Erzincan Üniversitesi
Fen Bilimleri Enstitüsü Dergisi
2025, 18(1), 73-97
ISSN: 1307-9085, e-ISSN: 2149-4584
Araştırma Makalesi

Erzincan University
Journal of Science and Technology
2025, 18(1), 73-97
DOI: 10.18185/erzifbed.1453842
Research Article

Convolutional Neural Network Based Emotion Recognition from Facial Expressions Using Different Feature Engineering Methods

Şengül Bayrak [1*] Fatıma Amiry [2] Anisah Kaso [3] Mina Çakır [4]

[1,3,4]Department of Software Engineering, Faculty of Nature and Science, İstanbul Sabahattin Zaim University, Istanbul, Türkiye

[2]Department of Data Science, Rome Business School, Rome, Italy

Received: 16/03/2024, **Revised:** 02/09/2024, **Accepted:** 19/09/2024, **Published:** 28/03/2025

Abstract

The impact of advancing technology has led to an increased interest in the automatic detection of human emotions in various industries. Emotion recognition systems that use facial images are important for meeting the needs of various industries in a wide range of application areas, such as security, marketing, advertising, and human-computer interactions. In this study, automatic facial expression detection was performed for seven different emotions (anger, disgust, fear, happiness, neutrality, sadness, and surprise) from facial image data. The process of the study was as follows: (i) preprocessing the image data with image grayscale and image enhancement methods, (ii) feature extraction by application of gradient histogram, Haar wavelet transform, and Gabor filtering methods to the preprocessed image, (iii) modeling the feature sets obtained using the three different feature extraction methods using a Convolutional Neural Network method, (iv) calculating the most successful feature extraction method for the detection of the seven different emotions with the Convolutional Neural Network. The experimental results showed that the Gabor filter feature extraction method had an accuracy rate of 83.14%. Comparison of the present results with other studies confirms that the developed model contributes to the literature by improving the recognition rate, dataset size, and feature engineering methods.

Keywords: Gabor filter, Haar Wavelet, Gradient Histogram, emotion recognition from facial expression, Convolutional Neural Network

Farklı Özellik Mühendisliği Yöntemleri Kullanarak Yüz İfadelerinden Evrişimsel Sinir Ağı Tabanlı Duygu Tespiti

Öz

Gelişen teknolojinin etkisiyle, insan duygularının otomatik olarak algılanması çeşitli sektörlerde büyük ilgi görmektedir. Yüz görüntülerinden duygu tanıma sistemleri, güvenlik, pazarlama, reklamcılık ve insan-bilgisayar etkileşimi gibi çok çeşitli uygulama alanlarında çeşitli endüstrilerin ihtiyaçlarını karşılamak için önemlidir. Bu çalışmada, yüz görüntüsü verilerinden 7 farklı duygunun (kızma, iğrenme, korku, mutlu, nötr, üzgün ve şaşkın) otomatik ifade tespiti gerçekleştirilmiştir. Çalışmanın işlem adımları şöyledir: (i) görüntü verilerinin görüntü gri tonlama ve görüntü iyileştirme yöntemleri ile ön işleme uygulanması, (ii) ön işlem uygulanan görüntüye Gradient Histogram, Haar dalgacık dönüşümü ve Gabor filtre yöntemlerinin uygulanarak özellik çıkarımı yapılması, (iii) üç farklı özellik çıkarma yöntemi için elde edilen özellik setlerinin Evrişimsel Sinir Ağı yöntemi ile modellenmesi, (iv) yedi farklı duygunun tespitinde en başarılı özellik çıkarım yönteminin Evrişimsel Sinir Ağı ile hesaplanmasıdır. Yapılan deneysel çalışmalar sonucunda Gabor filtresi özellik çıkarma yönteminin %83,14 doğruluk oranı ile başarılı olduğu tespit edilmiştir. Bu yöntemlerin sonuçları ile diğer çalışmaların sonuçları karşılaştırıldığında, geliştirilen modelin tanıma oranı, veri kümesi boyutu ve özellik mühendisliği yöntemleri açısından fark oluşturarak literatüre katkı sağlamaktadır.

Anahtar Kelimeler: Gabor filtresi, Haar Dalgacığı, Gradyan Histogramı, yüz ifadesinden duygu tanıma, Evrişimsel Sinir Ağı

*Corresponding Author: bayraksengul@ieee.org
Sengul Bayrak https://orcid.org/0000-0002-4114-4305,
FatımaAmiry https://orcid.org/0009-0009-4106-5955,
Anisah Kaso https://orcid.org/0009-0002-6068-9412
Mina Çakır https://orcid.org/0009-0004-5946-1183

1. Introduction

Emotion detection through facial expression analysis is the process of discerning an individual's emotional state by examining the expressions exhibited on their face. These expressions often reveal fundamental emotions, such as happiness, sadness, anger, fear, disgust, and surprise. Emotion detection from facial expressions is significant for a multitude of industries, including security, behavior analysis, and risk assessment, where it aids in threat detection. It also plays a pivotal role in marketing and advertising by enabling the measurement of consumer reactions, the evaluation of product impact, and the optimization of marketing strategies [1].

Paul Ekman, a seminal figure in the emotion detection field, has dramatically influenced various scientific disciplines through his pioneering work on micro-expressions, lie detection, and the correlation between emotions and facial expressions. Ekman's research has included the creation of a facial muscle map that delineates the basic emotions of happiness, disgust, anger, fear, surprise, and sadness (excluding a neutral expression) [2]. This universalized framework for facial expressions has revolutionized human–computer interactions, particularly as recent studies have leveraged artificial intelligence techniques to enable computers to perceive and accurately interpret emotions. These studies have extended the application of emotion detection to individuals with schizophrenia [3], autism spectrum disorder [4], and depression [5].

The importance of emotion detection methods lies in their ability to facilitate accurate mutual comprehension between humans and computers. Facial Expression or Emotion Recognition has emerged as a significant facet of machine learning by offering diverse applications in healthcare, human-computer interaction, and gaming. However, a review of previous research reveals that emotion classification remains challenging, mainly due to the lack of computerized training images for emotions and the static nature of images [6]. Previous studies [6, 7, 8, 9] have underscored the complexities arising from the considerable overlap among Ekman's emotion classes, such as the similarities between fear and surprise in eyebrow movement or the shared characteristics of happiness and anger in eye-contour wrinkling.

In human–computer interactions, facial emotion recognition systems hold promise for personalizing user experiences, analyzing emotional responses, and delivering adaptive interactions. Achieving this potential requires the integration of computer vision, artificial intelligence, and machine learning techniques. As depicted in Table 1, recent research underscores the critical role of feature extraction methods in distinguishing between Ekman's distinct emotions. Consequently, further studies are needed.

