Computational Engineering and Technology Innovations

www.ceti.reapress.com

Comput. Eng. Technol. Innov. Vol. 1, No. 2 (2024) 76–85.

Paper Type: Original Article

# Real-Time Facial Emotion Recognition: Insights and Comparative Analysis

**Seyed Ali Noorkhah[1,*], Zhang Hao[2], Haifa Alqahtani[3]**

[1] Department of Civil Engineering, Science and Research Branch, Islamic Azad University, Tehran, Iran; a.noourkhah@aihe.ac.ir.

[2] Institute of Education, Guizhou Normal University, Guizhou Province, China; haozhang_guizhou@sina.com.

[3] Analytic in the Digital Era, Collage of Business and Economic, United Arab Emirates University, United Arab Emirates; h.alqahtani@uaeu.ac.ae.

**Citation:**

| | |
|---|---|
| *Received: 02 February 2024*<br>*Revised: 17 March 2024*<br>*Accepted: 24 April 2024* | Noorkhah, S. A., Hao, Zh., & Alqahtani, H. (2024). Real-time facial emotion recognition: insights and comparative analysis. *Computational engineering and technology innovations, 1*(2), 76-85. |

## Abstract

Facial emotion recognition [1] is very useful these days and has various applications in product feedback, virtual assistants, safe and personalized cars, video game testing, monitoring expressions in an interview, law enforcement, surveillance, and monitoring. The orientation, position, and movement of the various facial muscles near the eyes, lips, nose, and chin are among the factors that affect a real-time emotion. To identify the facial emotion, it typically requires the feature extractor to detect the feature, and the trained classifier produces the label based on the feature. This paper discusses and compares various real-time methods for detecting facial expressions, taking into account a number of factors such as false negative rate, recall, precision, accuracy, false positive rate, specificity, etc. The results are produced after training the model on images of seven basic emotions (happy, sad, angry, surprised, disgusted, neutral, and fearful) in the dataset.

**Keywords:** Emotion recognition, CNN, AlexNet, HOG-ESR, Affdex CNN, SVM of HOG.

## 1 | Introduction

A facial expression is composed of one or more motions or positions of the facial muscles. According to a debunked theory, these movements show observers how someone is feeling.

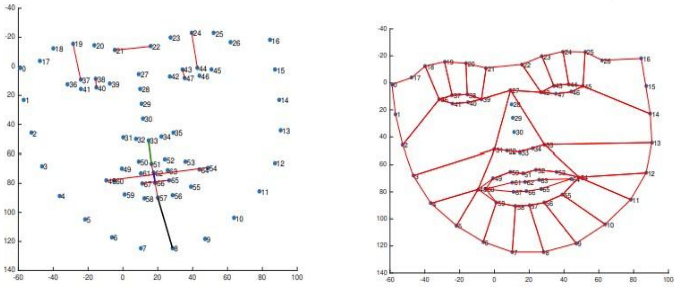

**Fig. 1. Geometrical features computed from the facial landmarks a. 11 distance features, b. 26 area features.**

Corresponding Author: a.noourkhah@aihe.ac.ir

10.48314/ceti.v1i2.29

Licensee System Analytics. This article is an open-access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0).

Nonverbal communication can also take the form of facial expressions. Whether consciously or unconsciously, people can express themselves through their facial expressions, and the neurological systems that do so depend on the circumstances. A cortical pathway in the brain frequently underlies voluntary facial expressions [2], which are also socially conditioned. On the other hand, it is thought that involuntary facial expressions are intrinsic and go through the subcortical region of the brain. The eyes are frequently seen as key components of facial expressions [2]. Blinking rate, for example, may be utilized to determine whether someone is lying or just nervous. Moreover, making eye contact is seen as a crucial component of interpersonal.

The propriety of maintaining eye contact or breaking it depends on the culture, though. Human social interaction depends on facial expressions. The movement of the facial [3] fascia and the muscles that are connected to it causes them. These muscles move the skin, which results in the formation of folds and creases and the movement of facial features like the mouth and brows.

The second pharyngeal arch in the embryo gives rise to these muscles. The muscles, including the internal and external pterygoid muscles, the masseter and temporalis muscles, and others, also have a minor impact on facial expression. Starting with the first pharyngeal arch, these muscles grow.

