Eurasian J. Sci. Eng., 2023, 9(3), 209-219
https://doi.org/10.23918/eajse.v9i3p18

Published by Tishk International University
Available at: https://eajse.tiu.edu.iq/submit/

RESEARCH ARTICLE Open Access

Improving the Accuracy of Human Emotion Recognition through CNN Layering Architecture

Shakir Fattah Kak[1]*

[1]Department of Information Technology, College of Informatics, Akre University for Applied Sciences, Duhok, Iraq

Article History
Received: 25.07.2023
Revised: 09.03.2024
Accepted: 24.03.2024
Published 26.03.2024
Communicated by: Dr. Orhan Tug
*Email address:
shakir.fattah@auas.edu.krd
*Corresponding Author

Abstract:
In accordance with the proverb "a picture is worth a thousand words," facial expressions have the ability to convey a wide range of emotions depending on the circumstances. Observing facial expressions during face-to-face interactions allows us to infer the thoughts and feelings of others. Therefore, the development of facial expression recognition systems holds great significance in the field of artificial intelligence. Recognizing facial expressions involves several fundamental stages. Firstly, the images used in the system undergo pre-processing. Following this, important features are collected and extracted from the dataset images. Finally, the resulting expressions from the images are classified using prediction techniques. This study employed convolutional neural networks (CNNs), a well-established image processing method with a layered architecture, to develop an efficient model for accurately recognizing facial expressions with optimal accuracy. To thoroughly analyze the proposed model, the following datasets were utilized: Facial Expression Recognition 2013 (FER2013), Extended Cohn-Kanade (CK+), and Japanese Female Facial Expressions (JAFFE). These datasets consist of supervised data and cover seven different emotions: neutral, happy, angry, sad, fear, disgust, and surprise. The findings demonstrate that the suggested technique consistently outperforms contemporary methods across all of the aforementioned datasets, achieving notable improvements in accuracy.
Keywords: *Supervised Data, Deep Learning, CK+48, Fer2013, JAFFE, Recognition of Facial Expression.*

1. Introduction

The primary way that people naturally express their emotional states to others is through their facial expressions. Numerous studies have revealed that more than 50% of our emotions are represented directly through facial expressions, a rate 10 times higher than the percentage of emotions communicated through the tones of oral words [1]. Artificial intelligence technology is a large area of research at present to study human psychological emotions in order to automatically recognize these feelings [2, 3]. Deep learning is a form of machine learning, and artificial intelligence can be attained through machine learning. The hidden multi-layer architecture is adopted in the concept of deep learning structuring. Deep learning integrates low-level characteristics to create a higher-level, more abstract representation of attribute categories or features to identify scattered properties in data [4]. Recently, Convolutional Neural Network (CNN) technology has been used to extract the basic properties of digital images. Its effectiveness exceeds that of traditional support vector machines, and it continues to work well in various environments and under rapid changes [2]. Face acquisition, facial feature extraction, and building a classifier are Mostly the three sequential stages of recognizing facial expressions [5]. The procedure for recognizing the emotional expression of human faces goes through three main stages, which include: first, face acquisition, then, after that, facial feature extraction, and finally, constructing the system classifier [5]. The topic of emotion recognition has attracted a lot of researchers who have conducted extensive studies on it [6]. Deep learning techniques have begun to gain traction in the field of emotion identification in recent years because of the tremendous advances in computer processing power and network architecture design [6]. The convolution neural network model has produced excellent target detection and facial expression recognition results. It also demonstrated that as the network layer depth increases, the network's ability to recognize targets and

facial expressions improves, as demonstrated by AlexNet's layered design [6-8]. Clinical psychology, neurology, pain assessment, psychiatry, lie detection, and multimodal human-computer interaction are just a few of the significant fields that might benefit from emotional recognition technology [9]. This inspired us to develop and design an efficient Convolutional Neural Network layer architecture-based facial expression recognition system, which we discuss in this research. In this research, some of the available layers were used to build a robust CNN structure and adopted after verifying the effectiveness of this new proposed structure on standards-related datasets in emotional expression recognition systems.

The organization of the remaining sections of the paper is as follows: Section 2 delves into related works, whereas Section 3 presents the methodology of the proposed model. Section 4 details the facial databases used in the research, and Section 5 demonstrates the outcomes of the study. Lastly, Section 6 concludes the paper.

