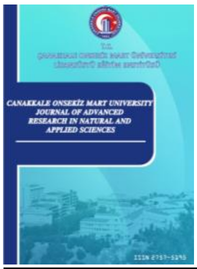
e-ISSN 2757-5195



Çanakkale Onsekiz Mart University
Journal of Advanced Research in Natural and Applied Sciences

Open Access

doi.org/10.28979/jarnas.1056664 2022, Vol. 8, Issue 4, Pages: 736-752 dergipark.org.tr/tr/pub/jarnas

Sentiment Analysis from Face Expressions Based on Image Processing Using Deep Learning Methods

Orhan Emre Aksoy[1], Selda Güney[2*]

[1-2]Department of Electrical & Electronics Engineering, Faculty of Engineering, Başkent University, Ankara, Turkey

Article History
Received: 12.01.2022
Accepted: 13.08.2022
Published: 15.12.2022
Research Article

Abstract − In this study, the classification study of human facial expressions in real-time images is discussed. Implementing this work in software have some benefits for us. For example, analysis of mood in group photos is an interesting instance in this regard. The perception of people's facial expressions in photographs taken during an event can provide quantitative data on how much fun these people have in general. Another example is context-aware image access, where only photos of people who are surprised can be accessed from a database. Seven different emotions related to facial expressions were classified in this context; these are listed as happiness, sadness, surprise, disgust, anger, fear and neutral. With the application written in Python programming language, classical machine learning methods such as k-Nearest Neighborhood and Support Vector Machines and deep learning methods such as AlexNet, ResNet, DenseNet, Inception architectures were applied to FER2013, JAFFE and CK+ datasets. In this study, while comparing classical machine learning methods and deep learning architectures, real-time and non-real-time applications were also compared with two different applications. This study conducted to demonstrate that real-time expression recognition systems based on deep learning techniques with the most appropriate architecture can be implemented with high accuracy via computer hardware with only one software. In addition, it is shown that high accuracy rate is achieved in real-time applications when Histograms of Oriented Gradients (HOG) is used as a feature extraction method and ResNet architecture is used for classification.

Keywords − *Classification, convolutional neural network, deep learning, emotion analysis, image processing*

1. Introduction

One of the biggest problems encountered within the scope of face based, computer-generated image applications are the variety of expressions in facial images (Yang & Kriegman, 2002). Although many system suggestions have been put forward, regarding the resistance to facial expressions, the problem could not be resolved (Frank, 2019). The face, which is easy to access and has a high distinctiveness, has been the subject of researchers for many years. Facial diagnosis and recognition studies have contributed not only to the development of security surveillance systems, but also to the development of neurology and psychology sciences (Yang & Kriegman, 2002). Negative aspects of the examination of facial images are mainly related to head posture, lighting conditions and facial expressions. Today, accessing and processing three-dimensional (3D) facial information has become extremely easy. This situation contributed significantly to the solution due to the pose and illumination, which does not change due to the nature of the data. Similar developments are also observed in facial expression

[1] orhanemreaksoy@gmail.com
[2] seldaguney@baskent.edu.tr
*Corresponding Author

recognition (Göngör & Tutsoy, 2018; Tutsoy et al., 2017; Zhao et al., 2015; Kuo et al., 2018; Jung et al., 2015; Abdullah, et al., 2021). Göngör & Tutsoy proposed a facial emotion recognition system to use in humanoid robot. They used eigenface method to extract the features. Then these features were classified by using Artificial Neural Network (ANN). They achieved promising results as a result of this study (Göngör & Tutsoy, 2018) . The results appear to be open to improvement. With the widespread use of deep learning, Zhao et al. analysed a deep learning method for facial emotion recognition in 2015 (Zhao et al., 2015). But they didn't examine the real time applications in facial emotion recognition. Kuo et. al. analysed deep learning methods both real time and non-real time applications. The results showed that the accuracy rate decrease for the real time applications (Kuo et al., 2018).

The recognition of facial expressions is used in wide area such as, the fields of security and visual surveillance and security, medical diagnosis, emotional research, law enforcement, and education (Li et al., 2020).

Facial expressions that provide important clues about people's emotional states, cognitive activities, and interest orientations; is also among the main factors affecting face-to-face communication. This makes machine-human interaction as important as expressions. Classification of facial expressions is also very important in terms of precautions that can be taken against these changes (Li & Liam, 2021). If the subject is handled in terms of security applications, it will be possible for face detection and recognition systems to give higher accuracy results by classifying two-dimensional (2D) face images according to the expressions they contain (Bhattacharya, 2021).

Facial expressions can be considered as concrete indications of individuals' feelings, understanding, characters and psychological states, and they are also an important tool in people's relations with each other. Mehrabian states that listeners are significantly influenced by their facial expressions (Mehrabian, 2016). Darwin argued that humans and animals contain some innate emotions in their nature and that these emotions are reflected as facial expressions. After this first study on facial expression analysis, it was revealed by Ekman and Friesen in 1971 that there are 6 basic emotions and each of them is reflected with different facial expressions; these are listed as anger, disgust, fear, happiness, sadness and surprise (Karaboyacı, 2009).

The use of still images and sequences of images has enabled facial expression analysis to cease to be a research topic that remains only within the boundaries of psychology. Developments in this area such as face detection, face tracking and face recognition in images have led to an increase in the number of studies on facial expression analysis. Facial expressions are formally determined by the deformation of certain regions on the face. By contractions of the facial muscles, especially the eyebrows, eyelids, lips, facial skin, facial features such as nose and temporary undergo changes because of the location of these temporary changes to a few seconds, the intensity and determination of the dynamics, in terms of detection and classification of facial expressions is of major importance. The extent and duration of such changes; it can change in line with the person, age, race, and gender, making the problem very stratified and difficult to solve. Pantic and Rothkrantz identified 3 main problems that emerged in studies on facial expression analysis. These are detection of facial parts in the image, extracting the features of facial expressions and facial expressions classification (Pantic & Rothkrantz, 2001).

It is seen that the studies in the literature also focus on solving the three main problems mentioned above. Kotsia and Pitas used support vector machines (SVM) for facial expression analysis over image sequences (Kotsia & Pitas, 2005). Within the scope of the study, they carried out, expression classification is made according to the displacement with the following images by referring to the points placed in the first image in the image sequence. Classify six different facial expressions, Otsuka used classifiers based on the Hidden Markov Model (HMM) (Pantic & Rothkrantz, 2001). Within the scope of their studies using Gabor features in expression analysis, suggested the AdaSVM method developed by them (Islam & Al-Murad, 2017); Cohn and Kanade achieved extremely high success rates in the facial expression dataset (Lucey et al., 2010). "Feature selection", which is one of the three main problems defined by Pantic and Rothkrantz, is a part of the extraction of features of facial expressions, and it is a research topic in which studies on classification of expressions are carried out continuously (Pantic & Rothkrantz, 2001).

