International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 08 | Aug 2020 www.irjet.net p-ISSN: 2395-0072

Emotion Recognition and Depression Detection using Deep Learning

Neha S[1], Nivya[2], Pooja HC Shekar[3], Keerthana S Kumar[4], Asha VG[5]

[1,2,3,4] Dept.of Computer Science, Jyothy Institute of Technology, Bangalore, India
[5]Assistant Professor, Dept.of Computer Science, Jyothy Institute of Technology, Bangalore, India

---

Abstract - *Emotion recognition systems based on facial gesture enable real-time analysis, tagging, and inference of cognitive affective states from a video recording of the face. It is assumed that facial expressions are triggered for a time period when an emotion is experienced, and so emotion detection can be achieved by detecting the facial expression related to it. Out of all the major 6 emotions present, depression plays a vital role. Depression is classified as a mood disorder. It may be described as feelings of sadness, anger or loss that interfere with a person's everyday activities. People experience depression in different ways. In certain cases, depression may lead to fatal cases. In order to avoid all of these, depression must be detected at the earliest and victim must be treated with appropriate remedies. The objective of the project is to analyse the emotion of a user using real-time video. This is achieved using Convolutional Neural Networks [CNN]. If the emotion is analysed as depression, then it has to be treated at the early stages. As the symptoms worsen, the mental ability of an individual goes out of control which leads to a disorder. If the emotion is analysed as to be depression, then a chatbot pop-up appears on the screen where the user can share his/her feelings with the chatbot achieved using Tkinter library. This helps to boost up the user's mood, analyse the level of depression and to help the user come out of this mood. If the user's emotion is found to be sad, then a continuous evaluation is done in order to classify between sadness and depression.*

***Key Words*: Emotion Recognition, Depression, Convolutional Neural Networks, Tkinter**

1. INTRODUCTION

1.1 Depression:

Emotions tend to rule our daily lives. Throughout life, our emotions influence the choices that we make.

According to the book "Discovering Psychology" by Don Hockenbury and Sandra E. Hockenbury, an emotion is a complex psychological state that involves three distinct components: a subjective experience, a physiological response, and a behavioral or expressive response. In simple terms emotions involve things other than just feelings. It is the way that someone experiences an emotion. It involves bodily reactions, like the racing of heart because of excitement. It also involves expressive movements, like facial expressions and sounds—for example, when you say "woah" because you are fascinated by something. Emotions further involve behaviors, like yelling at someone when you are angry. There are 6 basic emotions that a person experiences- happy, sad, fear, disgust, anger, and surprise.

Depression is a disorder of impaired emotion regulation. It is an abnormal emotional state that affects our thinking, perceptions, and behavior in pervasive and chronic ways. Depression does not necessarily branch out from a difficult or challenging situation, a loss, or a change of circumstance as a trigger. In fact, it often occurs in the absence of any such triggers. Depression is a common illness worldwide with 264 million people affected by it. When its long-lasting and with moderate or severe intensity, depression may lead to serious health condition. People tend to suffer a lot, and this creates an impact in their lives. They function poorly at work, school and at home. At its worst, depression can lead to suicide. Depending upon the severity of symptoms and its number, a depressive episode can be categorized as mild, moderate, or severe.

Mental health professionals use the American Psychiatric Association Diagnostic and Statistical Manual of Mental Disorders (DSM-5 criteria) to help determine if someone is sad or depressed. At least 5 of the following symptoms should hold good for a period of at least 2-weeks and a minimum of 1 of the symptoms must be diminished interest/pleasure or depressed mood. The DSM-5 criteria include nine potential symptoms of depression:

- lack of interest and enjoyment in activities you used to find pleasurable
- trouble sleeping, or sleeping too much
- trouble eating, or eating too much, coupled with weight gain or weight loss
- irritability, restlessness, or agitation
- extreme fatigue
- unwarranted or exaggerated feelings of guilt or worthlessness
- inability to concentrate or make decisions
- suicidal thoughts or actions, or thinking a lot about death and dying

1.2 Convolutional Neural Network

Deep learning based facial expression recognition is one of these methods to detect emotion state (e.g., anger, fear, neutral, happiness, disgust, sadness and surprise) of human. This method aims to detect facial expressions automatically to identify emotional state with high accuracy. In this method, labeled facial images from facial

expression dataset are sent to CNN and CNN is trained by these images. Then, proposed CNN model decides which facial expression is detected.