The amount of research has grown significantly over the past few years, and various methodologies have led to various algorithms. The two primary techniques for extracting emotion from an image involve training the model with a CNN through images and training the model with a CSV file that highlights the pixels present in the image (Mini XCEPTION).

The observations are recorded for the FER-2013 dataset, which consists of 35,887 48 x 48 grayscale images in total. These images are used to train and test the model. Of these, 28.709 images are used for training, and the remaining [4] 7.178 images are used for testing.

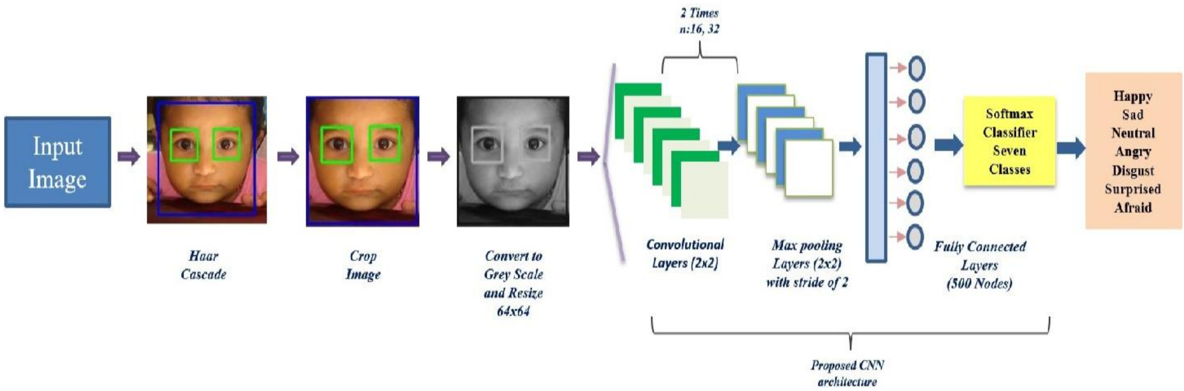


**Fig. 2. Steps in FER.**

The majority of the algorithms used are real-time algorithms, which means they can recognize human facial features and then emotions in real time, which is crucial in many fields today that deal with recognizing human emotions. The approach of using HOGs helped improve the accuracy of the model [5], as in FER-CNN, the accuracy was not high, but in HOG-ESRs, it is quite high. Earlier, a lot of work was done on analyzing and comparing the algorithms, but it was done for four emotions. In this paper, the analysis is done for seven emotions using traditional approaches, and some other parameters are discussed along with their significance.

# 2 | Methodology

In this paper, the algorithms discussed are: 1) AlexNet CNN 2) FER-CNN 3) SVM of HOG 4) HOG-ESR 5) Affdex CNN.

## 2.1|FER-CNN

FER-CNN is a deep learning [1] method with two convolutional neural network layers and one fully connected layer. Max pooling, which pulls the maximum value from the pool and creates the output, has been used in both layers of the pool size (2.2). ReLU is the activation function employed in both top layers. In order to avoid overfitting and effectively train the model, the dropout function has also been used. The softmax function is used to obtain network outputs in the fully connected layer. The pictures are then categorized using labels like angry, happy, sad, disgusted, surprised, neutral, and fearful. There are 28.709 trained images in total, of which 7.215 are trained to detect the happy emotion, 3.995 to detect the angry emotion, 4.850 to detect the sad emotion, 4.097 to detect the fearful emotion, and 3.171 to detect the surprised emotion. Similar to this, 436 images are trained to detect the disgustful emotion, while 4.965 images are trained to detect the neutral emotion.

While training the model, we could achieve a training accuracy of around 97.63% for 200 epochs. Having this much higher accuracy while training the model is generally not considered good as there are high chances of overfitting, and the model will perform poorly on the test data. During the evaluation of the model, the validation accuracy achieved was 17%, i.e., the model is overfitting. To remove this, we used data augmentation and again evaluated the model on the testing data; this time, we could achieve a validation accuracy of around 71%.