In Donato's approach, techniques for facial recognition were initially used (Yang & Kriegman, 2002). In the face image, there is also the identity information along with the expression; to perform a detailed expression analysis, it is useful to remove the identity information from the image. By taking sequences of images from the video, a difference image is created, and expressions were classified according to different spatial analyzes with optical flow estimation. Again, in Pantic and Rothkrantz's approach, facial movement units (Action Units) were determined by fitting point-based face models to the image; These movements have been used to classify emotional expressions (Pantic & Rothkrantz, 2001). In Colmenarez's work, feature points are also used for face modeling (Colmenarez et al., 1999). On the other hand, it should be said that the excessive dependence on the accuracy of the feature points is a disadvantage for such approaches. Numerous studies have been conducted on facial expression tracking and expression classification. These studies generally consist of three basic steps: face detection, facial feature extraction and facial expression recognition (Dhavaliker and Kulkarni, 2014). In determining the facial region, features such as skin color and the geometric structure of the face are considered. The fact that the real world is 3D in computer vision and 2D image processing brings with it some difficulties. It is extremely difficult to solve the face detection problem with a computer model. Therefore, face detection must also be accurate for an accurate and reliable facial expression recognition system. In the studies, two feature selection approaches are used to determine the ones that carry the most information about the expression analysis of the selected features (Güneş & Polat, 2009). The common features of the approaches can be listed as transforming the multiple classification problem into a binary classification problem, applying the feature selection, and then combining the selected features and using them for classification. For the real time facial expression recognition system, various studies have been done by using deep learning algorithms (Buhari et al., 2022; Umer et al., 2022; Bisogni et al., 2022). Umer et al. examined the effects of data augmentation methods to classify facial emotions for different database (Umer et al., 2022). Also they obtained successful accuracy results but they didn’t examine the results of real time application. Similarly, Bisogni et al analyzed deep learning approaches for the facial emotion recognition problem. As a result of these approaches, successful results were obtained for different data sets (Bisogni et al., 2022). However, they did not calculate accuracy for real-time systems as in the study of Umer et al.. Buhari et al. proposed a new method to recognize real-time facial emotion. In their studies, the accuracy rate was obtained as 87% in SMIC dataset (Buhari et al., 2022). Although good results have been obtained for the real-time application, it seems that there are results that are open to improvement. Saurav et al. improved dual integrated Convolutional Neural Network (CNN) architectures and they obtained the accuracy results for the different database (Saurav et al., 2022). Their study didn’t include real time face emotion recognition. All these studies indicate the need to increase accuracy for real-time systems in facial expression recognition. It seems that it is difficult to achieve high classification success in real-time systems due to various environmental (light, camera) factors (Buhari et al., 2022; Devrim et al., 2019, Engin et al., 2017). The classification accuracy in this regard needs to be improved. For these reasons, the aim of this study is to increase the success of facial emotion recognition for real-time systems.

In this study, different algorithms were evaluated to find best method for real-time facial recognition system. In this context, detailed analyzes are included in this study. After feature extraction in image processing, successful results can be obtained with traditional machine learning methods (Nagaraj & Banala, 2021). For this reason, in this study traditional machine learning algorithms and convolutional neural networks were compared. Also different databases were used to compare the effects of the databases in the algorithms. In addition to these, two practices were applied to all databases and algorithms to compare real-time and non-real time applications. The rest of this study is organized as follows: the dataset, deep learning models, methods, and techniques are given in Sect. 2. The results of real time applications are reported in Section 3. Also the discussion of the results is given in Section3. Section 4 consist of conclusion.

2. Materials and Methods

Facial recognition system is an integrated system consisting of software and the best deep learning facial system recognition. In this study, we have analyzed the different machine learning algorithms to understand which method is proper for the real time emotion recognition system. While the public datasets were used in the creation of the models, the test phase was carried out with 2 different practices to objectively test the success of these algorithms. In practice 1, all models were tested using images consisting of 70 image data (Mena-Chalco et al., 2009) not used in the training set. In practice 2, images captured from webcam were used to test the models.

The aim of this study is to create a real-time face recognition system with the Python programming language. Since it worked with the three different databases, a laptop computer with powerful hardware was chosen to run the software. To obtain the best results, the application and different parameters and tests were repeated several times on the Spyder and Anaconda development environment using Python language on an 8-core Intel Core i7 processor HP laptop. Python is an object-oriented, interpretative, modular, and interactive high-level programming language. Simple syntax based on indentations makes the language easy to learn and remember. This gives it the feature of being a language where programming can be started without wasting time with the details of syntax (Python Software Foundation, 2012). It is using only CPU cores for all other phases except the face recognition phase. The laptop's GPU is not used because, as mentioned above, the goal is to develop a system that can run with very few computational resources. Also, Logitech C270 720p webcam was used externally from the laptop to capture images and videos to test the system on a standard laptop. In this study, the camera was set to be directly opposite and this program was run in a bright environment to obtain more accurate results.

TensorFlow is an open-sourced end-to-end platform, a library for multiple machine learning tasks, while Keras is a high-level neural network library that runs on top of TensorFlow, whereas OpenCV is a computer vision library (Terra, 2022). OpenCV is library developed specifically for computer vision algorithms. Tensorflow is framework for machine learning problems. Though it is suited for more general problems as well, such as: classification, clustering and regression, you can do image recognition with TensorFlow. Keras is a high-level neural networks API, written in Python and capable of running on top of Tensorflow. It concludes that Tensorflow being a machine learning framework is widely adapted for deep learning tasks in research and industrial community. (Mokhtari, 2021) Most of the open-sourced deep learning models are written in either Tensorflow. With OpenCV, you can test simple face detection techniques like SVM and Histograms of Oriented Gradients (HOG).

In this study, three different public databases were used to train the model. These are FER2013, JAFFE and CK+ databases. These datasets were chosen because they are widely used public datasets. Chen et al. analyze and compare facial expression recognition methods. In this study, they propose some evaluation dimensions and discuss possible directions for future research by using CK+ database (Chen & Wu, 2018). Feature extraction is important issue in real-time application in terms of processing time. To extract the feature, HOG method was used. Different machine learning algorithms were used to classify these extracted features. These are k-Nearest Neighbor (k-NN), SVM and CNN methods. Also transfer learning was used in CNN architecture to improve the performance of the network.

2.1 FER2013 Database

The FER2013 dataset used to create the model in this study includes 35,887 face images. These images are in gray format with 48x48 pixels. Label parts of these photos are included in this data set. The photographs that contain 7 different emotions -happiness, sadness, surprise, disgust, anger, fear and neutral - described in the previous sections are a data set is open source. Some images used in the FER2013 data set are given in Figure 1.

Figure 1. FER2013 dataset (Verma & Verma, 2020)

2.2 JAFFE Database

The JAFFE database is a publicly available dataset of 213 facial expression images of 10 Japanese women, called the Japanese Female Facial Expression (JAFFE). Each subject displayed 7 basic emotions, and each expression included 3 - 4 images per subject. This database, which contains 30 angry, 29 disgust, 33 fear, 30 happiness, 31 sad, 30 surprise and 30 neutral expressions and it is a grayscale and has a resolution of 256 x 256 pixels. All facial images were created under similar, highly controlled lighting conditions, without any environmental factors such as hair or glasses. Some images used in the JAFFE data set are given in Figure 2.

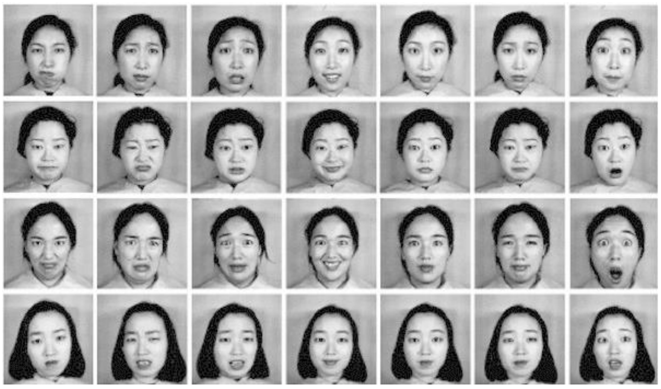


Figure 2. JAFFE dataset (Akinrotimi, 2018)

2.3 CK+ Database

The expanded Cohn-Kanade (CK+) dataset includes 593 images from a total of 123 different subjects aged 18 to 50 years, of various genders. Each image is in 640x480 pixel resolution and mostly in gray format. 327 of these images were tagged with one of the seven expressions. The CK+ database is considered the most widely used laboratory-controlled facial expression classification database available and is used in most facial expression classification methods. Some images used in the CK+ data set are given in Figure 3.

