Turkish Online Journal of Qualitative Inquiry (TOJQI)
Volume 12, Issue 3, June 2021:171- 181

Research Article

Robust Facial Expression Classification using CNN and Multi-Layer Perceptron Network classifiers

Katta Pushpa Pujitha[1], Dr. G. Harinatha Reddy[2]

Abstract

Facial expression is the display of people's emotions, and it plays an important part in everyday conversation. As a result, Facial Expression Recognition (FER) is becoming an extremely important activity in modern society. Face identification, image extraction, and classification are the three phases of FER. While many FER systems are proposed to recognize only some face expressions, this paper proposes a recognition system for automatic face expression which recognizes all eight essential facial expressions (Normal, Cheerful, Anger, Scorn, Surprise, Sad, Terror, and Disgust). The approach is tested with the Extended CK+ dataset. The algorithm Viola-Jones is used for facial recognition in the presented process. A descriptor used for characteristics of descriptive face pictures is the Histogram of Oriented Gradients (HOG). The Primary Component Analysis (PCA) is used to achieve the most critical features to minimize feature dimensions. Finally, using two different classifiers, Multi-Layer Perceptron Neural Network (MLPNN) and Convolutional Neural Network (CNN), the presented approach categorized facial expressions, and the results were compared.

Keyword: *Facial Expression Recognition, Viola-Jones algorithm, Convolutional Neural Network, Multilayer Perceptron Neural Network*

[1]P.G. Scholar, Dept. of Elect. & Comm. Engg., N.B.K.R. Inst. of Sci. & Tech., A.P., India.
[2]Professor & Head, Dept. of Elect. & Comm. Engg., N.B.K.R. Inst. of Sci. & Tech., A.P., India.
Received Accepted

Introduction

Facial Expression Recognition is concerned with classifying facial expressions based on face pictures. It offers an accurate and reliable method for recognizing human emotions. The identification of human feelings from facial expressions has a long history of importance, inspired by Darwin's groundbreaking work. Facial expression recognition has many uses in a variety of areas. For e.g., measuring facial expressions in centimeters provides information that can be used to detect deception on any level [1]. Face detection has been a hotly debated research subject for the last two decades. Access monitoring, sophisticated human computer interaction, video detection, and predictive image indexing are only a few of the implementations [2]. The face, also

known as "the organ of feeling," is the most important "channel" of nonverbal communication. A series of facial expressions, such as a joyful smile, a sad or disapproving frown, wide-open eyes in surprise, or a curled lip in disgust, can communicate many subtle meanings. If machines can understand these signals, the robustness and harmony of human-machine interaction can increase. One of the most powerful nonverbal channels for Human Machine Interaction is Facial Expression Recognition (FER) [3].

Facial expressions are essential in human-to-human communication because they convey sentiment and meaning. Humans can recognize gestures and comprehend the feelings of others, while computers need massive computations to distinguish distinct expressions from their face. Machines that can interpret people's facial expressions can greatly aid humans. For example, if robots are capable of understanding people's intentions through facial expression recognition, they may provide more friendly service to humans. Furthermore, Facial Expression Recognition (FER) has promising applications in a variety of fields such as computer interfaces, health management, autonomous driving, and so on [4]. Feed-Forward Neural Networks are the most popular and are used in a wide range of applications. There are two types of network architectures based on the form of interactions between neurons: “Feed-Forward Neural Networks (FFNN)” and “Recurrent Neural Networks (RNN)”. If there is no "feedback" from the neurons' outputs to the network's inputs, the network is referred to as a "feed-forward neural network [5]." It was also shown that one hidden layer is sufficient to represent any Boolean function with arbitrary precision (where the accuracy is computed by the number of nodes in the hidden layer) and the hidden layer suffice for the arbitrary accuracy of every continuous feature (when the accuracy is determined by the number of nodes in the hidden layer). [6].

