International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 05 | May - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

ARTIFICIAL INTELLIGENT SYSTEM FOR AUTOMATIC DEPRESSION LEVEL ANALYSIS THROUGH VISUAL AND VOCAL EXPRESSIONS

Sarita Pandharinath Tambe[1] and Dr. R. G. Dabhade[2]
Student, Department of VLSI & Embedded System (E&TC)[1]
Professor, Department of VLSI & Embedded System (E&TC)[2]
Matoshri College of Engineering and Research Centre, Eklhare, Nashik, Maharashtra, India

Abstract:

Depression is a major mental health disease of human which is rapidly affecting lives worldwide. Early detection and intervention are crucial for effective treatment and management. Depression is regularly identified with thoughts of suicide. Significant depression can bring about a determination of social and physical side effects. It could remember changes in rest, craving, vitality level, concentration, day by day conduct or confidence. In recent years, deep-learned applications concentrated on neural networks have shown superior performance at hand-crafted apps in various areas. This system presents an innovative artificial intelligence (AI) system designed for automatic depression level analysis by analyzing visual and vocal expressions. Leveraging advances in computer vision and natural language processing, the proposed system extracts relevant features from facial expressions and speech patterns to assess depression severity levels accurately. Deep-learned apps that settle the above issues that may precisely assess the degree of voice and face depression. In the proposed method, Convolutionary Neural Networks (CNN) is developed for learning deep-learned features and descriptive raw waveforms for visual expressions. Second.

Keywords: Artificial Intelligence, Depression, Vocal, Facial Expressions, *Deep Learning*

1. INTRODUCTION

In many situations humans who are depressed are totally ignorant of their disturbed mental condition. They are unable to identify the cause of constant unhappiness in them and eventually such students fall into a state of mind where they start having suicidal tendencies. In some cases students do know that they are suffering from depression, but they are hesitant to seek any kind of help from anyone mainly due to the wrongly conceived notion of 'humiliation' associated with depression. It is better to identify the signs of depression at initial stages of depression. Depression if identified in the initial stages, just a simple one hour talk with a counselor may be of immense help for the student. This may totally change the negative state of mind of that student into a positive one. Such a student can be given good counseling of how to deal with mental stress and can be guided to follow the right path to success. The most important form of non-verbal communications is facial expressions of a person. Many studies have been done for finding out the facial expressions that are related to depression. The current work is mainly undertaken to find out the presence of depression in college students by studying their facial features. This system mainly uses different image processing techniques for face detection, feature extraction and classification of these features as depressed or non-depressed. The system will be trained with features of depression. Then videos of different students with frontal face will be captured using a web camera. Then the facial features of these faces will be extracted for prediction of depression. Based on the level of depression features the student will be classified as depressed or non-depressed.

- Facial mood detection according to time series image inputs
- Predict mood level based on score or weight with class label.
- Successfully implement the test model based

on training set as supervised learning approach.

- Execute the proposed system maximum accuracy.

II. Literature Survey

This review provides an overview of the use of facial expression analysis in detecting depression. It covers various methodologies, including machine learning and deep learning techniques, for extracting features from facial images and analysing them to infer depression levels. Facial Expression Analysis for Depression Detection: A Comprehensive Review (2019) by Li et al. [1]. Focusing on vocal expression analysis, this review explores the use of acoustic features extracted from speech recordings for automatic depression detection. It discusses the effectiveness of different feature extraction methods and machine learning algorithms in identifying vocal cues indicative of depression. Automatic Depression Detection Using Vocal Acoustic Features: A Comprehensive Review (2020) by Alghowinem et al. [2]. This survey paper provides insights into the multimodal fusion techniques used for depression analysis, particularly integrating visual and vocal modalities. It discusses various fusion strategies, such as early fusion, late fusion, and hybrid fusion, and their impact on improving depression detection accuracy. Multimodal Fusion for Depression Analysis: A Survey (2021) by Zhang et al. [3].

Focusing on deep learning approaches, this review examines the application of deep neural networks in detecting depression from various modalities, including visual and vocal expressions. It discusses the advantages and challenges of deep learning models in capturing complex patterns associated with depression. Deep Learning for Depression Detection: A Review (2018) by D'Mello and Kory [4]. This review paper provides an in-depth analysis of automatic depression detection using speech analysis techniques. It discusses the role of different speech features, such as prosodic, spectral, and temporal features, in identifying vocal markers of depression and highlights recent advancements in machine learning algorithms for speech analysis. Automatic Detection of Depression in Speech: A Review (2019) by Cummins et al [5]. Focusing on multimodal analysis, this study investigates the effectiveness of integrating vocal and visual features for depression detection. It explores feature fusion techniques and evaluates the performance of multimodal models in capturing complementary information from different modalities. Multimodal Analysis of Vocal and Visual Features for Depression Detection (2020) by Zhao et al [6]

III. Problem Statement

Depression is a widespread mental health disorder affecting millions of individuals worldwide, with profound implications for personal well-being and societal functioning. However, timely detection and intervention remain significant challenges, often hindered by subjective assessment methods and limited access to mental health resources. Traditional approaches to depression assessment rely heavily on self-reporting and clinician evaluations, which may be prone to biases and inaccuracies. To address these challenges, this project aims to develop an artificial intelligence (AI) system capable of automatically analysing depression levels through the integration of visual and vocal expressions. By leveraging advancements in computer vision and natural language processing, the proposed system seeks to extract relevant features from facial expressions and speech patterns to assess depression severity levels accurately.

