POLİTEKNİK DERGİSİ

JOURNAL of POLYTECHNIC

ISSN: 1302-0900 (PRINT), ISSN: 2147-9429 (ONLINE)
URL: http://dergipark.org.tr/politeknik

Deep feature selection for facial emotion recognition based on BPSO and SVM

BPSO ve SVM'ye dayalı yüzde duygu tanıma için derin özellik seçimi

***Yazar(lar) (Author(s))**: Kenan DONUK[1], Ali ARI[2], Mehmet Fatih ÖZDEMİR[3], Davut HANBAY[4]*

ORCID[1]: 0000-0002-7421-5587

ORCID[2]: 0000-0002-5071-6790

ORCID[3]: 0000-0003-3563-054X

ORCID[4]: 0000-0003-2271-7865

To cite to this article: Donuk K., ARI A., ÖZDEMİR M. F. ve HANBAY D., "Deep feature selection for facial emotion recognition based on BPSO and SVM", ***Journal of Polytechnic***, 26(1): 131-142, (2023).

Bu makaleye şu şekilde atıfta bulunabilirsiniz: Donuk K., ARI A., ÖZDEMİR M. F. ve HANBAY D., "Deep feature selection for facial emotion recognition based on BPSO and SVM", ***Politeknik Dergisi***, 26(1): 131-142, (2023).

Erişim linki (To link to this article): http://dergipark.org.tr/politeknik/archive

DOI:10.2339/politeknik.992720

Deep Feature Selection for Facial Emotion Recognition Based on BPSO and SVM

Highlights

- *CNN model design and feature extraction for emotion recognition problem*
- *Use of the publicly available Fer+ facial expression dataset*
- *Feature selection with Binary Particle Swarm Optimization (BPSO) algorithm*
- *Classification of features with Support Vector Machine (SVM)*

Graphical Abstract

The contribution of the Binary Particle Swarm Optimization algorithm to the performance and speed of facial expression recognition was examined by classifying the features obtained as a result of the application of the BPSO algorithm to the last feature layer of the CNN model with the Support Vector Machine.

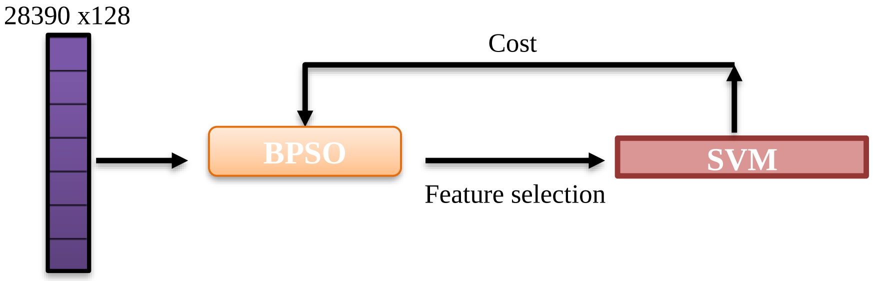

Figure. BPSO+SVM algorithm working procedure

Aim

The contribution of the application of the BPSO+SVM combination to the deep features on the accuracy and speed of face recognition was investigated.

Design & Methodology

Python software and Google Colab environment were used in CNN design and training.

Originality

The BPSO+SVM combination has been applied to CNN-based features.

Findings

The face recognition performance measured with the CNN model was 84.28%, the face recognition performance of the CNN+SVM combination was 84.81%, and the face recognition performance with the CNN+BPSO+SVM application was 85.74%. Runtime performances per image were measured as 0.52ms, 1.82ms, 1.12ms, respectively.

Conclusion

With the BPSO application to the CNN-based features of the Fer+ dataset, higher accuracy was measured than the CNN and CNN+SVM methods. The combination of CNN+BPSO+SVM has been shown to be better than CNN+SVM in terms of runtime performance.

Declaration of Ethical Standards

The author(s) of this article declare that the materials and methods used in this study do not require ethical committee permission and/or legal-special permission.

Deep Feature Selection for Facial Emotion Recognition Based on BPSO and SVM

Araştırma Makalesi / Research Article

Kenan DONUK[1*], Ali ARI[2], Mehmet Fatih ÖZDEMİR[2], Davut HANBAY[2]

[1]Cizre Meslek Yüksekokulu, Bilgisayar Programcılığı Bölümü, Şırnak Üniversitesi, Türkiye
[2]Mühendislik Fakültesi, Bilgisayar Müh. Bölümü, İnönü Üniversitesi, Türkiye

(Geliş/Received : 08.09.2021 ; Kabul/Accepted : 14.09.2021 ; Erken Görünüm/Early View : 27.09.2021)

ABSTRACT

Facial expressions, which are important social communication tools in our daily life, provide important information about the mental state of people. Research is being done to obtain this information accurately. The importance of these researchs in the field of human-computer interaction is increasing. Many methods have been used for the recognition of universal facial expressions such as neutral, happiness, surprise, sadness, anger, disgust, and fear by intelligent systems with high accuracy. Emotion recognition is an example of difficult classification due to factors such as ambient light, age, race, gender, and facial position. In this article, a 3-stage system is proposed for emotion detection from facial images. In the first stage, the CNN-based network is trained with the Fer+ dataset. The Binary Particle Swarm Optimization algorithm is applied for feature selection to the feature vector in the fully connected layer of the CNN network trained in the second stage. Selected features are classified by Support Vector Machine. The performance of the proposed system has been tested with the Fer+ dataset. As a result of the test, 85.74% accuracy was measured. The results show that the combination of BPSO and SVM contributes to the classification accuracy and speed of the FER+ dataset.

Keywords: Facial emotion recognition, convolutional neural network, binary particle swarm optimization, support vector machine.

BPSO ve SVM'ye Dayalı Yüzde Duygu Tanıma için Derin Özellik Seçimi

ÖZ

Günlük hayatımızda önemli sosyal iletişim aracı olan yüz ifadeleri, insanların ruhsal durumu hakkında önemli bilgiler vermektedir. Bu bilgiyi doğru bir şekilde elde etmek için araştırmalar yapılmaktadır. Bu araştırmaların insan-bilgisayar etkileşimi alanındaki önemi giderek artmaktadır. Nötr, mutluluk, şaşkınlık, üzüntü, öfke, iğrenme, korku gibi evrensel yüz ifadelerinin akıllı sistemler tarafından yüksek doğrulukla tanınması için birçok yöntem kullanılmıştır. Duygu tanıma, ortam ışığı, yaş, ırk, cinsiyet ve yüz pozisyonu gibi faktörler nedeniyle zorlu bir sınıflandırma örneğidir. Bu makalede, yüz görüntülerinden duygu tanıma için 3 aşamalı bir sistem önerilmiştir. İlk aşamada, tasarlanan CNN tabanlı ağ Fer+ veri seti ile eğitiliyor. İkinci aşamada, eğitilmiş olan CNN ağının tam bağlı katmanındaki özellik vektörüne özellik seçimi için İkili Parçacık Sürü Optimizasyon algoritması uygulanıyor. Seçilen özellikler Destek Vektör Makinesi tarafından sınıflandırılır. Önerilen sistemin performansı Fer+ veri seti ile test edilmiştir. Test sonucunda %85,74 doğruluk ölçülmüştür. Elde edilen sonuçlar İkili Parçacık Sürü Optimizasyon algoritması ve Destek Vektör Makinesi birleşiminin FER+ veri setinin sınıflandırma doğruluğuna ve hızına katkısını ortaya koymuştur.

Anahtar Kelimeler: Yüz duygu tanıma, evrişimsel sinir ağı, ikili parçacık sürü optimizasyonu, destek vektör makinesi.

