A machine learning based image and audio system for depression detection among students for counseling

DEEP JINDAL[1], NIGEL CRASTO[2], NOEL JOE[3], PROF. IMRAN ALI MIRZA[4]

[1,2,3,4]Department of computer Engineering, Don Bosco Institute of Technology, Mumbai, India.

[1]deephiteshjindal@gmail.com, [2]nigelcrasto12@gmail.com, [3]noeljoe643@gmail.com,

[4]imran.dbit@dbclmumbai.org

Abstract: The project hinges on several domains of computational technology i.e., Machine learning, Deep learning neural networks and Artificial Intelligence. The aim of the project is to invent a product that is helpful to prevent depression and other mental disorders using domains of computer technology. The system used in this project is waterfall model wherein we follow a sequential pattern of activities which includes the following: 1. We get a clear thought of our requirements for the project and assemble the parts for it. Aspects of deep learning such as convolutional neural network (CNN) was implemented to produce an outcome of highest probable mental disorders such as anxiety and depression. This was established through audio and visual analyzer, further integrated on a web application for better access. 2. Create a software architecture using TensorFlow to tentatively diagnose emotions and track behavioural variations. 3. Implementing the aforementioned architecture in the blueprint of the project by coding. 4. Run the code on test cases and check for bugs to make sure the product is running to the user's satisfaction. 5. Periodic maintenance and easy accessibility of the web application on which both models are integrated. Ahead in this paper, literature survey showed previously published studies that used computational advancements for predicting mental health status. Our study achieved a 75% accuracy at establishing a mental health condition, in concordance with the aforementioned studies.

Keywords: Machine Learning (ML), Internet of Things (IOT), Convolutional Neural Networks (CNN), Artificial Intelligence, Face and Object Detection

I. INTRODUCTION

Depression is commonly prevalent among adults across the globe. According to WHO, about 300 million people suffer from mental illness. Furthermore, in 1998 estimated an increase in depression burden in developed and developing countries (WHO, 2017). Depression is one of the leading causes of neuropsychiatric diseases. Nowadays, people have been facing an earlier onset of depression due to lifestyle changes. Dealing with mental health is a major challenge these days. Mental illnesses affect relations with peers and family. Often depression is characterized by ongoing feeling of sadness and aloofness to sometimes a pathological suicide. Many times, the sufferer does not reach out for counselling due to fear of societal judgements and tries to cope with mental illnesses alone. Research suggests that early prognosis of depression can better mental health[3](Duffy2000). How effectively can they overcome mental illness by reaching for the right kind of help is still unknown to them. Keeping in mind the difficulties people face, their ability to overcome and ease the fear related to mental disorders, we must employ technology for prognostication.

Our study focuses on silent sufferers of mental illnesses like depression, anxiety, panic and stress. The main objective is to make it easier for them to reach out for help and be counselled, which will not only help them but make it easier for therapists and psychiatrists to monitor them and provide required support. Mental illness such as depression and anxiety are merely biochemical processes and can often be completely cured through right guidance and treatment.

1.Depression: Depression (major depressive disorder) is a common and serious medical illness characterized by aloof behaviour, low self-esteem and lack of motivation to achieve a goal.

2. Anxiety: Anxiety is your body's natural response to stress. It's a feeling of fear or apprehension about forth coming life situations or uncertainties. It is marked by fast heart rate, rapid breathing, sweating and fatigued state.

Recent studies have shown computational technology can be used to prognosticate mental disorders (refer). In our study, we wanted to implement a model which would turn out to be beneficial to our user (i.e., the person suffering one of the various mental illnesses) by not only helping them communicate but also help in detection of depression

 DOI : 10.35291/2454-9150.2022.0417

through their distinctive behavioural and cognitive patterns in daily activities. With the help of our initiative using facial expression detection and activity tracking, it can also guide them in aspects such as "How can one reach for his/her required source of help?" such as a counsellor or psychiatrist. The application would help counsellors in order to detect mental illnesses quicker and easier. The outcome of our project is based on three of the widely discussed topics that are the backbone of computer technology and have a plethora of applications in the real world:

1. Convolutional Neural Networks (CNN): First appeared in the 20th century, these networks are a pack of layers that filter images using feature extraction techniques [5] (Fukushima et al 1980). Convolutional layers convolve the input and pass its result to the next layer and so on until the output layer is reached. The product will take an input from the user and convolve it for a better prediction without losing distinctive characteristics of the image.

