International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 03 | Mar 2020 www.irjet.net p-ISSN: 2395-0072

A Survey Paper on Emotion Recognition and Depression Detection using Machine Learning Algorithms

Keerthana S Kumar[1], Neha S[2], Nivya[3], Pooja HC Shekar[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 period of time 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 survey is to analyse and predict depression at its early stages. As the symptoms worsen, the mental ability of an individual goes out of control which leads to a disorder. The proposed system concentrates on detection of a person's emotion and if the emotion is found to be sad a continuous evaluation is monitored in order to classify between sadness and depression. This improves detection of depression at early stages using a chatbot. This is achieved using MTCNN and Keras library.

***Key Words*: Emotion recognition, Depression, MTCNN, Keras.**

1. INTRODUCTION

Emotions plays a vital role in our existence. There are different types of emotions that influence the way we live and interact with others. We are at most times in our life ruled by our emotion that reflects through the choices we make, actions we take and the perception we have. Traditionally these different types of emotions we experience are divided into 6 categories namely happiness, sadness, disgust, fear, surprise and anger.

While most of us experience the emotions of happiness, surprise and fear in an ample ratio we also tend to experience the feeling of being sad, lonely or depressed at times. These are some of the normal reaction to losing things, life struggles or to an injured self-esteem. Prolonged period of these feelings can attribute to physical symptoms that prevent individuals from leading a normal and active life while providing indication towards depression.

Sadness is a normal human emotion. Sadness are generally triggered by events that are hurtful, disappointing or challenging. Depression 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 men, it manifests often as tiredness, irritability and anger. In women it's in the form of sadness, worthlessness, and guilt. Younger children exhibit school refusal, anxiety when separated from parents, and worry about parents dying. Depressed teenagers tend to be irritable, sulky, and get into trouble in school. They also frequently have co-morbid anxiety, eating disorders, or substance abuse. In older adults, depression may manifest more subtly as they tend to be less likely to admit to feelings of sadness or grief and medical illnesses.

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:

- feeling depressed throughout each day on most or all days
- 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

In many situations, individuals are ignorant of their disturbed mental conditions while in others, people are aware of them but hesitant to seek help due to the wrong

conceived notion of humiliation associated with depression.

The system proposed in this paper mainly consists of a face detection and emotion recognition system that detects the basic emotions stated previously by using MTCNN and Keras library respectively since face plays an important role in social communication and is the fastest means of communication while conveying any type of information. The system focuses on recognizing sadness which plays a critical role in devising our results. On detection of sadness a questionnaire on the lines of DSM-5 criteria symptoms is displayed to the user using a chatbot. The answers of the individuals are taken into consideration to detect the presence of the symptoms and a conclusion that's states whether clinical depression exists is inferred. Chatbot suggested helpline numbers in case of a positive result or make personal recommendation to uplift an individual's mood in case of a negative result, which is momentary sadness.

2. LITERATURE SURVEY

The authors Kai-Biao He, Jing Wen, Bin Fang proposed Adaboost Algorithm based on MB-LBP features and skin color segmentation for face detection. In this paper, face detection is based on Adaboost algorithm and extracted MB-LBP features are used to train Adaboost classifier instead of Haar-features. To reduce the false alarm rate, skin color is also combined with Adaboost. Firstly, a skin Gaussian model is built in Cg-Cr color space and some constraints are used to get the face of the candidates. Secondly, the correlation coefficient is calculated. The results of the experiment show whether the system has achieved the high detection rates or low false positives.[1]

This paper was written by Nicu Sebe, Michael S. Lew, Ira Cohen, Ashutosh Garg, Thomas S. Huang on emotion recognition using Cauchy naïve based classifier using Cauchy distribution as the model distribution as a method for recognizing emotions based on the facial expressions in video sequences. The Cauchy versus gaussian assumption experiments were carried out by taking a random example from the training set and carry out an initial classification. The Cauchy distribution proved better which was then used to classification of the testing set. Person dependent experiments and person independent experiments were performed which conclusively showed that the Cauchy based classifier provided significantly better results. The system gave an accuracy of 80%, if a person displays a negative or positive facial expression. And when the emotion recognition problem was changed to mood recognition problem, the classification results proved to be significantly higher.[2]