1. INTRODUCTION

Facial expressions are important tools that reflect the mental state of a person. For this reason, facial emotion recognition (FER) has become an important tool for many research areas such as improving the teaching-learning process, detecting mental disorders, customer satisfaction, lie detection, fear detection and Autism [1-7]. The Facial Action Coding System (FACS), which was first introduced by Ekman and Freisen in 1978 and is a pioneering system in the definition of facial movements until today, defined universal facial expressions (anger, disgust, fear, happiness, sadness, and surprise) with 46 Action Units (AUs) on the face. Facial expressions are expressed with combinations of these action units [8]. Today, AUs are still used in research on many different subjects [9-13].

FER studies are examples of a difficult classification because of difficulties in facial expression recognition such as age, race, gender, face position, low resolution, background, skin diseases, brightness, and mask. Emotions are difficult to recognize and follow by humans. Thus, automatic and uninterrupted recognition of emotions by intelligent systems in human-computer interaction has become an important issue in machine learning. Many different methods have been used in the field of FER from past to present. Emotion detection in machine learning is divided into 3 categories. These are geometric, appearance and deep learning-based methods. Geometric-based methods are to obtain the feature vector

*Sorumlu Yazar (Corresponding Author)
e-posta : kenandonuk@sirnak.edu.tr

representing the facial expression by taking into account the positions and shapes of the turning points (eye, mouth, eyebrow, nose) on the face [14]. There are many geometric-based studies in the FER field [15-21]. Appearance-based methods are the use of features obtained from image processing techniques and calculations on pixel values. Methods such as Local Binary Patterns (LBP) [22-24], Histogram of Oriented Gradients (HOG) [25], Gabor Filters [26,27], Local Phase Quantization (LPQ) [28], Scale-Invariant Feature Transform (SIFT) [29,30] are examples of appearance-based methods. Thanks to the developing technology, deep learning-based methods have made progress in FER studies. Deep learning-based methods have emerged as Deep Neural Networks (DNN) with the deepening of Artificial Neural Networks (ANN) [31].

Traditional and the latest technology deep learning algorithms have been developed from past to present. The most popular of these algorithms is CNN. The first CNN architecture was the LeNet-5 network [32]. The structure of the network consists of convolution, pooling and full link layers. It has been reported that Deep Convolutional Neural Networks outperform traditional methods [33]. There are CNN-based studies in different fields in the literature [34-36].

There are also studies conducted with CNN architecture in the FER field [37-46]. FER studies consist of three phases. These are face detection, feature extraction from facial expression, and classification of these features.

In this study, a three-stage system is proposed in emotion classification. CNN-based deep learning model was used to extract the features of the images in the FER+ dataset. The dominant features were selected with the BPSO algorithm. The features selected with BPSO were classified by SVM method. The results show that the proposed system can be used to increase accuracy and speed in emotion classification. The performance of the proposed system has been tested on the FER+ dataset. The classification accuracy on the test set was calculated as 85.74%.

The rest of this paper is organized as follows; In the second chapter, studies in the literature in the field of FER are examined. In Chapter 3, the proposed system is explained in detail. In Chapter 4, the contribution of the proposed system to the classification accuracy and speed is examined. In Chapter 5, the results of the study are given.

2. RELATED WORKS

Niu et al. proposed a method to achieve high performance with traditional methods without using Deep Convolution Network (DNN). They applied traditional approaches such as Local Binary Patterns (LBP) and Oriented FAST and Rotated BRIEF (ORB) to the face area. The features obtained as a result of the approaches were combined. Combined features are given to Support Vector Machines (SVM) for classification. The proposed method was applied to Japanese Female Facial Expressions (JAFFE), The Extended Cohn-Kanade (CK+) and MMI facial expression datasets. In the experiments performed according to different feature sets, accuracies such as 92.4% and 88.5% from the JAFFE dataset, 99.2% and 93.2% from the CK+ dataset, 84.2% and 79.8% from the MMI dataset were obtained. These results showed the effectiveness of the proposed method [23]. Josephine Julina et al. analyzed facial features using Histogram of Oriented Gradients (HOG) and Local Binary Model (LBP). They tried to classify happy, sad and angry facial expressions from video frames with the system they established [47]. Christou et al. designed a CNN-based architecture with 5 convolutional layers. With the architecture they designed, they achieved an accuracy of 88.78% on the FER2013 dataset. They demonstrated the effect of data augmentation with an accuracy rate of 91.12% obtained by applying data augmentation to the FER2013 data set [43]. Mollahosseini et al. proposed a deep neural network architecture that includes the Inception layer structure. By combining the network outputs of the Inception layer, classification is performed with two fully connected layers. They have achieved a significant success in regional feature performance with the use of network within the network. The performance of the architecture has been tested on MultiPIE, MMI, CK +, DISFA, FERA, SFEW and FER2013 datasets. Evaluation of the results was made using the cross-validation technique with two different experiments. The success of the proposed architecture against the latest technology architectures has been demonstrated [45]. Bhandari et al. investigated the effect of using facial edge data in emotion recognition on performance. They designed an architecture consisting of two-tower and convolutional layers. In the first tower, semantic expressions on the face are extracted as features. In the other tower, features are extracted from the edge information obtained with the canny edge detection algorithm. The feature vectors of the two towers terminate in the classification layer. The architecture has been tested on the JAFFE and FER2013 datasets. This two-tower architecture provided a 6% increase in test accuracy in both datasets compared to the single-tower CNN model using raw images [41]. Gupta et al. designed a CNN network based on AlexNet architecture, which takes 30 temporal and spatial features from video datasets as input. These features collected in the EmotionalVLAN layer are transmitted to the classification layer. The proposed architecture has been tested on RML, BAUM-1s, eNTERFACE05, MMI, and FER2013 datasets. The proposed architecture achieved better results than the previously applied methods [40]. Li et al. tried to overcome the accuracy problem caused by overfitting in deep networks with the CNN-based ResNet-50 architecture. For the training and testing of the architecture, a dataset of 700 images was prepared from 20 subjects of different ages, careers and races. The overall crossover average of the network trained with 10 cross-validations is 95.39%. Architecture compared with

Biogeography-based optimization (BBO), Cat Swarm Optimization (CSO), Haar Wavelet Transform (HWT) and better results were obtained [48]. Liang et al. proposed a bidirectional long-short-term memory (BiLSTM) fusion network. In this network consisting of 3 parts, Deep Spatial Network (DSN) is used for spatial display features, Deep Temporal Network (DTN) is used to benefit from temporal information, and BiLSTM network is used for the formation of six basic expressions. Comparative experiments with 13 different methods on CK+, Oulu-CASIA and MMI datasets showed that the proposed method can achieve high performance [42]. Bargal et al. proposed an architecture to classify 8 emotional expressions (including contempt). The Acted Facial Expressions in the Wild (AFEW) 6.0 cropped video dataset combinations were used to train the architecture. In the proposed architecture, VGG13, VGG16 and ResNET networks are trained. The feature vectors of the trained networks are combined. The obtained feature vectors are added to the Statistical coding module (STAT), which produces a single feature vector, and classified via SVM. Dataset classification success resulted in validation 59.42% and testing 56.66% accuracy rates [49-51]. Li et al. revealed that the same emotion is expressed differently between human faces. They stated that this situation led to errors in emotion classification. Therfore, they suggested adding face identity information to the model to help emotion recognition models. The proposed architecture consists of two CNNs. These CNNs used DeepID for face identification and Resnet for emotion recognition. The output properties of these networks are combined and given as input to the fully connected layers. The proposed architecture has been evaluated in the CK+ and FER+ databases. They achieved 99.3% and 84.3% accuracy in the CK+ and FER+ datasets, respectively. Only the emotion classification trials in Resnet18 and Resnet18 + FC networks resulted in 83.1% and 83.4, respectively [33].

