2006 IEEE/RSJ International Conference on Intelligent Robots and Systems, 2006, pp. 5286–5291.
[10] H. Takemura, N. Zentaro, and H. Mizoguchi, "Development of vision based person following module for mobile robots in/out door environment," in 2009 IEEE International Conference on Robotics and Biomimetics (ROBIO), 2009, pp. 1675–1680.
[11] N. Bellotto and H. Hu, "Multisensor integration for human-robot interaction," IEEE J. Intell. Cybern. Syst., vol.1, no. 1, p. 1, 2005.
[12] W. Burgard, A. B. Cremers, D. Fox, D. Hähnel, G. Lakemeyer, D. Schulz, W. Steiner, and S. Thrun, "The interactive museum tour-guide robot," in AAAI/IAAI, 1998, pp. 11–18.
[13] M. Scheutz, J. McRaven, and G. Cserey, "Fast, reliable, adaptive, bimodal people tracking for indoor environments," in Intelligent Robots and Systems, 2004.(IROS 2004). Proceedings. 2004 IEEE/RSJ International Conference on, 2004, vol. 2, pp. 1347–1352.
[14] N. Bellotto and H. Hu, "People tracking and identification with a mobile robot," in Mechatronics andAutomation, 2007. ICMA 2007. International Conference on, 2007, pp. 3565–3570.
[15] S. Jia, L. Zhao, and X. Li, "Robustness improvement of human detecting and tracking for mobile robot," in Mechatronics and Automation (ICMA), 2012 International Conference on, 2012, pp. 1904–1909.
[16] M. Kristou, A. Ohya, and S. Yuta, "Target person identification and following based on omnidirectional cameraand LRF data fusion," in RO-MAN, 2011 IEEE, 2011, pp. 419–424.
[17] T. Wilhelm, H. J. Böhme, and H. M. Gross, "Sensor fusion for vision and sonar based people tracking on amobile service robot," in Proceedings of the International Workshop on Dynamic Perception, 2002, pp. 315–320.
[18] D. Calisi, L. Iocchi, and R. Leone, "Person following through appearance models and stereo vision using amobile robot.," in VISAPP (Workshop on on Robot Vision), 2007, pp. 46–56.
[19] A. Ess, B. Leibe, and L. Van Gool, "Depth and appearance for mobile scene analysis," in Computer Vision,2007. ICCV 2007. IEEE 11th International Conference on, 2007, pp. 1–8.
[20] A. Ess, B. Leibe, K. Schindler, and L. Van Gool, "A mobile vision system for robust multi-person tracking," in Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on, 2008, pp. 1–8.
[21] Y. Salih and A. S. Malik, "Comparison of stochastic filtering methods for 3D tracking," Pattern Recognit., vol. 44, no. 10, pp. 2711–2737, 2011.
[22] R. Muñoz-Salinas, E. Aguirre, and M. García-Silvente, "People detection and tracking using stereo vision and color," Image Vis. Comput., vol. 25, no. 6, pp. 995–1007, 2007.
[23] J. Satake and J. Miura, "Robust stereo-based person detection and tracking for a person following robot," in ICRA Workshop on People Detection and

Tracking, 2009.

[24] J. H. Lee, T. Tsubouchi, K. Yamamoto, and S. Egawa, "People Tracking Using a Robot in Motion with Laser Range Finder," in 2006 IEEE/RSJ International Conference on Intelligent Robots and Systems, 2006, pp. 2936– 2942.

[25] M. Lindstrom and J. O. Eklundh, "Detecting and tracking moving objects from a mobile platform using a laser range scanner," in 2001 IEEE/RSJ International Conference on Intelligent Robots and Systems, 2001. Proceedings, 2001, vol. 3, pp. 1364–1369 vol.3.

