

## Facial Expression Recognition Using Data Augmentation and Transfer Learning

Layla A. Al Hak<sup>\*ID</sup>, Wasan Ahmed Ali<sup>ID</sup>, Samah J. Saba<sup>ID</sup>

College of Science, Computer science Department, Diyala University, Diyala 32001, Iraq

Corresponding Author Email: [laylaeabdalhaq@uodiyala.edu.iq](mailto:laylaeabdalhaq@uodiyala.edu.iq)



Copyright: ©2024 The authors. This article is published by IETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.18280/isi.290338>

### ABSTRACT

**Received:** 19 September 2023

**Revised:** 10 January 2024

**Accepted:** 23 January 2024

**Available online:** 20 June 2024

#### Keywords:

VGG-16, FER2013, facial emotion recognition, deep learning, FER, data augmentation

The world of technology has dramatically benefited from artificial intelligence's rapid development. Deep learning and machine learning algorithms have attained remarkable success in various applications, including recommendation system, pattern recognition, classification system, etc., where traditional algorithms have failed to fulfill the requirements of humans in real time. A person's thoughts, conduct, and feelings are greatly influenced by their emotions. In this work proposed method for FER is based on existing VGG16 and concatenates additional layers. The model consists of the preprocessing stage based on the median filter and data augmentation, and classification based on pertained VGG16 with transfer learning added four layers to the existing vgg16 that already trained on the ImageNet dataset. The model proposed attained an accuracy of 90% on FER2013 test set. It is observed that the proposed method outperformed the previous works.

## 1. INTRODUCTION

Communication is significantly influenced by emotions. Numerous tasks benefit from recognizing facial emotions, including security monitoring, criminal justice systems, e-learning, smart card applications, consumer satisfaction identification, and social robot [1]. The primary building components of the conventional emotion identification system are emotion classification, face detection, and feature extraction [2].

To address the problems with conventional techniques, deep learning networks employ an end-to-end learning process. The size of the data is very critical in deep learning [3, 4]. To enhance the deep learning performance, researchers are utilizing normalizations, translations, data augmentation, adding noise, scaling methods, and cropping [5] to increase the data size. CNNs are the best-proven methods in classification and segmentation tasks [6, 7].

One of the main advantages of CNN is the automatic feature extraction [8].

Automated FER is still an exciting and challenging issue in computer vision. Since people differ in how they show their expressions, FER is considered a complex issue for machine learning methods [9].

In this work, we produce an architecture network depend on VGG16 with transfer learning for facial expression recognition. We used the FER2013 dataset to analyze our design. The obtained findings demonstrate that the suggested method is particularly useful in image expression recognition using Fer 2013 dataset, resulting in enhancement in analysis of facial expression.

The paper is organized as follow: related work detailed in Section 2, Methodology in Section 3, suggested method in Section 4, and conclusion in Section 5.

## 2. RELATED WORK

Different researches have been presented for automatic FER.

### 2.1 Related work based on deep learning methods

Singh and Nasoz [9], demonstrate FER categorizing depending on static photos utilizing CNNs without feature extraction or preprocessing tasks. Based on a seven-class classification assignment, the authors achieved 61.7% accuracy on FER2013, while a state-of-the-art categorization achieved 75.2% accuracy.

Debnath et al. [10], is proposed FER system depend on Convolution Neural Network (CNN) called a Convnet of four layers to recognize seven facial emotions utilizing the fusion of convolution neural networks (CNN), local binary pattern (LBP) attributes, rotated BRIEF (ORB), and Oriented FaST. The method conducted on three dataset CK+, Jaffee. The results achieved are 98.13%, and 92.05% respectively.

Zahra et al. [11], this study illustrates the system design that can recognize and anticipate the facial emotion classification using feature extraction in realtime with the OpenCV library, specifically Keras and TensorFlow. The Raspberry Pi-based study design comprises of three key processes: facial feature extraction, face detection, and facial expression categorization. In research employing the CNN approach and FER-2013, the prediction outcomes of facial expressions were 65.97%.

Khan [12] used the state of arts ImageNet models and updates the classification layer with Progressive SpinalNet and SpinalNet architecture to enhance accuracy. The classification is done using the dataset FER2013, which is publicly available on Kaggle and contains over 35.000 face

image datasets for seven different emotions. The final model with Progressive SpinalNet and SpinalNetsur passed all existing single stand-alone model research on FER2013 after finishing the training procedure and fine-tuning its hyperparameter. VGG SpinalNet, one of the proposed designs, has the highest single network accuracy of 74.45%.