Some studies on the FER+ dataset used in the study are as follows. Barsoum et al. evaluated emotion recognition performance on a proposed VGG network. They used the FER+ dataset in the VGG architecture. By making use of the probabilistic distribution of emotion labels in the FER+ dataset, Majority Voting (MV), Multilabel Learning (ML), Probabilistic Label Drawing (PLD) and Cross-Entropy Loss (CEL) methods were used during the training. Accuracies of 83.852%, 83.966%, 84.986% and 84.716% were obtained from these methods, respectively [52]. Miao et al. proposed the shallow CNN (SHCNN) architecture to overcome the performance problem in traditional methods and the overfitting problem in deep learning. They used temporal features for classification. TV-L1 used optical flow method to extract temporal features in video frames. They applied the SHCNN architecture to classify optical flow images. The proposed architecture has been tested on FER2013, FER+, CASME, CASME II, SAMM datasets. It has been demonstrated that the architecture shows a positive performance compared to the latest technology. The success of this architecture in the FER+ dataset is 86.54% [39]. Lian et al. followed the DenseNet-BC architecture with three dense blocks in emotion classification. Class Activation Maps (CAM) were used to identify distinctive image regions in different facial expressions. They trained the system by dividing the face into 7 regions. They concluded that the most effective region in emotion classification is the mouth. The proposed architecture has been tested on the FER+ dataset. They achieved a higher result than the training of facial region combinations with 81.93% classification success in classification over the whole face [44]. Georgescu et al. presented an architecture in which they combine features extracted by three CNN-based models and a handcrafted model. Combined features are classified by SVM. CNN-based models used 16-layer VGG-face, 8-layer VGG-f, VGG-13 and Bag-Of-Visual-Words (BOVW) model as a handcrafted model. Scale-Invariant Feature Transform (SIFT) and k-means clustering were used to extract local features from images in the BOVW model. They tested the proposed architecture with FER, FER+ and AffectNet datasets. The classification success in the FER+ dataset is 87.76% [29]. Huang compared the accuracy of the architectures obtained by combining Resnet and VGG architectures in the early or late stages of education. It has been determined that the performance obtained by combining different architectures in the late stage is higher. It used linear SVM as the classification method. Test accuracy of 87.40% was obtained in the experiments on the late fusion architecture with the FER+ dataset [53].

3. PROPOSED METHOD

The proposed system consists of CNN training, BPSO feature selection and SVM classification stages. At the end of the CNN training, feature vectors (28390x128) of the data set images are obtained from the fully connected layer (128 neurons) before the classification layer. The BPSO algorithm is applied for the optimization of these feature vectors. Out of the 128 features represented for each image by the BPSO algorithm, the most valuable features giving the highest classification accuracy are obtained from the BPSO-SVM combination. The best feature vector obtained at the end of BPSO-SVM is classified in SVM. The flow chart of the proposed system is shown in Fig. 1.

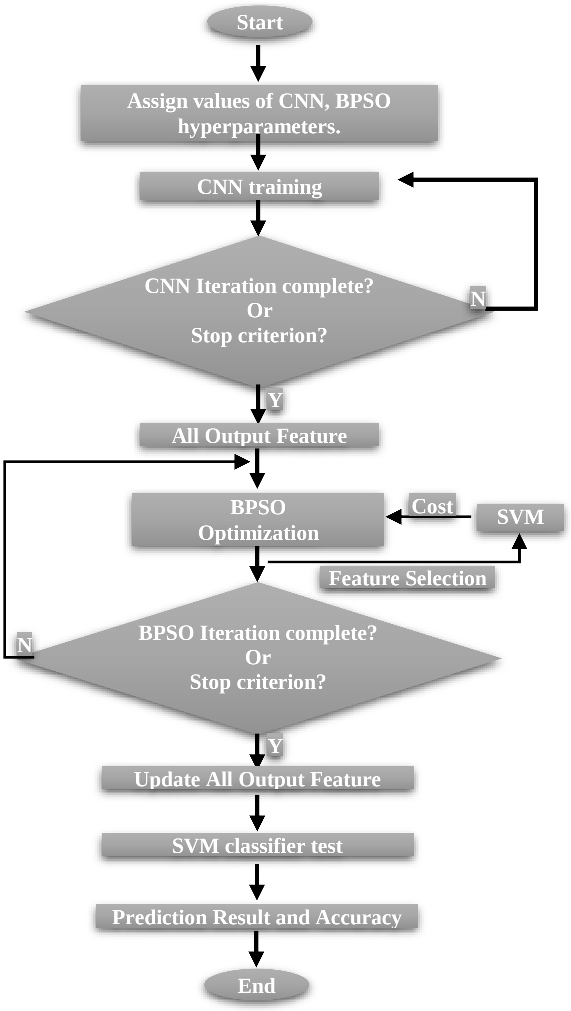

Figure 1. Flow chart of the proposed system

Table 1. Emotion expression datasets

Dataset
CASME [55]
CASME II [56]
SAMM [57]
CFD [58,59]
RAF-DB [60,61]
ADFES [62,63]
DEAP [64]
IEMOCAP [65,66]
JAFFE [67,68]
Google Facial Expression Comparison Dataset [69]
AffectNet [70,71]
Extended Cohn-Kanade Dataset (CK+) [72]
SFEW [73]
AFEW [74]
CMU MultiPIE [75]
DISFA [76]
GEMEP-FERA [77]
MMI [78]
Oulu-CASIA [79]
BAUM-1 s [80]
eNTERFACE'05 [81]
FER+ [52]
Fer2013 [54]

3.1. Fer2013 and Fer+ Database

The standard original FER-2013 dataset introduced at The International Conference on Machine Learning (ICML) was obtained through the Google image search API. There are 35887 grayscale face images in the dataset. Gray face images with 48x48 dimensions were labeled with 7 different emotion expressions (Neutral, happiness, surprise, sadness, anger, disgust, fear). Label accuracy in Fer2013 is lower than the FER+ dataset [54].

Unlike the FER-2013, in the FER+ dataset, which is an extension of the FER-2013 dataset, the emotion expressions corresponding to the images were labeled by 10 crowdsourced labelers instead of a single-label output. By choosing a single emotion for each image, a statistical emotion distribution was obtained for each facial expression. The FER+ dataset, unlike the original dataset, has 8 emotion labels including "contempt". The gold standard method was used for labeling quality [52]. FER+ dataset images were preprocessed and cropped. The FER+ dataset contains 35887 grayscale images, including training (28709), publicTest (3589), and privateTest (3589). Some facial expression datasets are shown in Table 1 below.

3.2. CNN architecture

FER+ dataset prepared with 10 crowdsourced labelers was used in the training of our CNN-based architecture. The images in the dataset are more reliable and have 10 different emotion labels, unlike the original FER-2013 dataset. These emotions are neutral, happiness, surprise, sadness, anger, disgust, fear, contempt, unknown, and NF (non-face). The emotional data distribution of the FER+ dataset is shown in Fig. 2.

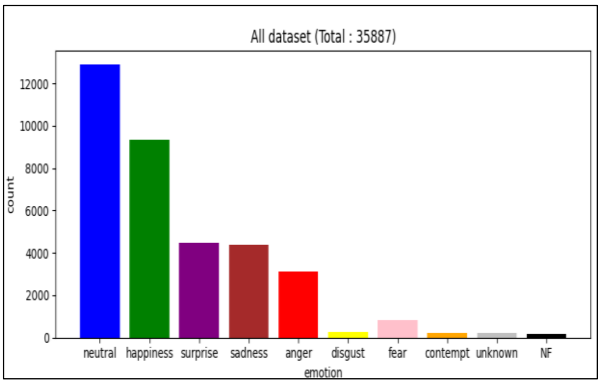

Figure 2. FER+ emotion tag data distribution

In order to train our CNN architecture more accurately in the data set, “unknown” and “NF” labeled images were removed from the data set. As a result of clearing the dataset from "unknown" and "NF" images, the total number of images in the dataset is 35488. For the training, validation and testing of the CNN architecture, the data set is divided into subsets of 80%, 10%, 10%, respectively. Before training, data augmentation was made in the training set. Dataset and emotion label data statistics Fig. 3 and Table 2, respectively.