1.1.Literature Review

The available literature shows that the first studies on understanding human emotions came from the fields of psychology and sociology, as well as medicine. Ralph Adolphs et al. [1] investigated whether damage to a specific brain region, the amygdala, affects face-based emotion recognition In the last twelve years, difficulties in recognizing facial expressions by computers have been recognized, raising the importance of accurate recognition of Ekman's six emotions, especially with artificial intelligence. Shinde and Pande [10] modeled the features obtained by a 2D Gabor filter with Support Vector Machine (SVM) and Neural Network architecture in an appearance model for preprocessing and feature extraction and considered a hybrid model and application of a neural network classifier. Kaburlasos et al. [11] masked the face image by applying a Viola-Jones face detector and then performed segmentation using an orthogonal moment method with a K-Nearest Neighbour (KNN) classifier and obtained 61.81% accuracy for neutral, 60.42% for fear, 61.36% for happiness, and 51–58% accuracy for other emotions. Piparsaniyan et al. [12] applied Principle Component Analysis (PCA) analysis with a geometric feature-based method for face images and introduced a facial emotion recognition method based on Gabor-based features and an efficient Bayesian classifier for multi-class classification. The proposed method achieved an overall accuracy of 96.73%, especially for the JAFFE database. Burkert et al. [13] obtained 99.6% accuracy on the CKP face dataset and 98.36% accuracy on the MMI dataset using Deep CNN in a 3-step system they proposed using convolution with 64 different filters, pooling normalized by Local Response Normalization (LRN), and two FeatEx blocks. Yu and Zhang [14] modeled the feature set obtained by Standard Histogram Equalization, Linear Plane Fitting, Normalization, and Zero Mean Unit Variance Vector processing steps with CNN and reported accuracies of 68.12% for anger, 21.95% for fear, 83.16% for happiness, 68.97% for a neutral expression, 54.55% for sadness, and 62.16% for surprise. Li et al. [15], by segmenting the facial region into areas, also showed that the method provided good performance for micro-expression detection of emotional states in the CASME spontaneous micro-expression database.

Matlovic et al. [16] used EEG data to capture participants' brain activity and achieved 53% accuracy in emotion classification. Xiang and Zhu [17] exploited the intrinsic correlation between face detection and facial expression recognition and concluded that the reliability coefficient of Face Emotion Recognition (FER) based on face detection with a Multi-Task Cascaded Convolutional Networks (MTCNN) model is low for both patient and healthy participant groups. Greche et al. [18] successfully used the L1 norm (Manhattan distance) and the L2 norm (Euclidean distance) for newly labeled data with 0% redundancy with the face tracking of the Kinect sensor. Kumar et al. [19] improved contrast using the Haar Wavelet Transform method with a genetic algorithm and Fuzzy-C Means and achieved a detection rate for phrase recognition of nearly 100%. Chang et al. [20], who used image intensities and ExpoNent CNN imaging conditions to develop a 3D morphable model for the data obtained with EmotiON-17 emotion recognition tests, reported much faster results than those achieved with ExpoNent alternatives. Jain et al. [21] produced an extended deep neural network for facial emotion recognition that yielded average results similar to those reported in other studies in the literature. Xie et al. [22] proposed a model using a deep multipath convolutional neural network with CNN and obtained 43.38% accuracy for JAFFE data, 24.44% for SFEW data, 49.10% for CK+ data, 52.75% for TFEID data.

Hammed et al. [23] achieved 50% accuracy using filters for noise reduction, localizing and extracting the face region for face detection, normalizing the color and size of the images, and enhancing the image with Histogram Equalization. Porcu et al. [24] achieved 83.30% accuracy with a VGG16 CNN model with data augmentation. Tsai et al. [25] developed CNN and an adaptive exponentially weighted average ensemble (AEWAE) model with a Face Frontalization method and a Face cropping algorithm and achieved accuracies of 62.19% for anger, 96.55% for disgust, 71.31% for fear, 86.36% for happiness, 56.23% for sadness, 84.06% for surprise, and 67.52% for a neutral expression. Almeida and Rodrigues [26] used a VGG 16-CNN model with a Haar-like feature selection technique and found accuracies of 90.6% for anger, 96.55% for disgust, 77.1% for fear, 100% for happiness, 86.2% for sadness, 92.5% for surprise, 82.1% for a neutral expression, 91.8% for a stress condition, and 92.2% for an unstressed condition. Shabbir and Rout [27] obtained 65.08% accuracy with a DCNN Optimized Tuning Strategy model with data augmentation. Kadakia et al. [28] used the GrabCut algorithm to reduce noise in the dataset and achieved accuracies of 99% for anger, 95% for fear, 100% for happiness, 94% for sadness, 98% for surprise, and 82.1% for a neutral expression using a combination of VGG16 deep learning and a Local Interpretable Model-Agnostic Explanations (LIME) model. Lee et al. [29] developed pre-trained VGG-16, VGG-19, and Xception models with the ImageNet database to attain a final average accuracy of 93.56±1.38%. Yaddaden et al. [30] modeled the feature set obtained from Local Binary Pattern and HOG using principal component analysis and locally linear embedding methods with a multi-class SVM classifier and obtained an accuracy rate of 87.26–95.49%.

These past studies in the field of facial expression recognition have highlighted several challenges, such as insufficient training data and the overlap of emotions with similar expressions. In particular, the large overlap between the classes proposed by Ekman makes classification difficult. In the present study, we aimed to extract meaningful features that would enable us to distinguish between seven different emotions using different techniques. Our goal HOG feature extraction methods with CNN architecture, with the belief that a CNN model could reveal a feature extraction method that would give the most meaningful features for all seven different emotions and that would overcome the current classification difficulties. The critical studies referenced for this research included one that employed Histogram Equalization for detecting local data in facial regions using filters [36]. The CNN architectures modeled for detecting different emotional expressions also provided valuable insights for this study [37-39] and served as significant references in this research [40-42].

Table 1: Literature review on emotion detection from facial expression.

