

## Research Article

# Impact of Activation, Optimization, and Regularization Methods on the Facial Expression Model Using CNN

**Irfan Ali Kandhro,<sup>1</sup> Mueen Uddin<sup>2</sup>, Saddam Hussain<sup>3</sup>, Touseef Javed Chaudhery,<sup>4</sup> Mohammad Shorfuzzaman<sup>5</sup>, Hossam Meshref,<sup>5</sup> Maha Albalhaq,<sup>6</sup> Raed Alsaqour,<sup>7</sup> and Osamah Ibrahim Khalaf<sup>8</sup>**

<sup>1</sup>Department of Computer Science, Sindh Madressatul Islam University, Karachi, Pakistan

<sup>2</sup>College of Computing and IT, University of Doha for Science and Technology, Doha, Qatar

<sup>3</sup>Department of Information Technology, Hazara University, Mansehra 21120, KP, Pakistan

<sup>4</sup>National University of Modern Languages, Islamabad, Pakistan

<sup>5</sup>Department of Computer Science, College of Computers and Information Technology, Taif University, Taif 21944, Saudi Arabia

<sup>6</sup>Department of Information Technology, College of Computer and Information Sciences, Princess Nourah Bint Abdulrahman University, P.O. Box 84428, Riyadh 11671, Saudi Arabia

<sup>7</sup>Department of Information Technology, College of Computing and Informatics, Saudi Electronic University, 93499, Riyadh, Saudi Arabia

<sup>8</sup>Al-Nahrain University, Al-Nahrain Nanorenewable Energy Research Center Baghdad, Baghdad, Iraq

Correspondence should be addressed to Saddam Hussain; saddamicup1993@outlook.com

Received 26 February 2022; Accepted 30 April 2022; Published 16 June 2022

Academic Editor: Muhammad Zubair Asghar

Copyright © 2022 Irfan Ali Kandhro et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

When it comes to conveying sentiments and thoughts, facial expressions are quite effective. For human-computer collaboration, data-driven animation, and communication between humans and robots to be successful, the capacity to recognize emotional states in facial expressions must be developed and implemented. Recently published studies have found that deep learning is becoming increasingly popular in the field of image categorization. As a result, to resolve the problem of facial expression recognition (FER) using convolutional neural networks (CNN), increasingly substantial efforts have been made in recent years. Facial expressions may be acquired from databases like CK+ and JAFFE using this novel FER technique based on activations, optimizations, and regularization parameters. The model recognized emotions such as happiness, sadness, surprise, fear, anger, disgust, and neutrality. The performance of the model was evaluated using a variety of methodologies, including activation, optimization, and regularization, as well as other hyperparameters, as detailed in this study. In experiments, the FER technique may be used to recognize emotions with an Adam, Softmax, and Dropout Ratio of 0.1 to 0.2 when combined with other techniques. It also outperforms current FER techniques that rely on handcrafted features and only one channel, as well as has superior network performance compared to the present state-of-the-art techniques.

## 1. Introduction

According to everyone's common knowledge, the advancement of computer technology has greatly facilitated the advancement of various sectors, including artificial intelligence and pattern classification [1]. In order to achieve a natural connection, there must be an amicable relationship between the human and the machine. Mehrabian [2] found

that facial emotions carry 55% of the useful information in communication, whereas sound and language convey just 38% and 7%, respectively, of this information. As a result, facial expressions convey a great deal of emotional information. Facial emotion recognition has been investigated extensively in the past few decades, and it has gained more and more scholars' attention in the process [3–6].

When it comes to nonverbal communication, facial expression recognition is a powerful tool for conveying emotions, states, and intentions. Numerous studies have been undertaken on autonomous facial expression in sociable robotics, data analytics, human-computer vision, medical therapy, and driver fatigue surveillance because of its significance. Automated FER has since been extensively examined and used to encode facial expression characteristics [7–9]. Indeed, it is possible to tackle the challenge by identifying basic expressions under regulated conditions, such as frontal faces and side pose emotions. There were six basic emotions described in the twentieth century by Ekman and Friesen based on cross-cultural research [10]. In psychology, the term “basic expression” refers to a set of facial expressions that can be used to indicate a wide range of human emotions. Advances in neuroscience and psychology have shown that these six emotions are culturally specific rather than universally applicable [11].

There are still problems with the action coding system and continuous approach when it comes to describing emotions in real-world situations [12–15]. The effect model, on the other hand, does not show how complex or sensitive our effective displays are. Complexity and changes in head pose, lighting, and occlusion are to bear responsibility. People have different ways of expressing themselves, and the background brightness and background color, the position of the image, and many other factors can all change the way an image is analyzed. The fact that unposed expressions are often subtle also affects how the image is analyzed. So, the reliable automated FER system is very important in the applications [16, 17].

The performance, speed, and intelligence requirements of traditional machine learning algorithms can no longer be met in the age of big data. This is especially true in the domains of identification, classification, and target detection, where deep learning has showed exceptional information processing abilities to get better at classification and prediction in the long term, deep learning can develop more abstract high-level features and attribute information [18]. Image features can be consistently extracted using the convolutional neural network (CNN) [19], which is a deep learning architecture. It has been widely used both in academic circles and in actual business applications, especially in the field of computer vision [20–27].

