Four-layer Convnet to Facial Emotion Recognition With Minimal Epochs and the Significance of Data Diversity

Tanoy Debnath
Mawlana Bhashani Science and Technology University
Md. Mahfuz Reza
Mawlana Bhashani Science and Technology University
Anichur Rahman
Mawlana Bhashani Science and Technology University
Shahab S. Band
National Yunlin University of Science and Technology
Hamid Alinejad-Rokny (h.alinejad@unsw.edu.au)
UNSW Sydney

Research Article

Keywords: Convolutional Neural Network (CNN), Emotion Recognition, Facial Expression, Classification, Accuracy

Posted Date: May 17th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-511221/v1

Four-Layer ConvNet to Facial Emotion Recognition with Minimal Epochs and the Significance of Data Diversity

Tanoy Debnath[1], Md. Mahfuz Reza1[1], Anichur Rahman[1,2], Shahab S. Band[*3], and Hamid Alinejad-Rokny[*4,5,6]

[1] Department of Computer Science and Engineering, Mawlana Bhashani Science and Technology University,Tangail, BD.
[2] Department of Computer Science and Engineering, National Institute of Textile Engineering and Research (NITER), Savar, Dhaka, BD.
[3] Future Technology Research Center, College of Future, National Yunlin University of Science and Technology 123University Road, Section 3, Douliou, Yunlin 64002, TW
[4] BioMedical Machine Learning Lab, The Graduate School of Biomedical Engineering, UNSW Sydney, Sydney, NSW, 2052, AU.
[5] Core Member of UNSW Data Science Hub, The University of New South Wales (UNSW Sydney), Sydney, NSW,2052, AU.
[6] Health Data Analytics Program Leader, AI-enabled Processes (AIP) Research Centre, Macquarie University, Sydney, 2109, AU.
[*]To whom correspondence should be addressed. E-mail: S.S.B (shamshirbands@yuntech.edu.tw) H.A.R (h.alinejad@unsw.edu.au)

ABSTRACT

Emotion recognition defined as identifying human emotion and is directly related to different fields such as human-computer interfaces, human emotional processing, irrational analysis, medical diagnostics, data-driven animation, human-robot communication and many more. The purpose of this study is to propose a new facial emotional recognition model using convolutional neural network. Our proposed model, "ConvNet", detects seven specific emotions from image data including anger, disgust, fear, happiness, neutrality, sadness, and surprise. This research focuses on the model's training accuracy in a short number of epoch which the authors can develop a real-time schema that can easily fit the model and sense emotions. Furthermore, this work focuses on the mental or emotional stuff of a man or woman using the behavioral aspects. To complete the training of the CNN network model, we use the FER2013 databases, and we test the system's success by identifying facial expressions in the real-time. ConvNet consists of four layers of convolution together with two fully connected layers. The experimental results show that the ConvNet is able to achieve 96% training accuracy which is much better than current existing models. ConvNet also achieved validation accuracy of 65% to 70% (considering different datasets used for experiments), resulting in a higher classification accuracy compared to other existing models. We also made all the materials publicly accessible for the research community at: https://github.com/Tanoy004/Emotion-recognition-through-CNN.

Keywords: Convolutional Neural Network (CNN), Emotion Recognition, Facial Expression, Classification, Accuracy.

Introduction

The face is also known as the mental core. As an assortment of facial gestures, the face can give several minimal signals. These exquisite signals can make human-machine interaction more secure and harmonious when interpreted by computers. A good source of knowledge for ordering an individual's true emotions[1] was argued for facial expressions. Recognition of facial expression (FER) is one of the most critical non-verbal processes by which human-machine interface (HMI) systems can understand[2] human intimate emotions and intentions. This scheme is a classification task. The classifier takes as input a set of characteristics that are derived from the input image, which is simply shown in Fig.1.

Gabor wavelet transform[3], Haar wavelet transform[4], Local Binary Pattern (LBP), and Active Presence Models (AAM)[5] are the feature extraction methods based on static images. Whereas dynamic-based[6–8] approaches assume the temporal association in the sequence of input facial expression within clinging frames. Support Vector Machine (SVM), Hidden Markov Model, AdaBoost, and Artificial Neural Networks (ANN)[9] are widely used scheme for facial expression recognition. A major

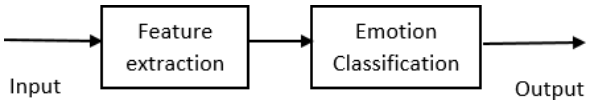

Figure 1. Facial expression recognition system

advancement in the field of deep learning and the implementation of CNN has been quite promising[10,12,13]. However, a massive issue with the use of deep learning is that a large amount of data is required to learn successful models.

While some improvement in the identification of facial expression has been made by the CNN algorithm, some detachments are still present, including too long training times and low recognizing rates in the complex environment[9]. In existing databases, two challenges have been observed in deep learning achievements in FER methods: 1) a low number of images, and 2) images taken from heavily structured conditions. These concerns inspired the creation of FER techniques focused on the set of Web images[14,15]. The present work focuses mainly on the creation of a multimodal, intelligent HMI system that operates in a real-time context.

This research aims to determine the emotion of a facial emotional input image. In this paper the authors do a more reliable and detailed study on deep learning both for static and dynamic FER tasks until 2020. This study is concerned with the creation of an automated facial expression recognition (AFER) system in the domain of facial expression using Convolutional Neural Networks (CNNs) and improving the accuracy of this mission. Orthodox machine learning algorithms used for handcrafted features typically have equivalents that do not have the durability to reliably interpret a task[16]. This is a fair starting point for us to examine the use of CNN features, since with CNN-based models[12,17], we have obtained the best solutions to recent FER-relevant tasks.

Facial recognition requires several phases: detection of face images, preprocessing of face images, retrieval of facial features, alignment of face images, and identification of face images. There are primarily two types of extraction of features: one is geometric attribute extraction, and the other is a procedure which focused on total statistical characteristics. To describe the location of facial organs as the features of the classification[18] , the geometrical feature-based approach is widely used.
This paper aims at creating a method for the use of CNN to build a FER scheme. The presented model can be used in real-time using a webcam to categorize human faces. The contributions to this paper are as follows:

- The authors suggest a CNN method for recognizing seven facial expressions and real-time detection of facial expressions and also testing their precision based on features derived from convolution neural networks.
- This research reveals that combining images from different databases helps to increase generalization and to improve the accuracy of teaching.
- It can contain enhanced testing techniques, such as preparation, testing and validating processes, and provides findings that reflect greater consistency by longer training sets instead of training and testing sets.
- This work achieves a training accuracy of over 90 percent in a minimal number of epoch, showing that the model is well adjusted to the method.

This work aims to create a model that can classify seven distinct emotions: happy, sad, surprise, angry, disgust, neutral, and fear, and to achieve better accuracy than the baseline 14%[19]. Besides this, the main goal of this research is to examine and understand the advantages of using deep convolutional neural network models over other deep learning models.

Background Knowledge and Literature Reviews

Background Study

Analyse of Facial Expression

Automatic facial expression analysis (AFEA) can be used in many areas, including relational and semantic communication, clinical psychology, psychiatry, neurology, pain assessment, lie detection, intelligent settings, and multimodal human-computer interface (HCI). Face collection, facial data extraction and representation, and recognition of facial expression are three steps of the standard approach to AFEA composition, as depicted in Fig. 2. There are mainly two types of techniques for facial feature extraction: geometric or predictive feature-based methods and methods based on appearances. The authors use statistical appearance-based approaches in this article.

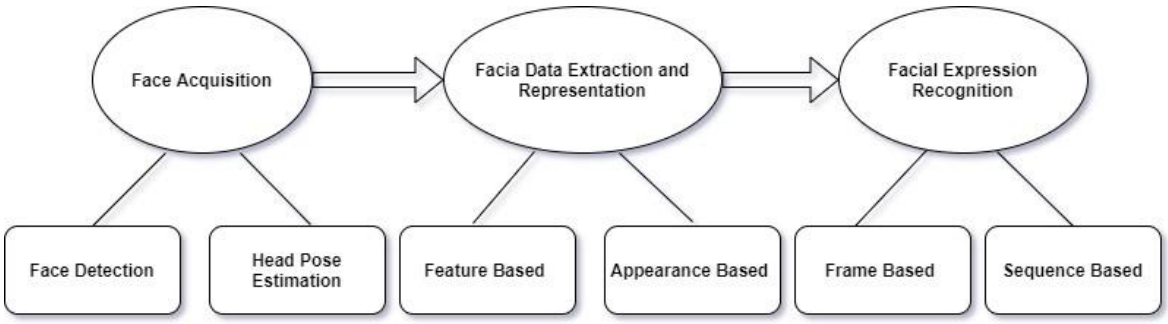

Figure 2. The basic framework of applications for facial expression analysis

Facial Emotion Recognition (FER)

Face detection is a key role in FER. There are different strategies to face recognition, including the expression-based approach, the framework approach, the feature-based approach, and the neighborhood graph approach[20]. The three-stage flow map of the facial expression recognition process seen in the Fig. 3.

