Emotion Recognition, Depression Detection and Consultancy using Deep Learning

Mr. Anilkumar Hulsure*, Parth Ashtikar, Mitali Nilapwar*****

Prajwal Kulkarni****, Shubham Kambale*******, Dr. Suvarna Patil********

Artificial Intelligence and Data Science, Dr. D.Y. Patil Institute of Engineering Management and Research, Akurdi, Pune-44

Abstract- In some psychological balances, tension is often seen as an obstacle. If a person cannot reconcile what is expected of him with his ability to solve, stress can arise and place a heavy burden on the brain. There are two types of difficulties. Depression can often be defined as a disruption of mental balance. One of the important types of research in biomedical engineering is the study of depression because depression will be easy to prevent. Depression presents a significant challenge to global health, demanding innovative solutions for early detection and intervention. Traditional assessment methods relying on self-reporting and clinical evaluations are of- ten subjective and may lead to under diagnoses. To address this is- sue, this paper aims to develop a comprehensive system leveraging advanced technologies, including facial expression analysis, tone of speech recognition, and related modalities. The primary objective is to create a user-friendly tool capable of detecting signs of depression accurately, even in cases where individuals conceal their emotional struggles. By providing early intervention and support, the paper aims to improve mental health outcomes, ultimately contributing to the well-being of individuals affected by depression checked.

Index Terms- CNN, facial expression, psychological

I. Introduction

Depression and anxiety are common worldwide. It is necessary to pay attention to the negative effects of depression on patients' health and economic problems. To support depression testing, the social psychology community uses signal processing, computer vision, and machine learning to identify the language and behavior of people with depression and predict what patterns should indicate a depressive state. These studies we analyzed the relationship between objective measures of voice, speech, and nonverbal behavior and clinical assessments of depression severity for the purpose of stress assessment. Despite the progress made in recent years, there are still several open research areas in depression research that need to be addressed: information about speech, e.g., speech, facial expressions (AU) Non-verbal. Words in conversational context can affect a person's sleep, mood, thoughts, and other aspects of their life. For the diagnosis of depression, it is very important to investigate various functional aspects such as voice, vision and language and create various combinations [1].

Due to privacy issues, only depression data is currently available and there are almost no previous studies on depression. There are also inconsistencies in these widely used depression data. The fact that they have different languages, different times, different materials and different purposes makes it difficult to combine them to increase the number of models and therefore use deep patterns. Using some data enhancement methods to increase the number of models is necessary to improve the performance of the model.

Depression is a state of negative emotions and lack of interest in work. From this perspective, research on depression should be associated with psychology. However, current research on depression and mental health is not independent. We hypothesize that the combination of melancholia prediction and concurrent perspective will lead to a stronger analysis of melancholia.

Problem Description:

1. Imagine teaching a computer to understand how people feel and to recognize when someone might be really sad. We all have emotions like happiness, sadness, anger, and even computers can learn to recognize these emotions as they learn how to play. Now, think about how people sometimes feel so deep and sad, and it is hard to talk to them. This is called depression. We want to teach computers to look for signs of depression by listening to how people talk, reading what they write, and observing how they work.
2. The problem we are trying to solve is like teaching a computer to be better friends who can understand our emotions and know when we might need some help. Thus, a computer can help us to talk to someone who can comfort us when we are really sad. It's like having a little friend who cares about our feelings and wants to be there for us when we're going through tough times. your world is becoming more technologically advanced, and so is our prospects for better mental health. We tackle a challenge that touches the core of the human experience: understanding emotions, harnessing the power of deep learning and recognizing the symptoms of depression [3].
3. This generation could make our devices more knowledge and responsive, developing a brand-new degree of connection among humans and machines. Our goal is to leverage Deep Learning to discover hidden signs and symptoms of depression in how human beings talk, write, and behave. This method

building algorithms which can pick out up on adjustments that might imply someone is suffering emotionally, even though they do not show it outwardly. The coronary heart of our paper lies in combining the artwork of expertise feelings with the technology of information and algorithms. By teaching machines to interpret human alerts and behaviors, we are establishing doors to early interventions and advanced intellectual health care. But this is greater than a technical adventure[2].