Study	Model	Result (%)
[31]	SVM	93.00 Anger: 100 Disgust: 96.77 Fear: 93.54
[11]	Bayesian	Happy: 96.77 Neutral: 96.77 Sad: 96.66 Surprise: 96.77
[12]	CNN	CKP: 99.60 CK+ dataset: 95.60
[13]	Genetic Algorithm + Fuzzy C-means	JAFEE dataset: 88.08
[32]	Sparse Representation-based Classification (SRC)	Average of scores:94% CK+: 43.38 JAFFE: 99.32 TFEID (6 Class): 93.65% TFEID(7 Class): 93.36 FER 2013:71.10
[15]	Deep Attentive Multi-path CNN	SFEW: 42.30 BAUM-2i(6 class):67.92 BAUM-2i(7 class):61.52
[16]	Generative Adversarial Network	83.30
[33]	SVM	CK :97.14 JAFFE : 92.53 TFEID : 98.9% CK+ :97.94% MMI : 83.12%
[34]	SVM	TFEID: 97.01, JAFFE:98.59, KDEF:96.54, CK+:100, Oulu-CASIA :100
[19]	Transfer learning in DCNN	KDEF dataset: 76.24 SFEW dataset: 52.26
[35]	SVM	78.37

This study leveraged a dataset comprising 14,000 images representing seven emotions and employed Gabor filtering, Haar wavelet, and HOG preprocessing techniques. The extracted features were subsequently subjected to CNN modeling, allowing us to identify the most effective method for detecting the seven distinct emotions.

The remainder of this article unfolds as follows: Section 2 introduces our proposed methodology, Section 3 presents our experimental results, and the subsequent sections provide a discussion, conclusions, and future research directions.

2. Material and Methods

2.1. Dataset

In this study, we utilized a FER dataset, comprising 14,000 image samples with dimensions of 48×48 pixels and encompassing seven emotional expressions: anger, disgust, fear, happiness, neutrality, sadness, and surprise. The dataset maintains a balanced distribution, containing precisely 2000 images for each emotion class [43].

2.2.Feature Engineering Process

We initially performed image grayscale conversion and applied image enhancement procedures to facilitate feature extraction from the dataset. These enhancement techniques involve filtering operations aimed at accentuating specific image details. Following these preprocessing steps, the resulting images were subjected to three distinct feature extraction methods: Gabor filtering, Haar wavelet transformation, and HOG. Our aim was to identify the most effective technique for extracting significant features representing emotional expressions within facial images.

2.2.1. Feature Engineering Process with Gabor Filter

After applying the preprocessing steps to the data, the Gabor filter was used to extract features from the data. The Gabor filter is a linear filter formed by the product of a harmonic function and a Gaussian function.

$$x' = x\cos\theta + y\sin\theta \tag{1}$$

$$y' = -x\sin\theta + y\cos\theta \tag{2}$$

From equation (1) and equation (2), the Gabor filter is implemented by equation (3).

$$g(x, y; \lambda, \theta, \psi, \sigma, \gamma) = \exp\left(\frac{(x'^2 + \gamma^2 2\, y'^2)}{2\theta^2}\right) \cos\left(\frac{2\pi\,(x')}{\lambda} + \psi\right) \tag{3}$$

where λ is the wavelength factor of the cosine value, θ is the direction of the Gabor function, ψ is the phase offset, and γ is the spatial angle of view. When Gabor filters with various orientations are applied to the target image, they produce different angular components derived by convolving the filtered images with the target image and averaging them [44, 45]. Figure 1 illustrates the outcome of applying Gabor filters to a sample image in our study.

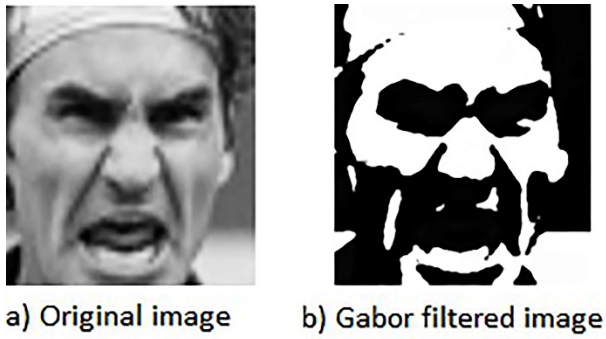

Figure 1. A sample of Gabor Filter image.

2.2.2. Feature Engineering Process with Haar Wavelet

The Haar wavelet transform is a fundamental and straightforward wavelet transform technique with extensive applications, particularly in image processing. The Haar wavelet is non-continuous, rendering it non-differentiable, but it excels in the analysis of signals featuring abrupt transitions. The Haar transform involves a staged sampling of matrix rows that enables low-resolution samples to effectively capture the localized features inherent in high-resolution signals [46]. The Haar wavelet's mother wave function ψ(t) can be defined as in Equation 4.

$$\psi(t) = \begin{cases} 1 \quad 0 \leq t < \frac{1}{2}, \\ 1 \; \frac{1}{2} \leq t < 1, \\ 0 \; otherwise \end{cases} \tag{4}$$

The scaling function φ(t) can be defined in Equation 5.

$$\phi(t) = \begin{cases} 1 \quad 0 \leq t < \frac{1}{2}, \\ 1 \; \frac{1}{2} \leq t < 1, \\ 0 \; otherwise \end{cases} \tag{5}$$

In terms of image decomposition, the algorithm used in one dimension transforms a two-element vector into $[x(1), x(2)]^T$ and $[y(1), y(2)]^T$ as in Equation 6.

$$\begin{bmatrix} y(1) \\ y(2) \end{bmatrix} = T \begin{bmatrix} x(1) \\ x(2) \end{bmatrix} \quad where\ T = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \tag{6}$$

A matrix with rows that are orthogonal to each other is orthogonal. The inverse of this type of matrix is the same as its transpose; therefore, $T -1 = T\ TT$ The vector x can be recovered by associating it with the vector y in Equation 7 [47].

$$\begin{bmatrix} x(1) \\ x(2) \end{bmatrix} = T^T \begin{bmatrix} y(1) \\ y(2) \end{bmatrix} \tag{7}$$

In this study, using the 2D data in Equation 8, the 2×2 matrices x and y are transformed by *T*. First, the columns of the *x* matrix are multiplied by *T*, and then the rows of the result are multiplied by *T* to obtain *y = TxTT*. In the next step, *x = TTyT* and *y = TxTT* are found.

$$x = TT_yT \tag{8}$$

Figure 2. A sample of Haar Wavelet image.

To apply this transformation to the entire image, pixels are grouped into 2×2 blocks, and Equation 7, along with the subsequent steps, is applied to each block. To visualize the outcome, the top-left components of y within the 2×2 blocks are aggregated to create the top-left sub-image shown in Figure 2. The same procedure was repeated for the components in the other three positions.