## 2.2|AlexNet CNN

Here, the pre-trained model of Alexnet CNN has been modified a bit and used to find how well it works on the FER-2013 dataset. Alex Krizhevsky and his research team first put forth the pre-trained model of Alexnet CNN [6], which contained three fully connected layers and five convolutional layers. The AlexNet network's convolution layer's convolution Kernel size is a single value, and the generated feature map features aren't diverse, so there won't be enough feature extraction. The three flaws-a single size convolution kernel, an excessive number of parameters in the fully connected layer, and a change in data distribution-were fixed.

In the conventional convolution layer, a single-sized convolution kernel [6] convolves the input data for many feature maps. Multichannel convolution technology is an extension of traditional convolution and consists of the addition of multiple convolution kernels of various sizes to a single convolution layer. The main issue was working with too many parameters, and to solve this problem, researchers came up with the idea of the GAP Layer which is the Global Average Pooling layer; this layer has been used in place of the fully connected layer.

For the training purpose, the batch size chosen was 64, the number of steps in each epoch was 5000, and the image size for the facial expression was set to 48 x 48. 0.01 was the starting value for the learning rate. The network was made more efficient using Adam Optimizer. By substituting four different-sized convolution kernels (1*1, 3*3, 5*5, and 7*7) for the traditional single-sized convolution kernel, the proposed method introduces multichannel convolution technology. More features can be extracted, and the diversity of the convolution kernel is strengthened when multichannel convolution is compared to a single-sized convolution kernel. Additionally, batch normalization and global average pooling are included. The upgraded AlexNet network has thus greatly improved facial expression recognition accuracy.

## 2.3|HOG-ESRs

Studies in the past have demonstrated that the Histogram of Oriented Gradient (HOG) can be used to extract facial features and that the accuracy and robustness of the algorithm [7] can be improved by using ensemble methods.

The suggested approach that is used on the FER-2013 dataset is based on ESRs (ensembles with shared representations) and HOG features. Here, four branches of convolutional neural networks were constructed. A model with four convolutional branches and four convolutional branches is used for each network branch. The CNN model of each network branch is built on the original pixel data. Prior to the mixed feature set advancing to the full connection layer, a histogram of gradient features is added to the final convolution layer. Prior to sending the composite features into the first convolution layer network, also known as the single branch network of ESRs, the convolution layer in the new model of HOG-ESRs links the convolution layer's features to the HOG features. The primary element of the classification task used by the convolution layer is the original pixel data.

## 2.4 | Affdex CNN

According to [1], the API for the Affdex SDK can recognize seven emotion metrics, twenty facial expression metrics, 13 emojis, and four appearance metrics.

More than 4 million facial videos and hundreds of thousands of facial frames made up the training set. This data, which depicts spontaneous, in-person facial expressions made under difficult circumstances like changing lighting, shifting head positions, and variations in facial features caused by race, age, gender, facial hair, and eyeglasses, represents more than 85 different nations. As can be seen, this tool for analyzing facial expressions is significantly more complex than earlier algorithms suggested in this paper. However, it was used to find out more about how closely businesses could use human-made emotion recognition techniques.

## 2.5 | SVM of HOG

SVM is a powerful and widely used supervised learning algorithm with its own principles and techniques. It focuses on finding an optimal decision boundary that maximizes the margin between classes rather than isolating each class separately.

The training algorithm in SVM, known as the optimization algorithm, aims to find the hyperplane with the maximum margin. It does not explicitly decrease the distance between classes. Instead, it optimizes the decision boundary based on the support vectors, which are the data points closest to the decision boundary [8].

Support Vector Machines (SVM) with HOG feature descriptor is a popular approach for object detection and recognition tasks in computer vision.

HOG is a feature descriptor that captures the local shape and gradient information in an image. It divides the image into small cells, computes the gradient magnitude and orientation within each cell, and then constructs a histogram of the orientations within each block of cells. This histogram represents the distribution of edge orientations in that block, providing a concise representation of the local texture.

For each image in the dataset, the HOG feature descriptor is computed. This involves dividing the image into cells, calculating the gradient magnitude and orientation for each cell, and constructing histograms of orientations within each block.

Then, the HOG descriptors are converted into a feature vector representation. This typically involves concatenating the histograms from all the cells and blocks, resulting in a high-dimensional feature vector. Next, the SVM classifier is trained using the feature vectors obtained from the positive and negative examples. The SVM aims to find the optimal hyperplane that separates the positive and negative examples in the feature space.

