A Review on Facial Expression Recognition System using Deep Learning

Rashmi Yadav [1], Lalit K P Bhaiya [2], Ghanshyam Sahu [2]

[1]*Student, Dept of CSE, Bharti college of engineering and technology, New Delhi, India.*

[2]*Assistant Professor, Dept of CSE, Bharti college of engineering and technology, New Delhi, India.*

Corresponding Author: i4usmriti@gmail.com

Abstract: - Human emotions are spontaneous and conscious mental states of feeling that are accompanied by physiological changes in the face muscles implying face expression. Some important emotions are happy, sad, anger, disgust, fear, surprise, neutral, etc. In non-verbal communication, facial expressions play a very important role because of the inner feelings of a person that reflect on faces. A lot of studies have been carried out for the computer modeling of human emotion. However, it's far behind the human vision system. In the area of computer vision, academic research in deep learning, specifically research into convolutional neural networks, received a lot of attention with the fast growth of computer hardware & the arrival of the Big Data era. Many researches & studies on emotion recognition & deep learning methods are carried out to identify emotions. This article presents a survey of Face Expression Recognition (FER) methods, including 3 key phases as pre-processing, extraction of features & classification. This survey discusses the many kinds of FER methods followed by categories & methods of emotional recognition. It also gives a brief overview of the deep learning approaches used in the FER classification system for facial emotion.

Key Words: — ***Detecting emotion, Emotion Recognition, Facial emotions. Facial Expression, Facial emotional recognition, Deep learning.***

I. INTRODUCTION

Emotions constitute a psychological state of the human mind and thinking processes, and they play a crucial part in the communication that takes place between individuals. Nonverbal information, such as that provided by facial expressions and gestures, is an important component of human communication. Expression recognition requires the extraction of a number of facial features from the face of a specific individual in order to function properly. The classification of a person's facial traits into one of the many different categories of emotion is required for emotion recognition. In spite of the fact that an emotion may be expressed, it is still necessary to recognize patterns in order to fulfil the role of recognising the emotion. The primary purpose of this study is to investigate the feasibility of employing facial recognition techniques in order to identify emotional expressions on people's faces.

Manuscript revised August 04, 2023; accepted August 05, 2023. Date of publication August 07, 2023.

This paper available online at www.ijprse.com

ISSN (Online): 2582-7898; SJIF: 5.59

The ability to recognize facial emotions from digital photographs of faces opens the door to a number of potential applications in the field of human-computer interaction. For instance, facial recognition can aid medical professionals in determining whether a patient is in discomfort or in conducting mental health evaluations. It combines facial detection measurement with expression recognition all in one convenient package. The first step of the recognition system for facial expression in photographs is known as facial recognition. Input frames contain images of faces. Identifying face position may be accomplished by a variety of facial feature extraction strategies [1].

The recognition of different facial expressions is a developing field that has applications in areas such as neuromarketing. In human communication, one of the most important things that helps us comprehend the intentions of others is the facial expressions that they use. Even photographs of the same individual might be different depending on the sort of variation that was used. Even though it has been studied from a far earlier time, when it comes to training and testing, few studies that highlighted fair work are ignored. The ability to recognize human emotions mostly depends on one's ability to read facial expressions.

The human mind organizes its thoughts and principles in a hierarchical structure. People will begin by learning fundamental concepts, and then write those concepts to express more complicated ideas. The human mind is analogous to a DNN in that it is made up of multiple neuron layers that operate as role detectors. As their levels develop, these role detectors are able to recognize more abstract qualities. The process of generalizing information in a more complicated fashion is made easier for computers by this. The most important advantage of DL is that it provides a condensed representation of a wider array of functions than the low networks that are utilized by the more typical type of learning [3].

