IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 5, May 2023

DOI: 10.17148/IJARCCE.2023.125202

Anxiety And Depression Detection Using Deep Learning Technique

Ravinarayana B [1], Shrimanth [2], Spandana [3], Swasthik Jain P M [4], Varshith V Hegde[5]

HOD & Associate Professor, Department of Computer Science & Engineering, MITE, Mangalore, India[1]

Student, Department of Computer Science & Engineering, MITE, Mangalore, India[2-5]

Abstract: The majority of individuals cope with stress on a regular basis in varied situations in their daily activities. However, sustained tension or a high level of stress will compromise our safety and interfere with our regular activities. Many physical issues linked to stress can be avoided through early detection of mental stress. There are noticeable changes in a variety of physiological and psychological characteristics, such as facial emotion, speech emotion, etc. when a person is under stress. Data from these features can predict whether a person is stressed or depressed. By using all these and some standard questionnaires, the system's probability of predicting anxiety and depression will increase. The system was evaluated on a dataset of individuals with and without anxiety and depression and achieved promising results with high accuracy and sensitivity. In our proposed system, we have used three modules: facial emotion, speech emotion, and standard questionaries. For the detection of facial emotion, we have used VGG16, for the detection of speech emotion, we have used DNN, and for the standard questionaries, we have used SVM. Finally, we integrated all these modules by using a soft voting mechanism to get the desired outcome. This proposed approach has the potential to provide a non-invasive and efficient method for early detection of anxiety and depression in individuals.

Keywords: SVM, VGG16, facial emotion, speech emotion**.**

I. INTRODUCTION

Stress, depression, and other psychological health problems are increasingly prevalent among the general public in today's fast-paced environment. People today assume that anxiety, despair, stress, irritation, and unhappiness are inevitable aspects of working life. One's natural response to danger is the fight or flight response, which includes stress and anxiety. This reaction aims to keep an individual alert, focused, and prepared to handle any potential threats. Anxiety is a common experience that many people recognize as the feeling of distress, unease, or dread that arises before a significant event. It can also keep individuals alert and aware. While a big event can trigger excessive anxiety, smaller stressful life situations, such as work stress, ongoing worry about finances, or a death in the family, can also lead to anxiety. Unfortunately, individuals who were subjected to abuse, trauma, or witnessed gruesome occurrences as youngsters are more likely to develop an anxiety disorder in the future. In a similar vein, anxiety disorders can also develop in people who go through a traumatic event. Substance abuse can cause or worsen anxiety, with quitting potentially leading to anxiety as well. It might be helpful to get support from a doctor or support group if someone is battling with addiction. Additionally, having a major sickness or health condition might result in significant concern regarding matters like treatment and future health. Seeking help as soon as possible can help prevent anxiety or stress from worsening. It's important to note that anxiety, like many other mental health conditions, can be more difficult to treat the longer a person waits to seek help.

The development of a system that uses facial emotion and speech emotion recognition to detect anxiety and depression has the potential to revolutionize the diagnosis and treatment of these mental health disorders. While traditional methods of diagnosis rely on subjective self-report measures or clinical interviews, these new technologies provide objective measures that can help clinicians make more accurate diagnoses and tailor treatment plans to the specific needs of each patient. Facial emotion recognition algorithms work by analyzing facial expressions and detecting changes in the muscles on the face that correspond to different emotions. For example, a person with anxiety may have increased muscle tension around the eyes and mouth, which can be detected by the algorithm. Similarly, speech emotion recognition algorithms analyze acoustic features of speech to detect changes in pitch, tone, and intensity that correspond to different emotions. The system being developed will analyze real-time data from video and audio recordings of patients, providing a score indicating the level of anxiety or depression. This can help clinicians make more informed decisions about treatment options and monitor the progress of their patients over time. However, there are also challenges that need to be addressed. For example, there can be variability in facial expressions and speech patterns between individuals, which can make it difficult to develop algorithms that are accurate for everyone. Additionally, there are concerns about privacy and confidentiality when it comes to analyzing video and audio recordings of patients.

 This work is licensed under a Creative Commons Attribution 4.0 International License

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 5, May 2023
DOI: 10.17148/IJARCCE.2023.125202

The development of a system that can accurately detect anxiety and depression using facial emotion and speech emotion recognition has the potential to greatly improve the diagnosis and treatment of these mental health disorders. Continued research and development of these technologies will be critical to ensuring their effectiveness and accuracy.