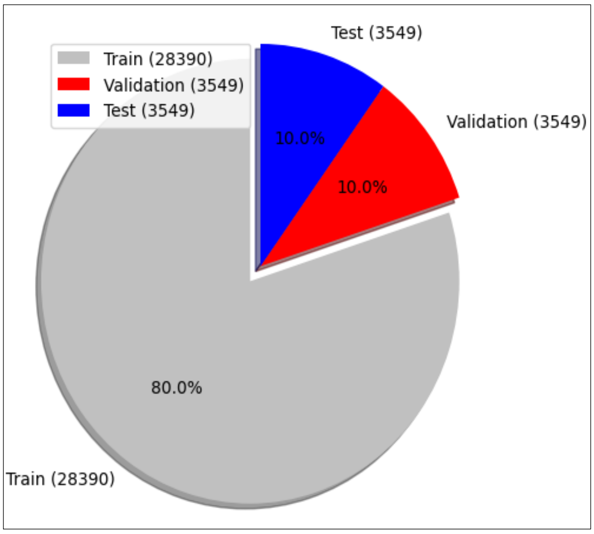

Figure 3. Training, validation and testing data

Table 2. Emotion expression data statistic

	Train	Valid	Test
Neu	10299	1305	1302
Hap	7519	933	903
Sur	3597	442	423
Sad	3453	451	467
Ang	2471	306	334
Dis	200	16	32
Fea	678	75	66
Con	173	21	22
Total	**28390**	**3549**	**3549**

In the CNN model of the proposed system, 6 convolutions were applied to the data set images with a spatial size of 48x48. The matrices obtained after the convolutional layers were flattened and transferred to 1 full link layer (128). 128 features in the fully connected layer are classified with the Softmax activation function in the classification layer with labels between 0-7. In the CNN model, max pooling was applied 3 times. Dropout is applied before the classification layer. ReLU activation function and batch normalization were used in all layers except the classification layer. The CNN model used is shown in Fig. 4.

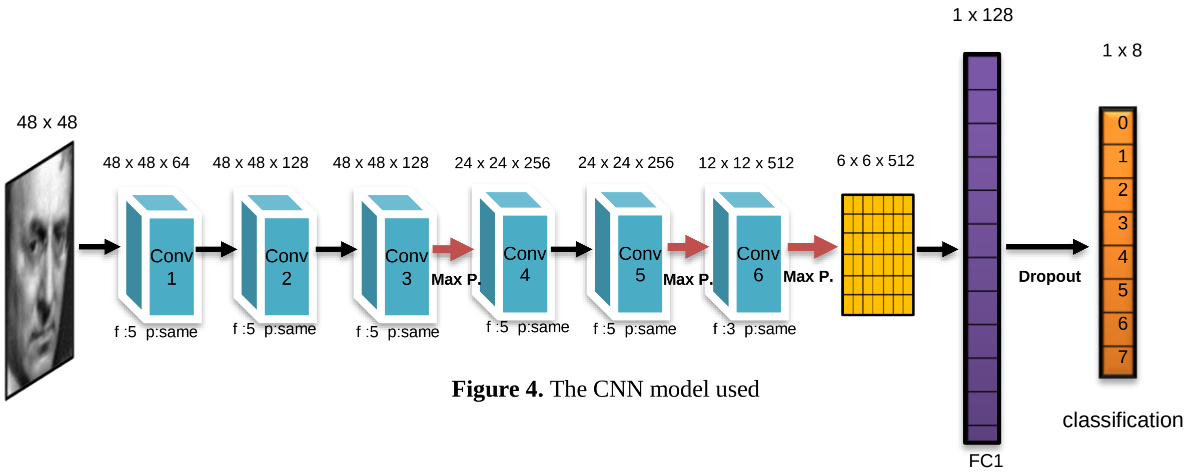

Figure 4. The CNN model used

Hyper parameters such as Max pooling, Dropout, ReLU and Batch normalization, Data augmentation, Optimization algorithm used in the CNN architecture are important parameters for the correct training and generalization ability of the network. Max pooling will not change the number of channels of the post-convolutional features, but will halve their spatial size due to the 2x2 filter applied. This way, dominant features in the image are selected. Max pooling offers ease of calculation as it reduces the spatial dimension. In the dropout stage, some of the neurons carrying the feature information are randomly discarded. So, over-learning is prevented by increasing the generalization ability in the training of the network. ReLU activation function is used to learn nonlinear situations from complex information. If the input value of this function is negative, the output is zero. If the input value is positive, the input value is given as the output. Batch normalization is used to reduce the effect of gradient losses. Data augmentation has been applied to enable the network to learn better. In order to minimize the error in the training of the network, one of the gradient descent based optimization algorithms "Adam" [82] was preferred. Binary Cross Entropy (Loss function) is used as error function.

3.3. PSO and BPSO

Heuristic optimization algorithms offer solutions close to nonlinear mathematics and real life problems. In the literature, heuristic optimization algorithms have been applied in many different areas [83-87]. Particle Swarm Optimization (PSO), one of these algorithms, is a heuristic algorithm put forward by Kennedy and Eberhart in 1995 [88]. This algorithm is used in the optimization of nonlinear problems. The inspiration for this algorithm is swarm of birds. Bird swarm act together and in communication for their vital needs. The individual in the best position to meet the need is the leader of the swarm. Birds in constant motion change their position according to the position of the swarm leader and their best-ever position. In this way, they try to reach the best goal that will meet the need. In solving non-linear problems, just like in bird swarm, a swarm of particles is created to obtain the best solution. Each particle in the swarm represents a solution. The particle that gives the best solution in the search space is the global best particle (GBest). The individual solutions of the particles that give the best solution represent the individual best particle (PBest). Particles update their velocities at each position change, taking the GBest and PBest solutions into account. With their new speed, they get a new position. This situation continues until the best solution for the problem is found [84,88,89]. Velocity and position update are given in Eq. (1) and equation Eq. (2), respectively. The particle's position update is shown in Fig. 5.

$$vn[t+1] = w[t]vn[t] + c1r1(xL,n[t] - xn[t]) + c2r2(xG,n[t] - xn[t]) \quad (1)$$

$$xn[t+1] = xn[t] + vn[t+1] \quad (2)$$

vn[t+1] :The new velocity of the particle

w[t] :Inertia coefficient

vn[t] :The previous velocity of the particle

c1 :Coefficient of remembering own best position

r1 :Random number from [0-1]

xL,n[t] :The best position of the particle so far

xn[t] :The previous position of the particle

xG,n[t] :Swarm leader's position

r2 :Random number from [0-1]

c2 :Coefficient of remembering the position of the swarm leader

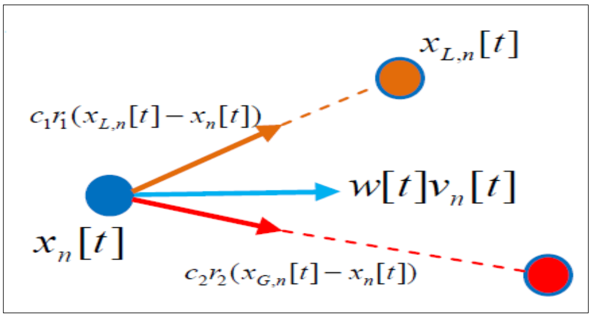

Figure 5. Particle location update [84]

Binary Particle Swarm Optimization (BPSO), a functional algorithm, is used in problems such as size reduction and feature selection. In BPSO, unlike PSO, candidate particles for solution are denoted by "0" and "1". After the velocity update of the particles in the binary search space, the velocities determined in determining the new positions of the particles are converted into probability values between 0-1 by passing a Sigmoidal function. Then, the velocity values obtained are compared with a random number and the new position of the particle is updated as "0" or "1". Equations where the position is calculated in BPSO are given in Eq. (3) and Eq. (4) [89,90].

$$Sig(vn[t+1]) = \frac{1}{1+e^{-(vn[t+1])}} \quad (3)$$

$$xn[t+1] = \begin{cases} 0, if\ rand() \geq Sig(vn[t+1]) \\ 1, if\ rand() < Sig\ (vn[t+1]) \end{cases} \quad (4)$$

4. RESULTS AND DISCUSSIONS

Training was completed in 49 iterations by applying an early stop to the CNN model to prevent overfitting. The CNN model resulted in 90.63% in training and 84.19% in validation. The model performance of the model in the test set was measured as 84.28%. The accuracy and loss graphs of the CNN model are shown in Fig. 6 and Fig. 7, respectively. CNN training and testing details are shown in Table 3.