Figure 3. CK+ dataset (Lucey et al., 2010**)**

2.4 Feature Selection Method Used in the Application

In this study, a real-time facial expression recognition software was studied. Different models were composed by using different machine learning methods. These models are basically based on recognizing 7 basic emotions. This software is based on the determination of facial expressions and their classification through decision mechanisms. For the classification of facial expressions, the main features expressing emotions should be extracted. In this study, these processes are carried out with the method named HOG. An example image with the HOG method is given in Figure 4. The histogram of directed gradients is a feature selection method used in computer vision and image processing for object detection, a feature descriptor often used for object detection. It relies on the property of objects within an image to have the distribution of density gradients or edge directions. Gradients are calculated within one image per block. A block is considered a pixel grid from which gradients are created from the magnitude and direction of the change in pixel intensities within the block.

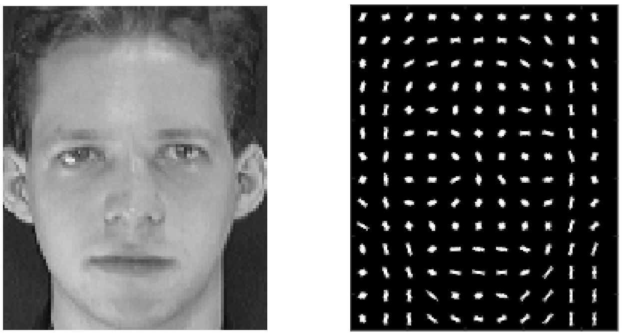


Figure 4. HOG sample face photo (Islam & Al-Murad, 2017)

A window size of 128 x 128 pixel was used for face images because it covers human faces (Yang & Kriegman, 2002). In HOG methods, identifiers are calculated on blocks of 8 x 8 pixels. For each pixel on the 8 x 8 block, these descriptive values are quantized into 9 bins, where each bin represents a directional gradient angle and the value in that bin, which is the sum of the magnitudes of all pixels with the same angle. Also, the histogram is then normalized to a block size of 16 x 16, which means four 8 x 8 blocks are normalized together to minimize light conditions. This mechanism reduces the accuracy drop due to a change in light.

In this study, one of the deep learning libraries of the Python programming language, Keras - TensorFlow was used in this software. The block diagram of the proposed method is shown in Figure 5. When the program is run, it takes face images from the webcam and after a preprocessing called preprocessing, face detection is made. Then, features are extracted with the HOG method. The features are performed automatically with the designed convolutional neural network. By extracting features of face, seven universal emotions can be classified as happiness, sadness, anger, disgust, surprise, neutral and fear.

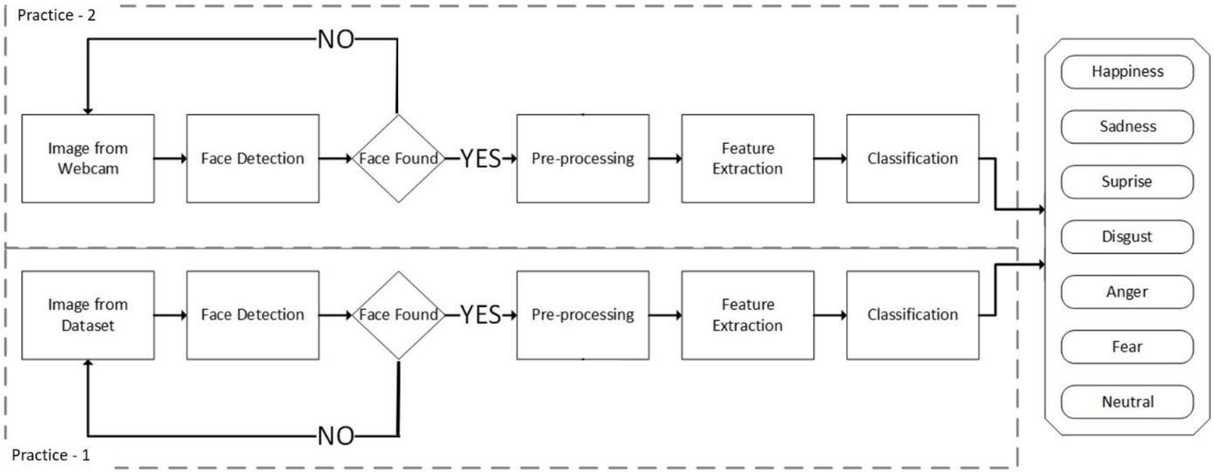


Figure 5. The proposed emotion recognition system from facial expression based on image processing

2.5 Methods

In this study, 3 different machine learning methods were used to classified. SVM and k-NN are traditional machine learning algorithms. CNN is one of the deep learning algorithm. Thus, in this study, traditional methods and deep learning can be compared for emotion recognition application.

SVM is a supervised machine learning algorithm used in classification and regression problems. In classification applications, SVM method is a binary classifier. In the SVM algorithm, each data is represented as a point in n-dimensional space, with the value of a particular coordinate. It then finds the hyperplane that best separates the two classes. The algorithm classifies the data according to which side of this hyperplane it is (Khoong, 2021).

k-NN is one of the supervised machine learning algorithms. So the algorithm uses the labeled dataset. The class of the new observation is determined according to the classes of the k nearest observations. The new observation is classified according to the class with the highest votes among its k nearest neighbors (Srivastava,2018).

CNN is a deep learning network that is very successful in image processing. It is developed by analogy from the visual cortex. CNN architecture processes the image by passing it through various layers. The commonly used layers are convolution layer, pooling layer, flattened layer, fully connected layer and softmax layer. Different architectures can be obtained stacking of these layers (Verma &Verma, 2020) Alexnet is the first large-scale CNN architecture to succeed in the Imagenet competition (Çınar, 2018). Alexnet architecture composes of 5 convolution layers and 3 fully connected layers. It was trained with approximately 1.3 million images with 1000 classes. Residual Network (ResNet) is another CNN architecture. The main basic element of the Resnet architecture is residual blocks. To prevent vanish gradient problem, residual links are used in this architecture (Özdemir et al., 2019). There are different ResNet architectures with different number of layers. In this study, Resnet50 consisting of 50 layers was preferred. While training neural networks, losses occur in the feature map due to convolution and pooling operations. Dense net architecture has been proposed to prevent this. It is a CNN architecture consisting of dense blocks. The layers within the blocks are densely interconnected (Özdemir et al., 2019). And lastly, InceptionV3 CNN architecture was used. This architecture proposes an initial model that combines multiple different sized convolutional filters into a new filter (Islam & Al-Murad, 2017). This model reduces the number of parameters to be trained. This also reduces the computational complexity of the model.

2.6 Transfer Learning

Reuse of a previously trained model on a new problem is defined as transfer learning in machine learning. Transfer learning is the reuse of information learned by one network in another dataset to improve the performance of another network (Sharma, 2021).

A machine uses knowledge learned from a previous assignment to increase insight into a new task in transfer learning. When training a classifier to predict whether an image contains anything, information acquired during training can be used to distinguish faces. If a simple classifier is trained to predict whether an image contains a smiling human face, the model's training information can be used to identify other facial expressions.

Transfer learning is commonly applied in sentiment analysis (Li & Lima,2021). Neural networks typically aim to detect edges in the first layer, forms in the middle layer, and task-specific features in the later layers(Çınar, 2018). The early and central layers are used in transfer learning, and the later layers are only retrained. As the newly created model is trained to recognize facial expressions at earlier levels, subsequent layers are retrained to understand what distinguishes a smiley from other expressions (Akinrotimi, 2018).

