International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 09 Issue: 05 | May 2022 www.irjet.net p-ISSN: 2395-0072

EMOTION RECOGNITION SYSTEMS: A REVIEW

Shilpa M[1], Prof. Hema S[2]

[1]PG Student, Dept. of Electronics & Communication Engineering, LBSITW, Kerala, India
[2] Assistant Professor, Dept. of Electronics & Communication Engineering, LBSITW, Kerala, India

---

Abstract - *Emotions are state of feelings that can be associated with certain situations. Emotion recognition plays an important role in today's world. It has been an important research area in the recent years. It has a wide range of applications in the field of healthcare, biometric security, education etc. Emotions can be recognized through handwriting, facial expression, speech, posture etc. Different methods can be used for emotion recognition based on its application. This paper gives a brief review of some existing emotion recognition methods by some deep learning and machine learning techniques. The features extracted and the algorithms used in each paper were also briefly discussed.*

Key Words: **Convolutional Neural Network (CNN), Mel Frequency Cepstral Coefficients (MFCC), Emotion recognition, Support Vector Machine (SVM), Recurrent Neural Network (RNN)**

1. INTRODUCTION

Emotions are associated with one's thoughts, feelings, responses, pleasure etc. There were large range of emotions that can be seen in each individuals. It can vary depending on a situation. Emotion recognition is gaining popularity day by day. Applications of emotion recognition includes in the field of medicine, e-learning, monitoring, entertainment, marketing, customer services, security measures etc.

Artificial Intelligence (AI) is a technology that makes smart machines capable of performing tasks that require human intelligence. The availability of large quantities of data and new algorithms made AI an emerging research area in recent years. Through AI, it is possible to recognize emotions by various algorithms.

Emotional state of a person can be accessed through various ways such as by handwriting, facial expressions, voice analysis, ECG signals, body postures, etc. The main steps involved in emotion recognition:

1) Input feature extraction

2) Emotion classification. Features extracted for each method varies depending upon the input provided for emotion classification.

This paper presents a review of emotion recognition systems through various machine learning and deep learning methods.

2. REVIEW ON EMOTION RECOGNITION SYSTEMS

Akriti Jaiswal et al. [1] proposed a facial emotion detection using deep learning. Here the images were given as an input to a CNN network. Feature extraction was done by two submodels by sharing the input and they were of same kernel size. The output obtained through it were flattened into vectors and it is given to a fully connected layer which will classify the emotions.

A. Christy et al. [2] proposed an emotion recognition through speech signals. Here the speech signals splits into short frames. Then feature extraction from each frame was performed using MFCC and Modulation Spectral features. Then the extracted features were used for the classification of emotions. Here the classification was done by using decision tree, random forest, SVM and CNN. CNN has shown more accuracy in recognizing emotions compared to others. Here only limited samples were taken.

Dhara Mungra et al. [3] proposed an emotion recognition system through facial expressions. Emotion recognition was performed initially by some specific image pre-processing steps and by using CNN. This method uses haar cascade for face detection and histogram equalization for increasing the contrast of the image. Also data augmentation was done subsequently for increasing the size of the dataset. Then the images were given to the CNN model for the classification of emotions. This model gives more testing accuracy when using both histogram equalization and data augmentation than without using both histogram equalization and data augmentation.

A deep learning approach for facial expression recognition was proposed by Gozde Yolcu et al. [4]. Firstly, three separate CNN were trained to segment three facial components and the output from these CNN forms a face iconize image. This image is combined with raw facial image which is used as the input for the last CNN. This CNN recognizes various facial expressions.

Akash Saravanan et al. [5] proposed a facial emotion recognition using CNN. Here four different models were used to compare the results; a decision tree model and three neural network models. The neural network models were feed-forward neural network, simple CNN and proposed CNN. Feed-forward neural network predicts the angry expression for every input. Simple CNN model predicts the happy expression for every input. The proposed CNN model mainly consists of six two-dimensional convolutional layers, two max pooling layers and two fully connected layers. Each of its convolutional layer differ in filter size. Upon tuning the hyperparameter, highest accuracy was achieved for the proposed CNN using Adam optimizer. But this model have difficulty in predicting the disgust emotion due to less amount of data in the dataset.