2. Related Works

Many algorithms based on deep learning have been studied recently due to the growth of large data sets and the advancement of hardware technology. To improve the recognition rate of facial expressions and the effective features of the new combination of layers, it was intended to create a model composed of deep learning layers and train them with big data. These advances also affect the FER segment, so independent learning of the extracted facial features has resulted in more reliable and efficient feature recognition.

In 2019, [3] presented an article titled "Human Face Expression Recognition Based on Deep Learning-Deep Convolutional Neural Network ", proposing facial expression recognition, which consists of effective data processing, neural network structure, and efficient loss function. The researcher proposed a CNN model consisting of three sections as follows: convolutional layers next come the convolutional layer followed by the fully connected layer, which is followed by the ReLu layer after each section. In July 2016, [10] published a research that demonstrates the facial expression recognition system by applying the concept of deep learning technology. The proposed method extracted the facial features to identify facial emotions, even if there are any changes in face position, such as their rotation. They used the idea of cropping the necessary parts as a movement block, excluding the unnecessary parts. In April 2019, [11], presented an article that combines the outcome of the softmax layer of two facial characteristics, by considering the mistake related to the moment's most noteworthy feeling (Top-2) expectation result. They provide a method that makes use of the autoencoder approach to produce face pictures with neutral emotion. Without any sequence information, it can extract the dynamic face characteristics from both neutral and emotional photos. The effectiveness of the proposed technique was tested on the CK+ and JAFFE datasets. Similarly, in 2020, [12] issued a publication titled "Research on Facial Expression Recognition Based on Neural Network." The authors proposed a CNN model based on the classic AlexNet, employing the cross-connection technique, and added to the fully connected layer the processed pixels at the bottom. The authors examined several value selections of CNN hyperparameters, such as the number of convolution kernels in each layer and the size of the receptive field. Another paper titled "A novel approach for facial expression recognition" [13] was published in 2020. This research goes through two main stages: the first is the process of identifying facial expressions only for all existing images from the images in the FER datasets used in this research, which are the CK+ and JAFFE using a Gabor filter technique, and the second stage is extracting the best features of the specific facial expressions from the previous stage using the PCA technique. In addition to the above, the SVM technique was used in the process of classifying the facial expressions of the images used in this research. Also, in a study in 2021 [14], the authors explored Facial Emotion Recognition using the MobiExpressNet model. They provide MobiExpressNet, a novel, lightweight Deep Learning model for FER. They suggested a model that depends on depth-wise separable convolutions to restrict complexity, and they used a rapid down-sampling strategy in conjunction with a few layers in the

architecture to make the model size as minimal as possible. The size and Floating-point Operations Per Second of the MobiExpressNet model is demonstrated to be more than 5 times less than the smallest MobileNet model, making the created model particularly appealing for real-time applications. The effectiveness of the proposed technique was tested on the Fer2013 dataset. Moreover, in April 2020, [15], presented an article titled "A Face Expression Recognition Using CNN & LBP". The researchers followed a work procedure that included several steps, starting by splitting the FER datasets (CK+, YALE FACE, and JAFFE) into both the testing set (30%) and training set (70%). Next comes using the LBP and CNN for feature extraction and the SVM for the classification of the extracted features. Finally, the Soft Max layer is used to classify the FER dataset images.

As a result, the use of feature extraction and classification techniques to classify facial expressions, in general, has received significant study in recent years, but there is still more work to be done to investigate more reliable and accurate techniques to increase the efficiency of systems that rely on this principle in their work.

3. The Proposed Model Methodology

The suggested architecture of the CNN model is specifically created for carrying out tasks related to image classification, whereby the model acquires the skill of identifying patterns and characteristics in the given input images. The researchers develop numerous models based on CNN structures consisting of several layers after running trials on them and then selecting the best among them to achieve an excellent accuracy rate in the field of recognizing human expressions of emotion through the face, and the research is still ongoing. CNNs are a kind of neural network that is quite effective in applications like face expression recognition systems. It consists of many layers, which are employed to learn directly from the expression of face picture pixels. Generally speaking, the testing and training phases of most recognition systems, particularly human expression recognition, entail the use of an existing related dataset or the creation of a new one. In this research, the datasets employed are JAFEE, FER2013, and CK+ for testing the proposed CNN layering architecture. A huge training set that uses a face expression database determines how accurate CNNs are. For the CNN layers to create a multidimensional array with new numbers, they need input data in the form of a multidimensional array. The output data from the previous layer serves as the input for the subsequent layer in the CNN algorithm.