In the study by Al-Asbaily and Bozed [13] utilized a system that fused the Classic neural networks and VGG16 model. Where the VGG16 model is utilized for feature extraction and classic neural network was utilized to classification on the FER2013 database, the system attained an accuracy of 89.31%.

The method suggested by Wang [14], involves continual confrontation training between the generator and discriminator structures of Generative Adversarial Networks to enable improved extraction of visual characteristics from a detected input set. Then, high-accuracy face expression recognition is achieved. For simulation verification, the experimental section employs CK+, JAFEE, and the FER2013 dataset. The recognition approach has clear advantages in datasets of various sizes. The rates of average accuracy of recognition are 95.6%, 96.6%, and 72.8%, respectively.

### 3. METHODS

#### 3.1 Dataset

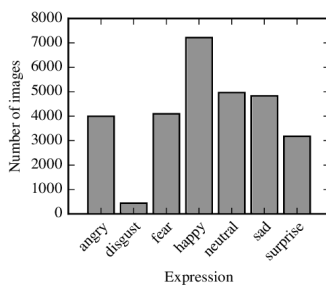


Figure 1. FER2013 dataset expression distribution [15]

The data utilized for the model was the FER2013 database obtained from the challenge of kaggle on FER2013 [15]. The dataset is utilized to combine the facial expression classification model. It comprises 35,887 images, divided into 3589 experiment and 28,709 pictures of trains. To indicate the final test contains 3589 test images. Figure 1 depicts the FER2013 dataset expression distribution.

#### 3.2 Median filter

Image filtering is a technique for reducing noise or artifacts, sharpening the contrast between adjacent regions, highlighting the contours with a particular orientation, and detecting edges. It involves convolving a kernel (square matrix) with an image. Median filtering is a nonlinear method of removing noise from images. It is widely used because it effectively reduces noise while preserving edges. It has been proven that median filtering is a dependable technique for eliminating impulsive noise without impairing edge features, and it is durable in the presence of high noise [16, 17].

#### 3.3 Image augmentation

Image data augmentation is a method that artificially increases training data size by updating dataset images. Data can result in more skillful deep learning neural networks, and augmentation approaches can provide variations of the images, allowing fit models to apply their learning to new images. Image data Generator class in the Keras deep learning neural networks Toolllkit enables us for the model fitting utilizing image data augmentations.

Deep learning algorithm, such as CNN, can learn characteristics that are independent of their placement in the image. Nonetheless, augmentation can help with this transform invariant learning method helps the model learn characteristics which are also transformed invariant, such as top-to-bottom to the left-to-right ordering, levels of light in images, and more.

Data augmentation is a basic image preprocessing method implemented online or offline. Offline augmentation methods are utilized to boost small dataset sizes, whereas online augmentation approaches are mainly used to increase the size of large datasets. Image data augmentation approaches generate from original data more training data while requiring no more storage memory. In most cases, generated photos are short batches destroyed after model training. The common methods to produce new images are: (1) Flip vertically or horizontally; (2) translate; (3) Crop randomly; (4) Scale inward or outward; (5) Rotate at some degrees; (6) Add Gaussian noise to avoid over fitting and improve the capability of learning [18, 19]. Example on image augmentation demonstrated in Figure 2.

In the proposed system we used flip vertically, shift, and rotate, zoom methods for data augmentation.

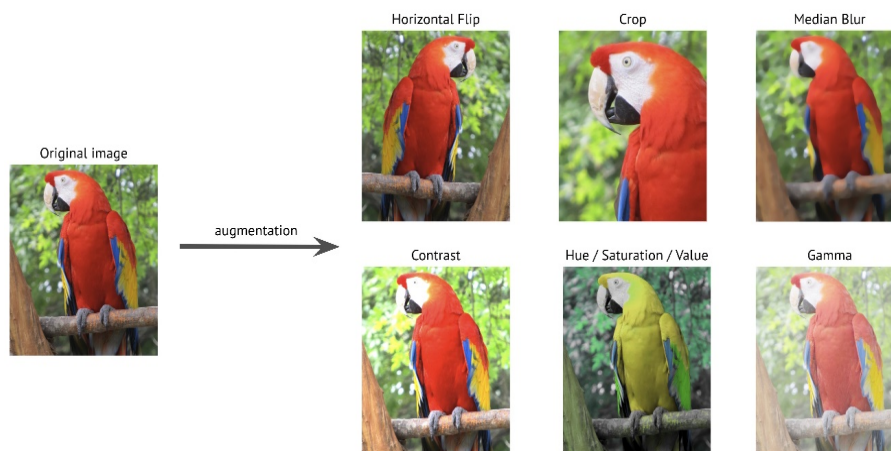


Figure 2. Example on image augmentation [18]

