

# A Novel Efficient Patient Monitoring FER System Using Optimal DL-Features

Mousa Alhajlah\*

College of Applied Computer Science, King Saud University (Almuzahmiyah Campus), Riyadh, 11543, Saudi Arabia

\*Corresponding Author: Mousa Alhajlah. Email: mhajlah@ksu.edu.sa

Received: 20 May 2022; Accepted: 29 September 2022

**Abstract:** Automated Facial Expression Recognition (FER) serves as the backbone of patient monitoring systems, security, and surveillance systems. Real-time FER is a challenging task, due to the uncontrolled nature of the environment and poor quality of input frames. In this paper, a novel FER framework has been proposed for patient monitoring. Preprocessing is performed using contrast-limited adaptive enhancement and the dataset is balanced using augmentation. Two lightweight efficient Convolution Neural Network (CNN) models MobileNetV2 and Neural search Architecture Network Mobile (NasNetMobile) are trained, and feature vectors are extracted. The Whale Optimization Algorithm (WOA) is utilized to remove irrelevant features from these vectors. Finally, the optimized features are serially fused to pass them to the classifier. A comprehensive set of experiments were carried out for the evaluation of real-time image datasets FER-2013, MMA, and CK+ to report performance based on various metrics. Accuracy results show that the proposed model has achieved 82.5% accuracy and performed better in comparison to the state-of-the-art classification techniques in terms of accuracy. We would like to highlight that the proposed technique has achieved better accuracy by using 2.8 times lesser number of features.

**Keywords:** Facial expression recognition; deep learning; transfer learning; feature optimization

## 1 Introduction

Automated FER is being explored by many researchers working in the area of computer vision. It has been a hot research area for the last decade. FER has gained enormous fame because of its utility in various areas like security and automated surveillance, patient monitoring, mood identification biometrics, access control, smart cards, law enforcement, etc. Passwords and pins are difficult to remember and can be stolen or forgotten. Similarly, cards can be corrupted or unreadable and keys can be misplaced but an individual's biological traits cannot be misplaced, forgotten, or stolen [1]. It highlights the utmost significance of this research. Patients admitted to hospitals in the Intensive Care Unit (ICU) require special attention. The automated patient monitoring system can be very



This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

helpful for doctors and staff, especially at night time. Due to increased rush in hospitals and lack of staff it has become very difficult to pay full attention to critical patients [2]. FER systems have a wide range of applications for patients' mental disorders, checking their conditions with the help of emotions is highly desired. Affective expressions can be shown in emotions, and these emotions are expressed in various ways by faces, gestures, postures, voices, and actions. It can also affect physiological parameters.

In [3], the authors have highlighted the problem of classifying uncommon expressions like disgust and fear. Authors claimed that classifying common expressions like happiness and surprise is easier to classify than rare expressions like disgust and fear. They showed two factors that are responsible for this problem, one is intra-class variation and another is the imbalanced dataset. To handle the first issue, the proposed model used semantic and spatial enhancement techniques to enhance expression features of fine-level and high-level semantic features. For the second issue, a novel technique is proposed in which they collected images from different databases instead of generating new images from existing training samples.

The authors' in [3] proposed an efficient lossless compression algorithm for facial images in expression recognition. The proposed method consists of two operations, one is a data preparing operation and another one is a bitstream encoding operation. For the first step facial image is divided and processed into two parts reference and residual images which represent similar and dissimilar parts of the left and right sides of the face. In the second step, these two images are processed into the bit-stream encoding operation for getting the final version of the bit-stream.

A two-channel based deep convolutional neural network model for facial recognition is proposed [4] in which the gradient features and texture features were merged for combining local texture and locally shaped features. Experiments were performed on the Japanese Facial Expression (JAFPE) and FER2013 database which contain seven different sample expressions. The proposed model achieved 76.86% accuracy in the FER2013 database and 88.75% in the JAFPE database which is far above then other state-of-the-art models.

In [5], a technique is proposed in which major and minor detail of facial expressions has been captured using spatial pyramid Zernike features and Law's texture features which were merged. The presented technique used a multilayer perceptron and a feed-forward artificial neural network. JAFPE and KDEF datasets were used for experimental purposes. The proposed model achieved an accuracy of 95.86% and 88.87% on the JAFPE and Karolinska Directed Emotional Faces (KDEF) datasets. A method for human facial expression recognition is proposed in [6] in which authors employed active learning and a Support Vector Machine (SVM) algorithm. Active learning was used to detect Action Units (AU) of facial expression while SVM was used for classifying. Benitez-Quiroz et al. [7] used the geometric as well as texture features and merged them near vital points to improve the feature depiction of each face area. Authors' of [8] used joint region learning to detect AU's. In the proposed strategy, they selected 49 vital points near the nose, mouth, and eyes. After that author' extracted SIFT features to symbolize each area. Kuo et al. [9] proposed a model which is based on CNN for expression recognition. That model consisted of two convolutional layers and two fully connected layers.

