SRTTU

Journal of Electrical and Computer Engineering Innovations
(JECEI)
Journal homepage: http://www.jecei.sru.ac.ir

JECEI

Research paper

Facial Expression Recognition through Suboptimal Filter Design Using a Metaheuristic Kidney Algorithm

***E. Ghasemi* [1], *S. M. Razavi* [1], *S. Mohamadzadeh* [1,*], *M. Taghipour-Gorjikolaie* [2]**

[1] *Department of Electrical and Computer Engineering, University of Birjand, Birjand, Iran.*

[2] *School of Engineering, London South Bank University, London, UK.*

Article Info

Article History:
Received 06 January 2024
Reviewed 18 March 2024
Revised 10 April 2024
Accepted 30 April 2024

Keywords:
Suboptimal filter
Kidney algorithm
Nearest neighbor classification
Neural network
Filter bank
Facial expression recognition

*Corresponding Author's Email Address:
s.mohamadzadeh@birjand.ac.ir

Abstract

Background and Objectives: The recognition of facial expressions using metaheuristic algorithms is a research topic in the field of computer vision. This article presents an approach to identify facial expressions using an optimized filter developed by metaheuristic algorithms.

Methods: The entire process of feature extraction hinges on using a filter suboptimally configured by metaheuristic algorithms. Essentially, the purpose of utilizing this metaheuristic algorithm is to determine the suboptimal weights for feature extraction filters. Once the suboptimal weights for the filter have been determined by the metaheuristic algorithm, suboptimal filter sizes have also been determined. As an initial step, the k-nearest neighbor classifier is employed due to its simplicity and high accuracy. Following the initial stage, a final model is presented, which integrates results from both filterbank and Multilayer Perceptron neural networks.

Results: An analysis of the existing instances in the FER2013 database has been conducted using the method proposed in this article. This model achieved a recognition rate of 78%, which is superior to other algorithms and methods while requiring less training time than other algorithms and methods.In addition, the JAFFE database, a Japanese women's database, was utilized for validation. On this dataset, the proposed approach achieved a 94.88% accuracy rate, outperforming other competitors.

Conclusion: The purpose of this article is to propose a method for improving facial expression recognition by using an optimized filter, which is implemented through a metaheuristic algorithm based on the KA. In this approach, optimized filters were extracted using the metaheuristic algorithms kidney, k-nearest neighbor, and multilayer perceptron. Additionally, by employing this approach, the suboptimal size and number of filters for facial state recognition were determined in order to achieve the highest level of accuracy in the extraction process.

Introduction

This principle emphasizes that every state (words, silence, activity) in human interaction carries a meaningful value, and that communication plays an important role in establishing civil society, according to Watzlawick, a psychologist [1].Generally, communication is composed of 7% verbal communication, 38% para-verbal communication (such as tone analysis), and 55% nonverbal communication (such as facial expressions, gestures, eye contact, etc.) Therefore, nonverbal communication is significant to many aspects of our daily lives and to the effectiveness of human interaction in general [2].

A positive emotion, such as trust and agreement, is more likely to develop than a negative emotion, such as fear, mistrust, etc [3]. Consequently, facial expressions are the main means by which we communicate our emotions to the external world and interact with others, and we assess how others engage with us through them. Computers are capable of automatically recognizing facial expressions and emotions [4]. A rich and robust human-computer interaction is contingent upon automatic emotion recognition systems, namely machines and devices that can recognize mood. Recent years have seen a growing interest in integrating emotion recognition into modern human-computer interfaces. Research in this field has also found applications in a variety of fields, including animation, medicine, and security [5]. Robot-assisted therapy for children with Autism Spectrum Disorders (ASD) [6] or for elderly patients with dementia has gained attention from researchers in the field of healthcare [7], [8]. In recent years, humanoid and pet robots have become increasingly popular for therapeutic purposes, including companionship, sports coaching, and daily life assistance [9].

As research progresses in this area, machine learning techniques are being used in order to enable such robots to increasingly engage and read patient states. As a result, while efficient facial recognition systems have been developed, systems that can automatically detect emotions have not been developed [10], [11]. In order to recognize facial expressions of individuals, multiple methods have been proposed. A number of advances have been made in the field of facial recognition based on machine learning approaches as facial features are extremely complex. The extraction of high-dimensional and suboptimal features, as well as the classification of the data, becomes feasible through the utilization of heuristic algorithms and machine learning. An innovative machine learning model has been proposed to accurately and rapidly detect facial expressions in this study.

Related Work

Since their introduction in the late 1990s, Convolutional Neural Networks (CNNs) have demonstrated considerable potential in image processing [12]. An ordinary CNN typically consists of a convolutional layer, a hidden layer, and a fully connected layer. This ensures efficiency in processing static images in computer vision. In the past, the utilization of CNN was limited due to the scarcity of training data and computational power. After 2010, the exponential growth in computational power and the availability of larger datasets have transformed CNNs into highly suitable tools for feature extraction and image classification

Various techniques have been proposed for enhancing performance. For instance, the sigmoid activation function has been replaced by the rectified linear unit (ReLU) activation function to address issues such as gradient explosion and facilitate training acceleration. Various methods, such as average pooling and maximum pooling, are employed for sampling inputs and aiding in generalization [14]. Regularization and data augmentation are used to prevent overfitting. Batch normalization is employed to mitigate the issue of vanishing gradients [15].

The advancements and extensive research in Convolutional Neural Networks (CNNs) have rendered them highly desirable tools for addressing tasks in image processing, pattern recognition, and feature extraction. When a large dataset of facial expressions, known as the FER2013 dataset, was introduced at ICML in 2013, various types of CNNs achieved significant classification accuracies ranging from 65% to 72.7% [16]-[24]. For instance, Liu et al. trained three separate CNN models and grouped them to enhance their performance [16]. Minaei et al. employed an end-to-end deep learning framework utilizing an attention-based correlation network [17]. Tang et al. replaced the SoftMax layer with a Support Vector Machine in a deep neural network [22]. Shi et al. proposed a new modification representation module (ARM) to replace the hidden layer [23]. Pramerdorfer et al. compared the performance of three different architectures, VGG, Inception, and Resnet, through a comprehensive analysis [24].

Liu et al. ensembled three CNNs and demonstrated that multiple groups enhance the performance of different models [16]. Pramerdorfer et al. ensemble eight CNNs together and achieved improved performance [24].

Proposed Methodology

According to the methodology employed in this paper, facial expressions are classified into seven distinct groups using the FER2013 database. Following the block diagram depicted in Fig. 1, the proposed method consists of carrying out the following stages of facial expression recognition step-by-step.

A. Dataset Preprocessing

Data preprocessing is one of the fundamental stages in facial expression recognition, and its proper implementation significantly aids in improving the recognition rate. As the starting point in this article, the FER2013 database is invoked. It consists of 28,709 training and 3,589 testing and evaluation data with gray-scale backgrounds, which have been normalized by reading them. A series of seven emotional states have been used to categorize the images in the database.

B. Face Detection and Extraction

As a first step, all images have been analyzed using the MTCNN algorithm, which has been applied to detect and crop faces in a 32x32 pixel measurement area [25].

