International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435(Online) Volume 12, Issue 08, August 2023)

A Review on Emotion Recognition using Machine Learning and Facial Expressions

Aishwarya Vishwakarma[1], Dr. Vikas Sakalle[2]
[1]Research Scholar, [2]CSE Department, LNCT University, Bhopal
aishwarya.vishwakarma@gmail.com[1], vikassakalle@gmail.com[2]

***Abstract*-- Facial Expression Recognition (FER) is a computer vision task aimed at identifying and categorizing emotional expressions depicted on a human face. Various researchers have worked in this area and used machine learning and deep learning models for this task. In this present study we have reviewed the work done by researchers in last five years (2018-2023). We have observed that 30% researchers have used machine learning and 70%researchers have used deep learning algorithms for the task, Most of the researchers have used CK+, MIMI, and JAFFE data set, remaining researchers used other datasets. We have also observed that most of the researchers have obtained accuracy greater than 80% and their model is able to recognize more than 4emotions**.

I. INTRODUCTION

Facial actions convey information about a person's emotion, intention, and physical state, and are vital for use in studying human cognition and related processes (W. -S. Chu,2017). Facial expressions have been studied due to the great deal of applications that are required for interpreting human communication through facial gestures, such as automaticbehavior recognition, human-computer interaction, pain assessment, health-care, surveillance, deceit detection and sign language recognition. Two main approaches can be considered to expression recognition methodologies: a static approach that processes each image in a sequence separately, and a sequence based or dynamic approach where the expression evolution can be captured(D. Acevedo et al, 2017).Expression recognition has long been a challenging task in the field of computer vision. In previous works, researchers usually at- tempt to solve the FER problem by detecting expression-related AUs or training a classifier with extracted expressional features. In recent years, deep learning algorithm is applied to expression analysis, which is regarded as a promising approach to recognize expressions (SiyueXie et al,2019).

Facial expressions are very useful for expressing emotions and intentions and primarily for interacting with other people. New facial expression recognition systems have been implemented in several fields, including psychology, computer graphics, consumer neuroscience, media testing & advertisement, psychotherapy, medicine, and transportation security.

Psychologists were the first among other people who realized and investigated the importance of facial expressions and the need for recognizing emotions through computer applications (Reddy et al, 2019).

Automatic facial expression recognition is an interesting and challenging problem which has important applications in many areas like human-computer interaction. It helps to build more intelligent robots which has the ability to understand human emotions.

Various researchers have done significant work in this area (latest reference).In this work we have reviewed the work done by various researchers in last five years. (2018-2023).

Apart from introduction the paper has been organized as follows: Section 2 contains work done by the researcher, Section 3 presents our observations section 4 contains conclusion.

Emotion detection using facial images during 2018-2023

In this section we have mentioned the work done by the researchers during 2018-2023 for emotion detection using machine learning and deep learning and facial data. We have observed their work from the point of datasets use, machine learning/deep learning method using and number of emotions recognized.

Zeng et al, 2018mentioneda novel idea for FER to automatically distinguish the expressions with high accuracy. The results indicate that the presented idea can achieve a high recognition accuracy of 95.79% on the extended Cohn-Kanade (CK+) database for seven facial expressions, which performs the other three state of-the-art methods by as much as 3.17%, 4.09% and 7.41%, respectively.

Jain et al, 2020 proposed a method that includes the five major processes as Extended Boundary Background Subtraction (EBBS), Multi-Angle Texture Pattern+STM, Densely Extracted SURF+Local Occupancy Pattern (LOP), Priority Particle Cuckoo Search Optimization (PPCSO) and Long Short-Term Memory -Convolutional Neural Network (LSTM-CNN). Initially, the EBBS algorithm subtracts the background and isolates the foreground from the images which overcome the illumination and pose variation.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435(Online) Volume 12, Issue 08, August 2023)

Then, the MATP-STM selects the texture patterns and DESURF-LOP selects the suitable key features of the facial points. The PPCSO algorithm selects the suitable features from the MATP-STM feature set to speed up the classification.