A two-branch-based neural network having six layers was used for facial recognition in [10] which extracts global features as well as local information. A dense CNN was proposed in [11] in which all convolutional layers were fully connected with succeeding layers to classify facial expressions. A CNN model for high inter-expression variations has been proposed by authors of [12] in which they described that learning features are not inclined by facial expressions variations. They used expression and identity-sensitive loss to prove it. Li et al. [13] proposed a lightweight network for the FER task named

Facial Expression Recognition Network Auto-FERNet. The authors designed an unpretentious and effective relabeling method which is based on the similarity of Facial Expressions to improvement in the ambiguity problem caused by natural dynamics. The authors' proposed method achieved an accuracy of 73.78% in the FER2013 dataset.

In [14] Agrawal et al. presented two CNN models and evaluated the performance by analyzing the effect of the size of kernel and filters on accuracy using the FER2013 database. In 2020, Li et al. [15] presented the novelty by introducing the attention module in the model. The technique was tested against different datasets and achieved an accuracy of 75.8% on the FER2013 dataset.

It has been observed from the literature that FER systems suffer from computational complexity and low accuracy. There are multiple factors for these problems mainly because of intra-class variance and the existence of imbalanced data among different expressions in the dataset. FER-2013 dataset is a challenging dataset as images taken in this dataset suffer low contrast, illumination variation, background and pose changes and also most of the samples contain partial images. These factors make it more challenging for multiclass classification. The difference in cultural variations in expression also plays its role in causing lower accuracy [16,17] and is a hindrance in the development of a robust system. In this paper, we have developed an efficient system for the classification of expression by applying preprocessing to handle contrast and illumination changes, and data augmentation to handle data misbalancing issues. The system is trained using well-known CNN models. Hybrid features are extracted and optimized to achieve better accuracy in lesser time.

The main contributions of the proposed work are summarized as follows:

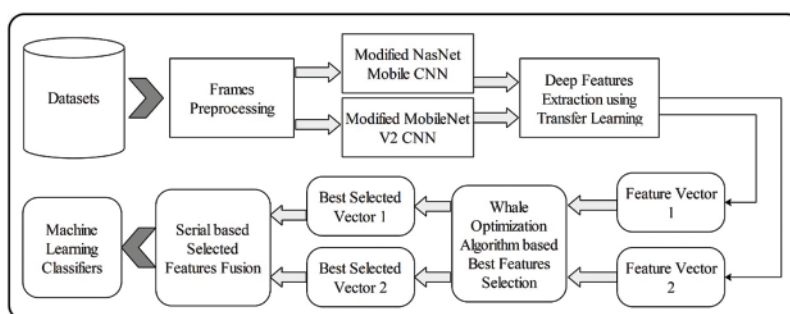
- The proposed framework is reliable across all types of expressions, especially against disgust and fear.
- The deep CNN features are extracted through average pooling layers of NASNet Mobile and MobileNetV2 models using transfer learning.
- Features extracted through both models are then fused using serial-based fusion and are optimized using the whale optimization technique.
- The optimized features show the robustness and perform equally well using a lesser number of features.
- The system captures between-class variations and minimizes the within-class variations.
- The proposed framework is computationally efficient and most likely be used for real-time applications.

This paper is organized in Section 2, and an overview of the proposed framework is presented. The proposed technique steps preprocessing, feature extraction, feature fusion, and optimization are described in subsections. The experimental setup along with related discussion is provided in Section 3. In Section 4, the conclusion and future work is discussed.

## 2 Proposed System Architecture

The system starts by taking the image dataset as input and performing some preprocessing. Firstly, data is normalized and Contrast Limited Adaptive Histogram Equalization (CLAHE) is applied to handle contrast and illumination variations. Two modified CNN models MobileNetV-2 and NasNet Mobile are used for training input data. Features are extracted from both these trained networks separately. We have used these trained features and tested them using 10-fold cross-validation. Individual performance of both these networks is tested. Then features are combined ending up a large feature vector. A large feature vector takes more time in training and testing. These features

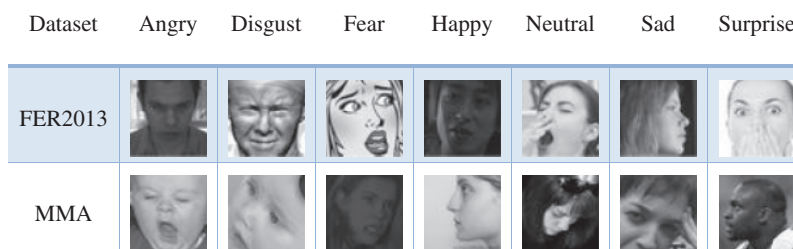
are optimized using well known improved Whale optimizer. Optimal numbers of features are then used for classification. The architecture of the proposed methodology is shown in Fig. 1. There was a significant improvement in the accuracy of the classification results.



**Figure 1:** Proposed system architecture

## 2.1 Preprocessing

The dataset used for experiments suffers from contrast, illumination variation, pose changes, occlusion, and background irrelevant data and can be observed from the data samples shown in Fig. 2. It requires necessary preprocessing before feature extraction. We have applied CLAHE for handling the contrast variations in the image. It performs local enhancement by preserving the local variations.



**Figure 2:** Sample images from both datasets

CLAHE deals with the issue of over-enhancement. It works on the betterment of contrast and also sustains the normal variation of classes which is of the most significance for improved classification.