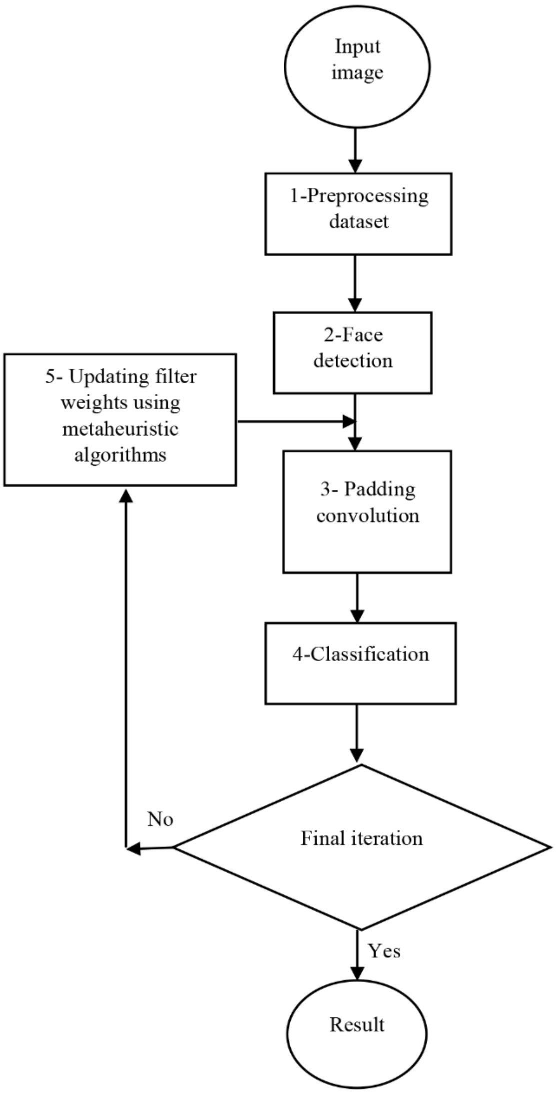


Fig. 1: Proposed methodology structure.

C. Application of Convolution Filters

A kernel can be processed over an image using two convolution methods: Valid Padding Convolution and Same Padding Convolution. We use Valid Padding Convolution in this article to minimize the amount of computational time required. Fig. 2 show that Scaling with Pooling.

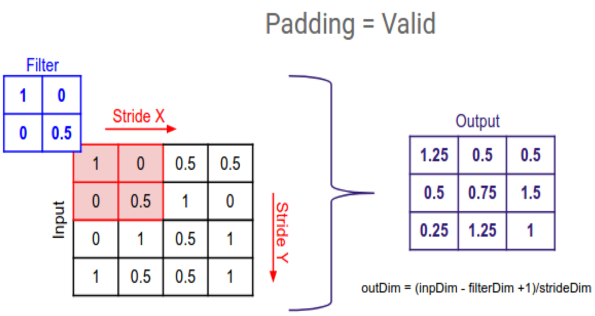


Fig. 2: Scaling with pooling.

D. Classification

In this article, it is proposed to start with the k-nearest neighbor (k-NN) classification with a k-value of 3. After the results from single filters and filter banks have been obtained, the neural network model, known as MLP (Multilayer Perceptron), which is a powerful classifier, is utilized. Furthermore, the Kidney Algorithm (KA), which belongs to the class of newer metaheuristic algorithms, is employed. The objective is to use the Kidney Algorithm (KA) to extract a single suboptimal filter or filter bank, thereby enhancing the recognition rate. As a result, by applying the filters, new features are extracted to determine how the output changes. Initially, KNN was employed, and when the best result for the nearest neighbor classifier and the overall algorithm for single-filter were obtained, a metaheuristic algorithm was utilized to extract a filter bank using an MLP neural network. In this article, employing the metaheuristic algorithm, four suboptimal single filters with sizes of 3×3, 5×5, 7×7, and 9×9 (pixels) have been extracted, and the recognition results on the data have been compared. The same method was then used to extract a filter bank, and the final stage involved implementing an MLP neural network to enhance recognition.

E. Updating Filter Weights Using Metaheuristic Algorithms

The filter weights can be related through an integral equation. This section illustrates how filter weights function as pattern recognition adaptors. Convolutional filters are commonly described based on the pattern-matching operator. Take note that, in this context, f(x) represents the pixel values or the values of lower layers of the network. The higher the f(x) value near a point, the higher the convolution value because discrete convolution is equivalent to a point-wise multiplication between the filter weights and the underlying filter values. In geometric terms, the dot product operator represents a scaled measure of vector similarity. The output of a focused convolution receives two vectors: one representing weights, the other input values. The respective values are then multiplied together and added together. This is comparable to the operation performed by the dot product operator. The magnitude of one or both vectors in the dot product can be increased by increasing their magnitudes. However, when the magnitudes are constant, the maximum value of the dot product is obtained when the vectors are in the same direction [26].

The self-adaptive filter, based on the convolution operator, aids in extracting features such as edges, blurring, and sharpening. Numerous filters are at our disposal for feature extraction, including edge detection, blurring, and sharpening. Furthermore, it has been demonstrated that the matrix numbers of the filters undergo changes based on their respective operations. The filter extracts the main feature of the original image. Thus, our objective is to determine the suboptimal weights for the filter, ensuring suboptimal recognition results. We aspire to extract a suboptimal filter, capable of extracting the best features from the image. Fig. 3 illustrates the structure of a convolutional filter [27].

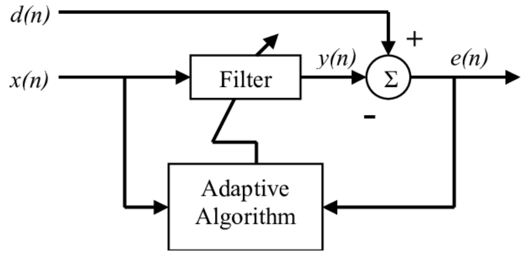


Fig. 3: Block diagram of adaptive filter [27].

F. Meta-heuristic Algorithm of the Kidney

KA is a population-based metaheuristic approach sharing certain characteristics with other population-based algorithms. Additionally, as its name implies, it emulates certain renal system methods. The primary components of the kidney process are described here. In the initial stage of KA, a random population of selected solutions is generated, and the objective performance is calculated for all of them. In each iteration, by moving towards the best-found solution so far, a new solution is generated for all candidate solutions. Then, utilizing the filtering operator, high-quality candidate solutions are directed to the FB filter repository within the population, while the remaining solutions are transferred to W. By examining the mechanisms of absorption, secretion, and excretion in the overall biological process of the kidney, the stages of the search are simulated. If a candidate solution is assigned to W in some embedded conditions of the algorithm, the algorithm provides another opportunity for this solution to improve itself before it can be transferred to FB. If this opportunity is not satisfactory, the solution is discarded from W, and a random solution is added to W. On the other hand, if, after filtration, a candidate solution is assigned to FB and the quality of this solution is not better than the worst solution found, it is discarded from FB. However, if the solution is better than the worst one, the worst solution will be discarded from FB. Eventually, the available solutions in FB are ranked, and the best solution is updated. The level of filtration is updated, and FB and W are merged. This iterative process continues until the termination criterion is achieved. The block diagram of the overall algorithm is presented in Fig. 4 [28].

Considering that the primary objective of using metaheuristic algorithms is optimization, these algorithms have been applied as a proposed approach, in which the coefficients of the extracted filters responsible for facial state recognition are optimized by the metaheuristic algorithm. These optimized coefficients obtained by metaheuristic algorithms are the primary reason for the superior recognition rate when compared to other methods. According to the results of various experiments, the size of the filters and the suboptimal number of filters were selected in the proposed method for feature extraction and achieving the superior recognition rate. In order to optimize the proposed method in various aspects, a number of optimization techniques have been employed.

Results and Discussion

A. Dataset

The FER2013 database consists of facial images with corresponding emotions, developed by Ian Goodfellow and colleagues [29]. This collection consists of 35,887 images in 8-bit grayscale format, with a size of 48×48 pixels, depicting facial emotions. The dataset is divided into three categories: 28,709 training data, 3,589 testing data, and 3,589 validation data.

All images in the database are labeled in such a way that each image falls into one of the seven main categories of facial emotions: anger, disgust, fear, happiness, sadness, surprise, and neutral.