A Convolutional Neural Network (CNN) is a Deep Learning algorithm which can take in an input image, assign importance in the form of learnable weights and biases to various aspects/objects in the image and be able to differentiate one from the other. The pre-processing required in a CNN is much lower as compared to other classification algorithms. CNNs work better for image recognition tasks since they can capture spatial features of the inputs due to their large number of filters.

There are various techniques that can be kept in mind while building a deep neural network and is applicable in most of the computer vision problems. Below are few of those techniques which are used while training the CNN model below.

1.Data Augmentation: More data is generated using the training set by applying transformations. It is required if the training set is not enough to learn representation. The image data is generated by transforming the actual training images by rotation, crop, shifts, shear, zoom, flip, reflection, normalization etc.

2. Kernel Regularizer: It allows to apply penalties on layer parameters during optimization. These penalties are incorporated in the loss function that the network optimizes. Argument in convolution layer is nothing but L2 regularization of the weights. This penalizes peaky weights and makes sure that all the inputs are considered.

3. Batch Normalization: It normalizes the activation of the previous layer at each batch, i.e. applies a transformation that maintains the mean activation close to 0 and the activation standard deviation close to 1. It addresses the problem of internal covariate shift. It also acts as a regularizer, in some cases eliminating the need for Dropout. It helps in speeding up the training process.

4. Global Average Pooling: It reduces each feature map into a scalar value by taking the average over all elements in the feature map. The average operation forces the network to extract global features from the input image.

4. Depth wise Separable Convolution: These convolutions are composed of two different layers: depth-wise convolutions and point-wise convolutions. Depth-wise separable convolutions reduce the computation with respect to the standard convolutions by reducing the number of parameters.

1.3 Chatbot

A chatbot is an artificial intelligence software that can house a conversation with a user in natural language through messaging applications, websites, mobile apps or through the telephone. A chatbot is one of the most advanced and promising expressions of interaction between humans and machines.

A personal companion chatbot is used to adhere to the personal needs of the user. This can be done in many ways like setting up reminders according to the schedule of the user, doing different task over a network according to the needs of the user or simply helping a user to share his feelings with it and acting accordingly.

A companion chatbot that is button based is implemented in this paper where a set of questions that contributes in detecting depression is presented in front of the user with some options attached to it in form of radio button. The answers of the user are recorded and analysed when the user chooses the appropriate button as the answer. The chatbot makes recommendation to the user in order to make him feel better after analysing his answers.

The system focuses on recognizing sadness which plays a critical role in devising our results. On detection of sadness, a chatbot, which uses a questionnaire on the lines of DSM-5 criteria symptoms, is displayed to the user. The answers of the users are taken into consideration to detect the presence of the any of the symptoms of depression or general sadness and a conclusion is drawn to state whether clinical depression exists. The chatbot suggests helpline numbers in case acute depression is inferred or make personal recommendations to uplift an individual's mood in case a general negative mood is inferred, like momentary sadness.

2. LITERATURE SURVEY

This paper was written by Aliaa A. A. Youssif, Wesam A. A. Asker which presents a computer vision system for automatic facial expression recognition (AFER). There are three major steps in AFER, the first step being the detection of the face in the scene. The second step is to extract the facial features that showing the facial expression and the third step is to classify the facial display shown on the face. The face detection uses the open source code library (OpenCV) that employs a face detection algorithm based on Viola & Jones features. Then, the Facial Features Extraction is done where the segmentation process is performed first to divide the face image into three areas: mouth, nose and two eyes and two eyebrows. Second, the facial characteristic points (FCPs) are in each face component using mouth, nose, eyes and eyebrows FCPs extraction techniques. The feature extraction process is applied to face image to produce a feature vector that consists of two types of features: geometric features and appearance features which shows a pattern for facial expression classes. After this, the feature vector is given as an input into the radial basis function artificial neural network to recognize the facial expressions The results show that the AFER system classifies the facial expressions accurately with recognition rates between 90% and 99% in a person-dependent dataset and between 83% and 100% in a person- independent dataset.[1]