SVM with HOG feature descriptor has been widely used in various applications such as pedestrian detection, face detection, and object recognition. It provides a robust and effective method for capturing local texture information and making accurate predictions.

# 3 | Experiment and Analysis

## 3.1 | FER-CNN

The test data from the FER2013 dataset [9], which contains 7178 images, is used to test the model. In general, OpenCV makes use of the device's camera to carry out real-time emotion detection. Using the haarcascade frontal-face XML file, the first task is to find a human face and then draw a bounding box around it. The rectangle is first formed around the detected face after the device's image has been preprocessed. The image is then cropped into a frame measuring 48 x 48 pixels, which is then made into a monochromatic version and fed into the model for feature extraction and emotion classification from the available 7 emotions. The matrix for the same is shown in *Table 1* above.

**Table 1. Feature extraction and emotion classification from the available 7 emotions.**

| | Angry | Disgust | Fear | Happy | Sad | Surprise | Neutral |
|---|---|---|---|---|---|---|---|
| Angry | 0.61 | 0.02 | 0.11 | 0.04 | 0.13 | 0.01 | 0.08 |
| Disgust | 0.22 | 0.65 | 0.04 | 0.03 | 0.02 | 0.02 | 0.02 |
| Fear | 0.09 | 0.01 | 0.56 | 0.03 | 0.17 | 0.05 | 0.09 |
| Happy | 0.01 | 0.00 | 0.02 | 0.91 | 0.02 | 0.01 | 0.03 |
| Sad | 0.05 | 0.01 | 0.11 | 0.05 | 0.61 | 0.01 | 0.16 |
| Surprise | 0.02 | 0.00 | 0.08 | 0.05 | 0.02 | 0.81 | 0.02 |
| Neutral | 0.04 | 0.01 | 0.05 | 0.05 | 0.13 | 0.01 | 0.71 |

*Predicted label.
**Overall accuracy: 69.428%.

From the above matrix, different metrics have been evaluated for each label as follows:

**Table 2. Different metrics have been evaluated from Feature extraction and emotion classification from the available 7 emotions.**

| | Accuracy | False Positive Rate | Recall | False Negative Rate | Specificity | Precision | Matthews Correlation Coefficient | F-1 Score |
|---|---|---|---|---|---|---|---|---|
| Angry | 0.8778 | 0.0716 | 0.61 | 0.39 | 0.9283 | 0.0716 | 0.5296 | 0.5980 |
| Disgust | 0.9850 | 0.0083 | 0.65 | 0.35 | 0.9916 | 0.0083 | 0.7484 | 0.7647 |
| Fear | 0.8823 | 0.0683 | 0.56 | 0.44 | 0.9316 | 0.0683 | 0.4979 | 0.5685 |
| Happy | 0.9301 | 0.0416 | 0.91 | 0.09 | 0.9583 | 0.0416 | 0.8171 | 0.8425 |
| Sad | 0.8619 | 0.0816 | 0.61 | 0.39 | 0.9183 | 0.0816 | 0.5079 | 0.5809 |
| Surprise | 0.9682 | 0.0183 | 0.81 | 0.19 | 0.9816 | 0.0183 | 0.8199 | 0.8437 |
| Neutral | 0.8874 | 0.0667 | 0.71 | 0.29 | 0.9333 | 0.0667 | 0.6163 | 0.6729 |

We can see from the above table that the emotion happy has an individual accuracy of 93.01%, and at the same time, the false negative rate is only 0.09, i.e., The FNR is very low, meaning that fewer images belong to the happy state and are classified as some other emotion. Alternatively, we can say that if an image belongs to the happy emotion state, then there is a high chance that it will be classified as happy. For the same emotion, the recall value is 0.91, meaning that most of the images are correctly classified.

Similarly, from the values calculated above, we can clearly see how our model works for identifying a particular emotion label. For the sad emotion, the individual accuracy is not good, and the FNR is also quite high for it, meaning that around 40% of images are falsely classified. Also, the FPR value is highest among the other emotions, meaning that around 8% of images don't belong to the sad category but are classified as sad. Looking at the recall values, it seems that only the images belonging to the "happy" emotion are correctly classified, as the rest of the images belonging to other emotions have lower recall values.