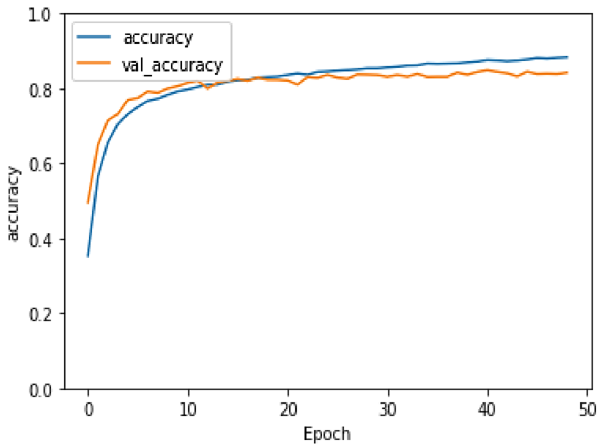

Figure 6. Model accuracy graph

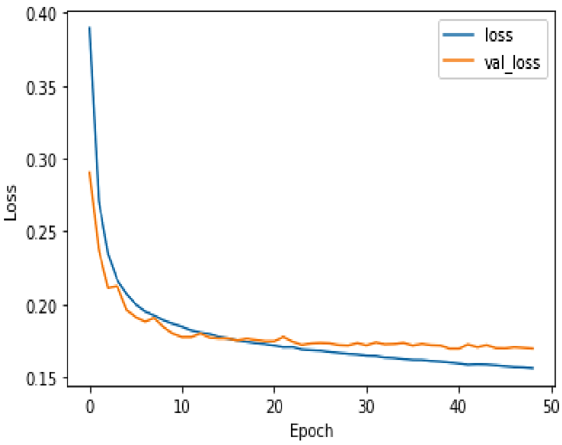

Figure 7. Model loss graph

Table 3. CNN training and testing details

Dataset	Accuracy
Train	90.63%
Validation	84.19%
Test (data augmentation) (*lr=0.0001*)	**84.28%**
Test (without data aug.) (*lr=0.0001*)	79.91%
Test (data augmentation) (*lr=0.001*)	79.49%

Confusion matrix, which is a reliable measure, was used to measure the classification performance of our CNN model. Fig. 8 shows the confusion matrix of the test dataset. When the confusion matrix given in Fig. 8 is examined, the classification result for "happiness" is 91.58, "contempt" is 31.82%, and "disgust" is 40.62%.

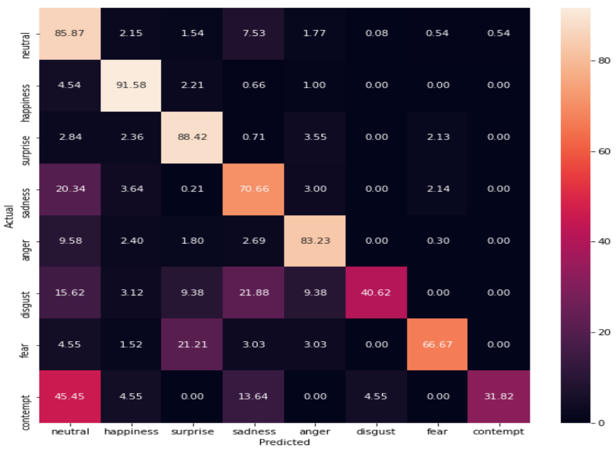

Figure 8. Test dataset confusion matrix

This is due to the fact that the "happiness" data has the highest data, and the "contempt" and "disgust" data have the least data. Another thing to note is that 45.45% of the data labeled "contempt" is classified as "neutral". Classification of “contempt” data as “contempt” has a rate of 31.82%. The reason for this is that the facial expressions in the images with the labels "contempt" and "neutral" are very similar.

4.1. BPSO+SVM

At this stage, the feature vectors of all training images belonging to the FC1 layer of the CNN-based architecture, whose training has been completed, are transmitted to the BPSO unit. The BPSO algorithm generates a population of candidate particles to find the most valuable feature vector. Each particle is a vector consisting of the values "0" and "1". These particles are sequentially applied to the CNN feature vector. The index values of the particle with the value of "1" are found and the features belonging to these index values in the feature vector are filtered. The resulting new feature vector is transmitted to the SVM [91] classifier. An error value for all particles is obtained with the RBF kernel method of the SVM classifier [92]. This process continues until the best particle is found. The CNN test performance of the proposed system is 84.28%. Test success of the CNN feature vector with SVM is 84.81%. As a result of applying the BPSO + SVM combination to the FC1 layer, the most valuable features were extracted from 128 features of each image. Thus, the number of features was reduced from 128 to 59. As a result of testing with 59 features, the accuracy obtained was increased by 0.93%, reaching 85.74% accuracy. The working procedure of the BPSO+SVM algorithm applied to the FC1 layer and the BPSO error graph with 10 iterations and 35 particles are shown in Fig. 9 and Fig. 10, respectively.

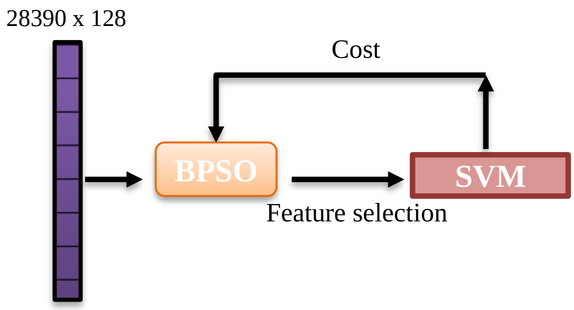

Figure 9. BPSO+SVM algorithm working

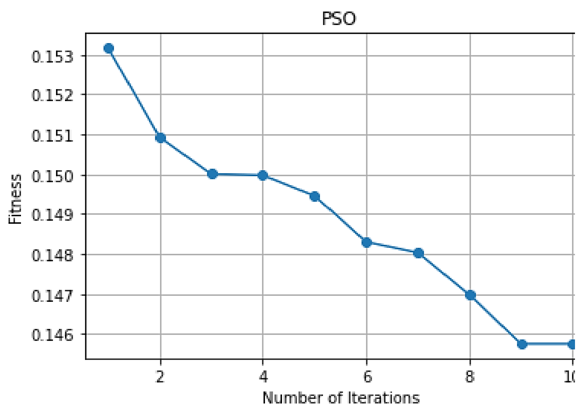

Fig. 10. BPSO error graph

While the classification rate per CNN+SVM image was 1.82 ms, the classification rate of CNN+BPSO+SVM was reduced to 1.12 ms. This situation is shown in Table 4. In addition, the literature comparison of the accuracy of the proposed model is given in Table 5.