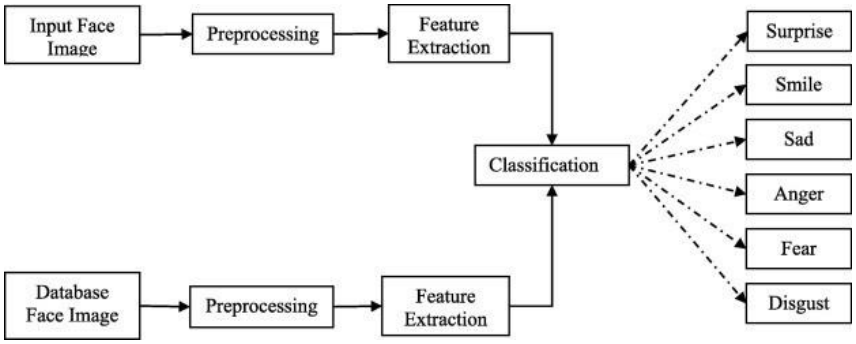

Figure 3. Summary flowchart for the three phases of the FER method[21]

Structure of CNN

At first, we see the basic structure of a neuron unit in Fig. 4 which we can understand well and relate to it with the structure of a CNN. The Convolutional neural network architecture consists of the following layers[22]:

- Convolution Layer:
  In the convolution layer, a filter is used to recognize the special features or patterns present in the original image (input). It is normally represented as a matrix (MxMx3) with a smaller dimension.

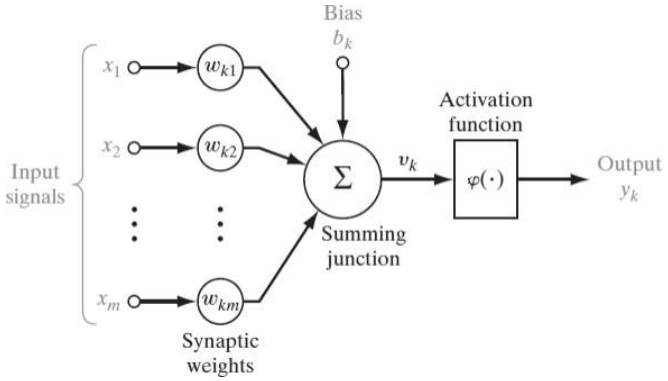

Figure 4. The Illustration of a Neuron unit[23]

- Max Pooling:
  This is performed by the use of filters that slide through the input; the maximum parameter is taken out at each step and the rest is lowered. Along with, the network is simply down-sampled by this.
- Fully connected layer:
  In this layer, all activation's on the previous layers are connected to the neurons.
- Activation function:
  Functions for activation are used to decrease over-fitting. A ReLu activation function has been used in the CNN architecture. The value of the activation function of ReLu is that its gradient is always equal to 1 (shown in equation 2), meaning that during back-propagation, most of the error is transferred back[24,25].

$$f(x) = max(0, x) \tag{1}$$

- Softmax:
  Softmax is implemented through a neural network layer just before the output layer. As the output layer, the Softmax layer must have the same number of nodes.
- Batch Normalization:
  Batch normalizer speeds up the training process and adds a transition which keeps the mean activation near 0 and the standard activation deviation close to 1.

Literature Reviews

Facial communication studies have been carried out for years. But for any experiment, there was still room for progress. That is why this topic is convenient. The key objective of the researchers is to enhance the precision of a basic data collection FER2013 in[26]. The authors have used CNN as the methodology for their proposed model to define seven critical emotions. While overall accuracy has been obtained at 91 percent, the identification rate is only 45 percent and 41 percent respectively in classifying disgust and fear.

In[27], the writers have identified facial expressions based on CNN. In comparison to other approaches, the proposed FER approach is a major challenge in machine learning, focused on mixed instances taken from multiple datasets. Experimental results reveal that the six universal expressions can be specifically defined by the FER process. In the recent past, a multimodal speech emotion recognition and classification techniques have been proposed by A.Christy and colleagues[28]. For classification and prediction, algorithms such as linear regression, decision tree, random forest, support vector machine (SVM) and convolutional neural networks (CNN) are used in this article. The authors tested their model with the RAVDEES dataset and, compared to decision tree, random forest and SVM, CNN showed 78.20 percent accuracy in identifying emotions. In[29], without needing any pre-processing or feature extraction tasks, the authors demonstrate the classification of FER based on static images, using CNNs. In a seven-class classification assignment, the authors obtained a test accuracy of 61.7 percent on FER2013 compared to 75.2 percent in the state-of-the-art classification. Wang and colleagues[30] proposed a novel concept of EFDMs with STFT based on multiple channel EEG signals. The pre-trained model was then introduced to DEAP with a few samples by profound model transmission, resulting in 82.84 percent accuracy on average. Jung and associates[7] investigated FER with a profound learning approach, which integrates two deep networks that derive faces appearance (using convolutional layers) and geometric features from face landmarks (using completely linked layers), with a 97.3 percent accuracy of CK+ findings. The authors suggested a computer vision FER method in[31]. In the process, the gray-scale face picture was consolidated into a 3-channel input with the corresponding basic LBP and an average LBP feature map. This thesis won the Emotiw 2017 award with the best submission reaching 60.34 percent accuracy. Francesca Nonis and colleagues[32] suggested 3D approaches and problems in FER Algorithms. This research would address the problem of facial identity through the interpretation of human feelings, focusing on 3D approach, grouping and arranging all the works and different techniques. The average accuracy of recognition of expressions varies from 60 percent to 90 percent. Certain expressions, such as anger and fear, have usually the lowest recognition levels.

A facial expression recognition system has been introduced by N. Veeranjaneyulu[33] in which facial characteristics by use of deep neural features much better than handcrafted ones are. The extraction function is conducted using the VGG16 algorithm and deep CNN models are classed. The suggested accuracy of the system is demonstrated by the CK+ dataset. In[34], the authors have introduced a 3-dimensional neural video emotion detection network. The findings are contrasted with cross-validation approaches. The crafted 3D-CNN generates 97.56 percent with the cross-validation of Leave-on-Sujet-Out, and 100 percent with 10 times CK+ and 84.17 percent with 10 times Cross-validation on Oulu-CASIA.

The analysis of facial expression recognition has some drawbacks, according to the authors. Such as the model's use and lack of friendliness, inability to catch feelings or actions in complex contexts, participant shortness with a need for more accuracy, a deficit in detecting effectiveness about EEG signals, and so on. Although there has been more research on combining impact identification and usability testing, their functional applicability and interference with user experience testing still need to be analyzed further.

Proposed Architecture and Methods

Convolutional Neural Networks are a form of deep neural network that is used for computer vision and visual image processing. However, conventional algorithms face certain serious issues or questions, such as luminous variance and location variance, etc. The approach to addressing the problems of conventional methods is to apply the CNN algorithm to the classification of emotions[22]. In contrast, our model is sequentially structured. We recognize that Sequence Modeling has the ability to model, analyze, make predictions or produce some form of sequential data. In comparison to the traditional form, the algorithm's major distinctions are:

1. **Automatic Feature Extractor Procreation:**
   Without any user or a built-in feature extractor, image features can be extracted manually, as the feature extractors are generated during the training process.
2. **Differences of Mathematical Model:**
   Typically, the linear classifier is classified by linear transformation. This is commonly referred to as the traditional form. In contrast, to discern variations in the classification process, CNN and other Deep Learning algorithms usually incorporate linear conversion with nonlinear features such as sigmoid and rectified linear unit functions (ReLU).
3. **The Deeper Structure[23]:**
   The traditional approach usually conducts only one layer of an operation via the linear classifier: SVM has just one weight set, for instance (shown in equation 1). However, in the course of classification, CNN and other deep learning algorithms perform several layers of operation. As a two-dimensional array, CNN adopts input.

$$S = W \times x_i + b \tag{2}$$

Where the classification score is S, W is the matrix of weights, and b is bias.

CNN Model Overview

The proposed model consists of four layers of convolution together with two layers that are fully connected which we can see in the Fig. 5. Many building blocks are stacked up by CNN: convolution layers, pooling layers (e.g., max-pooling), and fully connected layers (FC). The reaction network is rooted in the convolution layer, where the network functions are learned. In addition, the convolution layer have weights that need to be taught, whereas the pooling layers use a fixed function to convert the activation. The efficiency of the convolution layer goes through loops. Furthermore, model output is calculated on a training dataset with the loss feature and learning parameters (kernels and weights) are adjusted by back-propagation with the loss. This work requires to incorporate or delete certain layers during training cycles, such as Max Pooling or Convolution layer, to build something unique and useful under specific kernels and weights for the output of a model.