The proposed structure aims to improve the accuracy of detection and recognition of human emotions that appear on faces by using the standard global data set, selecting the data and then processing it, then training the model and conducting tests on it, which ultimately leads to choosing the appropriate model for this idea. Determining the appropriate model is based on obtaining good accuracy and being stable as a solution to the problem in terms of extracting features from the images of the data set through the use of generalization, activation, and classification of the model to respond correctly and effectively to reach the purpose of testing the proposed model.

3.1 Split the Data Set

An important basic process in validating a proposed model or setting up human emotion recognition systems is to split the data set used into two groups. The first group is called the test group, which works to confirm the effectiveness of the proposed model, while the second group is called the training group.

3.2 Building Blocks of CNN Layers

A CNN technology uses convolutions to handle the arithmetic in the background, which is a significant distinction between a CNN and a standard neural network. At least one layer of the CNN employs convolution rather than matrix multiplication. A function is taken from two functions and returned through convolutions. The CNN model often consists of several layers, as presented in Figure 1.

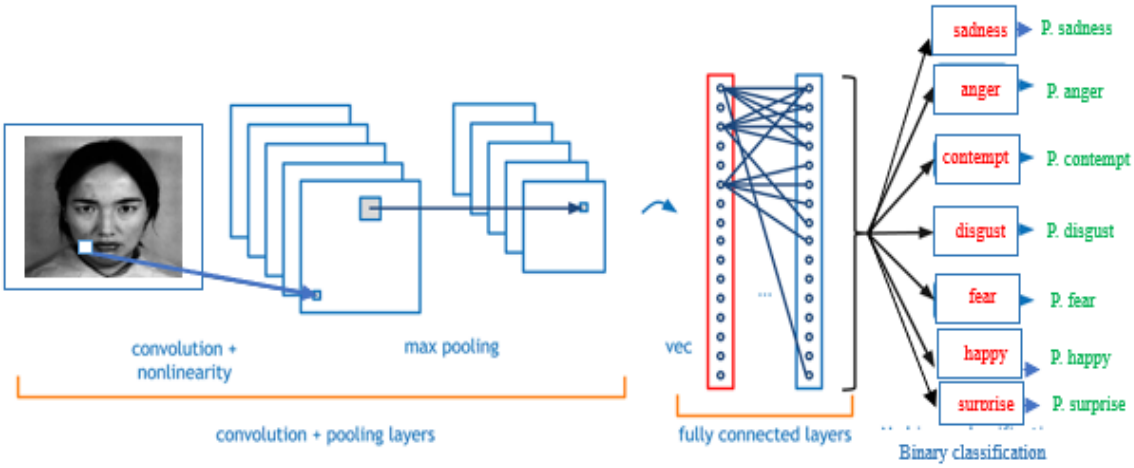

Figure 1: Structure layer levels of CNN.

For the CNN layers to create a multidimensional array with new numbers, they need input data in the form of a multidimensional array. The output data from the prior layer serves as the input data for the subsequent layer in the CNN algorithm. Initially, as shown in Figure 1, the matrix image is sent as input data to the CNN, which will be passed through its various layers. The convolution layer receives the input image as a first layer, and the output data are the extracted features from the input image that uses a kernel and an image filter as inputs and are then mathematically processed. Next, if the image is too huge, it can be used either as a sum pooling, average pooling, or max pooling layer, which will assist in lowering the number of parameters. Max pooling has been utilized in this model. The image is repeated for the convolution layer, where features are extracted. Additionally, in extracting image features, other useful layers with different parameter values can be employed at this level, such as the batch normalization layer, relu Layer, and max pooling layer. After the feature extraction phase, the classification phase started, consisting of several useful layers, such as the fully connected layer based on the dataset subjects, the softmax layer, and the classification layer. The proposed CNN layering (16 layers) architecture is illustrated in Figure 2.