As the neural network training takes place on a large scale, there would be a dramatically decrease in the number of relevant network training algorithms [7], including the number of network parameters. A large-scale optimization task may be considered as learning, for instance, a large number of hidden layer weights in the neural network (MLP). The Viola-Jones algorithm was used in this analysis for face detection and was developed in 2001 by Paul Viola and Michael Jones. PCA was used to map this into a feature vector to remove redundant data and to obtain useful information. The PCA technique in this article reduces the dimensional characteristics of HOG. Finally, for speech processing, CNN and MLP approaches are used as classifiers.

Related Works

Hivi Ismat Dino et al. [8] introduced the technology for recognition of face automatic speech which were able to identify all eight fundamental facial phrases (usual, joyful, angry, disgusted, surprised, sad and afraid) and only few other FER devices were suggested. The approach is tested with the Extended CK+ dataset. The algorithm Viola-Jones is used for facial recognition in the presented process. The Histogram of Oriented Gradients (HOG) is used as a descriptor to exclude features from images of expressive faces. The Principal Component Analysis (PCA) is used to

decrease the dimensionality of functions to achieve the most important features. Finally, three separate classifiers compared the results of the presented method: K-Nearest Neighbor, Support Vector Machines (SVM) and MLPNN.

K. Khorasani et al. [9] suggested a methodology and applied on a database of 60 photographs of males, each with five different facial expressions. For network preparation, the other 20 photographs of men are used for widespread and testing purposes. In network verification and generalization for four face expressions, the results of the trained network are tested with uncertainty matrices (smile, anger, sorrow, and surprise). The highest identification standards for teaching have been discovered and generalizing pictures are 100 percent and 93.75 percent (without rejection), respectively.

Daniel Svozil et al. [10] derived that the objective functions partial derivatives with respect to the weight and threshold coefficients are calculated. These derivatives are useful for the neural network's adaptation mechanism. It is explored how to train and generalize multi-layer feed-forward neural networks. The basic back-propagation algorithm is improved upon. A multi-layer feed-forward neural network example is provided for predicting carbon-13 NMR chemical shifts of alkanes. Additional neural network uses in chemistry are discussed. The benefits and drawbacks of multilayer feed-forward neural networks are discussed.

Nisha Thomas et al. [11] cast-off a program to recognize human facial expressions is an intriguing research project. In this paper, a framework for identifying facial expressions based on feed forward neural networks is presented with the aim of creating an intelligent system. The suggested scheme identifies four major emotions: positive, depressed, surprised, and neutral. The neural network achieved a score of 0.0198 against a target of 0.0200.

Yegui Xiao et al. [12] proposed a new FER solution built on the Feed Forward Network (FFN) decision tree and a full-size facial picture 2-D DCT (Discrete Cosine Transform) approach. A number of facial expressions with representatives of "smile", "surprise" and other expressions of "pain" and "tradition", are distinguished by the initial NN-based node on the Decision tree. This node will assist in resolving the confusion between the tier members of the two classes. Each group has two NN-based nodes that separate their two members according to the first node.