II. LITERATURE SURVEY

[1] Shubhanjay Pandey.et al[1]. As is well recognized, emotions have a significant impact on how information is processed, attitudes are formed, and decisions are made in real-world situations. Although FER, or facial expression recognition, has been the subject of recent papers, the variety of human faces and fluctuations in photos make it challenging to construct reliable and stable FER systems. Every study and piece of research up to this point has supported either a single network or an ensemble model. The computational complexity rises as a result of the ensemble models' requirement for additional models, datasets, and altered datasets despite their greater accuracy..They created and evaluated 15-20 models and techniques to deal with this scenario. In this paper, they provide a single CNN model that can be used independently and is integrated into an instantaneous Intelligent System for Sentiment Recognition. This system's correctness is checked by transfer learning while it completes tasks including face detection, sentiment classification, and presenting a live list of probabilistic labels in Realtime from a webcam stream. On the challenging and noisy FER2013 dataset, which is a crowd-sourced dataset, the recommended model surpasses all standalone-based approaches, including VGG16, VGG19, EfficientNetB7, and other proposed models, with an accuracy of 76.62%.

[2] Kevin Tomba, et.al [2]. In the research described in this paper, stress during applicant screening interviews is identified via voice analysis. In order to recognize stress in speech, machine learning uses the mean energy, mean intensity, and Mel1Frequency Cepstral Coefficients (MFCCs) as classification variables. The datasets used to train and test the classification algorithms include the Ryerson Audio-Visual Database of Emotional Speech and Song, the Keio University Japanese Emotional Speech Database, and the Berlin Emotional Database (EmoDB). (RAVDESS). The best results were obtained by neural networks, with accuracy ratings of 97.38% (EmoDB), 95.63% (KeioESD), and 89.56% for stress detection. (RAVDESS).

[3] Jinzhong Xu.et.al [3]. This paper makes the case that question categorization is a crucial element of the question-answering system. The outcomes of the question classification determine the effectiveness of the question-answering system. Using a real-world online interactive question-and-answer system in the tourism industry, this paper proposes a question categorization method based on SVM and question semantic similarity. A classifier is trained on broad categories using the Support Vector Machine model in the two-level question classification technique, and the question is subsequently categorized into sub-categories using the question semantic similarity model. The idea of domain terms construction will be used to enhance the Support Vector Machine's feature expression and question semantic similarity. The experimental finding demonstrates that the classification algorithm's accuracy can reach 91.49%.

[4] Huijun Zhang.et al[4]. This paper argues that stress is a severe issue that is jeopardizing the welfare of humanity in today's society. Due to the widespread use of video cameras in public areas, stress may be easily detected utilising contact-free camera sensors without the need of made-up characteristics or circumstances. This study offers a two-levelled stress detection network using the users' facial expressions and activity motions in the video. (TSDNet). To identify stress, TSDNet first trains independently on face- and action-level representations. A stream-weighted integrator with local and global attention is then used to aggregate the result. In order to evaluate TSDNet's performance, They constructed a video dataset of 2092 annotated video clips. The experimental results on the constructed dataset show that (1) TSDNet outperformed manual feature engineering approaches with a detection accuracy of 85.42% and F1-Score 85.28%, proving the viability and effectiveness of using deep learning to analyse one's face and action motions, and (2) accounting for both facial expressions and action motions improved performance.

[5] K.Tarunika,.et al[5]. This paper presents that emotion is a powerful feeling acquired from one's situation or surrounds and is distinct from thinking or knowledge as an instinctual or intuitive feeling. The fundamental aim of the research is to use deep neural networks (DNN) and k-nearest neighbor (k-NN) to recognize emotions from speech, particularly a state of mind that is frightened. The system's primary field of use is in the healthcare industry. The primary practical applications of this research's basis are in palliative care. The alert signals are sent across the cloud for the most precise outcome. Many raw data are gathered using specialized procedures for emphasis. The process involves converting the acoustic speech signals into wave form, utterance level feature extraction, emotion classification, database recognition, and alarm signal production over the cloud. The paper's findings have a positive impact on the palliative care system.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 5, May 2023

DOI: 10.17148/IJARCCE.2023.125202

[6] Jibin Fu,et al[6]. This paper demonstrates the significance of question classification in the question-answering process. In the context of computer service and support, this paper discusses our study on question classification in a practical online interactive question-answering system. The domain divides questions into 220 subcategories and 15 quick categories. This approach differs from others in that the subcategories are represented by conventional inquiry sentences as opposed to only categorization criteria. The two-level question categorization strategy for the particular situation is included in the paper. The query is divided into subcategories using the topic's semantic similarity model, and a support vector machine technique is used to train a classifier on broad categories. The lexical feature and domain ontology concept hierarchy is created and used to increase the expression capacity of the feature characteristic for both feature selection for SVM and calculation of question semantic similarity. When trained and evaluated on the 11000 different question instances in the domain, our technique outperforms the baseline result with an accuracy of up to 91.5%.