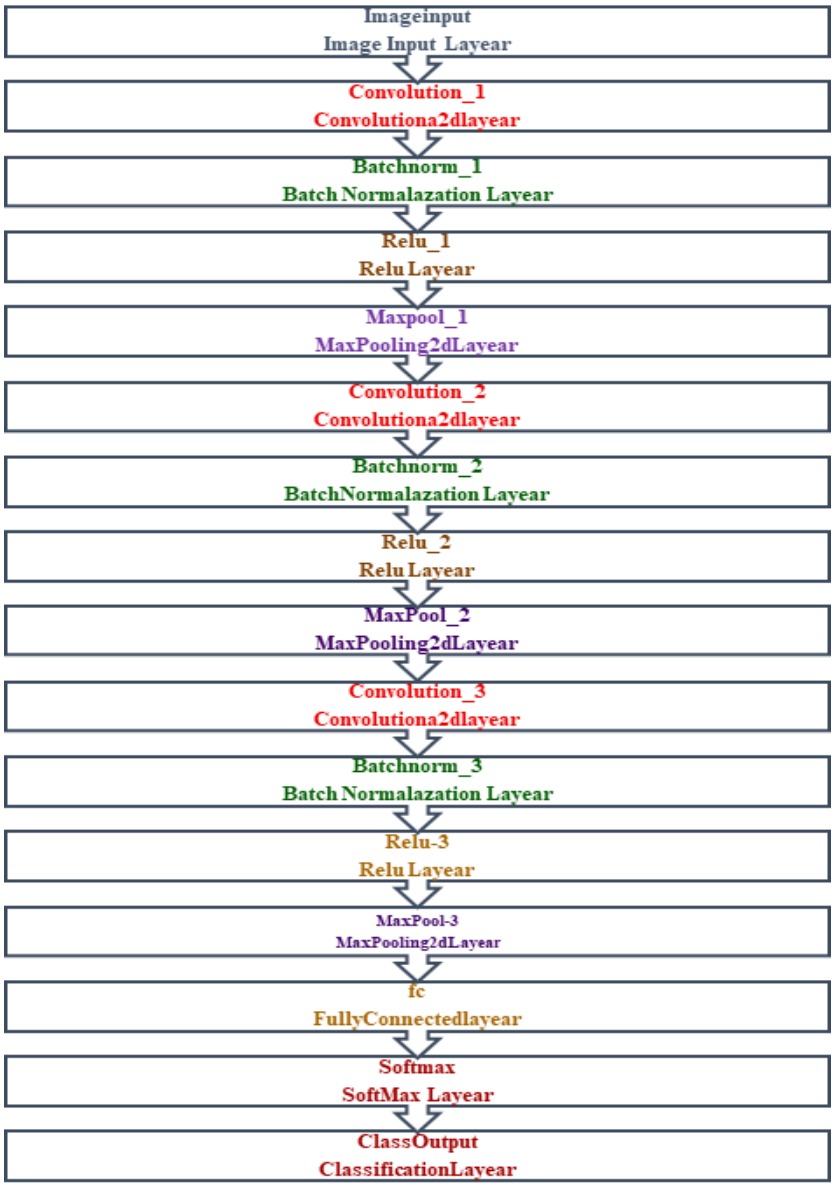

Figure 2: CNN layering architecture model.

3.3 Dataset Image Features Extracting

The process of choosing and modifying essential information from raw data to generate a set of features appropriate for machine learning algorithms is known as feature extraction. Using a set of measured data as a starting point and properly employing machine learning techniques such as feature extraction, pattern recognition, and image processing to obtain efficient and non-repetitive derivative values that make further learning possible in later periods. Machine learning techniques have provided high-precision tools to deal with this process. In this model, the convolution process is utilized to detect image features via image matrix multiplication through a filter. Because neural networks are extremely sensitive to anomalous input, batch normalization assistance are needed to get normalizing outcomes. In the field of artificial intelligence and deep learning, activation functions play a crucial role in determining the output of neural networks. Generalization is the process through which a model can perform better when applied to newly discovered data from the same distribution as that used to build the model. Furthermore, generalization and activation play crucial roles in the process of learning and cognitive development. It may control how much the training dataset learns from the network model by selecting the activation feature in the hidden layer. For this purpose, the relu layer is used in the proposed structure. Additionally, the maxPooling2dLayer is used in this model. In the field of deep learning, a max pooling 2D layer plays a critical role in reducing the spatial dimensions of input data

while preserving its most important features. This specialized layer extracts the maximum value within a specified window, called a kernel, as it moves across the input data. The purpose is to efficiently reduce computational cost and control overfitting, which are common issues when working with large-scale or high-resolution data. Implementing a max pooling 2D layer in neural networks is essential to improving the performance and efficiency of the overall model. Here, maxPooling2dLayer is one of the most commonly utilized pooling techniques in order to reduce network complexity.