In regards to the distribution of emotions, 35,887 images are divided as follows: anger with 4,953 images, disgust with 547 images, fear with 5,121 images, joy with 8,989 images, sadness with 6,077 images, surprise with 4,002 images, and neutrality with 9,816 images. Fig. 5(a) illustrates a sample of images from the FER2013 database. In recent years, due to the popularity of FER2013 and its free accessibility, various methods have been employed to develop high-accuracy models for facial emotion recognition.

JAFFE is a grayscale dataset collected from psychological experiments. Expression is a mixture of different facial emotions collected in a laboratory-controlled environment. This dataset comprises 213 images of seven different facial expressions. Fig. 5(b) shows seven images from the dataset.

B. Performance Parameters

One of the most crucial steps following the construction and design of a model is its evaluation and analysis. After analysis, the results can be categorized into two groups: positive and negative. Subsequently, the quality of the algorithm can be assessed using relevant indicators. After analysis, the data can be divided into four groups in terms of categorization. TP (true positive.e., positive and classified as positive), FP (false positive i.e., negative but classified as positive), TN (true negative i.e., negative and classified as negative), and FN (false negative i.e., positive but classified as negative).

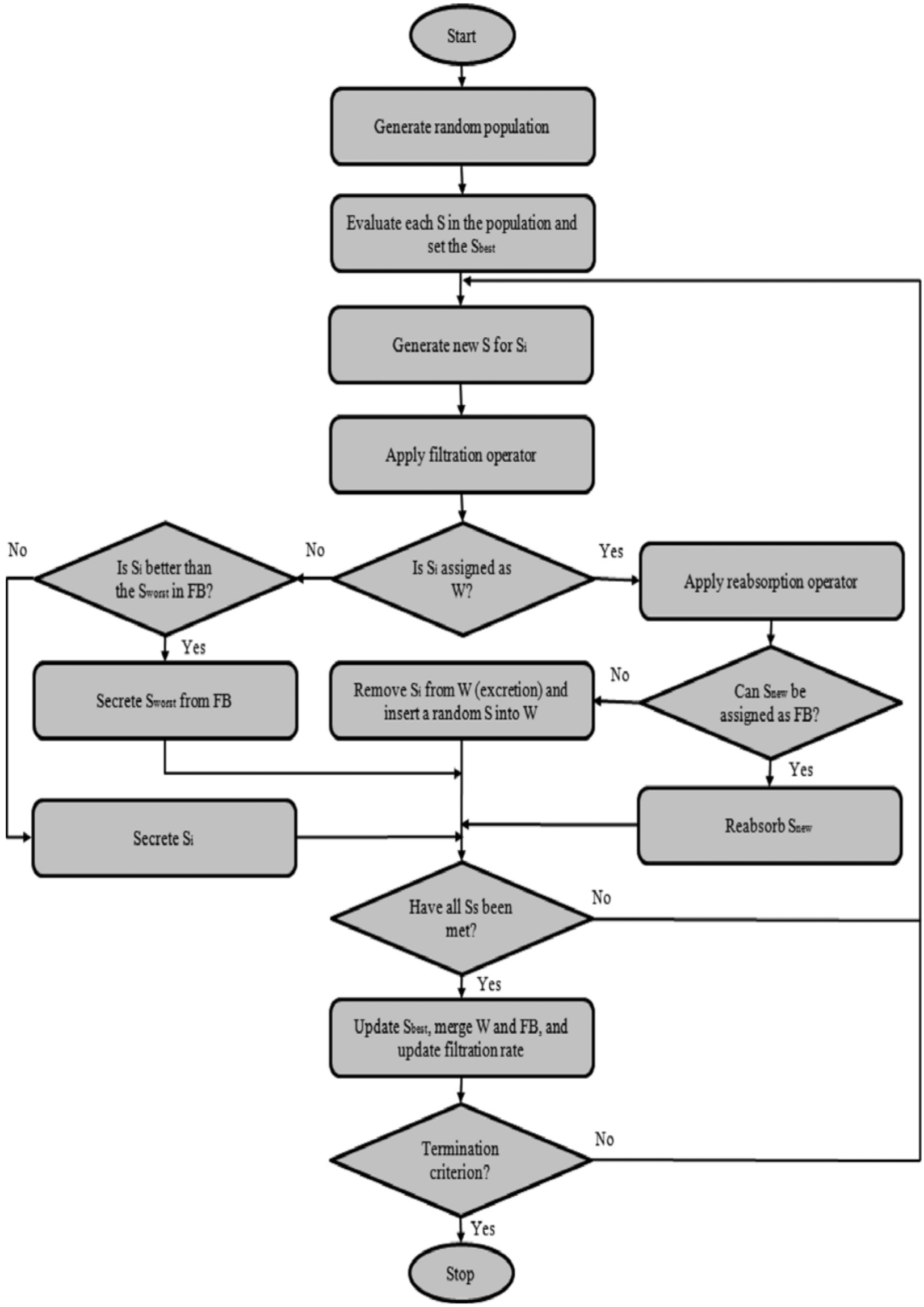
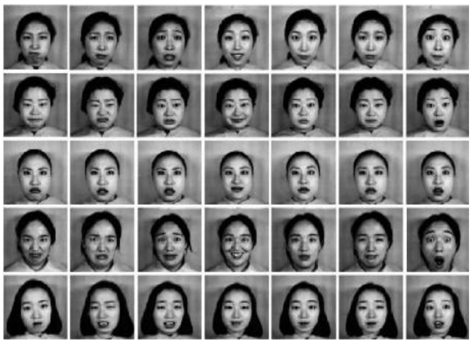


Fig. 4: Flowchart of KA [28].

(a)



(b)

Fig. 5: Datasets (a) FER2013[29] (b) JAFFE.

The expressions used to evaluate various performance parameters are given in (1) to (4) [30].

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

$$Recall = \frac{TP}{TP + FN} \tag{2}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{3}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{4}$$

Table 3: The specific details of the experiment

Experiment specifications	Details
Metaheuristic algorithm	kidney
Population size	20
Iterations	250
Valid Padding Convolution	stride=1
Classifier	KNN(K=3) & MLP
Filter size	3x3 , 5x5 , 7x7, 9x9

This experiment, which was conducted based on Python software, was executed on the Google Colab platform, with the following specifications, which were used in the experiment. [31]

C. The Extraction of a Suboptimal Single Filter Using the KA Algorithm

Table 2, Table 3, Table 4, and Table 5 present the Confusion Matrix for the extraction of single filters of sizes 3x3, 5x5, 7x7, and 9x9 using the KA algorithm on the nearest neighbor classifiers. In Fig. 6, Fig. 7, Fig. 8, and Fig. 10, the optimization progress and extracted filters are shown.