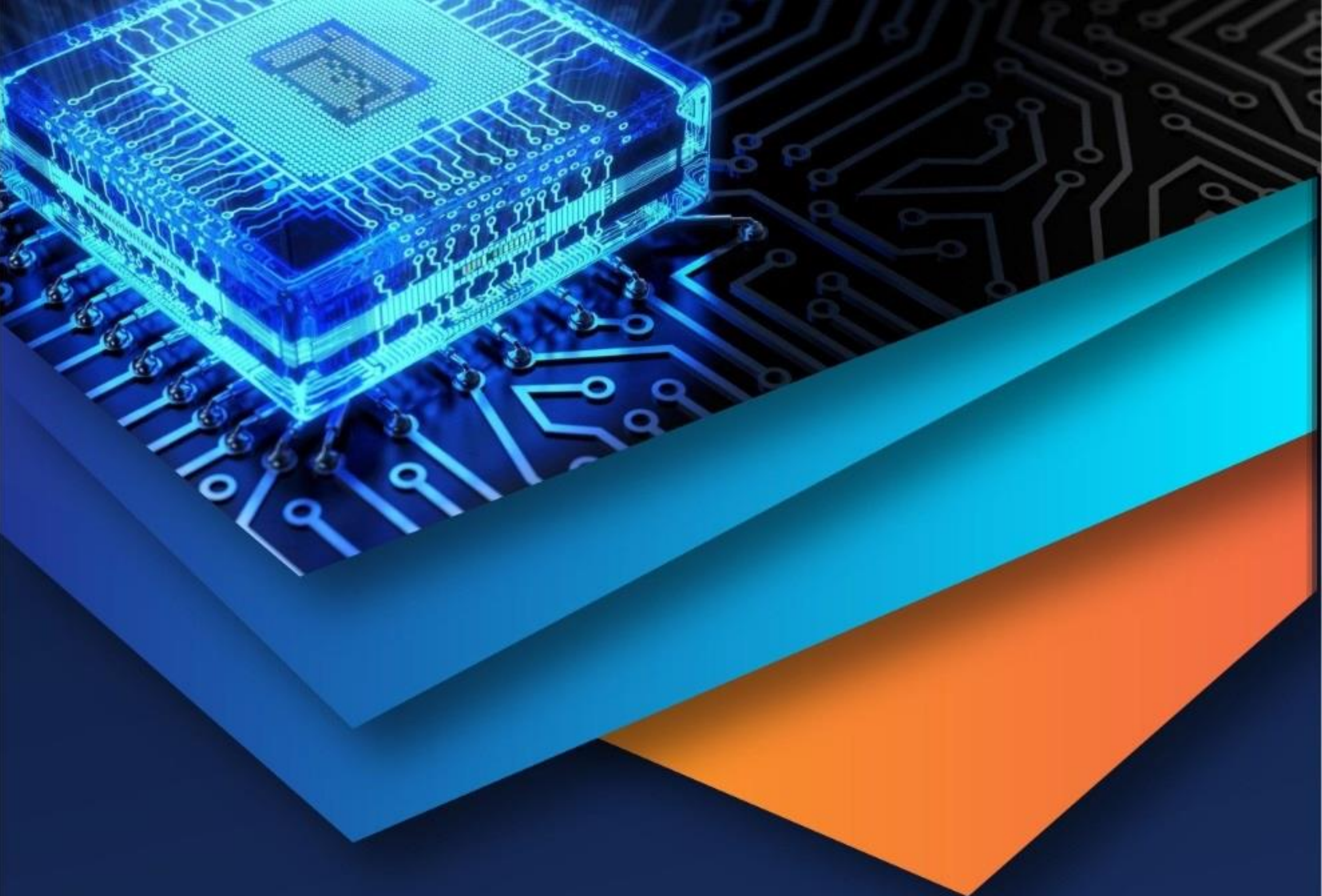
[26] M. Kristou, A. Ohya, and S. Yuta, "Panoramic vision and lrf sensor fusion based human identification and tracking for autonomous luggage cart," in Robot and Human Interactive Communication, 2009. RO-MAN 2009. The 18th IEEE International Symposium on, 2009, pp. 711–716.

[27] C. Schlegel, J. Illmann, H. Jaberg, M. Schuster, and R. Wörz, "Vision Based Person Tracking with a Mobile Robot.," in BMVC, 1998, pp. 1–10.

[28] N.-Y. Ko, D.-J. Seo, and Y.-S. Moon, "A Method for Real Time Target Following of a Mobile Robot Using Heading and Distance Information," J. Korean Inst. Intell. Syst., vol. 18, no. 5, pp. 624–631, Oct. 2008.

[29] K. S. Nair, A. B. Joseph, and J. I. Kuruvilla, "Design of a low cost human following porter robot at airports."

[30] V. Y. Skvortzov, H.-K. Lee, S. Bang, and Y. Lee, "Application of Electronic Compass for Mobile Robot in an Indoor Environment," in 2007 IEEE International Conference on Robotics and Automation, 2007, pp. 2963–2970.