I. Data set

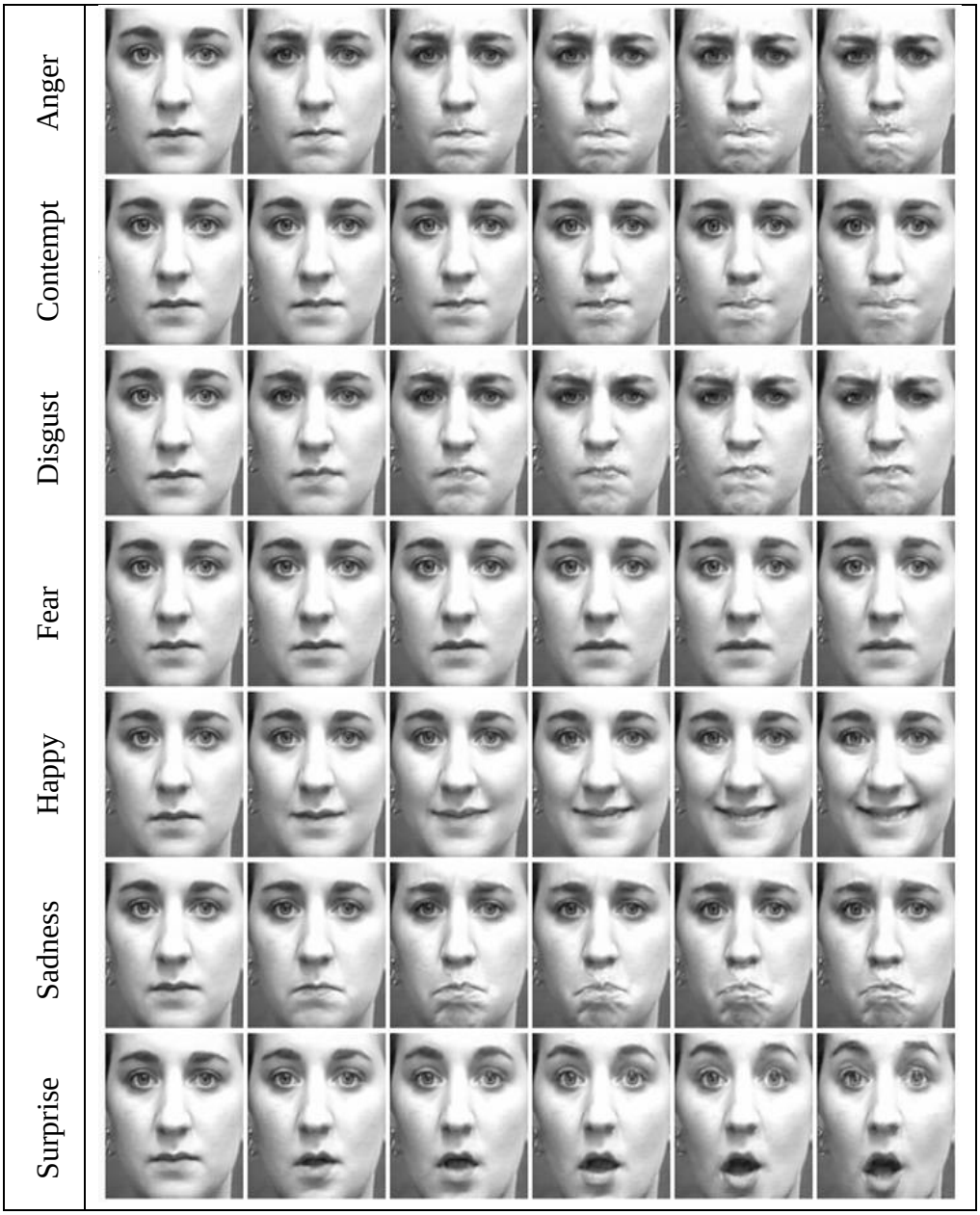

Fig. 1. CK+ Facial Expression Database with seven expressions extending from surprise, sadness, happy, fear, disgust, contempt, anger

The Extended Cohn-Kanade (CK+) Dataset [13] shown in Fig. 1. was used in this article. It comprises 123 individual face images (grey and or color values), which includes 593 video sequences in the age group of 18 to 50 years. Video sequences are captured at 30 FPS (Frames Per Second) with resolution of 640 x 490 or 640 x 480 pixels from the expression Neutral to seven target expressions like anger, contempt, disgust, fear, happiness, sadness and surprise.