Facial expressions convey vital data about a person's feelings. The field of face recognition is one that is continuously undergoing development and change. Approaches to feature extraction may be categorized according to whether they concentrate on the mobility or deformation of the face and facial characteristics, or if they function locally or holistically, respectively. The purpose of "extraction of face features" is to locate the representation of a face that is the most accurate. Researchers over the past decade have become interested in the concept of facial expression recognition due to its wide variety of potential applications. Deep learning is a subfield of machine learning that has emerged in recent years as a discipline that is not only relevant but also inspiring. Deep learning is a type of machine learning that is the most effective, trackable, time-consuming, and cost-effective overall. Deep learning has made significant improvements or efficiencies in many fields of use. Some of these fields include cancer identification, natural language processing, object detection, stock market research, computer vision, facial recognition, speech recognition, smart cities, and many others. Deep learning has been widely used in many fields. The identification of face expressions in this piece is accomplished by the application of deep learning strategies rather than manually crafted characteristics. Recent advances in recognition tasks have been made possible by an algorithm called deep learning. As a result, world records have been broken in the majority of identification challenges. Deep neural networks, such as deep-neural networks (DNN) and convolutional neural networks (CNN), are often the sorts of deep networks that are utilized to tackle problems with recognition. In the design of convolutional neural networks (CNN), feature extraction and classification have been identified as important components [4, 5].

The remaining sections of the paper are organized as follows: The second portion provides a more in-depth examination of emotion recognition, including the classifications of emotions and the strategies employed in ER. In Section III, a concise introduction to the Facial Expression Recognition system is provided. In Section IV, a variety of different approaches to deep learning are discussed. In Section V, the work that was done in FER's literature is presented utilizing a variety of deep learning approaches. At long last, the process is finished off with Section VI.

II. EMOTION RECOGNITION

Being able to feel and express emotions is essential to the human experience. In the social lives of humans, it comes naturally to understand how the other person is feeling emotionally, but when it comes to computers, this is a lot more challenging task. Expressions of emotion may be unimodal, such as speech, facial expressions, text, gesture, etc.; bimodal, such as speech and facial; speech and text; brain signals and facial; etc.; multimodal, such as audio, video, physiological signals, and so on; or unimodal, bimodal, or multimodal. Unimodal expressions of emotion include speech; bimodal expressions include facial expressions, text, gesture, and so on.

Affective computing is a subfield of computing that focuses on the research and development of systems and devices that are capable of perceiving, interpreting, and simulating human emotions. Emotion detection is a primary area of focus within this subfield of computing. Affect is a term that refers to the awareness of one's feelings. Affective computing may have been initially introduced by Professor Picard in 1995, and he may have characterized it as "that computing which connects to, emerges from, and actively influences feelings." It is able to teach the system that it creates information through the application of artificial intelligence. Examples of data that may be used to define feelings include an individual's facial characteristics, the fluency with which they articulate themselves, the physiological signals that their body emits, and articulated movements. Expressions of joy, surprise, contempt, sadness, rage, and terror can be shown on their faces. Facial expressions can reveal a person's emotions. Let's take a more in-depth look at these data on categorization [6, 7].

Table.1. The table below provides key emotions FE

Emotion	Facial Expression
Anger	he mind, Dilated nostrils, Lipswere strongly pressed
Happiness	Raised cheeks, widened side of lips
Surprise	Strong eyebrows curved, A more visible eye

Disgust	Lips Raised, Wrinkle's nose
Sadness	Skin triangulated under the eyebrow; Upper eyelid pulled in.
Fear	Growing eyebrows, Open the mouth

2.1 Categories of Emotions Recognition

2.1.1 Facial Expressions

One of the most straightforward methods also happens to be one of the oldest: reading someone's emotions from their face. Due to the fact that it is believed to link to important facets of feelings, it has a lengthy history. As a direct consequence of this, a variety of systems for classifying emotions on the basis of facial expressions have been developed. All of these methodologies are predicated on the idea that emotional signals may be inferred from the spatial configuration of particular items and areas on a face [8].

2.1.2 Speech

When someone is speaking, the sound that they make not only communicates the meaning of what they are saying, but it also provides information about how they are feeling. There are several characteristics of speech that may be utilized to determine what kind of speech it is. One of the most important sources of power in expressiveness is the vibrating of the vocal chords. The rate at which the vibrations of the cords are carried along by an acoustic signal is referred to as its fundamental frequency. It is a reference to the tone of a voice. The changes in pitch and volume, taken combined, make up the prosody. The acoustic characteristics of a voice include its pitch, its strength, its interval of speaking, and its spectral aspects [9].