Sajid et al, 2019 has proposed a method in that Raspberry Pi camera captures the video stream and detects faces using the Viola Jones algorithm. The face area is pre-processed by using Gabor filter and median filter prior to feature extraction. Both Oriented FAST and Rotated BRIEF (ORB) features are then extracted and the support vector machine (SVM) classifier is used for training.

Xie et al, 2019 have proposed a model that contains two novel modules: an attention-based Salient Expressional Region Descriptor (SERD) and the Multi-Path Variation-Suppressing Network (MPVS-Net). SERD can adaptively estimate the importance of different image regions for FER task, while MPVS-Net disentangles expressional information from ir- relevant variations. By combining SERD and MPVS-Net, DAM-CNN is able to highlight facial expression- suitable features and generate a variation-robust representation for expression classification.

Nazir et al, 2019 have been proposed the transformation of HOG features to frequency domain can make this descriptor one of the most suitable to characterize illumination and orientation invariant facial expressions. Discrete cosine transform (DCT) has been applied to transform the features into frequency domain and obtain the most important discriminant feature, these features are fed to the well-known classifier to identify the underlying emotions from expressive facial images.

Li et al., 2020had proposed new face cropping and rotation strategies and simplification of the convolutional neural network (CNN) to make data more abundant and only useful facial features can be extracted. recognition accuracies of 97.38% and 97.18% were obtained for 7-class experiments on the CK+ and JAFFE databases.

Saad Saeed et al, 2022 mentioned an algorithm, automated framework for facial detection using a convolutional neural network (FD-CNN) is proposed with four convolution layers and two hidden layers to improve accuracy. FD-CNN is performed in three major steps that include preprocessing, feature extraction, and classification. By using this proposed method, an accuracy of 94% is obtained in FER. To validate the proposed algorithm-fold cross-validation is performed. After validation, sensitivity and specificity are calculated which are 94.02% and 99.14%, respectively.

Furthermore, the f1 score, recall, and precision are calculated to validate the quality of the model which is 84.07%, 78.22%, and 94.09%, respectively.

Ali I. Siam et al. 2022developedanapproach includes four phases: preprocessing, key point generation, key point selection and angular encoding, and classification. MediaPipe face mesh algorithm has been used to generate key point, which is based on real-time deep learning. They achieved accuracy of 97%, which ensures superiority among the efforts in this field.

Mayuri Arul VinayakamRajasimman et al. 2022 implemented a novel Robust Facial Expression Recognition using an Evolutionary Algorithm with Deep Learning (RFER-EADL) model. RFER-EADL aims to identify various kinds of emotions using computer vision and DL models. Primarily, RFER-EADL performs histogram equalization to normalize the intensity and contrast levels of the images of identical persons and expressions. Next, the deep convolutional neural network-based densely connected network (DenseNet-169) model is exploited with the chimp optimization algorithm (COA) as a hyper parameter-tuning approach. Finally, teaching and learning-based optimization (TLBO) with a long short-term memory (LSTM) model is deployed for facial expression recognition and classification. The designs of COA and TLBO algorithms aided in the optimal parameter selection of the Dense Net and LSTM models, respectively.

Table 1:
work done by researchers.

AUTHOR	DATASET	METHOD	IMOTION	ACCURACY
Reddy et al,2019	CK+	SVM	6	97.3 %
Zeng et al,2018	(CK+)	DSAE	7	95.79
Jain et al,2020	CK+ MMI	LSTM-CNN	6	96.17 98.72
Sajid et al,2019	(CK+), MMI, and JAFEE	SVM	7	99.1 on MMI accuracy (92) on JAFFE accuracy (90) on CK+ 99.6 by using only 32 features with KNN
Xie et al,2019	CK+ JAFFE	DAM-CNN	6	With 10-fold cross-validation accuracy 95.88 99.32
Nazir et al,2019	CK+ MMI JAFFE	CNN	7	99.1 on MMI accuracy (92) on