Comprehensive surveys on automatic expression analysis have been published in recent days. These surveys have focused on and established a set of standard algorithms for automated facial expression [28]. The convolutional neural network (CNN) has enabled significant performance in related tasks [29–32]. Current studies extensively use CNN methods to analyze various features and extract essential details from facial expressions while evaluating different datasets [31, 33]. These works differ considerably in terms of the CNN structure, preprocessing, training, testing, and validating protocols. Furthermore, it is not plausible to compare the model’s performance in a single experiment based on the reported results found on existing issues and bottlenecks in existing CNN architectures, consequently increasing the FER model’s performance.

The followings are some of our significant contributions:

- (i) We focused on the performance of the facial expression recognition model on optimization, regularization, and activation parameters by reviewing the existing convolutional neural network-based methods.
- (ii) We compare convolutional neural network methods by highlighting the differences empirically under consistent settings.
- (iii) Moreover, based on this, we identified the dodge and directions for improving the model’s performance.
- (iv) Eventually, we confirm rationally that overcoming issues such as bottlenecks enhances the performance significantly.
- (v) The proposed convolutional neural network architecture model achieves state-of-the-art facial expression recognition as shown in experiments performed on various datasets.

This paper is divided into the following section. Section 1 describes an introduction about the research being done. Section 2 discusses the related work where past and present FER techniques have been discussed. Section 3 describes the proposed FER model with a detailed description of various components. The experimental results are discussed in Section 4, and Section 5 concludes the paper.

## 2. Related Work

For its superior performance in image processing, computer vision, and image classification, the CNN approach has been widely adopted in deep learning for these applications. A halftone image classification and image processing system, developed by Zhang et al. [34], was built to assess significant aspects of videos/images, and it performed exceptionally well. Unsupervised learning was used to extract features from halftone images, and he proposed using stacked sparse autoencoders (SAE). According to Khorrami et al. [35], it is possible to get good results from CNN if they are taught to look at a face and determine which elements influence CNN’s predictions. An initial step is to train a zero-biased CNN using facial expression data, which the authors use the expanded Cohn-Kanade (CK+) dataset and the Toronto Face Dataset as benchmarks. They next perform a qualitative examination of the network, observing the spatial patterns that most intensely stimulate the different types of neurons in the convolutional layers and show how they mimic facial action units (FAUs). A final step in this process is to verify that all FAUs visible in filter visualization correlate with facing movements in the CK+ dataset. A flexible hypothesis pooling approach for image multiclassification was developed by Wei et al. [36]. The model accepts any number of object segment assumptions made about the object as the input. Each hypothesis was linked to a shared CNN in a step-by-step fashion. Finally, the model varied hypothesis outcomes are averaged with max pooling, resulting in classic predictors for multilabel predictions that are based on classic

predictors. Face alignment, face detection, and face recognition are only a few of the issues in FER-related applications, among others. Characters can now express themselves in human expressions, thanks to a new method developed by Aneja et al. [37]. It begins with a training phase in which two CNNs are trained to recognize human and stylized character expressions. After that, they develop a shared embedding feature space by employing a technique called transfer learning. This involves learning the mapping between characters and persons. This embedding additionally enables the retrieval of pictures based on human expressions as well as pictures based on character expressions. To obtain human-like character expressions, the authors used a perceptual model. Finally, the authors use the newly acquired stylized character expression dataset to evaluate their method on various retrieval tasks. The authors also give evidence that the ranking order of the proposed attributes has a strong correlation with the ranking order that was provided by an expert on facial expressions and mechanical Experiments.

Li et al. [38] developed a CNN-based cascade model with a robust discriminative capability to maintain high performance while dealing with the problem of changes in visual properties due to changes in expressions, pose, and lighting in original face recognition. A deep learning model for face adjustment and alignment with landmark characteristics and recurrent recognition has been proposed by Parka et al. [39]. Chen et al. [40] use deep learning to demonstrate an effective method for recognizing smiles in the wild. Unlike previous works that collected handcrafted features from face pictures and trained a classifier to undertake smile detection in a two-step process, deep learning may be able to merge feature learning and classification into a single model rapidly. The authors use the deep convolutional network to overcome this popular deep learning model. Smile CNN is a deep convolutional network created by the authors to feature learning and smile detection simultaneously. Although a deep learning model is typically designed to handle “big data,” experimental results show that the model can also effectively manage “small data.” They are now looking at the discriminative power of the learned features, which are derived from the cell activations of our SmileCNN’s last hidden layer. We demonstrate that the learned features have the excellent discriminative capability to train an SVM or AdaBoost classifier. Experiments on the GENKI4K database show that the proposed method can achieve promising performance in smile recognition. Pang et al. [41] provided a solution for visual target tracking tasks based on the CNN DL algorithm. On real-time visual tracking, the model obtained state-of-the-art performance. Another difficulty for video analysis is detecting human activity, which is currently being investigated in many studies. Ronao et al. [42] demonstrated an efficient and effective human activity detection system based on smartphone sensors that took advantage of the inherent properties of activities and ID time-series signals to identify them. On several experimental databases, this approach produced state-of-the-art outcomes that were previously unattainable.