2.2.3. Feature Engineering Process with HOG

The gradient orientations in each region of an image are computed to calculate the shapes and patterns of the objects. The magnitude (g) and orientation θ of each pixel are then calculated as in Equation 9. Equation 10 represents the vertical gradient, and g_x is the horizontal gradient [48, 49].

$$q = \sqrt{g_x^2 + g_y^2} \tag{9}$$

$$\theta = arctan\frac{g_y}{g_x} \tag{10}$$

The image is divided into cells to compute the spatial information of the oriented gradients and to provide robustness to noise. The gradient magnitudes of the pixels for each cell are then calculated. Since the values of the histograms are based on the magnitudes of the gradients, any change directly affects the values of the histograms. Therefore, block normalization is applied to the histograms. The normalized histograms are then combined, as in Equation 11, to create the final feature vector. In Equation 11, b denotes the number of blocks and $\oplus$ stands for concatenation. Figure 3 shows the result of applying HOG to a sample image.

$$f^{HOG} = [f_1^{HOG} \oplus \cdots \oplus f_b^{HOG}] \tag{11}$$

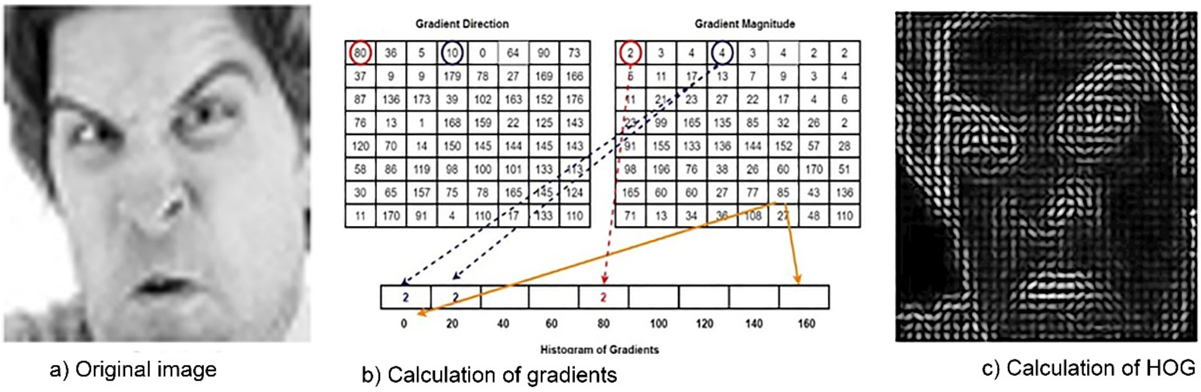

Figure 3. A sample of HOG chart.

2.3. CNN Architecture

The most commonly used CNN architectures for analyzing visual images are specifically designed to process pixel data and utilize a mathematical operation known as a convolution in at least one layer, rather than relying on general matrix multiplication, as shown in Figure 4 [41, 42]. This study employs a specialized CNN for emotion detection from facial images. This type of custom CNN tailors the layers, filters, and hyperparameters to address the specific requirements of the task, often resulting in a model that is more efficient and better suited to the dataset.

In this work, the input layer receives image data, reshaped to a fixed size of 48×48 pixels for each feature extraction method. The convolutional layers extract features from the input image by applying filters. The numbers and sizes of the filters, along with the stride and padding values, were customized for our specific problem. Nonlinear activation functions, typically Rectified Linear Unit (RELU), were applied after each convolution to introduce nonlinearity into the model. Pooling layers were then employed to reduce the spatial dimensionality of the data, thereby reducing the computational load and mitigating overfitting. Following the convolutional and pooling layers, the extracted features were passed into fully connected layers, ultimately mapping the learned features. Dropout layers were incorporated to prevent overfitting and randomly deactivating neurons during training. The output layer employed a softmax activation function for classification tasks to produce the final prediction across seven classes.

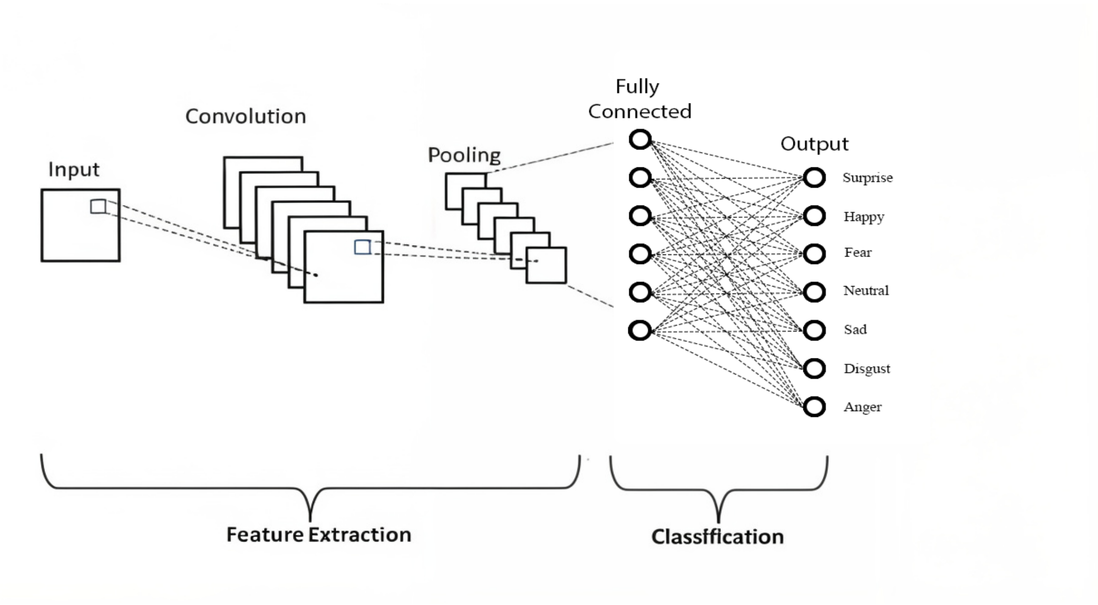

Figure 4. The general CNN architecture.

2.4. Evaluation Metrics

This study used accuracy, precision, sensitivity, and F1-score values to evaluate the classification performance based on the feature sets [50, 51].

Accuracy is the ratio of correctly predicted observations to total observations. It is one of the most commonly used and intuitive metrics, mainly when the data are balanced, as in Equation 12.

$$Accuracy = \frac{True\ Positives + True\ Negatives}{Total\ Observations} \tag{12}$$

Precision is the ratio of correctly predicted positive observations to the total predicted positive observations. It assesses the accuracy of the model's optimistic predictions, as in Equation 13.

$$Precision = \frac{True\ Positives}{True\ Positives + False\ Positives} \tag{13}$$

Recall is the ratio of correctly predicted positive observations to the total number of observations in the actual positive class. This indicates how well the model captures positive instances, as in Equation 14.

$$Recall = \frac{True\ Positives}{True\ Positives + False\ Negatives} \tag{14}$$

The F1-score is the harmonic mean of precision and sensitivity. Balancing these two metrics provides a particularly useful performance measure, especially in imbalanced datasets, as shown in Equation 15.