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435(Online) Volume 12, Issue 08, August 2023)

				JAFFE accuracy (90) on CK+ 99.6 by using only 32 features with KNN
Li et al,2020	CK+ JAFFE	CNN	7	97.38 0.14 SD 97.18 0.30 SD
Saad Saeed et al,2022	CK+	K-FOLD	8	84.07%,78.22% , and 94.09%
Ali I. Siam ET AL,2022	CK+ JAFFE *RAF-DB*	MLP	7	97%
Mayuri Arul Vinayakam Rajasimman et al,2022	CK+	RFER-EADL	7	99.21%
A-Hyeon Jo,2023	Korean speech emotion database	Bi-LSTM and CNN-	8	96%
Elif Yildirim, 2023	FER-2013 and CK+	CNN	4	77.68%
Li-Min Zhang,2023	RAVDEES, CASEA	CNN and Bi LSTM	7	87.91%
SaiyedUmer et al,2021	KDEF, GENKI-4k , CK+	CNN	7	94.6%
Hadhami Aouani et al,2020	RML emotion dataset	SVM	6	93%
Ali Shariq Imran et al,2020	KAGGLE SENTIMENT140, EMOTIONAL TWEETS	LSTM	6	81%
Awais Salman Qazi et al,2022	CK+ JAFFE	CNN	7	CK+(92.66) JAFFE(94.94)
Zhenjie Song et al,2021	FER2013	CNN	7	74%
Mohammed F. Alsharekh et al,2022	FER-2013, CK+, and KDEF	CNN	7	94.04
D Y Liliana,2019	CK+	CNN	8	92.81

Table 1 represents the work done by researchers for emotion detection in last five years. We can observe from Table 1 that 70% researchers have used deep learning model and used CK+ dataset,MIMI,JAFFE and 30% researchers have used machine learning model like SVM. And used CK+ and other dataset also.

A-Hyeon Joet al. 2023mentioned a two-stream-based emotion recognition model based on bidirectional long short-term memory (Bi-LSTM) and convolutional Neural networks (CNNs) using a Korean speech emotion database, and the performance is comparatively analyzed. The data used in the experiment were obtained from the Korean speech emotion recognition database built by Chosun University. Two deep learning models, Bi-LSTM and YAMNet, which is a CNN-based transfer learning model, were connected in a two-stream architecture to design an emotion recognition model. Consequently, the speech emotion recognition performance of Bi-LSTM and YAMNet was 90.38% and 94.91%, respectively. However, the performance of the two-stream model was 96%, which was a minimum of 1.09% and up to 5.62% improved compared with a single model.

ElifYildirimet al., 2023 proposed a facial expression-based prediction model to eliminate user bias in questionnaire-based assessment systems and to minimize false entries to the system. They measured the neural, behavioral, and physical manifestations of emotions with a mobile app and recognize emotional experiences from facial expressions. They used three different situations to test whether there couldbe any factor other than the food that could affect a person's mood. They asked users to watchvideos, listen to music or do nothing while eating. They employed three ConvolutionalNeural Network (CNN) architectures, fine-tuned VGG16, and Deepface to recognize emotionalresponses during eating. The experimental results showed that the fine-tunedVGG16 provides remarkable results with an overall accuracy of 77.68% for recognizingthe four emotions. They worked on four emotions.