The Impact on Worldwide of Labour Force Employment Since Last 19 Years (2000-2019) by using Different Classification Techniques of Pattern Recognition

Gajanand A. Boywar[1], Chandrakant Patil[1], Pratik S. Jaiswal[1], Mukta G. Dhopeshwarkar[1]

[1]Department of Computer Science & I.T, Dr. Babasaheb Ambedkar Marathwada University, Aurangabad, Maharashtra - 431004 India.

---***---

Abstract. The acknowledgment of work power interest rate and joblessness rate has turned into a significant part pondering in Employments Sector. Example acknowledgment turns into a significant part as the entire investigation comprise of statistics. In the present examination, distinctive classifier of example recognition have been utilized. In which we have found extraordinary result by utilizing DecisionTree classifier, HoeffdingTree classifier, J48 classifier, LMT classifier, RandomForest classifier, RandomTree classifier, and in conclusion REPTree classifier i.e., 33.4298%, 39.725%, 82.1274%, 79.3054%, 98.4986%, 98.4624%, 66.4797% individually. When we apply DecisionTree classifier the result is poor interestingly with other classifier for example 33.4298%. what's more, Other hand the RandomForest classifier result is high conversely with other classifier for example 98.4986%. in case on the off chance that we apply classifier of same kind data later on, we could extraordinary results by the utilization of above classifier. Just DecisionTree classifier will be considered as uncommon. **Keyword.** Tree Classifier, DecisionTree, HoeffdingTree, J48, LMT, RandomForest, RandomTree, and REPTree.

1 Introduction

International Labor Organization (ILO) programmed on labor force participation rate and unemployment rate is part of large international effort on demographic estimated and projection to which several UN agencies contribute. The principle target of ILO program is to give individual from States, International organizations and the general population everywhere with the most extensive, itemized and practically identical work and Unemployment of the work power for nations and regions, the world all in all and its primary topographical area. the essential information are single-year work power interest rates by sex (percent)group and populace matured 15 years and over[9].

2 Data Information

The work control adventure dataset gather the data over the word by assembled in region cunning, nation wise and Area skillful. The dataset in collected between years 2000 to 2019[8-12]. In addition, the dataset pace of energy of utilized and jobless by sex (percent). The given break down dataset is amass from International Labor Organization(ILO),Geneva, Key Indicators of the Labor Market(KILM ninth discharge) and the ILOSTAT database, last got to January 2019. Additionally, Population developed 15 years and over, with the exception of if for the most part footnoted.

1. Information avoids Armenia, Azerbaijan, Cyprus, Georgia, Israel and Turkey.

2. Caucasus alludes to Armenia, Azerbaijan, Cyprus, Georgia, Israel furthermore, Turkey.

3. Populace matured 16 years and over.

4. Populace matured 14 years and over.

5. Occupant populace (by right).

6. Barring the institutional populace.

7. For factual purposes, the information for China do exclude those for the Hong Kong Special Administrative Region (Hong Kong SAR), Macao Special Administrative Region (Macao SAR) and TaiwanProvince of China.

8. Populace matured 15 to 69 years.

9. Populace matured 16 to 65 years.

10. Populace matured 15 to 74 years.

11. Nationals, occupants.

12. Populace matured 18 to 64 years.

13. Populace matured 15 to 64 years.

14. People present (true).

15. True populace.

16. Populace matured 17 years and over.

17. Fundamental city or metropolitan zone.

18. Excluding a few territories.

3 Methodology

Right off the bat, the way toward acquiring and fusing model information has been significantly changed so as to completely incorporated the dataset assemble from United Nations static division. At that point it was not in line way, so we need that to clean the database base in an effective way. For that first, we are perfect the no of occasions semantic goofs which are shown in the database classification[13]. Beginning their ahead, we refreshed all characteristics in a certifiable way. When we have done the whole above database cleaning process the database is set up to perform specific activities on it.