Table 4. CNN+BPSO+SVM vs CNN+SVM

Methods	Acc.	Working time (per image)
Our Proposed *(CNN+BPSO+SVM)*	85.74%	1.12 ms
(CNN+SVM)	84.81%	1.82 ms
(CNN)	84.28%	0.52 ms

Table 5. Literature Comparison of Accuracy

Methods	Accuracy
PSR [93]	89.56 ± 0.03%
SENet Teacher [94]	88.80 ± 0.3%
CNNs and BOVW + local SVM [29]	87.76%
ResNet + VGG Late Fusion (Linear SVM) [53]	87.40%
ESR-9 [95]	87.15 ± 0.1%
CNNs and BOVW + global SVM [29]	86.96%
VGG (Linear SVM) [53]	86.90%
SHCNN [39]	86.54%
ResNet (Linear SVM) [53]	86.20%
fine-tuned VGG-f [29]	86.01%
Our Proposed (CNNs+BPSO+SVM)	85.74%
VGG16-PLD [52]	84.99 ± 0.37%
fine-tuned VGG-face [29]	84.79%
VGG16-CEL [52]	84.72 ± 0.24%
VGG-13 [29]	84.41%
TFE-JL [33]	84.30%
VGG16-ML [52]	83.97 ± 0.36%
VGG16-MV [52]	83.85 ± 0.63%
ResNet18 + FC [33]	83.40%
ResNet18 [33]	83.10%
Lian et al. [44]	81.93%
pre-trained VGG-face [29]	81.73%
BOVW [29]	80.65%

When Table 5 is examined, although the accuracy of the proposed system is behind the latest technology systems, it has achieved a better result than some studies with architectures such as VGG and ResNet. This study does not claim the best accuracy rate with the fer+ dataset. The main point of the study is to draw attention to the contribution of feature selection to the accuracy and speed of emotion classification. While the accuracy of the classification test with CNN was 84.28%, the accuracy of the proposed system was measured as 85.74%. The proposed system contributes about 1.5% to the classification with CNN.

5. CONCLUSIONS

In this paper, a system is proposed to increase the accuracy and speed of emotion expression classification. The proposed system consists of CNN, BPSO and SVM. With the CNN model we created, 128 features representing the images of the FER+ dataset were extracted. By using these features, classification was performed with SVM. Classification with SVM achieves better accuracy than the CNN model. To further increase the accuracy, the BPSO heuristic optimization method is used. Using BPSO, the features with the best classification accuracy were selected from 128 features representing the images. At the end of the BPSO, the number of features was reduced to approximately half the size. It was reclassified by the SVM method with the features obtained with BPSO. As a result of the classification made with the combination of BPSO + SVM, the accuracy and speed values gave better results than the classification made with SVM alone. In this paper, the contribution of the proposed system to the classification accuracy and speed of the FER+ dataset, which is one of the emotion expression datasets, is revealed.

ACKNOWLEDGEMENT

This study was supported by Inonu University Scientific Research Projects Coordination Unit (BAP) with the project coded FDK-2020-2110.

DECLARATION OF ETHICAL STANDARDS

The author(s) of this article declare that the materials and methods used in this study do not require ethical committee permission and/or legal-special permission.

AUTHORS' CONTRIBUTIONS

Kenan Donuk: Performed the experiments and wrote the manuscript.
Ali ARI: Analysed the results.
Mehmet Fatih ÖZDEMİR: Analysed the results.
Davut Hanbay: Analysed the results.

CONFLICT OF INTEREST

There is no conflict of interest in this study.

REFERENCES

[1] Bouhlal M., Aarika K., AitAbdelouahid R., Elfilali S. and Benlahmar E., "Emotions recognition as innovative tool for improving students' performance and learning approaches", ***Procedia Computer Science***, 175: 597-602, (2020).

[2] Simcock G., McLoughlin L. T., De Regt T., Broadhouse K. M., Beaudequin D., Lagopoulos J. and Hermens D. F., "Associations between facial emotion recognition and mental health in early adolescence", ***International Journal of Environmental Research and Public Health***, 17(1), (2020).

[3] Bouzakraoui M. S., Sadiq A. and Alaoui A. Y., "Appreciation of Customer Satisfaction Through Analysis Facial Expressions and Emotions Recognition", ***Proceedings of 2019 IEEE World Conference on Complex Systems, WCCS 2019***, 1-5, (2019).

[4] Owayjan M., Kashour A., Al Haddad N., Fadel M. and Al Souki G., "The design and development of a lie detection system using facial micro-expressions", ***2012 2nd International Conference on Advances in Computational Tools for Engineering Applications, ACTEA 2012,*** 33-38, (2012).

[5] Zloteanu M., "Reconsidering Facial Expressions and Deception Detection", In Handbook of Facial Expression of Emotion, 3: 238-284, **FEELab Science Books & Leya**, (2020).

[6] Praditsangthong R., Slakkham B. and Bhattarakosol P., "A fear detection method based on palpebral fissure", ***Journal of King Saud University - Computer and Information Sciences***, (2019).

[7] Harms M. B., Martin A. and Wallace G. L., "Facial emotion recognition in autism spectrum disorders: A review of behavioral and neuroimaging studies", ***In Neuropsychology Review,*** 20(3): 290-322, (2010).

[8] Ekman P. and Friesen W., "Facial action coding system: a technique for the measurement of facial movement", (1978).

[9] Fan Y., Lam J. C. K. and Li V. O. K., "Demographic effects on facial emotion expression: an interdisciplinary investigation of the facial action units of happiness", ***Scientific Reports***, 11(1): 5214, (2021).

[10] Ma J., Li X., Ren Y., Yang R. and Zhao Q., "Landmark-Based Facial Feature Construction and Action Unit Intensity Prediction", ***Mathematical Problems in Engineering***, (2021).

[11] Ko H., Kim K., Bae M., Seo M.-G., Nam G., Park S., Park S., Ihm J. and Lee J.-Y., "Changes in facial recognition and facial expressions with age", (2021).

[12] Taubert J. and Japee S., "Using FACS to trace the neural specializations underlying the recognition of facial expressions: A commentary on Waller et al. (2020)", ***Neuroscience and Biobehavioral Reviews,*** 120: 75–77, (2021).

[13] Nadeeshani M., Jayaweera A. and Samarasinghe P., "Facial emotion prediction through action units and deep learning", ***ICAC 2020 - 2nd International Conference on Advancements in Computing, Proceedings,*** 293-298, (2020).

[14] Sari M., Moussaoui A. and Hadid A., "Automated Facial Expression Recognition Using Deep Learning Techniques: An Overview", ***In International Journal of Informatics and Applied Mathematics***, 3(1): 39-53, (2020).

[15] Murugappan M. and Mutawa A., "Facial geometric feature extraction based emotional expression classification using machine learning algorithms", ***PloS One***, 16(2), (2021).

[16] S.Bavkar S., S. Rangole J. and U. Deshmukh V., "Geometric Approach for Human Emotion Recognition using Facial Expression", ***International Journal of Computer Applications,*** 118(14): 17-22, (2015).

[17] Ghimire D. and Lee J., "Geometric feature-based facial expression recognition in image sequences using multi-class AdaBoost and support vector machines", ***Sensors (Switzerland)***, 13(6): 7714-7734, (2013).

[18] Perez-Gomez V., Rios-Figueroa H. V., Rechy-Ramirez E. J., Mezura-Montes E. and Marin-Hernandez A., "Feature selection on 2d and 3d geometric features to improve facial expression recognition", ***Sensors (Switzerland),*** 20(17), 1–20, (2020).

[19] Ounachad K., Oualla M. and Sadiq A., "Geometric feature based facial emotion recognition", ***International Journal of Advanced Trends in Computer Science and Engineering***, 9(3):3417-3425, (2020).

[20] Liu X., Cheng X. and Lee K., "GA-SVM based Facial Emotion Recognition using Facial Geometric Features", ***IEEE Sensors Journal***, 1-1, (2020).