This information learned by a deep network is used by adapting the model to improve the performance of another network and to develop a network that is not suitable for the training set. The situations where the model trained and used for a different data set is called model adaptation. The purpose of this method; the knowledge of the previous model is used by creating a new model on top of the models at a certain level that have learned the subject well (Goodfellow et al., 2017)

3. Results and Discussion

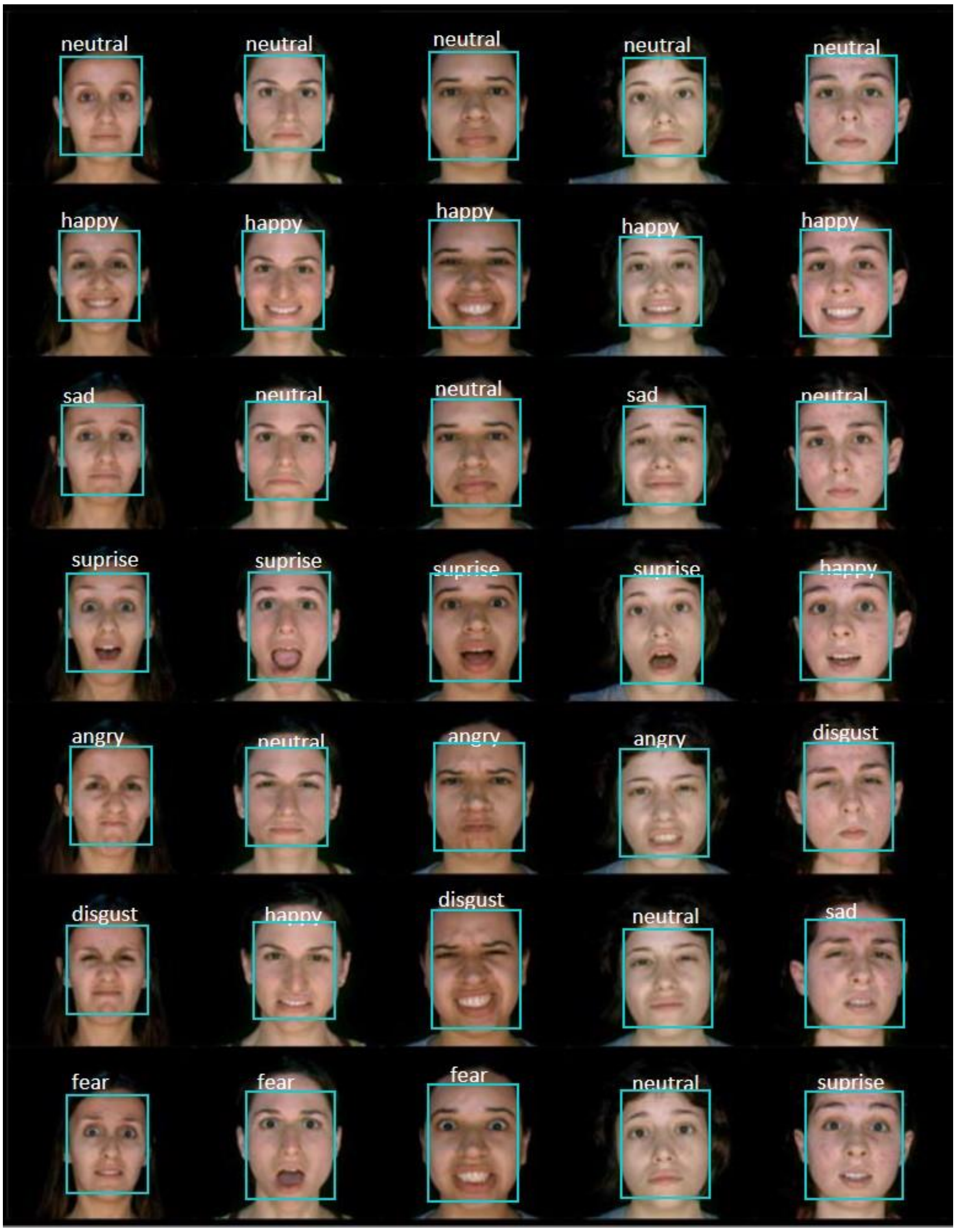
JAFFE, FER2013 and CK+ databases as facial expression databases are applied to observe the difference between the management applied in this study and the methods in the literature. OpenCV library for face reading and face recognition processes were used. Studies were carried out with k-NN and SVM classifiers and AlexNet, ResNet, DenseNet and Inception learning algorithms. All the results are given in Table 1.

Table 1
The comparison of accuracy rate and performances of different algorithms for practice 1

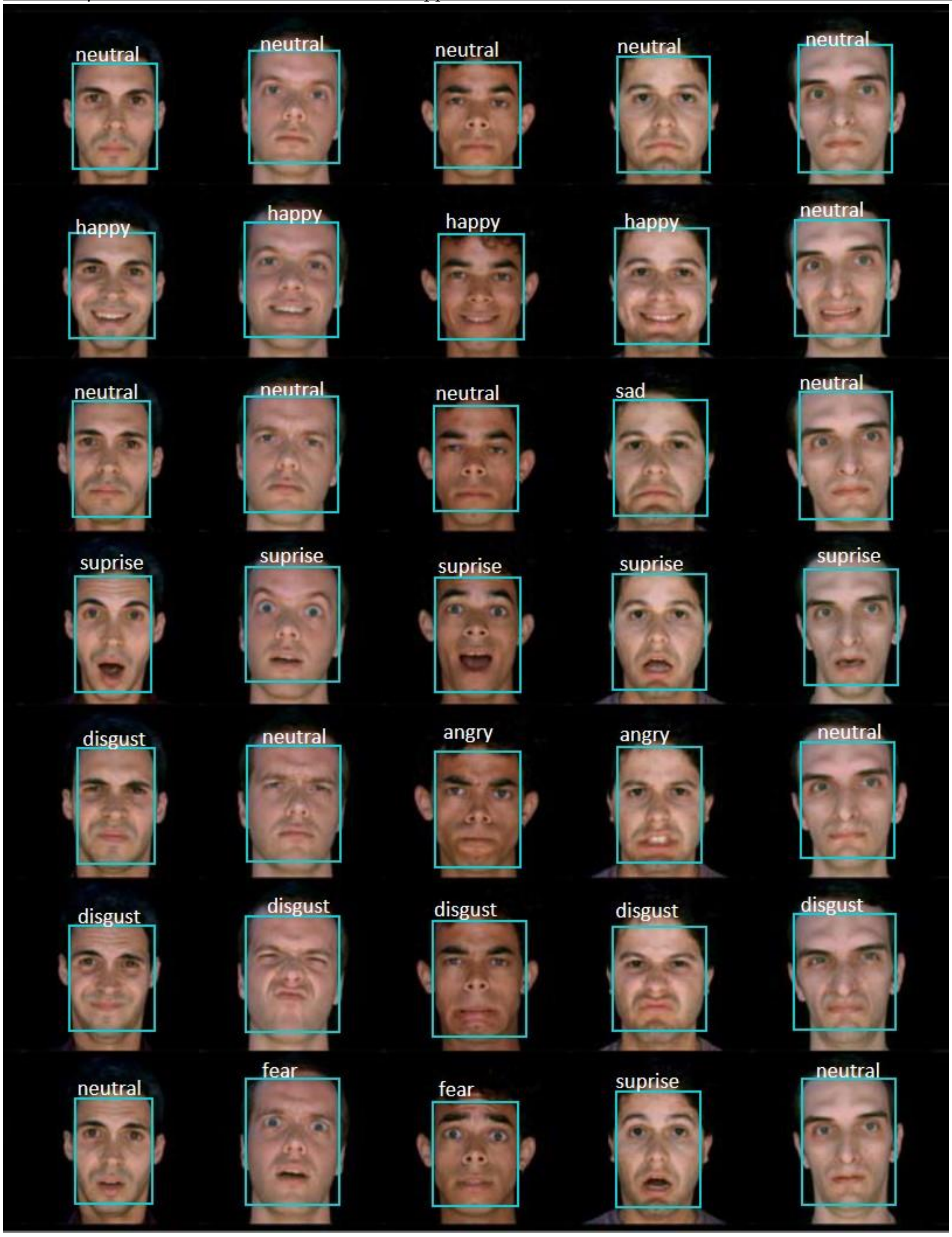
	Database	**Practice-1 Accuracy**	**Practice-2 Accuracy**	**Training Time (sec)**
SVM	**JAFFE**	73,08%	74,80%	11
	FER2013	71,88%	73,22%	12
	CK+	72,74%	74,45%	11
KNN	**JAFFE**	72,58%	74,98%	12
	FER2013	77,39%	78,92%	14
	CK+	73,94%	76,71%	13
AlexNet	**JAFFE**	81,67%	83,48%	30
	FER2013	83,42%	84,70%	34
	CK+	82,26%	84,11%	32
ResNet50	**JAFFE**	90,34%	91,64%	22
	FER2013	93,85%	94,51%	26
	CK+	92,11%	93,12%	25
DenseNet121	**JAFFE**	89,82%	90,26%	28
	FER2013	92,04%	92,30%	31
	CK+	90,79%	91,54%	30
InceptionV3	**JAFFE**	84,58%	87,52%	16
	FER2013	86,73%	89,30%	19
	CK+	85,07%	87,85%	17