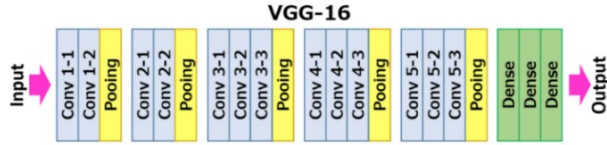


Figure 3. VGG16 architecture [23]

### 3.4 Transfer learning

Transfer learning another interesting model for preventing over fitting [20]. It operates by training a network on a large database such as Imagenet and then utilizing those weights as the initial weight in a new categorization task. The convolutional layer weights are typically copied rather than the complete network. This is highly useful because a lot of image datasets contain low-level spatial properties that may be taught more effectively with massive data. Understanding the link between domains of transmitted data is a work in progress.

Transfer learning is divided into fine tuning and feature extraction approach. The pre-trained weights and layers of VGG16 are included in our built model. We elected to freeze all pre-trained layers in this scenario. This model will extract features from frozen pre-trained layers and train a Fully-Connected layer for predictions [21, 22].

### 3.5 VGG16

A Convnet, also known as a Convolution Neural Network, is a type of artificial neural networks. Which is comprised of an input, multiple hidden, and an output layers. VGG16 is a type of CNN. This model authors enhanced the network depth with a small (3×3) convolution filter and assessed the

networks, which achieved considerable enhancement over prior-art setups [23]. Resulting in around 138 trainable parameters because they raised the depth to 16-19 weight layer. Figure 3 illustrates the VGG16 architecture.

### 3.6 Evaluation metrics

Specificity, F1 score, Accuracy, Sensitivity (Recall), and Precision are the performance metrics employed in this paper. True-positive (TP), false-negative (FN), false-positive (FP), and true-negative (TN), metrics are utilized to define these measures [23-25].

$$Accuracy(ACC) = \frac{Tp+Tn}{Total\ popouation} \quad (1)$$

$$recall = \frac{Tp}{Tp+Fn} \quad (2)$$

$$precision = \frac{Tp}{Tp+Fp} \quad (3)$$

$$F1\text{-score} = \frac{2 \times Recall \times Precision}{Recall + Precision} \quad (4)$$