When the characteristics of contrast are not similar throughout the image, histogram equalization failed to handle such a situation in a better way. CLAHE is an advanced variation of Adaptive Histogram Equalization (AHE) in which the computation of enhancement is altered by commanding a user-stated measure to the height of the local histogram. Therefore, enhancement is lowered in uniform regions of the image which ultimately avoids over-enhancement of noise and decreases the shadowing effects of edges in unlimited AHE.

Fig. 2 shows the sample of images taken from both the datasets and variations and challenges are quite evident in the sample. The dataset contains a sample of images taken from the FER-2013 Dataset [18].

After preprocessing there is an improvement, in contrast, illumination, and background. There is not only visual improvement but also better accuracy has been observed after performing these steps.

Considerable enhancement can be seen in the sample images shown in Fig. 3. We have presented a comparison of the original images with enhanced images.



**Figure 3:** Sample of images after contrast enhancement

To handle the problem of data imbalance, two methods have been applied to balance the dataset. Firstly, the dataset is balanced by adding relevant expression images from other datasets, and secondly, we have applied augmentation. The accuracy of the dataset is improved using augmentation and helps in better training. It also overcomes the problem of overfitting. Furthermore, the dataset gets balanced.

## 2.2 Features Extraction Using CNN Models

### 2.2.1 Mobilenetv2

MobileNetv2 is an efficient neural network architecture, designed specifically considering the needs of mobile and resource-constrained devices. It retains the same accuracy as other models but significantly reduces the resource requirements. It has replaced the full convolutional operator with the Depth wise Separable Convolutions. These are the basis for any efficient neural network [19]. It has been used for solving many problems and has proven to be efficient both in terms of time and accuracy [20]. The architecture contains the initial fully convolution layer with 32 filters, followed by 19 residual bottleneck layers. ReLU6 is used as the non-linearity because it is robust when used with low-precision computation. kernel size  $3 \times 3$  is used as is standard for modern networks and utilizes dropout and batch normalization during training. It uses 3.4 million parameters and is computationally less expensive.

### 2.2.2 NASNet-Mobile

NASNet Mobile is a lightweight model which provides promising results. It uses only 3.2 million parameters. It has been applied for solving many problems and is computationally the least expensive and can be used for real-time systems. Vladimir Nekrasov et.al. used it for real-time segmentation [21]. Because of its fast convergence and efficiency Bharati et al. utilized this network for the estimation of face attributes [22].

### 2.2.3 Improved Whale Optimization Method

The whale optimization algorithm (WOA) is a swarm intelligence-based algorithm that impersonates the social behavior of naturally seen bubble-net-feeding humpback whales [23]. It is a very competitive optimization algorithm that is used by a variety of researchers for feature optimization. The WOA algorithm avoids the local optima and considers the optimization algorithm as a black box then evaluates the global optima using a precise mechanism. In any given optimization problem, WOA generates and improves upon a set of features. This development is motivated by the behavior of humpback whales to surround and swirl movement around their target [24]. WOA has main three steps including shrinking, encircling, and spiral updating. Shrinking and encircling are exploitation phases while the spiral updating step is searching for prey and is known as an exploration phase [25]. We have utilized an improved WOA algorithm [26] for the optimization of features.

## 3 Experimental Setup and Results

The proposed framework has been evaluated using a rigorous set of experiments using a publicly available well-known and most challenging FER dataset taken from Kaggle FER2013. Experiments are evaluated using numerous performance metrics like sensitivity, false negative rate, precision Rate, F-1 score, and accuracy. Experimental results along with their analysis are presented in this section. A comparison with existing state-of-art techniques is also presented.

### 3.1 Experimental Settings, Dataset, and Performance Measures

FER2013 dataset consists of 35886 facial expression images divided into seven different expressions named angry, disgust, fear, happy, sad, surprised, and normal. All the images are of resolution  $48 \times 48$  and are grayscale images. Sample of images collected from the dataset are shown in Fig. 2. Experiments are conducted using Core i7, 10th generation having 16GB RAM and GPU having 8GB GPU. MATLAB 2020a is used for performing experiments and training/testing the FER 2013 dataset. Thought experiments minibatch size of 32, a learning rate of 0.0001, a momentum of 0.6, and maximum epochs of 500 are set. For experiments, we have a 50:50 train-to-test ratio. To reduce biasness, we employed 10-fold cross-validation for all the experiments. For the evaluation of classification performance several different classifiers, namely Fine Tree, Naïve Bayes, Gaussian SVM, SVM (Quadratic), SVM (Cubic), K-Nearest Neighbor (KNN) with its variants i.e., Fine-KNN, Cosine-KNN, Weighted-KNN, and Ensemble-Bagged Tree are run and results are presented.

Performance metrics like False Negative (FN), False Positive (FP), True Negative (TN), and True Positive (TP) can be derived from the confusion matrix and the formula to calculate these metrics is given in the equation given below.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

$$Specificity = \frac{TN}{TN + FP} \quad (3)$$

$$Recall = \frac{TP}{TP + FN} \quad (4)$$

$$F1 - score = 2 \times \frac{Recall \times Precision}{Recall + Precision} \quad (5)$$



Initially, the image dataset consists of 35887 images with distribution among seven different expressions as shown in Table 1. After modification and augmentation, the dataset increased to 57799 with a distribution among seven expressions shown against the augmented dataset column in Table 1.