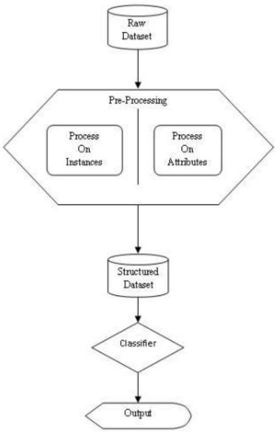


Fig.1. Illustration of Dataset process

DecisionStump is usually used in conjunction with a boosting algorithm. Does regreesion(based on mean-squared error) or classification(based on entopy). Missing is treated as separate value[1]. A HoeffdingTree (VFDT) is an incremental, anytime decision tree induction algorithm that is capable of learning from massive data streams, assuming that the distribution generating examples does not change over time[2]. J48 classes for generating a pruned or unpruned C4 which based on 5 decision tree[3]. LMT is used to classifier for building 'logostic model trees'. Which are classification trees with logistic regression functions at the leaves[4]. The RandomForest classifier used for create class for constructing a forest of random trees[5]. RandomTree used to create class for constructing a tree that considers chosen attributes at each node. Whoch performs no pruning. Also has an option to allow estimation of class probabilities(or target mean in the regression case) based on hold-out set (backfitting)[6]. The REPTree is a fast decision tree learner which builds a decision/regression tree using information gain/variance and prune it using reduced-error pruning[7].

4 Implementation

In the Labour Force Participation rate and unemployment rate(LFPRAUR) dataset the 5528 number of occasions subsequent to apply this classifier, the case in grouped in a different result. In DecisionStump classifier 33.4298% , a HoeffdingTree classifier is 39.725%, the J48 classifier is 62.1274%, the LMT classifier is 79.3054, the RandomForest is 98.4986%, the RandomTree classifier is 98.4624%,and at the last is REPTree Classifier is 66.4797%.

Table1. Result of DecisionStump

	No. Of instance	Result
Correctly Classified Instances	1848	33.4298%
Incorrectly Classified Instances	3680	66.5702%
Kappa statistic	0.2001	
Mean absolute error	0.2273	
Root mean squared error	0.3371	
Relative absolute error	81.8315%	
Root relative squared error	90.4608%	
Total Number of Instances	5528	

After performing the DecisionStump classifier operation on dataset with 5528 number the result is shows as in table1 which state that the Correctly Classified Instances 1848 which is in 33.4298 %, Incorrectly Classified Instances 3680 which is 66.5702 %, Kappa statistic is 0.2001, Mean absolute error is 0.2273,Root mean squared error0.3371, Relative absolute error is 81.8315 %, Root relative squared error is 90.4608 %.

Table1.1. Confusion Matrix of DecisionStump

a	b	c	d	e	f
922	1	0	0	0	0
5	926	0	0	0	0
916	0	0	0	0	0
5	916	0	0	0	0
803	113	0	0	0	0
17	904	0	0	0	0

Where, a = Labour force participation Total, b = Unemployment rate Total, c = Labour force participation Male, d = Unemployment rate Male, e = Labour force participation Female, f = Unemployment rate Female.

Table2. Result of HoeffdingTree

	No. Of instance	Result
Correctly Classified Instances	2196	39.725%
Incorrectly Classified Instances	3332	60.275%
Kappa statistic	0.2759	
Mean absolute error	0.234	
Root mean squared error	0.3374	
Relative absolute error	84.2287%	
Root relative squared error	90.5386%	
Total Number of Instances	5528	

Subsequent to playing out the HoeffdingTree classifier activity on dataset with 5528 number the outcome is appears as in table2 which express that the Correctly Classified Instances 2196 which is in 39.725 %, Incorrectly Classified Instances 3332 which is 60.275 %, Kappa measurement is 0.2759, Mean outright mistake is 0.234, Root mean squared blunder 0.3374, Relative supreme blunder is 84.2287 %, Root relative squared blunder is 90.5386 %.

Table2.1. Confusion Matrix of HoeffdingTree

a	b	c	d	e	f
719	5	62	0	137	0
1	927	0	0	3	0
497	1	376	0	42	0
1	918	0	0	2	0
601	119	22	0	174	0
9	906	0	0	6	0

Where, a = Labour force participation Total, b = Unemployment rate Total, c = Labour force participation Male, d = Unemployment rate Male, e = Labour force participation Female, f = Unemployment rate Female.

Table3. Result of J48

	No. Of instance	Result
Correctly Classified Instances	4540	82.1274%
Incorrectly Classified Instances	988	17.8726%
Kappa statistic	0.7855	
Mean absolute error	0.07	
Root mean squared error	0.187	
Relative absolute error	25.1833%	
Root relative squared error	50.183%	
Total Number of Instances	5528	

In the wake of playing out the J48 classifier activity on dataset with 5528 number the outcome is appears as in table3 which express that the Correctly Classified Instances 4540 which is in 82.1274 %, Incorrectly Classified Instances 988 which is 17.8726 %, Kappa measurement is 0.7855, Mean outright mistake is 0.07, Root mean squared blunder 0.187, Relative total blunder is 25.1833 %, Root relative squared blunder is 50.183 %.