### 3. Proposed Work

This research suggested a unique and novel architecture for evaluating facial expressions using a convolutional neural network implemented in a simulation environment. The process provided in our proposed model initially obtains a fresh raw image, referred to as image acquisition, from a variety of various datasets to ensure that our model is not biased toward any particular dataset. During this phase, our suggested model begins evaluating the selected image to detect the presence of a face. If it recognizes a face in the chosen image using a cascade classifier, it is passed to the 2nd phase for further preprocessing, where it is further refined. Image preprocessing is carried out during this phase, with several distinct stages being done, as illustrated in Figure 1. The facial expressions detected will be improved by utilizing various tools and techniques such as crop, rotate, flip, and stretch from the detected face. Afterward, the selected facial expressions are registered, and landmarks are recognized through normalization and magnification techniques. Micro and macro spotting are carried out on the landmarks chosen to extract the relevant and essential features in the following process phase. The final phase involves feeding the proposed CNN model with more data to forecast and categories the class.

### 4. Methodology

The proposed model comprises the following components used to perform various functionalities while evaluating the facial expression process from an image. These components are explained below.

**4.1. Database.** The Japanese female facial expression (JAFFE) dataset contains 213 samples of posture expression from 10 Japanese ladies, which are well-managed and laboratory-controlled. Six primary facial expressions are represented in three-quarters of each female image. These include smiles and scowls as well as angry, fearful, surprised, and disgusted looks. Since there are so few samples per participant and expression, the dataset poses a significant challenge. Laboratory-managed databases like the Extended Cohn-Kanade (CK+) are frequently consulted in the evaluation of FER systems. There are 593 video clips from 123 subjects in the dataset. Between 10 and 60 frames, the expressions change from neutral to ecstatic. There were 327 sequences from 118 people that were categorized with 7 basic emotion labels: contempt; anger; happiness; surprise; sorrow; and disgust [43, 44].

**4.2. Preprocessing.** Unconstrained scenarios frequently feature differences in lighting, noise, head postures, and backgrounds that have little to do with facial expressions. Preprocessing is needed to align and standardize the visual semantic input before training the FER model on CNN. The following processes are included in the preprocessing phase, as outlined below.

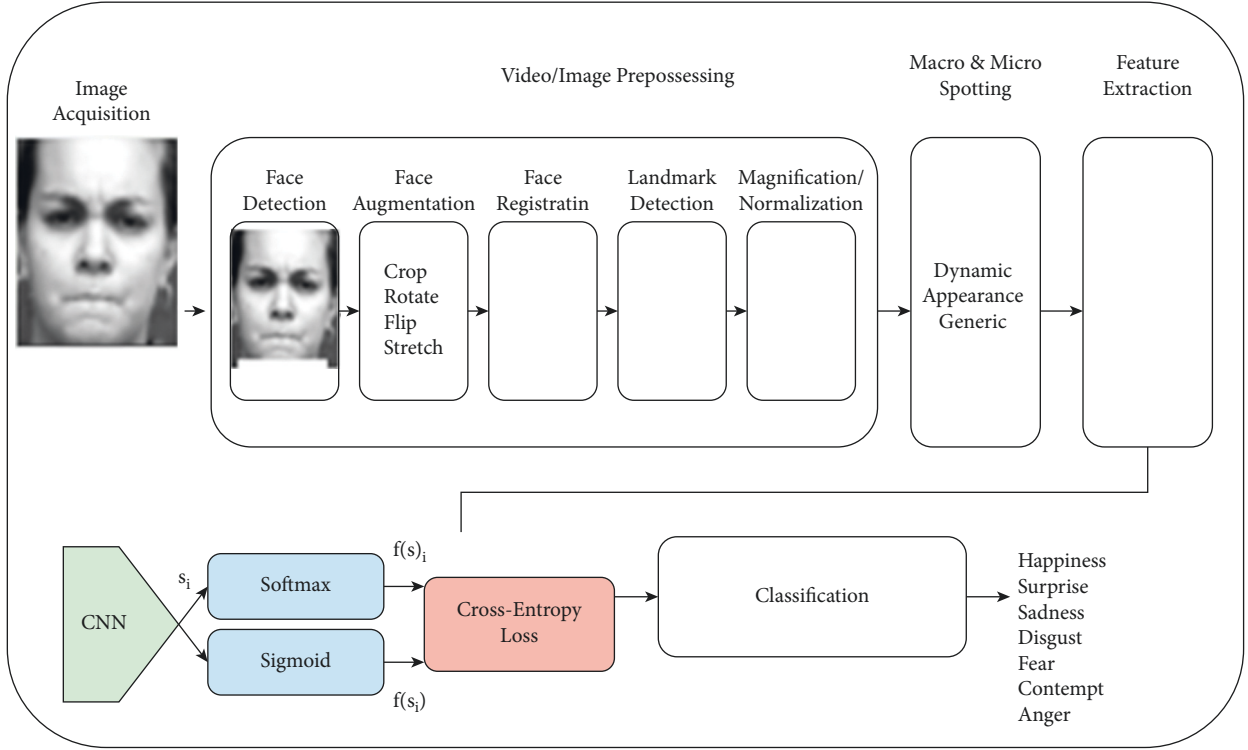


FIGURE 1: Proposed system architecture.