Proposed System

Each person's emotions are represented in the dataset. The dataset is made up of image sequences that have been translated from video frames. The dataset contains a large number of incomplete mark sequences. In our experiment, we considered and used the branded feeling. A standard emotion was used as the first image frame in the dataset, and a named emotion was used as the last image frame. Our study is aimed at naming 8 human emotionals – neutral, annoying, disgusting, fearful, happy, sorrowful and surprising. Fig. 2 displays the current scheme block diagram.

i. Preprocessing

Any image processing technique's first step is referred to as this. We used the algorithm of Viola-Jones to detect faces in this analysis. In the original files, 8-bit gray or 24-bit color values were digitized into 640x490.

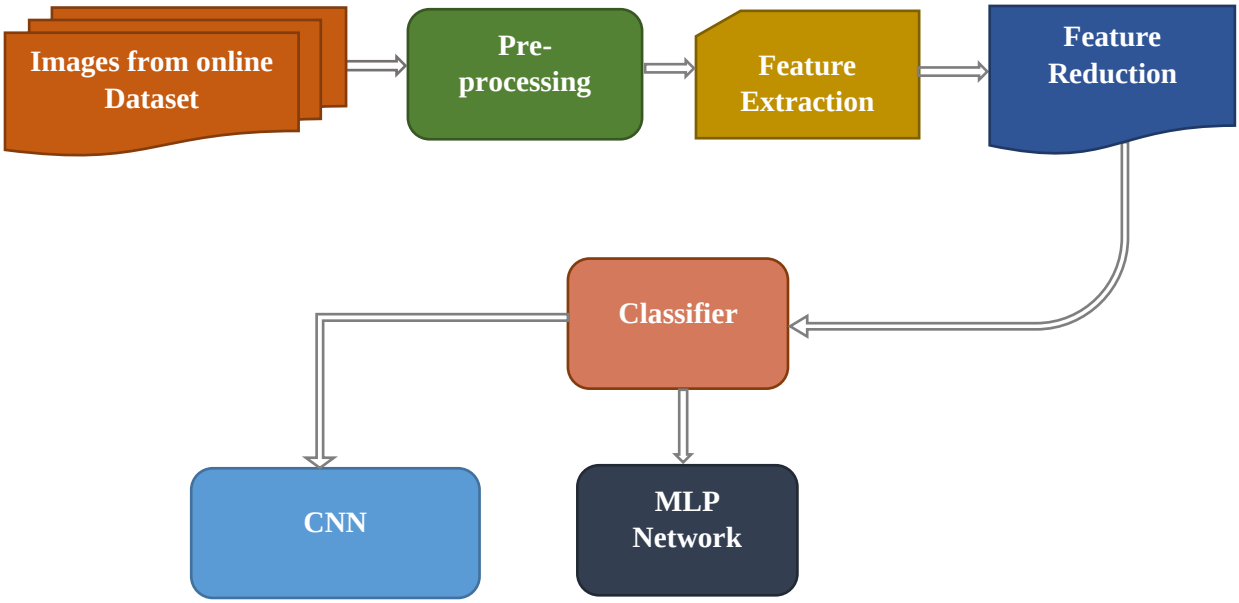

Fig. 2. Block Diagram of proposed Facial Expression classification Model

Viola-Jones is one of the most commonly deployed face recognition algorithms thanks to its robust and real-time facial detection capabilities. The four phases of the algorithm include the collection of hair functions, integrated image production, ad boost preparation and cascade category. After identification, the images were normalized and resize to 256 gray pixel arrays. The total number of feelings labeled and faces identified as seen in Table 2. Fig. 3 represents the expressions of various emotions on the sensed tongue. Using the following formula the hair transformation of a number of samples is calculated:

Algorithm 1. Viola-Jones Algorithm

1. *Take into account the number of n/2 pairs (a, b)*
2. *For each pair,*
   compute (a+b)/sqrt(2),
   which will constitute the first half of the output set of these values.
3. *For a pair compute (a - b)/sqrt(2);*
   these are the second half values.
4. *For the first half of the array, repeat steps 1–4.*
   (The duration of the series should be two-powered.)