The authors Enrique Correa, Arnoud Jonker, Michael Ozo and Rob Stolk proposed their paper of emotion recognition using Convolutional Neural Network. This method includes a few hundred high resolution photos to tens of thousands smaller images. In order to increase the accuracy of the emotions detected the size of the training dataset must be increased from 9000 images to 20000 images from FERC. The results obtained are compared with other methods such as SVM and LVQ. It produces an accuracy of 90% happy,80% neutral and 77% surprised. [2]

The authors Kartika Candra Kirana, Slamet Wibawanto and, Heru Wahyu Herwanto, in their paper proposed emotion detection using Viola Jones Algorithm. Though Viola Jones is commonly used for face detection, here Viola Jones algorithm is used for both face detection and emotion recognition. Rectangular feature and cascading AdaBoost algorithm are applied as the main concept of the Viola-Jones Algorithm in both the processes. These processes use Russel's Circumplex to classify the emotions as this has a better efficiency in classifying the emotions. This method consists of 3 stages: initially an image is captured from a video; the unwanted rectangular areas are deleted and then the emotion in the picture is recognized. The prediction provided an accuracy of 74%.[3]

The authors Aafiya Shaikh, Dipti More, Ruchika Puttoo, Sayli Shrivastav proposed a model of chatbot which works as an android application. The user has to login to the application using email and a password. The details are being used for user authentication purpose. Once the user logs in, the real processing of the data takes place on the server. The input is taken from the user, sent it to the server for processing using Recurrent Neural Network (RNN). RNN uses encoding and decoding mechanism for implementing a chatbot. The architecture is an end to end mechanism and based on that a decoder is being built. The attention mechanism is also implemented where it the essential characteristics of the sentences are being extracted.[4]

3. REQUIREMENTS SPECIFICATION

3.1 Hardware Requirements

Processor: Any Processor above 500 MHz
Ram: 8 GB
Hard Disk: 4 GB
Dedicated Graphics Card
Input device: Standard Keyboard, Mouse and Camera.
Output device: High Resolution Monitor.

3.2 Software Requirements:

Operating System: Windows 7 or higher
Programming language: Python and related libraries.
Software: Anaconda Version 3.6
Dataset is obtained from Kaggle. The data consists of 48x48 pixel grayscale images of faces.
The training set consists of 28,709 examples. The public test set used consists of 3,589 examples. The final test set consists of another 3,589 examples.

3.3 Functional Requirements

In the functional requirements, we focus on documenting the operations and activities that our project is supposed to perform, and they include the following.
R1: The user initially faces the camera and emotion is being recognized.
R2: The user's emotion is analysed using a count variable.
R3: If the user is happy then the process is aborted.
R4: If the user is found to have an emotion of sad or fear then a chatbot like application called deskbot pops up.
R5: If the user is ready to take up the test then a series of questions appear to analyse the mood of the user and guide him accordingly.

3.4 Non-Functional Requirements

In terms of non-functional requirements, we should be mainly focusing on the performance requirements of our system.
S1: The model should be trained with a better accuracy dataset in order to classify among the various emotions.
S2: The quality of camera should be good so that it captures the video at a better quality.
S3: The processor should be of i5 or more, else the video capturing will not happen.

4. IMPLEMENTATION

4.1 Convolutional Neural Network

The architecture of a ConvNet is analogous to that of the connectivity pattern of Neurons in the Human Brain and was inspired by the organization of the Visual Cortex. The Convolutional Layer and the Pooling Layer together form the i-th layer of a Convolutional Neural Network. Depending on the complexities in the images, the number of such layers may be increased for capturing low levels details even further, but at the cost of more computational power.