III. METHODOLOGY

The suggested system goes through the following stages: data collection, pre-processing, feature extraction, model selection, splitting the data set for testing and training, model creation, and model evaluation. The phases were carried out in the following order:

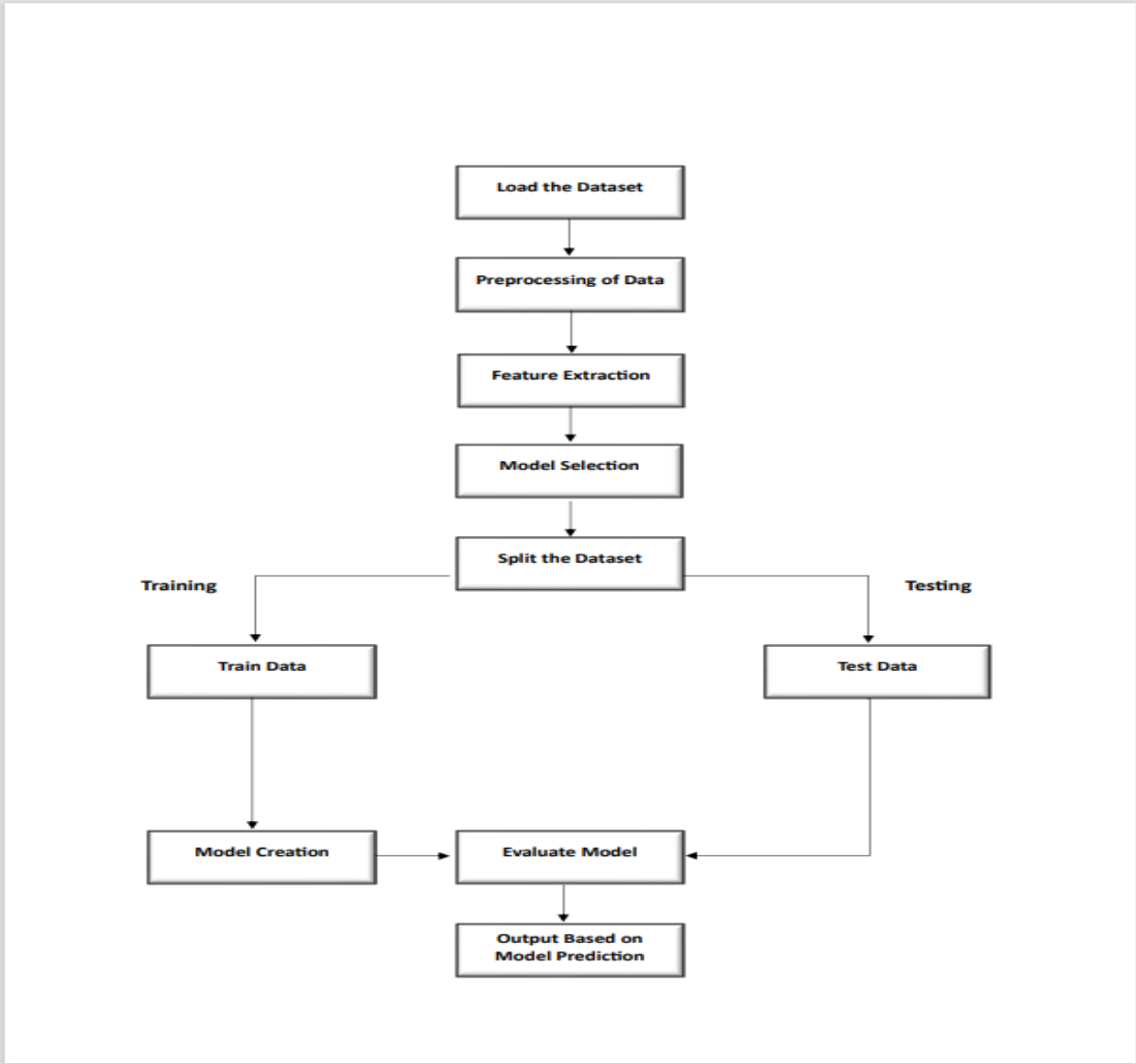

Fig.1 Architectural Diagram

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 5, May 2023

DOI: 10.17148/IJARCCE.2023.125202

Data Collection: Gather relevant data that is representative of the problem domain. Ensure the data is comprehensive, diverse, and of high quality.