2. Machine Learning (ML): In machine learning, we enable the computing machine to learn aspects related to the real world. The machine will eventually be able to detect and predict the outcome of a particular expression

or text. The product with the output image will predict result for the user.

3.IOT (Internet of Things): Internet of things is a trend in the modern technological world and a boon to society. Nowadays, various aspects such as wireless network connectivity and portable processors drive our world. Our project consists of two modules that will be effectively used in processing data, capturing images, displaying location of the person, tracking prosodic variations. This domain proves as a good link to the aforementioned domains- CNN and ML.

II. LITERATURE SURVEY

A. Problem Statement and Objective:

Currently, detection of depression is based on Beck's Depression Inventory (BDI), self-disclosed details and counsellor's clinical experience [7] (Olaya Contreras et al 2010). However, the prognosis is highly dependent on doctor-patient relation and former's proficiency, making the diagnosis subjective. Consequently, resulting in misdiagnosis and thereby increased risk of major depression and mortality [4] (Fogel et al 2006). A study conducted by [6] Li et al (2016) showed that judgements over mental illness often makes depression a social problem. Owing to the need for early and accurate diagnosis as well as social problems, automatic detection of depression via computing are helpful.

With the advancements and wide application of computing and ML, there are screening tools to detect whether a patient has depressive symptoms, which serve as primary care. Our study introduces a combined solution system that will serve as a solution to the increased burden of mental illness. Through two models: audio visual analyzer constituting our system architecture, we examine the sufferer. Two metrics have been applied face detection and prosodic change recognition for emotion detection application. Additionally, through our application we hope to ease communicational and diagnostic tasks. Our project domain is machine learning and we have decided to make use of the concept of a CNN built using TensorFlow to process image data and perform the recognition task after being trained with datasets.

B. Survey of Existing Systems:

Application of Deep neural networks (DNN) continued to exponentially increase with advancements and developments in architectural design structure of DNN (Bachman et al 2017; Roy et al 2019). Previously,[10] Cohn et al (2009) used automated facial image analysis and audio signal processing, compared the former analysis to manual FACS annotation for depression detection. Their results had high accuracy of 88% and 79% for FACS coding and audio analysis respectively. Thus, suggesting usage of ML in early detection of mental disorders. There have been wide variety of studies in successful clinical prognosis of emotions using deep learning approach, CNN- which is the state-of-the-art in ML. Long back, the Big Five Model studied images to establish personality attributes (McCrae and John 1992). Since the appearance of CNN in 1980, there have been several studies that have employed CNN based networks for medical applications, people detection, image and audio detection, and acoustic classification [1] (Deng et al 2013; Golik, Schutler ad Ney 2015). Recently, [8] Namboodiri and Venkataraman (2018) used an automatic system inclusive of Gabor filter and SVM classifier to successfully detect depression using images. The study was based on a large scale dataset of student's face analysis to classify facial images through the system which yielded an accuracy of 64.38%. The severity of depression was measured by the extraction of negative emotion captured in a video by Gabor filter. A population-based study implemented a combined visual-audio model that collected data based on responses to mood stimulus. Through a 3D facial report processed by Kinect and 2D frontal face image capture by optical camera, the level and risk of depression was measured. They showed a higher detection of depression risks on the basis of optimistic and pessimistic stimulus in females compared to male as well as established the better functioning of combined system architecture over 2D or 3D alone [9] (Guo et al 2019). Further, to study time series of learned facial expressions over stimulus-generated emotion to detect depression, they employed the long short-term memory (LSTM) model. Two module LSTM model, analyzed timeline of static learned facial expression and facial motion by 2D images and 3D EUs and point respectively [9] (Guo et al 2021).

 DOI : 10.35291/2454-9150.2022.0417

The latter showed exceedingly better accuracy at identifying depression.

An analysis by [11] Guntuku et al (2019) on social media platforms such as twitter revealed that a language prediction algorithm for survey-reported depression and anxiety when applied on two datasets followed by feature extraction determined user's depression or anxiety and demographics which improves prediction performance of depression. Previously, a study was conducted on online depression communities where data was collected from on Weibo and Twitter.