**Table 1:** FER-2013 dataset distribution among expressions

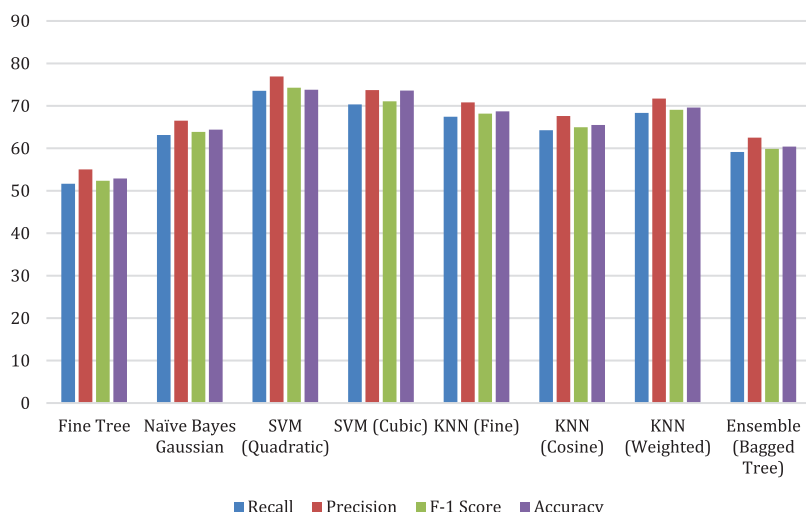
Expression	Original dataset (No. of images)	Augmented dataset (No. of images)
Angry	4953	8963
Disgust	547	5089
Fear	5121	8704
Happy	8989	8989
Neutral	6198	8372
Sad	6077	8955
Surprise	4002	8727
Total	35887	57799

### 3.2 Experimental Results

The performance of the classification technique using different performance measures is shown in Table 2 and Fig. 4. Dataset has been trained on 50 percent and tested using 50 percent. Testing results are presented. A comprehensive set of experiments is conducted using different training to test ratios. Dataset used is challenging in terms of occlusion, illumination, contrast, and partial images. Through preprocessing technique, these issues have been addressed. However, it has been observed that performance deteriorated mainly because of an imbalance in the dataset, especially in cases of disgust, fear, and surprise. Due to the use of augmentation and data updating, there is considerable improvement in accuracy. The deep learning model MobileNetV-2 is a pre-trained CNN model that is fine-tuned by the use of replacement of three layers namely a fully connected layer, updated SoftMax, and a revised classification layer. 1280 features are extracted and used for classification.

**Table 2:** Classification performance using MobileNetv2

Classifier	Recall (%)	Precision (%)	F-1 score (%)	Accuracy (%)
Fine tree	51.65	55.02	52.37	52.9
Naïve Bayes Gaussian	63.15	66.52	63.87	64.4
SVM (Quadratic)	73.55	76.92	74.27	73.8
SVM (Cubic)	70.35	73.72	71.07	73.6
KNN (Fine)	67.45	70.82	68.17	68.7
KNN (Cosine)	64.25	67.62	64.97	65.5
KNN (Weighted)	68.35	71.72	69.07	69.6
Ensemble (Bagged tree)	59.15	62.52	59.87	60.4



**Figure 4:** Illustration of MobileNetV2 results with different classifiers

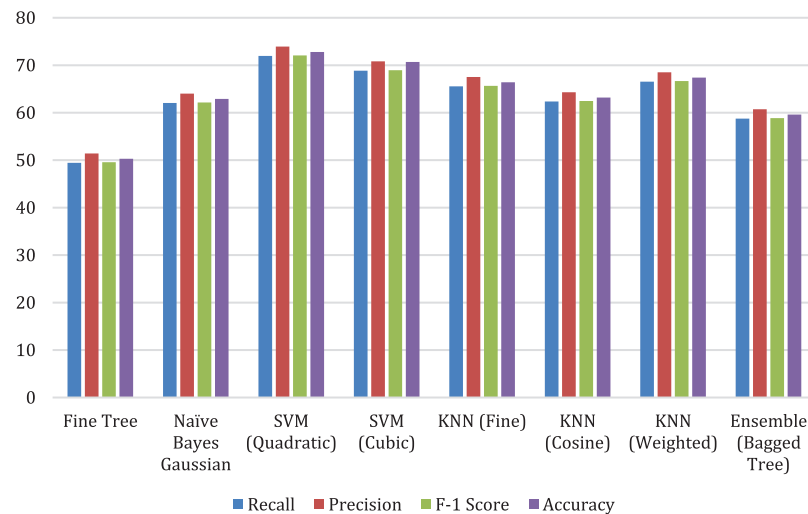
A feature set of 1280 is used for classification. In Table 2 classification performance using MobileNetV2 is shown. It is evident from the results that consistent performance has been observed. It is because of the data balance and preprocessing. Quadratic SVM and Cubic SVM have shown better performance in terms of accuracy, precision, recall, and F-1 score.

To reduce the complexity of the algorithm resultant features extracted using MobileNetV2 are optimized by using improved WOA. 458 optimal features are used for classification and passed to the classifiers and the results are presented in Table 3 and Fig. 5. It is evident that almost similar accuracy is achieved using a smaller number of features. Hence reduced training and testing time resulting in a faster system.