## 4. PROPOSED METHOD

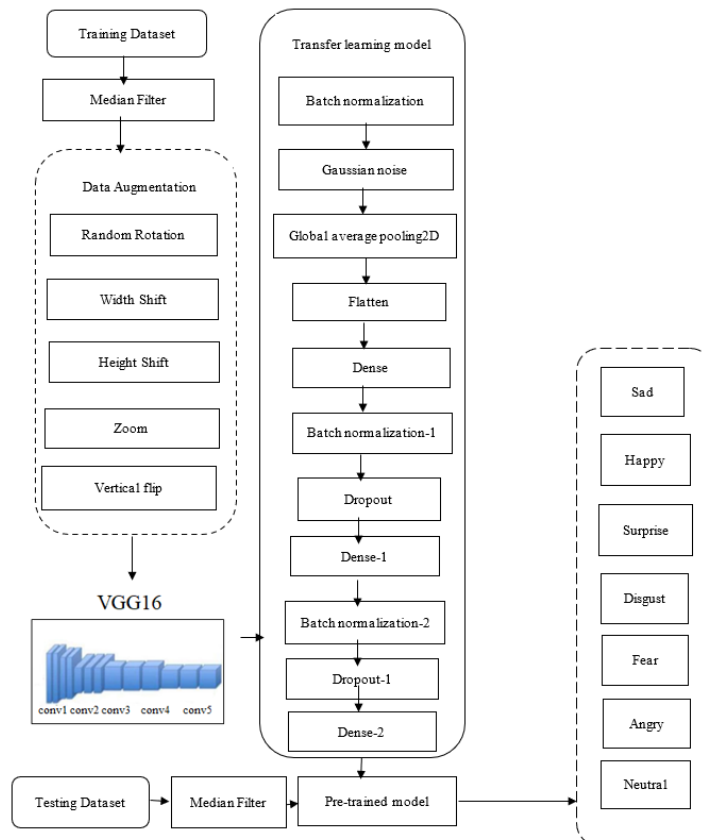


Figure 4. Proposed method

In this work proposed method for FER based on VGG16 with transfer learning. The model consists of preprocessing stage based on the median filter and data augmentation, and classification based on pertained VGG16 with transfer learning added a four layer to the existing vgg16 that already trained on ImageNet dataset that has one thousand classes. We froze the existing version layers and trained new layers on FER2013 dataset. A Keras Image Data Generator generates additional training data from the original data to prevent overfitting. It is done online by looping over in small blocks throughout every iteration of the optimizer. To help produce artificial images, there are some graphic parameter (e.g. shift, Rotation, Add Gaussian noise, flipping). Figure 4 illustrates the proposed method design.

We used Python jupyter notebook to simulate the dataset on our model for analysis. Lenovo intel core i7, Ram 16, hard 512 SSD are used to build the model.

## 5. RESULTS AND DISCUSSION

We used Python jupyter notebook to simulate the dataset on our model for analysis. The dataset is divided into 70%training and 30%testing. The parameters are loss function 0.01, epochs=100, and Adam as optimizer. Because VGG16 layers in this model of transfer learning were trained on Imagenet, the FER2013 dataset testing and training procedure has total parameters of 14,883,399, trainable parameters of 13,146,119, and non-trainable parameters of 1,737,200.

It is shown that the last epoch training loss is 0.48, the accuracy is 96.41%, the loss of testing is 1.3794, and the testing accuracy is 90%. Figure 4 demonstrates the proposed model's efficiency on the FER2013 database. Table 1 shows that the transfer learning-based model outperformed some well-known prior research papers on the same dataset.

This work is implemented on python Jupyter notebook using Keras and tensorflow library. The data is consisted of 7 classes

(sad, angry, surprise, happy, fear, neutral, and disgust). The dataset composed of 35,887 images distributed on the seven classes as illustrated in Figure 1. The data is preprocessed using the median filter and then data augmentation method is applied. The data augmentation is performed on training set to boost the training set size and avoid overfitting. The images are generated using five data augmentation method: height shift (0.1), Random rotation with range (10), zoom (0.1), width shift (0.1), and horizontal flip. Figure 5 illustrates the data distribution on classes after data augmentation process.

Then weights are initialized based on pre-trained VGG16 on Imagenet, FER2013 dataset testing and training procedure has total parameters of 14,883,399, trainable parameters of 13,146,119, and non-trainable parameters of 1,737,200. The summary of VGG16 is demonstrated in Table 1, and the summary of build transfer learning model is demonstrated in Table 2.