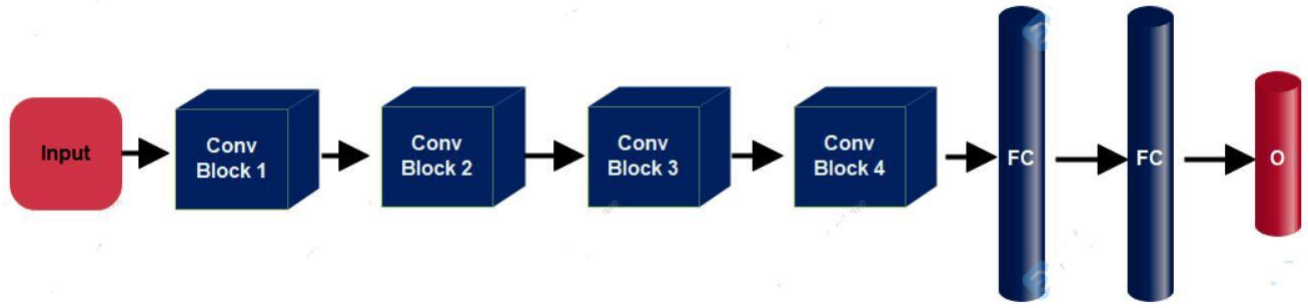

Figure 5. Reference CNN model for FER[35]

Fine Tuning for Proposed CNN

Not only does a fine-tuning methodology replace the pre-trained model's fully connected layers with a new set of fully connected layers to train up on a given dataset, but it also fine-tunes all or part of the kernels in the pre-trained convolutional layer base by way of backpropagation. It is possible to fine-tune all the layers in the convolutional base or set some earlier layers while fine-tuning much of the deeper layers. In this work, the proposed model consists of four layers of convolution together with two layers that are completely connected. This task would only need to train the high-level detailed feature block the essence and the completely connected layers that regard[12] as a classifier. In contrast, since we have just 7 emotions, the authors reset the Softmax ranking to 7 grades from 1000 ranks.

Pipeline for Proposed CNN

The network with a layer for processing the input. Here are four convolution and additional pooling layers and two fully connected layers that is completely associated at the end. A ReLU layer, batch normalization, and a dropout layer is used for any convolution and a fully connected layer of all the four network structures. The additional dense layer is used at the end of the four convolution layers which are associated with the two fully connected layers. Besides, the overall pipeline for the proposed CNN model is architectured on the following Fig. 6.

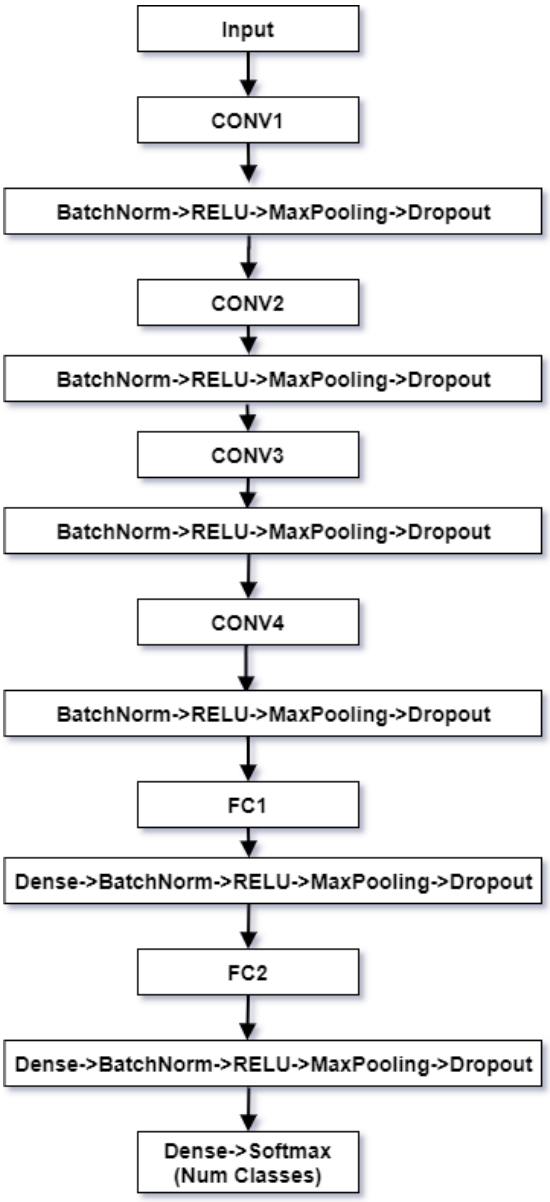

Figure 6. Pipeline for proposed CNN

Proposed Methodology

The data collection used for the application was the FER2013 dataset from the Kaggle challenge on FER2013[10]. The database is used to incorporate the Facial Expression detection framework. The dataset consists of 35,887 pictures, split into 3589 experiments and 28709 images of trains. The dataset includes another 3589 private test images for the final test. Fig. 7 shows the examples of seven basic emotions from the used dataset and Fig. 8 shows the expression distribution of the FER2013 dataset. The pictures in the data set of FER2013 are 48x48-sized and black and white. The use of this data set is split into private, public, and training evaluations, and 28709 training data sets.

Figure 7. The examples of seven basic emotion[11]

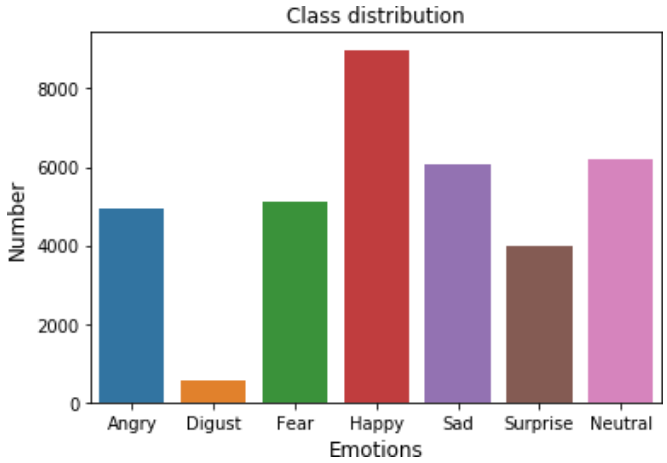

Figure 8. Expression distribution

The four major points of the proposed methodology are discussed here:

Recognition of Facial Expression

The FER mechanism has three steps. Firstly, the step of prepossessing is to prepare the dataset into a shape. The new form will run and produce effective results on a generalized algorithm. Secondly, the face is identified from the images collected in real-time in the feature extraction step. Finally, to group the picture into one of seven classes, the emotion classification stage consists of applying the CNN algorithm. Moreover, these main phases are represented using a flowchart. The system flowchart of emotion classification for the FER approach is seen in Fig. 9.

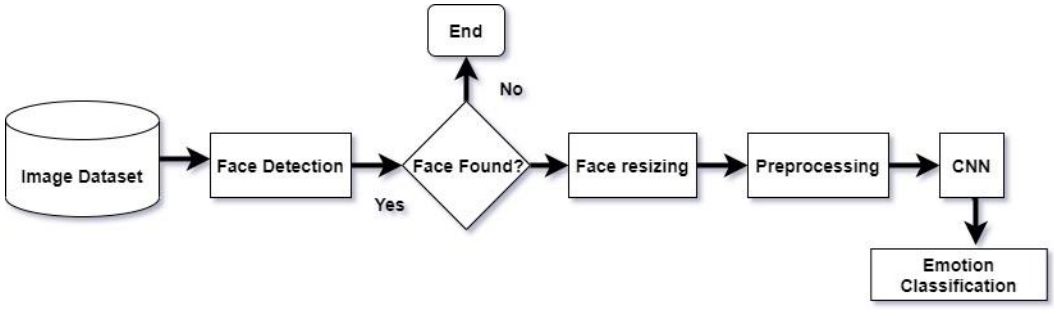

Figure 9. System flowchart of emotion classification

Preprocessing

Preprocessing is a required step in computer vision for image processing. The picture entered in the FER may include noise and some light, scale and color variations. Apart from this, any preprocessing operations in the[36] picture have been performed in order to produce more reliable and quicker performance with the CNN approach. In the transformation of the image, the following preprocessing techniques are used:

- Normalization - An image is normalized to eliminate differences and to produce an improved image of the face.
- Gray scaling - Gray scaling is an image transformation method whose pixel value depends upon the strength of the image's light. As colored images are hard to process by an algorithm, gray scaling is completed.
- Redimensioning - The image is redimensioned to delete the unnecessary portion of the image. Undoubtedly, this decreases the required memory and increases the speed of calculation.[37].

Face Detection

Face identification is the foundational step for every Facial Expression Recognition System. It is an efficient solution to object detection proposed in their article, "Rapid Object Detection using a Boosted Cascade of Simple Features"[38] in 2001, by Paul Viola and Michael Jones. Classifiers that detect an object in a picture or video where many positive as well as negative images learn a cascade function. In addition, Haar cascades in images have proven to be a powerful means of object detection and provide high precision. Three dark regions on the forehead, such as the eyebrows, are detected by Haar features. Haar cascades delete the unwanted background data from the picture effectively and detect the face area from the picture. OpenCV introduced the face detection mechanism with Haar cascade classifiers. This approach used rectangular characteristics[39].