**Table 3:** Experimental results using optimized features on trained MobileNetV2 models

Classifier	Recall (%)	Precision (%)	F-1 score (%)	Accuracy (%)
Fine tree	49.45	51.42	49.56	50.3
Naïve Bayes Gaussian	62.05	64.02	62.16	62.9
SVM (Quadratic)	71.95	73.92	72.06	72.8
SVM (Cubic)	68.85	70.82	68.96	70.7
KNN (Fine)	65.55	67.52	65.66	66.4
KNN (Cosine)	62.35	64.32	62.46	63.2
KNN (Weighted)	66.55	68.52	66.66	67.4
Ensemble (Bagged tree)	58.75	60.72	58.86	59.6





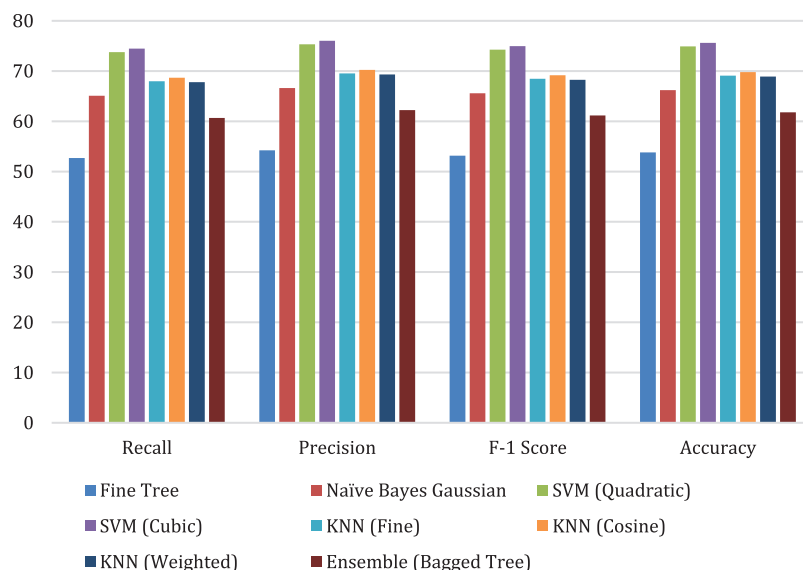
**Figure 5:** Illustration of MobileNetV2 optimized results with different classifiers

The preprocessed dataset is passed to another lightweight CNN model NasNetMobile. The transfer learning model is trained and 1280 features are extracted. Results are compiled using these features and are shown in Table 4 and Fig. 6. The performance of the NasNetMobile CNN model has been evaluated using a Fine tree, Naïve Bayes Gaussian, SVM (Quadratic), SVM (Cubic), KNN (Fine), KNN (cosine), KNN (weighted), and Ensemble (Bagged Tree) classifiers. The accuracy rate of 53.8%, 66.2%, 74.9%, 69.1%, 69.8%, 68.9%, and 61.8% have been achieved using Fine tree, Naïve Bayes Gaussian, SVM (Quadratic), SVM (Cubic), KNN (Fine), KNN (cosine), KNN (weighted), and Ensemble (Bagged Tree) classifiers respectively.

**Table 4:** Experimental results using the NasNetMobile model

Classifier	Recall (%)	Precision (%)	F-1 score (%)	Accuracy (%)
Fine Tree	52.68	54.23	53.17	53.8
Naïve Bayes Gaussian	65.08	66.63	65.57	66.2
SVM (Quadratic)	73.78	75.33	74.27	74.9
SVM (Cubic)	74.48	76.03	74.97	75.6
KNN (Fine)	67.98	69.53	68.47	69.1
KNN (Cosine)	68.68	70.23	69.17	69.8
KNN (Weighted)	67.78	69.33	68.27	68.9
Ensemble (Bagged tree)	60.68	62.23	61.17	61.8

The best result Cubic SVM has performed better and 75.6% has been achieved as compared to other classifiers.



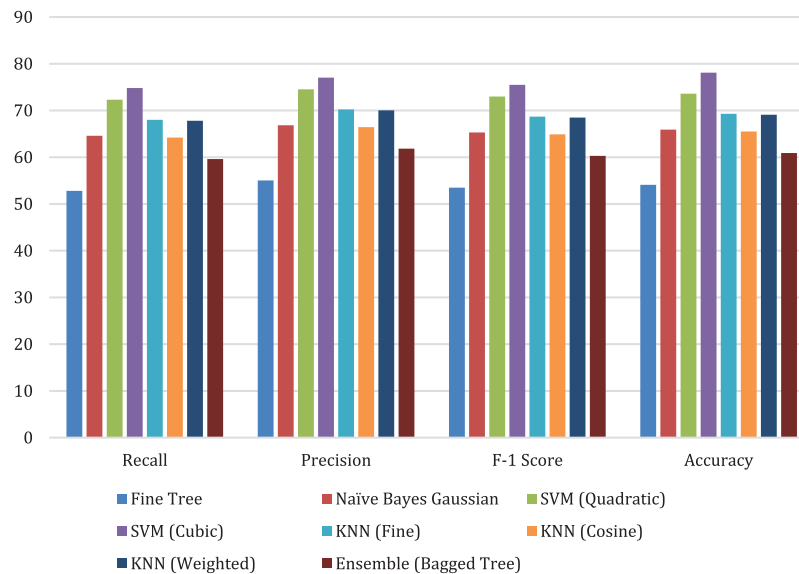
**Figure 6:** Illustration of NasNetMobile results with different classifiers

Features extracted using NasNetmobile are also optimized using WOA and optimal features are passed to the classifiers for testing. Results are presented in [Table 5](#) and [Fig. 7](#).