IV. SYSTEM DEVELOPMENT

DESIGN METHODOLOGY:

The proposed research to design and implement a system for depression level prediction using deep learning, the visual features has extracted from users face and predicts the scale of depression.

PROPOSED SYSTEM:

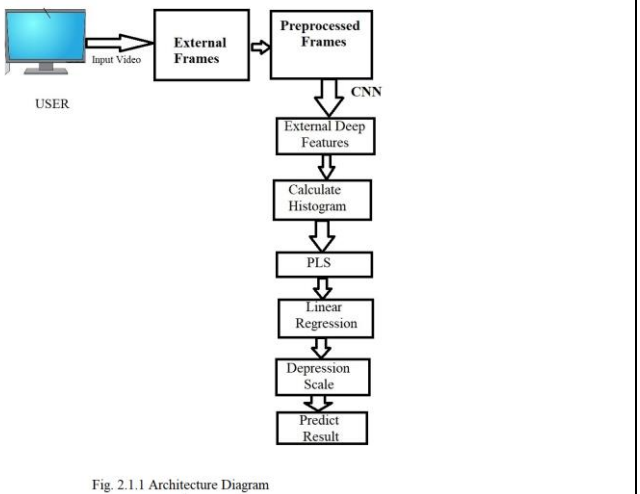

Fig. 2.1.1 Architecture Diagram

Following is the methodology used in proposed system:

- The image data were collected from Kaggle.
- The collected dataset is divided into 2 parts. i.e.:- 80% for training and 20% for testing
- Various Techniques like Preprocessing, feature extraction is applied
- CNN was used for classification
- Web application is been developed using Php and bootstrap for frontend and Python for backend.
- The user captured image is passed and captured images feature are extracted.
- Extracted Features will be matched with the trained model, depending on nearby match the predicted output isbeen obtained

Hardware Requirements:

- Processor – I3 / I5
- CPU
- GPU
- Networking
- Accelerators
- Server
- Cooling System
- Speed – 2.6 GHz
- RAM – 4 GB(min)
- Hard Disk - 128 GB SSD
- Monitor – SVG
- Power Supply

The hardware requirements for an Artificial Intelligence system for Automatic Depression Level Analysis through Visual and Vocal Expressions can vary depending on the scale of the project, the complexity of the algorithms, and the desired performance.

Software Requirements:

- **Operating System - Windows / Linux**
- **Front End – python Django**
- **Database - My SQL 5.0**

The software requirements for an Artificial Intelligence system for Automatic Depression Level Analysis Through Visual and Vocal Expressions will involve a combination of programming languages, libraries, frameworks, and development tools. Here's a list of key software components:

Programming Languages:

Python:

Python is widely used for developing machine learning and deep learning applications due to its extensive libraries and frameworks.

Image processing in Python

As the name says, image processing means processing the image and this may include many different techniques until we reach our goal. The final output can be either in the form of an image or a corresponding feature of that image. This can be used for further analysis and decision making.

V. RESULT AND CONCLUSION

A.RESULT:

The results of an artificial intelligence system for automatic depression level analysis through visual and vocal expressions typically involve the system's accuracy, precision, recall, and F1 score in detecting depression levels based on visual and vocal cues.

Automatic identification of depression attracted in expanding attention from specialists in brain research, software engineering and related controls. Subsequently, vocal and visual expressions are utilized to recognize depression level. In this proposed work these efforts by presenting review of depression detection systems and discuss best practices and most promising approaches to this task. Behaviour of a depressed individual shows relative change as far as discourse design, outward appearances and head development when contrasted with a non-depressed individual. The audio recordings are pre-processed and the audio segments of the user's speech are retrieved. The features are then extracted from audio segments and normalized. Finally, Neural Network is trained to detect depression or predict depression severity level.