They showed depression among age groups by semantic analysis, motion detection and four classifiers of text. The results concluded that linguistic method analysis reduced inaccuracy of emotion detection on online sites. The limitation of these studies is the negligence of possible situations at the time of the text, therefore leading to consequential action intermixing. Action intermixing can be avoided in the field of image processing as it brings the possibility of time series capture of image and video via 2D and 3D processors. In the study by De and Saha (2015), they demonstrated five different methods to recognise human emotion. In a study, these human emotions were taken into consideration based on the muscle involvement and analysed using sub-image feature, further classified using algorithms like AdaBoost hidden Markov model and support vector machine [13] (Kudiri, Said and Nayan 2012). An analysis by [14] Chen and Cheng (2015) based on an edge detection algorithm allowed image preprocessing, image recognition processing to locate the eyes and lips, individually marked and extract the edge shape feature. Additionally, eventually the system was trained by using a face database to achieve other face expression identification. Some facial expressions can be easily and readily processed via a novel DNN architecture for face expression recognition (FER) [12] (Mallohasseine, Chan and Mahoor 2016). They examined the neural network function using MultiPIE and FERA to perform cross database classification while training on databases that have limited scope for FER. A meta-analysis of images by [15] Jaiswal, Raju and Deb (2020) demonstrated a method to successfully identify seven emotions such as anger, disgust, neutral fear, happy, sad, and surprise using deep leading techniques and image processing.

C. Scope:

By using this application, we intend to provide a very basic requirement to detect depression in people. We also aim to guide them towards legitimate counselling centres and experts who can actually help them and prevent the disease from worsening. All this will be done with the help of booming computation technology and resources.

III. MODELS AND SETUPS

Project Architecture

Using the modified approach to emotion detection as [16] Gopika et al (2021), we employed a system implementing two convolutional neural network (CNN) in detection of emotions such as anxious, depressed or healthy, through two models: video analyser and voice analyser. Both models are easily implemented on a webpage (See Fig 1). The objective of this system is to produce accurate prognostication of depression and anxiety, as well as establish a mental health status in order to take measures to better mental health if needed. We will use the behaviour dataset and self-disclosed persona attributes encoding video-level analysis to build a CNN classifier of mental disorders. The video-level behavioural dataset has 3 labels: facial expression units (EUs), head and eye movement. Histograms and graph spectra represented features of both modules. The voice analysis processed the suprasegmental changes in phonetics, thus reflecting mental health status. The 5- and 18-layered architecture of CNN converted the analysis from video and voice module respectively, into lower dimension without losing characteristic features, thus produced an output of highest emotion prospect.

A. User Authentication

The first stage will authenticate the user ID in order to make an account for a user for which the collected data will be fed in the database for future references and other activities of the client. The authentication takes place through various modes such as email id verification, Aadhar card verification etc. after which it is fed into the machine learning module (See Fig. 2).

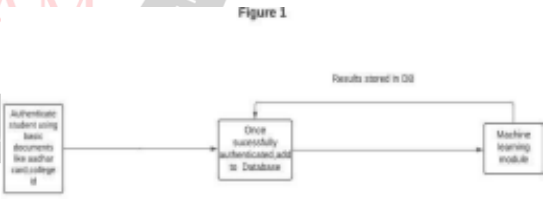

Fig 1: Project architectural design.

B. Voice Analyzer:

Using an identifier, the emotion behind the audio-file was analysed. From the collected datasets emotions are classified into categories like happy, sad, afraid and angry. To carry out this analysis of emotion and extraction of speech a Python tool Librosa was used. This tool performs audio and music analysis. Librosa tools are often used to gain MFCC features (Mel Frequency Cepstral Coefficient). Spectrograms play a vital role in representing the data and picture the power of the given signal. Several audio files were taken under consideration. The .zip format of audio files was chosen to make analysis and extraction of emotion easier. These audio files are then tested for functionality and feature to note the wavelengths and plot the graph.

 DOI : 10.35291/2454-9150.2022.0417

Audio files were differentiated on several factors. The basic classification between female and male voices were performed by identifiers. Each factor considered is tagged with a given value. The data model is trained after segregation into train and test groups. The most suitable choice for classification used is CNN as it portrays the highest efficiency.