4. It is approximately compassion and empathy, the usage of era to light up the path for people who might be strolling in darkness. Our paintings should probably exchange lives, supplying solace to the youngster going through emotional turbulence, the adult battling internal struggles, and the elder experiencing loneliness. Challenges will absolutely stand up; however, each impediment is a possibility to refine our solutions. This paper aims to address the critical issue by leveraging advanced technology, including facial expression analysis and speech recognition, to develop a system that can accurately detect signs of depression. The goal is to create a non-invasive, objective, and scalable solution that can be used in various settings, from healthcare facilities to mobile applications. By doing so, the paper seeks to improve the early detection of depression, enhance access to mental health care, and ultimately contribute to better mental health outcomes for individuals worldwide[10].

II. IDENTIFY, RESEARCH AND COLLECT IDEA

In this phase of paper, the aim was to identify, research, collect ideas relevant to emotion recognition, depression detection and consultancy using deep learning methodologies in which the following steps were undertaken:

1) Investigated recent studies on deep learning in mental health.
2) Identified obstacles like data availability and model interpretability.
3) Attend conferences, webinars and collaborated with mental health and deep learning specialists for insights.
4) Evaluated existing datasets for training deep learning models.

III. STUDIES AND FINDINGS

Depression is a serious problem that can affect people in many ways. Many treatments are available for people with depression, but the challenge is predicting people who do not know they are depressed. Therefore, we prepared some models to predict human depression, our main models are described in the article: a) using machine learning classifiers and WEKA, b) using graph view and machine learning, c) risk taking. Depression makes people sick not only physically but also mentally. It has many effects on people's quality of life. Additionally, depression does not have to be at a high level to affect a person's life. Many studies have been done on predicting depression, and three main methods have been studied to determine which is most accurate based on previous studies. After looking at machine learning classifiers, feature reduction method, cross-validation method, relatedness, the best and most accurate method is Bayesian network classifier for percentile selection.

A. System Architecture

The module architecture will be as:

- Prepossessing:

In this mode the machine will follow the given input. On the front end, the machine uses training data and removes noise from the given words. Then adjust the size of the file set.

- Feature Extraction:

In this module user will give Emotions happy, sad neutral, etc. That attribute gives to machine then extract the audio to text and detect emotions.

- Classification:

Show users the results of training data using the CNN algorithm (convolutional neural network). Machine learning can predict whether a person is depressed through feedback. We use machine learning with CNN (Convolutional Neural Network) here to get more accurate result.

B. Data Flow

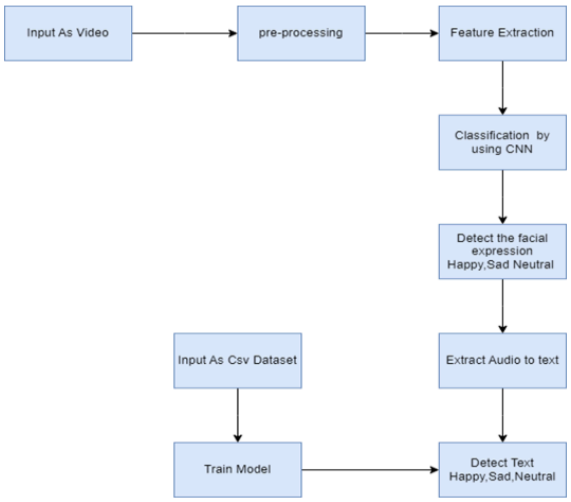

Fig 1: Flowchart of input algorithm and output

Data Acquisition: The system begins by gathering data relevant to depression detection. This data can include:
Video recordings of users' faces.
Audio recordings of users' speech.
Surveys or questionnaires to assess users' mental state.