- (1) Face detection is the foremost step in computer vision that locates a face area from the image. It states to find the face coordinates in the image, whereas localization refers to demarcating the extent of the face. The Viola–Jones (V&J) face detector is a classic and widely employed detection method [45].
- (2) In a deep learning-based FER system, data augmentation is essential. Whereas, large samples are needed to train the CNN model and provide generalizability to a specific recognition challenge. It is necessary to flip and crop the input images before they can be used in the machine learning process.
- (3) Face registration is a traditional preprocessing step in the face recognition task. Registration is done to align sample faces to a reference face. The subjects need not cooperate with data acquisition in the natural FER systems.
- (4) The eyes, mouth, nose, and eyebrows are all instances of facial landmarks that are used to locate and represent the most important parts of the face. We find the position of the subject's head and neck in the picture and note any distinctive features of their faces ROI.
- (5) The FER performance can be hampered by differences in illumination and head positions, which can result in considerable picture fluctuations. As a result, we present two common methods for normalizing faces to minimize these variations: normalization of lighting and head pose [46].

## 5. Convolutional Neural Networks

Deep learning methods use machine learning algorithms for forming high-level abstractions in emotions and pattern recognition in images and text [32]. The learning levels take the results of the previous levels as input, which are then converted and transformed into intuitions to train further and validate the classification model. The CNN model processes various types of information. The functionality of CNN creates complex layers to represent and process the complex data. Figure 1 shows the architecture of our face emotion detection FED model. The input of our model is a  $(48 \times 48)$  grayscale image with some regularizations and optimization methods for the training system to understand and analyze the features. The output shows results in a single output class from seven emotions. The CNN network is composed of three convolutional layers, numbered C1 through C3, three MaxPooling layers, numbered P1 through P3, and four Relu activation functions, numbered R1 through R4. Additionally, as can be seen in Figure 1, all of these layers have complete connectivity between the input and output.

The C1 convolutional layer used filters on the input image, the size of the input image is  $48 \times 48$ , and learnable kernel/filters size used is  $3 \times 3$  and produced 32 metrics with the size of  $(62 \times 62)$ . The results of the previous layer are used as input to the Relu activation layer, which fundamentally transforms the small range when the gradient is nonzero. The output Relu layer is used to the MaxPooling layer, which uses  $2 \times 2$  learnable kernels. The output size of MaxPooling P1 matrices is  $(48 \times 48)$ . Similarly, these results are passed

TABLE 1: CNN configuration.

Layer (type)	Output shape	Param #
conv2d (Conv2D)	(None, 128, 128, 6)	456
max_pooling2d (MaxPooling2D)	(None, 64, 64, 6)	0
conv2d_1 (Conv2D)	(None, 64, 64, 16)	2416
Activation (activation)	(None, 64, 64, 16)	0
max_pooling2d_1 (MaxPooling2)	(None, 32, 32, 16)	0
conv2d_2 (Conv2D)	(None, 30, 30, 64)	9280
max_pooling2d_2 (MaxPooling2)	(None, 15, 15, 64)	0
Flatten (flatten)	(None, 14400)	0
Dense (dense)	(None, 128)	1843328
Dropout (dropout)	(None, 128)	0
dense_1 (dense)	(None, 7)	903

TABLE 2: CNN results with CK+ and JAFEE.

Dataset	Epochs	Batch size	Test CNN Acc (%)	Test CNN loss
CK+	15	8	93	0.07
	20	16	94	0.06
	25	32	95	0.05
	30	64	94	0.06
	35	128	93	0.07
	40	256	96	0.07
	45	512	80	0.20
	50	1024	97	0.03
JAFEE	15	8	45	0.55
	20	16	48	0.52
	25	32	60	0.40
	30	64	49	0.51
	35	128	65	0.35
	40	256	42	0.58
	45	512	60	0.40
	50	1024	52	0.48

TABLE 3: Hyperparameters testing accuracy on the JAFEE dataset.

Activation	Optimization	Dropout	Testing accuracy
Softmax	Adam	0.1	0.94%
Softmax	AdaGrad	0.1	0.91%
Softmax	Nadam	0.1	0.92%
Softplus	Adamax	0.1	0.91%
Sigmoid	Adam	0.1	0.90%
Relu	Adam	0.1	0.93%
Relu	AdaGrad	0.1	0.91%
Relu	Nadam	0.1	0.90%
Relu	Adamax	0.1	0.90%
Softmax	Adam	0.2	0.95%
.....	.....	.....	.....
hard_sigmoid	Adamax	0.4	0.88%

through the next convolutional layer C2 layer, with the same parameters. Finally, C3 and P3 layers use learnable kernels of size  $(3 \times 3)$  and  $(2 \times 2)$ . Next, the flatten layer gives 4608 values; the first layer gives 2304 hidden units, and the other 1152. Lastly, the output layers generate seven classes, as shown in Table 1.

**5.1. Hyperparameters.** For the facial expression database, our suggested model utilizes a variety of hypermeters. Before

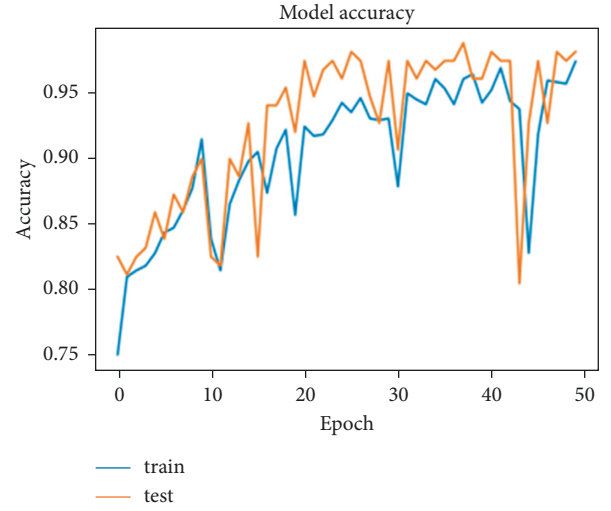


FIGURE 2: CNN model CK+'s accuracy.