Fig. 3. Sample of nominal face images from the database.

Table 1. Face Emotions detected in whole

Value	Count	Percentage
Neutral	316	49.9 %
Rage	44	7.2 %
Contempt	17	2.7%
Disgust	54	8.8 %
Terror	23	3.7 %
Happiness	67	10. 8%
Sorrow	25	4.0 %
Surprise	80	12.9 %
Total	**626**	**100. 0%**

ii. Feature extraction

This is the main step of the FER approach. FER's efficiency is based on the strategies used at this stage. Local Binary Pattern (LBP), Gabor filter, PCA, and other feature retrieval methods are used to. For feature extraction, we used the HOG algorithm proposed by Dalal and Higgs in this work.

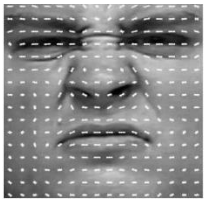

Fig. 4. HOG characteristics extracted from a grayscale input image

The basic concept behind the HOG descriptor is that assigning gradient strength or edge directions represents a local subject's shape and presence in a picture. HOG is a histogram that incorporates histograms with different gradient directions. Local histograms may be construct-normalized to improve accuracy. You can select a larger image area called the block, calculate the power measurement and normalize the cells inside the block by using the derived value. Fig. 4. shows the histogram extract of directed gradients. The cumulative number of HOG features omitted is 8109 and it is a big number.

iii. Feature Reduction

Furthermore, reduced dimensions contribute to loss of information, while PCA helps to decrease information losses. PCA is a common technique for data reduction in signal processing and computer pattern recognition. Since the pattern typically includes a lot of redundant information, PCA is used to convert it into a functional vector that eliminates redundant information and leaves only valid information. The functionality derived is to differentiate the input patterns. A 256x256 pixel face picture provides 8109 HOG characteristics in two measurements, which can be used as a one-dimensional characteristic vector. To reduce dimensionality, PCA was used on the HOG function; we suggest retaining 90% of the original detail. On each study, the final feature size was 245 pixels.

iv. Classifier

Using a special classifier, the derived attributes from facial representations are allocated to the corresponding classes of face expressions. MLP Neural Networks is the most commonly used classifiers for classification. In this work, we used MLP evaluated its performance.

Multi-Layer Perceptron

Because of their high accuracy, neural networks are commonly used. When using Multi-Layer Perceptron (MLP) neural networks to solve problems, one of the most important factors to note is the number of hidden neurons. However, it is impossible to predict a suitable structure that will ensure convergence. As a result, many researchers concentrate their studies on positive learning.

The face recognition method built on the "MLP" architecture consists of two steps: preparation and checking. The "MLP" networks used in this study were trained using the backpropagation algorithm. The output signal moves into the sigmoid activation system, and the backpropagation network goes into supervised learning. The two equations below were used to train the "MLP" network:

$$\Delta wji(t) = -\eta \frac{\partial Ep(t)}{\partial Wij(t)} \qquad (1)$$

$$Ep = \frac{1}{2}\sum_k (dpk - spk)^2 \qquad (2)$$

Wherever k is the learning rate of the MLP outputs and the number of neurons. For the p^{th} training vector, the predicted and real network outputs are dp and p. Ep reflects the p^{th} sequence network malfunction (Mean Square Error "MSE") dp and sp are expected and real p^{th} training vector network outputs.