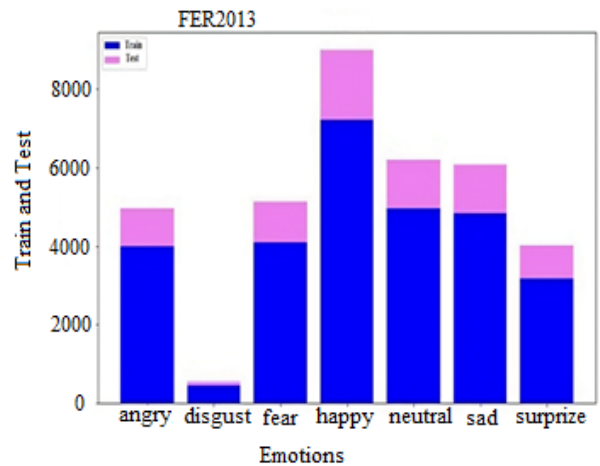


Figure 5. The FER2013 dataset expression distribution after data augmentation

Table 1. Summary of the VGG16

Layer(type)	Output Shape	Param#
Input1(inputlayer)	(None,48,48,3)	0
Block1_conv1(conv2d)	(None,48,48,64)	1792
Block1_conv2(conv2d)	(None,48,48,64)	36928
Block1pool(maxpooling2d)	(None,24,24,64)	0
Block2_conv1(conv2d)	(None,24,24,128)	73856
Block2_conv2(conv2d)	(None,24,24,128)	147584
Block2pool(maxpooling2d)	(None,12,12,256)	0
Block3_conv1(conv2d)	(None,12,12,256)	295168
Block3_conv2(conv2d)	(None,12,12,256)	590080
Block3_conv3(conv2d)	(None,12,12,256)	590080
Block3pool(maxpooling2d)	(None,6,6,512)	0
Block4_conv1(conv2d)	(None,6,6,512)	1180160
Block4_conv2(conv2d)	(None,6,6,512)	2359808
Block4_conv3(conv2d)	(None,6,6,512)	2359808
Block4pool(maxpooling2d)	(None,3,3,512)	0
Block5_conv1(conv2d)	(None,3,3,512)	1180160
Block5_conv2(conv2d)	(None,3,3,512)	2359808
Block5_conv3(conv2d)	(None,3,3,512)	2359808
Block5pool(maxpooling2d)	(None,3,3,512)	0

FER2013 dataset accuracy, precision, recall, loss, F-measure versus epoch graph is presented in Table 2 and Figure 6. It is shown that the last epoch training loss is 0.48, the accuracy is 96.41%, the loss of testing is 1.3794, and the testing accuracy is 90. Table 3 shows that the transfer learning-

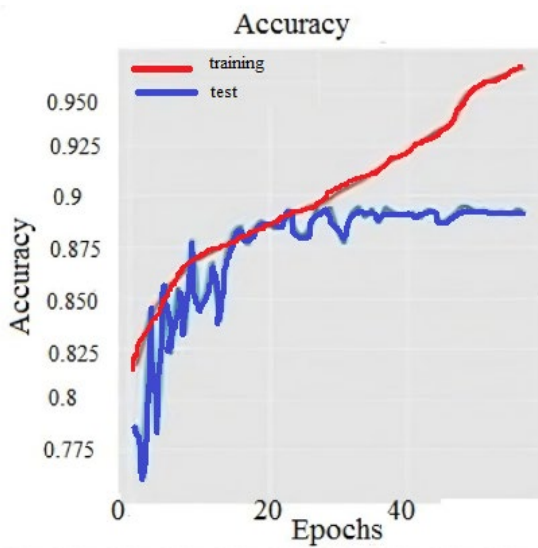
based model outperformed some well-known prior research papers on the same dataset. Singh and Nasoz [9] are achieved an accuracy of 75.2% with CNN. Zahra et al. [11] is used CNN Algorithm based Raspberry Pi for FER prediction and achieved an accuracy of 65.97%. Saleh et al. [7] is used VGG

SpinalNet and achieved an accuracy of 74.45%. Wang [14] is used GAN and achieved 72.8% accuracy. Al-Asbaily and Bozed [13] was used VGG16 model for feature extraction and a classic neural network was used to classification and

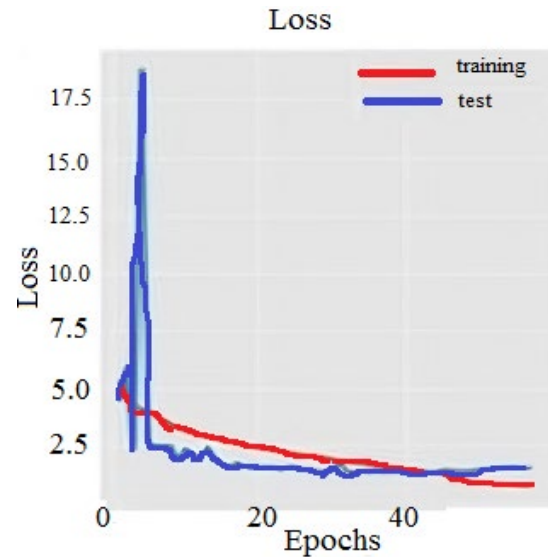
achieved 89.31% accuracy. Whereas our method used data augmentation and transfer learning with VGG16 and achieved an accuracy of 90%.