In this study, accuracy and time analyzes of two different applications were carried out. In practice 1, facial expressions belonging to 10 different people (Mena-Chalco, 2009) were shown the model to test. In practice 2, it was tried to perform the emotion recognition function from facial expression in real time by using web-cam image. The accuracy rates obtained during these applications and the time spent in each training round in machine / deep learning methods are given in seconds. All transfer learnings were performed in 100 rounds to get the best results. The epoch, batch size, momentum and learning rate are selected as 100, 32, 0.9 and 0.01 respectively in all transfer learning architectures. Adam optimization algorithm was used to update network weights. To obtain overfitting problem dropout was applied and dropout rate was selected as 0,2. In terms of accuracy, it is observed that deep learning networks have a very high accuracy rate.

In the first application, all methods, and their effects on three different databases (JAFFE - FER2013 - CK+) were tested with a total of 70 images of 10 different people. The expressions shown in Figure 6 shows the emotions detected by the application.



(a)



(b)

Figure 6. The faces images used in the practice-1 for test a) first five-person b) last five person (Mena-Chalco, 2009)

In second practice, the webcam images shown in Figure 7 classified for the facial expression in real – time. When the model was trained with the FER2013 database and classified by using SVM method, the confusion matrix shown in Figure 8 is obtained for practice 2. The color bar in Figure 8 indicates that if the color in the confusion matrix is approaching red, the value is approaching 100%. If it is going towards the dark blue color, it indicates that it is approaching 0%. The model achieved the highest success in recognizing the "neutral" expression. The lowest accuracy rate was recorded in the facial expressions of "surprise".

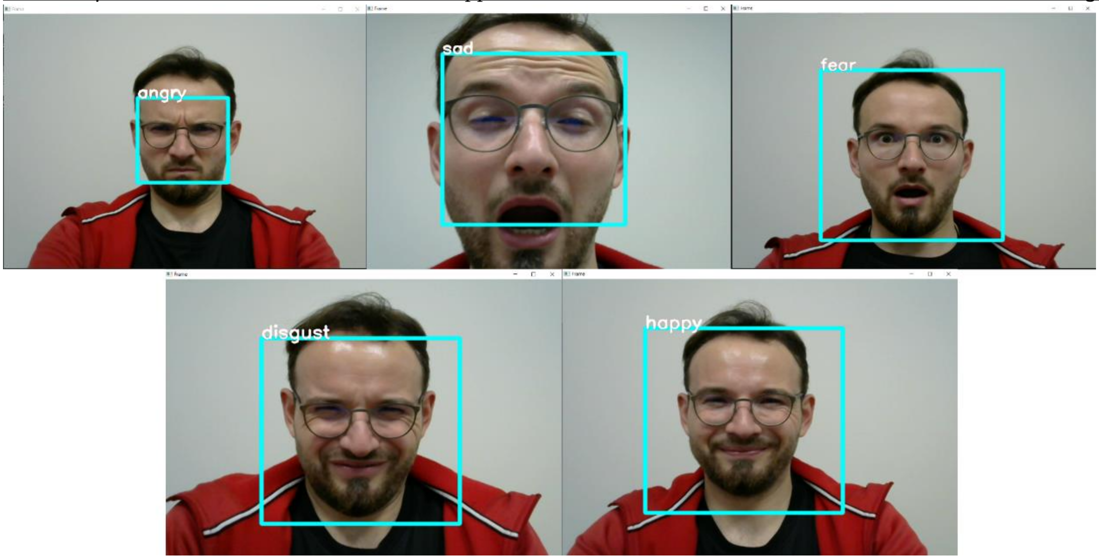


Figure 7. Screenshot images (angry, sad, fear, disgust and happy) from real-time face emotion recognition application

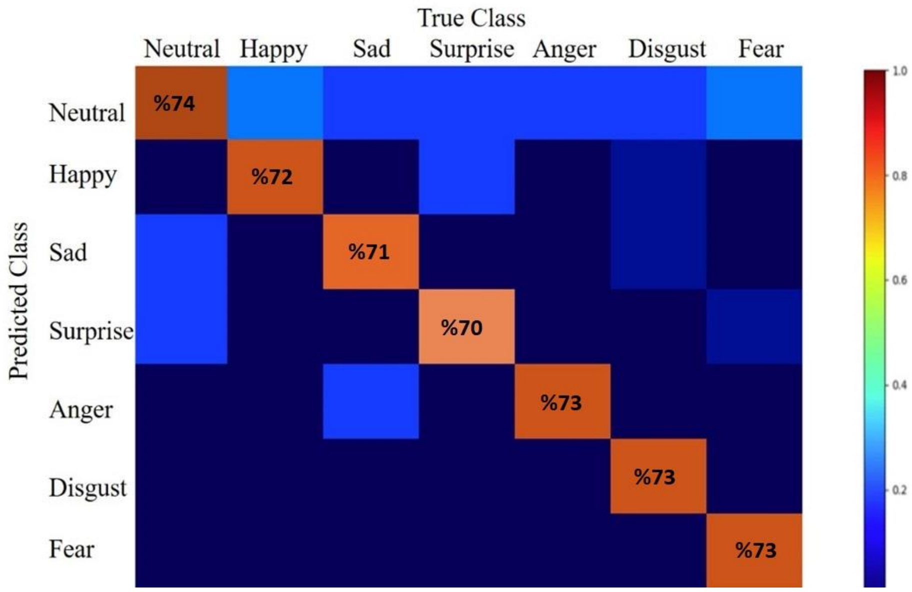


Figure 8. The confusion matrix when the model is trained with FER2013 database and the SVM algorithm is used for classification in practice 2

In another model, the highest accuracy rate was obtained for the practice 2 when the model was trained with FER2013 database in ResNet architecture, which is one of the transfer learning methods. In this model, the lowest accuracy rate was observed in the expressions "disgust" and "surprise", while the highest accuracy rate was observed in the expression "neutral". The confusion matrix of this study is also given in Figure 9.