3.4 Training The Proposed Model for Evaluation

To obtain high accuracy for recognizing facial expressions based on the databases used in this research, the values of the variables in the proposed model layers had to be changed continuously. For example, the hyperparameters whose values have been modified in this research are the learning rate, number of epochs, batch size, and the convolutional kernel's n-th degree. To verify the effectiveness of the suggested structure at various division rates, the training set and test set are divided based on the subjects in the datasets (anger, happiness, contempt, disgust, fear, sadness, and surprise), which are shared by all the datasets used in this approach. At this point, separate portions of the training and test sets have been allotted. For example, the training dataset now contains 90% of the photos, while the test dataset has been given 10%. According to the accurate prediction of the validation test dataset with the training dataset, the suggested model structure's performance was assessed.

Assessing the Model's Performance

The proposed model effectively splits the dataset into two separate portions: one for training the model and one for assessing its performance using the test dataset. Figure 3 displays an example of applying the confusion matrix using the proposed model on the CK+ dataset.

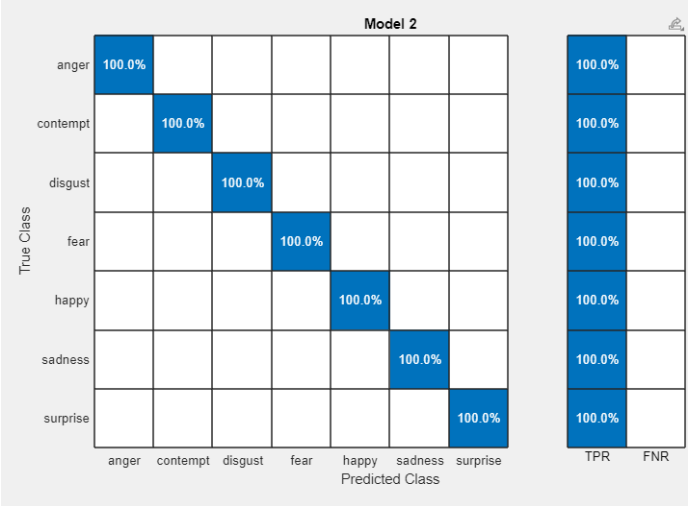

Figure 3: Confusion Matrix.

A Confusion Matrix is a performance measurement technique that provides valuable insights into the model's classification performance by presenting a matrix of actual versus predicted classes. In the case of facial expression recognition, the Confusion Matrix showcases the model's ability to correctly identify different facial expressions, such as happiness, sadness, anger, and more, with unparalleled precision. The 100% accuracy rate underscores the effectiveness of the deep learning algorithms employed, highlighting the model's proficiency in capturing subtle nuances and variations in facial features. This achievement signifies a breakthrough in the field of computer vision, particularly in applications related to human emotion recognition. The high-performing CNN architecture demonstrates its effectiveness in correctly classifying facial expressions captured in the CK+ dataset. The confusion matrix reveals a strong diagonal line, indicating a significant number of correct predictions across various emotional categories such as happiness, sadness, anger, surprise, and more.

This outcome underscores the model's ability to accurately distinguish and classify different facial expressions, thereby highlighting the success of the proposed CNN layering architecture on the CK+ dataset.

In essence, the Learning Process Rate's success in achieving a perfect accuracy rate in facial expression recognition points towards a highly effective and refined model. This achievement demonstrates the efficacy of deep learning methodologies in enhancing the model's capacity for nuanced interpretation of facial cues. Figure 4 demonstrate the process of using the proposed model on the CK+ dataset, the 100% accuracy rate underscores the efficiency of the learning process, highlighting its robust ability to capture and comprehend diverse facial expressions, including happiness, sadness, anger, and more.