If we want to rank the performances for different categories as per the F1 score, it is found that the classifier works well for surprise emotion, followed by happy, disgust, neutral, anger, sad and fear.

## 3.2 | AlexNet CNN

The confusion matrix [10] is:

**Table 3. AlexNet CNN confusion matrix.**

| | Angry | Disgust | Fear | Happy [11] | Neutral | Sad [11] | Surprised |
|---|---|---|---|---|---|---|---|
| Angry | 90.36 | 4.31 | 0.01 | 0.08 | 1.32 | 0.32 | 1.14 |
| Disgust | 4.79 | 85.32 | 2.19 | 0.01 | 2.41 | 2.01 | 0.60 |
| Fear | 0.14 | 2.16 | 86.25 | 0.01 | 1.54 | 5.07 | 2.38 |
| Happy | 0.01 | 0.01 | 0.35 | 98.87 | 0.72 | 0.14 | 0.01 |
| Neutral | 2.22 | 7.41 | 3.92 | 0.04 | 83.26 | 0.01 | 3.49 |
| Sad | 0.00 | 0.32 | 2.85 | 1.64 | 2.03 | 95.74 | 0.01 |
| Surprised | 3.10 | 1.24 | 0.03 | 0.00 | 1.78 | 1.00 | 92.21 |

*Predicted label.
**Overall accuracy: 90.958 %.

**Table 4. Different metrics have been evaluated from the AlexNet CNN confusion matrix.**

| | Accuracy | False Positive Rate | Recall | False Negative Rate | Specificity | Precision | Matthews Correlation Coefficient | F-1 Score |
|---|---|---|---|---|---|---|---|---|
| Angry | 0.9705 | 0.0171 | 0.926 | 0.07 | 0.9828 | 0.8980 | 0.8975 | 0.9119 |
| Disgust | 0.9557 | 0.0258 | 0.876 | 0.12 | 0.9741 | 0.8466 | 0.8385 | 0.8613 |
| Fear | 0.9730 | 0.0156 | 0.884 | 0.11 | 0.9843 | 0.9021 | 0.8758 | 0.8930 |
| Happy | 0.9948 | 0.0029 | 0.987 | 0.01 | 0.9970 | 0.9823 | 0.9824 | 0.9849 |
| Neutral | 0.9714 | 0.0164 | 0.829 | 0.17 | 0.9835 | 0.8946 | 0.8393 | 0.8609 |
| Sad | 0.9754 | 0.0144 | 0.933 | 0.06 | 0.9855 | 0.9180 | 0.9125 | 0.9255 |
| Surprise | 0.9780 | 0.0128 | 0.928 | 0.07 | 0.9871 | 0.9235 | 0.9133 | 0.9258 |

Out of all the algorithms analyzed in this paper, Alexnet CNN has the highest accuracy. It can be noticed that the individual accuracy of all the emotions is quite good, and at the same time, the FNR values are low, supporting the fact that this model really performs well.

Considering the happy emotion, the FPR and FNR values are lowest among all other emotions, and the specificity, recall, and precision values are highest among all other emotions, meaning that AlexNet's performance for happy emotion images is exceptionally high.

It is noticed that the individual accuracy for neutral emotion is the lowest among all other emotions and has the highest value for FNR, meaning that out of all other emotions, AlexNet's performance is the lowest for this particular emotion, but still, the performance of this model for neutral emotion is better than other models.

From the metrics calculated above, we can see that the FNR for neutral emotion is 0.17 (the highest), meaning that there are images that are neutral but are not classified as neutral; this is happening at a rate of 17%, i.e., out of 100 neutral images, 17 images are not classified as neutral. To rank the classifier on different categories, we'll look at the F-1 scores calculated above; the classifier works well on happy emotions followed by surprise, sad, anger, fear, disgust and neutral.