**Table 5:** Experimental results using deep learning (NasNetMobile) model with optimized features

Classifier	Recall (%)	Precision (%)	F-1 score (%)	Accuracy (%)
Fine tree	52.8	55.04	53.49	54.1
Naïve Bayes Gaussian	64.6	66.84	65.29	65.9
SVM (Quadratic)	72.3	74.54	72.99	73.6
SVM (Cubic)	74.8	77.04	75.49	78.1
KNN (Fine)	68	70.24	68.69	69.3
KNN (Cosine)	64.2	66.44	64.89	65.5
KNN (Weighted)	67.8	70.04	68.49	69.1
Ensemble (Bagged tree)	59.6	61.84	60.29	60.9

Dataset has diversity and is challenging. One model extracts feature using a different process. There is a chance of irrelevant information that may lead to misclassification. For this, we have optimized features extracted using both models. Optimized Features are fused using serial fusion. It has been observed that there is a lot of improvement in the performance of the classification. Furthermore, results are consistent using all the performance measures. Importantly, Cubic SVM has performed best and 82.5% accuracy is achieved. Results of the proposed technique are shown in [Table 6](#) and [Fig. 8](#).



**Figure 7:** Illustration of NasNetMobile optimized results with different classifiers

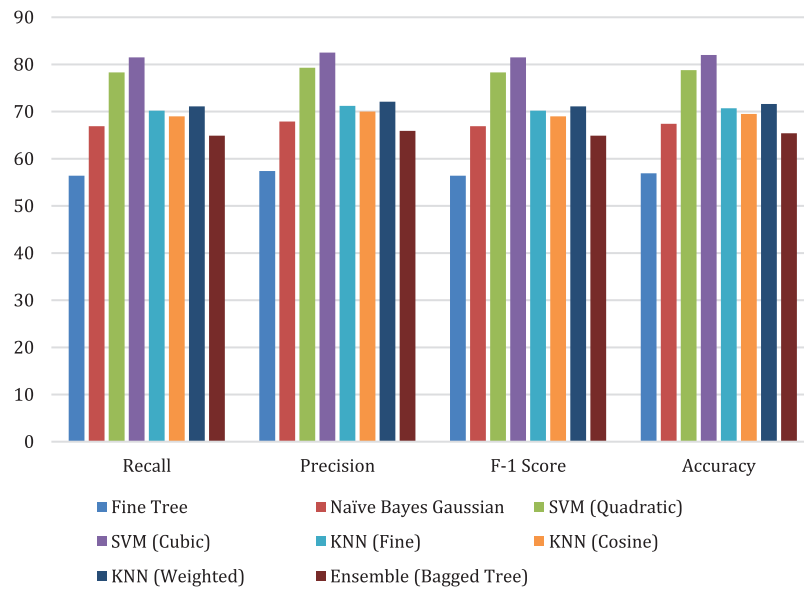
**Table 6:** Proposed technique experimental results using optimal fused features

Classifier	Recall (%)	Precision (%)	F-1 score (%)	Accuracy (%)
Fine Tree	56.4	57.4	56.4	56.9
Naïve Bayes Gaussian	66.9	67.9	66.9	67.4
SVM (Quadratic)	78.3	79.3	78.3	78.8
SVM (Cubic)	81.5	82.5	81.5	82.0
KNN (Fine)	70.2	71.2	70.2	70.7
KNN (Cosine)	69	70	69	69.5
KNN (Weighted)	71.1	72.1	71.1	71.6
Ensemble (Bagged tree)	64.9	65.9	64.9	65.4

Fig. 9, illustrates the confusion matrix of the different expressions. Here, we get 53.4%, 81.2%, 50.7%, 76%, 52.8%, 46.3% and 81.5% accuracy rate for angry, disgust, fear, happy, sad and surprise expression respectively. We noticed that the angry expression is mostly confused with fear and sad expression. The sad expression is the most challenging facial expression classified with a lower accuracy rate of 46.3% as compared to other expressions. This expression is misclassified as fear and neutral.

### 3.3 Comparison with the State-of-the-Art Techniques

A comprehensive set of experiments are performed using the different train-to-test ratios. Consistent results are observed. Results are shown in Table 7 are the best classifier using a lesser number of features. It is evident from the Table 7 that the proposed technique has outperformed all the existing techniques.