**Table 2.** Summary of transferred learning model

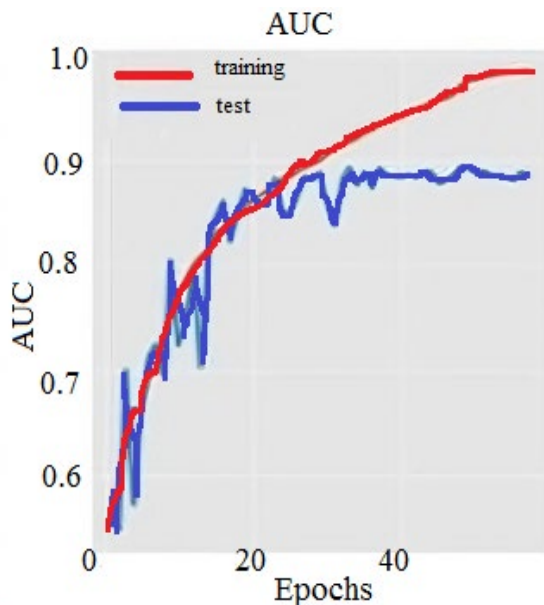
Layer (Type)	Output Shape	Param#
Vgg16 (functional)	(None,1,1,512)	14714688
Batchnormalization (batch normalization)	(None,1,1,512)	2048
Gaussiannoise (gaussiannoise)	(None,1,1,512)	0
Globalaveragepooling2d (globalaveragebooling2d)	(None,512)	0
Flatten (flatten)	(None,512)	0
Dense (dense)	(None,256)	131328
Batchnormalization_1 (batch normalization)	(None,256)	1024
Dropout (dropout)	(None,256)	0
Dense1 (dense)	(None,128)	32896
Batchnormalization_2 (batch normalization)	(None,128)	512
Dropout (dropout)	(None,128)	0
Dense2 (dense)	(None,7)	903
Total parameters: 14883399		
Trainable parameters: 13146119		
Nontrainable parameters: 1737280		



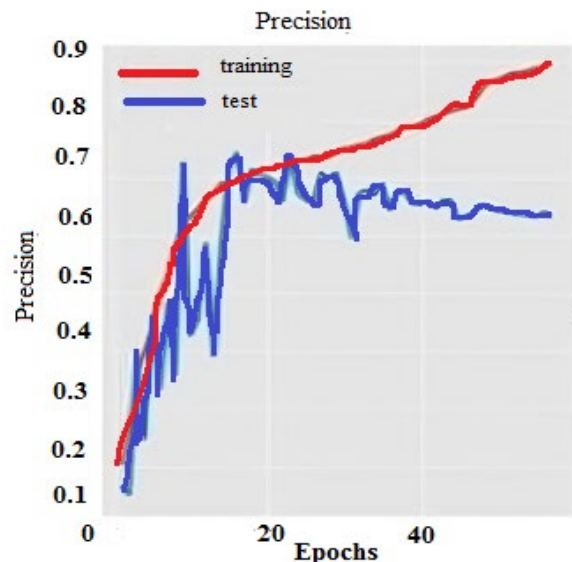
(a) System accuracy



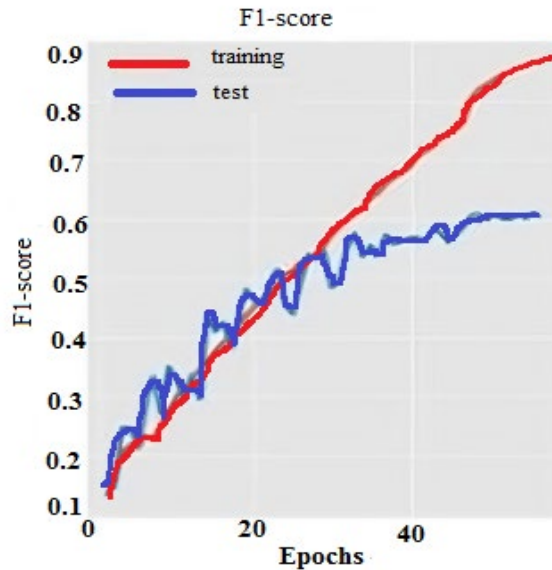
(b) System loss



(c) System AUC



(d) System precision



(e) System F1-score

Figure 6. Proposed method results

Table 3. Test results of HOG-CNN model based on FER2013 dataset

Class	Accuracy (%)
Zahra et al. [11]	65.97
Saleh et al. [7]	74.45%
Singh and Nasoz [9]	75.2%
Wang [14]	72.8%
Our	90%