Table 1: confusion matrix for extracting the suboptimal 3x3 single filter using the KA algorithm on KNN classifier (%)

Confusion Matrix		Actual Class						
		angry	disgust	fear	happy	sad	surprise	neutral
Predicted Class	angry	53.3	10.1	9.5	2.8	9.63	3	9.68
	disgust	9.4	72	4.6	2	2	3.5	2.5
	fear	15.15	14	60	1.5	17.82	13.2	6.2
	happy	2.5	2.6	2.5	85	2	6	4
	sad	10.7	1	7.4	2.2	55.5	1.1	16.16
	surprise	2	0	13	3	1.5	74.7	4.94
	neutral	6.95	0	3	4.5	11.55	1.5	56.52

Table 2: confusion matrix for extracting the suboptimal 5x5 single filter using the KA algorithm on KNN classifier (%)

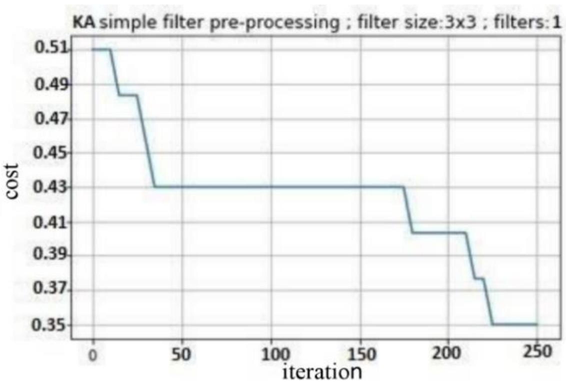
Confusion Matrix		Actual Class						
		angry	disgust	fear	happy	sad	surprise	neutral
Predicted Class	angry	57.66	9.6	8.43	2.2	8.63	1.6	8.88
	disgust	8.41	75	3.5	1.3	1.6	1.32	1.6
	fear	15.15	13.14	62.45	1.4	17.6	12.55	6.3
	happy	2.14	2	2.2	86.5	2.3	3.9	4.12
	sad	10.8	0	7.2	2.5	58.8	1.1	16.2
	surprise	1	0	12.12	1.5	1.5	77.9	1.99
	neutral	4.84	0.26	4.1	4.6	9.57	1.63	60.91

Table 4: confusion matrix for extracting the suboptimal 7x7 single filter using the KA algorithm on KNN classifier (%)

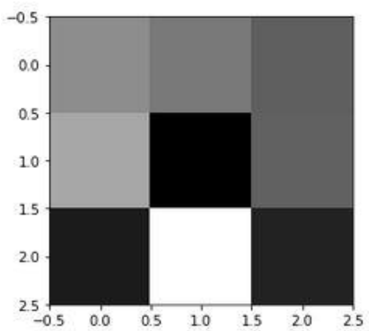
Confusion Matrix		Actual Class						
		angry	disgust	fear	happy	sad	surprise	neutral
Predicted Class	angry	52.3	11.4	9.5	1.8	8.38	2	8.18
	disgust	10.4	70	4.6	3.5	2	1.5	4
	fear	15.15	13	58	2	17.82	13.2	6.7
	happy	3.5	2.6	3.5	83	3	3.5	5
	sad	9.7	1	6.9	2.2	54	1.1	14.66
	surprise	3	1	12.5	4	2.5	77.2	6.44
	neutral	5.95	1	5	4.5	12.8	1.5	55.02

Table 5: confusion matrix for extracting the suboptimal 9x9 single filter using the KA algorithm on KNN classifier (%)

Confusion Matrix		Actual Class						
		angry	disgust	fear	happy	sad	surprise	neutral
Predicted Class	angry	49	12.4	10	2.3	9.38	2	6.68
	disgust	11.7	68	5.6	4.5	5	2.1	5.5
	fear	16.15	14	55.5	3.5	16.22	12.6	7.7
	happy	4	2.6	4.5	80.5	1.5	4.5	6
	sad	10.2	2	7.8	1.7	51	2.1	12.66
	surprise	3.3	0	11.6	5	2.5	75.2	8.94
	neutral	5.65	1	5	2.5	14.4	1.5	52.52

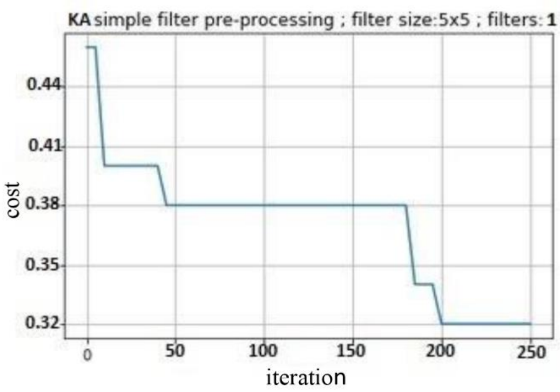


(a)

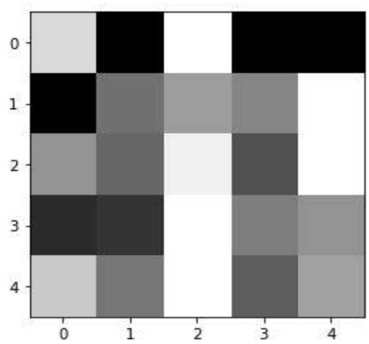


(b)

Fig. 6: (a) optimization progress, (b) adaptive filter 3x3.



(a)



(b)

Fig. 7: (a) optimization progress, (b) adaptive filter 5x5.

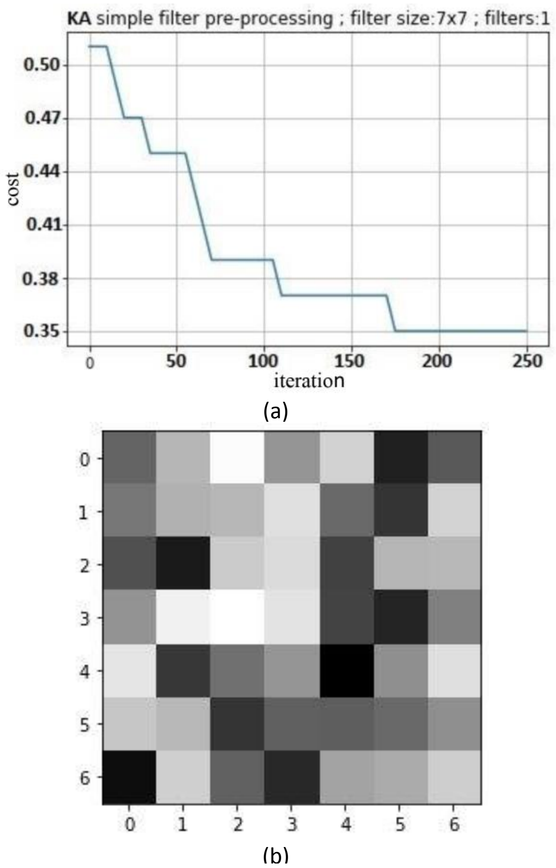


Fig. 8: (a) optimization progress, (b) adaptive filter 7x7.

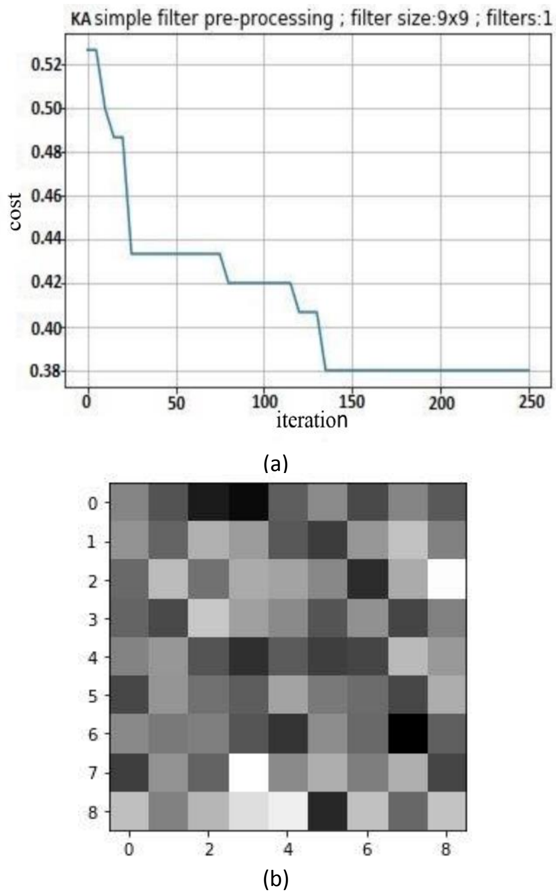


Fig. 9: (a) optimization progress, (b) adaptive filter 9x9.