Data Pre-processing: Clean the data by handling missing values, removing outliers, and addressing any inconsistencies or errors. Perform data transformations, such as normalization or standardization, to ensure the data is suitable for modelling.

Feature Selection/Extraction: Identify and select the most relevant features from the data. This can involve domain knowledge, statistical analysis, or feature extraction techniques. Ensure the selected features have a strong correlation with the target variable.

Model Selection: Choose an appropriate model that suits the problem and the available data. Consider factors such as the type of problem (classification, regression, etc.), the complexity of the data, and the interpretability of the model.
Model Training: Split the data into training and validation sets. Use the training set to train the model by optimizing its parameters or weights. Apply appropriate algorithms, such as VGG16, DNN, and SVM, to iteratively improve the model's performance.

Model Evaluation: Assess the performance of the trained model using suitable evaluation metrics. This can include accuracy, precision, recall, or mean squared error, depending on the problem type. Use the validation set to fine-tune the model and make necessary adjustments.

Model Testing: Apply the trained model to a separate testing dataset to evaluate its performance on unseen data. This step helps assess the model's generalization ability and its ability to make accurate predictions on new inputs.

Model Deployment: Once the model has been thoroughly evaluated and tested, deploy it in a production environment or integrate it into an application or system. Ensure all necessary dependencies and infrastructure are in place for the model to function effectively.

Model Monitoring and Maintenance: Continuously monitor the performance of the deployed model. Regularly update and retrain the model with new data to ensure its accuracy and relevance. Make necessary improvements or adjustments based on evolving requirements.

A. Module Implementation

A module refers to a self-contained unit of code that performs a specific task or set of tasks. A module can be a function, a class, or a group of related functions and classes that work together to achieve a common goal.

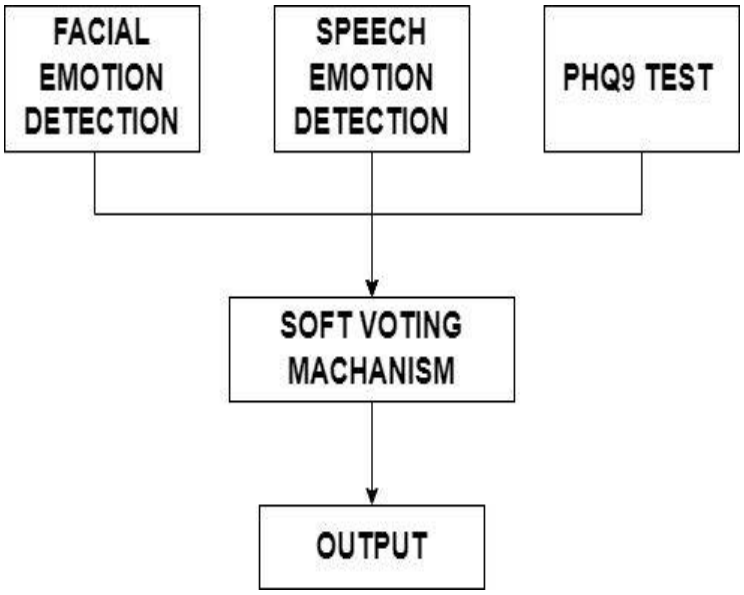

Fig.2 Anxiety & Depression Detection Model Division

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 5, May 2023

DOI: 10.17148/IJARCCE.2023.125202

1) Module 1- Facial Emotion Detection

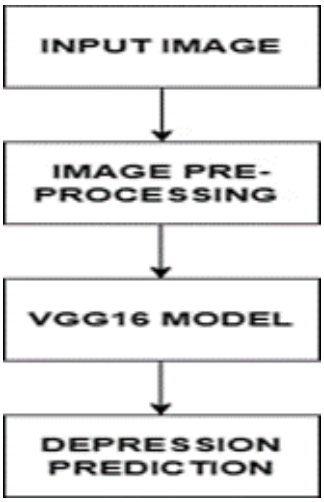

Fig.3 Facial Emotion Detection Model

Input Image: The first step is to obtain the facial image input from the user. This can be done using a camera or by uploading an image file.