C. Image and Video Classifier

The next stage is image analysis, images of students are taken from a database and face detection is done using ROI or other algorithms and feature extraction is done. Testing is done through video recordings, frames are extracted and converted to images and then emotion recognition is completed. These videos and images were captured from the feed of the user's camera. In the case of images, the faces in the frame from the feed were spotted using Machine Learning techniques. In case of videos, breakdown of the video into several individual frames takes place. Each image from these frames is put through the Machine Learning model. Image classification is done through the Haar Cascade method (See Fig. 2). This Machine Learning model helps the cascade function in filtering the positive and negative images from a large dataset. The Haar cascade model is trained for automated facial detection. It modifies the captured image to required criteria before processing for prediction. The image is then classified into one of the seven stated emotional categories. This classification is done on the basis of highest match probability with one of the seven expressions. After this, the image is processed through the CNN and the result with highest expression match probability is displayed.

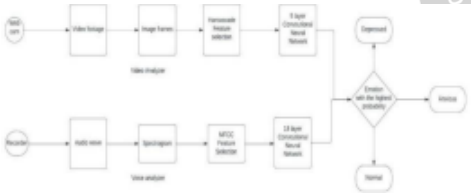

Fig 2: Classifier

D. Dataset

The FER (Facial Expression Recognition) dataset is used for analysis through facial expressions. The dataset was obtained from Keggle. For FER, the data was collected in the form of grayscale images. These expressions were categorised into seven categories namely Angry, Disgust, Happy, Sad, Fear, Surprise and Neutral. The data contains the train.csv model. This model is categorised into emotion and pixels (See Fig. 3). The emotional category consists of the seven emotions. These are numbered from 0-6. The test.csv model is trained to hold the pixel value and match with the emotional value column.

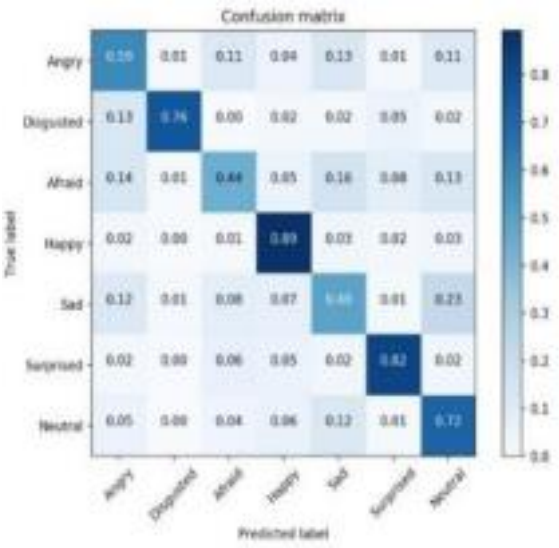

Fig 3: FER database

The speech recognition is done using the RAVDESS (Ryerson Audio-Visual Database of Emotional Speech and Song) database. With deep learning techniques, voice extraction and classification, emotion recognition and several other parameters are classified in RAVDESS.

IV. EXPECTED OUTCOMES

Depression has become a drastically increasing medical issue over the recent years. Though it cannot be completely avoided, early detection can help in prevention and control of mental health. Many doctors and counsellors are trying to spread awareness and help people open up and fight their depression. Major parts of society dealing with this illness fail to detect it in time and fall prey to serious consequences. Thus, with the help of advancing technology in all fields we have made this project. Machine learning, Deep learning, Neural networks and several such models have been put together in order to detect depression and cause a positive change in the society. On integrating these models, we aim to make depression detection easier and available to a larger section of society. The project turns out to be challenging in several ways in terms of technology and adaptation of this concept for the people. The research can be applied in early detection of this illness and prevention of it. The aim of this project lies in benefiting the society and the future scope of a fully functional and operational depression detection software for creating a positive impact on society and saving a life.

We expect to build a fully functional app which has three major modules: Authentication, audio-visual analyzer resulting in a CNN outcome and classifier. First, we authenticate users who want to use the application using standard authentication procedures. After successful authentication we import the data using a scraping technique from websites like Twitter, Reddit, Quora where people are the most vulnerable. We then clean the data i.e., remove the character repetitions, remove stop

 DOI : 10.35291/2454-9150.2022.0417

words, negation etc. and classify them into positive, negative texts using the Naive Bayes Classifier. We enter the images into the app and after ROI location, feature extraction and emotion recognition we classify them as depressed or not. We used an interactive questionnaire like PHQ-9,GAD etc.. On effective recognition of depression, the college counsellor is alerted.