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.[3]

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. [4]

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%.[5]

Support Vector Machine for Face emotion detection on real time basis was written by the authors E.M. Bouhabba, A.A. Shafie and R.Akmeliawati. This paper deals with the real time method as the solution to the problem of facial classification in video sequences. An automatic facial feature tracker is employed to locate and extract the facial features. The displacements of the facial feature in the

video stream are taken as the inputs to a SVM classifier. The method is evaluated using the recognition accuracy for many interactions.[6]

Face detection and expression recognition using fuzzy rule interpolation written by Williams. D. Ofor, Nuka. D. Nwiabu, Daniel Matthias. The system comprises of face detection by Viola Jones algorithm, feature extraction by local binary pattern and expression recognition by fuzzy rule interpolation. The steps include face acquisition, face detection where Viola jones algorithm works in four steps- haar-like features, integral image, Adaboost and cascade classifier. Here a sub window of the input image is taken to detect the face by rescaling the images repeatedly, feature extraction- a particular feature from the image is extracted to categorize the expression into various category. LBP technique extracts the textual local info about the image and expression classification- here sparse situation is taken that is incapable of generating conclusions as the values of the observation does not match with the standard values used for comparison. The system uses Cohn Kanade database. The system gave back an accuracy of 71%.[7]

This paper written by M. Shamim Hossain and Ghulam Muhammad proposes a high-performance emotion recognition system for mobile applications. The Bandlet transform is applied to some selective frames, which are extracted from the video, to give some subband images. Local binary patterns (LBP) histogram is calculated from the subband images. This histogram describes the features of the frames. A Gaussian mixture model (GMM)based classifier is used as a classifier. The proposed emotion recognition system is evaluated using several databases. The contribution of this paper is as follows: (i) the use of the Bandlet transform in emotion recognition, (ii) the use of the Kruskal-Wallis (KW) and GMM classifier feature selection to reduce the time requirement during a test phase. Databases used are The JAFFE and CK database. The accuracy is 99.8% and 99.7% respectively. Predictions can be inaccurate for those patients who experience these tremors in later stages of the disease. [8]

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.[9]

Extraction of Facial Features for Depression Detection among Students written by Dr. D. Venkataraman and Namboodiri Sandhya Parameswaran aims at detecting depression in college students as in most cases these students are ignorant of their mental state. The paper takes two factors into consideration. One is the questionnaire conducted with the student, the other is the video of the student that has been taken of the student. Depression detection will be done by analyzing the amount of happy, contempt and disgust features throughout the video frames. The depression is detected by the amount of negativity in the video. Algorithms like Viola Jones, SVM and Gabor filters are used to achieve the goals.[10]

3. METHODOLOGY

Boosting based detector proposed by Viola and Jones showed real time performance was achieved by learning a sequence of simple Haar-like rectangle features. But since Haar like feature set is large and contains a mass of redundant information, this paper we present Multi Block Local Binary Pattern (MB-LBP). MB-LBP operator is defined by comparing the central rectangle's average intensity with those of its neighborhood rectangles. For a feature, calculate the sum of all the Positive samples and sum of all the negative samples. The skin color detection is done using the Cascaded Detector.[1]

In this paper, a simplified model where 12 features were used that shows the magnitude of the 12 Facial motion measurements defined and the total combination of the features give the seven basic facial expressions that must be classified. For each class of facial expression, the corresponding training set was considered and a Cauchy naïve Bayes model and a gaussian distribution model with the appropriate parameters were estimated using the ML framework. A random sample was taken, and classification was performed. The model which produced the best results in classification was taken to classify the rest of classes. There were 2 sets of experiments conducted. Person dependent experiment was conducted over 5 people, which consisted of 6 classes of facial expressions for each person and 1 expression was taken for training. Person independent experiment had sequences of one subject used as the test sequences and the sequences of the remaining four subjects was used as training sequences. It was successfully proved that Cauchy assumption provided significant improvement than the Gaussian assumption with an 80% accuracy if a person displayed a positive or a negative expression.[2]