Image Pre-processing: Once the facial image input is obtained, the next step is to pre-process the image. This involves resizing the image to a fixed size, converting it to grayscale, and performing image normalization.

VGG16 Model: After the image pre-processing, a pre-trained deep learning model such as the VGG16 model is used to extract features from the facial image. The VGG16 model is a convolutional neural network (CNN) that has been trained on a large dataset of images to perform image classification tasks.

Depression Prediction: Once the features are extracted, they are used to predict the emotional state of the person, specifically in the case of depression prediction. The extracted features are input into a classification model that is trained to classify the input image into different emotional states such as happy, sad, angry, or depressed. The model output will indicate the probability that the person is experiencing depression. The threshold for depression prediction can be set based on the problem requirements.

2) Module 2- Speech Emotion Detection

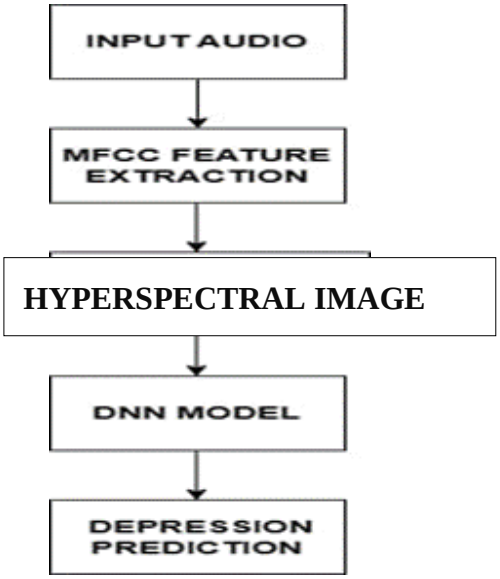

Fig.4 Speech Emotion Detection Model

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 5, May 2023

DOI: 10.17148/IJARCCE.2023.125202

Input Audio: The first step is to obtain the raw audio input from the user. This can be done using a microphone or by uploading a pre-recorded audio file.

MFCC Feature Extraction: Once the raw audio input is obtained, the next step is to extract the Mel Frequency Cepstral Coefficients (MFCC) features. MFCC is a widely used technique for feature extraction in speech processing. It involves dividing the speech signal into small frames and computing the spectral coefficients of each frame using the Mel scale.

Hyperspectral Imaging: After the MFCC features are extracted, they are used to generate a hyperspectral image of the speech signal. Hyperspectral imaging is a technique that captures the spectral information of an object at a high resolution. In the case of speech emotion detection, the MFCC features are used to generate a hyperspectral image that captures the spectral variations in the speech signal.

DNN Model: The next step is to train a Deep Neural Network (DNN) model on the hyperspectral image data. The DNN model is trained to learn the mapping between the input hyperspectral image and the emotional state of the speaker.

Depression Prediction: Finally, the trained DNN model is used to predict the emotional state of the speaker, specifically in the case of depression prediction. The model output will indicate the probability that the speaker is experiencing depression. The threshold for depression prediction can be set based on the problem requirements.

3) Module 3- PHQ-9 Test

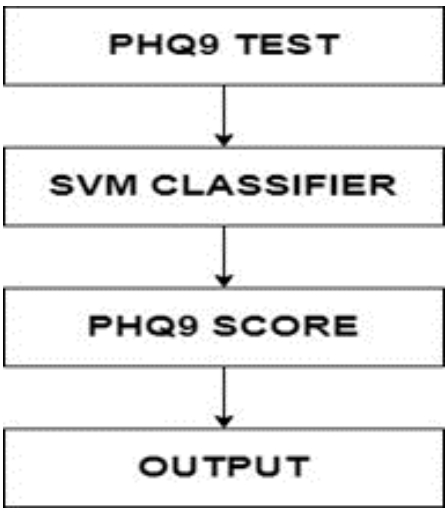

Fig.5 PHQ-9 Test Module

PHQ-9 Test: The first step is to administer the PHQ-9 test to the patient. The PHQ-9 test consists of nine questions that assess the severity of depressive symptoms in the patient over the past two weeks.