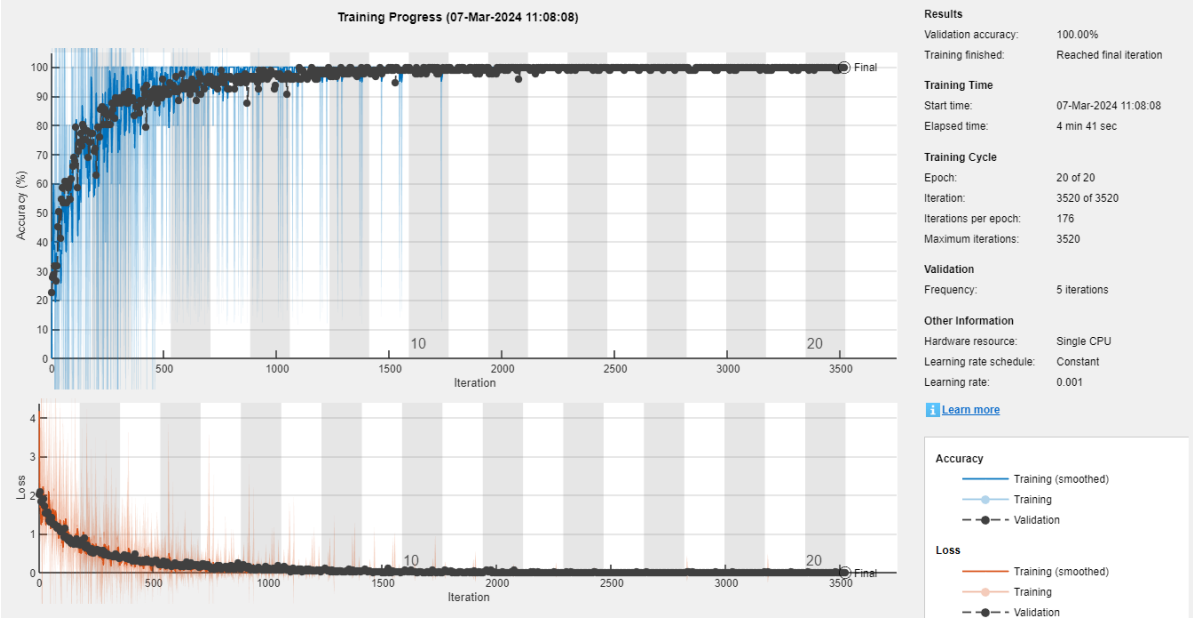

Figure 4: Learning rate process-CK48 dataset.

4. Facial Emotional Databases

One of the most important characteristics of this deep learning technology is the ability to modify and propose a new structure using the available layers so that a new structure is formed that fits the current work. For this, researchers need to face emotional databases to train the neuronal network with models to ensure the effectiveness of the newly proposed method and compare it with the results of other methods. In this research, three FER datasets were used to test the proposed method for improving facial expression recognition, as follows:

4.1 Japanese Female Facial Expression (Jaffe) Dataset

The JAFFE consists of 213 grayscale face photos captured by 10 Japanese women, each displaying one of seven basic human emotions: happiness, neutrality, fear, sadness, disgust, surprise, or anger [13]. Examples of collected JAFFE are shown in Figure 5.

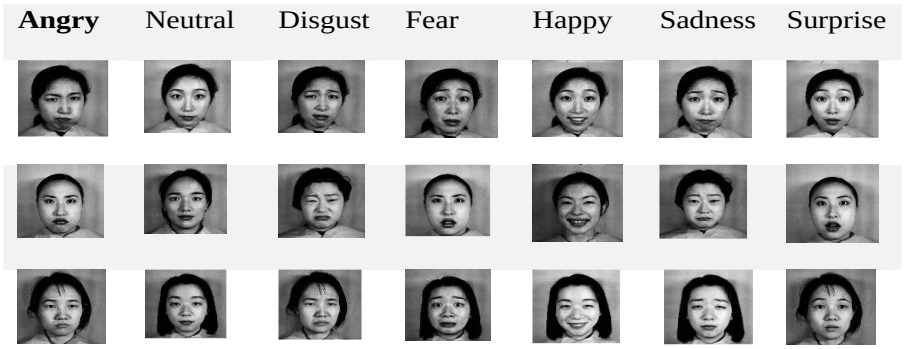

Figure 5: Examples of the JAFFE emotional dataset.