[21] Chouhayebi H., Riffi J., Mahraz M. A., Yahyaouy A., Tairi H. and Alioua N., "Facial expression recognition based on geometric features", ***2020 International Conference on Intelligent Systems and Computer Vision, ISCV 2020,*** 1-6, (2020).

[22] Ravi R., Yadhukrishna S. V. and Prithviraj R., "A Face Expression Recognition Using CNN LBP", ***Proceedings of the 4th International Conference on Computing Methodologies and Communication, ICCMC 2020,*** 684–689, (2020).

[23] Niu B., Gao Z. and Guo B., "Facial Expression Recognition with LBP and ORB Features", ***Computational Intelligence and Neuroscience***, (2021).

[24] Lakshmi D. and Ponnusamy R., "Facial emotion recognition using modified HOG and LBP features with deep stacked autoencoders", ***Microprocessors and Microsystems***, 82: 103834, (2021).

[25] Jumani S. Z., Ali F., Guriro S., Kandhro I. A., Khan A. and Zaidi A., "Facial Expression Recognition with Histogram of Oriented Gradients using CNN", ***Indian Journal of Science and Technology,*** 12(24): 1-8, (2019).

[26] Verma K. and Khunteta A., "Facial expression recognition using Gabor filter and multi-layer artificial neural network", ***IEEE International Conference on Information, Communication, Instrumentation and Control, ICICIC 2017,*** 1-5, (2018).

[27] Mehta N. and Jadhav S., "Facial emotion recognition using log gabor filter and PCA", ***Proceedings - 2nd International Conference on Computing, Communication, Control and Automation, ICCUBEA 2016,*** 1-5, (2017).

[28] Borui Z., Liu G. and Xie G., "Facial expression recognition using LBP and LPQ based on Gabor wavelet transform", ***2016 2nd IEEE International Conference on Computer and Communications, ICCC 2016 - Proceedings,*** 365–369, (2017).

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

[30] Shi Y., Lv Z., Bi N. and Zhang C., "An improved SIFT algorithm for robust emotion recognition under various face poses and illuminations", ***Neural Computing and Applications***, 32(13): 9267-9281, (2020).

[31] Hinton G. E. and Salakhutdinov R. R., "Reducing the dimensionality of data with neural networks", ***Science,*** 313(5786):504-507, (2006).

[32] LeCun Y., Bottou L., Bengio Y. and Haffner P., "Gradient-based learning applied to document recognition", ***Proceedings of the IEEE***, 86(11): 2278-2323, (1998).

[33] Li M., Xu H., Huang X., Song Z., Liu X. and Li X., "Facial Expression Recognition with Identity and Emotion Joint Learning", ***IEEE Transactions on Affective Computing,*** (2018).

[34] Ari A. and Hanbay D., "Deep learning based brain tumor classification and detection system", ***Turkish Journal of Electrical Engineering and Computer Science,*** 26(5): 2275-2286, (2018).

[35] Türkoğlu M. and Hanbay D., "Plant disease and pest detection using deep learning-based features", ***Turkish Journal of Electrical Engineering and Computer Science***, 27(3): 1636-1651, (2019).

[36] Uzen H., Turkoglu M. and Hanbay D., "Texture defect classification with multiple pooling and filter ensemble based on deep neural network", ***Expert Systems with Applications,*** 175: 114838, (2021).

[37] Liu S., Li D., Gao Q. and Song Y., "Facial Emotion Recognition Based on CNN", ***Proceedings - 2020 Chinese Automation Congress, CAC 2020,*** 398-403, (2020).

[38] Chirra V. R. R., Uyyala S. R. and Kolli V. K. K., "Virtual facial expression recognition using deep CNN with ensemble learning", ***Journal of Ambient Intelligence and Humanized Computing***, 1: 3, (2021).

[39] Miao S., Xu H., Han Z. and Zhu Y., "Recognizing facial expressions using a shallow convolutional neural network", ***IEEE Access***, 7: 78000-78011, (2019).

[40] Gupta R. and Vishwamitra L. K., "Facial expression recognition from videos using CNN and feature aggregation", ***Materials Today: Proceedings***, (2021).

[41] Bhandari A. and Pal N. R., "Can edges help convolution neural networks in emotion recognition?", ***Neurocomputing,*** 433: 162-168, (2021).

[42] Liang D., Liang H., Yu Z. and Zhang Y., "Deep convolutional BiLSTM fusion network for facial expression recognition", ***Visual Computer,*** 36(3): 499-508, (2020).

[43] Christou N. and Kanojiya N., "Human facial expression recognition with convolution neural networks", ***In Advances in Intelligent Systems and Computing***, 797: 539-545, (2019).

[44] Lian Z., Li Y., Tao J., Huang J. and Niu M., "Region Based Robust Facial Expression Analysis", ***2018 1st Asian Conference on Affective Computing and Intelligent Interaction, ACII Asia 2018***, (2018).

[45] Mollahosseini A., Chan D. and Mahoor M. H., "Going deeper in facial expression recognition using deep neural networks", ***2016 IEEE Winter Conference on Applications of Computer Vision, WACV 2016***, 1-10, (2016).

[46] Lv Y., Feng Z. and Xu C., "Facial expression recognition via deep learning", ***Proceedings of 2014 International Conference on Smart Computing, SMARTCOMP 2014,*** 303-308, (2014).

[47] Josephine Julina J. K. and Sharmila T. S., "Facial Emotion Recognition in Videos using HOG and LBP", ***2019 4th IEEE International Conference on Recent Trends on Electronics, Information, Communication and Technology, RTEICT 2019 - Proceedings***, 56-60, (2019).

[48] Li B. and Lima D., "Facial expression recognition via ResNet-50", ***International Journal of Cognitive Computing in Engineering***, 2: 57-64, (2021).

[49] Bargal S. A., Barsoum E., Ferrer C. C. and Zhang C., "Emotion Recognition in the Wild from Videos using Images*",* (2016).

[50] Simonyan K. and Zisserman A., "Very deep convolutional networks for large-scale image recognition", ***3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings,*** (2015).

[51] He K., Zhang X., Ren S. and Sun J., "Deep residual learning for image recognition", ***Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition***, 770-778, (2016).

[52] Barsoum E., Zhang C., Ferrer C. C. and Zhang Z., "Training Deep Networks for Facial Expression Recognition with Crowd-Sourced Label Distribution", ***ICMI 2016 - Proceedings of the 18th ACM International Conference on Multimodal Interaction,*** 279-283, (2016).

[53] Huang C., "Combining convolutional neural networks for emotion recognition", ***2017 IEEE MIT Undergraduate Research Technology Conference, URTC 2017,*** 1-4, (2018).

[54] Goodfellow I. J., Erhan D., Luc Carrier P., Courville A., Mirza M., Hamner B., Cukierski W., Tang Y., Thaler D., Lee D. H., Zhou Y., Ramaiah C., Feng F., Li R., Wang X., Athanasakis D., Shawe-Taylor J., Milakov M., Park J. and et al., "Challenges in representation learning: A report on three machine learning contests", ***Neural Networks,*** 64: 59-63, (2015).

[55] Yan W. J., Wu Q., Liu Y. J., Wang S. J. and Fu X., "CASME database: A dataset of spontaneous micro-expressions collected from neutralized faces", ***2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition, FG 2013,*** (2013).

[56] Yan W. J., Li X., Wang S. J., Zhao G., Liu Y. J., Chen Y. H. and Fu X., "CASME II: An improved spontaneous micro-expression database and the baseline evaluation", ***PLoS ONE,*** 9(1): e86041, (2014).

[57] Davison A. K., Lansley C., Costen N., Tan K. and Yap M. H., "SAMM: A Spontaneous Micro-Facial Movement Dataset", ***IEEE Transactions on Affective Computing,*** 9(1): 116-129, (2018).