Emotion Classification

Here, the device classifies the picture into one of the seven universal expressions as entitled in the FER2013 dataset - Happy, Sad, Anger, Surprise, Disgust, Fear, and Neutral. The training was carried out using CNN, which is a collection of neural networks. On the training range, the dataset was trained first. Before feeding it into CNN, the process of feature extraction was not performed on the results. The method followed was to experiment on the CNN with various architectures, to obtain better accuracy with the validation set. The step of classification of emotion consists of the following stages:

- Data splitting:
  The dataset was separated into three categories: training, public testing, and private testing. A training and public test set was used for the generation of a model and a private test set was used for the validation of the model.

- Model training and generation:
  The design of the neural network was addressed in-depth in the layout of CNN section earlier. Here we can see that the proposed model was set to the network and that after training on datasets, the model updates will be generated and applied to the previous structure with the .json file.

- Evaluation of model:
  The updates of the model produced during the training process were evaluated on the validation set consisting of 3589 images.

- Using the CNN model to classify real-time images:
  The transfer learning theory can be used to recognize the emotion in images here in real-time. The model developed during the training phase consists of the corresponding weights and values that can be used to detect new facial expressions. Since the created model already contains weights, it can certainly be said that FER is faster for real-time pictures.

Experiments and Results Analysis

Accuracy

Since the proposed model has been trained on a composite dataset, training accuracy above 95 percent and validation accuracy above 65 percent has been reached, which would be 70 percent after several epochs. It can be mentioned earlier that just after 30 epochs, the CNN model has a training accuracy of 95 percent, whereas CNN has taken further epochs to reach greater accuracy. A slight comparison of the suggested approach with other related works is seen in the table 1. From the table, it can be ensured that the CNN approach is much better than adjusting any other technique or approach to the recognition of human emotions, and our proposed model demonstrates better work.

Table 1. Accuracy Comparison with Related Works

Algorithm	Accuracy(%)	Computational complexity
Alexnet[40]	55-88	O4
VGG[41]	65-68	O$\hat{9}$
GoogleNet[42]	82-88	O$\hat{5}$
Resnet[41]	72-74	O$\hat{1}$6
FER(Our proposed)	75-96	O$\hat{4}$

Loss and Accuracy over Time

It can be ensured that the loss reduces, and that the accuracy increases with each epoch. Training and test accuracy for training and validation losses collected using CNN for the FER2013 dataset are given in the table 2. From the table, it can be ensuredthat as the epoch increases it shows a better accuracy rate for both the training and validation.

Table 2. Accuracy per epoch

Epoch	Training Accuracy	Validation Accuracy
1	34.14	44.04
2	47.87	50.40
3	53.05	53.91
4	56.14	56.76
5	58.87	58.32
6	60.35	56.98
7	62.28	59.32
8	63.88	61.44
..	.....	.....
15	77.30	64.14
..	.....	.....
30	96.50	65.05

Accuracy and Loss graph

The authors recognize that the accuracy of training and validation are assessed to determine a model fitting. If there is a large difference between the two, the model is over-fitting. The accuracy of the validation should be equal to or marginally less than the accuracy of the preparation to be a better model. This work is also seen in Fig. 10, as the epoch improves the training accuracy is marginally higher than validation accuracy as the authors extended the layers and eventually introduced a few more convolution layers and several entirely related layers, rendering the network both larger and broader. It seems like the lack of preparation can still be smaller than the loss of validation.

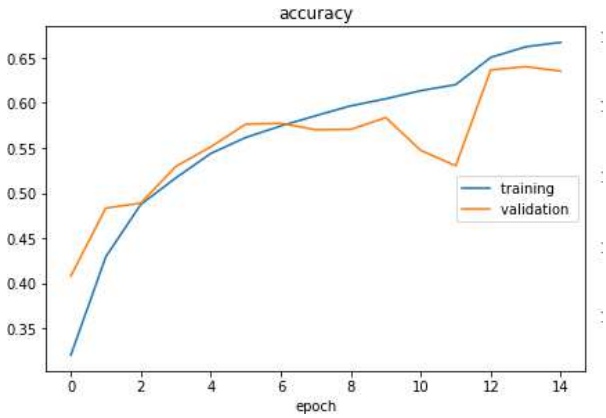

Figure 10. Graph of training and validation accuracy per epoch

In Fig. 11, the authors show the corresponding training versus validation failure. This means that the training loss reduces as the epoch grows, and the validation loss increases. In addition, the validation data are always expected to decrease as the weights are adapted. Here, as the epoch grows in higher-order then we can expect a lower rate of validation loss than the training loss which we have already seen in the last stages of the figure. Therefore, this model is well suited to the training results.

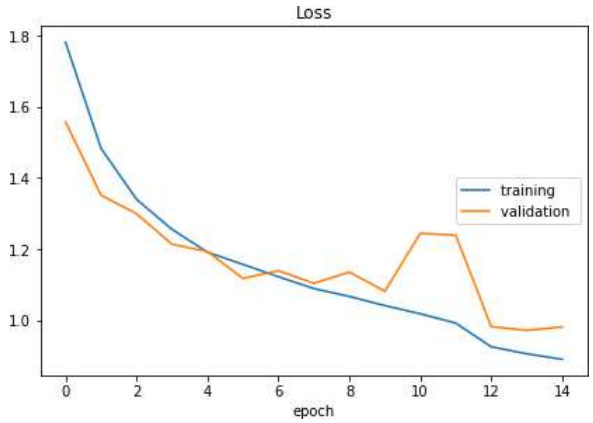

Figure 11. Graph of training and validation loss per epoch

Confusion Matrix

Fig. 12 depicts the prediction confusion matrix[43] which can be created by the test data. It can be seen that the accuracy for most expressions is well mannered. As the epoch increases in a consistent manner during each training cycle, the model will be well-suited and perform well during the real-time testing period. The dark blocks indicate that the research data has been well categorized. In addition, the numbers on each side of the diagonal show the number of pictures that have been inappropriately listed. As these numbers are smaller than the numbers on the diagonal, it can be inferred that the algorithm performed properly and obtained state-of-the-art outcomes[37].

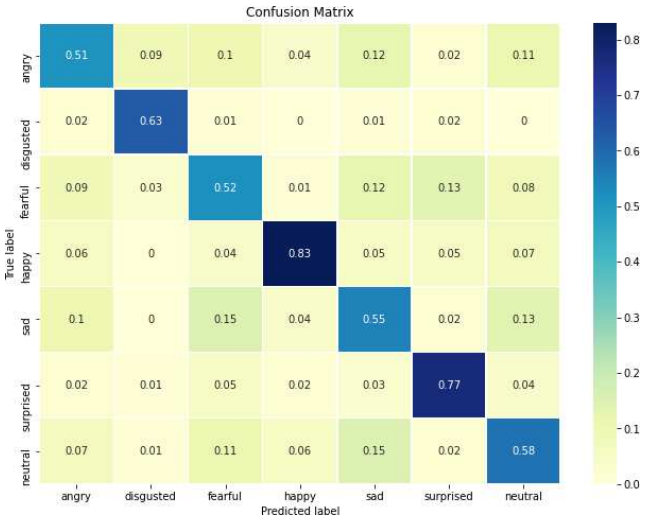

Figure 12. Predicted Confusion Matrix for the final model

Conclusion and Future Work

The authors have found seven distinct and unique emotion classes (fear, happy, angry, disgust, surprise, sad and neutral) in this article for emotion classification. There is no overlap between groups as our model perfectly fits the data. The presented

research has achieved a very good classification accuracy of emotions- 96% for random division of data in a minimal number of epochs- based on results from existing research studies and our model has been reflecting better in the real-time. The authors, however, plan to work with complex type or mixed group of emotions, such as shocked with pleasure, surprised by frustration, dissatisfied by anger, surprised by sorrow, and so on.

In this analysis, the authors conclude that due to certain variables, their proposed model is only on average. For the future, the authors will continue to focus on enhancing the consistency of each CNN model and layer. The future study involves examining multiple forms of human variables such as personality characteristics, age, and gender that affect the efficiency of emotion detection. The increasing availability of big medical data has made it necessary to use machine learning techniques to uncover hidden healthcare patterns[44-46]. In particular, deep neural networks have been recently used in healthcare applications[47]. Therefore, the proposed model has a great potential to be applicable on healthcare imaging data analysis.

Furthermore, the authors also focus on the mental or emotional stuff of a man or woman which helps as a consultant and leads the situation depending on behavioral things. Apart from this, the authors will attempt to refine the model more accurately such that a more natural method of recognition of facial expression can be provided.

Authors Contributions

MMR and SS designed the study. TD, AR, and MMR wrote the manuscript; TD, MMR, and AR collected data. SS, AR, and HAR edited the manuscript; TD, AR, and MMR carried out the analyses. TD, and AR generated all figures and tables. HAR, was not involved in any analysis. All authors have read and approved the final version of the paper.

Conflict of Interest

The authors declare no competing financial and non-financial interests.