Muktha Sharma et al. [6] proposed a method to analyze the emotions. Here the emotion recognition is done by the fusion of duplex features from the face. The proposed approach consist of three phases: Region of interest (ROI) extraction, Fusion of duplex features and Classification. Firstly, the eye centers were located using a novel eye center detection algorithm and then the face region was extracted from background region of the image. The face region is then subdivided into seven regions to build up a facial expression. Features were extracted from each regions. These features were then fused to form a single feature vector and these feature vectors were used to train the system and finally used to classify the images to predict the facial expression. But the recognition rate of this approach is less for the images having larger head deflection of the subjects.

Emotion detection through face was proposed by Charvi Jain et al. [7]. Here the face detection was done by using Viola Jones algorithm. Face detection was followed by feature extraction. Here the features eye and lips were extracted and it was analyzed for the classification of emotions. Here the author compared the classification accuracy using Fisherface classifier, SVM Classifier, Gabor Filter followed by SVM classifier, Histogram of Gradient (HOG) followed by SVM classifier, Discrete Wavelet Transform (DWT) and HOG followed by SVM classifier, DWT followed by SVM classifier. The HOG followed by SVM classifier gives more accuracy compared to other methods.

Emotion recognition through speech signals was proposed by Adib Ashfaq et al. [8]. Here the audio signal is sampled and it is divided into several frames. For each frame of the speech signal, the extracted MFCC feature vectors were used to detect the underlying emotions of the speech. Each of the frames were classified using trained model. Different frames of a speech may be classified as different emotions. But the speech as a whole conveys only one emotion. So by using the classified frames, a decision has to be made about the emotion of the full speech. To achieve this, we used a majority voting mechanism on the classified frames. While classifying each frame of the unknown instance, a vote is assigned to that particular emotion class. Thus each of the frames were assigned an emotion value. After classifying all the frames of the signal, the emotion which has the maximum number of votes was considered to be the emotion of the full speech signal. The accuracy of the model depends on how many full speech signals were correctly classified using this majority voting mechanism. Logistic Model Tree classifier is used for classification purpose. But this method shows misclassification for certain emotions.

An emotion recognition model based on facial recognition is proposed by D. Yang et al. [9]. Firstly, the given input image will be converted to grayscale and then the face, eye and mouth detection is done through haar cascade algorithm. After the detection, eye and mouth regions were cropped out to perform edge detection. The edge detection is carried out by sobel edge detection method. Then feature extraction which is followed by classifier learning will be taken place and thus the emotions were classified. But the proposed method doesn't consider the illumination and pose of the image.

Emotion recognition from speech signals were analyzed by Esther Ramdinmawii et al. [10]. Here the speech signals were analyzed to obtain the production characteristics of four emotion states. The analysis is done by using the features: instantaneous fundamental frequency, formant frequencies, dominant frequencies, zero-crossing rate and the signal energy. But the analysis shows that there is an overlap between happy and anger emotions.

Anna Esposito et al. [11] proposed a method to assess the depression, anxiety and stress by handwriting and drawing. Here emotional states of participants were assessed by Depression-Anxiety-Stress Scales questionnaire. Some of the tasks were recorded through a digitizing tablet such as pentagon drawing, house drawing, circle drawing, clock drawing, words copied in handprint and one sentence copied in cursive writing. From the collected data, the author computed certain measurements related to timing, ductus and position of the writing device. Then this set of measurement is analyzed and classified using a random forest classifier. Here the set of extracted features is restricted to timing.

Abdul Malik et al. [12] proposed an emotion recognition by speech using spectrogram and deep CNN. The proposed method extracted the features from spectrogram through the CNN. The proposed CNN architecture mainly consists of three

convolutional layers, three fully connected layers and a softmax layer which classifies the emotions. The author compares the result between the proposed CNN model and fine-tuned pre-trained Alexnet model. Satisfactory result were obtained for the former one.