2.1.3 Physiological Signals

Although the physiological signal has no direct bearing on the experience of feeling, the information that may be gleaned from signals can be utilized to either target or identify feelings. It's possible to split it up into two different groups depending on where you're coming from. [10] The first category of signals originates from the peripheral nervous system and includes things like heart rate, electrocardiograms, electromyograms, and skin conductance. The second category originates from the central nervous system and includes things like brain signals (EEG).

2.2 Multimodal Information

Multimodal Information In order to recognize feelings, you may utilize a modified version of any of the approaches described earlier in this article. It is claimed that data is multimodal when more than one type of data is used for emotion identification. Data or information used for the purpose of feeling detection is also considered to fall under this category. For example, the combination of voice signals with facial expressions can provide an audiovisual signal, which is the most up-to-date sort of data [11, 12].

2.3 Techniques used for Emotion Recognition

2.3.1 Principal Component Analysis

[12] The Principal Constituents Analysis (PCA) is a technique for analyzing patterns in data and expressing them in a way that emphasizes their separation and similarity. This tool may be found here. In order to recognize facial emotions using Eigen faces, PCA is utilized to first eliminate characteristics that are contained within the input pictures. They begin by generating a functioning dataset so that they may evaluate the results of the experiment against it. If the facial picture has been pre-processed, it is matched to a training dataset, which has previously been computed. However, they separated the training dataset into six simple categories according to universal voice (happy, surprised, disgusted, sad, angry, and afraid).

2.3.2 Local Binary Pattern

The LBP-based feature extraction approach [12] is utilized due to the fact that it possesses a superior light invariance attribute and a low level of computing complexity. After the neighborhood value is thresholder by the center value, the output is regarded as a binary number. If a certain quantity of canter pixels is superior to any other number of neighboring pixels, put 1; otherwise, write 0. It successfully encodes knowledge about neighborhoods in these many ways.

2.3.3 Active Appearance Model

The statistical approach for shape and texture modelling as well as feature extraction is known as the Active Appearance model [12]. It is often considered to be the best option for computer vision software. The development of statistical appearance models is accomplished by AAM through the combination of two models: the model of form dissimilarity and the model of texture variance. The AAM's output is a training face picture series form and texture grouping model. This model is produced

as a byproduct of the AAM. "Textures" are a target image's pixel intensities.

2.3.4 Facial Action Coding System (FACS)

In 1976, researchers Paul Ekman and Wallace Friesen devised a method for evaluating individuals based on the characteristics of their facial features. A technique known as FACS [12] is based on the study of the relationship between the contraction of facial muscles and the way a person's face looks. On the face were Upper and Lower Face Action Modules that may be activated. Action Units are a term used to describe the changes that occur in the face as a result of the contraction of a single muscle or a group of muscles. 46 AUs are used to describe the latest developments in face characteristics, while 12 AUs are used to indicate eye contact direction and head position respectively.

2.4 Haar Classifier

For its owing high detection accuracy & real-time output, the Haar classifier-based approach [12] is selected for face recognition. The amount of a function is a variation of a no. pixel values in white & black areas, & it is made up of black & hite related rectangles. The use of Integral images Improves the computational speed of function measurement. depending on a specific mission. RBM is normally used for learning features [17-20].

III. LITERATURE REVIEW

A current work examines studies and research findings of recognizing the face expression based on research literature both at home and abroad in recent years, from feature extraction expression, classification, and face recognition. These studies and findings are based on the recognition of facial expressions.

K. Tsai et al. (2018) This article presents a proposal for a FER system that is built on top of an advanced neural CNN. It does this by employing facial frontalization and positioning approaches, which help to reduce the impact of background noise and components that are not very apparent. Precisions are raised in addition to the hierarchical structure and adaptive exponentially weighted average ensemble that are already there. The proposed system surpasses the state-of-the-art FER techniques, as shown by simulations conducted on a variety of datasets, and it has the potential to identify a person's expression [21].