1. Facial Expression dataset:
There are many open accesses facial expression dataset in literature. We have used dataset for facial expression from Kaggle and the data has 48x48 pixel grayscale images of faces. The training set consists of 28,709 examples with seven emotions (happy, sad, surprised, fearful, angry, disgusted, and neutral).
2. Image Preprocessing:
The face circumference was detected using the Haar Cascade library from the pictures. Then, these detected rectangular facial expressions were clipped and recorded. Also, the images were converted to gray images and was

placed in neural networks. This process was done to avoid unnecessary density in the neural networks.

3. Convolutional neural network architecture:

The proposed CNN architecture is aimed to educate the pixel values in the rectangular region containing facial expressions. This occurs in 3 stages after which it is fed into the fully connected layers. The CNN structure consists facial expression data and includes the 3 stages each of which has 2 convolutional layers with 'relu' activation function followed by max-pooling layers, and 3 fully connected layers with 'relu' and softmax activation function. After all operations of convolutional layers and max-pooling layers, each frame feeds to the fully connected layers and prediction of frames was processed with classifier as seven different facial emotional state.

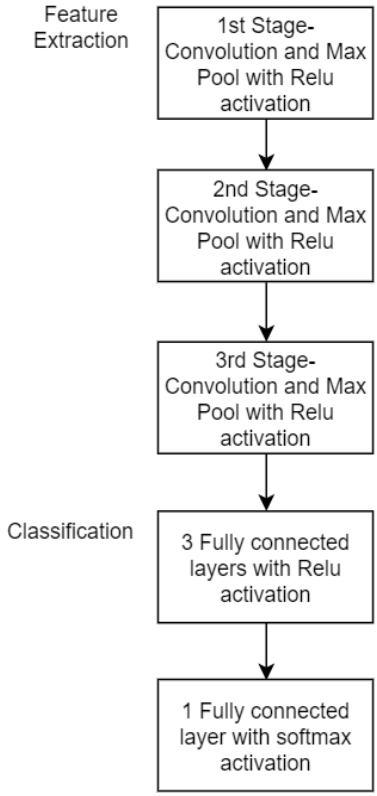

Fig. 4.1.1 - CNN Model Architecture- The layers used in the construction of the CNN Model

4. Network training:

The neural networks were implemented using Keras with a TensorFlow backend running in Python. The model was trained for 50 epochs.

5. Real time testing:

After the training of proposed CNN architecture, the trained model was tested in real time. First, human faces were detected by the computer camera using Haar Cascade library. After that, the detected images were sent to the model and the classes they belong to were queried. As a result of the predictions, the possibility of belonging to which class the facial expression was shown on the webcam screen.

OpenCV is used to draw a rectangular boundary around the face detected and the emotions recognised is displayed on the screen with an emoticon indicator on the window with the percentage confidence of that emotion.

Table 4.1.1 - Modules used

Modules used	Functionality
Data Pre-processing	The data required to train and test the model is processed and divided into training and test dataset. The images are reshaped for keras model.
Model definition	The CNN Model code is defined using the required ML Libraries with 3 fully formed layers after the operations of 3 convolutional and 3 max pooling layers.
Real time Emotion Recognition	The real time classification of emotion is done with the help of OpenCV to access video with a webcam and the emotion recognition classification is shown on the screen

4.2 ChatBot

A chatbot is a computer program that simulates human conversation through voice commands or text chats or both. The type of chatbot which is implemented is a button-based chatbot. A button based chatbot has a set of button-like options where the user can click on any one of the options which is suitable and reply with their answer.

Building a chatbot:

The button based chatbot is implemented using Tkinter and other python functions. Tkinter is one of the most frequently used GUI toolkits. There are various python libraries which are used to implement radio-buttons, buttons, and text boxes. The questions which are displayed in a chatbot are implemented using textboxes. Each question has 4 options given where those are radio buttons. A user can select any option and reply which are used for analysis.

