International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 04 | April - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

D-SCAN : DEPRESSION DETECTION

Assistant Professor Mrs. Deepti N.N

Abhishek Sharma, Anusha Devi. E, Ashwini. R, Sidram. S.B

Department of Computer Science and Engineering
Ragiv Gandhi Institute Of Technology
Bangalore- 560032, Karnataka

Abstract - Depression is a serious illness that affects millions of people globally. From child to senior citizen are facing depression. Major area is occupied by adults, college going students and teenagers also. In recent years, the task of automation depression detection from speech has gained popularity. We provide a comparative analyses of various features for depression detection by evaluating how a system built on text-based, voice-based, and speech-based system. Detecting texts that express negativity in the data is one of the best ways to detect depression. In this paper, this problem of depression detection on social media and various machine learning algorithms that can be used to detect depression have been discussed.

***Key Words*:** Depression, Face detection, Audio detection, Video detection, Healthcare innovation, Result.

1.INTRODUCTION

Depression is a dysfunctional behavior that can influence anybody regardless of old enough, gender, status, and so forth. It extremely brunt a person's life affecting what they think about themself, their sleeping cycle, eating cycle, etc. It is the worst state of a person's mind when they feel sad and loses interest in nearly doing every productive thing and they can't simply move from that state. Factors like Social, Biological, and psychological factors are responsible for causing depression.

We provide a comparative analyses of various features for depression detection. we evaluate how a system built on text-based and speech-based system. We find that a combination of features drawn from both speech, video and text lead to the best system performance. Using a machine learning approach to detect depression will surely help social media users for detecting and predicting depression risk. Textual data being the most widely used form of communication offers a bunch of characteristics which makes it the best choice for doing data analysis, for emotion AI. depressed face expression has the same characteristics of a sad expression, such as the upward slanted eyebrows etc. but the main difference is that there is no major frown involved. Traditionally, depression is detected using standardized scales requiring patients' subjective responses or clinical diagnoses given by attending clinicians methods that have some shortcomings.

2. MOTIVATION

Main motivation of this depression detection system is to detect depression symptoms in early stage of depression, so that person will get alert and take necessary action, consult doctor in early stage itself, so that later stages will be avoided, and stage that is Suicide will get avoided. However, at early stages of depression, 70% of the patients would not consult doctors, which may take their condition to advance stages. Human experts will have privileged knowledge that codes the facial, text and audio features

3. Body of Paper

The analysis in this section focuses on the dataset used for training and testing in a deep learning process, as well as the preprocessing steps applied to the dataset and the configuration of the Convolutional Neural Network (CNN) model. Understanding the properties and characteristics of the speech signal is crucial in this analysis, as the speech signal is a complex and dynamic signal that changes rapidly over time.

I. Face Detection:

The primary goal of the face detection technique is to identify and locate faces in an image while minimizing external disturbances and other factors. The face detection process involves the following steps:

1) Image pyramid: The image is decomposed into multiple scales using an image pyramid. This technique helps extract features while reducing noise and other factors. The frame is smoothed and down-sampled iteratively, resulting in a frame.

2) Histogram of Oriented Gradients (HOG): HOG is a feature descriptor commonly used to detect objects in images. It quantifies the occurrences of gradient orientations within localized portions of an image. By utilizing HOG, the face within the image can be described using a set of intensity gradient distributions.

3) Linear Classifier: Once the Histogram of Oriented Gradients features are extracted, a linear classifier is employed to classify whether a given region of the image contains a face or not.

II. Audio Detection:

Voice is an index that reflects the characteristics of the vocal cords, and speech is an index that includes speech speed and hesitation. The voice detection process involves the following steps:

1) Our proposed approach for depression detection of audio signals is based on a graph neural network

2) This test utilizes the Librosa library to analyze voice patterns and extract Mel-frequency cepstral coefficients (MFCCs).

3) These features help identify changes in voice tone and pitch that are indicative of emotional states such as sadness or anxiety.