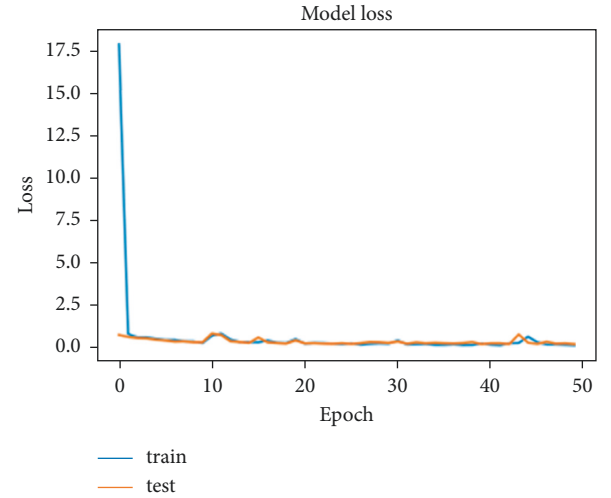


FIGURE 3: Loss of CNN model CK+.

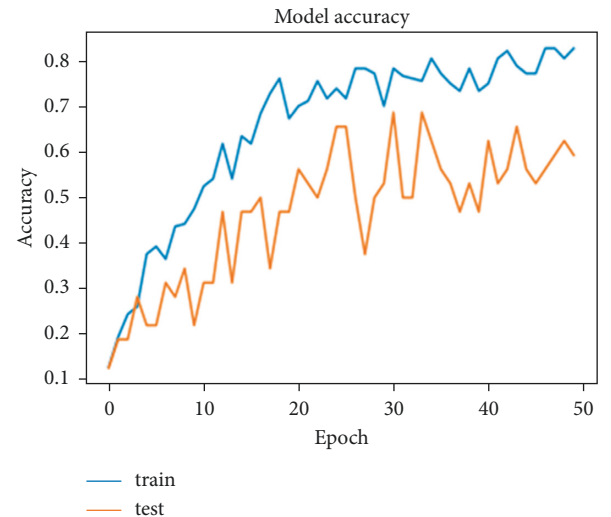


FIGURE 4: Accuracy of the CNN model with CK + dataset.

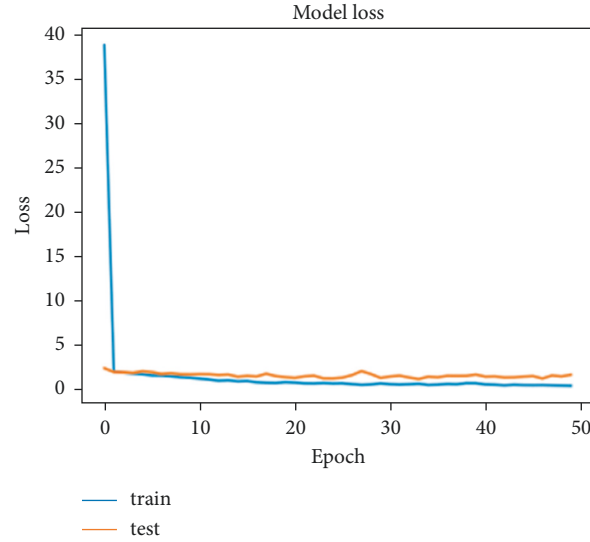


FIGURE 5: Loss of the CNN model with JAFFE dataset.

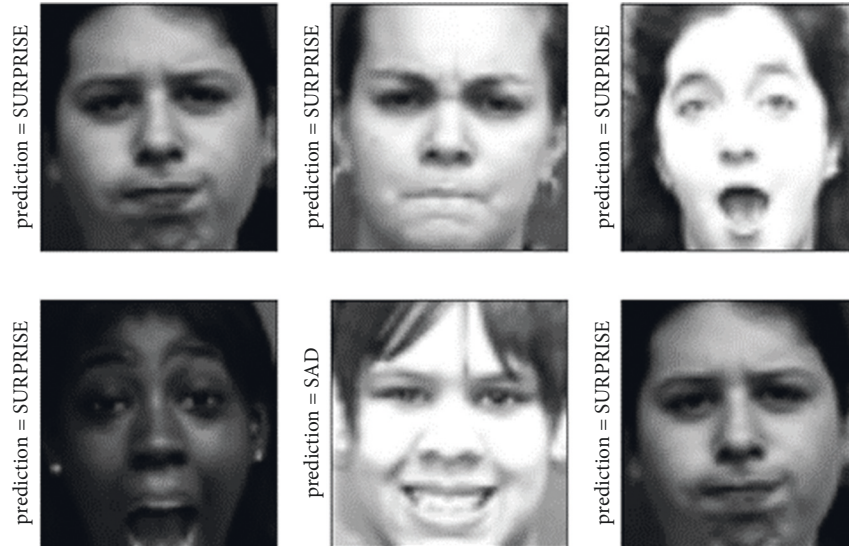


FIGURE 6: Predication of model CK+.

going into the specifics of the model's performance on several databases, we briefly covered our training approach. Each dataset in our experiment was used to train our model; however, we made an effort to keep the architecture and hyperparameters consistent across models. From scratch, each model received 50 training epochs. We use a random Gaussian with zero and standard derivation to set the network weights in the beginning. Neuronal outputs were shaped by the activation function in the model, i.e., (Adam, AdaGrad, Nadam, and AdaMax). The output findings are shifted nonlinearly in response to their magnitude. By increasing in amplitude, signals spread and took on the shape of the final prediction of the network.