**Figure 8:** Illustration of proposed method results using optimized features with different classifiers

True Class	angry	53.4%	4.0%	12.4%	5.3%	10.1%	12.9%	2.4%
	disgust	5.3%	81.2%	2.3%	1.6%	3.0%	3.7%	1.3%
	fear	12.3%	2.3%	50.7%	3.7%	8.6%	15.8%	8.7%
	happy	5.8%	1.8%	4.0%	76.0%	7.2%	5.1%	2.5%
	neutral	7.9%	2.8%	7.3%	6.2%	52.8%	14.2%	1.8%
	sad	12.6%	3.1%	15.0%	4.4%	15.6%	46.3%	1.9%
	surprise	2.7%	4.9%	8.5%	2.8%	2.7%	1.9%	81.5%

**Figure 9:** Confusion matrix showing the results of the different facial expression classes

**Table 7:** Comparison of state-of-the-art techniques

Reference/Technique	Accuracy (%)
Li et al. [15]	75.8
Lee et al. [25]	65
Wang et al. [27]	73.75
Jiang et al. [28]	70.8
Hussein et al. [29]	81

(Continued)

**Table 7:** Continued

Reference/Technique	Accuracy (%)
Kaviya et al. [30]	65
<b>Proposed</b>	<b>82.5</b>

#### 4 Conclusion

Facial expression recognition and classification is a widely used research domain and has its utility in a variety of disciplines. In this paper novel, framework FER problem has been investigated by focusing on patient monitoring. The proposed model has been developed and tested using the most challenging dataset FER2013. Preprocessing is done by handling contrast enhancement and illumination variations. The dataset is balanced using augmentation and addition of new images especially disgust, fear, and sad expressions. Two efficient lightweight deep learning models MobileNetV2 and NasNet-Mobile are employed for significant feature extraction. Features extracted from these two models are optimized using improved WOA and are fused for classification. Improved results are achieved using 10-fold cross-validation. In future work, we plan to investigate the proposed model for Ulcer classification [31], medical records classification [32,33], and face analysis [34,35].

**Acknowledgement:** Researchers Supporting Project Number (RSP2022R458), King Saud University, Riyadh, Saudi Arabia.

**Funding Statement:** Researchers Supporting Project Number (RSP2022R458), King Saud University, Riyadh, Saudi Arabia.

**Conflicts of Interest:** The authors declare that they have no conflicts of interest to report regarding the present study.

#### References

- [1] Y. Ma, X. Wang and L. Wei, "Multi-level spatial and semantic enhancement network for expression recognition," *Applied Intelligence*, vol. 51, pp. 8565–8578, 2021.
- [2] N. Samadiani, G. Huang, B. Cai, W. Luo and C. -H. Chi, "A review on automatic facial expression recognition systems assisted by multimodal sensor data," *Sensors*, vol. 19, pp. 1863, 2019.
- [3] C. Fan, F. Li, Y. Jiao and X. Liu, "A novel lossless compression framework for facial depth images in expression recognition," *Multimedia Tools and Applications*, vol. 80, pp. 24173–24183, 2021.
- [4] H. Gao, L. Sun and J. -X. Wang, "Phygeonet: Physics-informed geometry-adaptive convolutional neural networks for solving parameterized steady-state PDEs on irregular domain," *Journal of Computational Physics*, vol. 428, pp. 110079, 2021.
- [5] V. G. Mahesh, C. Chen, V. Rajangam, A. N. J. Raj and P. T. Krishnan, "Shape and texture aware facial expression recognition using spatial pyramid zernike moments and law's textures feature set," *IEEE Access*, vol. 9, pp. 52509–52522, 2021.
- [6] L. Yao, Y. Wan, H. Ni and B. Xu, "Action unit classification for facial expression recognition using active learning and SVM," *Multimedia Tools and Applications*, vol. 80, pp. 24287–24301, 2021.
- [7] C. F. Benitez-Quiroz, R. Srinivasan and A. M. Martinez, "Emotionet: An accurate, real-time algorithm for the automatic annotation of a million facial expressions in the wild," in *Proc. Computer Vision and Pattern Recognition*, Las Vegas, NV, USA, pp. 5562–5570, 2016.
- [8] K. Zhao, W. -S. Chu, F. De la Torre, J. F. Cohn and H. Zhang, "Joint patch and multi-label learning for facial action unit detection," in *Proc. CVPRW*, Boston, MA, USA, pp. 2207–2216, 2015.