III. Text Detection:

Social media platforms have become a vast ocean of human expression in the current digital age, offering an unprecedented window into society as a whole. Here normal text as well as social media text are detected. The text detection process involve following steps:

4) Natural Language Processing: Natural Language Processing (NLP) and Machine Learning (ML) approaches have demonstrated significant promise in recent years for the diagnosis and understanding of depression from textual data. When utilising NLP and ML to detect depression, there are frequently a number of crucial phase.

5) By using the NLPK and CNN frameworks, the system can analyze the user's text input to detect emotions such as sadness, hopelessness, or anxiety.

6) This allows for a deeper understanding of the user's emotional narrative and provides insights into their state of mind based on their language use..

IV. Integration and Process:

1)Input – Audio, Speech and text (Data Set): The first step of the system is to input an audio speech and text

2)Pre-Processing: The second phase of the system deals with quality enhancement of the input signals of the audio speech and text. It may include silence removal, Pre-emphasis, noise removal, windowing and unwanted pauses, etc.

3)Feature Extraction: Feature extraction involves the analysis of the speech signal and text. The speech signal contains large number hidden information which reflects the emotional characteristics. It is considered as an important phase of the system as extraction of relevant and significant features heavily impact on the final recognition. Some of the features extracted by various researchers are MFCC (Mel-Frequency Cepstral Coefficients), LFPC(Log Frequency Power Coefficients), pitch, energy, and voice quality.

4)Classification: The fourth step is the main step of the system in which the audio speech and text is classified into different emotions based on the features extracted from the audio speech using CNN classifier. With the help of the features extracted, the audio speech is classified into different emotions. Then detect person audio is stress or not.

Depression detection through emotions: A depressed face expression has the same characteristics of a Sad expression, such as the upward slanted eyebrows etc. but the main difference is that there is no major frown involved. Also a sad face may have eyes lowered looking downward showing the helpless, dejected mood. On the contrary a depressed person can put forth a face devoid of depression. This depicts a case of concealed expression of depression, i.e. the depressed face may not be a sad face, and instead the personmay put forth a happy face to conceal depression. Individual person are classified as neutral or negative,based on a curated word-list to detect depression tendencies.

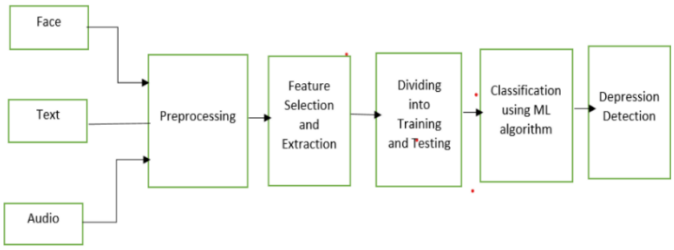

Table -1: Sample Table format

4. LITERATURE REVIEW

In machine learning, there are supervised machine learning classification algorithms for example Support Vector Machine, K Nearest Neighbor, Naive Bayes, Decision Tree, etc. Based on the different tasks and available data we can use those algorithms. The accuracy of the model can be improved using ensemble learning methods.

The first step in the process is to detect the face in the input image, followed by the identificationof facial features such as eyes and mouth. These features are then subjected to specific filters andtransformations [1]. The image classification system typically involves two stages: feature extraction and classification. In this paper [2], we utilize the deep learning library "Keras" provided by Google to develop a robust convolutional neural network (CNN) for facial emotion detection. We employ two different datasets and train our proposed network, evaluating its validation accuracy and loss accuracy. This study [3] focuses on addressing the challenges associated with Emotion Recognition Datasets and explores different parameters and architectures of Convolutional Neural Networks (CNNs) for the detection of seven emotions in human faces: anger, fear, disgust, contempt, happiness, sadness, and surprise. The proposed model achieves an accuracy of 91%, enabling effective tracking of human emotions through facial expressions [4]. High boost filtering is employed as a specialized technique to reduce image noise while preserving low-frequency components. This paper aims to compare different deep learning architectures available in Keras for emotion detection[5]. It utilizes Deep Facial Features in images by employing Transfer Learning from popular pre-trained models such as VGG16, ResNet152V2, InceptionV3, and Xception. Bottleneck features are generated for the input images, and the performance of these models is evaluated using a dataset that combines the Cohn-Kanade Dataset (CK+) and the Japanese female facial emotion (JAFFE) dataset. It combines digital image processing, video processing, pattern recognition, and other related technologies [6]. Research has revealed that the human brain's ability to recognize faces involves the coordinated activity of thousands of neurons. Machine learning and deep learning algorithms have harnessed this understanding to develop intricate networks that mimic neural functions. The rise of social media [7] has led to an overwhelming amount of textual data, necessitating an approach to extract meaningful