Li-Min Zhanget al, 2023proposed a speech emotion recognition method based ongender classification. First, they use MLP to classify the original speech by gender. Second, basedon the different acoustic features of male and female speech, they analyze the influence weights ofmultiple speech emotion features in male and female speech, and establish the optimal feature setsfor male and female emotion recognition, respectively. Finally, they train and test CNN and BiLSTM, respectively, by using the male and the female speech emotion feature sets.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435(Online) Volume 12, Issue 08, August 2023)

Saiyed Umer et al, 2021 proposed a novel facial expression recognition system in this paper. The objective of this paper is to identify thetypes of expressions in the human face region. The implementation of the proposed system is divided into four components. In the first component, a region of interest as face detection has been performed from the captured input image. Forextracting more distinctive and discriminant features, in the second component, a deep learning-based convolutional neuralnetwork architecture has been proposed to perform feature learning tasks for classification purposes to recognize the typesof expressions. To enhance the performance of the proposed system, in the third component, some novel data augmentation techniques have been applied to the facial image to enrich the learning parameters of the proposed CNN model. In the fourth component, a trade-off between data augmentation and deep learning features have been performed for fine-tuningthe trained CNN model..

Hadhami Aouani et al, 2020 proposed an emotion recognition system based on speech signals in two stage approach namely feature extraction and classification engine. Firstly two set of features are investigated which are the first one we extract 42 dimensional vector of audio features including 39 coefficients of Mel Frequency Cepstral Coefficients (MFCC), Zero Crossing Rate (ZCR), Harmonic to Noise Rate (HNR) and Teager Energy Operator (TEO). And the second one, we propose the method Auto-Encoder for the selection of pertinent parameters from the parameters previously extracted. Secondly, we use the Support VectorMachines (SVM) as a classifier method. Experiments are performed on the Ryerson Multimedia Laboratory (RML).

AliShariqImran et al, 2020showed a studyto analyze reaction of citizens from different cultures to the novel Coronavirus and people's sentimentabout subsequent actions taken by different countries. Deep long short-term memory (LSTM) models usedfor estimating the sentiment polarity and emotions from extracted tweets have been trained to achieve state of-the-art accuracy on the sentiment140 dataset.

Awais Salman Qazi et al, 2022proposed anew architecture design of a convolutional neural network (CNN) for the FER system and containsfive convolution layers, one fully connected layer with rectified linear unit activation function, anda SoftMax layer. Additionally, the feature map enhancement is applied to accomplish a higherdetection rate and higher precision.

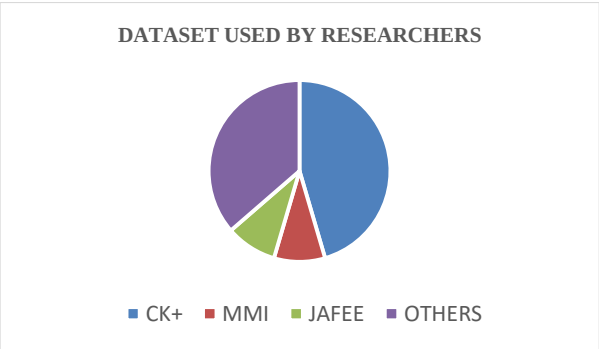
Results indicate that the proposed CNN achieves 92.66%accuracy with mixed datasets, while the accuracy for the cross dataset is 94.94%.

Zhenjie Song et al, 2021proposed a novel featurefusion dual-channel expression recognition algorithm based on machine learning theoryand philosophical thinking. Specifically, the feature extracted using convolutional neuralnetwork (CNN) ignores the problem of subtle changes in facial expressions. The firstpath of the proposed algorithm takes the Gabor feature of the ROI area as input. Inorder to make full use of the detailed features of the active facial expression emotion area, first segment the active facial expression emotion area from the original face image, and use the Gabor transform to extract the emotion features of the area. Focus on the detailed description of the local area. The second path proposes an efficient channelattention network based on depth separable convolution to improve linear bottle neckstructure, reduce network complexity, and prevent overfitting by designing an efficientattention module that combines the depth of the feature map with spatial information.

Mohammed F. Alsharekh, 2022mentioned a technique which is using aCNN model to classify emotions from facial images. The proposed algorithm is an enhanced network architecture developed to process aggregated expressions produced by the Viola–Jones(VJ) face detector.