Table3.1. Confusion Matrix of J48

a	b	c	d	e	f
879	1	26	0	17	0
0	746	0	77	1	107
26	0	886	0	4	0
0	310	0	556	3	52
35	6	7	1	862	5
4	266	1	26	13	611

Where, a = Labour force participation Total, b = Unemployment rate Total, c = Labour force participation Male, d = Unemployment rate Male, e = Labour force participation Female, f = Unemployment rate Female.

Table4. Result of LMT

	No. Of instance	Result
Correctly Classified Instances	4384	79.3054%
Incorrectly Classified Instances	1144	20.6946%
Kappa statistic	0.7517	
Mean absolute error	0.0924	
Root mean squared error	0.2052	
Relative absolute error	33.249%	
Root relative squared error	55.0739%	
Total Number of Instances	5528	

Subsequent to playing out the LMT classifier activity on dataset with 5528 number the outcome is appears as in table4 which express that the Correctly Classified Instances 4384 which is in 79.3054 %, Incorrectly Classified Instances 1144 which is 20.6946 %, Kappa measurement is 0.7517, Mean supreme blunder is 0.0924, Root mean squared mistake 0.2052, Relative outright mistake is 33.249 %, Root relative squared mistake is 55.0739 %.

Table4.1. Confusion Matrix of LMT

a	b	c	d	e	f
841	0	28	1	53	0
1	492	0	232	2	204
10	0	900	0	6	0
0	113	0	661	1	146
104	0	14	1	786	11
5	99	0	105	8	704

Where, a = Labour force participation Total, b = Unemployment rate Total, c = Labour force participation Male, d = Unemployment rate Male, e = Labour force participation Female, f = Unemployment rate Female.

Table5. Result of RandomForest

	No. Of instance	Result
Correctly Classified Instances	5445	98.4986%
Incorrectly Classified Instances	83	1.5014%
Kappa statistic	0.982	
Mean absolute error	0.0682	
Root mean squared error	0.1268	
Relative absolute error	24.5602%	
Root relative squared error	34.029%	
Total Number of Instances	5528	

Subsequent to playing out the RandomForest classifier activity on dataset with 5528 number the outcome is appears as in table 5 which express that the Correctly Classified Instances 5445 which is in 98.4986 %, Incorrectly Classified Instances 83 which is 1.5014 %, Kappa measurement is 0.982, Mean supreme mistake is 0.0682, Root mean squared blunder is 0.1268, Relative total mistake is 24.5602 %, Root relative squared mistake is 34.029 %.

Table5.1. Confusion Matrix of RandomForest

a	b	c	d	e	f
922	0	0	0	1	0
0	898	0	22	0	11
0	0	916	0	0	0
0	18	0	896	0	7
0	0	0	0	916	0
0	15	0	8	1	897

Where, a = Labour force participation Total, b = Unemployment rate Total, c = Labour force participation Male, d = Unemployment rate Male, e = Labour force participation Female, f = Unemployment rate Female.

Table6. Result of RandomTree

	No. Of instance	Result
Correctly Classified Instances	5443	98.4624%
Incorrectly Classified Instances	85	1.5376%
Kappa statistic	0.9815	
Mean absolute error	0.0051	
Root mean squared error	0.0506	
Relative absolute error	1.8452%	
Root relative squared error	13.5837%	
Total Number of Instances	5528	

In the wake of playing out the RandomTree classifier movement on database with 5528 number the result is appeared as in table6 which express that the Correctly Classified Instances 5543 which is in 98.4624%, Incorrectly Classified Instances 83 which is 1.5376 %, Kappa estimation is 0.9815, Mean preeminent error is 0.0051, Root mean squared screw up 0.0506, Relative all out bungle is 1.8452%, Root relative squared bumble is 13.5837%.

Table6.1. Confusion Matrix RandomTree

a	b	c	d	e	f
923	0	0	0	0	0
0	931	0	0	0	0
0	0	916	0	0	0
0	49	0	872	0	0
1	0	0	0	915	0
0	34	0	0	1	886

Where, a = Labour force participation Total, b = Unemployment rate Total, c = Labour force participation Male, d = Unemployment rate Male, e = Labour force participation Female, f = Unemployment rate Female.