In their 2019 publication, J.-H. Kim and colleagues offer a novel FER system that is based on hierarchical deep learning. A suggested approach for the CK+ and JAFFE datasets, both of which are generally regarded as proven data sets for the recognition of facial expressions. The results of a 10-fold cross-validation show that the CK+ data set has a precision of 96.46%, whereas the JAFFE data set only has a precision of 91.27%. In comparison to other methods, the hierarchical deep network structure that was developed demonstrates an increase in accuracy of up to 3% and an improvement of 1.3% in the average CK+ data set. The JAFFE data sets see an improvement in accuracy of up to 7%, with an increase of about 1.5% being validated as the norm. [22].

S. Lin et al. (2019) propose in this paper a continuous model for facial expression identification based on a deep learning approach that combines CNN and RNN to analyse and identify continuous face expressions over some time to enhance traditional image recognition technique [23]. This model is based on a deep learning technique that recognises facial emotions based on a deep learning technique that combines CNN and RNN.

P. Kaviya and T. Arumugaprakash (2020) This being said, Researchers Babajee et al. (2020) investigate the recognition of human facial expressions with the use of the deep learning approach and the CNN algorithm. The system employs a labelled data set consisting of about 32,298 photos with a variety of different face expressions for training and testing purposes. A noise reduction face detection subsystem that contains a feature extraction component is included in the pre-training process. The results of their research demonstrate that they were able to identify all seven essential human emotions with an accuracy of 79.8% [25], despite the fact that they did not use any optimisation approaches.

L. Cai et al. (2020) An approach that takes into account both verbal and non-verbal cues, as described in this article, is suggested. They do it by using CNN and LSTM to learn the emotional aspects of speech. The concurrent extraction of facial expression characteristics using a variety of small-scale kernel convolution blocks. In the end, DNNs were employed to combine the aspects of the face and the spoken emotion. An IEMOCAP data set evaluated a multimodal model for emotion identification. When compared to a model that just included facial expressions or voice, the suggested model's recognition accuracy was found to be 10.05 and 11.27 percent more accurate, respectively, than the previous model. [26].

(2020) J. Awatramani and N. Hasteer. The findings of this research indicate that children who have autism spectrum disorder may be taught to recognise human emotions by

employing and putting into practise the core architecture of CNN. An existing data set from the body of literary research was used to validate a model, which was shown to have an accuracy of 67.5%. [27].

Y. Maeda et al. (2020) In this article, the researchers suggest a method of deep learning to classify the range of human emotions conveyed by facial expressions and to generate emotional responses in robots based on the Markov model's emotional states. They use CNN, which is a form of deep learning, to learn human facial photos with distinct emotional traits and to recognise human emotions in human encounters. This operation takes place here. The emotional behaviour of a robot is returned to humans on the basis of the human feeling that it has learned through deep learning. An experiment including interaction was carried out with a real-life communication robot for the purpose of this study, and the results of that experiment are provided here [28].

Table.2. Comparative Study of Literature Survey

Year	Author's Name	Title	Technique	Outcome
2018	K. Tsai, J. Ding and Y. Lee	Frontalization with Adaptive Exponentially-Weighted Average Ensemble Rule for Deep Learning-Based Facial Expression Recognition	Advanced neural CNN	Simulations on various datasets reveal that the system proposed outperforms state of art FER approaches & may well detect a person's expression
2019	J. -H. Kim, B. -G.Kim, P. P. Roy and D. -M. Jeong	Efficient Facial Expression Recognition Algorithm Based on Hierarchical Deep Neural Network Structure	Hierarchical deep learning	A finding of 10-fold cross-validation demonstrates 96.46% of the preciseness inthe CK+ data set and 91.27% of the precision in the JAFFE data set. Up to 7% of accuracy in JAFFE data sets is improved & the average increase is confirmed by around 1.5%
2019	S. Lin, Y. Tseng, C. Wu, Y. Kung, Y. Chen and C. Wu	A Continuous Facial Expression Recognition Model based on Deep Learning Method	Deep learning technique	analyze & detect continuous face expressions over some time to enhance conventional image recognition technique
2020	P. Kaviya and T. Arumugaprakash	P. Kaviya and T. Arumugaprakash	Deep learning technique	For FER-2013 & 60% for custom data sets, the proposed model obtains a final accuracy of 65%
2020	P. Babajee, G. Suddul, S. Armoogum and R. Foogooa	Identifying Human Emotions from Facial Expressions with Deep Learning	Deep learning method	The outcomes of their work show an accuracy of 79.8% to identify all 7 fundamental human emotions withoutusing optimization methods
2020	L. Cai, J. Dong and M. Wei	Multi-Modal Emotion Recognition from Speech and Facial Expression Based on Deep Learning	CNN & LSTM	The total accuracy of recognition of their proposed model has been improved by 10.05% & 11.27% respectively compared to a single modal of speech & face expression.
2020	J. Awatramani and N. Hasteer	Facial Expression Recognition using Deep Learning for Children with Autism Spectrum Disorder	CNN	A model has been validated using an existing literature data set, with an accuracy of 67.5%
2020	Y. Maeda, T. Sakai, K. Kamei and E. W. Cooper	Human-Robot Interaction Based on Facial Expression Recognition Using Deep Learning	Deep learning	operate here to learn human facial picturesusing CNN, which is a type of deep learning, with different emotional features& to identify the human emotions in human interactions.