Table -1: Review on different emotion recognition systems

Year & Reference	Algorithm	Dataset	Description	Limitation/Future Scope
2020 [1]	CNN	FERC-2013, JAFEE	Feature extraction from the input images was done by two sub-models by sharing the input and the performance evaluation is done in terms of validation accuracy, computational time, etc.	-
2020 [2]	Decision tree, Random forest, SVM, CNN	RAVDESS	Feature extraction from each frame of the speech signal is performed using MFCC and Modulation Spectral features.	Future scope indicates for more number of samples.
2020 [3]	CNN	FER-2013	Face detection is done using haar cascade algorithm. Histogram equalization and data augmentation is also done in this method.	Future scope indicates that the images can be taken from more sources and other features can be incorporated.
2019 [4]	CNN	RaFD, MUG	Three separate CNN were trained to segment three facial components and the output from these CNN's are combined with raw facial image to recognize various facial expressions	-
2019 [5]	Decision tree, Feed-forward neural network, CNN	FER-2013	Proposed CNN model uses Adam optimizer	This model have difficulty in predicting the disgust emotion due to less amount of data in the dataset.
2019 [6]	CNN	Dataset created from authors, CK+, MMI, JAFEE	Face region of the image is subdivided into seven regions and features extracted from these regions were fused to form a single feature vector to predict the facial expression.	Recognition rate of this approach is less for the images having larger head deflection of the subjects.
2019 [7]	Fisherface, SVM	CK+	Face detection is done using Viola Jones algorithm. Also the features eyes and lips were extracted and analyzed.	-
2019 [8]	Logistic Tree Model	Emo-DB, RAVDESS	MFCC feature were extracted for each frame of the speech signal. Each of the frames were assigned an emotion value. Finally the emotion which has the maximum number of votes is considered to be the emotion of full speech signal.	Misclassification occurs for certain emotions. Future work tends to extract contextual information from speech signal.
2018 [9]	Neural Network Classifier	JAFEE	Eye and mouth detection is done by haar cascade algorithm. These regions were cropped out to perform edge detection through sobel edge detector.	This method doesn't consider the illumination and pose of the image.

2017 [10]	-	German and Telugu Emotion database	The features instantaneous fundamental frequency, formant frequency, dominant frequency, zero crossing rate and the signal energy were analyzed in the speech signal.	Overlap between certain emotions. Future wok tends to incorporate systems to differentiate the emotions.
2017 [11]	Random Forest Classifier	EMOTHAW	Emotional states of participants were assessed by Depression-Anxiety-Stress-Scales questionnaire and some tasks were recorded through a digitizing tablet. The author then computed certain measurements related to timing, ductus and position of the writing device from the collected data for analysis.	Extracted features were restricted to timing. Future scope indicates to incorporate more features.
2017 [12]	CNN	Berlin dataset	The method extracted the features from the spectrogram of the speech signal	Future work tends to use more data with more complex model.
2017 [13]	CNN, LSTM	Berlin database	Speech signal is converted to 2D representation and it is given as an input to CNN and subsequently to LSTM network for the classification of emotions.	Future scope indicates multimodal emotion recognition task.
2015 [14]	SVM, CNN	Candid image facial expression dataset, CK+	Two feature based baseline approaches: LBP followed by SVM and SIFT followed by SVM were compared with CNN architecture.	Future work tends to incorporate live video analysis and the integration of engineered and learned features
2015 [15]	LIBSVM	Berlin dataset	MFCC and MEDC features were extracted from the input speech signal.	-

Wootaek et al. [13] proposed a speech emotion recognition method. This method is based on the concatenation of CNN and RNN. The speech signal was transformed to two dimensional (2D) representation using Short Time Fourier Transform (STFT). The transformed output was given as an input to CNN and subsequently to the LSTM network for the classification of emotions. Future scope indicates multimodal emotion recognition task.

Facial expression recognition for candid images was proposed by Wei Li et al. [14]. Here two feature based baseline approaches were compared with CNN architecture. The baseline approaches were Local Binary Pattern (LBP) followed by SVM and Scale-Invariant Feature Transform (SIFT) followed by SVM. The CNN model uses data augmentation technique to generate sufficient amount of data samples. The CNN mainly consist of input layer, three convolutional layer and an output layer. These baseline approaches and the CNN model were tested with Extended Cohn-Kanade (CK+) dataset and candid image facial expression (CIFE) dataset. The proposed CNN architecture gives highest accuracy when compared with baseline approaches.

A speech emotion recognition method was proposed by Y. D. Chavhan et al. [15]. The input speech given is in .wav file format. MFCC and MEDC (Mel Energy Spectrum Dynamic Coefficients) features were extracted from the input speech signal. The extracted features were given to the LIBSVM (Library for Support Vector Machines) classifier for the classification of emotions. The classifier uses Radial Basis Function (RBF) kernel. The method shows the recognition results for the gender dependent and gender independent system. The results shows that the gender dependent system gives the highest accuracy when compared with gender independent system.