information from it. Aimens system proposes a solution to detect emotions such as happy, sad, and angry from contextual conversations. The system employs the Long short-term memory (LSTM) model with word2vec and doc2vec embeddings, achieving anF-score of 0.7185. The field of Speech Emotion Recognition [8] has seen an influx of techniques, with a focus on recognizing discriminant features in speech signals to classify emotions. Recently, deep learning techniques have shown promise in this field. In this study, a modified Deep Stride Convolutional Neural Network (DSCNN) architecture was proposed, with fewer convolutional layers to increase computational speed while maintaining accuracy. Speech emotion recognition involves the identification of human emotions and emotional statesthrough speech analysis [9]. The emotions considered in this experiment include neutral, joy, and sadness. Emotions are influenced by physical characteristics such as muscular tension, skin elasticity, blood pressure, heart rate, breath, and speech. Regarding speech emotion recognition (SER) using voice, recognition accuracy increases as more data are employed. In particular, in the case of deep learning, a large amount of data is essential [10]. The Open SMILE software is used for feature extraction, and the emotional features are obtained by calculating statistics using 12 functions from 16 original acoustic features.

5. ALGORITHMS.

7) Convolutional Neural Network: Convolutional Neural Networks specialized for ap- plications in image and video recognition. CNN is mainly used in image analysis tasks like Image recognition, Object detection Segmentation. There are Four types of layers in Convolutional Neural Networks

a) Convolutional Layer: In a typical neural network each input neuron is connected to the next hidden layer. In CNN, only a small region of the input layer neurons connect to the neuronhidden layer.

b) Fully-Connected layer: Fully Connected Layers form the last few layers in the network. The input to the fully connected layer is the output from the final Pooling or Convolutional Layer, which is flattened and then fed into the fully connected layer.

2) Mel-Frequency Cepstral Coefficients: MFCC are popular features extracted from speech signals for use in recognition tasks. In the source-filter model of speech, MFCC are understood to represent the filter (vocal tract). The frequency response of the vocal tract is relatively smooth, whereas the source of voiced speech can be modeled as an impulse train.

3) Support Vector Machine: In machine learning, support-vector machines (SVMs, also support-vector networks) are supervised learning models with associated learning algorithms that analyze data used for classification and regression analysis. Given a set of training examples, each marked as belonging to one or the other of two categories, an SVM training algorithm builds a model that assigns new examples to one category or the other, making it a non-probabilistic binary linear classifier (although methods such as Platt scaling exist to use SVM in a probabilistic classification setting). An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on the side of the gap on which they fall. Two types of SVM Linear SVM:

Linear SVM is used for linearly separable data, which means if a dataset can be classified into two classes by using a single straight line, then such data is termed as linearly separable data, and classifier is used called as Linear SVM classifier.

Non-linear SVM: Non-Linear SVM is used for non-linearly separated data, which means if a dataset cannot be classified by using a straight line, then such data is termed as non-linear data and classifier used is called as Non-linear SVM classifier

6. RESULTS

In the proposed project, D-Scan offers a holistic approach to depression detection by utilizing three different types of tests: facial expression detection, voice-based emotion detection, and text-based emotion detection. Each of these modalities plays a crucial role in accurately assessing a user's emotional state and providing them with appropriate recommendations to improve their well-being. After testing, the D-Scan system provides personalized recommendations to the user based on their emotional state. These recommendations may include music, videos, or yoga practices tailored to help alleviate depression and improve overall mental health. This personalized approach is a key advantage of the D-Scan project, as it ensures that the user receives support that is specifically tailored to their needs. Additionally, the project's use of advanced machine learning techniques ensures that the system is efficient, scalable, and adaptable to various contexts, making it a valuable tool for mental health professionals and individuals alike.