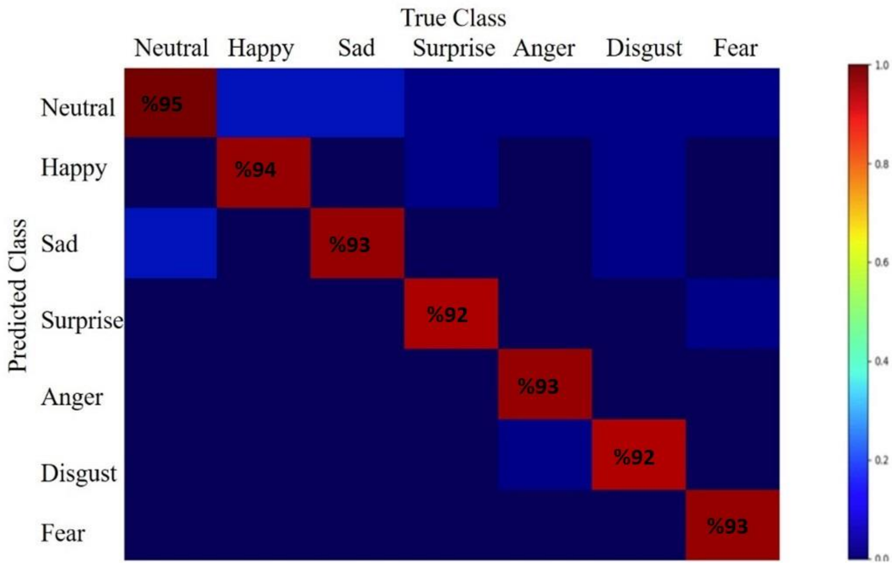


Figure 9. The confusion matrix when the model is trained with FER2013 database and ResNet architecture is used for classification in practice 2.

The confusion matrix, which includes the accuracy rates according to the seven different expressions that are considered as a basis. Neutral expression was mostly observed as the expression with the highest accuracy rate. The main reason for this is that a neutral facial expression can be created on the face without activating any facial muscles. Many different muscles of the face must be activated and contracted at the same time in the other facial expression. These contractions occur differently on each face. These differences detected better as the dataset gets larger and random. Thus, as the data set grows, the expression detection rate also increases. The FER2013 dataset contains the most data from the datasets used in this study. As can be seen from Figure 10, the highest success was achieved in all methods except SVM when FER2013 data was used in training stage. SVM is a method that is successful even when the number of features is high and the number of measurements is low (Güneş & Polat, 2009) For this reason, this study also achieves a certain success in data sets with a small number of data such as JAFFE dataset.

In practice 2, tests were carried out in real time and the expressions taken by the webcam The results of our application show that the tests were more successful in real-time application. The reason for this is that preprocessing, and feature extraction processes are successful in real-time applications when suitable conditions are created.

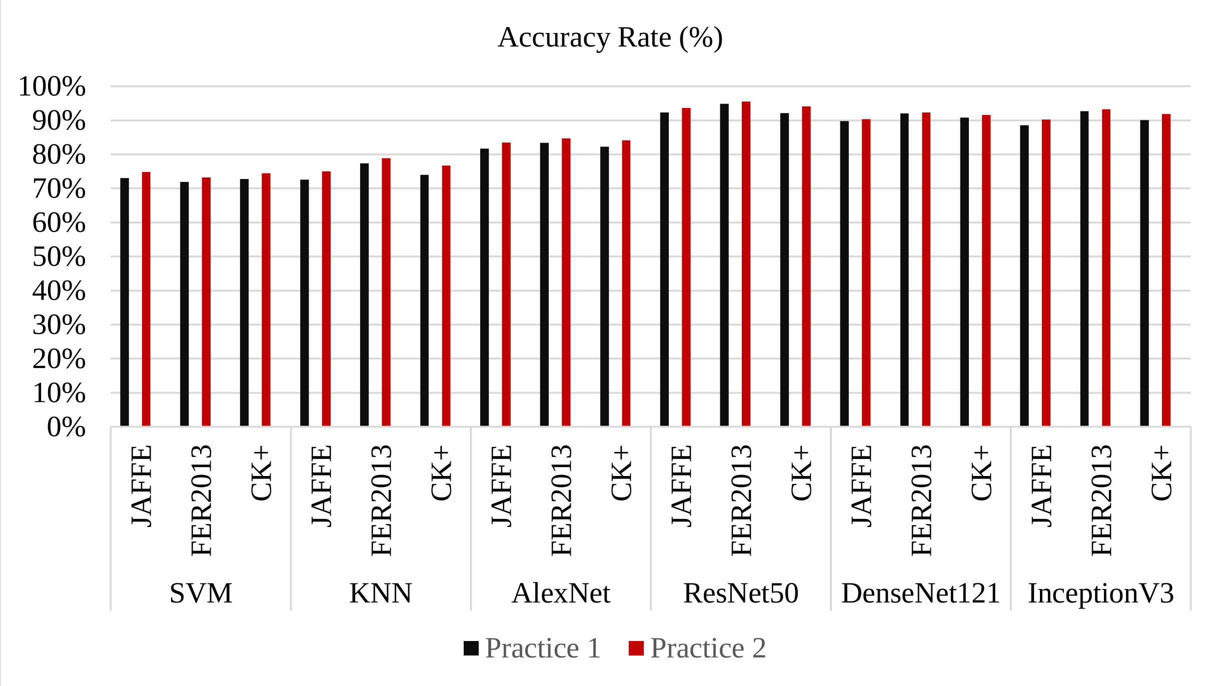


Figure 10. The comparison of accuracy rates of different algorithms for practice 1 and 2.

In the Figure 10, which method gives better results in which application and which method achieves higher accuracy are detailed. ResNet50 was observed in the FER2013 dataset with an accuracy rate of 94.51%. Also classification successes given in Table 2 are analyzed according to each facial expression. Neutral expression is mostly observed as the expression with the highest accuracy rate like the practice 1. For the same reason as in application 1, it is thought that the success obtained in neutral expression is high.

Table 2
Classification accuracies of each facial expression for practice 2.

Method	Dataset	Neutral (%)	Happy (%)	Sad (%)	Suprise (%)	Anger (%)	Disgust (%)	Fear (%)
SVM	**JAFFE**	73,63	73,87	71,24	71,92	72,94	71,63	73,36
	FER2013	73,45	71,39	70,75	69,83	72,22	72,98	72,46
	CK+	73,15	73,48	71,13	71,26	71,91	71,56	72,54
KNN	**JAFFE**	73,46	73,43	72,91	71,82	72,67	71,83	71,92
	FER2013	78,64	77,84	76,97	76,58	77,42	76,24	76,43
	CK+	74,65	74,18	74,81	72,26	75,27	72,74	73,36
AlexNet	**JAFFE**	82,29	82,45	80,91	79,83	80,19	80,13	81,59
	FER2013	83,86	83,89	83,56	81,76	83,47	83,61	83,72
	CK+	82,93	82,79	81,65	81,26	82,73	81,88	82,46
ResNet50	**JAFFE**	90,83	90,61	90,14	90,50	89,73	89,84	90,89
	FER2013	94,83	93,42	92,84	91,76	92,51	91,79	92,23
	CK+	92,61	92,17	91,68	91,13	92,09	91,49	91,88
DenseNet121	**JAFFE**	89,42	89,35	89,07	88,59	89,29	89,12	89,20
	FER2013	92,18	91,28	91,07	90,69	91,68	90,83	92,05
	CK+	90,12	90,25	90,92	90,87	91,26	90,45	91,62
InceptionV3	**JAFFE**	86,17	85,97	85,57	84,74	85,21	85,15	85,23
	FER2013	88,59	88,21	86,65	86,41	86,26	85,42	85,33
	CK+	86,13	85,70	85,42	84,69	84,57	84,12	84,24

When the results obtained in this study are compared with the results of the literature, it seems that successful results have been obtained with the proposed method. The results are given in Table 3. Feature extraction with the HOG method and then classification of the obtained feature extracted image with the resnet50 architecture has been successful in all datasets. When the literature results are examined, there are higher successful results. However, since these results do not include real-time application, a full comparison cannot be made. The results show that the overall success of the proposed method is high for all data sets.

3. 2 Application Restrictions

For software to achieve higher accuracy and to work with better performance, some of the restriction steps we encountered in our system are listed below.