## 3.3 | HOG-ESRs

The confusion matrix is:

Table 5. HOG-ESRs confusion matrix.

| | Neutral | Happy | Sad | Surprised | Fear | Disgust | Angry |
|---|---|---|---|---|---|---|---|
| Neutral | 89.6 | 2.1 | 6.1 | 1.1 | 0.2 | 0.0 | 0.8 |
| Happy | 2.3 | 94.1 | 0.6 | 1.7 | 0.0 | 0.0 | 0.7 |
| Sad | 21.1 | 2.6 | 73.5 | 0.0 | 0.4 | 0.2 | 3.4 |
| Surprised | 5.4 | 3.2 | 11.4 | 89.3 | 1.2 | 0.0 | 0.4 |
| Fear | 7.2 | 2.4 | 0.0 | 31.1 | 45.3 | 0.0 | 2.5 |
| Disgust | 0.0 | 6.1 | 1.7 | 12.7 | 0.0 | 56.4 | 25.4 |
| Angry | 7.4 | 3.1 | 0.0 | 1.3 | 0.0 | 0.1 | 86.4 |

*Predicted Label.

**Overall accuracy: 76.316%.

Table 5. Different metrics have been evaluated from the HOG-ESRs confusion matrix.

| | Accuracy | False Positive Rate | Recall | False Negative Rate | Specificity | Precision | Matthews Correlation Coefficient | F-1 Score |
|---|---|---|---|---|---|---|---|---|
| Neutral | 0.8816 | 0.0722 | 0.896 | 0.10 | 0.9277 | 0.6736 | 0.7352 | 0.7694 |
| Happy | 0.9454 | 0.0324 | 0.946 | 0.05 | 0.9675 | 0.8283 | 0.8654 | 0.8835 |
| Sad | 0.9428 | 0.0330 | 0.726 | 0.27 | 0.9669 | 0.7877 | 0.7172 | 0.7557 |
| Surprise | 0.8681 | 0.0812 | 0.805 | 0.19 | 0.9187 | 0.6508 | 0.6659 | 0.7198 |
| Fear | 0.9945 | 0.0029 | 0.511 | 0.48 | 0.9970 | 0.9617 | 0.6751 | 0.6681 |
| Disgust | 0.9990 | 0.0005 | 0.551 | 0.44 | 0.9994 | 0.9947 | 0.7131 | 0.7094 |
| Angry | 0.9080 | 0.0551 | 0.878 | 0.12 | 0.9448 | 0.7224 | 0.7604 | 0.7930 |

HOG-ESRs stand fourth in terms of accuracy when analyzed against other methods. From the metrics obtained above, it is clear that this algorithm has the highest individual accuracy for Disgust emotion, but the FNR is also 0.44, meaning that it wrongly classifies a significant number of images. One thing to understand is why the individual accuracy and the FNR are both high at the same time; the reason behind this is that the sum of true negatives and true positives is quite high, i.e., there is a high number of images that do not belong to the Disgust emotion, and they are correctly classified as Not Disgust meaning that these images belong to other emotion tags and thus are not classified as Disgust, due to this reason the individual accuracy is high.

One more thing to notice is that for the "Fear" emotion, the FPR is 0.0029. The FNR is 0.48 (the highest), i.e., fewer images do not belong to the Fear emotion and are classified into the Fear emotion tag. More images belong to the Fear emotion but are not classified into the Fear emotion tag. Due to this, the recall value is only 0.511 (the lowest) in spite of having high individual accuracy and precision.

Here are the rankings based on F-1 scores as follows: happy comes first, followed by angry, neutral, sad, surprised, disgusted and fear.

## 3.4 | SVM of HOG

The Confusion matrix is:

**Table 6. SVM of HOG confusion matrix.**

| | Surprise | Fear | Neutral | Happy [11] | Disgust | Anger | Sad [11] |
|---|---|---|---|---|---|---|---|
| Surprise | 0.87 | 0.03 | 0.04 | 0.04 | 0.00 | 0.02 | 0.00 |
| Fear | 0.01 | 0.98 | 0.00 | 0.01 | 0.00 | 0.00 | 0.00 |
| Neutral | 0.01 | 0.00 | 0.88 | 0.04 | 0.02 | 0.01 | 0.04 |
| Happy | 0.00 | 0.02 | 0.04 | 0.92 | 0.00 | 0.01 | 0.01 |
| Disgust | 0.00 | 0.00 | 0.01 | 0.06 | 0.85 | 0.07 | 0.01 |
| Anger | 0.06 | 0.02 | 0.05 | 0.00 | 0.07 | 0.74 | 0.06 |
| Sad | 0.02 | 0.00 | 0.02 | 0.05 | 0.01 | 0.01 | 0.89 |

*Predicted label.
**Overall accuracy: 87.571%.