Pre-processing: Once collected, the data undergoes pre-processing to ensure uniformity and quality for analysis by the deep learning models. This may involve:

For video data: Resizing frames to a standard dimension and normalizing pixel values.
For audio data: Extracting relevant features like pitch and mel-frequency cepstral coefficients (MFCCs).
For survey/questionnaire data: Formatting and cleaning the data for compatibility with the machine learning models.
Feature Extraction: Pre-processed data is then fed into deep learning models to extract features. Here, two separate models are likely used:

A CNN (Convolutional Neural Network) for visual features: The CNN extracts feature from the video data, likely focusing on facial features that might indicate depression symptoms.
An RNN (Recurrent Neural Network) for audio features: The RNN extracts features from the audio data, possibly focusing on speech patterns or prosody that can be associated with depression [7].
Classification: The extracted features from both the CNN and RNN are then used to train a classification model, possibly another neural network or a Support Vector Machine (SVM). This model is designed to classify whether a user exhibits signs of depression based on the extracted features.

Output: The system generates an outcome based on the classification model's prediction. This could be a binary classification (depression or not), or it could indicate a depression likelihood score.

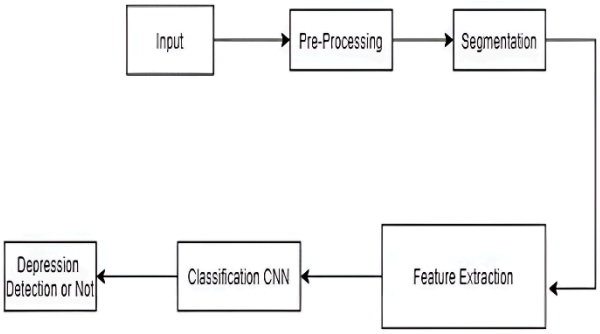

Fig 2: Steps to identify emotion recognition and depression detection

C. Exponential Graph

The exponential curve depicts the rapidly growing exploration of Deep Learning (DL) techniques in the domain of mental health. This surge reflects the increasing potential of DL to revolutionize how we approach:

Emotion Recognition: By analyzing facial expressions, speech patterns, and even physiological signals, DL models can effectively recognize emotions. This can be instrumental in early detection of mental health issues and for developing personalized interventions[6].

Depression Detection: DL algorithms, trained on vast datasets of speech, facial features, and user interactions, can analyze patterns associated with depression. This can aid in early diagnosis, potentially leading to more timely treatment[8].

Chatbot Consultancy: DL-powered chatbots can provide initial support and mental health resources. These chatbots can leverage emotion recognition to tailor their responses and offer guidance or suggest connecting with a mental health professional.

The exponential growth signifies the immense potential of DL in transforming mental healthcare. As research continues and datasets grow, we can expect even more sophisticated and effective DL-based applications to emerge in the coming years.