Table 4. Confusion matrix

Confusion Matrix								
True Label	Angry	125	15	135	212	199	126	126
	Disgust	16	3	15	26	25	7	17
	Fear	106	20	148	213	204	155	132
	Happy	202	25	213	400	301	241	216
	Neutral	145	26	170	301	211	178	172
	Sad	141	18	175	317	264	185	143
	Surprise	97	9	122	206	173	124	100
		Angry	Disgust	Fear	Happy	Neutral	Sad	Surprise
		Predicted Label						

The normalized confusion matrix is demonstrated in Table 4.

dataset, utilizing different neural network such as Alexnet.

## 6. CONCLUSIONS

FER systems are used in many applications such as education, medical diagnosis, etc. Models based transfer learning allow us for knowledge transferring from one model to another to enhance and accelerate the system performance. We removed the top levels of VGG16 and placed our layers above them. We have already trained the VGG16 model on Imagenet that has one thousand classes; new layers were trained on FER2013 dataset after freezing the existing framework layers. The Keras Image Data Generators tool is used to generate from original data more training data to avoid the model overfitting. Our proposed model achieved 90% accuracy, 60% recall, F-measure of 63%, and 66% precision. For future work, the proposed model will test using different

## REFERENCES

- [1] Kołakowska, A., Landowska, A., Szwoch, M., Szwoch, W., Wrobel, M.R. (2014). Emotion recognition and its applications. *Human-Computer Systems Interaction: Backgrounds and Applications*, 3: 51-62. [https://doi.org/10.1007/978-3-319-08491-6\\_5](https://doi.org/10.1007/978-3-319-08491-6_5)
- [2] Chowdary, M.K., Nguyen, T.N., Hemanth, D.J. (2023). Deep learning-based facial emotion recognition for human-computer interaction applications. *Neural Computing and Applications*, 35(32): 23311-23328. <https://doi.org/10.1007/s00521-021-06012-8>
- [3] Saeed, N.A., Al-Ta'i, Z.T.M. (2020). Heart disease prediction system using optimization techniques. In *New Trends in Information and Communications Technology Applications: 4th International Conference, NTICT 2020*,