4.2 CK+ FER Dataset

The Extended Cohn-Kanade, often known as the CK+ database, contains 593 photos altogether from 123 individuals that represent a human facial expression based on the subject's perception of each of the seven fundamental feelings. Ck+ dataset examples are shown in Figure 6 based on the seven emotional categories [16].

Figure 6: Examples of the CK+ dataset.

4.3 FER2013 Dataset

It is considered an open international image database, where they were collected through the Google search engine. These images express the expressions of people collected for Kaggle competitions. These images were collected in a real environment, not prepared in advance, with different lighting and makeup and different categories of people such as children, youth, and the elderly, but with a 48*48 grayscale for all images. The collected images (totaling 35887 images) were divided into seven categories of expression of human feelings, as follows: For happy expression, 8989 images were collected; 6198 images were collected for neutral; 6077 images were collected for sadness, 547 images were collected for disgust; 4953 images were collected for angry; 4002 images were collected for surprise; and finally, 5121 pictures were collected for the scary category. [16, 17]. Figure 7 shows partial images of FER2013.

Figure 7: Examples of the FER2013 dataset.

5. Results

In this research, three databases were used in the process of the experiment: JAFFE (Japanese Female Facial Expression Database), CK + (Cohn–Kanade Plus), and Facial Expression Recognition Dataset 2013 (Fer 2013). The suggested network was assessed on the matching classification task by utilizing the datasets we trained, and the outcomes were contrasted with pertinent studies in the domain. These findings are succinctly presented in Table 1.

Table 1: Proposed Model and Related Works Results.

Author-Reference	Technology used	Dataset	Results
Linging Liu [3]	Deep convolution neural network	Fer2013	49.8%
Lopes et al.[10]	Combination of Convolutional Neural Network and specific image pre-processing	CK+ and JAFFE	98.80% and 82.10%
JI-HAE KIM et al. [11]	hierarchical deep neural network structure	CK+ and JAFFE	96.5% and 91.3%
Zhiheng Zhang., and Ming Li[12]	classic AlexNet, Improved AlexNet	JAFFE	88% and 96%
Boughida Adil, and et al. [13]	Gabor filter features, PCA, and SVM for classification	JAFFE and CK+	95.11% and 92.19%
Shane F. Cotter[14]	MobiExpressNet	Fer2013	67.96%
Rahul Ravi, and et al. [15]	CNN	CK+, JAFFE, and YALE FACE	97.32%, 77.27% and 31.82
Proposed architecture	CNN layering architecture	Fer2013, JAFFE and CK+	69.1%, 97.27%, and 100%.

As shown in the table above, this study achieved the highest accuracy compared to the other studies that used similar datasets and technology. It is worth noting that the accuracy results of these studies may be influenced by various factors, such as the size and quality of the datasets, the preprocessing techniques used, and the parameters of the machine learning algorithms. Nevertheless, based on the results of this comparison, it appears that the study has achieved higher accuracy than the other studies that have used similar datasets and technology.

6. Conclusion

In conclusion, this study has effectively demonstrated the efficacy of facial expression recognition through cutting-edge machine learning methodologies. The achieved results showcase a remarkable level of accuracy in the identification and classification of diverse facial expressions across the CK+, Fer2013, and JAFFE datasets. This accomplishment holds significant implications for various academic and practical domains, including psychology, human-computer interaction, and security systems. Moreover, the inclusion of preprocessing techniques, such as data augmentation and normalization, has played a pivotal role in augmenting the models' performance. These techniques contribute to the overall robustness and reliability of the facial expression recognition system developed in this research, enhancing its potential for broader practical applications. The success of

the study is attributed to the careful use of a powerful dataset comprising a large number of images and the implementation of cutting-edge deep learning algorithms, in particular convolutional neural networks. It is worth mentioning that the system, which is proposed in the present paper, has controlling management exceeding the state-of-the-art.

7. Authors' Contribution

"I confirm that the manuscript has been read and approved by author."

8. Conflict of Interest

The authors declare that there is no conflict of interest for this paper.

References

[1] S. Begaj, A. O. Topal, and M. Ali, "Emotion Recognition Based on Facial Expressions Using Convolutional Neural Network (CNN)," in *2020 International Conference on Computing, Networking, Telecommunications & Engineering Sciences Applications (CoNTESA)*, 2020, pp. 58-63. https://ieeexplore.ieee.org/document/9302866 .