**Table 7. Different metrics have been evaluated from the SVM of the HOG confusion matrix.**

| | Accuracy | False Positive Rate | Recall | False Negative Rate | Specificity | Precision | Matthews Correlation Coefficient | F-1 Score |
|---|---|---|---|---|---|---|---|---|
| Surprise | 0.9713 | 0.0166 | 0.870 | 0.13 | 0.9833 | 0.8969 | 0.8642 | 0.8832 |
| Fear | 0.9801 | 0.0116 | 0.980 | 0.02 | 0.9883 | 0.9333 | 0.9489 | 0.9560 |
| Neutral | 0.9545 | 0.0266 | 0.880 | 0.12 | 0.9733 | 0.8461 | 0.8395 | 0.8627 |
| Happy | 0.9438 | 0.0333 | 0.920 | 0.07 | 0.9666 | 0.8214 | 0.8463 | 0.8679 |
| Disgust | 0.9712 | 0.0166 | 0.850 | 0.15 | 0.9833 | 0.8947 | 0.8514 | 0.8717 |
| Angry | 0.9650 | 0.0200 | 0.740 | 0.26 | 0.9800 | 0.8604 | 0.7674 | 0.7956 |
| Sad | 0.9657 | 0.0200 | 0.890 | 0.11 | 0.9800 | 0.8811 | 0.8664 | 0.8855 |

This model stands second in terms of overall accuracy; one thing to notice here is that this model has shown exceptionally high performance in determining the Fear emotion; unlike other models, the individual accuracy, precision, and recall all have comparatively good values that indicate better performance. The FPR and the FNR both values are low indicating that there are fewer images wrongly classified.

We can say that this is the best model for detecting fear emotions; the values of the above metrics show that this model works better for fear emotions than happy emotion images. Also, this model works greatly on Sad emotions as the values of the FNR and the FPR are both convincing, and it has good individual accuracy.

The value of specificity for Angry and Sad emotions is the same, and the values of the FPR (1-specificity) are also the same, meaning that there is an equal ratio of the images that do not belong to these emotions and are not classified into this emotion (true negatives) to the total number of images having another emotion label (not angry or sad).

Rankings of the classifier are as follows:

First comes Fear, then sadness, followed by surprise, disgust, happiness, neutrality and anger.

## 3.5 | Affdex CNN

**Table 8. Affdex CNN confusion matrix.**

| | Accuracy | False Positive Rate | Recall | False Negative Rate | Specificity | Precision | Matthews Correlation Coefficient | F-1 Score |
|---|---|---|---|---|---|---|---|---|
| Happy | 0.9848 | 0.0101 | 0.965 | 0.0345 | 0.9899 | 0.9695 | 0.9567 | 0.9675 |
| Sad | 0.8133 | 0.1320 | 0.848 | 0.1515 | 0.8679 | 0.6816 | 0.6700 | 0.7560 |
| Angry | 0.9521 | 0.0303 | 0.709 | 0.2903 | 0.9697 | 0.8864 | 0.7352 | 0.7882 |
| Fear | 0.9846 | 0.0101 | 0.928 | 0.0714 | 0.9899 | 0.9684 | 0.9316 | 0.9480 |

*Overall accuracy: 86.30%

Rankings for the classifier (f1-score) are as follows:

Happy comes first, then fear, followed by anger and sadness.

# 4 | Conclusion

After making all the necessary calculations, it was discovered that AlexNet CNN has the highest accuracy of all the models, followed by the SVM of HOG, Affdex CNN, HOG-ESRs, FER-CNN, and MLP ANN. For images of happy emotion, it could be argued that AlexNet CNN, Affdex CNN, MLP ANN, and HOG-ESRs perform best, whereas FER-CNN performs best for images that belong to the surprise emotion tag, and SVM of HOG performs best for fear emotion.

**Table 9. Comparison of different models in facial feature recognition.**

| Models | Accuracy |
|---|---|
| AlexNet CNN | 90.958 % |
| SVM of HOG | 87.571 % |
| Affdex CNN | 86.307 % |
| HOG-ESRs | 76.316 % |
| FER-CNN | 69.428 % |


# Acknowledgments

The authors express their gratitude to open-source datasets and computational resources that were essential for performing the real-time evaluations in this study. Special thanks go to the creators of the FER-2013 dataset and the open-access machine learning communities that supported this research.