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{15}$$

A confusion matrix was used to depict the performance of each model in terms of correct and incorrect predictions for each class in this study.

2.5. The application steps

The process steps applied in this study are shown in Figure 5.

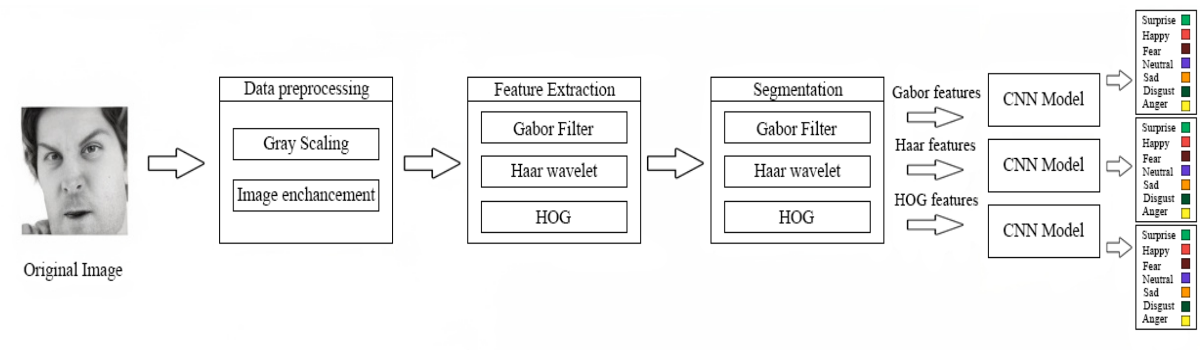

Figure 5. Experimental flow chart.

Figure 5 shows an original image with a facial expression obtained in the first step. In this step, color images were converted to grayscale and enhanced with various techniques to improve image quality. In the second step, feature extraction was performed. The methods used were Gabor filtering, Haar wavelet transformation, and HOG. These methods were applied to extract various meaningful features from the facial expression in the image. In the third step, the image was segmented according to the extracted features. Segmentation was also performed using Gabor, Haar, and HOG. The extracted and segmented feature sets (Gabor, Haar, and HOG features) were trained with a custom CNN model. The results of each model classified the emotions shown (surprise, happiness, fear, neutrality, sadness, disgust, and anger). For each method, the accuracy and success results are also provided. This flow analyzes facial expressions using different preprocessing and feature extraction techniques to perform emotion recognition using CNN. The goal is to achieve the best results with each method.

3.Results and Discussion

As shown in Figure 6, the segmentation process was applied to images with a Gabor filter, Haar wavelet, and HOG. Because emotional expressions are usually concentrated in some areas of the face, segmentation becomes an essential step in that it separates different regions of the face, such as the eyes, nose, and mouth, thereby allowing for a more detailed analysis of expressions in these regions.

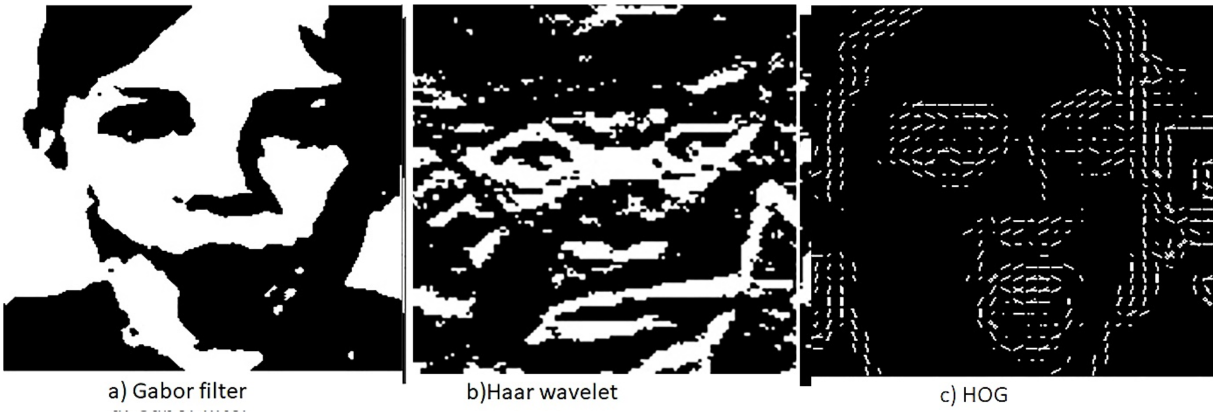

Figure 6. Segmentation result obtained from an image using Gabor filter (a), Haar wavelet (b), HOG (c) methods.

In Figure 6, a sample image with feature engineering processes applied has been segmented after applying Gabor filter (a), Haar wavelet (b), and HOG (c) methods.

3.1. Modeling Results with CNN

The image size of the three feature extraction methods in CNN modeling is 48 × 48 pixels. After modeling CNN on our dataset of 14,000 images (consisting of 2000 images from each of the seven basic emotion classes), 70% were used for training and 30% for testing. To prevent the model from overfitting during training and to evaluate the overall performance of the model, 30% of the training data were divided into validation data. Therefore, 70% of the training data were used for training the model and 30% for validation. The hardware specifications for the computer used for the experimental studies were an Intel(R) Core (TM) i7-6600U processor with a base frequency of 2.60 GHz and a maximum speed of 2.81 GHz. The system had 16 GB of installed RAM (15.6 GB usable). The operating system was Windows 10 Pro, 64-bit architecture, version 19045.4780. Spyder 3.5 was used for the software environment during the experimental studies. When the validation loss ceased to improve, the training process was stopped using the early stopping feature of the models with 100 iterations, which helped prevent the model from overfitting the training data. This saved computational resources and time while still achieving a well-performing model. In Table 2, the values given are the optimum values that provided the most successful results for CNN modeling in this study.

Table 2. Optimum parameter values for CNN modeling.

Parameters	Value
Convolutional kernel size	(3, 3) (5, 5), (3, 3)
Number of convolution units	64, 128, 512, 512
Max pooling kernel size	(2, 2)
Number of epochs	50
Activation function	RELU
Optimizer	Adam
Learning rate	0.001
Initial learning rate	0.1
Dropout rate	0.25
Weight decay	0.0001
Fully connected layers (number of neurons)	256, 512
Fully connected layers (number of layers)	2

3.1.1.CNN modeling results of the feature set obtained with Gabor filter

The Gabor-filtered images were segmented and then modeled with CNN. The model performance is shown in Figure 7.