- Baghdad, Iraq, pp. 167-177. [https://doi.org/10.1007/978-3-030-55340-1\\_12](https://doi.org/10.1007/978-3-030-55340-1_12)
- [4] Mehmood, R.M., Du, R., Lee, H.J. (2017). Optimal feature selection and deep learning ensembles method for emotion recognition from human brain EEG sensors. *Ieee Access*, 5: 14797-14806. <https://doi.org/10.1109/ACCESS.2017.2724555>
- [5] Pitaloka, D.A., Wulandari, A., Basaruddin, T., Liliana, D. Y. (2017). Enhancing CNN with preprocessing stage in automatic emotion recognition. *Procedia computer science*, 116: 523-529. <https://doi.org/10.1016/j.procs.2017.10.038>
- [6] Xu, M., Cheng, W., Zhao, Q., Ma, L., Xu, F. (2015). Facial expression recognition based on transfer learning from deep convolutional networks. In 2015 11th International Conference on Natural Computation (ICNC), Zhangjiajie, pp. 702-708. <https://doi.org/10.1109/ICNC.2015.7378076>
- [7] Saleh, A.A., Saeed, N.A., Saleh O.A. (2024). Kin Cancer Detection And Diagnosis Using Deep Neural Network, *International Journal of Computer Science*, 12(1): 3423-3432. <https://www.ijcsjournal.com/book?bid=517>
- [8] Fathallah, A., Abdi, L., Douik, A. (2017). Facial expression recognition via deep learning. In 2017 IEEE/ACS 14th International Conference on Computer Systems and Applications (AICCSA), Hammamet, Tunisia, pp. 745-750. <https://doi.org/10.1109/AICCSA.2017.124>
- [9] Singh, S., Nasoz, F. (2020). Facial expression recognition with convolutional neural networks. In 2020 10th Annual Computing and Communication Workshop and Conference (CCWC), Las Vegas, NV, USA, pp. 0324-0328. <https://doi.org/10.1109/CCWC47524.2020.9031283>
- [10] Debnath, T., Reza, M.M., Rahman, A., Beheshti, A., Band, S.S., Alinejad-Rokny, H. (2022). Four-layer ConvNet to facial emotion recognition with minimal epochs and the significance of data diversity. *Scientific Reports*, 12(1): 6991. <https://doi.org/10.1038/s41598-022-11173-0>
- [11] Zahara, L., Musa, P., Wibowo, E.P., Karim, I., Musa, S.B. (2020). The facial emotion recognition (FER-2013) dataset for prediction system of micro-expressions face using the convolutional neural network (CNN) algorithm based Raspberry Pi. In 2020 Fifth International Conference on Informatics and Computing (ICIC), Gorontalo, Indonesia, pp. 1-9. <https://doi.org/10.1109/ICIC50835.2020.9288560>
- [12] Khan, A.R. (2022). Facial emotion recognition using conventional machine learning and deep learning methods: Current achievements, analysis and remaining challenges. *Information*, 13(6): 268. <https://doi.org/10.3390/info13060268>
- [13] Al-Asbaily, S.A., Bozed, K.A. (2022). Facial emotion recognition based on deep learning. 2022 IEEE 2nd International Maghreb Meeting of the Conference on Sciences and Techniques of Automatic Control and Computer Engineering (MI-STA), Sabratha, Libya, pp. 534-538. <https://doi.org/10.1109/MI-STA54861.2022.9837630>
- [14] Wang, J. (2021). Improved facial expression recognition method based on GAN. *Scientific Programming*, 2021: 1-8. <https://doi.org/10.1155/2021/2689029>
- [15] FER-2013. <https://www.kaggle.com/datasets/msambare/fer2013>, accessed on Sep. 7, 2022.
- [16] Hwang, H., Haddad, R.A. (1995). Adaptive median filters: New algorithms and results. *IEEE Transactions on Image Processing*, 4(4): 499-502. <https://doi.org/10.1109/83.370679>
- [17] Perreault, S., Hébert, P. (2007). Median filtering in constant time. *IEEE Transactions on Image Processing*, 16(9): 2389-2394. <https://doi.org/10.1109/TIP.2007.902329>
- [18] Shorten, C., Khoshgoftaar, T.M. (2019). A survey on image data augmentation for deep learning. *Journal of Big Data*, 6(1): 1-48. <https://doi.org/10.1186/s40537-019-0197-0>
- [19] Gu, S., Pednekar, M., Slater, R. (2019). Improve image classification using data augmentation and neural networks. *SMU Data Science Review*, 2(2): 1.
- [20] Weiss, K., Khoshgoftaar, T.M., Wang, D. (2016). A survey of transfer learning. *Journal of Big Data*, 3(1): 1-40. <https://doi.org/10.1186/s40537-016-0043-6>
- [21] Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., Li, F.F. (2009). Imagenet: A large-scale hierarchical image database. In 2009 IEEE conference on computer vision and pattern recognition, Miami, FL, USA, pp. 248-255. <https://doi.org/10.1109/CVPR.2009.5206848>
- [22] Zamir, A.R., Sax, A., Shen, W., Guibas, L.J., Malik, J., Savarese, S. (2018). Taskonomy: Disentangling task transfer learning. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3712-3722)..
- [23] Dubey, A.K., Jain, V. (2020). Automatic facial recognition using VGG16 based transfer learning model. *Journal of Information and Optimization Sciences*, 41(7): 1589-1596. <https://doi.org/10.1080/02522667.2020.1809126>
- [24] Saeed, N.A., Al-Ta, Z.T.M. (2019). I," Feature selection using hybrid dragonfly algorithm in a heart disease predication system,". *International Journal of Engineering and Advanced Technology*, 8(6): 2862-2867. 10.35940/ijeat.F8786.088619
- [25] Sanchez-Mendoza, D., Masip, D., Lapedriza, A. (2015). Emotion recognition from mid-level features. *Pattern Recognition Letters*, 67: 66-74. <https://doi.org/10.1016/j.patrec.2015.06.007>

## NOMENCLATURE

CNN	CONVOLUTION NEURAL NETWORK
FER	FACIAL EMOTION RECOGNITION
TN	TRUE NEGATIVE
FP	FALSE POSTIVE
VGG 16	Visual Geometry Group
LBP	Local Binary Patterns
GAN	Generative Adversarial Networks
ORB	Oriented FAST and Rotated BRIEF
CK+	Extended Cohn-Kanade