Algorithm 2. MLP Neural Network Algorithm

1. *From the "TD,"*
   choose (N cl ini) classes (N classes = N cl ini),
   then (N input = 80N classes).
2. *Creation of the "MLP" compound,*
   composed of secret neurons (N hid = Nh ini).
3. *Assign arbitrary values to the neuron relation and bias weights.*
4. *To obtain the predefined parameter, use the backpropagation*
   algorithm to train the "MLP" with
   (N data) input patterns chosen from the "TD".
5. *Go on to step 6 whether the training algorithm will reduce the "MSE"*
   to "" or to step 9.
6. *If sum of N input N is greater than zero in N:*
   the total number of "TD" patterns would be in step 7, or step 12.
7. *Increase N-cl ini input pattern number by increasing*
   the number of "TD" classes (N-class = N-classes + N-cl ini) (N input)
8. *Increase output neurons number N out = N groups,*
   then set new output neurons to weight 8;
9. *Keep the last weight ties in a safe place.*
10. *Add one more secret neuron to the mix (N hid=N hid + 1).*
11. *Set the weights of the current secret neuron and proceed to Step 4.*
12. *Select the architecture with the least amount of generalization error*
    and the fewest number of neurons.

The proposed method was validated using the CK+ dataset, with 69 percent of the participants being female, 13 percent Afro-American, 81 percent Euro-American, and 6% from other ethnic

groups. CK+ is made up of 593 image sequences of 123 subjects' facial expressions, 326 of which have eight emotion codes, and the image resolution is 640x490 or 640x480. After preprocessing, there were 626 labelled images of eight feelings.