3. CONCLUSION

Emotions has an important role in our day to day life. Emotion recognition is the process of detecting human emotions in various aspects. It is important as it has applications in many fields. Thus the paper reviewed some emotion recognition systems through some deep learning and machine learning approaches.

ACKNOWLEDGEMENT

We would like to thank the Director of LBSITW and the Principal of the institution for providing the support for our work.

REFERENCES

[1] Akriti Jaiswal, A.Krishnama Raju, Suman Deb, "Facial emotion detection using deep learning", 2020 International Conference for Emerging Technology (INCET), IEEE, August 2020.

[2] M.D. Anto Praveena, A. Jesudoss, S. Vaithyasubramanian, A. Christy, "Multimodal speech emotion recognition and classifcation using convolutional neural network techniques", Springer, International Journal of Speech Technology, Volume: 23, pp: 381–388, June 2020.

[3] Dhara Mungra, Anjali Agrawal, Priyanka Sharma, Sudeep Tanwar, Mohammad S. Obaidat, "PRATIT: a CNN-based emotion recognition system using histogram equalization and data augmentation", Springer, Multimedia tools and applications Volume: 79, pp: 2285-2307, January 2020.

[4] Gozde Yolcu, Ismail Oztel, Serap Kazan, Cemil Oz, Kannappan Palaniappan, Teresa E. Lever, Filiz Bunyak, "Facial expression recognition for monitoring neurological disorders based on convolutional neural network", Springer, Multimedia tools and applications, Volume: 78, pp: 31581–31603, November 2019.

[5] Dr. K. S. Gayathri, Akash Saravanan, Gurudutt Perichetla, "Facial emotion recognition using Convolutional Neural Networks", arXiv:1910.05602v1 [cs.CV], October 2019.

[6] Mukta Sharma, Anand Singh Jalal, Aamir Khan, "Emotion recognition using facial expression by fusing key points descriptor and texture features", Springer, Multimedia tools and applications, Volume: 78, pp: 16195-16219, June 2019.

[7] Charvi Jain, Kshitij Sawant, Mohammed Rehman, Rajesh Kumar, "Emotion Detection and Characterization using Facial Features", 2018 3rd International Conference and Workshops on Recent Advances and Innovations in Engineering (ICRAIE), IEEE Conference Record : 43534, May 2019.

[8] Adib Ashfaq A. Zamil, Sajib Hasan, Isra Zaman, Jawad MD. Adam, Showmik MD. Jannatul Baki, "Emotion Detection from Speech Signals using Voting Mechanism on Classified Frames", 2019 International Conference on Robotics, Electrical and Signal Processing Techniques (ICREST), IEEE, February 2019.

[9] D. Yang, Abeer Alsadoon, P.W.C. Prasad, A.K. Singh, A. Elchouemi, "An emotion recognition model based on facial recognition in virtual learning environment", 6th International Conference on Smart Computing and Communications, ICSCC, Procedia Computer Science, Elsevier, Volume:125, pp: 2–10, January 2018.

[10] Esther Ramdinmawii, Abhijit Mohanta, Vinay Kumar Mittal, "Emotion recognition from speech signal", TENCON 2017 - 2017 IEEE Region 10 Conference, December 2017.

[11] Likforman Sulem, Anna Esposito, Marcos Faundez Zanuy, Stephan Clemencon, Gennaro Cordasco, "EMOTHAW: A Novel Database for Emotional State Recognition From Handwriting and Drawing", IEEE Transactions on Human-Machine Systems, Volume: 47, Issue: 2, pp: 273-284, April 2017.

[12] Abdul Malik Badshah, Jamil Ahmad, Nasir Rahim, Sung Wook Baik, "Speech emotion recognition from spectrograms with deep convolutional neural networks", 2017 International Conference on Platform Technology and Service (PlatCon), IEEE, February 2017.

[13] Wootaek Lim, Daeyoung Jang, Taejin Lee, "Speech emotion recognition using convolutional and recurrent neural networks", Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA), IEEE, January 2017.

[14] Wei Li, Min Li, Zhong Su, Zhigang Zhu, "A deep learning approach to facial expression recognition with candid images", 2015 14th IAPR International Conference on Machine Vision Applications (MVA), July 2015.

[15] Y. D. Chavhan, B. S. Yelure, K. N. Tayade, "Speech emotion recognition using RBF kernel of LIBSVM", 2015 2nd International Conference on Electronics and Communication Systems (ICECS), IEEE, June 2015.