The list of various functions used in the code is as follows:

Table 4.2.1- Functions used

Function Name	Description
ch(),des(), des1()	Destruction of windows
window()	Creation of window and

	setting orientation of the window
I4(), func2(), func1(), func()	Functions to collect basic information of user
page2(), page3(), sad(), dep(), dep1()	Function to display different questions based on the analysis made by the system
last()	Function to record the feelings of the user

Analysis:

The system detects symptoms of depression for a user by relying on a variable, score that is hidden to the user. The value of the score variable determines the next set of questions that needs to be presented to the user and thus helps in making prediction that tells the emotion of the user which is one amongst the two categories- sad or depressed. Each option associated with a particular question is governed by a number that contributes towards the score variable. The first page of questionnaire consists of few basic questions. A score of 16 leads the user to take up a depression test. If the user gets a score of 16 or above in it, depression is detected, and appropriate helpline number is displayed for further help. A score lesser than 16 leads to a result that says the user is sad, and the user is made recommendation to uplift his mood. On the other hand, a score of 4 on the basic questionnaire page leads to a conclusion of the user being happy.

Lastly any score between the range of 4 and 16 takes the user to a sadness test where the user is further directed forward into the sadness test or taken into the depression test based on his score value.

5. RESULTS AND DISCUSSION

We trained our Convolutional Neural Network model using Kaggle database which includes seven emotions (happiness, anger, sadness, disgust, neutral, fear and surprise) The detected face images are resized to 48×48 pixels, and converted to grayscale images then were used for inputs to the CNN model. We achieved an accuracy rate of 64% for the emotion recognition.

In this paper, we introduced a button-based desktop application to detects whether the user is depressed or not and this application employs a questionnaire which the user must answer. Based on the resulted answer the user is detected if he/she is depressed or not. The analysis of whether a person is momentary sadness or showing symptoms of depressions was given according to internals scores awarded to the answers of each question when tested with a number of different scenarios.