The entire demonstration of the CNN model is exceedingly complex and nonlinear because of the activation function. To reduce the model's error, the optimization functions Softmax, Softplus, Sigmoid, and Relu work

together. In comparison to other optimization algorithms, Adam's performance appears to be the best with a learning rate of 0.003 and weight decay. With GPU, our model can be trained in 10 minutes on the FER database (using JAFFE and CK+). There are few samples; thus, the model can be completed within ten minutes. The regularization or shrinking strategy was also employed to avoid overfitting the model, i.e., setting coefficients to 0. (dropout used between 0.1 and 0.4). Softmax activation is used for multi-classification in the dense layer after experimenting with a variety of combinations.

## 6. Results and Discussion

Experiments on two important facial expression recognition datasets are presented in this section, which includes the results of our model's evaluation. A brief overview of



database concerns and challenges is presented, followed by an evaluation of the FER model's capabilities on two real-world datasets, CK+ and JAFFE, using a variety of hyperparameters. As FER work moves its focus to challenging environmental circumstances, many researchers are turning to deep learning apparatus in order to deal with challenges such as illumination variance, occlusions, nonfontal head positions identification bias, and recognition of low-intensity expressions. When using deep learning, it is essential to have a significant number of training examples available in order to accurately catch tiny expression-related bends. The primary challenge for deep FER systems is a deficiency in the quantity and quality of the training data.

We have split the dataset into 70% and 30% for the training and testing on both datasets. Table 2 shows the comparison of the model on CK+ and JAFFE datasets with epoch and batch size parameters, as table data depicted the performance of model outstanding with batch 128, 1024, and epoch 35, 50 on both databases. On the CK+ dataset, our model achieved 97% testing accuracy with batch size 1024 and epoch 50 and a loss of 0.03. On the other hand, the JAFFE dataset with batch size 1024 and epoch 50 models recorded 65% and loss of 0.35.

Table 2 shows the experimental results on the CK+ database's various basic and advanced parameters. The model is tested on 80 different combinations of parameters with distinct batch sizes and the number of epochs.

The approaches of regularization, optimization and activation are all used in this model. These are the activation functions that have been used: (Adam, AdaGrad, Nadam, Adamax, and hard Sigmoid), and the optimization procedures that have been used: (Softmax, Softplus, Sigmoid, and Relu). There is a difference between the regularization dropout ratio and (0.1 to 0.4). Softmax, Adam, and dropout values of 0.1 and 0.2 provide state-of-the-art accuracy for the FER model, which is the most accurate model currently available. When Softmax and Adam were used, the model claimed 94% testing accuracy with a dropout value of 0.1. This is depicted in Table 3.

Figures 2 and 3 illustrate the performance of a basic CNN model on CK+ in terms of loss and accuracy during testing on 50 iterations with 32 batch sizes. The accuracy of the simple CNN model is increasing concerning epochs. Similarly, the loss is displayed during the learning that is decreasing in each epoch.

Besides, Figures 4 and 5 show the training 97% and testing 70% accuracy on the CNN model using the JAFFE dataset and similarly, training loss of 0.07 and validation loss of 2.02.

Figure 6 shows that the model was able to learn from the CK+ database [47].

## 7. Conclusions

The paper proposed a deep neural network (DNN) architecture for facial expression recognition (FER) based on convolutional neural networks (CNN). The CNN is one of the most representative network structures in the FER system and image processing in the deep learning field. The

paper examined the emotional interpretation learning capacity of three distinct techniques, namely, activation, optimization, and regularization. Specifically, it has been noted that just two comparisons are completed with exceptional precision in terms of training, testing, and validation out of a total of 80 comparisons performed. There was a remarkable 0.1 difference in the model's accuracy when using Adam, Relu, and dropout data from other models. Using the FER2013 dataset, we ran a thorough experiment that yielded 97% training accuracy and 70% testing accuracy. A net loss of 0.05 and 2.01 is also shown. Using a simple CNN model, it is also discovered that the HOG operator's findings are ineffective when the image size is tiny and the image quality is unclear.

## Data Availability

The data used in this research can be obtained from the corresponding authors upon request.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

This research was funded by Princess Nourah Bint Abdulrahman University Researchers Supporting Project number (PNURSP2022R97), Princess Nourah Bint Abdulrahman University, Riyadh, Saudi Arabia, and the Taif University Researchers Supporting Project number (TURSP-2020/79), Taif University, Taif, Saudi Arabia.