V. CONCLUSION

Depression has become a drastically increasing medical issue over the recent years. Though it cannot be completely avoided, early detection can help in prevention and control of mental health.Many doctors and counsellors are trying to spread awareness and help people open up and fight their depression.Major part of society dealing with this illness fail to detect it in time and fall prey to serious consequences. Thus with the help of advancing technology in all fields we have made this project. Machine learning, Deep learning, Neural networks and several such models have been put together in order to detect depression and cause a positive change in the society. Integrating these models we aim to make depression detection easier and available to a larger section of society. The project turns out to be challenging in several ways in terms of technology and adaptation of this concept for the people. The research can be applied in early detection of this illness and prevention of it. The aim of this project lies in benefiting the society and the future scope of a fully functional and operational depression detection software for creating a positive impact on society and saving a life.

REFERENCES

[1] Deng, L.; Li, J.; Huang, J.T.; Yao, K.; Yu, D.; Seide, F.;Seltzer, M.; Zweig, G.; He, X.; Williams, J.; et al. (2013). Recent advances in deep learning for speech research atMicrosoft. In Proceedings of the 2013 IEEE International Conference on Acoustics, Speech, and Signal Processing(ICASSP 2013), Vancouver, BC; pp. 8604–8608.

[2] Golik, P.; Tüske, Z.; Schlüter, R.; Ney, H. (2015). Convolutional neural networks for acoustic modelling of raw time signal in LVCSR. Proceedings of the SixteenthAnnual Conference of the International Speech Communication Association, Dresden, Germany, 6–10 September.

[3] Duffy A. (2000). Toward effective early intervention and prevention strategies for major affective disorders: A review of antecedents and risk factors. Can J Psychiatry; 45:340–8.

[4] Fogel, J., Eaton, W., and Ford, D. (2006). Minor depression as a predictor of the first onset of major depressive disorder over a 15−year follow−up. Acta Psychiatr. Scand. 113, 36–43. doi: 10.1111/j.1600-0447.2005.00654.x

[5] Fukushima, K. (1980). Neocognitron: A self organising neural network model for a mechanism of pattern recognition unaffected by shift in position. Biol. Cybern,36, 193–202.

[6] Li, X., Jing, Z., Hu, B., Zhu, J., Zhong, N., Li, M., et al. (2017). A resting-state brain functional network study inMDD based on minimum spanning tree analysis and hierarchical clustering. Complexity 2017:9514369.

[7] Olaya-Contreras, P., Persson, T., and Styf, J. (2010). Comparison between the beck depression inventory and psychiatric evaluation of distress in patients on long term sick leave due to chronic musculoskeletal pain. Journal of multidisciplinary healthcare, 3:161.

[8] Namboodiri Sandhya Parameswaran,D Venkatraman (2015). A computer vision based image processing for depression detection among students for counselling.

[9] Weitong Guo,Hongwu Yang,Zhenyu Liu,Yaping Xu,Bin Hu(2019). Deep neural networks for depression recognition based on 2D and 3D Facial expressions under emotional stimulus tasks.

[10] Jeffrey F. Cohn,Tomas Simon Kruez , Iain Matthews , Ying Yang,, Minh Hoai Nguyen,Margara Tejera Padilla2 ,Feng Zhou , and Fernando De la Torre (2009). Detecting Depression from facial actions and vocal prosody.

[11] Sharath Chandra Guntuku, Daniel Preotiuc Pietro, Johannes C. Eichstaedt, Lyle H. Ungar (2019). What Twitter profile and posted images reveal about depression and anxiety.

[12] Ali Mollahosseini, David Chan , and Mohammad H. Mahoor (2016). Going deeper in facial recognition using deep neural networks.

[13] Krishna Mohan Kudiri, Abas Md Said, M Yunus Nayan(2012). Emotion Detection Using Sub-Image based features human facial expressions.

[14] Xiaoming CHEN and Wushan CHENG (2015). FacialExpression Recognition based on edge detection.

[15] Akriti Jaiswal, A. Krishnama Raju, Suman Deb (2020).Facial emotion detection using deep learning

[16].A Robust Emotion Extraction System from EEG Signal Dataset using Machine Learning Muthulakshmi P and Gopika R. Volume 3, Issue 2, (2021)

 DOI : 10.35291/2454-9150.2022.0417