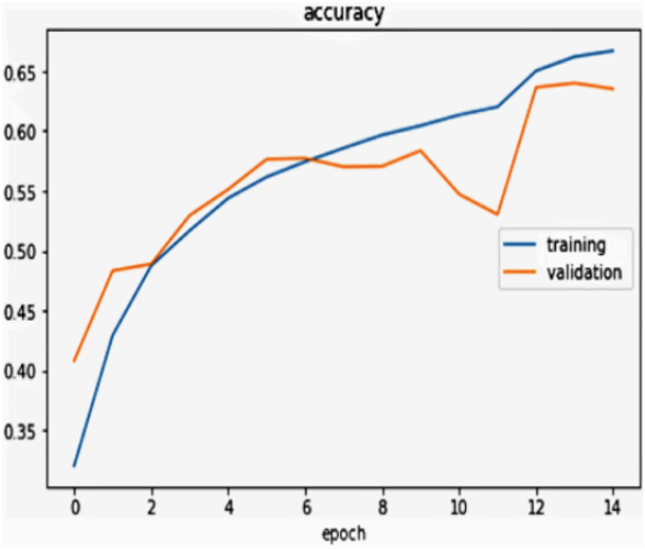

Fig 3: Exponential Graph

D. Proposed Algorithm

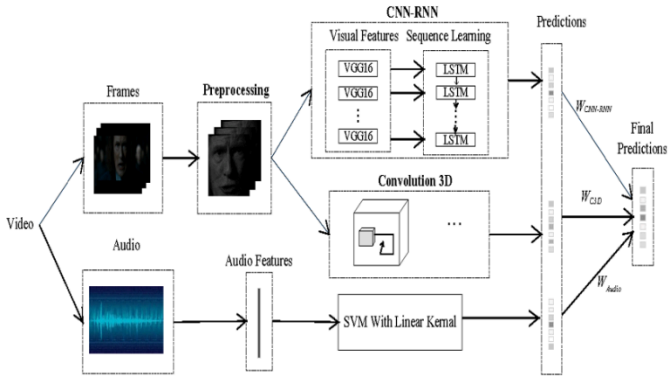

Fig 4: Proposed Algorithm

- Preprocessing: The system starts by pre-processing the input video. This likely involves resizing the frames to a uniform size and normalizing the pixel values.

- Visual Features: A pre-trained CNN, , is employed to extract visual features from each frame of the video. CNNs are adept at recognizing patterns in images and can efficiently capture facial features in each frame.

- Describing the process by which CNNs and sentiment analysis function:

  For the CNN:

- Input Preparation:
  Imagine you have grayscale images (like black and white photos) of size 48x48 pixels, showing faces.
  The model starts by looking at these images and trying to find patterns or features that represent different aspects of the face.

- Finding Features and Making Predictions:
  The model uses a series of steps called convolution and pooling to understand the images better. It gradually learns to recognize more complex patterns like lines, shapes, and textures.
  Each time it learns something new, it tries to remember it. But to avoid memorizing too much and getting confused, it randomly forgets some things during training. This helps it become better at recognizing faces in new pictures it has not seen before.

  For Natural Language Processing using Sentiment Analysis:

- Getting Text Ready:
  Extracting audio from the video input and classifying text from the audio.
  First, we break down the text into smaller pieces, like individual words or phrases. Then we clean it up a bit, removing unnecessary stuff like punctuation or capital letters.
  Once it is tidy, we convert each word into numbers so the computer can understand it. We use special tricks to represent each word in a way that captures its meaning.

- Understanding and Guessing Feelings:
  Now the computer reads through these numbers and tries to figure out what the words mean when put together. It learns to understand the overall tone or feeling of the text.
  Finally, it guesses how positive or negative the text is. It is like asking the computer, "Do you think this comment is happy, sad, or something in between?" And based on what it learned; it gives an answer.

- Sequence Learning: A type of RNN, specifically Long Short-Term Memory (LSTM) network, is then used to process the sequence of visual features extracted from each frame. LSTMs are capable of modeling temporal dependencies, meaning they can account for the order in which the frames appear and how emotions change over time in the video.

- CNN-RNN: The integration of these two deep learning architectures is known as CNN-RNN. Here, the visual features are presumably fed into another CNN, possibly to incorporate additional spatial information from the previous CNN layer. The resulting features are then incorporated into the LSTM for sequential learning.

- Audio Features: The system can also incorporate audio features extracted from the video. Here, an SVM with a linear kernel is used for classification. SVMs are a type of machine learning model that excels at categorizing data, and linear kernels are functions used in SVMs to learn the relationship between the input data and the categories.

- Fusion and Prediction: Finally, the visual and audio features are presumably merged to make the final emotion predictions. The specific fusion method is not illustrated in the image but could involve concatenating the features from the two modalities (visual and audio) before feeding them into a final classifier.

Overall, this CNN-RNN algorithm leverages the strengths of both CNNs and RNNs to recognize emotions from videos. CNNs effectively capture visual features from individual

frames, while RNNs model the temporal dependencies between frames, essential for understanding the flow of emotions in a video.