This paper proposed a system for classifying seven basic emotions with two types of features, namely, geometric and appearance features. The face detection and segmentation were performed using a face detection algorithm based on Viola and Jones features. The facial feature extraction consisted of 2 parts. The geometric

feature extraction was done where 19 features were extracted from the face. The segmentation divided the face into mouth, 2 eyes and 2 eyebrows and then the facial characteristic points were located and extracted in these regions and later, using the FCPs and Euclidean distance, certain lengths were calculated. The appearance features were extracted using canny edge detection and the face map was finally represented as feature vector with 64 components. Finally, the facial feature vector was given as input to the radial basis function neural network for classification where the RBF neural network had one hidden layer of neurons with RBF activation functions for describing the receptors and was carried out for person dependent and person independent experiments.[3]

The proposed method consists of emotion recognition through a live video. The biggest appearing face in a live video is being tracked, extracted and scaled to a dimension of 48*48 and treated as an input to the model. In order to recognize the emotion from a live video, the model must be trained with sufficient amount of dataset. In this we would use 3 types of dataset such as FERC-2013, RaFD,CK+. Since this is based on neural network, the models are trained along with the network. Here neuron layers are created instead of creating neurons at every stage. There are 3 convolutional layers present where 2 layers are attached with max pooling in order to reduce the chance of overfitting. For the 1st convolutional layer, the 48*48 image is being sent as the input. All the networks present in the model are trained with 60 epochs. [4]

The proposed method consists of 4 steps in order to recognise the emotion. The various methods are being described as follows: Image Acquisition: In this phase, the emotion of a student's face is being recorded using a DSLR camera and these are stored as .MOV files of dimensions 1280*720 pixels. Pre-processing: After the students' emotions are recorded in a video format, these videos are converted into images. Hence, we obtain 100 images out of which 50 are used for training and the other 50 for testing. Face detection and Emotion recognition: In this study, Viola Jones algorithm is modified and is used for both face detection and emotion recognition. The modified algorithm consists of evaluation of integral of the image. The weight of the sub window is compared with threshold value. If weight of sub window is less than the thresh hold value, then the window is rejected. Else the sub window is being processed for face recognition. The same procedure is repeated for emotion recognition as well. If the value of the filtered face is less than the updated thresh hold value, then the expression is set as interest. Else the expression is set as bored. Evaluation: A set of evaluation parameters such as accuracy, precision, recall was applied onto the faces that were being detected in the earlier stages.[5]

The model in this paper consists of the following modules, Implementation of Face Detection method: Face detection is done using Haar Cascade Classifier which is already implemented in EmguCV. It consists of a classifier file stored as xml file. This xml file is used for training and testing. The Classifier outputs a 1 if the region is likely to show the object (i.e., face) and 0 otherwise. Implementation of Support Vector Machine: The user requests for training examples with labels and then combined with the displacements output by the feature extraction phase and added as a new example to the training set. Unseen expressions to be classified pass the same feature extraction process and are assigned label of the target expression that most closely matches their displacement pattern by the SVM classifier. Evaluation: The proposed system was evaluated by considering the four basic emotions which includes neutral, joy, sad and surprise.[6]

The system includes the following modules: Face Acquisition and Face Detection: Images are captured by camera or obtained from a database of some standard images which are automatically pre-processed. Face is detected using Viola Jones algorithm. The main window is divided into sub windows where the invariant detector rescales the images repeatedly. Feature Extraction and Expression Classification: Local Binary Pattern along with Fuzzy Rule Interpolation is used to do this. The textual image local information is extracted. A sparse situation in hand is dealt with where a CRI (Compositional Rule of Interference) fails to draw conclusion. Here we use the FIVE (Fuzzy Interpolation in Vague Environment). Some excerpts conclusions are as follows: R1 = Lip Tightened ∧ Lip pressor ∨ Upper lip raiser ∧ Brow lowerer, then angry

R2 = Lip Tightened ∧ Outer brow raised ∧ Inner brow raiser, then fear

R3 = Lip corner pulled ∧ Lips parted, then happy

Expression Display: This module displays the required result which is one amongst the seven standard expressions.[7]