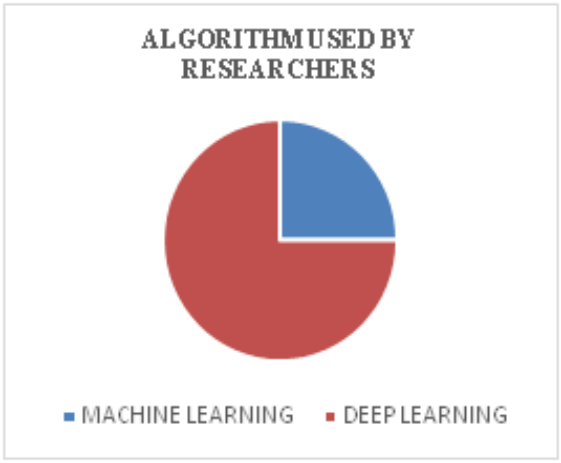
D Y Liliana, 2019 proposed Convolutional Neural Network (CNN) approach toFER task. This task is performed by detecting the occurrence of facial Action Units (AUs) as a subpart of Facial Action Coding System (FACS) which represents facial expression. This research uses the CK+ dataset which is used for FER experiment. Thesystem performance shows average accuracy rate of 92.81%.

Figure 1 shows that graphical representation of our observation. We can see that maximum researchers have used CK+ dataset after that MIMI and JAFFE



International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435(Online) Volume 12, Issue 08, August 2023)

Figure 2 shows that graphical representation of our observation. We can see that maximum researchers have used deep learning algorithm rather than machine learning algorithm.



II. CONCLUSION

In this study we have reviewed the work done by the researchers for emotion detection using facial expressions and machine learning and deep learning in last five years. We have observed that by using machine learning and deep learning algorithm more than 80% accuracy have been achieve and CK+,JAFFE,MIMI dataset have been used frequently.

REFRENCES

[1] Reddy, C.V.R.; Reddy, U.S.; Kishore, K.V.K. Facial Emotion Recognition Using NLPCA and SVM. Traitement du Signal 2019, 36,13–22. [CrossRef]

[2] Zeng, N.; Zhang, H.; Song, B.; Liu,W.; Li, Y.; Dobaie, A.M. Facial expression recognition via learning deep sparse autoencoders.Neurocomputing2018, 273, 643–649. [CrossRef]

[3] Jain, D.K.; Zhang ZHuang, K. Multi angle optimal pattern-based deep learning for automatic facial expression recognition.PatternRecognit. Lett. 2020, 139, 157–165. [CrossRef]

[4] Sajjad, M.; Nasir, M.; Ullah, F.U.M.; Muhammad, K.; Sangaiah, A.K.; Baik, S.W. Raspberry Pi assisted facial expression recognition framework for smart security in law-enforcement services. Inf. Sci. 2019, 479, 416–431. [CrossRef]

[5] Xie, S.; Hu, H.; Wu, Y. Deep multi-path convolutional neural network joint with salient region attention for facial expression recognition. Pattern Recognit. 2019, 92, 177–191. [CrossRef]

[6] Nazir, M.; Jan, Z.; Sajjad, M. Facial expression recognition using histogram of oriented gradients based transformed features.Clust. Comput. 2018, 21, 539–548. [CrossRef]

[7] Li, K.; Jin, Y.; Akram, M.W.; Han, R.; Chen, J. Facial expression recognition with convolutional neural networks via a new face cropping and rotation strategy. Vis. Comput. 2020, 36, 391–404. [CrossRef]

[8] Saeed S, Shah AA, Ehsan MK, Amirzada MR, Mahmood A, Mezgebo T. Automated Facial Expression Recognition Framework Using Deep Learning. J Healthc Eng. 2022 Mar 31;2022:5707930. doi: 10.1155/2022/5707930. PMID: 35437465; PMCID: PMC9013309.