- Real-time recognition becomes very difficult for 720p and higher images. This is because the time it takes to execute the face position step is also significantly extended on the platform. Therefore, our system's real-time recognition is limited to images in sizes from 180p to 720p.
- The system can detect and recognize faces from about 50 to 100 cm from the webcam. Outside of this range, the system will perform poorly.
- Since no extra flash or lighting system is used in the system, the angle of the camera and the amount of light in the environment also affect the performance of the system. For best performance, the camera should be positioned so that it does not reflect overhead lighting or daylight.

Table 3
The comparison of accuracy rates with literature for each dataset

Study	Dataset	Feature Extraction Method	Classification Method	Real Time	Accuracy (%)
Saxena, A., 2020	JAFFE	MLP	ANN	x	73
Saxena, A., 2020	JAFFE	Haar	ADABOOST	x	92
Saxena, A., 2020	GWI	Spectral Embedding	GABOR	x	62.3
Chen, Y., 2018	CK+	-	SVM	x	79
Sebe, N., 2004	-	Optical Flow	K-NN	✓	86
Devrim, M., O., 2019	JAFFE	Viola Jones	DenseNet	✓	90,18
Devrim, M., O., 2019	JAFFE	Viola Jones	InceptionV3	✓	87,50
Devrim, M., O., 2019	JAFFE	Viola Jones	ResNet	✓	91,52
Özdemir, M., A., 2019	FER2013	BPSO	ResNet	x	83,40
Engin, D., 2017	CK+	LBP	ResNet	✓	93,21
Engin, D., 2017	CK+	LBP	InceptionV3	✓	85,77
Zhao, X. 2015	JAFFE	CNN	RELM*	x	96,8
Zhao, X. 2015	CK+	CNN	RELM*	x	86,5
Zhao, X. 2015	FER2013	CNN	RELM*	x	62,5
This Study	JAFFE	HOG	Resnet50	✓	91,64
This Study	FER2013	HOG	Resnet50	✓	94,51
This Study	CK+	HOG	Resnet50	✓	93,12

*Regularized Extreme Learning Machine

4. Conclusion

Computer-based expression recognition models have a wide range of uses in daily life. In this study, a facial expression recognition system was developed to detect seven facial expressions based on facial mimics in real time. Different classification methods, were compared in terms of accuracy rates and training times with training images in databases which have different number images. Spyder software, Keras, Tensorflow and OpenCv libraries have been installed in the computer environment. Among the classification methods, k-NN, SVM, AlexNet, ResNet, DenseNet and Inception were preferred. JAFFE, FER2013 and CK+ were preferred as databases because these are open source. A comprehensive analysis was performed for the different datasets and different methods. According to the results of the applications made in these different methods and databases, the highest accuracy rate was observed in

the model using the FER2013 dataset and the ResNet. With CNN architectures, much more successful results have been obtained in this application compared to traditional methods. It was observed that the accuracy rate increased as the amount of data in the data set increased. When the feature is extracted with HOG and then classified with ResNet, good results have been obtained with this proposed method. The highest accuracy rate was obtained as 95,1% with the proposed method in real time application.
By using this application on different datasets, using with different classifiers during training, and increasing the training data, the system can be improved for the future studies. The results of the performance analysis carried out within the scope of this study can be improved by testing more classification methods on different data sets.

Authors Contribution

Selda Güney: Conceived of the presented idea.
Orhan Emre Aksoy: Verified the analytical methods.
All authors developed the theory, performed the computations, discussed the results and contributed to the final manuscript.

Conflicts of Interest

The authors declare no conflict of interest.

References

Abdullah, S. M. S., and Abdulazeez, A. M. (2021). Facial expression recognition based on deep learning convolution neural network: A review. *Journal of Soft Computing and Data Mining*, 2(1), 53-65. Retrieved from: https://publisher.uthm.edu.my/ojs/index.php/jscdm/article/view/7906

Akinrotimi, A, O. (2018). Facial Emotion Recognition Using Principal Component Analysis and Support Vector Machine. *Computing, Information Systems, Development Informatics & Allied Research Journal*, 9(2), 12-20. Retrieved from: https://www.researchgate.net/publication/336120020_Facial_Emotion_Recognition_Using_Principal_Component_Analysis_and_Support_Vector_Machine

Bhattacharya, S. (2021). A Survey on: Facial Expression Recognition Using Various Deep Learning Techniques. *Advances in Intelligent Systems and Computing, Advanced Computational Paradigms and Hybrid Intelligent Computing*, 619–631. doi:https://doi.org/10.1007/978-981-16-4369-9_59

Bisogni, C., Castiglione, A., Hossain, S., Narducci, F, and Umer S. (2022). Impact of Deep Learning Approaches on Facial Expression Recognition in Healthcare Industries. *IEEE Transactions on Industrial Informatics*, 18(8), 5619-5627. Retrieved from: https://ieeexplore.ieee.org/document/9674818

Buhari A. M., Ooi, C. P., Baskaran, V.M., Phan, R. CW, Wong, K. and Tan, W-H. (2022). Invisible emotion magnification algorithm (IEMA) for real-time micro-expression recognition with graph-based features. *Advances in Soft Computing Techniques for Visual Information-based Systems*, 81, 9151–9176. doi:https://doi.org/10.1007/s11042-021-11625-1

Chen, Y., Wu, H. (2018). A Comparison of Methods of Facial Expression Recognition. *2018 WRC Symposium on Advanced Robotics and Automation (WRC SARA)*, 261-268. Retrieved from: https://ieeexplore.ieee.org/document/8584202

Coşar, S., (2008). *Facial Feature Point Tracking Based On A Graphical Model Framework* (Master Thesis). Retrieved From: https://vpa.sabanciuniv.edu/vpadb/findfile_get_anonymous_paper.php?f=1954

Çınar, A., C., (2018) .Deep Learning . Retrieved From:
https://www.ahmetcevahircinar.com.tr/2017/08/11/imagenet-classification-with-deep-convolutional-neural-networks/

Colmenarez, A., Frey, B. and Huang, T.S. (1999). A probabilistic framework for embedded face and facial expression recognition. *Proceedings 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Cat. No PR00149)*, 592-597, doi: 10.1109/CVPR.1999.786999.

Devrim, M., O. (2019). *Farklı Derin Evrişimsel Sinir Ağlarının Yüz İfadesi Tanıma İşleminde Karşılaştırılması* (Master Thesis). Retrieved from: https://tez.yok.gov.tr/UlusalTezMerkezi/tezDetay.jsp?id=vLQ0gfV9vTaOpH16k6kjMA&no=ZgJjiVHnbVeqxh9M0L2OHg

Dhavalikar, A., Kulkarni, R., K., (2014). Face detection and facial expression recognition system. *International Conference on Electronics and Communication Systems (ICECS),* 1-7. Retrieved From: https://ieeexplore.ieee.org/document/6892834

Engin, D., (2017). *Facial Expression Pair Matching* (Master Thesis) Retrieved from: https://tez.yok.gov.tr/UlusalTezMerkezi/tezDetay.jsp?id=2bQ8F7EJQFFi10i_DbMMkQ&no=poooj2n8rQ268jwtzPwoqw

Frank, M., G., (2001). *Facial Expressions*. International Encyclopedia of the Social & Behavioral Sciences, Pergamon, 5230-5234. doi:https://doi.org/10.1016/B0-08-043076-7/01713-7

Goodfellow, I., Bengio, Y., Courville, A. (2016). *Deep Learning. Regularization for Deep learning*. 228-270, Cambridge, Massachusetts: MIT Press. Retrieved From: http://www.deeplearningbook.org

Güneş, T., Polat, E. (2009). Feature Selection In Facial Expression Analysis And Its Effect On Multi-Svm Classifiers. *23rd International Symposium on Computer and Information Sciences*, 1-5, doi:10.1109/ISCIS.2008.4717874.