IV. CONCLUSION

The state, mood, and present feeling of a person may be deduced from their facial display of emotions, which is a form of nonverbal communication. Understanding a person's emotions may require looking at them from different vantage points throughout their life. The proportion of different feelings experienced at each stage is quite variable. This paper provides an overview of face expression recognition systems as well as the many research that have been conducted. Face recognition, the extraction and categorization of facial characteristics are all included in these systems. There are a variety of methods that may be utilized to achieve desirable recognition rates. Methods with a higher recognition rate achieve higher levels of performance. These methods offer workable answers to the problem of recognizing people's facial expressions and have the potential to function well in constrained settings. Recognizing human facial expressions is one of the most important and difficult aspects of social communication. It's also one of the most rewarding. Humans typically use their face expressions, which are both a natural and straightforward manner of communicating their emotions and intentions. Facial expressions are crucial components of non-verbal communication because they convey feelings and emotions. This paper examines the research that has been conducted and published in the field of facial expression recognition, as well as the many techniques that have been utilized for facial expression recognition.

REFERENCES

[1]. Annu George," Artificial Intelligence Facial Expression Recognition for Emotion Detection: Performance and Acceptance", Issues in Information Systems Volume 21, Issue 4, pp. 81-91, 2020.

[2]. R. G. Harper, A. N. Wiens, and J. D. Matarazzo, Nonverbal communication: The state of the art. John Wiley & Sons, 1978.

[3]. V. PreamSudha," A Survey on Deep Learning Techniques, Applications and Challenges", International Journal of Advance Research in Science and Engineering, IJARSE, Vol. No.4, Issue 03, March 2015.

[4]. Jung, H., Lee, S., Park, S., Kim, B., Kim, J., Lee, I., & Ahn, C. (2015). Development of deep learning-based facial expression recognition system. 2015 21st Korea-Japan Joint Workshop on Frontiers of ComputerVision (FCV).

[5]. Aleix Martinez," A Model of the Perception of Facial Expressions of Emotion by Humans: Research Overview and Perspectives", Journal of Machine Learning Research 13 (2012) 1589-1608.

[6]. Wisal Hashim Abdulsalam, Rafah Shihab Alhamdani, Mohammed Najm Abdullah," Facial Emotion Recognition: A Survey", International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 7, Issue 11, November 2018, ISSN: 2278 – 1323.

[7]. Gayathri.P&Matilda.S," Emotion Recognition: A Survey", International Journal of Advanced Research in Computer Science and Applications Issn 2321-872x Online Issn 2321-

8932 Print Impact Factor: 0.782 Volume 3, Issue 1, January 2015.

[8]. Ralph Adolphs, _Recognizing Emotion from Facial Expressions: Psychological and Neurological Mechanisms ‘, University of Iowa College of Medicine.

[9]. Dorota Kaminska and Tomasz Sapinski and Adam Pelikant,

[10]. Recognition of Emotional States in Natural Speech ‘, Institute of Mechatronics and Information Systems.