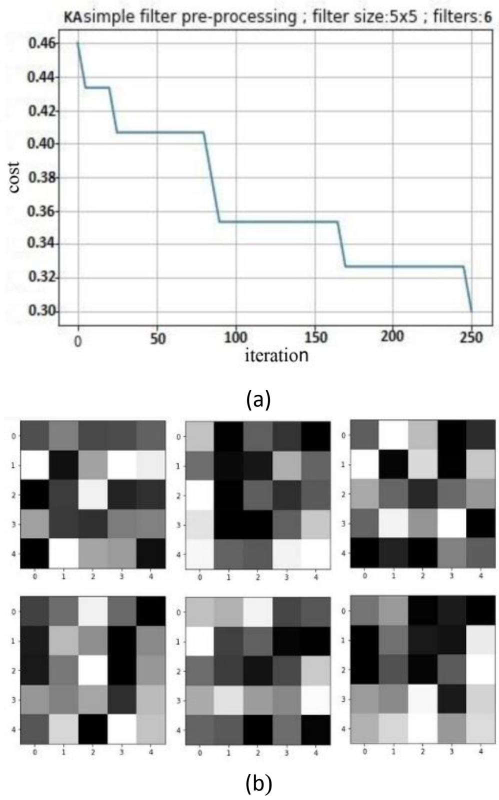


Fig. 10: (a) optimization progress, (b) filter bank 5x5(knn).

Table 6 specifies the recognition rate and number of extracted features for each filter. The Confusion Matrix shows which individual filters have performed more successfully in specific classes. In each category, the numbers on the main diagonal indicate the accuracy rate of each filter.

Table 6: Comparing the accuracy results of facial expression detection and the number of features extracted by the overall algorithm for filters of different sizes

Filter Size	Accuracy	number of features
3×3	65.3%	900
5×5	68%	784
7×7	65%	676
9×9	62%	576

As can be seen in Table 6, the smaller the filter size, the greater the number of features extracted, emphasizing finer details. This filter, therefore, has a better local search ability. Conversely, as the filter size increases, the number of features decreases, resulting in more emphasis on general characteristics. As a result, a larger filter size is preferred for global searching. Therefore, these results will be used to develop suboptimal filterbanks for other experiments in order to achieve a more favorable outcome. Table 6 details the accuracy results of various facial expression recognition models and the number of features extracted by the ka algorithm for optimized feature extraction.

The suboptimal adaptive filter with a size of 5×5 delivers the highest accuracy, achieving 68% and yielding a total of 784 features. Consequently, the next phase involves applying filter banks of 5×5 pixels to investigate their impact on facial expression recognition performance.

D. The Extraction of a Suboptimal 5x5 Filter Bank Using the KA Algorithm Via KNN

In the k-nearest neighbor classifier and KA algorithm, the performance accuracy was found to be 70%. Table 7 presents the confusion matrix of the optimized 5x5 filter bank (6 filters) on the knn classifier. The filter bank and optimization progress are depicted in Fig. 10. The number of features in this experiment is 4704.

Table 7: confusion matrix for extracting the suboptimal 5x5 filter bank using the KA algorithm on KNN classifier (%)

Confusion Matrix		Actual Class						
		angry	disgust	fear	happy	sad	surprise	neutral
Predicted Class	angry	59.66	8.6	8.43	1.5	9.23	1.1	7.39
	disgust	7.91	78	3.7	1	1	0.82	3.09
	fear	14.65	11.64	64.95	0.9	15.6	12.55	5.3
	happy	1.64	1.27	1.7	88	2	2.9	4.12
	sad	10.3	0	5.2	3	61.1	0.9	15.2
	surprise	2	0	13.12	1.3	3.25	80.1	1.24
	neutral	3.84	0.49	2.9	4.3	7.82	1.63	63.66

E. The Extraction of a Suboptimal 5x5 Filter Bank Using the KA Algorithm Via MLP

In the Multi-Layer Perceptron (MLP) neural network classifier and KA algorithm, the performance accuracy was found to be 78%. Table 7 presents the confusion matrix of the optimized 5x5 filter bank (6 filters) on the mlp classifier. The filter bank and optimization progress are depicted in Fig. 11. The number of features in this experiment is 4704. The Performance Parameters of the final model are presented in Table 7.

The accuracy of the proposed classification method and previous methods has been performed on the FER2013 database in Table 8. The best-reported accuracy is associated with the Ensemble ResMaskingNet method, combined with 6 other CNNs, achieving an accuracy of 82.76% [41]. In this article, a method is proposed for recognizing facial expressions using metaheuristic algorithms for optimizing feature extraction filters.

The method uses a continuous confrontation between optimized filter weights and nearest neighbor classifiers, as well as multi-layer perceptron neural networks to extract accurate features from the dataset and ensure accurate recognition of facial expressions. A stronger classifier is substituted in the proposed model by enhancing the weight structure of feature extraction filters. The computational cost for identifying the network model decreases. Ultimately, based on comprehensive datasets of various sizes, the proposed method is validated for its efficient performance in facial state recognition. It has been demonstrated that the proposed method offers distinct advantages in terms of precision of recognition, processing speed, and optimized filters. In this paper, an accuracy of 78% was achieved.

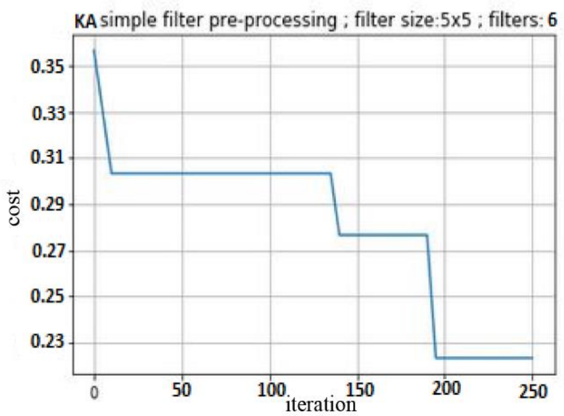
Table 9 compares the accuracy rates of different methods with the proposed approach, indicating that the proposed method has exhibited superior performance.

Table 8: confusion matrix for extracting the suboptimal 5x5 filter bank using the KA algorithm on MLP classifier (%)

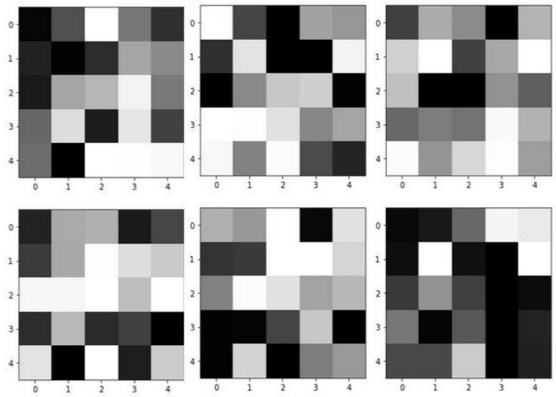
Confusion Matrix		Actual Class						
		angry	disgust	fear	happy	sad	surprise	neutral
Predicted Class	angry	67.76	7.23	6.23	0.5	5.8	0.1	4.23
	disgust	5.91	68.67	2.7	0.6	0	0.54	1.83
	fear	11.23	11.6	73.17	0.9	4.64	9.55	3.3
	happy	1.43	2	0.9	94	0.87	1.9	2.33
	sad	9.3	0.6	3.2	2	86.6	0.4	13.2
	surprise	3	2.25	9.12	0.3	0.5	85.67	1.24
	neutral	1.37	7.65	4.68	1.7	1.59	1.84	73.87