In this system, an artificial intelligent system was proposed for automatic depression scale prediction. This is based on facial and vocal expression in naturalistic video recordings. Deep learning techniques are used for visual feature extraction on facial expression faces. Based on the idea of MHH for 2-D video motion feature, we proposed FDHH that can be applied to feature vector sequences to provide a dynamic feature (e.g., EOH_FD, LBP_FD, LPQ_FD, deep feature V32_FD, etc.) for the video. This dynamic feature is better than the alternate approach of MHH_EOH that was used in previous research, because it is based on mathematical feature vectors instead of raw images. Finally, PLS regression and LR are adopted to capture the correlation between the feature space and depression scales. The experimental results indicate that the proposed method achieved good state-of-the-art results on the AVEC2014 dataset. When comparing the hand-crafted versus deep features taken from the correct layer shows significant improvement over hand-crafted. With regards to selecting the correct layer, it seems that features should be extracted directly from the convolution filters responses. Generally the earliest fully connected layer will perform be the best, although the performances are fairly close to call. Audio fusion contributed in getting state-of-the-art results using only the MFCC feature, demonstrating that a multimodal approach can be beneficial.

B. CONCLUSION

In conclusion, we presented a novel approach to optimize word-embedding for classification tasks. We performed a comparative evaluation on some of the widely used deep learning models for depression detection from tweets on the user level. We performed our experiments on publicly available datasets. Our experiments showed that our CNN-based models perform better than RNN-based models. Models with optimized embedding have managed to maintain performance with the generalization ability.

ACKNOWLEDGMENT

Apart from my efforts, the success of any work depends largely on the encouragement and guideline of many others. I would like to deeply thank the various people who, during the several months in which this endeavor lasted, provided us with useful and helpful assistance. I take this opportunity to express our gratitude to the people who have been instrumental in the successful completion of the Seminar task. Without their care and consideration, this task would likely not have matured. Thanks to all other supporting people who helped me to improve my knowledge and skill required to enhance my work for this seminar. Last but not the least; I am very much thankful to my parents who guided me in every step which I took.

REFERENCES

[1] A. Krizhevsky, I. Sutskever and G. E. Hinton, "ImageNet classification with deep convolutional neural networks", *Proc. Adv. Neural Inf. Process. Syst.*, pp. 1097-1105, 2012.

[2] Alghowinem et al, Automatic Depression Detection Using Vocal Acoustic Features: A Comprehensive Review (2020) [3] Alghowinem, Sharifa, Roland Goecke, Jeffrey F. Cohn, Michael Wagner, Gordon Parker, and Michael Breakspear. "Cross-cultural detection of depression from nonverbal behavior." In Automatic Face and Gesture Recognition (FG), 11th IEEE International Conference and Workshops on, vol. 1, pp. 1-8. IEEE, 2015

[4] Cummins et al Automatic Detection of Depression in Speech: A Review (2019)

[5] D'Mello and Kory Deep Learning for Depression Detection: A Review (2018)

[6] Girard, Jeffrey M., Jeffrey F. Cohn, Mohammad H. Mahoor, Seyedmohammad Mavadati, and Dean P. Rosenwald. "Social risk and depression: Evidence from manual and automatic facial expression analysis." In Automatic Face and Gesture Recognition (FG), 10th IEEE International Conference and Workshops on, pp. 1-8. IEEE, 2013

[7] Harati, Sahar, Andrea Crowell, Helen Mayberg, Jun Kong, and Shamim Nemati. "Discriminating clinical phases of recovery from major depressive disorder using the dynamics of facial expression." In Engineering in Medicine and Biology Society (EMBC), 38th Annual International Conference of the, pp. 2254-2257, IEEE, 2016

[8] Jha et al Challenges and Opportunities in Automatic Depression Detection Using Multimodal Data: A Review (2021)

[9] J. R. Williamson et al., "Vocal biomarkers of depression based on motor incoordination", *Proc. 3rd ACM Int. Workshop Audio/Vis. Emotion Challenge (AVEC)*, pp. 41-48, 2013.

[10] Maddage, Namunu C., Rajinda Senaratne, Lu-Shih Alex Low, Margaret Lech, and Nicholas Allen. "Video-based detection of the clinical depression in adolescents." In Engineering in Medicine and Biology Society, (EMBC). Annual International Conference of the IEEE, pp. 3723-3726. IEEE, 2009

[11] M. F. Valstar et al., "AVEC 2014: 3D dimensional affect and depression recognition challenge", *Proc. Int. Conf. ACM Multimedia Audio/Vis. Emotion Challenge Workshop*, pp. 3-10, 2014.

[12] M. Nasir, A. Jati, P. G. Shivakumar, S. N. Chakravarthula and P. Georgiou, "Multimodal and multiresolution depression detection from speech and facial landmark features", *Proc. 6th Int. Workshop Audio/Vis. Emotion Challenge (AVEC)*, pp. 43-50, 2016.

[13] Meng, Hongying, Di Huang, Heng Wang, Hongyu Yang, Mohammed AI-Shuraifi, and Yunhong Wang. "Depression recognition based on dynamic facial and vocal expression features using partial least square regression." In Proceedings of the 3rd ACM international workshop on Audio/visual emotion challenge, pp. 21-30, ACM, 2013