7. CONCLUSION.

In this paper, we have introduced an ensemble learning method that we are going to experiment on existing methods of depression detection. After that, we will get the model with more improved accuracy which can detect depression more accurately.

The project we have detected whether the person is depressed or not using Face, Text and Audio. We have used CNN (Convolutional Neural Network) for Face images training, for Face recognition we have used Random Cascade Algorithm. To detect depression using Text we have used SVM (Support Vector Machine) Algorithm. Lastly for Audio input, we have used MFCC for speech recognition.

We extract Low-level audio features to find the differences in the audio. Then, it is proved that the experiment is feasible through the division of different parts. Our work provides a new scheme for the follow-up study of depression recognition in the computer field.

D-Scan presents a novel approach to depression detection and intervention, offering a comprehensive and personalized solution that leverages state-of-the-art machine learning techniques. Its potential impact on mental health care is significant, providing a valuable tool for individuals and mental health professionals alike.

8. REFERENCES

[1] C. Jain, K. Sawant, M. Rehman and R. Kumar, "Emotion Detection and Characterization using Facial Features," 2018 3rd International Conference and Workshops on Recent Advances and Innovations in Engineering (ICRAIE), Jaipur, India, 2023, pp.

[2] A. Jaiswal, A. Krishnama Raju and S. Deb, "Facial Emotion Detection Using Deep Learning," 2020 International Conference for Emerging Technology (INCET), Belgaum, India, 2023, pp. 1-5.

[3] S. Begaj, A. O. Topal and M. Ali, "Emotion Recognition Based on Facial Expressions Using Convolutional Neural Network (CNN)," 2020 International Conference on Computing, Networking, Telecommunications & Engineering Sciences Applications (CoNTESA), Tirana, Albania, 2023, pp. 58-63.

[4] Ram Kumar Madupu, Chiranjeevi Kothapalli, Vasanthi Yarra "Automatic Human Emotion Recognition System using Facial Expressions with Convolution Neural Network" Fourth International Conference on Electronics, Communication and Aerospace Technology (ICECA- 2023).

[5] Hari Kishan Kondaveeti , Mogili Vishal Goud "Emotion Detection using Deep Facial Features" IEEE International Conference on Advent Trends in Multidisciplinary Research and Innovation (ICATMRI) 2022.

[6] Charvi Jain, Kshitij Sawant, Mohammed Rehman "The Application of Artificial Intelligence in Emotion Recognition" 2022 International Conference on Intelligent Computing and Human- Computer Interaction (ICHCI).

[7] E. Pranav, S. Kamal, C. Satheesh Chandran and M. H. Supriya, "Facial Emotion Recognition Using Deep Convolutional Neural Network," 2022 6th International Conference on

Advanced Computing and Communication Systems (ICACCS), Coimbatore, India, 2022, pp. 317-320.

[8] U. Rashid, M. W. Iqbal, M. A. Skiandar, M. Q. Raiz, M. R. Naqvi and S. K. Shahzad, "Emotion Detection of Contextual Text using Deep learning," 2022 4th International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT), Istanbul, Turkey, 2022, pp. 1-5.

[9] Susu Yan, Liang Ye, Shuai Han, Tian Han, Yue Li and Esko Alasaarela "Speech Interactive Emotion Recognition System Based on Random Forest".

[10] T. M. Wani, T. S. Gunawan, S. A. A. Qadri, H. Mansor, M. Kartiwi and N. Ismail, "Speech Emotion Recognition using Convolution Neural Networks and Deep Stride Convolutional Neural Networks," 2020 6th International Conference on Wireless and Telematics (ICWT), Yogyakarta, Indonesia, 2021.