## References

- [1] Y. Wang, Y. Li, Y. Song, and X. Rong, "The influence of the activation function in a convolution neural network model of facial expression recognition," *Applied Sciences*, vol. 10, no. 5, p. 1897, 2020.
- [2] A. Mehrabian and J. A. Russell, *An Approach to Environmental Psychology*, the MIT Press, Cambridge, MA, USA, 1974.
- [3] K. Wang, X. Peng, J. Yang, D. Meng, and Y. Qiao, "Region attention networks for pose and occlusion robust facial expression recognition," *IEEE Transactions on Image Processing*, vol. 29, pp. 4057–4069, 2020.
- [4] A. Wood, M. Rychlowska, S. Korb, and P. Niedenthal, "Fashioning the face: sensorimotor simulation contributes to facial expression recognition," *Trends in Cognitive Sciences*, vol. 20, no. 3, pp. 227–240, 2016.
- [5] G. Xu, H. Yin, and J. Yang, "Facial expression recognition based on convolutional neural networks and edge computing," in *Proceedings of the 2020 IEEE Conference on Telecommunications, Optics and Computer Science (TOCS)*, vol. 11, pp. 226–232, IEEE, Shenyang, China, December 2020.
- [6] E. Cambria, D. Das, S. Bandyopadhyay, and A. Feraco, "Affective computing and sentiment analysis," *A Practical Guide to Sentiment Analysis*, Springer, Cham, Switzerland, pp. 1–10, 2017.
- [7] M. Kenji, "Recognition of facial expression from optical flow," *IEICE - Transactions on Info and Systems*, vol. 74, no. 10, pp. 3474–3483, 1991.

- [8] Y. Yacoob and L. Davis, "Computing Spatio-temporal representations of human faces," in *Proceedings of the Computer Vision and Pattern Recognition, 1994. Proceedings CVPR'94., 1994 IEEE Computer Society Conference On*, pp. 70–75, IEEE, Seattle, WA, USA, June 1994.
- [9] M. J. Black and Y. Yacoob, "Tracking and recognizing rigid and non-rigid facial motions using local parametric models of image motion," in *Proceedings of the Computer Vision, 1995. Proceedings., Fifth International Conference on*, pp. 374–381, IEEE, Cambridge, MA, USA, June 1995.
- [10] P. Ekman and W. V. Friesen, "A new pan-cultural facial expression of emotion," *Motivation and Emotion*, vol. 10, no. 2, pp. 159–168, 1986.
- [11] R. E. Jack, O. G. B. Garrod, H. Yu, R. Caldara, and P. G. Schyns, "Facial expressions of emotion are not culturally universal," *Proceedings of the National Academy of Sciences*, vol. 109, no. 19, pp. 7241–7244, 2012.
- [12] A. A. Khan, M. Uddin, A. A. Shaikh, A. A. Laghari, and A. E. Rajput, "MF-ledger: blockchain hyperledger sawtooth-enabled novel and secure multimedia chain of custody forensic investigation architecture," *IEEE Access*, vol. 9, Article ID 103637, 2021.
- [13] A. Norouzi, M. S. M. Rahim, A. Altameem et al., "Medical image segmentation methods, algorithms, and applications," *IETE Technical Review*, vol. 31, no. 3, pp. 199–213, 2014.
- [14] H. Kolivand, M. S. Sunar, A. Altameem, A. Rehman, and M. Uddin, "Shadow mapping algorithms: applications and limitations," *Applied Mathematics and Information Sciences*, vol. 9, no. 3, pp. 1307–1315, 2015.
- [15] A. Amanat, M. Rizwan, A. R. Javed et al., "Deep learning for depression detection from textual data," *Electronics*, vol. 11, no. 5, p. 676, 2022.
- [16] B. Martinez and M. F. Valstar, "Advances, challenges, and opportunities in automatic facial expression recognition," in *Advances in Face Detection and Facial Image Analysis*, pp. 63–100, Springer, Berlin, Germany, 2016.
- [17] A. Dhall, R. Goecke, J. Joshi, J. Hoey, and T. Gedeon, "EmotiW 2016: video and group-level emotion recognition challenges," in *Proceedings of the 18th ACM International Conference on Multimodal Interaction*, pp. 427–432, ACM, Tokyo Japan, November 2016.
- [18] B. Sullivan and E. Charniak, "An introduction to deep learning," *Perception*, vol. 48, no. 8, pp. 759–761, 2019.
- [19] A. G. Howard, "Some improvements on deep convolutional neural network based image classification," 2013, <https://arxiv.org/abs/1312.5402>.
- [20] T. Hussain, D. Hussain, I. Hussain et al., "Internet of things with deep learning-based face recognition approach for authentication in control medical systems," *Computational and Mathematical Methods in Medicine*, vol. 2022, Article ID 5137513, 17 pages, 2022.
- [21] M. M. Aslam, L. Du, Z. Ahmed, M. N. Irshad, and H. Azeem, "A deep learning-based power control and consensus performance of spectrum sharing in the CR network," *Wireless Communications and Mobile Computing*, vol. 2021, Article ID 7125482, 16 pages, 2021.
- [22] C. A. ul Hassan, J. Iqbal, S. Hussain, H. AlSalman, M. A. A. Mosleh, and S. Sajid Ullah, "A computational intelligence approach for predicting medical insurance cost," *Mathematical Problems in Engineering*, vol. 2021, Article ID 1162553, 13 pages, 2021.
- [23] J. Kainat, S. Sajid Ullah, F. S. Alharithi, R. Alroobaea, S. Hussain, and S. Nazir, "Blended features classification of leaf-based cucumber disease using image processing techniques," *Complexity*, vol. 2021, Article ID 9736179, 12 pages, 2021.
- [24] M. J. Ibrahim, J. Kainat, H. AlSalman, S. S. Ullah, S. Al-Hadhrami, and S. Hussain, "An effective approach for human activity classification using feature fusion and machine learning methods," *Applied Bionics and Biomechanics*, vol. 2022, Article ID 7931729, 14 pages, 2022.
- [25] F. Shah, Y. Liu, A. Anwar et al., "Machine learning: the backbone of intelligent trade credit-based systems," *Security and Communication Networks*, vol. 2022, Article ID 7149902, 10 pages, 2022.
- [26] F. Shah, A. Anwar, I. ul haq, A. S. Hussain, S. Hussain, and S. Al-Hadhrami, "Artificial intelligence as a service for immoral content detection and eradication," *Scientific Programming*, vol. 2022, Article ID 6825228, 9 pages, Scientific Programming, vol. 2022, 2022.
- [27] M. M. Bukhari, B. F. Alkhamees, S. Hussain, A. Gumaei, A. Assiri, and S. S. Ullah, "An improved artificial neural network model for effective diabetes prediction," *Complexity*, vol. 2021, Article ID 5525271, 10 pages, 2021.
- [28] S. Li and W. Deng, "Deep facial expression recognition: a survey," *IEEE Transactions on Affective Computing*, vol. 13, p. 1, 2020.
- [29] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Advances in Neural Information Processing Systems (NIPS)*, vol. 60, pp. 1097–1105, 2012.
- [30] K. He, X. Zhang, H. Ren, and J. Sun, "Deep residual learning for image recognition," *CoRR*, vol. 1512, 2015.
- [31] F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: a unified embedding for face recognition and clustering," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1045–1051, Boston, MA, USA, June 2015.
- [32] Y. Tang, "Deep learning using support vector machines," in *Proceedings of the International Conference on Machine Learning (ICML) Workshops*, pp. 1–9, June 2013.
- [33] Z. Yu and C. Zhang, "Image based static facial expression recognition with multiple deep network learning," in *Proceedings of the ACM International Conference on Multimodal Interaction (MMI)*, pp. 435–442, Seattle Washington USA, November 2015.
- [34] Y. Zhang, E. Zhang, and W. Chen, "Deep neural network for halftone image classification based on sparse auto-encoder," *Engineering Applications of Artificial Intelligence*, vol. 50, pp. 245–255, 2016.
- [35] P. Khorrami, T. Paine, and T. Huang, "Do deep neural networks learn facial action units when doing expression recognition?" in *Proceedings of the IEEE International Conference on Computer Vision (ICCV) Workshops*, pp. 19–27, Santiago, Chile, December 2015.
- [36] Y. Wei, W. Xia, M. Lin et al., "HCP: a flexible CNN framework for multi-label image classification," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 9, pp. 1901–1907, 2016.
- [37] D. Aneja, A. Colburn, G. Faigin, L. Shapiro, and B. Mones, "Modeling stylized character expressions via deep learning," in *Proceedings of the Asian Conference on Computer Vision*, pp. 136–153, Springer, Taipei, Taiwan, November 2016.
- [38] H. Li, Z. Lin, X. Shen, J. Brandt, and G. Hua, "A convolutional neural network cascade for face detection," in *Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5325–5334, Boston, MA, USA, June 2015.