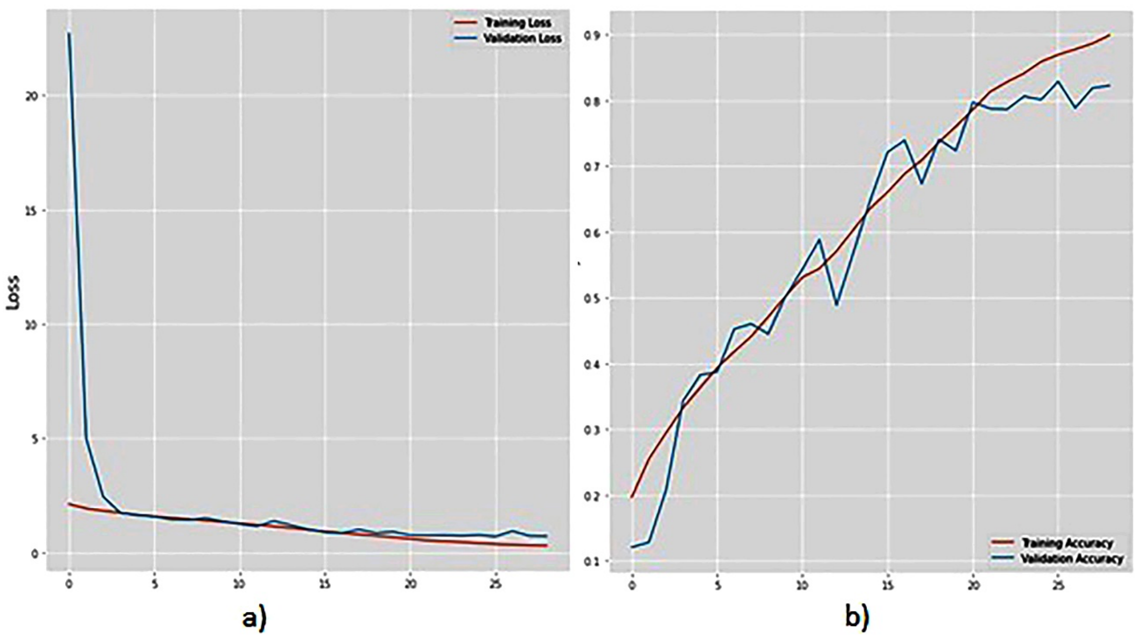

Figure 7. (a) Loss function for Training and Validation Accuracy, (b)Results of the Training Accuracy - Validation Accuracy.

As shown in Figure 7(a), training loss and validation loss decreased as the iteration continued. As depicted in Figure 7(b), training accuracy and validation accuracy increased as the number of iterations increased during the training process.

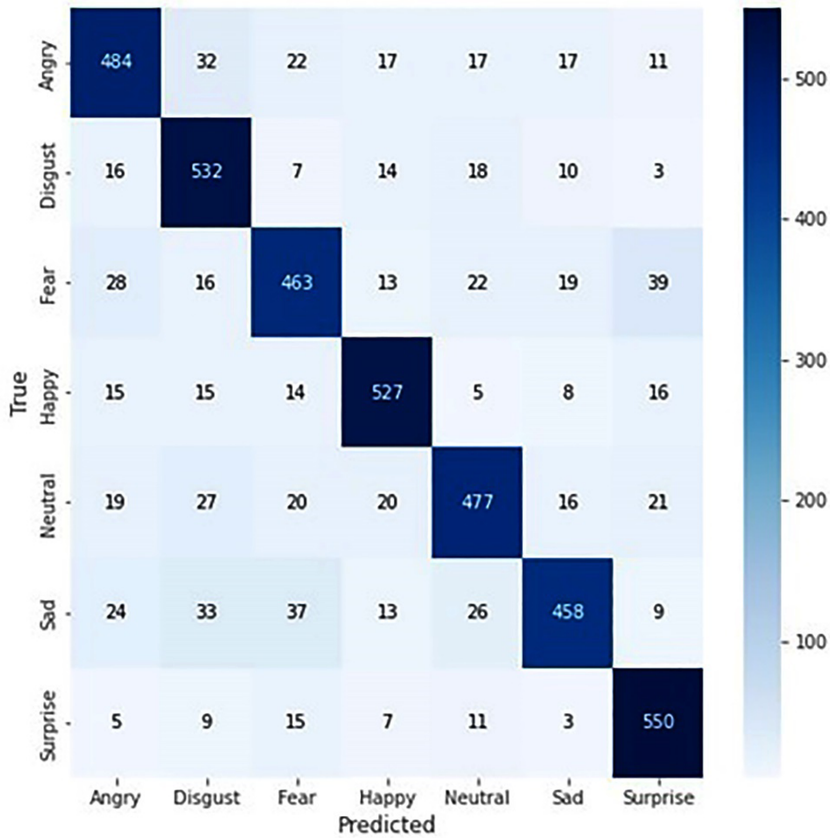

Figure 8. Confusion matrix for the Gabor filter.

Figure 8, which shows the modeling of the discriminative features obtained from facial image data with the Gabor filter with CNN, indicates matching between the angry class and disgust class, the fear class and surprise class, and the sad class and fear class. The most successful test was achieved with 92% accuracy for the surprise class. The average test performance of all seven classes was 83.14%.

3.1.2.CNN Modeling Results of the Feature Set Obtained with the Haar Wavelet

After applying the Haar wavelet, the images were segmented and modeled using CNN. The model performance is shown in Figure 9.