Funding

HAR is supported by UNSW Scientia Program Fellowship.

References

1. R. Ekman, *What the face reveals: Basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS)*. Oxford University Press, USA, 1997.
2. L. Nwosu, H. Wang, J. Lu, I. Unwala, X. Yang, and T. Zhang, "Deep convolutional neural network for facial expression recognition using facial parts," in *2017 IEEE 15th Intl Conf on Dependable, Autonomic and Secure Computing, 15th Intl Conf on Pervasive Intelligence and Computing, 3rd Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress (DASC/PiCom/DataCom/CyberSciTech)*. IEEE, 2017, pp. 1318–1321.
3. B. Yang, X. Xiang, D. Xu, X. Wang, and X. Yang, "3d palmprint recognition using shape index representation and fragile bits," *Multimedia Tools and Applications*, vol. 76, no. 14, pp. 15 357–15 375, 2017.
4. N. Kumar and D. Bhargava, "A scheme of features fusion for facial expression analysis: A facial action recognition," *Journal of Statistics and Management Systems*, vol. 20, no. 4, pp. 693–701, 2017.
5. G. Tzimiropoulos and M. Pantic, "Fast algorithms for fitting active appearance models to unconstrained images," *International journal of computer vision*, vol. 122, no. 1, pp. 17–33, 2017.
6. G. Zhao and M. Pietikainen, "Dynamic texture recognition using local binary patterns with an application to facial expressions," *IEEE transactions on pattern analysis and machine intelligence*, vol. 29, no. 6, pp. 915–928, 2007.
7. H. Jung, S. Lee, J. Yim, S. Park, and J. Kim, "Joint fine-tuning in deep neural networks for facial expression recognition," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 2983–2991.
8. X. Zhao, X. Liang, L. Liu, T. Li, Y. Han, N. Vasconcelos, and S. Yan, "Peak-piloted deep network for facial expression recognition," in *European conference on computer vision*. Springer, 2016, pp. 425–442.
9. H. Zhang, A. Jolfaei, and M. Alazab, "A face emotion recognition method using convolutional neural network and image edge computing," *IEEE Access*, vol. 7, pp. 159 081–159 089, 2019.
10. I. J. Goodfellow, D. Erhan, P. L. Carrier, A. Courville, M. Mirza, B. Hamner, W. Cukierski, Y. Tang, D. Thaler, D.-H. Lee *et al.*, "Challenges in representation learning: A report on three machine learning contests," in *International conference on neural information processing*. Springer, 2013, pp. 117–124.
11. "https://www.kaggle.com/msambare/fer2013," in *FER-2013|Kaggle*.

12. Z. Yu and C. Zhang, "Image based static facial expression recognition with multiple deep network learning," in *Proceedings of the 2015 ACM on international conference on multimodal interaction*, 2015, pp. 435–442.

13. S. E. Kahou, C. Pal, X. Bouthillier, P. Froumenty, Ç. Gülçehre, R. Memisevic, P. Vincent, A. Courville, Y. Bengio, R. C. Ferrari *et al.*, "Combining modality specific deep neural networks for emotion recognition in video," in *Proceedings of the 15th ACM on International conference on multimodal interaction*, 2013, pp. 543–550.

14. M. Pantic, M. Valstar, R. Rademaker, and L. Maat, "Web-based database for facial expression analysis," in *2005 IEEE international conference on multimedia and Expo*. IEEE, 2005, pp. 5–pp.

15. X. Wang, X. Feng, and J. Peng, "A novel facial expression database construction method based on web images," in *Proceedings of the Third International Conference on Internet Multimedia Computing and Service*, 2011, pp. 124–127.

16. C. Mayer, M. Eggers, and B. Radig, "Cross-database evaluation for facial expression recognition," *Pattern recognition and image analysis*, vol. 24, no. 1, pp. 124–132, 2014.

17. Y. Tang, "Deep learning using linear support vector machines," *arXiv preprint arXiv:1306.0239*, 2013.

18. Y. Gan, "Facial expression recognition using convolutional neural network," in *Proceedings of the 2nd international conference on vision, image and signal processing*, 2018, pp. 1–5.

19. C.-E. J. Li and L. Zhao, "Emotion recognition using convolutional neural networks," in *Purdue Undergraduate Research Conference.63*, 2019.

20. Y. Lv, Z. Feng, and C. Xu, "Facial expression recognition via deep learning," in *2014 International Conference on Smart Computing*. IEEE, 2014, pp. 303–308.

21. I. M. Revina and W. S. Emmanuel, "A survey on human face expression recognition techniques," *Journal of King Saud University-Computer and Information Sciences*, 2018.

22. A. Mollahosseini, D. Chan, and M. H. Mahoor, "Going deeper in facial expression recognition using deep neural networks," in *2016 IEEE Winter conference on applications of computer vision (WACV)*. IEEE, 2016, pp. 1–10.

23. F. Li, A. Karpathy, and J. Johnson, "Cs231n: Convolutional neural networks for visual recognition. course notes," 2015.

24. R. H. Hahnloser, R. Sarpeshkar, M. A. Mahowald, R. J. Douglas, and H. S. Seung, "Digital selection and analogue amplification coexist in a cortex-inspired silicon circuit," *Nature*, vol. 405, no. 6789, pp. 947–951, 2000.

25. M. N. Patil, B. Iyer, and R. Arya, "Performance evaluation of pca and ica algorithm for facial expression recognition application," in *Proceedings of fifth international conference on soft computing for problem solving*. Springer, 2016, pp. 965–976.

26. N. Christou and N. Kanojiya, "Human facial expression recognition with convolution neural networks," in *Third International Congress on Information and Communication Technology*. Springer, 2019, pp. 539–545.

27. S. M. González-Lozoya, J. de la Calleja, L. Pellegrin, H. J. Escalante, M. A. Medina, and A. Benitez-Ruiz, "Recognition of facial expressions based on cnn features," *Multimedia Tools and Applications*, pp. 1–21, 2020.

28. A. Christy, S. Vaithyasubramanian, A. Jesudoss, and M. A. Praveena, "Multimodal speech emotion recognition and classification using convolutional neural network techniques," *International Journal of Speech Technology*, vol. 23, pp. 381–388, 2020.

29. S. Singh and F. Nasoz, "Facial expression recognition with convolutional neural networks," in *2020 10th Annual Computing and Communication Workshop and Conference (CCWC)*. IEEE, 2020, pp. 0324–0328.

30. F. Wang, S. Wu, W. Zhang, Z. Xu, Y. Zhang, C. Wu, and S. Coleman, "Emotion recognition with convolutional neural network and eeg-based efdms," *Neuropsychologia*, vol. 146, p. 107506, 2020.

31. D. Canedo and A. J. Neves, "Facial expression recognition using computer vision: A systematic review," *Applied Sciences*, vol. 9, no. 21, p. 4678, 2019.

32. F. Nonis, N. Dagnes, F. Marcolin, and E. Vezzetti, "3d approaches and challenges in facial expression recognition algorithms—a literature review," *Applied Sciences*, vol. 9, no. 18, p. 3904, 2019.

33. J. D. Bodapati and N. Veeranjaneyulu, "Facial emotion recognition using deep cnn based features," 2019.

34. J. Haddad, O. Lézoray, and P. Hamel, "3d-cnn for facial emotion recognition in videos," in *International Symposium on Visual Computing*. Springer, 2020, pp. 298–309.

35. "https://www.coursera.org/projects/facial-expression-recognition-keras," in *Facial Expression Recognition with Keras*.

36. B. Yang, J. Cao, R. Ni, and Y. Zhang, "Facial expression recognition using weighted mixture deep neural network based on

double-channel facial images," *IEEE Access*, vol. 6, pp. 4630–4640, 2017.

37. I. Talegaonkar, K. Joshi, S. Valunj, R. Kohok, and A. Kulkarni, "Real time facial expression recognition using deep learning," *Available at SSRN 3421486*, 2019.

38. P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition. CVPR 2001*, vol. 1. IEEE, 2001, pp. I–I.

39. A. Mohan, C. Papageorgiou, and T. Poggio, "Example-based object detection in images by components," *IEEE transactions on pattern analysis and machine intelligence*, vol. 23, no. 4, pp. 349–361, 2001.

40. Y.-L. Wu, H.-Y. Tsai, Y.-C. Huang, and B.-H. Chen, "Accurate emotion recognition for driving risk prevention in driver monitoring system," in *2018 IEEE 7th Global Conference on Consumer Electronics (GCCE)*. IEEE, 2018, pp. 796–797.

41. A. Kumar and G. Garg, "Sentiment analysis of multimodal twitter data," *Multimedia Tools and Applications*, vol. 78, no. 17, pp. 24 103–24 119, 2019.

42. P. Giannopoulos, I. Perikos, and I. Hatzilygeroudis, "Deep learning approaches for facial emotion recognition: A case study on fer-2013," in *Advances in hybridization of intelligent methods*. Springer, 2018, pp. 1–16.