Table 9: Summary of results on FER2013

Method	Accuracy	Percision	Recall	F1-score
[16]CNN [16]	62.44	61.11	55.10	0.579
AlexNet	63.41	60.23	54.10	0.570
GoogleNet [20]	65.20	59.54	55.11	0.572
Human Accuracy [33]	65.5	62.33	60.22	0.612
VGG+SVM [19]	66.31	64.62	59.78	0.621
Conv+Inception [21]	66.40	65.14	60.32	0.626
Bag of Words [18]	67.40	66.50	62.41	0.643
Local Learning Bow [33]	67.48	66.74	60.78	0.636
Deep Emotion [34]	70.02	67.33	60.41	0.636
Attentional ConvNet [17]	70.02	68.38	65.32	0.668
EfficientNet	70.42	68.92	64.10	0.664
CNN+SVM [22]	71.20	69.08	65.52	0.672
ARM(ResNet-18) [23]	71.38	69.79	62.44	0.658
Inception [24]	71.60	68.19	66.14	0.671
Inception-v1	71.85	69.34	65.08	0.671
Ad-Corre [35]	72.3	69.68	67.14	0.683
ResNet [24]	72.40	70.30	66.50	0.683
SE-Net50 [36]	72.50	70.68	66.23	0.683
VGG [24]	72.70	70.92	67.38	0.691
Inception-v3	72.91	71.08	67.22	0.690
DenseNet-121	73.16	70.82	67.91	0.693
ResNet50 [36]	73.20	71.21	68.49	0.698
ResNet152	73.27	70.60	66.20	0.683
CNNs and BOVW + global SVM [37]	73.25	71.79	70.12	0.709
CBAM ResNet50	73.32	71.49	65.41	0.683
ResNet34v2	73.65	70.07	66.82	0.693
ResNet18 With Tricks [38]	73.7	71.33	67.27	0.692
Residual Masking Network [33]	74.14	71.60	73.20	0.723
LHC-NetC [39]	74.28	73.80	70.30	0.720
LHC-Net [39]	74.42	72.89	66.31	0.694
CNNs and BOVW + local SVM [37]	75.42	72.49	67.11	0.696
Segmentation VGG-19 [40]	75.97	73.38	68.52	0.708
Ensemble ResMaskingNet with 6 other CNNs [41]	76.82	73.17	68.23	0.706
Our Method	**78**	**77.62**	**72.08**	**0.747**



(a)

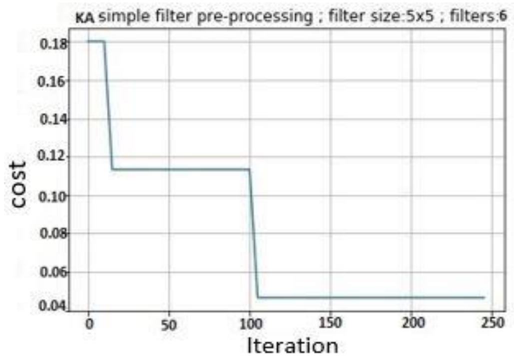


(b)

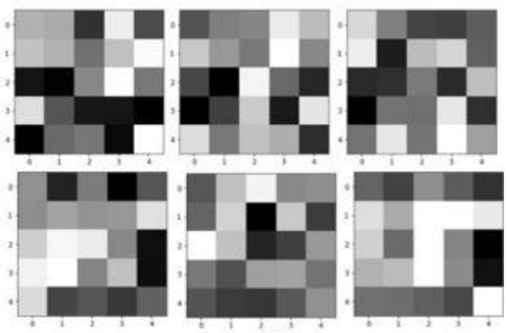
Fig. 11: (a) optimization progress, (b) filter bank 5x5(mlp) (fer2013).

F. The Final Proposed Model Was Evaluated on the JAFFE Dataset

It was shown that the proposed approach has outperformed other competitors on the JAFFE [42] database, which corresponds to a database of Japanese women, achieving an accuracy of 94.88 percent. presents the confusion matrix of the optimized 5x5 filter bank (6 filters) on the mlp classifier. The filter bank and optimization progress are depicted in Fig. 12.



(a)



(b)

Fig. 12: (a) optimization progress, (b) filter bank 5x5(mlp) (jaffe).

The accuracy of the proposed classification method has been compared with previous methods on the JAFFE accuracy of 94.88% was achieved. In Table 10, we compare the accuracy rates of different methods using the proposed approach, demonstrating that the proposed method has demonstrated superior performance on the dataset JAFFE when compared with other methods.

Table 10: Summary of results on JAFFE

Reference	Model	Accuracy
Sajjanhar et al. [42]	Pre-trained InceptionV3	75.88
	Pre-trained VGG19	94.71
	Pre-trained VGG-Face	86.67
Bhatti et al. [43]	RELM	91.67
Minaee et al. [44]	Attentional CNN	92.8
Kartheek et al. [45]	SVM	66.2
Sun et al. [46]	PCANet	71.38
	LDANet	70.18
Goutam Kumar Sahoo et al. [47]	Pre-trained AlexNet	65.62
	Pre-trained SqueezeNet	57.8
	Pre-trained VGG19	84.4
Ghasemi et al.	***Our Method***	***94.88***

G. Comparative Analysis of Training Time of the Proposed Method

Based on the performance accuracy of the proposed method on two databases, the proposed method is shown to be superior to other methods. The training time of the proposed model is analyzed using AlexNet, SqueezeNet, and VGG19 models on the JAFFE dataset. Table 11 shows that the proposed model requires less training time than the existing model. The model is trained with a training and validation data split of 70–30%.

Table 11: The performance of trained networks on the Jaffe dataset in terms of training time

Model	Training time (Sec)
AlexNet	342
SqueezeNe	284
VGG1	920
Our Method	276

H. Analyses of Facial Expressions with Confusion Matrix Results

An evaluation of a classifier can be conducted by using the confusion matrix. Essentially, the confusion matrix represents the results of classification based on the actual information available and is extremely useful in evaluating the performance of a classifier for one or more classes in particular. This matrix indicates the recognition rate of correctly identifying samples by its main diagonal.

Table 8 shows the confusion matrix resulting from the classification on the 2013-FER dataset. In facial expression recognition, the detection of the state of happiness is generally easier compared to other states due to the presence of more pronounced facial features. Comparatively to other classes in the dataset, the recognition of the state of happiness performed better than other classes. It is important to note that the state of sadness exhibits many similarities with other emotional states such as anger and fear. The manifestations of emotions such as anger, fear, and sadness are somewhat similar, so it is possible for diagnoses to be made in error. The proposed framework, however, has shown higher performance in the states of sadness and fear when compared to other states in the 2013-FER dataset despite lower accuracy rates.

Table 12 presents the evaluation results of the proposed structure on the JAFFE [48] dataset. The similarity between these two emotions in this dataset makes it difficult to distinguish between the sad and fearful expressions. In addition, there have been some errors in the detection of the surprise expression due to the similarity between the surprise and fear expressions. Overall, the proposed framework performed well on the dataset.

Table 12: confusion matrix for extracting the suboptimal 5x5 filter bank using the KA algorithm on MLP classifier (%)(JAFFE)

Confusion Matrix		Actual Class						
		angry	disgust	fear	happy	sad	surprise	neutral
Predicted Class	angry	96	2	2	0	0	0	1
	disgust	3	97	1	0	1	0	0
	fear	0	0	95	1	8.26	7.23	0
	happy	0	0	0	96	0	0	1
	sad	0	0	0	0	90.74	0	2
	surprise	1	1	1	3	0	92.77	0
	neutral	0	0	0	0	0	0	96

Conclusion

This article presents a method to improve facial expression recognition using an optimized filter and the KA metaheuristic algorithm. The method incorporates four filters with sizes 3×3, 5×5, 7×7, and 9×9. Each filter affects the number of features used. As the size of the filters increases, the number of features decreases. The selection of appropriate facial expression features plays an important role in increasing the accuracy of facial expression recognition. Different sizes of filters were extracted to improve facial expression recognition accuracy, and it was found that a 5×5 filter size yields the highest accuracy. In addition to improving accuracy by utilizing suboptimal filters for facial expression recognition, the suboptimal filter has also contributed to enhancing algorithm precision. This is because, after applying the filter, images are obtained that are not intelligible to us. By utilizing the nearest neighbor classifier and selecting the suboptimal k value, as well as the KA metaheuristic algorithm, we have been able to extract the best features for facial expression recognition. The suboptimal size of a single filter was therefore determined, and facial expression recognition was performed using filter banks, which achieved better accuracy than a single filter. In the final stage, instead of using the k-nearest neighbors (kNN) classifier, a perceptron neural network was employed, which is a powerful classifier, to optimize facial expression recognition.