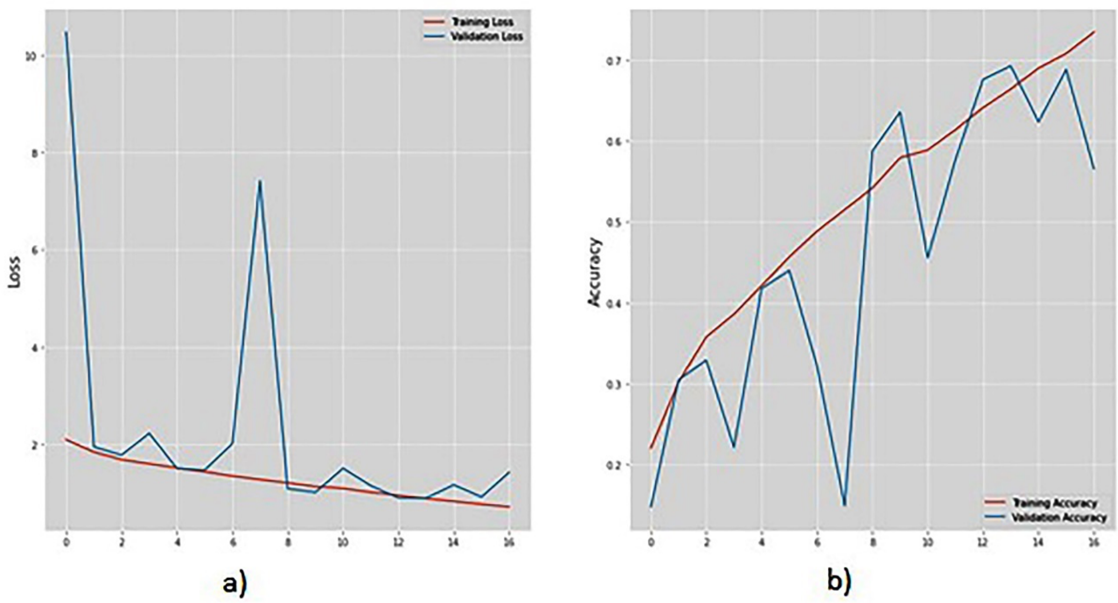

Figure 9. (a) Loss function for Training and Validation Accuracy (b)Results of the Training Accuracy - Validation Accuracy.

Figure 9(a) shows that, as the iteration continued, the training loss decreased, whereas the validation loss was volatile. Figure 9(b) shows that training accuracy and validation accuracy fluctuated as the number of iterations increased during the training process.

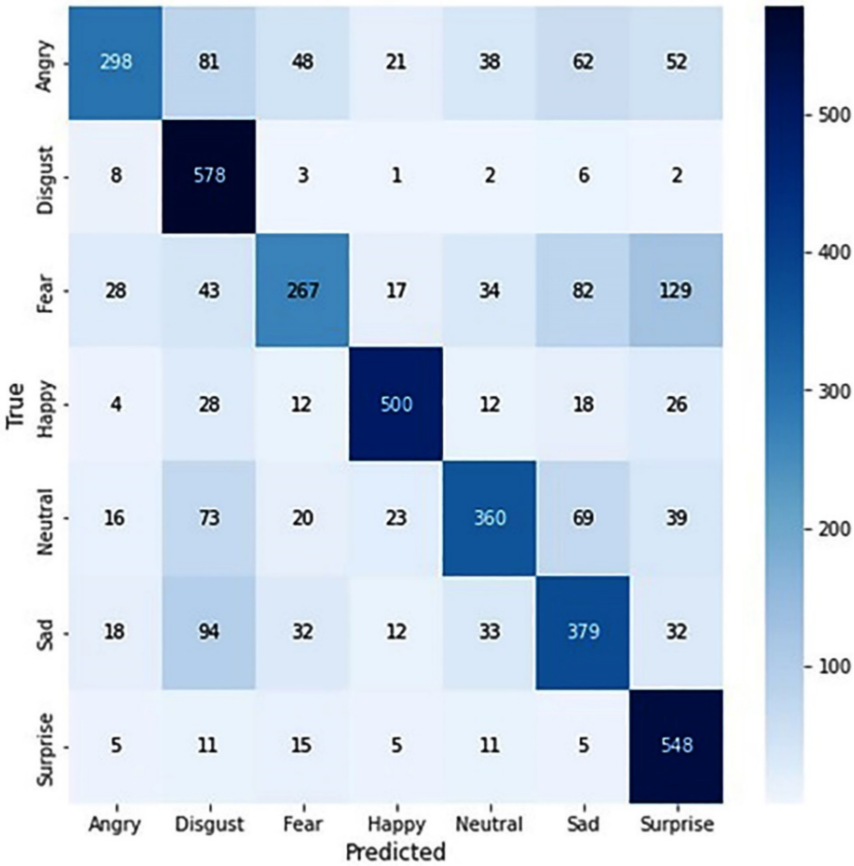

Figure 10. Confusion matrix for the Haar wavelet.

As shown in Figure 10, modeling the discriminative features obtained from facial image data with the Haar wavelet method with CNN gave mixed results for the angry class and disgust class, the fear class and surprise class, and the neutral, disgust, and sad classes. The most successful test was obtained for the emotion of disgust, with 96.33% accuracy. The average test performance of all seven classes was 69.71%.

3.1.3. CNN Modeling Results of the Feature Set Obtained with HOG

The model performance, after applying the HOG method, segmenting the resulting images, and modeling with CNN, is given in Figure 11.

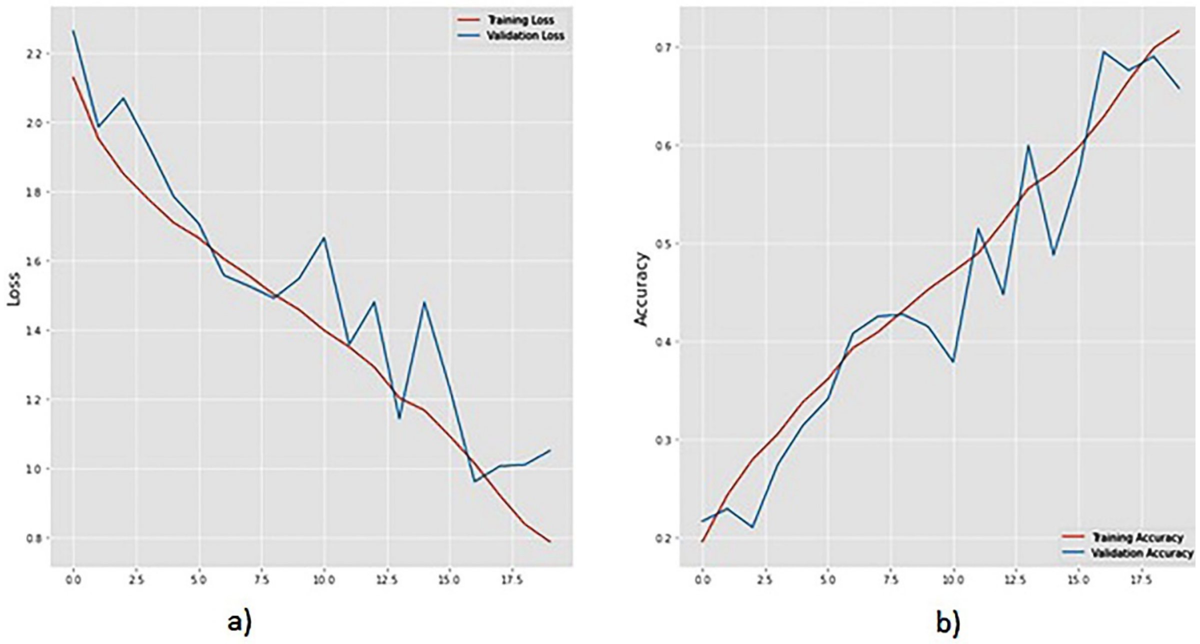

Figure 11. (a) Loss function for Training and Validation Accuracy (b)Results of the Training Accuracy - Validation Accuracy.