- [39] B. H. Park, S. Y. Oh, and I.-J. Kim, "Face alignment using a deep neural network with local feature learning and recurrent regression," *Expert Systems with Applications*, vol. 89, pp. 66–80, 2017.
- [40] J. Chen, Q. Ou, Z. Chi, and H. Fu, "Smile detection in the wild with deep convolutional neural networks," *Machine Vision and Applications*, vol. 28, no. 1-2, pp. 173–183, 2017.
- [41] S. Pang, J. J. del Coz, Z. Yu, O. Luaces, and J. Díez, "Deep learning to frame objects for visual target tracking," *Engineering Applications of Artificial Intelligence*, vol. 65, pp. 406–420, 2017.
- [42] C. A. Ronao and S.-B. Cho, "Human activity recognition with smartphone sensors using deep learning neural networks," *Expert Systems with Applications*, vol. 59, pp. 235–244, 2016.
- [43] M. Lyons, S. Akamatsu, M. Kamachi, and J. Gyoba, "Coding facial expressions with gabor wavelets," in *Proceedings of the Automatic Face and Gesture Recognition, 1998. Proceedings. Third IEEE International Conference on*, pp. 200–205, IEEE, Nara, Japan, April 1998.
- [44] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews, "The extended cohn-kanade dataset (ck+): a complete dataset for action unit and emotion-specified expression," in *Proceedings of the Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on*, pp. 94–101, IEEE, San Francisco, CA, USA, June 2010.
- [45] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Computer Vision and Pattern Recognition*, *CVPR*, vol. 200, pp. 201–213, 2001.
- [46] B. K. Kim, S. Y. Dong, J. Roh, G. Kim, and S. Y. Lee, "Fusing Aligned and non-aligned face information for automatic affect recognition in the wild: a deep learning approach," in *Proceedings of the IEEE Conf. Computer Vision and Pattern Recognition (CVPR) Workshops*, pp. 48–57, Las Vegas, NV, USA, July 2016.
- [47] J. R. Lee, L. Wang, and A. Wong, "Emotionnet nano: an efficient deep convolutional neural network design for real-time facial expression recognition," *Frontiers in Artificial Intelligence*, vol. 3, Article ID 609673, 2020.

Copyright of Computational Intelligence & Neuroscience is the property of Hindawi Limited and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.