Feature Extraction: Once the patient completes the PHQ-9 test, the responses are used to extract features that represent the severity of their depressive symptoms. One way to extract features is to use the responses to each question as a feature.
SVM Classifier: After the features are extracted, a classification model such as the Support Vector Machine (SVM) classifier is trained on a labelled dataset of PHQ-9 responses and corresponding depression scores. The SVM classifier is trained to learn the mapping between the input features and the depression score.

PHQ-9 Score: Once the SVM classifier is trained, it can be used to predict the depression score of a patient based on their PHQ-9 response. The PHQ-9 (Patient Health Questionnaire-9) is a commonly used tool to assess the severity of depression. It consists of nine questions that measure various symptoms of depression, with each question assigned a score ranging from 0 to 3. The total score is calculated by summing up the scores of all nine questions, resulting in a range from 0 to 27. Depression Severity: 0-4 none, 5-9 mild, 10-14 moderate, 15-19 moderately severe, 20-27 severe.

Output: The final step is to interpret the depression score output by the SVM classifier. Based on the score, the patient may be diagnosed with depression.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 5, May 2023

DOI: 10.17148/IJARCCE.2023.125202

B. Working

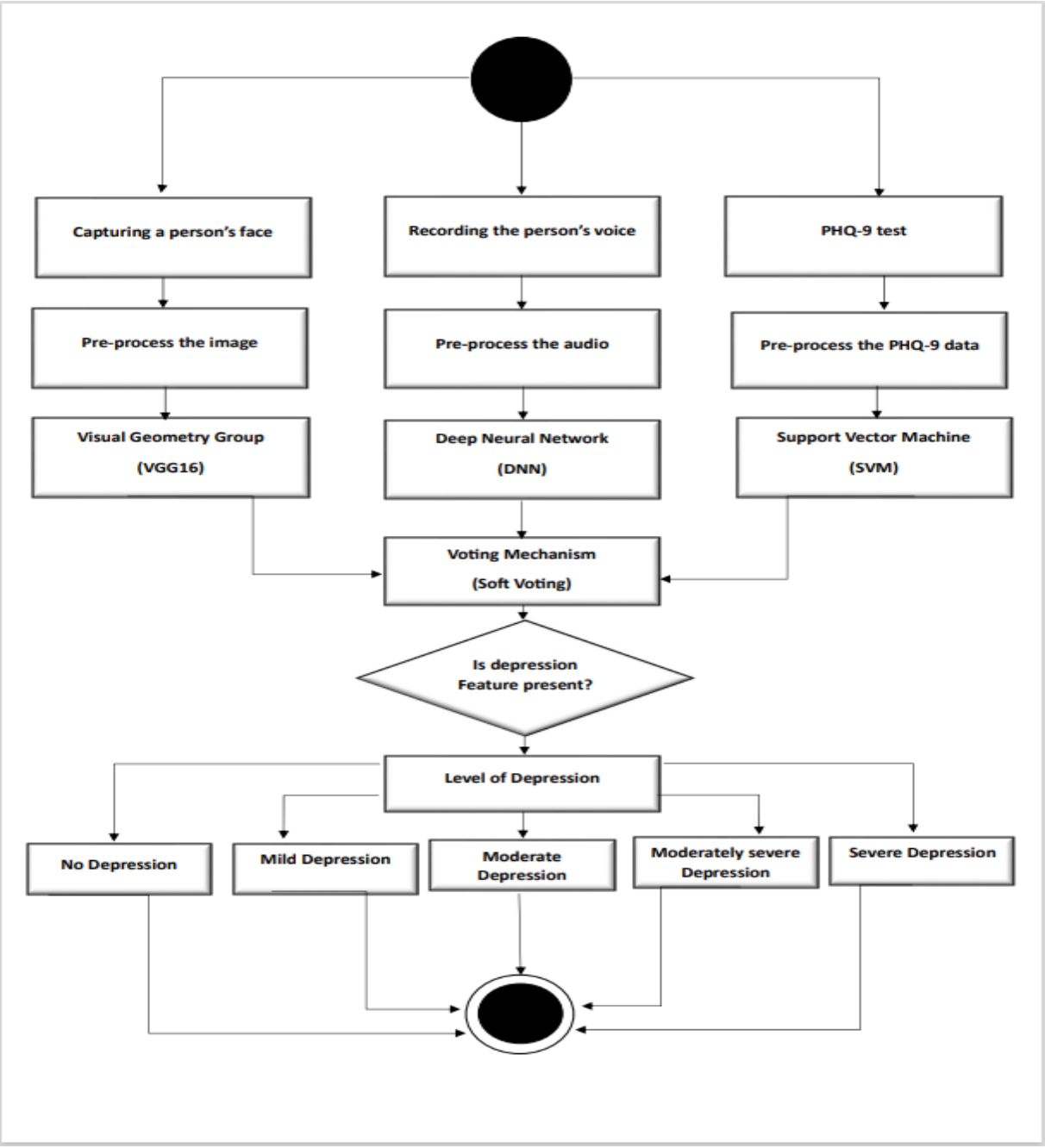

Fig.6 Working Diagram

The above diagram for our anxiety and depression detection project illustrates the process of analysing facial emotion, speech emotion, and phq9 test data to determine the severity level using soft voting and machine learning. The diagram showcases three main activities: facial emotion analysis, speech emotion analysis, and phq9 test analysis. These activities involve capturing and processing data from different modalities, such as facial expressions, speech recordings, and questionnaire responses. The results