[9] Siam, Ali &Soliman, Naglaa&Algarni, Abeer&Abd El-Samie, Fathi&Sedik, Ahmed. (2022). Deploying Machine Learning Techniques for Human Emotion Detection. Computational Intelligence and Neuroscience. 2022. 10.1155/2022/8032673.

[10] Mayuri, A V &Manoharan, Ranjith&Subramani, Neelakandan&Manimaran, A. &Galety, Mohammad. (2022). Robust Facial Expression Recognition Using An Evolutionary Algorithm with a Deep Learning Model. Applied Sciences. 13. 468. 10.3390/app13010468.

[11] Jo, A-Hyeon&Kwak, Keun-Chang. (2023). Speech Emotion Recognition Based on Two-Stream Deep Learning Model Using Korean Audio Information. Applied Sciences. 13. 2167. 10.3390/app13042167.

[12] Yildirim, E., Akbulut, F.P. &Catal, C. Analysis of facial emotion expression in eating occasions using deep learning. Multimed Tools Appl (2023).

[13] Zhang, L.-M.; Li, Y.; Zhang, Y.-T.; Ng, G.W.; Leau, Y.-B.; Yan, H. A Deep Learning Method Using Gender-Specific Features for Emotion Recognition. Sensors 2023, 23, 1355.

[14] Umer, S., Rout, R.K., Pero, C. et al. Facial expression recognition with trade-offs between data augmentation and deep learning features. J Ambient Intell Human Comput 13, 721–735 (2022).

[15] Aouani, Hadhami, and Yassine Ben Ayed. "Speech emotion recognition with deep learning." Procedia Computer Science 176 (2020): 251-260.

[16] Imran, Ali Shariq, et al. "Cross-cultural polarity and emotion detection using sentiment analysis and deep learning on COVID-19 related tweets." Ieee Access 8 (2020): 181074-181090.

[17] Qazi, A.S.; Farooq, M.S.; Rustam, F.; Villar, M.G.; Rodríguez, C.L.; Ashraf, I. Emotion Detection Using Facial Expression Involving Occlusions and Tilt. Appl. Sci. 2022, 12, 11797.

[18] Song, Zhenjie. "Facial expression emotion recognition model integrating philosophy and machine learning theory." Frontiers in Psychology 12 (2021): 759485.

[19] Alsharekh, M.F. Facial Emotion Recognition in Verbal Communication Based on Deep Learning. Sensors 2022, 22, 6105.

[20] Liliana, DewiYanti. (2019). Emotion recognition from facial expression using deep convolutional neural network. Journal of Physics: Conference Series. 1193. 012004. 10.1088/1742-6596/1193/1/012004

[21] W. -S. Chu, F. De la Torre and J. F. Cohn, "Learning Spatial and Temporal Cues for Multi-Label Facial Action Unit Detection," 2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017), Washington, DC, USA, 2017, pp. 25-32, doi:10.1109/FG.2017.13.

[22] D. Acevedo, P. Negri, M. E. Buemi, F. G. Fernández and M. Mejail, "A Simple Geometric-Based Descriptor for Facial Expression Recognition," 2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017), Washington, DC, USA, 2017, pp. 802-808, doi: 10.1109/FG.2017.101.

[23] SiyueXie, Haifeng Hu, YongboWu,Deep multi-path convolutional neural network joint with salient region attention for facial expression recognition, Pattern Recognition, Volume 92,2019,Pages 177-191,ISSN 0031-3203,

[24] Connie, Tee & Al-Shabi, Mundher & Cheah, Wooi & Goh, Michael. (2017). Facial Expression Recognition Using a Hybrid CNN–SIFT Aggregator. 139-149. 10.1007/978-3-319-69456-6_12.

[25] Siddiqi, M.H., Ali, R., Khan, A.M. et al. Facial expression recognition using active contour-based face detection, facial movement-based feature extraction, and non-linear feature selection. Multimedia Systems 21, 541–555 (2015). https://doi.org/10.1007/s00530-014-0400-2