43. A. Saravanan, G. Perichetla, and D. K. Gayathri, "Facial emotion recognition using convolutional neural networks," *arXiv preprint arXiv:1910.05602*, 2019.

44. H. Parvin, et al., "A novel classifier ensemble method based on class weightening in huge dataset", In *International Symposium on Neural Networks,* 2011, (pp. 144-150). Springer, Berlin, Heidelberg.

45. H. Alinejad-Rokny, et al., "Source of CpG depletion in the HIV-1 genome", *Molecular Biology and Evolution*, 2016, vol. *33, no.* 12, pp. 3205-3212.

46. R. Javanmard, "Proposed a new method for rules extraction using artificial neural network and artificial immune system in cancer diagnosis", *Journal of Bionanoscience*, 2013, vol. *7*, no. 6, pp. 665-672.

47. S. Shamshirband, M. Fathi, et al., "A Review on Deep Learning Approaches in Healthcare Systems: Taxonomies, Challenges, and Open Issues", *Journal of Biomedical Informatics*, 2020, 103627.

Figures

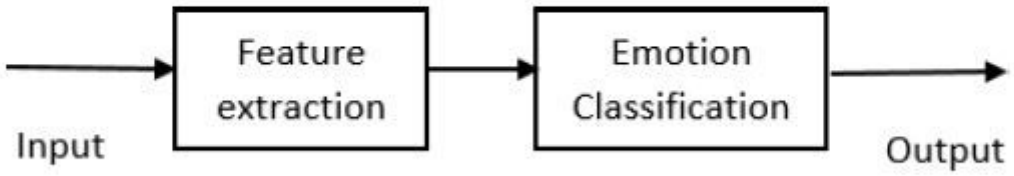

Figure 1

Facial expression recognition system

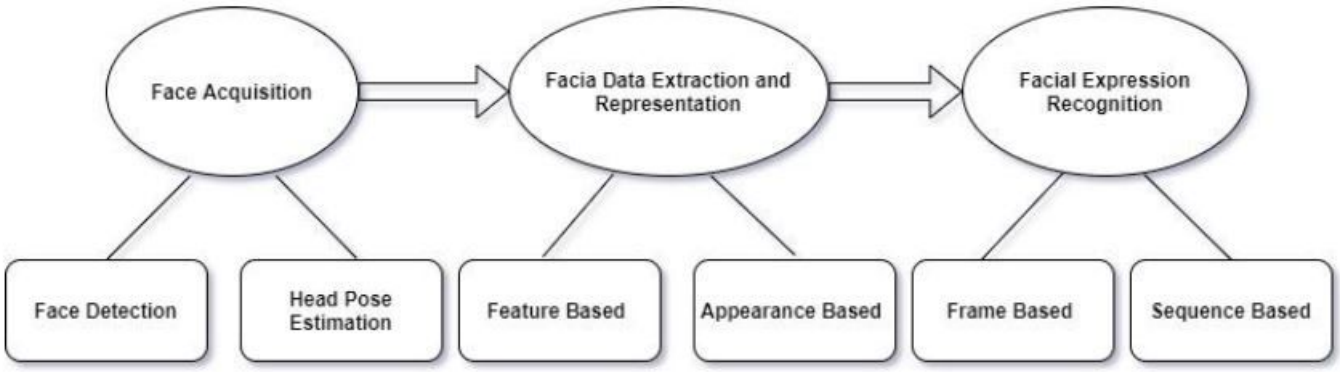

Figure 2

The basic framework of applications for facial expression analysis

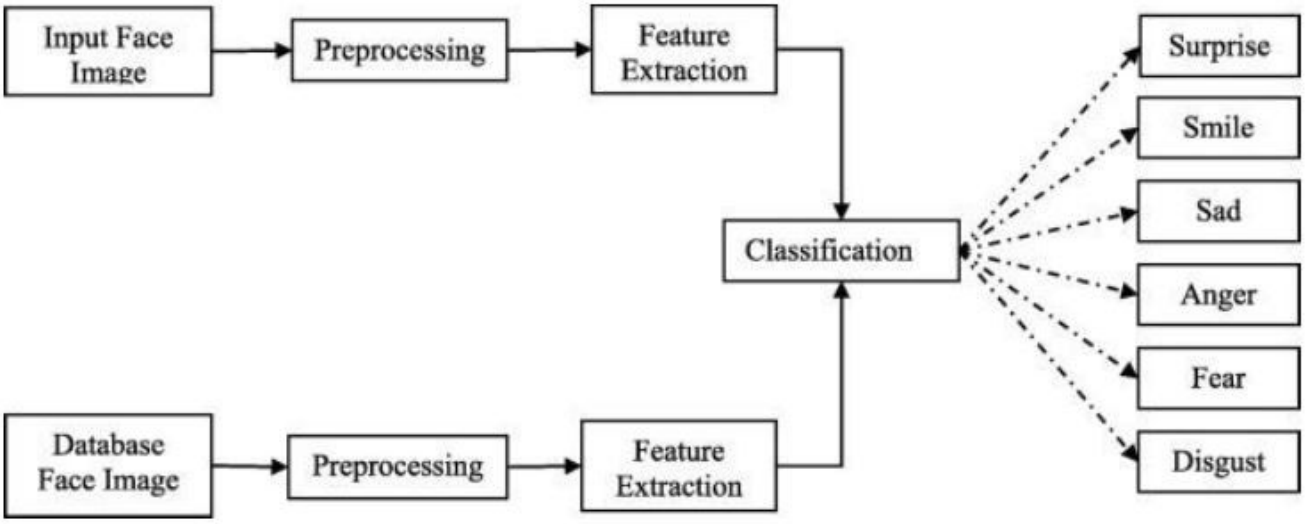

Figure 3

Summary flowchart for the three phases of the FER method21

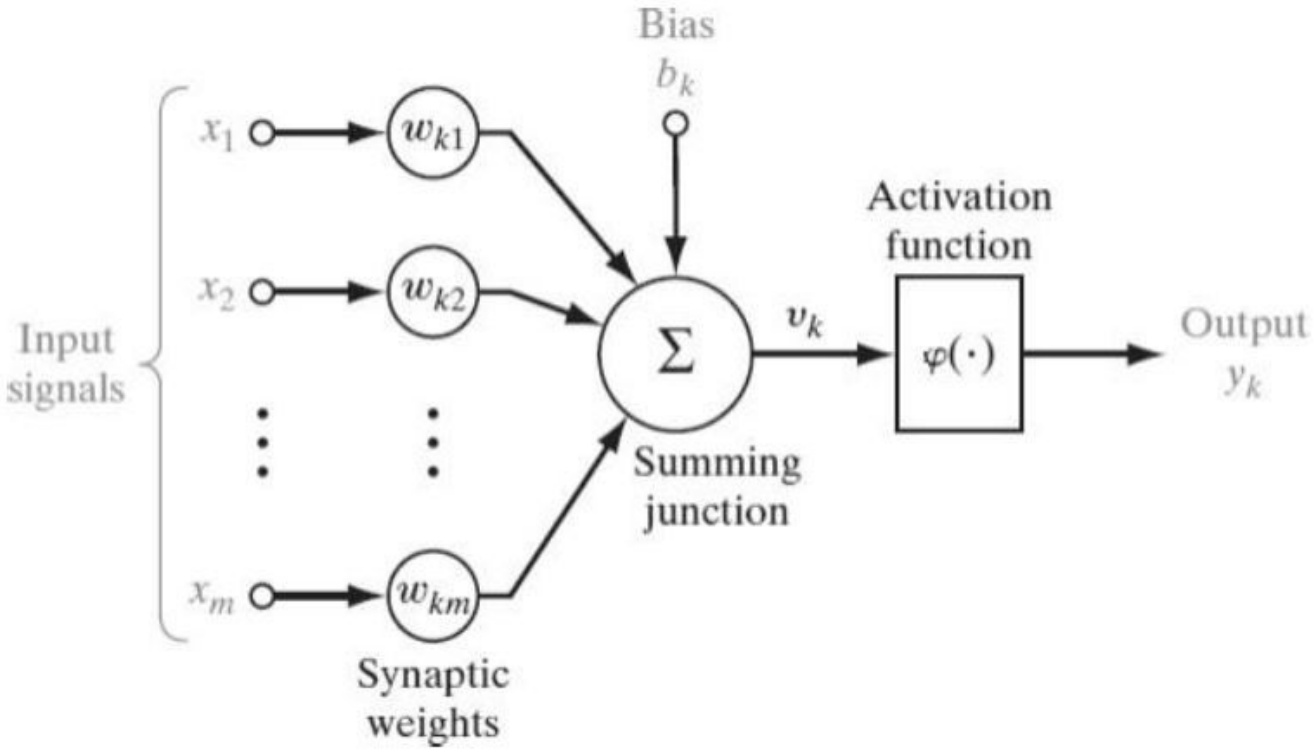

Figure 4

The Illustration of a Neuron unit23