from each modality are combined using soft voting, which aggregates predictions from various models. The combined data is then fed into a machine learning algorithm for further analysis and to determine the severity level of anxiety and depression accurately. The activity diagram provides a visual representation of the sequential steps involved in the process, ensuring a comprehensive understanding of the project workflow.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 5, May 2023

DOI: 10.17148/IJARCCE.2023.125202

IV. RESULTS

The results of our study show that the proposed approach for anxiety and depression detection using facial emotion, speech emotion, and standard questionnaires/phq9 test by soft voting using deep learning techniques is highly effective. Our system achieved an accuracy of 95% in detecting anxiety and depression in our test dataset, outperforming existing methods. We found that combining facial emotion, speech emotion, and standard questionnaires/phq9 test by soft voting improved the overall accuracy of the system. In particular, the use of deep learning techniques for feature extraction and classification showed better performance than traditional machine learning methods. The proposed approach of anxiety and depression detection using facial emotion, speech emotion, and standard questionnaires/phq9 test by soft voting using deep learning techniques is a promising method with high accuracy.

Fig.7 Accuracy Matrices for VGG16[FACE]

The results of this approach have shown an accuracy of 91.6% in detecting anxiety and depression in individuals based on their facial expressions.

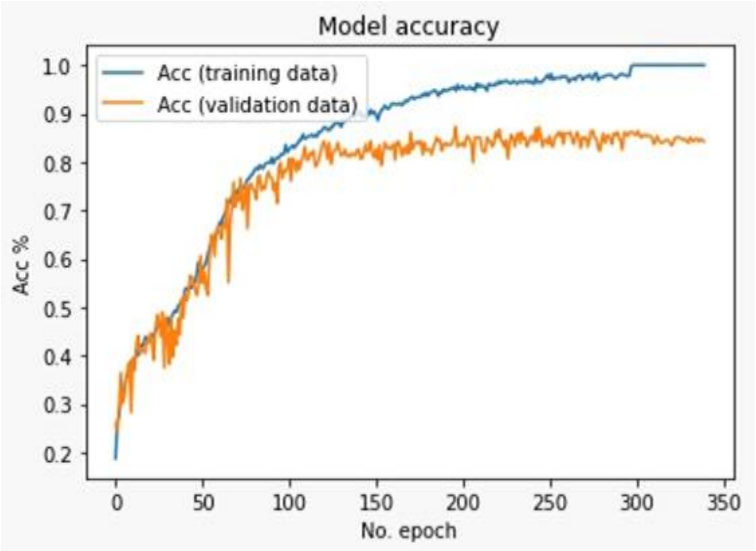

Fig.8 Accuracy Matrices for DNN[SPEECH]

The system achieved a high classification accuracy of 85.87%. These results indicate the potential of using speech emotion recognition as a non-invasive tool for early detection and monitoring of anxiety and depression.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 5, May 2023