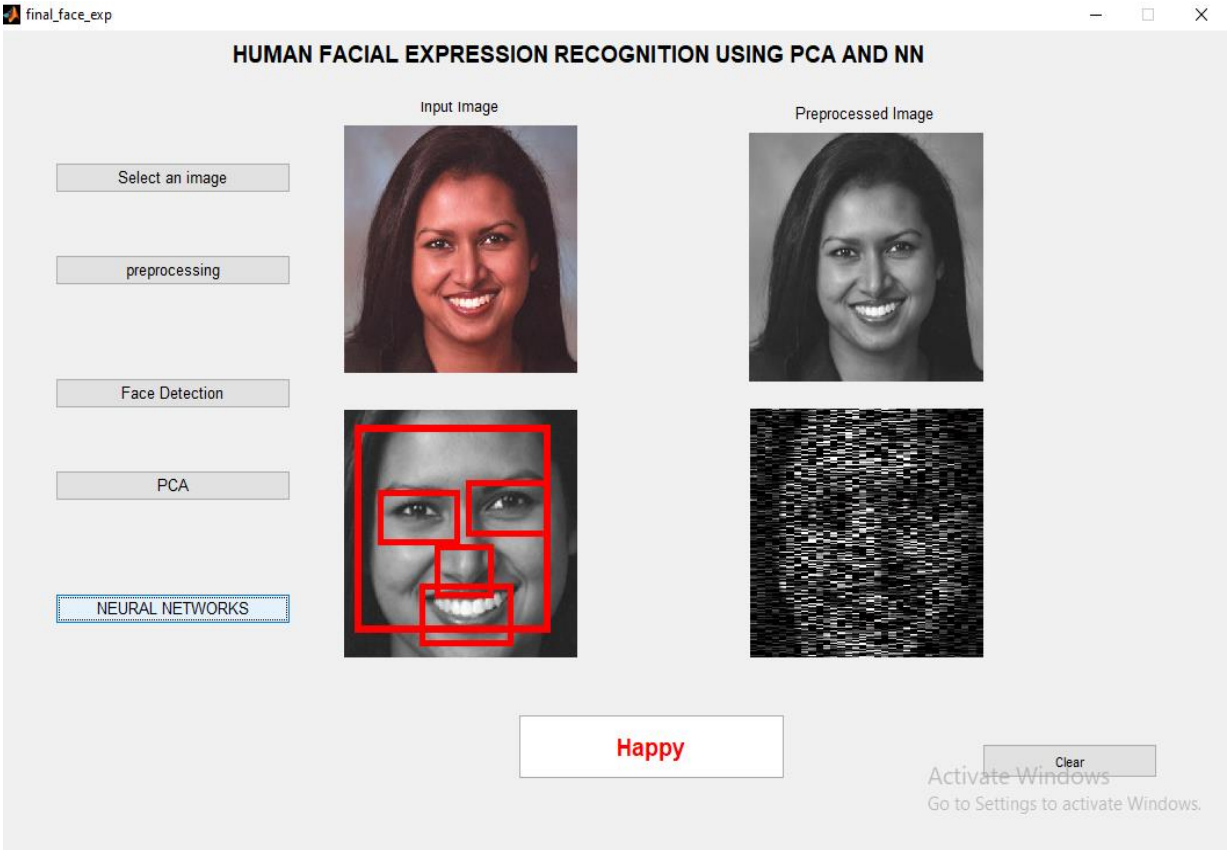

Fig. 5. Output screen shot of Facial Expression model using CNN and MLP

Table 2. Accuracy, Specificity, Recall and Precision of CNN and MLP feedforward neural network

Metrics	CNN average from 8 emotions	MLP average from 8 emotions
Specificity	96. 42 %	97.62 %
Recall	86.53 %	87.98 %
Precision	93.45 %	95.61 %
Accuracy	95.60 %	96.89 %

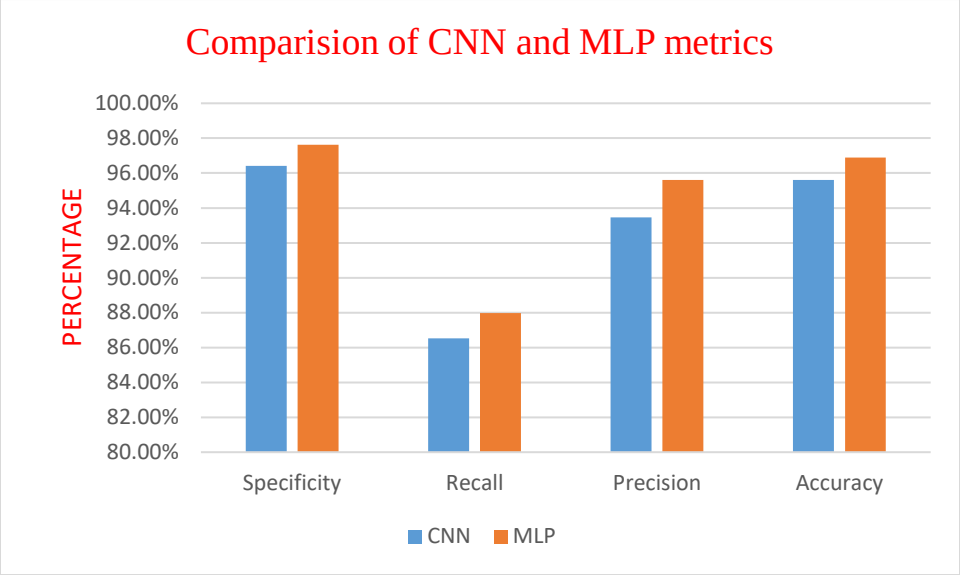

Fig. 6. Accuracy, Specificity, Recall and Precision of CNN and MLP feedforward neural network.

Table 3. Comparison of Accuracy for proposed method with the state of art methods

Model	**Accuracy**
Pramerdorfer et al. [14]	75.2%
Tang et al. [15]	71.2%
R. Zatarain-Cabada et al. [16]	83%
Proposed method using CNN	95.60%
Proposed method using MLP	96.89%

With proper description from table 3, the MLP feedforward neural network classifier has an accuracy rate of 96.89 percent, and CNN classification has an accuracy rate of 95.60 percent. We can infer from the above fig. 6 that the average accuracy score of the MLP feedforward neural network classifier using eight emotions is better than that of the CNN classifier.