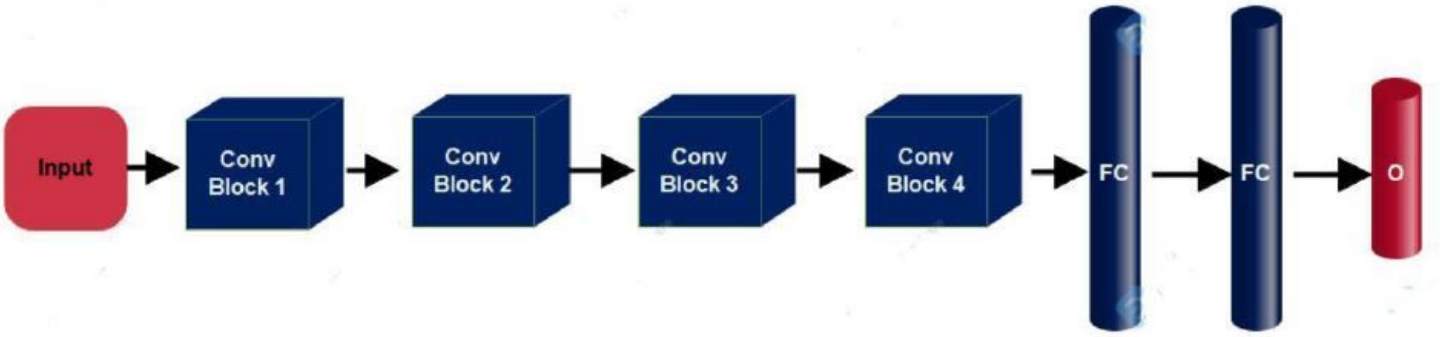

Figure 5

Reference CNN model for FER35

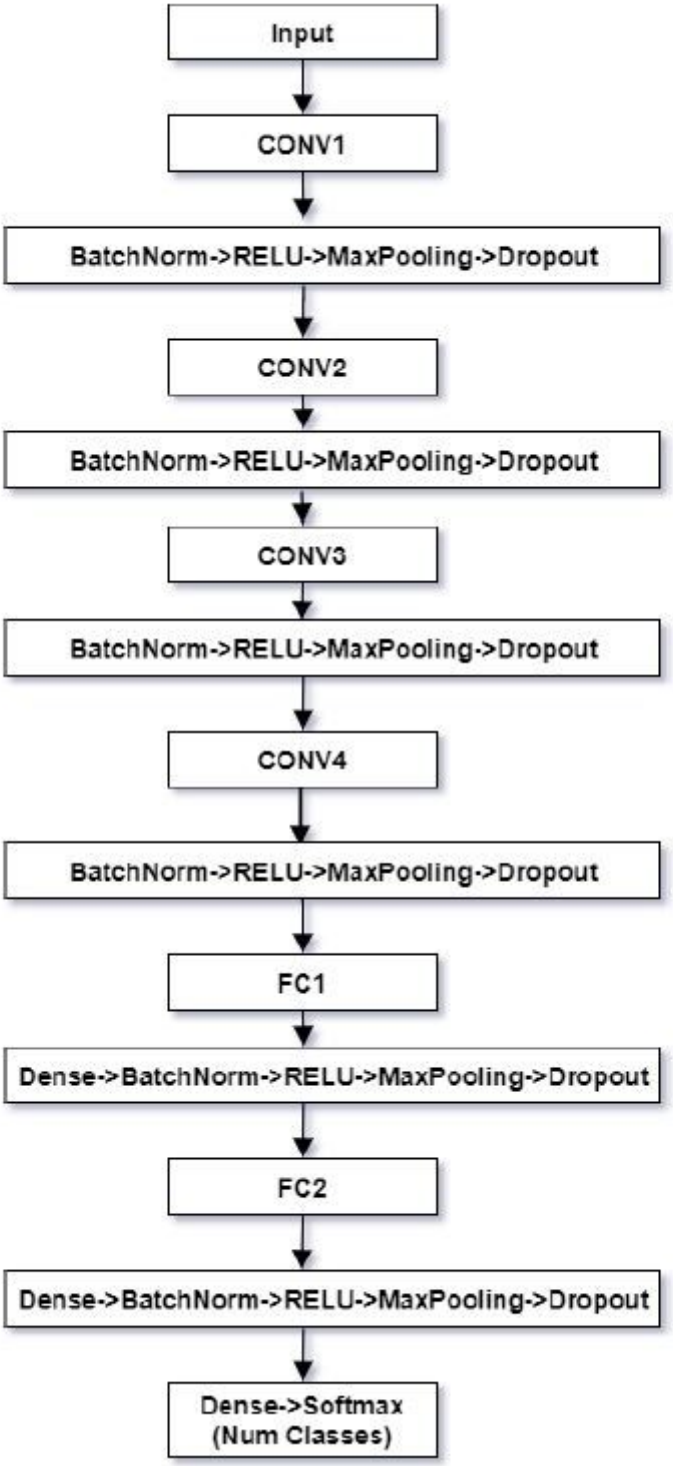

Figure 6

Pipeline for proposed CNN

Figure 7

The examples of seven basic emotion11

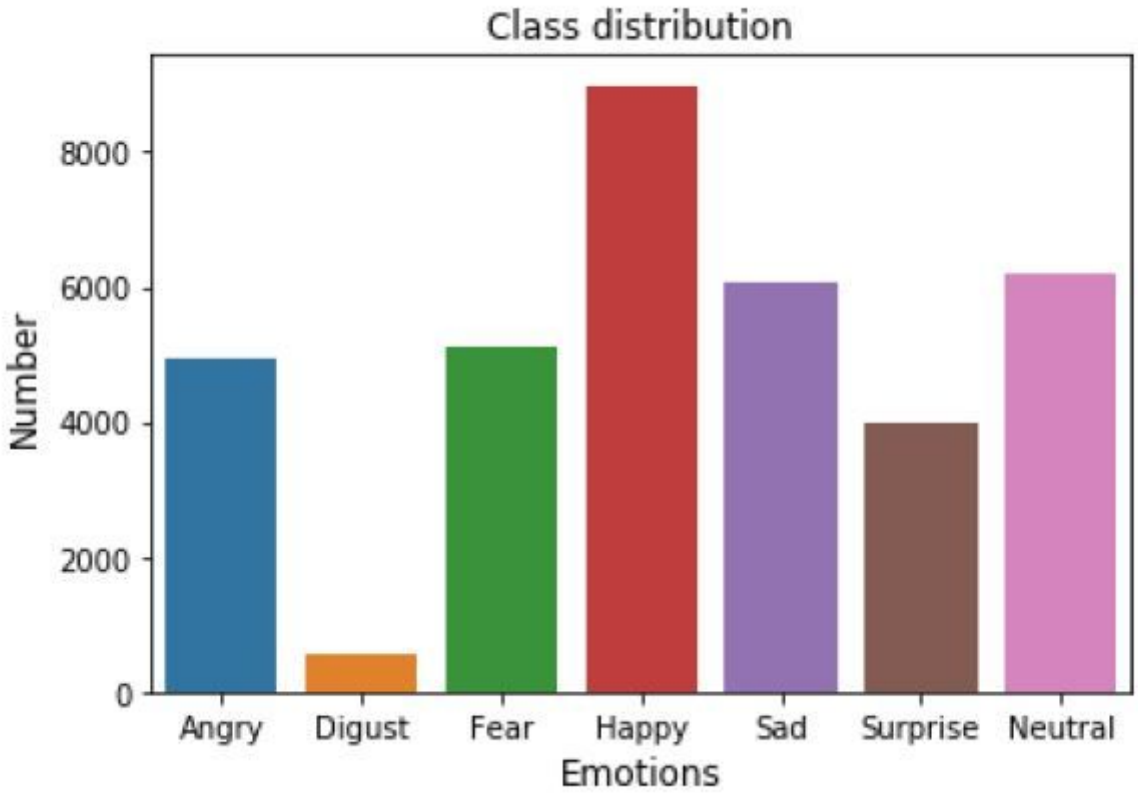

Figure 8

Expression distribution

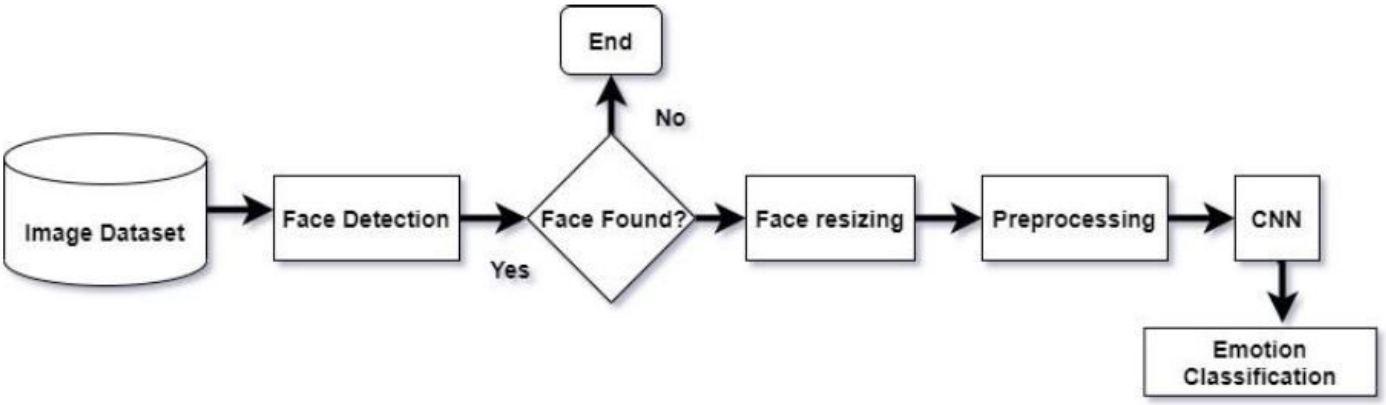

Figure 9

System flowchart of emotion classification

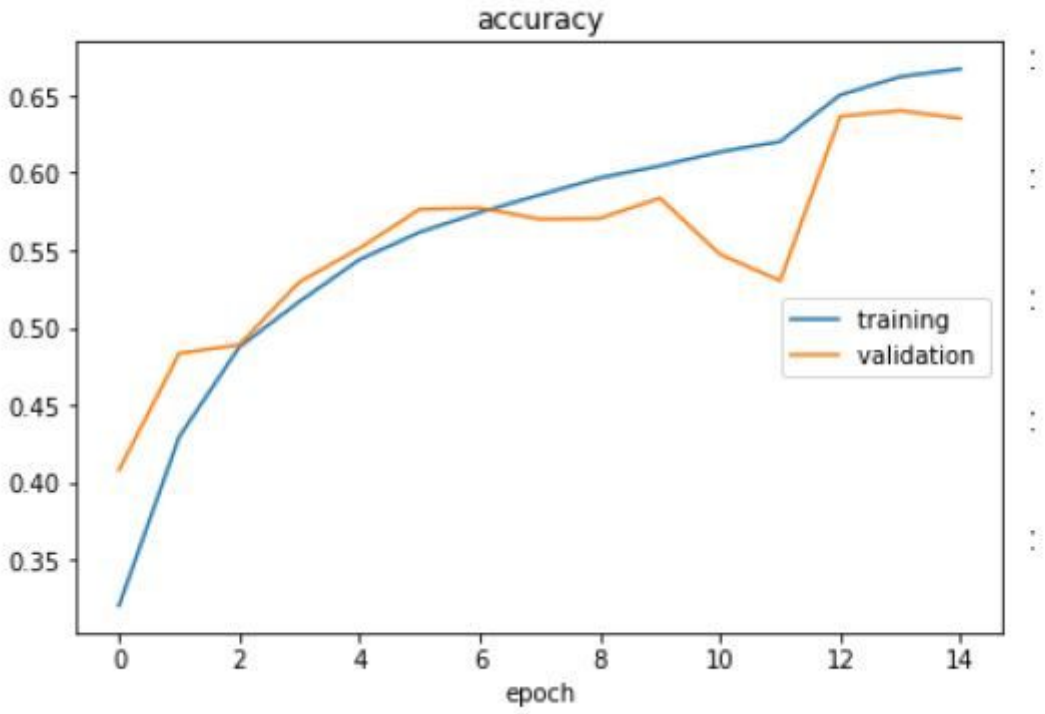

Figure 10

Graph of training and validation accuracy per epoch

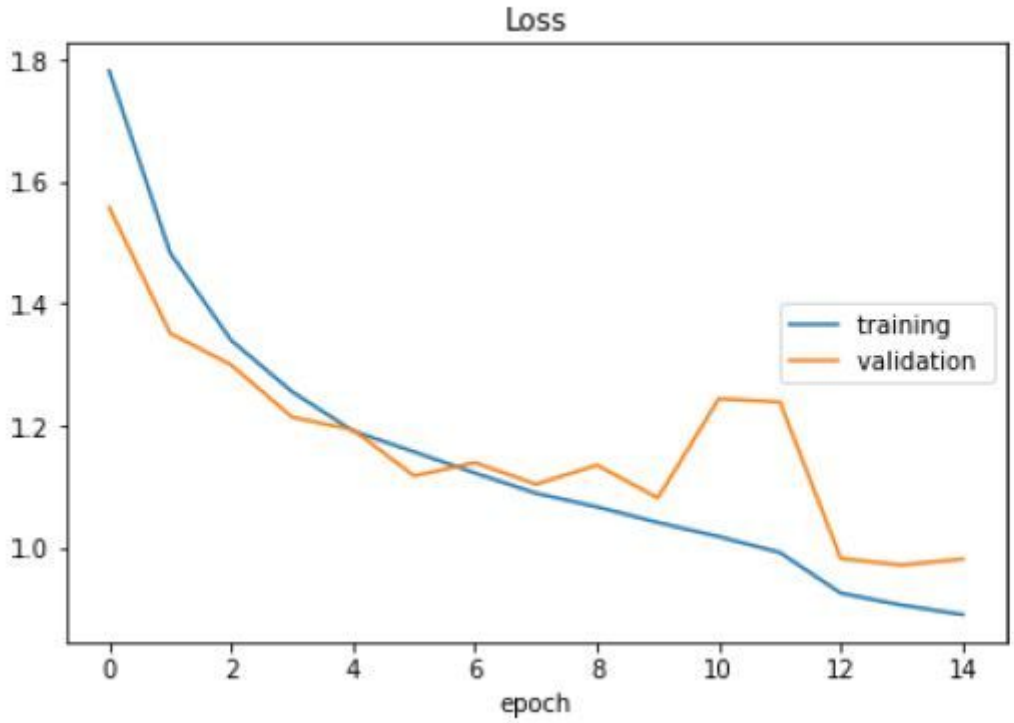

Figure 11

Graph of training and validation loss per epoch

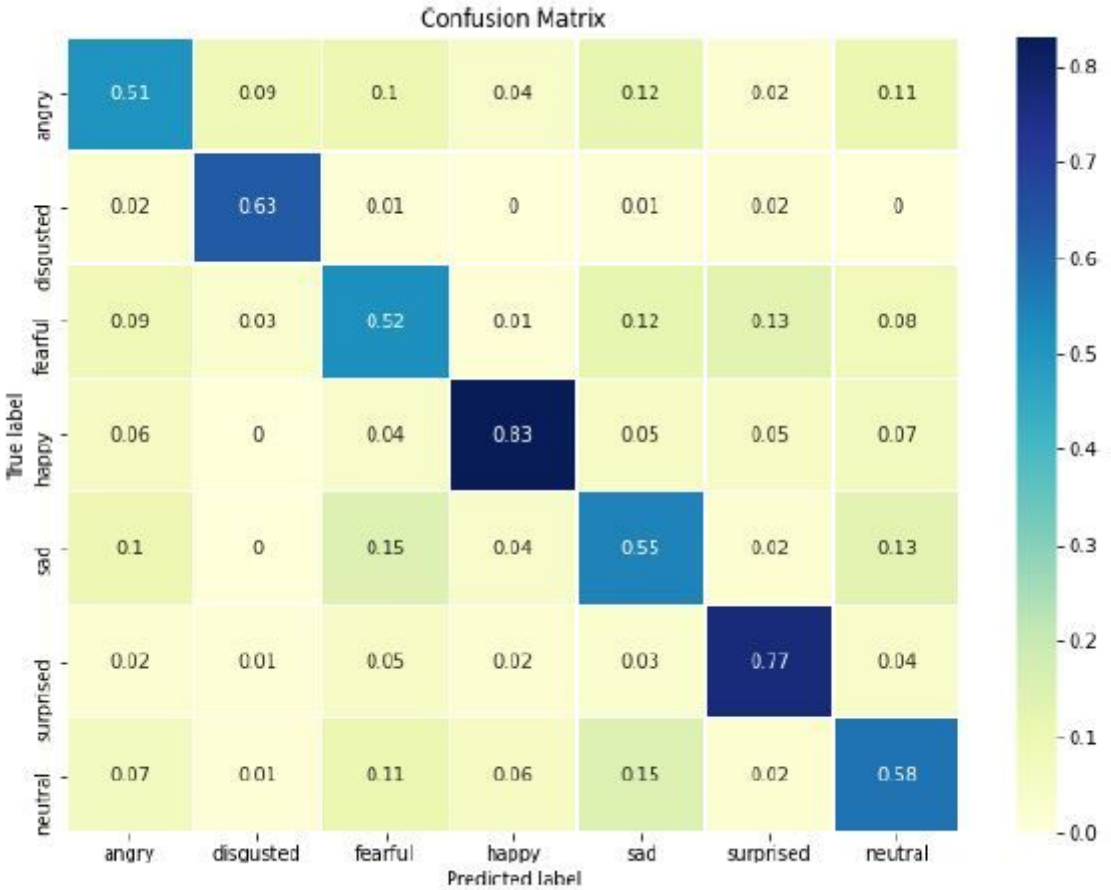

Figure 12

Predicted Confusion Matrix for the final model