Based on the obtained results, it appears that this approach has several advantages over other methods, including the reduced number of parameters, the decreased amount of training time, the ease of implementation, and the availability of filters with optimized parameters. These reasons substantiate the superiority of this method over others. Furthermore, the proposed method has shown higher accuracy in recognizing facial expressions than other methods. Our algorithm has demonstrated superior performance in comparison to the other methods.

Author Contributions

Dr. Razavi was the supervisor and Dr.Mohammadzade and Taghipour adviser of the current research paper. They sketched the research framework and the roadmap. Also, they analyzed the results. E. Ghasemi searched in authentic journals to gather all relevant papers. Also, he collected the data and wrote the manuscript. Dr. Razavi, Dr. Mohammadzade, Dr. Taghipour and E. Ghasemi interpreted the results.

Acknowledgment

We would like to express our gratitude to the Research Institute of Communication and Information Technology and the Development Headquarters of Connectivity and Communication Technologies of the Presidential Vice President for Science, Technology and Knowledge-Based Economy, who supported Mr. Ehsan Ghasemi's thesis, and also to the research supervisor Dr. Shaghayegh Naderi.

Conflict of Interest

The authors declare no potential conflict of interest regarding the publication of this work. In addition, the ethical issues including plagiarism, informed consent, misconduct, data fabrication and, or falsification, double publication and, or submission, and redundancy have been completely witnessed by the authors.

Abbreviations

CNN	Convolutional Neural Network
ASD	Autism Spectrum Disorders
RLUE	rectified linear unit
ICML	International Conference on Machine Learning
ARM	Advanced RISC Machines
MTCNN	Multi-task Cascade Convolutional Network
FER	Face Expression Recognition
KNN	K-Nearest Neighbor
MLP	Multi Layer Perceptron
KA	Kidney Algorithm
FB	Filtering Better solution
W	Worse solution
TP	True Positive
FP	False Positive
TN	True Negative
FN	False Negative

References

[1] S. Minaee, M. Minaei, A. Abdolrashidi, "Deep-emotion: Facial expression recognition using attentional convolutional network," Sensors, 21(9): 3046, 2021.

[2] S. Kaur, N. Kulkarni, "Emotion recognition-a review," Int. J. Appl. Eng. Res., 16(2): 103-110, 2021.

[3] E. Palagi, A. Celeghin, M. Tamietto, P. Winkielman, I. Norscia, "The neuroethology of spontaneous mimicry and emotional contagion in human and non-human animals," Neurosci. Biobehav. Rev., 111: 149-165, 2020.

[4] P. Singh, R. K. Mishra, S. Urolagin, V. Sharma, "Enhancing security by identifying facial check-in using deep convolutional neural network," in Proc. 3rd International Conference on Advances in Computing, Communication Control and Networking (ICAC3N): 1006-1010, 2021.

[5] J. He, "Algorithm composition and emotion recognition based on machine learning," Comput. Intell. Neurosci., 2022 .

[6] K. Richardson et al., "Robot enhanced therapy for children with autism (DREAM): A social model of autism," IEEE Technol. Soc. Mag., 37(1): 30-39, 2018.

[7] K. Osaka, R. Tanioka, F. Betriana, T. Tanioka, Y. Kai, R. C. Locsin, "Robot therapy program for patients with dementia: Its framework and effectiveness," Information Systems - Intelligent Information Processing, p. 87, 2021.

[8] D. Cruz-Sandoval, A. Morales-Tellez, E. B. Sandoval, J. Favela, "A social robot as therapy facilitator in interventions to deal with dementia-related behavioral symptoms," in Proc. ACM/IEEE International Conference on Human-Robot Interaction: 161-169, 2020.

[9] B. Szymona et al., "Robot-assisted autism therapy (RAAT). Criteria and types of experiments using anthropomorphic and zoomorphic robots. Review of the research," Sensors, 21(11): 3720, 2021.

[10] Z. Wang et al., "Masked face recognition dataset and application," arXiv:2003.09093v2, 2020.

[11] S. MacHiraju, S. Urolagin, R. K. Mishra, V. Sharma, "Face mask detection using keras, opencv and tensorflow by implementing mobilenetv2," in Proc. 3rd International Conference on Advances in Computing, Communication Control and Networking (ICAC3N): 1485-1489, 2021.

[12] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, "Gradient-based learning applied to document recognition," Proc. IEEE, 86(11):2278-2324, 1998.

[13] S. Fooladi, H. Farsi, S. Mohamadzadeh., "Detection and classification of skin cancer using deep learning, " J. Birjand Univ. Med. Sci., 26(1): 44-53, 2019.

[14] S. Albawi, T. A. Mohammed, S. Al-Zawi, "Understanding of a convolutional neural network," in Proc. International Conference on Engineering and Technology (ICET), 2017.

[15] B. Han, J. Sim, H. Adam, "BranchOut: Regularization for online ensemble tracking with convolutional neural networks," in Proc. 30th IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017.

[16] K. Liu, M. Zhang, Z. Pan, "Facial expression recognition with CNN Ensemble," in Proc. 2016 International Conference on Cyberworlds, CW 2016.

[17] S. Minaee, A. Abdolrashidi, "Deep-emotion: facial expression recognition using attentional convolutional network," arXiv:1902.01019v1, 2019.

[18] R. T. Ionescu, M. Popescu, C. Grozea, "Local learning to improve bag of visual words model for facial expression recognition," in Proc. ICML 2013 Workshop on Representation Learning, 2013.

[19] M. I. Georgescu, R. T. Ionescu, M. Popescu, "Local learning with deep and handcrafted features for facial expression recognition," IEEE Access, 7: 64827- 64836, 2019.

[20] P. Giannopoulos, I. Perikos, I. Hatzilygeroudis, "Deep learning approaches for facial emotion recognition: A case study on FER-2013," Smart Innovation, Systems and Technologies, 85: 1-16, 2018.

[21] A. Mollahosseini, D. Chan, M. H. Mahoor, "Going deeper in facial expression recognition using deep neural networks," in Proc. IEEE Winter Conference on Applications of Computer Vision (WACV), 2016.

[22] Y. Tang, "Deep learning using linear support vector machines," arXiv, 1306.0239, 2013.

[23] J. Shi, S. Zhu, Z. Liang, "Learning to amend facial expression representation via de-albino and affinity," arXiv Prepr. arXiv2103.10189, 2021.

[24] C. Pramerdorfer, M. Kampel, "Facial expression recognition using convolutional neural networks: state of the art," arXiv Prepr. arXiv1612.02903, 2016.

[25] K. Zhang, Z. Zhang, Z. Li, Y. Qiao, "Joint face detection and alignment using multitask cascaded convolutional networks," IEEE Signal Process. Lett., 23(10): 1499-1503, 2016.

[26] T. Shanmugaraja, R. Dhivaya Devi, M. Supriya, K. Murugan, T. Venkatesh, K. Siddarthraju, "A critical review on applications of adaptive filters," J. Crit. Rev., 7(6): 656-659, 2020.

[27] S. R. Ashokkumar, G. MohanBabu, S. Anupallavi, "A novel two-band equilateral wavelet filter bank method for an automated detection of seizure from EEG signals," Int. J. Imaging Syst. Technol., 30(4): 978-993, 2020.