[58] Ma D. S., Correll J. and Wittenbrink B., "The Chicago face database: A free stimulus set of faces and norming data", ***Behavior Research Methods***, 47(4): 1122-1135, (2015).

[59] "Chicago Face Database." Accessed May 19, 2021. https://chicagofaces.org/default/

[60] Li S., Deng W. and Du J. P., "Reliable crowdsourcing and deep locality-preserving learning for expression recognition in the wild", ***Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017,*** 2584-2593, (2017).

[61] Li S. and Deng W., "Reliable crowdsourcing and deep locality-preserving learning for unconstrained facial expression recognition", ***IEEE Transactions on Image Processing***, 28(1): 356-370, (2019).

[62] van der Schalk J., Hawk S. T., Fischer A. H. and Doosje B., "Moving Faces, Looking Places: Validation of the Amsterdam Dynamic Facial Expression Set (ADFES)", ***Emotion***, 11(4): 907-920, (2011).

[63] "Introduction - Amsterdam Interdisciplinary Centre for Emotion (AICE) - University of Amsterdam." Accessed May 19, 2021. https://aice.uva.nl/research-tools/adfes-stimulus-set/adfes-stimulus-set.html?cb.

[64] Koelstra S., Mühl C., Soleymani M., Lee J. S., Yazdani A., Ebrahimi T., Pun T., Nijholt A. and Patras I., "DEAP: A database for emotion analysis; Using physiological signals", ***IEEE Transactions on Affective Computing***, 3(1): 18-31, (2012).

[65] Busso C., Bulut M., Lee C.C., Kazemzadeh A., Mower E., Kim S., Chang J.N., Lee S. and Narayanan S. S., "IEMOCAP: Interactive emotional dyadic motion capture database", ***Journal of Language Resources and Evaluation***, 42(4): 335-359, (2008).

[66] "IEMOCAP" Accessed May 19, 2021. https://sail.usc.edu/iemocap/iemocap_release.htm

[67] Lyons M., Kamachi M. and Gyoba J., "The Japanese Female Facial Expression (JAFFE) Dataset", (1998).

[68] "(JAFFE) Dataset | Zenodo." Accessed May 19, 2021.https://zenodo.org/record/3451524#.YKTd3bczZp8.

[69] "Google Facial Expression Comparison Dataset – Google Research." Accessed May 19, 2021. https://research.google/tools/datasets/google-facial-expression/.

[70] Mollahosseini A., Hasani B. and Mahoor M. H., "AffectNet: A New Database for Facial Expression, Valence, and Arousal Computation in the Wild", ***IEEE Transactions on Affective Computing***, (2017).

[71] "AffectNet – Mohammad H. Mahoor, PhD." Accessed May 19, 2021. http://mohammadmahoor.com/affectnet/.

[72] Lucey P., Cohn J. F., Kanade T., Saragih J., Ambadar Z. and Matthews I., "The extended Cohn-Kanade dataset (CK+): A complete dataset for action unit and emotion-specified expression", ***2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops, CVPRW 2010***, 94-101, (2010).

[73] Dhall A., Goecke R., Lucey S. and Gedeon T., "Static facial expression analysis in tough conditions: Data, evaluation protocol and benchmark", ***Proceedings of the IEEE International Conference on Computer Vision***, 2106-2112, (2011).

[74] Dhall A., Goecke R., Lucey S. and Gedeon T., "Collecting large, richly annotated facial-expression databases from movies", ***IEEE Multimedia,*** 19(3): 34-41, (2012).

[75] Gross R., Matthews I., Cohn J., Kanade T. and Baker S., "Multi-PIE", ***Image and Vision Computing,*** 28(5): 807-813, (2010).

[76] Mavadati S. M., Mahoor M. H., Bartlett K., Trinh P. and Cohn J. F., "DISFA: A spontaneous facial action intensity database", ***IEEE Transactions on Affective Computing***, 4(2): 151-160, (2013).

[77] Bänziger T. and Scherer K. R., "Introducing the Geneva Multimodal Emotion Portrayal (GEMEP) Corpus", ***In Blueprint for affective computing: A sourcebook,*** 271–294, (2010).

[78] Pantic M., Valstar M., Rademaker R. and Maat L., "Web-based database for facial expression analysis", ***IEEE International Conference on Multimedia and Expo, ICME 2005***, 317–321, (2005).

[79] Zhao G., Huang X., Taini M., Li S. Z. and Pietikäinen M., "Facial expression recognition from near-infrared videos", ***Image and Vision Computing***, 29(9): 607-619, (2011).

[80] Zhalehpour S., Onder O., Akhtar Z. and Erdem C. E., "BAUM-1: A Spontaneous Audio-Visual Face Database of Affective and Mental States", ***IEEE Transactions on Affective Computing***, 8(3): 300–313, (2017).

[81] Martin O., Kotsia I., Macq B. and Pitas I., "The eNTERFACE'05 Audio-Visual emotion database", ***ICDEW 2006 - Proceedings of the 22nd International Conference on Data Engineering Workshops***, (2006).

[82] Kingma D. P. and Lei Ba J., "ADAM: A method for stochastic optimization", (2015).

[83] Fırat H. and Alpaslan N., "An effective approach to the two-dimensional rectangular packing problem in the manufacturing industry", ***Computers and Industrial Engineering***, 148:106687, (2020).

[84] Donuk K., Özbey N., Inan M., Yeroğlu C. and Hanbay D., "Investigation of PIDA Controller Parameters via PSO Algorithm", ***2018 International Conference on Artificial Intelligence and Data Processing, IDAP 2018***, (2019).

[85] Wulandhari L. A., Komsiyah S. and Wicaksono W., "Bat Algorithm Implementation on Economic Dispatch Optimization Problem", ***Procedia Computer Science***, 135: 275-282, (2018).

[86] Sarkar R., Barman D. and Chowdhury N., "Domain knowledge based genetic algorithms for mobile robot path planning having single and multiple targets", ***Journal of King Saud University - Computer and Information Sciences***, (2020).

[87] Sharma V. and Mir R. N., "An enhanced time efficient technique for image watermarking using ant colony optimization and light gradient boosting algorithm", ***Journal of King Saud University - Computer and Information Sciences***, (2019).

[88] Kennedy J. and Eberhart R., "Particle swarm optimization", ***Proceedings of ICNN'95 - International Conference on Neural Networks***, 4: 1942-1948, (1995).

[89] Kumari K., Singh J. P., Dwivedi Y. K. and Rana N. P., "Multi-modal aggression identification using Convolutional Neural Network and Binary Particle Swarm Optimization", ***Future Generation Computer Systems***, 118: 187-197, (2021).

[90] Dagar N. S. and Dahiya P. K., "Edge Detection Technique using Binary Particle Swarm Optimization", ***Procedia Computer Science***, 167: 1421-1436, (2020).

[91] Cortes C. and Vapnik V., "Support-vector networks", ***Machine Learning***, 20(3): 273–297, (1995).

[92] "Scikit-Learn 0.24.2 Dokümantasyon." Accessed May 21, 2021. https://scikit-learn.org/stable/auto_examples/svm/plot_rbf_parameters.html.

[93] Vo T. H., Lee G. S., Yang H. J. and Kim S. H., "Pyramid with Super Resolution for In-the-Wild Facial Expression Recognition", ***IEEE Access***, 8: 131988-132001, (2020).

[94] Albanie S., Nagrani A., Vedaldi A. and Zisserman A., "Emotion recognition in speech using cross-modal transfer in the wild", ***MM 2018 - Proceedings of the 2018 ACM Multimedia Conference***, 292-301, (2018).

[95] Siqueira H., Magg S. and Wermter S., "Efficient facial feature learning with wide ensemble-based convolutional neural networks", ***AAAI 2020 - 34th AAAI Conference on Artificial Intelligence***, 5800-5809, (2020).