Conclusion

In this study, a FER method was proposed based on the history of directed gradients and a variety of algorithms for machine learning. The Extended Cohn-Kanade (CK+) dataset was used as a suitable database to investigate the characterization of human facial expressions. PCA was used to limit the dimensionality of features. This paper describes a device that can interpret all eight basic facial expressions. CNN and the MLP feedforward neural network were used as classifiers for facial expressions. The findings of the experiments reveal that the MLP feedforward neural network is a stronger classifier, with an accurate classification rate of 96.89.

References

[1] Lu, Y. Z., & Wei, Z. Y. (2004, August). Facial expression recognition based on wavelet transform and MLP neural network. Int. Conf. on Signal Processing, 2004.. (Vol. 2, pp. 1340-1343). IEEE.
[2] Boughrara, H et al. MLP neural network using modified constructive training algorithm: application to face recognition. In International Image Processing, Applications and Systems Conference (pp. 1-6). IEEE.
[3] Gurukumar Lokku, H. R. G. (2020). Discriminative Feature Learning framework for Face Recognition using Deep Convolution Neural Network. Solid State Technology, 63(6), 18103-18115.
[4] Qiu, Y., & Wan, Y. (2019, December). Facial Expression Recognition based on Landmarks. In 2019 IEEE 4th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC) (Vol. 1, pp. 1356-1360). IEEE.
[5] Sazlı, M. H. (2006). A brief review of feed-forward neural networks.
[6] Guru Kumar Lokku, G. Harinatha Reddy, M. N. Giri Prasad, "Automatic Face Recognition for Various Expressions and Facial Details," International Journal of Innovative Technology and Exploring Engineering (IJITEE), Volume 8, Issue 9S3, Pages 264-268, July 2019. https://doi.org/10.35940/ijitee.I3048.0789S319
[7] Kumar, D. Maruthi, and K. Kannaiah. "A Conceal Fragment Visible Image Broadcast Through Montage Images with Revocable Colour Alterations." Emerging Trends in Electrical, Communications, and Information Technologies. Springer, Singapore, 2020. 673-682.
[8] Dino, H. I., & Abdulrazzaq, M. B. (2019, April). Facial expression classification based on SVM,KNN and MLP classifiers. In 2019 International Conference on Advanced Science and Engineering (ICOASE) (pp. 70-75). IEEE.
[9] Ma, L., & Khorasani, K. (2004). Facial expression recognition using constructive feedforward neural networks. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 34(3), 1588.
[10] Svozil, D., Kvasnicka, V., & Pospichal, J. (1997). Introduction to multi-layer feed-forward neural networks. Chemometrics and intelligent laboratory systems, 39(1), 43-62.
[11] Thomas, N., & Mathew, M. (2012, February). Facial expression recognition system using neural network and MATLAB. In 2012 Int. Conf. on Computing, Comm. and Applications (pp. 1-5). IEEE.
[12] Xiao, Y., Ma, L., & Khorasani, K. (2006, July). A new facial expression recognition technique using 2-D DCT and neural networks based decision tree. In The 2006 IEEE Int. Joint Conference on Neural Network Proceedings (pp. 2421-2428). IEEE.
[13] P. Lucey et.al, ''The Extended Cohn–Kanade dataset (CK+): A complete dataset for action unit and emotion-specified expression,'' in Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Workshops, San Francisco, CA, USA, Jun. 2010, pp. 94–101.
[14] Pramerdorfer, C., Kampel, M.: Facial expression recognition using convolutional neural networks: state of the art. Preprint arXiv:1612.02903v1, 2016
[15] Y. Tang, "Deep Learning using Support Vector Machines," in International Conference on Machine Learning (ICML) Workshops, 2013.
[16] R. Zatarain-Cabada, M. L. Barrón-Estrada, F. González-Hernández and H. Rodriguez-Rangel, "Building a Face Expression Recognizer and a Face Expression Database for an Intelligent Tutoring System," 2017 IEEE 17th International Conference on Advanced Learning Technologies (ICALT), 2017, pp. 391-393, doi: 10.1109/ICALT.2017.141.