[28] N. S. Jaddi, J. Alvankarian, S. Abdullah, "Kidney-inspired algorithm for optimization problems," Commun. Nonlinear Sci. Numer. Simul., 42: 358-369, 2017.

[29] I. J. Goodfellow et al., "Challenges in representation learning: A report on three machine learning contests," Neural Netw., 64: 59-63, 2015.

[30] M. Taheri, M. Rastgarpour, A. Koochari, "A novel method for medical image segmentation based on convolutional neural networks with SGD optimization," J. Electr. Comput. Eng. Innov., 9 (1): 37-46, 2020.

[31] S. Fooladi, H. Farsi, S. Mohamadzadeh, "Segmenting the lesion area of brain tumor using convolutional neural networks and fuzzy k-means clustering," Int. J. Eng., 36(8): 1556-1568, 2023.

[32] H. Bayati, M. B. Dowlatshahi, M. Paniri, "Multi-label feature selection based on competitive swarm optimization," J. Soft Comput. Inf. Technol., 9(4): 56-69, 2020.

[33] I. J. Goodfellow et al., "Challenges in representation learning: A report on three machine learning contests," Neural Netw., 64: 59-63, 2015.

[34] S. Minaee, M. Minaei, A. Abdolrashidi, "Deep-emotion: Facial expression recognition using attentional convolutional network," Sensors, 21(9): 3046, 2021.

[35] A. P. Fard, M. H. Mahoor, "Ad-corre: Adaptive correlation-based loss for facial expression recognition in the wild," IEEE Access, 10: 26756-26768, 2022.

[36] A. Khanzada, C. Bai, F. T. Celepcikay, "Facial expression recognition with deep learning," arXiv Prepr. arXiv2004.11823, 2020.

[37] M. I. Georgescu, R. T. Ionescu, M. Popescu, "Local learning with deep and handcrafted features for facial expression recognition," IEEE Access, 7: 64827-64836, 2019.

[38] Y. Khaireddin, Z. Chen, "Facial emotion recognition: State of the art performance on FER2013," arXiv Prepr. arXiv2105.03588, 2021.

[39] R. Pecoraro, V. Basile, V. Bono, "Local multi-head channel self-attention for facial expression recognition," Information, 13(9): 419, 2022.

[40] S. Vignesh, M. Savithadevi, M. Sridevi, R. Sridhar, "A novel facial emotion recognition model using segmentation VGG-19 architecture," Int. J. Inf. Technol., 15(4): 1777-1787, 2023.

[41] L. Pham, T. H. Vu, T. A. Tran, "Facial expression recognition using residual masking network," in Proc. 25Th International Conference on Pattern Recognition (ICPR): 4513-4519, 2020.

[42] A. Sajjanhar, Z. Wu, Q. Wen, "Deep learning models for facial expression recognition," in Proc. 2018 International Conference on Digital Image Computing: Techniques and Applications (DICTA): 1-6, 2018.

[43] Y. K. Bhatti, A. Jamil, N. Nida, M. H. Yousaf, S. Viriri, S. A. Velastin, "Facial expression recognition of instructor using deep features and extreme learning machine," Comput. Intell. Neurosci., 2021: 1-17, 2021.

[44] S. Minaee, M. Minaei, A. Abdolrashidi, "Deep-emotion: Facial expression recognition using attentional convolutional network," Sensors, 21(9): 3046, 2021.

[45] M. N. Kartheek, M. V. N. K. Prasad, R. Bhukya, "Windmill graph based feature descriptors for facial expression recognition," Optik (Stuttg)., 260: 169053, 2022.

[46] Z. Sun, H. Zhang, S. Ma, Z. Hu, "Combining filtered dictionary representation based deep subspace filter learning with a discriminative classification criterion for facial expression recognition," Artif. Intell. Rev., 55(8): 6547-6566, 2022.

[47] G. K. Sahoo, S. K. Das, P. Singh, "Performance comparison of facial emotion recognition: A transfer learning-based driver assistance

framework for in-vehicle applications," Circuits, Syst. Signal Process.: 1–28, 2023.

[48] M. J. Lyons, J. Budynek, S. Akamatsu, "Automatic classification of single facial images," IEEE Trans. Pattern Anal. Mach. Intell., 21(12): 1357-1362, 1999.

Biographies

Ehsan Ghasemi received his AD.Sc. degree in Electrical Engineering from university of Islamic Republic of Iran broadcasting, in 2014 and his B.Sc. degree in Electrical Engineering from Islamic Azad University, Birjand, in 2018 and M.Sc. degree in Electrical Engineering from the Birjand University, in 2021 respectively. He is currently a Ph.D. student at Birjand University to receive a Ph.D. degree in Electronics Engineering. His research interests include Computer Vision, Pattern Recognition, optimization algorithms and Artificial Intelligence.

- Email: ehsanghasemi91@birjand.ac.ir
- ORCID: 0009-0002-6185-9049
- Web of Science Researcher ID: NA
- Scopus Author ID: NA
- Homepage: NA

Seyyed Mohammad Razavi received his B.Sc. degree in Electrical Engineering from Amirkabir University of Technology, in 1994 and his M.Sc. and Ph.D. degree in Electrical Engineering from the Tarbiat Modares University, Iran, in 1996 and 2006 respectively. Now, he is a Full Professor in University of Birjand. His research interests include Computer Vision, Pattern Recognition and Artificial Intelligence.

- Email: smrazavi@birjand.ac.ir
- ORCID: 0000-0002-3493-7614
- Web of Science Researcher ID: AAF-7386-2021
- Scopus Author ID: 56214431100
- Homepage: https://cv.birjand.ac.ir/mrazavi/fa

Sajad Mohamadzadeh received the B.Sc. degree in Communication Engineering from Sistan & Baloochestan, University of Zahedan, Iran, in 2010. He received the M.Sc. and Ph.D. degree in Communication Engineering from South of Khorasan, University of Birjand, Birjand, Iran, in 2012 and 2016, respectively. Now, he works as Associte professor at department of electrical and computer engineering, University of Birjand, Birjand, Iran. His area research interests include Image and Video Processing, Deep Neural Network, Pattern recognition, Digital Signal Processing, Sparse Representation, and Deep Learning.

- Email: s.mohamadzadeh@birjand.ac.ir
- ORCID: 0000-0002-9096-8626
- Web of Science Researcher ID: NA
- Scopus Author ID: 57056477500
- Homepage: https://cv.birjand.ac.ir/mohamadzadeh/en

Mehran Taghipour-Gorjikolaie received the B.Sc. degree in Electrical Engineering from the University of Mazandaran, Iran, in 2008, and the M.Sc. (Hons.) and Ph.D. degrees in Electronic Engineering from the University of Birjand, Iran, in 2011 and 2016, respectively. He was a lecturer (assistant Professor) with the University of Birjand from 2016 to 2021, after that he was a research fellow in application AI, ML, and CV in healthcare and wellbeing at CVSSP, in the University of Surrey from 2021 to 2023, and now he is working as research fellow in London South Bank University. His research interests include the application of machine learning, artificial intelligence, computer vision and optimization algorithms.

- Email: mehran.taghipour-gorjikolaie@lsbu.ac.uk
- ORCID: 0000-0001-6132-8454
- Web of Science Researcher ID: NA
- Scopus Author ID: NA
- Homepage: NA

How to cite this paper:
E. Ghasemi, S. M. Razavi, S. Mohamadzadeh, M. Taghipour-Gorjikolaie, "Facial expression recognition through suboptimal filter design using a metaheuristic kidney algorithm," J. Electr. Comput. Eng. Innovations, 12(2): 425-438, 2024.

DOI: 10.22061/jecei.2024.10512.717

URL: https://jecei.sru.ac.ir/article_2096.html