Table7. Result of HoeffdingTree

	No. Of instance	Result
Correctly Classified Instances	3675	66.4797%
Incorrectly Classified Instances	1853	33.5203%
Kappa statistic	0.5976	
Mean absolute error	0.12	
Root mean squared error	0.2449	
Relative absolute error	43.1991%	
Root relative squared error	65.726%	
Total Number of Instances	5528	

In the wake of playing out the REPTree classifier activity on dataset with 5528 number the outcome is appears as in table7 which express that the Correctly Classified Instances 3675 which is in 66.4797%, Incorrectly Classified Instances 1853 which is 33.5203%, Kappa measurement is 0.5976, Mean total blunder is 0.12, Root mean squared mistake 0.2449, Relative outright blunder is 43.1991%, Root relative squared mistake is 65.726 %.

Table7.1. Confusion Matrix of REPTree

a	b	c	d	e	f
806	1	65	0	51	0
2	814	0	54	5	56
153	0	759	0	4	0
2	614	0	283	6	16
185	17	7	4	702	1
12	557	0	13	28	311

Where, a = Labour force participation Total, b = Unemployment rate Total, c = Labour force participation Male, d = Unemployment rate Male, e = Labour force participation Female, f = Unemployment rate Female.

5 Conclusion

After the few classifier tries the result on the work power cooperation, Unemployment rate, in which Male support, and there joblessness rate,also the Female work power investment and there joblessness rate discovered exceptional outcome by using DecisionTree classifier, HoeffdingTree classifier, J48 classifier, LMT classifier, RandomForest classifier, RandomTree classifier, and in end REPTree classifier.When we apply DecisionTree classifier the outcome is poor strikingly with other classifier for instance 33.4298%. in addition, Other hand the RandomForest classifier result is high then again with other classifier for instance 98.4986%. On the off chance that in case we apply classifier of same kind information later on, we could phenomenal outcomes by the use of above classifier. Just DecisionTree classifier will be considered as phenomenal.

6 References

1. United Nations static division http://data.un.org
2. Geoff Hulten, Laurie Spencer, Pedro Domingos: “Mining time-changing data streams”. In: ACM SIGKDD Intl. Conf. on Knowledge Discovery and Data Mining, 97-106, 2001.
3. Ross Quinlan (1993). C4.5: “Programs for Machine Learning”. Morgan Kaufmann Publishers, San Mateo, CA.
4. Niels Landwehr, Mark Hall, Eibe Frank (2005). Logistic Model Trees. Machine Learning. 95(1-2):161-205George H. John, Pat Langley: “Estimating Contiguous Distributions in Bayesian Classifiers”. In Eleventh Confrence on Uncertainly in Artificial Intelligence, San Mateo, 338-345,1995.
5. Leo Breiman (2001). Random Forests. Machine Learning. 45(1):5-32
6. ILO, Labour force projections, 1965-85 (1st edition, Geneva 1971).
7. ILO, Labour force projections, 1950-2010 (2nd edition, Geneva 1976).
8. ILO, Economically Active Population: Estimates and projections, 1950-2025 (3rd edition, Geneva 1986).
9. ILO, Economically Active Population Estimates and projections, 1950-2010 (4th edition, Geneva 1996).
10. ILO, Estimates and Projections of the Economically Active Population, 1980-2020 (5th edition, Geneva 2007, Update August 2008,Update December 2009).
11. The reduction from 193 to 189 reference areas is due to the inclusion of the data of Dominique, Martinique, French Guyana and Réunion into metropolitan France. The change in horizon from 2050 to 2030 allows to base the projections solely on econometricmethods.
12. ILO, Estimates and Projections of the Economically Active Population, 1990-2030 (2013 edition, Geneva 2013); ILO, Labour Force. Estimates and Projections, 1990-2050 (2015 edition, Geneva 2015).
13. Anupriya K. Kamble, Ramesh R. Manza, Yogesh M. Rajput, Kavita A. Khobragade ”Classification of Insulin Dependent Diabetes Mellitus Blood Glucose Level Using Support Vector Machine”, IOSR Journal of Computer Engineering (IOSR-JCE), e-ISSN: 2278-0661,p-ISSN: 2278-8727(2017).