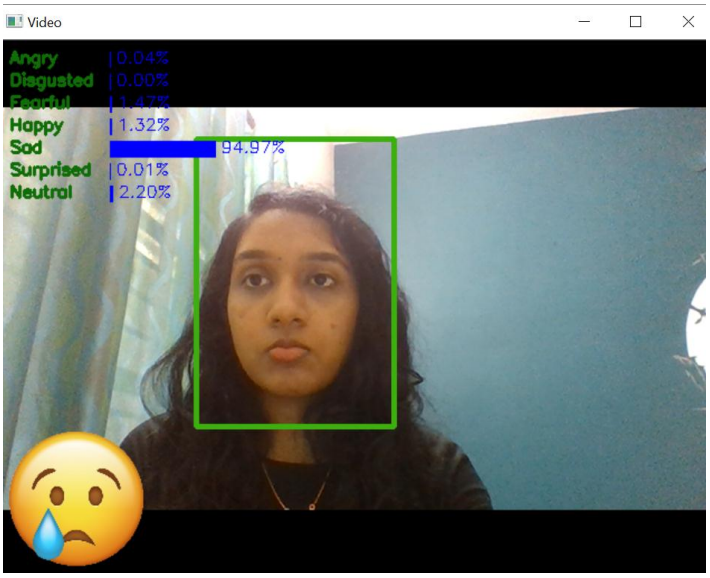

Fig 5.1- Emotion Recognition

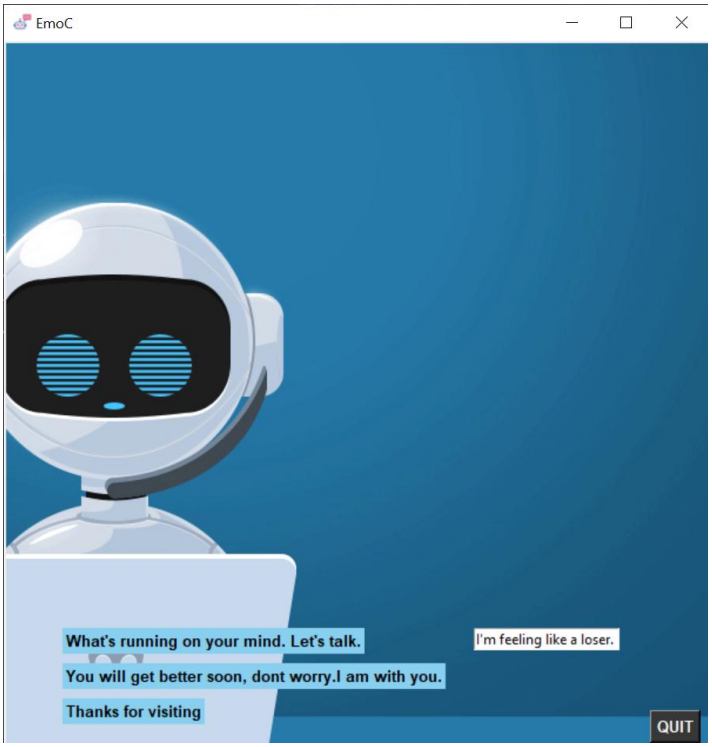

Fig 5.2- Chatbot implementation

6. CONCLUSION

An automated Facial Expression Recognition System has a wide range of applications in psychological research and human-computer interaction applications. The system plays a communicative role in interpersonal relations because they can reveal the affective state, cognitive activity, personality, intention, and psychological state of a person. The system has 3 modules-face detection that is implemented by Haar Cascade, emotion recognition which

is implemented by CNN using Keras that mainly focuses on detecting emotions that can reflect depression in an individual. Finally, the last module, a chatbot is used that is used to recognize depression that further helps to differentiate between sadness and depression.

Due to its efficiency or ease of implantation the above stated algorithms are selected for face detection and emotion recognition. A new approach by utilization of *DSM-5* criteria through a chatbot is used to analyze the symptoms of depression and thus conclude its presence.

7.FUTURE WORK

The project mainly deals with emotion recognition and a depression analyser which is implemented as a desktop application based chatbot. In the future trends this desktop application based chatbot can be actually implemented as a chatbot using Natural Language Processing.

As of the current working, the desktop application is local. In the future days, this application can be hosted on a website using an internet connection.

The current application is basically a screening test before consulting a doctor. In the future days, a video consultancy to doctor can be arranged if the user is detected to be depressed.

REFERENCES

[1] Automatic Facial Expression Recognition System Based on Geometric and Appearance Features- Aliaa A. A. Youssif, Wesam A. A. Asker

[2] Enrique Correa, Arnoud Jonker, Michael Ozo and Rob Stolk proposed their paper of emotion recognition using Convolutional Neural Network.

[3] Emotion Detection using Viola Jones Algorithm- Kartika Candra Kirana, Slamet Wibawanto and, Heru Wahyu Herwant

[4] A Survey Paper on Chatbots, Aafiya Shaikh, Dipti More, Ruchika Puttoo, Sayli Shrivastav, Swati Shinde

[5] Video based Emotion Recognition using Deeply Supervised Neural Networks by Yingruo Fan, Jacqueline C.K Lam, Victor O.K Li

[6] Dr. D.Venkataraman, Namboodri Sandhya Parameswaran, "Extraction of Facial Features for Depression Detection among Students",International Journal of Pure and Applied Mathematics,2018.

[7] Ma Xiaoxi Lin Weisi,Huang Dongyan,Dong MinGhui,Haizhou Li, "Facial Emotion Recognition",IEEE,2017.

[8] Facial expression recognition based on Local Binary Patterns: A comprehensive study by Caifeng Shana, Shaogang Gongb, Peter W. McOwanb

[9] Nicu Sebe, Michael S,Lew ,Ira Cohen, Ashutosh Garg, Thomas S Huang,"Emotion Recognition using Cauchy Naive Bayes Classifier",IEEE,2002.

[10] Kia-Biao He,Jing Wen,Bin Fang, "Adaboost algorithm using MB-LBP and skin color segmentation",IEEE,2011.

[11] Head Pose and Movement Analysis as an Indicator of Depression, Sharifa Alghowinem, Roland Goecke, Michael Wagner, Gordon Parker, Michael Breakspear

[12] E.M Bouhabba, A.A Shafie, R.Akmeliawati, "Support Vector Machine for Face Emotion Detection on real time basis",IEEE,2011.