[11]. Guillaume Chanel, Cyril Rebetez, MireilleBétrancourt, and Thierry Pun, (2011) Emotion Assessment from Physiological Signals for Adaptation of Game Difficulty ‘, IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans, Vol. 41, No. 6.

[12]. FoteiniAgrafioti, DimitriosHatzinakos, and Adam K. Anderson, (2012), ECG Pattern Analysis for Emotion Detection ‘, IEEE Transactions on Affective Computing, Vol. 3, No. 1.

[13]. DipikaRaval, MukeshSakle,” A Literature review on Emotion Recognition System Using Various Facial Expression”, IJARIIE- ISSN(O)-2395-4396, Vol-1 Issue-2 2015.

[14]. Anjali,” A Survey on Different Types of Techniques uses for Face Expression Recognition”, International Journal of Engineering and Techniques - Volume 6 Issue 1, February 2020.

[15]. Dhwani Mehta, Mohammad Faridul Haque Siddiqui and Ahmad Y. Javaid,” Facial Emotion Recognition: A Survey and Real-World User Experiences in Mixed Reality”, Sensors 2018, 18, 416.

[16]. I.Michael Revina,” A Survey on Human Face Expression Recognition Techniques”, Computer and Information Sciences.

[17]. Nithya Roopa. S,” Emotion Recognition from Facial Expression using Deep Learning”, International Journal of Engineering and Advanced Technology (IJEAT).

[18]. Shan Li and Weihong Deng∗, Member, IEEE,” Deep Facial Expression Recognition: A Survey”, arXiv:1804.08348v2 [cs.CV] 22 Oct 2018.

[19]. Shaveta Dargan1 · Munish Kumar,” A Survey of Deep Learning and Its Applications: A New Paradigm to Machine Learning”, Archives of Computational Methods in Engineering.

[20]. Senzhang Wang, Jiannong Cao, Fellow, IEEE, Philip S. Yu, Fellow, IEEE,” Deep Learning for Spatio-Temporal Data Mining: A Survey”, arXiv:1906.04928v2 [cs.LG] 24 Jun 2019.

[21]. K. Tsai, J. Ding and Y. Lee, "Frontalization with Adaptive Exponentially-Weighted Average Ensemble Rule for Deep Learning-Based Facial Expression Recognition," 2018 IEEE Asia Pacific Conference on Circuits and Systems (APCCAS), 2018, pp. 447-450.

[22]. J. -H. Kim, B. -G. Kim, P. P. Roy and D. -M. Jeong, "Efficient Facial Expression Recognition Algorithm Based on Hierarchical Deep Neural Network Structure," in IEEE Access, vol. 7, pp. 41273- 41285, 2019.

[23]. S. Lin, Y. Tseng, C. Wu, Y. Kung, Y. Chen and C. Wu, "A Continuous Facial Expression Recognition Model based on Deep Learning Method," 2019 International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS), 2019, pp. 1-2.

[24]. P. Kaviya and T. Arumugaprakash, "Group Facial Emotion Analysis System Using Convolutional Neural Network," 2020 4th International Conference on Trends in Electronics and Informatics (ICOEI)(48184),2020.

[25]. P. Babajee, G. Suddul, S. Armoogum and R. Foogooa, "Identifying Human Emotions from Facial Expressions with Deep Learning," 2020 Zooming Innovation in Consumer Technologies Conference (ZINC), 2020, pp. 36-39.

[26]. L. Cai, J. Dong and M. Wei, "Multi-Modal Emotion Recognition from Speech and Facial Expression Based on Deep Learning," 2020 Chinese Automation Congress (CAC), 2020, pp. 5726-5729.

[27]. J. Awatramani and N. Hasteer, "Facial Expression Recognition using Deep Learning for Children with Autism Spectrum Disorder," 2020 IEEE 5th International Conference on Computing Communication and Automation (ICCCA), 2020, pp. 35-39.

[28]. Y. Maeda, T. Sakai, K. Kamei and E. W. Cooper, "Human-Robot Interaction Based on Facial Expression Recognition Using Deep Learning," 2020 Joint 11th International Conference on Soft Computing and Intelligent Systems and 21st International Symposium on Advanced Intelligent Systems (SCIS-ISIS), 2020, pp. 1-6.