IV. CONCLUSION

This paper marks a significant milestone in the application of AI for mental health support. While challenges exist, the system's potential to positively impact individuals' mental well-being is promising, and ongoing efforts will continue to refine and expand its capabilities. In this paper we detect the facial emotions and Research Advancements: Stay abreast of the latest research in emotion recognition, depression detection, and AI technologies. Continuously integrate new findings and advancements into the paper to enhance accuracy and capabilities. Also various capabilities are possible such as

- Mobile Application: Develop a dedicated mobile application for the consultancy system to reach a broader user base, providing support through smartphones and ensuring accessibility.
- Multilingual Support: Expand the system to support multiple languages, making it more inclusive and accessible to a global audience.
- Machine Learning Model Interactions: Incorporate multiple machine learning models that can work collaboratively to enhance the accuracy and reliability of the system. AI Chatbot Integration: Consider integrating AI-driven chatbots for real-time interaction with users, providing immediate responses and support.

REFERENCES

1. Y. Kwon and N. Lobo, "Age classication from facial images," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Seattle, WA, USA, Jun. 1994, pp. 762–767.
2. Y. Kwon and N. Lobo, "Age classication from facial images," Comput. Vis. Image Understand., vol. 74, no. 1, pp. 1–21, 1999.
3. W. B. Horng, C. P. Lee, and C. W. Chen, "Classication of age groups based on facial features," Tamkang J. Sci. Eng., vol. 4, no. 3, pp. 183–192, 2001.
4. J. Hayashi, M. Yasumoto, H. Ito, Y. Niwa, and H. Koshimizu, "Age and gen- der estimation from facial image processing," in Proc. 41st SICE Annu. Conf., vol. 1. Osaka, Japan, Aug. 2002, pp. 13–18.
5. Lanitis, C. Taylor, and T. Cootes, "Toward automatic simulation of aging effects on face images," IEEE Trans. Pattern Anal. Mach. Intell., vol. 24, no. 4, pp. 442–455, Apr. 2002.
6. Fevre, Mark Le; Kolt, Gregory S.; Matheny, Jonathan,. "EuDepression, diDe- pression and their interpretation in primary and secondary occupational De- pression management interventions: which way first?". Journal of Managerial Psychology, 2006, 21 (6): 547 -565. doi:10.1108/0268390610684391.
7. Das S, O'Keefe J. "Behavioral cardiology: recognizing and addressing the pro- found impact of psychosocial Depression on cardiovascular health". Curr Atheroscler Rep. 2006; 8:111-
8. Shusterman V, Aysin B, Gottipaty V, Weiss R, Brode S, Schwartzman D, Anderson KP. "Autonomic Nervous System Activity and the Spontaneous Ini- tiation of Ventricular Tachycardia." ESVEM Investigators. Electrophysiologic Study Versus Electrocardiographic Monitoring Trial. J Am Coll Cardiol. 1998; 32:1891-9.73
9. Brodsky MA, Sato DA, Iseri LT, Wolff LJ, Allen BJ. "Ventricular Tach- yarrhyth- mia Associated with Psychological Depression." JAMA. 1987; 257:2064-7 PPCT-FIM: prepost computation tree based Frequent itemset mini

AUTHORS

First Author –Anilkumar, M.Tech, Dr.D Y Patil Institute of Engineering, Management & Research and anil.hulsure@dypiemr.ac.in.

Second Author – Parth Ashtikar, BE, , Dr.D Y Patil Institute of Engineering, Management & Research and parthdyp@gmail.com

Third Author – Mitali Nilapwar , BE, Dr.D Y Patil Institute of Engineering, Management & Research and mitaliunilapwar@gmail.com,

Correspondence Author – Dr.Suvarna Patil , Prajwal Kulkarni , Shubham Kambale,
suvarnapatil@dypiemr.ac.in.
prajwalkulkarni16pk@gmail.com,
shubhamappasaheb.kambale@gmail.com.
Dr.D Y Patil Institute of Engineering, Management & Research