- [9] C. -M. Kuo, S. -H. Lai and M. Sarkis, "A compact deep learning model for robust facial expression recognition," in *Proc. Conf. on Computer Vision and Pattern Recognition Workshops*, Salt Lake City, UT, USA, pp. 2121–2129, 2018.
- [10] S. Xie and H. Hu, "Facial expression recognition using hierarchical features with deep comprehensive multipatches aggregation convolutional neural networks," *IEEE Transactions on Multimedia*, vol. 21, pp. 211–220, 2018.
- [11] H. Ma and T. Celik, "FER-Net: Facial expression recognition using densely connected convolutional network," *Electronics Letters*, vol. 55, pp. 184–186, 2019.
- [12] Z. Meng, P. Liu, J. Cai, S. Han and Y. Tong, "Identity-aware convolutional neural network for facial expression recognition," in *Proc. FG*, Washington, DC, USA, pp. 558–565, 2017.
- [13] S. Li, W. Li, S. Wen, K. Shi and Y. Yang, "Auto-FERNet: A facial expression recognition network with architecture search," *IEEE Transactions on Network Science and Engineering*, vol. 8, pp. 2213–2222, 2021.
- [14] A. Agrawal and N. Mittal, "Using CNN for facial expression recognition: A study of the effects of kernel size and number of filters on accuracy," *The Visual Computer*, vol. 36, pp. 405–412, 2020.
- [15] J. Li, K. Jin, D. Zhou, N. Kubota and Z. Ju, "Attention mechanism-based CNN for facial expression recognition," *Neurocomputing*, vol. 411, pp. 340–350, 2020.
- [16] M. N. Noor, M. Nazir, S. Rehman and J. Tariq, "Sketch-recognition using pre-trained model," in *Proc. National Conf. on Engineering & Computing Technologies (NCECT)*, Islamabad, Pakistan, 2021.
- [17] M. N. Noor and F. Haneef, "A review on big data and social network analytics techniques," *Researchpedia Journal of Computing*, vol. 1, pp. 1–11, 2020.
- [18] I. J. Goodfellow, D. Erhan, P. L. Carrier, A. Courville and M. Mirza, "Challenges in representation learning: A report on three machine learning contests," in *Proc. ICONIP*, Daegu, South Korea, pp. 117–124, 2013.
- [19] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov and L. -C. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," in *Proc. Computer Vision and Pattern Recognition*, Salt Lake City, UT, USA, pp. 4510–4520, 2018.
- [20] F. Saxen, P. Werner, S. Handrich, E. Othman and L. Dinges, "Face attribute detection with mobilenetv2 and nasnet-mobile," in *11th Int. Symp. on Image and Signal Processing and Analysis (ISPA)*, 2019, Dubrovnik, Croatia, pp. 176–180, 2019.
- [21] V. Nekrasov, C. Shen and I. Reid, "Light-weight refinenet for real-time semantic segmentation," *IEEE Robotics and Automation Letters*, vol. 6, pp. 263–270, 2018.
- [22] S. Bharati, P. Podder, M. Mondal and N. Gandhi, "Optimized nasnet for diagnosis of COVID-19 from lung CT images," in *Proc. ISDA*, Springer, Cham, pp. 647–656, 2020.
- [23] S. Mirjalili and A. Lewis, "The whale optimization algorithm," *Advances in Engineering Software*, vol. 95, pp. 51–67, 2016.
- [24] T. Goel, R. Murugan, S. Mirjalili and D. K. Chakrabartty, "Automatic screening of COVID-19 using an optimized generative adversarial network," *Cognitive Computation*, vol. 20, pp. 1–16, 2021.
- [25] K. -C. Lee and P. -T. Lu, "Application of whale optimization algorithm to inverse scattering of an imperfect conductor with corners," *International Journal of Antennas and Propagation*, vol. 5, pp. 1–9, 2020.
- [26] N. Hussain, M. A. Khan, S. Kadry, U. Tariq and R. Mostafa, "Intelligent deep learning and improved whale optimization algorithm-based framework for object recognition," *Human-Centric Computing and Information Sciences*, vol. 11, pp. 34, 2021.
- [27] Y. Wang, Y. Li, Y. Song and X. Rong, "The application of a hybrid transfer algorithm based on a convolutional neural network model and an improved convolution restricted boltzmann machine model in facial expression recognition," *IEEE Access*, vol. 7, pp. 184599–184610, 2019.
- [28] P. Jiang, G. Liu, Q. Wang and J. Wu, "Accurate and reliable facial expression recognition using advanced softmax loss with fixed weights," *IEEE Signal Processing Letters*, vol. 27, pp. 725–729, 2020.
- [29] E. S. Hussein, U. Qidwai and M. Al-Meer, "Emotional stability detection using convolutional neural networks," in *2020 IEEE Int. Conf. on Informatics, IoT, and Enabling Technologies (ICIoT)*, NY, USA, pp. 136–140, 2020.

- [30] P. Kaviya and T. Arumugaparakash, "Group facial emotion analysis system using convolutional neural network," in *4th Int. Conf. on Trends in Electronics and Informatics (ICOEI)*, NY, USA, pp. 643–647, 2020.
- [31] Y. Masmoudi, M. Ramzan, S. A. Khan and M. Habib, "Optimal feature extraction and ulcer classification from WCE image data using deep learning," *Soft Computing*, vol. 26, pp. 7979–7992, 2022.
- [32] M. Ramzan, M. Habib and S. A. Khan, "Secure and efficient privacy protection system for medical records," *Sustainable Computing: Informatics and Systems*, vol. 35, pp. 100717, 2022.
- [33] T. Sadad, A. Hussain, A. Munir, M. Habib and S. Ali Khan, "Identification of breast malignancy by marker-controlled watershed transformation and hybrid feature set for healthcare," *Applied Sciences*, vol. 10, pp. 1900, 2020.
- [34] S. A. Khan, A. Hussain and M. Usman, "Reliable facial expression recognition for multi-scale images using weber local binary image based cosine transform features," *Multimedia Tools and Applications*, vol. 77, pp. 1133–1165, 2018.
- [35] S. A. Khan, M. Ishtiaq, M. Nazir and M. Shaheen, "Face recognition under varying expressions and illumination using particle swarm optimization," *Journal of Computational Science*, vol. 28, pp. 94–100, 2018.