Göngör, F., and Tutsoy, O. (2018). Eigenface based emotion analysis algorithm and implementation to humanoid robot. *In International Science and Academic Congress*, 1-17. Retrieved from: https://www.researchgate.net/publication/329557854_Eigenface_Based_Emotion_Analysis_Algorithm_and_Implementation_to_Humanoid_Robot

Islam, K., Al-Murad, A. (2017). Performance of SVM, CNN, and ANN with BoW, HOG, and Image Pixels in Face Recognition. *2nd International Conference on Electrical & Electronic Engineering (ICEEE)*, 1-4. Retrieved from: https://ieeexplore.ieee.org/document/8412925

Jung H. et al., (2015). Development of deep learning-based facial expression recognition system. *21st Korea-Japan Joint Workshop on Frontiers of Computer Vision (FCV)*, 1-4. doi: 10.1109/FCV.2015.7103729.

Karaboyacı, C., (2009). *Geometrical Feature Based Automated Facial Expression Analysis* (Master Thesis). Retrieved from: https://polen.itu.edu.tr/bitstream/11527/339/1/9597.pdf

Khoong, W., H., (2021). *When Do Support Vector Machines Fail*? Retrieved From: https://towardsdatascience.com/when-do-support-vector-machines-fail-3f23295ebef2

Kotsia, I., Pitas, I., (2008). Facial Expression Recognition in Image Sequences Using Geometric Deformation Features and Support Vector Machines. *IEEE Transactions on Image Processing*, 16(1), 172-187. doi: 10.1109/TIP.2006.884954. Retrieved From: https://ieeexplore.ieee.org/document/4032815

Kuo, C., Lai, S. and Sarkis, M. (2018). A Compact Deep Learning Model for Robust Facial Expression Recognition. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2202-2208, Retrieved from: https://ieeexplore.ieee.org/document/8575457

Lucey, P., Cohn, J., Kanade T. (2010) .The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression. Retrieved From: https://ieeexplore.ieee.org/document/5543262

Li, B. and Lima, D. (2021). Facial expression recognition via ResNet-50. *International Journal of Cognitive Computing in Engineering,* 2, 57-64. doi: https://doi.org/10.1016/j.ijcce.2021.02.002

Li, S. (2020). Deep Facial Expression Recognition: A Survey. *IEEE Transactions on Affective Computing*, 1-20. Retrieved From: https://ieeexplore.ieee.org/document/9039580

Mehrabian, A. (2016), https://en.wikipedia.org/wiki/Albert_Mehrabian, 01.06.2022.

Mena-Chalco, J.P., Macedo, I., Velho, L. and Cesar, R.M. (2009). 3D face computational photography using PCA spaces. *Visual Computer,* 25, 899-909. doi:https://doi.org/10.1007/s00371-009-0373-x

Mokhtari, N., I., (2021). Which is Better for Your Machine Learning Task. Retrieved From: https://towardsdatascience.com/which-is-better-for-your-machine-learning-task-opencv-or-tensorflow-ed16403c5799

Nagaraj, P., Banala, R., (2021). Real Time Face Recognition using Effective Supervised Machine Learning Algorithms. *J. Phys.: Conf. Ser.* 012007, 1-6. doi:10.1088/1742-6596/1998/1/012007 Retrieved From: https://iopscience.iop.org/article/10.1088/1742-6596/1998/1/012007/pdf

Özdemir, M., A., Elagöz, B., Alaybeyoğlu, A., Akan, A. (2019). Real Time Emotion Recognition from Facial Expressions Using CNN Architecture. *2019 Medical Technologies Congress (TIPTEKNO)* DOI: 10.1109/TIPTEKNO.2019.8895215. Retrieved from: https://ieeexplore.ieee.org/document/8895215

Özdemir and Hanbay. (2022). Deep feature selection for facial emotion recognition based on BPSO and SVM. *Journal of Polytechnic*, Early Access. Retrieved From: https://dergipark.org.tr/tr/pub/politeknik/issue/33364/992720

Pantic, M., Rothkrantz, J., (2001), Automatic Analysis of Facial Expressions: The State of the Art. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(12), 1424-1445. doi: 10.1109/34.895976. Retrieved from: https://ieeexplore.ieee.org/document/895976

Paleari, M., Velardo, C., Huet, B., Dugelay, J. (2009). Face dynamics for biometric people recognition. *2009 IEEE International Workshop on Multimedia Signal Processing*, 1-5, doi: 10.1109/MMSP.2009.5293300. Retrieved From: https://ieeexplore.ieee.org/document/5293300

Python Software Foundation (2012) Retrieved from: https://web.archive.org/web/20121024164224/http://docs.python.org/faq/general.html

Saurav, S., Gidde, P., Saini, R., and Singh, S. (2022). Dual integrated convolutional neural network for real-time facial expression recognition in the wild. *The Visual Computer*, 38, 1083–1096. Retrieved from: https://doi.org/10.1007/s00371-021-02069-7

Saxena, A., Khanna, A., Gupta, D. (2020). Emotion Recognition and Detection Methods. *Journal of Artificial Intelligence and Systems*, 2, 53-79. Retrieved from: https://iecscience.org/uploads/jpapers/202003/dnQToaqdF8IRjhE62pfIovCkDJ2jXAcZdK6KHRzM.pdf

Saraçbaşı, N. Z. (2021). *Face Recognition Using Facial Dynamics of Emotional Expressions* (Master Thesis). Retrieved From: https://tez.yok.gov.tr/UlusalTezMerkezi/tezSorguSonucYeni.jsp

Sebe, N., Cohen, I., Huang, T.S (2004). *Multimodal Emotion Recognition*. Handbook of Pattern Recognition and Computer Vision, 387-409. doi:https://doi.org/10.1142/9789812775320_0021

Sharma, P. (2021). Understanding Transfer Learning for Deep Learning**.** Retrieved from: https://www.analyticsvidhya.com/blog/2021/10/understanding-transfer-learning-for-deep-learning/

Srivastava, T., (2018). Introduction to k-Nearest Neighbors. Retrieved From: https://www.analyticsvidhya.com/blog/2018/03/introduction-k-neighbours-algorithm-clustering/

Terra, J., (2022). Key Differences Among the Deep Learning Framework. Retrieved From: https://www.simplilearn.com/keras-vs-tensorflow-vs-pytorch-article#:~:text=TensorFlow%20is%20an%20open%2Dsourced,because%20it's%20built%2Din%20Python.

Tutsoy, O., Güngör, F., Barkana, D. E., & Köse, H. (2017). An emotion analysis algorithm and implementation to NAO humanoid robot. *The Eurasia Proceedings of Science Technology Engineering and Mathematics*, 1, 316-330. Retrieved From: http://www.epstem.net/tr/download/article-file/381431

Umer, S., Rout, R.K., Pero, C., and Nappi, M. (2022). Facial expression recognition with trade-offs between data augmentation and deep learning features. *Journal of Ambient Intelligence and Humanized Computing*, 13, 721–735. doi:https://doi.org/10.1007/s12652-020-02845-8

Verma, G., Verma, H. (2020). Hybrid-Deep Learning Model for Emotion Recognition Using Facial Expressions. *The Review of Socionetwork Strategies*. doi:https://doi.org/10.1007/s12626-020-00061-6

Yang, M., H., Kriegman, D., (2002). Detecting Faces in Images: A Survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(1), 34-58. Retrieved from: https://ieeexplore.ieee.org/document/982883?arnumber=982883

Zhao, X., Shi X. and Zhang, S. (2015). Facial Expression Recognition via Deep Learning. *IETE Technical Review*, 32(5), 347-355, Retrieved From: https://ieeexplore.ieee.org/document/7043872