# Author Contributions

Seyed Ali Noorkhah directed the research design, implemented the algorithmic models (FER-CNN, AlexNet CNN, SVM-HOG), and managed the technical writing and comparative analysis.

Zhang Hao played a key role in refining the experimental setup, formulating evaluation metrics, and integrating the HOG-ESRs and Affdex CNN models, in addition to editing the manuscript.

Haifa Alqahtani focused on the analytical framework and investigated practical business applications of emotion recognition systems. She also provided valuable insights on result interpretation and ensured the final version of the manuscript adhered to academic standards.

All authors contributed equally to the manuscript's development, review, and approval of the final version.


# Funding

This research did not receive support from any external grants or funding agencies. All computational tools and resources employed in this study were provided by the authors' affiliated institutions.


# Data Availability

Further preprocessed data, experimental code, and evaluation scripts can be obtained from the corresponding author upon reasonable request.

# Conflicts of Interest

The authors affirm that there are no conflicts of interest concerning the publication of this paper. The research was carried out impartially and without any commercial or financial influence.

# References


[1] Kartali, A., Roglić, M., Barjaktarović, M., Đurić-Jovičić, M., & Janković, M. M. (2018). Real-time algorithms for facial emotion recognition: a comparison of different approaches. *2018 14th symposium on neural networks and applications (Neurel)* (pp. 1–4). IEEE. https://doi.org/10.1109/NEUREL.2018.8587011

[2] De, A., & Saha, A. (2015). A comparative study on different approaches of real time human emotion recognition based on facial expression detection. *2015 international conference on advances in computer engineering and applications* (pp. 483–487). IEEE. https://doi.org/10.1109/ICACEA.2015.7164792

[3] Murugappan M., & Mutawa A. (2021). Facial geometric feature extraction based emotional expression classification using machine learning algorithms. *PLOS one*, *16*(2), e0247131. https://doi.org/10.1371/journal.pone.0247131

[4] Lee, S., & Park, H. (2023). Effect of optimization techniques on feedback alignment learning of neural networks. *2023 international conference on artificial intelligence in information and communication (ICAIIC)* (pp. 227–231). IEEE. https://doi.org/10.1109/ICAIIC57133.2023.10067047

[5] Zhao, H., Wang, Q., Jia, Z., Chen, Y., & Zhang, J. (2021). Bayesian based facial expression recognition transformer model in uncertainty. *2021 international conference on digital society and intelligent systems (DSINS)* (pp. 157–161). IEEE. https://doi.org/10.1109/DSInS54396.2021.9670628

[6] Zhao, Y., & Chen, D. (2022). Expression recognition using improved alexnet network in robot intelligent interactive system. *Journal of robotics*, *2022*(1), 4969883. https://doi.org/10.1155/2022/4969883

[7] Zhong, Y., Sun, L., Ge, C., & Fan, H. (2021). HOG-ESRs face emotion recognition algorithm based on HOG feature and ESRs method. *Symmetry, 13(2). https://doi.org/10.3390/sym13020228*

[8] Singh S., Singh D., Y. V. (2020). Face recognition using HOG feature Extraction and SVM Classifier. *International journal of emerging trends in engineering research*, *8*(9). https://doi.org/10.30534/ijeter/2020/244892020

[9] Talegaonkar, I., Joshi, K., Valunj, S., Kohok, R., & Kulkarni, A. (2019). Real Time facial expression recognition using deep learning. *Proceedings of international conference on communication and information processing (ICCIP)*. https://dx.doi.org/10.2139/ssrn.3421486

[10] Sujanaa, J., Palanivel, S., & Balasubramanian, M. (2021). Emotion recognition using support vector machine and one-dimensional convolutional neural network. *Multimedia tools and applications*, *80*(18), 27171–27185. https://doi.org/10.1007/s11042-021-11041-5

[11] Zhao, Y., Yang, J., Du, J., Chen, Z., & Yang, W.-C. (2021). A lightweight classifier for facial expression recognition based on evolutionary svm ensembles. *2021 6th international conference for convergence in technology (I2CT)* (pp. 1–9). IEEE. https://doi.org/10.1109/I2CT51068.2021.9417940