[2] W. Mellouk and W. Handouzi, "Facial emotion recognition using deep learning: review and insights," *Procedia Computer Science,* vol. 175, pp. 689-694, 2020. https://doi.org/10.1016/j.procs.2020.07.101.

[3] L. Liu, "Human face expression recognition based on deep learning-deep convolutional neural network," in *2019 International Conference on Smart Grid and Electrical Automation (ICSGEA)*, 2019, pp. 221-224. https://ieeexplore.ieee.org/document/8901324 .

[4] L. Sun, C. Ge, and Y. Zhong, "Design and implementation of face emotion recognition system based on CNN Mini_Xception frameworks," in *Journal of Physics: Conference Series*, 2021, p. 012123. https://iopscience.iop.org/article/10.1088/1742-6596/2010/1/012123.

[5] A. Fathallah, L. Abdi, and A. Douik, "Facial expression recognition via deep learning," in *2017 IEEE/ACS 14th International Conference on Computer Systems and Applications (AICCSA)*, 2017, pp. 745-750. https://ieeexplore.ieee.org/document/8308363.

[6] S. Liu, D. Li, Q. Gao, and Y. Song, "Facial emotion recognition based on cnn," in *2020 Chinese Automation Congress (CAC)*, 2020, pp. 398-403. https://ieeexplore.ieee.org/document/9327432 .

[7] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Communications of the ACM,* vol. 60, pp. 84-90, 2017. https://doi.org/10.1145/3065386.

[8] N. Y. Abdullah and A. M. F. Alkababji, "Masked face with facial expression recognition based on deep learning," *Indonesian Journal of Electrical Engineering and Computer Science,* vol. 27, pp. 149-155, 2022. DOI: http://doi.org/10.11591/ijeecs.v27.i1.pp149-155.

[9] T. K. Arora, P. K. Chaubey, M. S. Raman, B. Kumar, Y. Nagesh, P. Anjani*, et al.*, "Optimal facial feature based emotional recognition using deep learning algorithm," *Computational Intelligence and Neuroscience: CIN,* vol. 2022, 2022. https://doi.org/10.1155/2022/8379202.

[10] A. T. Lopes, E. De Aguiar, A. F. De Souza, and T. Oliveira-Santos, "Facial expression recognition with convolutional neural networks: coping with few data and the training sample order," *Pattern recognition,* vol. 61, pp. 610-628, 2017. https://doi.org/10.1016/j.patcog.2016.07.026.

[11] J.-H. Kim, B.-G. Kim, P. P. Roy, and D.-M. Jeong, "Efficient facial expression recognition algorithm based on hierarchical deep neural network structure," *IEEE access,* vol. 7, pp. 41273-41285, 2019. https://ieeexplore.ieee.org/document/8673885 .

[12] Z. Z. M. Li, "Research on Facial Expression Recognition Based on Neural Network," presented at the International Conference on Computer Network, Electronic and Automation (ICCNEA), Xi'an, China, 2020. https://ieeexplore.ieee.org/document/9239777 .

[13] B. Adil, K. M. Nadjib, and L. Yacine, "A novel approach for facial expression recognition," in *2019 International Conference on Networking and Advanced Systems (ICNAS)*, 2019, pp. 1-5. https://ieeexplore.ieee.org/document/8807883 .

[14] S. F. Cotter, "MobiExpressNet: A deep learning network for face expression recognition on smart phones," in *2020 IEEE International Conference on Consumer Electronics (ICCE)*, 2020, pp. 1-4. https://ieeexplore.ieee.org/document/9042973.

[15] R. Ravi and S. Yadhukrishna, "A face expression recognition using CNN & LBP," in *2020 fourth international conference on computing methodologies and communication (ICCMC)*, 2020, pp. 684-689. https://ieeexplore.ieee.org/document/9076422.

[16] H. Huo, Y. Yu, and Z. Liu, "Facial expression recognition based on improved depthwise separable convolutional network," *Multimedia Tools and Applications,* pp. 1-18, 2022. **DOI:** https://doi.org/10.1007/s11042-022-14066-6.

[17] F. Y. Zhou Yue, Zeng Shangyou, Pan Bing, "Facial Expression Recognition Based on Convolutional Neural Network," presented at the 10th International Conference on Software Engineering and Service Science (ICSESS), Beijing, China, 2019. https://ieeexplore.ieee.org/document/9040730 .