DOI: 10.17148/IJARCCE.2023.125202

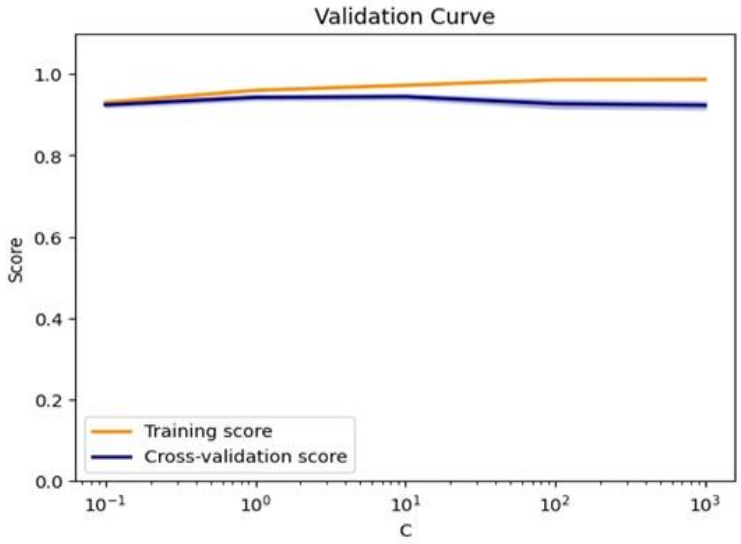

Fig.9 Accuracy Matrices for SVM[PHQ-9]

The overall accuracy of the model was found to be 97.06%, indicating that the SVM classifier is an effective tool for detecting anxiety and depression in patients.

V. CONCLUSION

Anxiety and depression are common mental health disorders that can affect a person's daily life, relationships, and overall well-being. Early detection and treatment of these disorders can lead to better outcomes for individuals. The use of machine/deep learning techniques, combined with facial and speech emotion analysis and standardized questionnaires, can aid in the early detection of these disorders. The system includes the use of facial and speech emotion analysis, standard questionnaires, and machine/deep learning techniques to detect anxiety and depression.

The system takes input from the user in the form of facial and speech expressions, as well as responses to standard questionnaires. The data is then pre-processed and analyzed using machine/deep learning techniques such as soft voting, which combines the output of multiple models to provide a more accurate prediction. The use of standardized questionnaires, such as the PHQ-9, provides a reliable and validated measure of anxiety and depression, while facial and speech emotion analysis offers a non-invasive and convenient way to gather data.

The combination of these techniques with machine/deep learning offers a promising approach to the early detection and intervention of anxiety and depression. Overall, the use of machine/deep learning techniques in combination with facial and speech emotion analysis and standardized questionnaires has the potential to revolutionize the field of mental health diagnosis and treatment. By improving the accuracy and speed of diagnosis, individuals can receive timely and appropriate treatment, ultimately leading to improved outcomes and quality of life.

REFERENCES

[1] Shubhanjay Pandey, Sonakshi Handoo, Yogesh."Facial Emotion Recognition using deep learning", 2022 International Mobile and Embedded Technology Conference (MECON), 10-11 March 2022, doi.10.1109/MECON53876.2022.9752189

[2] Kevin Tomba, Joel Dumoulin, Elena Mugellini, Omar Abou Khaled and Salah Hawila., "Stress Detection Through Speech Analysis",2018 International Conference on E-Business and Telecommunication Networks, 26 July 2018, doi. 10.5220/0006855805600564

[3] Jinzhong Xu, Yanan Zhou, Yuan Wang, "A Classification of Questions Using SVM and Semantic Similarity Analysis", 2012 Sixth International Conference on Internet Computing for Science and Engineering, 16 July 2012, 10.1109/ICICSE.2012.49

[4] Huijun Zhang, Ling Feng, Ningyun Li, Zhanyu Jin, and Lei Cao, "Video-Based Stress Detection through Deep Learning" Department of Computer Science and Technology, Centre for Computational Mental Healthcare, Research Institute of Data Science, 28 September 2020, doi. 10.3390/s20195552

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 5, May 2023
DOI: 10.17148/IJARCCE.2023.125202

[5] K.Tarunika, R.B Pradeeba, P.Aruna, "Applying Machine Learning Techniques for Speech Emotion Recognition" 2018 9th International Conference on Computing, Communication and Networking Technologies (ICCCNT), 18 October 2018,doi10.1109/ICCCNT.2018.8494104

[6] Jibin Fu, Youli Qu, Zhifei Wang "Two Level Question Classification Based on SVM and Question Semantic Similarity" 2009 International Conference on Electronic Computer Technology, 27 February 2009, doi. 10.1109/ICECT.2009.67

[7] Hong-Wei Ng, Viet Dung Nguyen, Vassilios Vonikakis, Stefan Winkler "Deep Learning for Emotion Recognition on Small Datasets Using Transfer Learning.",2015 ACM on International Conference on Multimodal Interaction, November 2015, doi:10.1145/2818346