In this paper, the proposed system consists of the following modules: Selecting Representative Frames: Images clicked by the camera of a mobile are first converted them into greyscale. Histograms for each of these frames are generated followed by calculation of chi square distance. A frame is selected when the distance between three adjacent frames are minimum. Face Detection and Bandlet Transform: Viola Jones is used to detect the face and geometric structure from a face is utilized. It calculates geometric flow in terms of Bandlet bases. LBP: The LBP uses a 3x3 sub band. The LBP histograms are calculated for all the blocks and then concatenated together to obtain the feature set for the image.KW Feature Selection: It adopts a null hypothesis that states that samples of different groups have same median. It returns a value p if the value of p is close to zero, indicating that the feature is selected. GMM Based Classifier: Features selected are fed into this classifier.

Log-likelihood scores are calculated for each emotion. The emotion with maximum score is given as output to the user.[8]

The proposed methodology consists of implementation of an android application where client server architecture has been implemented. It consists of 3 phases: Android application: An android application is created which takes input from the user and produces the output. The user has to login using the credentials which are used for authentication purpose. Server: The processing of the chatbot takes place on the server. The methodology used on the server side is Recurrent Neural Network. For the chatbot a sequence to sequence RNN is used which consists of 2 main components - encoder and a decoder. Architecture diagram: In this phase a chatbot is implemented with sequence to sequence model along with attention mechanism. Attention mechanism is the technique where the essential characteristics of the sentences are being extracted are the conversation are built upon that.[9]

The paper proposed a system that detects depression in college students. The system is trained with features of happy, contempt and disgust faces. Further videos of the students are collected while they undertake a questionnaire. Necessary features of the face are extracted from the video and compared to a test dataset. The face is detected using the Viola Jones algorithm. Feature extraction is done by the Gabor filters. Classification of the features are done by SVM. The accuracy achieved by this paper is 67%.[10]

4. CONCLUSIONS

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 MTCNN, emotion recognition which is implemented by Keras that mainly focuses on detecting percentage of sadness and levels of other emotions that can reflect depression in an individual. The two modules stated is used to recognize sadness. Finally, the last module, a chatbot is used that is used to recognize depression that further helps to differentiate between sadness and depression.

Many research papers are analyzed in order to select the best appropriate algorithms that can be used in the modules. 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.

REFERENCES

1. Adaboost algorithm using MB-LBP and skin color segmentation- Kai-Biao He, Jing Wen, Bin Fang .
2. Emotion Recognition Using a Cauchy Naive Bayes Classifier- Nicu Sebe, Michael S. Lew, Ira Cohen, Ashutosh Garg, Thomas S. Huang.
3. Automatic Facial Expression Recognition System Based on Geometric and Appearance Features- Aliaa A. A. Youssif, Wesam A. A. Asker.
4. Enrique Correa, Arnoud Jonker, Michael Ozo and Rob Stolk proposed their paper of emotion recognition using Convolutional Neural Network.
5. Emotion Detection using Viola Jones Algorithm- Kartika Candra Kirana, Slamet Wibawanto and, Heru Wahyu Herwant.
6. Support Vector Machine for Face emotion detection on real time basis - E.M. Bouhabba, A.A. Shafie and R.Akmeliawati.
7. Face detection and expression recognition using fuzzy rule Interpolation, Williams. D. Ofor , Nuka. D. Nwiabu, Daniel Matthias.
8. An Emotion Recognition System for Mobile Applications, M. Shamim Hossain and Ghulam Muhammad.
9. A Survey Paper on Chatbots, Aafiya Shaikh, Dipti More, Ruchika Puttoo, Sayli Shrivastav, Swati Shinde
10. Extraction of Facial Features for Depression Detection among Students, Dr.D.Venkataraman, Namboodri Sandhya Parameswaran
11. www.algorithmia.com/blog/introduction-to-emotion-recognition
12. www.google.com
13. https://www.medscape.com/answers/286759-14692/what-are-the-dsm-5-criteria-for-diagnosis-of-major-depressive-disorder-clinical-depression