Figure 11(a) shows that the training loss decreased steadily as the iteration continued, whereas the validation loss fluctuated. Figure 11(b) shows that training accuracy increased and validation accuracy fluctuated as the number of iterations increased during the training process.

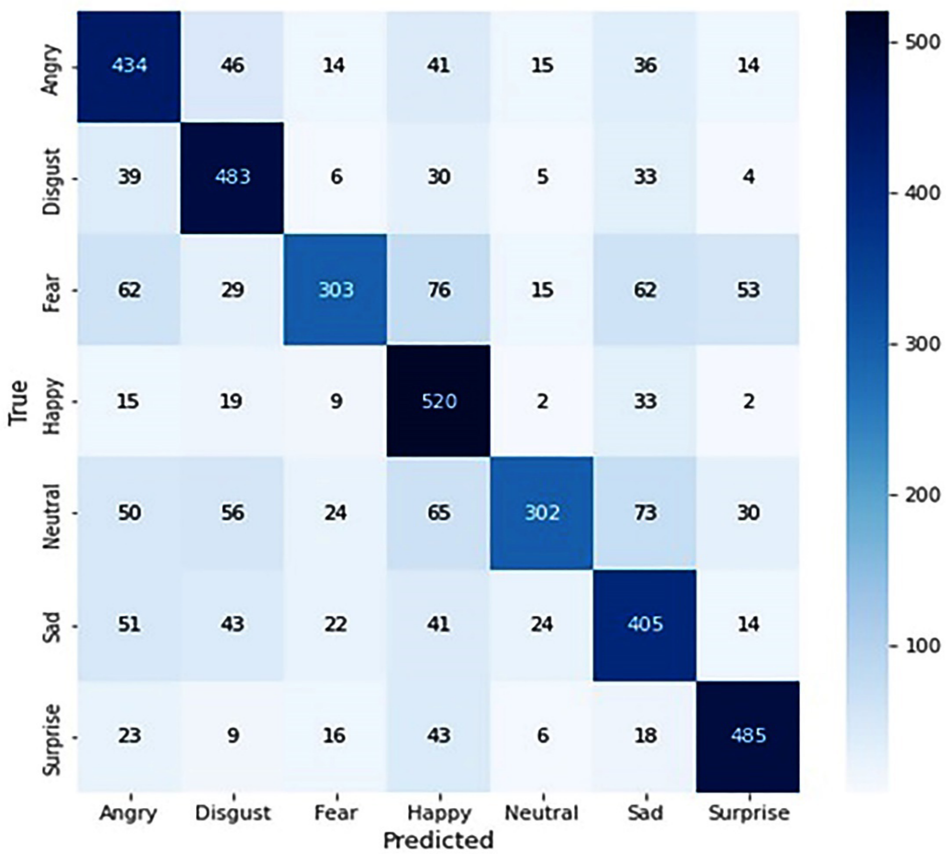

Figure 12. Confusion matrix for HOG.

As shown in Figure 12, as a result of modeling the discriminative features obtained from facial image data with the HOG method and CNN, the angry class was confused with the disgust class, the fear class was confused with the angry and sad classes, the neutral class was confused with the happy class, and the surprised class was confused with the happy class. The most successful test was achieved for the emotion of happiness, at 87% accuracy. The test performance for all seven classes was 69.85%.

3.2.Analysis of Experimental Results

Feature extraction methods, such as HOG, Haar wavelet transformation, and Gabor filtering, use structural information, such as edges, texture, and shape. Color information is usually unimportant with these methods; therefore, graying the images is important for improving the accuracy performance of the CNN algorithm by ensuring that significant structural information in the image is effectively extracted. In addition, color images take considerably longer to process because more information is contained within them. Graying reduced the computational complexity and made the image-processing process more efficient.

Table 3. Gabor Filter, Haar Wavelet, HOG test performances

	Gabor Filter				Haar Wavelet				HOG			
Emotions	Accuracy	Precision	Recall	F1-score	Accuracy	Precision	Recall	F1-score	Accuracy	Precision	Recall	F1-score
Angry	0.81	0.82	0.81	0.81	0.50	0.79	0.50	0.61	0.72	0.64	0.72	0.68
Disgust	0.89	0.80	0.89	0.84	**0.96**	0.64	**0.96**	0.77	0.80	0.71	0.81	0.75
Fear	0.77	0.80	0.77	0.79	0.45	0.67	0.45	0.54	0.51	0.77	0.51	0.61
Happy	0.88	**0.86**	0.88	0.87	0.83	**0.86**	0.83	**0.85**	**0.87**	0.64	**0.87**	0.73
Neutral	0.79	0.83	0.80	0.81	0.60	0.73	0.60	0.66	0.50	**0.82**	0.50	0.62
Sad	0.76	**0.86**	0.76	0.81	0.63	0.61	0.63	0.62	0.68	0.61	0.68	0.64
Surprise	**0.92**	0.85	**0.92**	**0.88**	0.91	0.66	0.91	0.77	0.81	0.81	0.81	**0.81**
Overall Accuracy	**83.14**				69.71				69.85			

As shown in Table 3, modeling with the Gabor filter resulted in the happy class having the most successful precision value, with a success rate of 0.86. The recall value was 0.92 for the surprise emotion, and the F-1 score was 0.88, indicating that surprise was the most successfully modeled emotion. As shown in Table 3, modeling with the Haar wavelet transformation resulted in the happy class having the most successful precision value at 0.86. A recall value of 0.96 was obtained for the disgust emotion. The F-1 score value was 0.85 for the happy class, indicating it was the most successfully modeled emotion. Table 3 shows that the neutral class had the best precision value, at 0.82, when modeled with HOG. The recall value was 0.87 for the happy emotion. The F-1 score value was 0.81 for the surprise class, indicating the most successful modeling. Table 3 shows the overall test performances obtained using the Gabor filtering, Haar wavelet transformation, and HOG methods.

Table 4. Comparison of recognition rates (%) of different data size and emotions.

Study	Data size	Classes	Accuracy (%)
[12] (2014)	210 images	7	96.75
[13] (2015)	2900 videos	6	98.63
[19] (2018)	210 images	7	88.08
[22] (2019)	327 images	7	43.38
[24] (2020)	4900 images	7	83.30
[33] (2018)	2900 videos	6	83.12
[27] (2023)	4900 images	7	76.24
[35] (2017)	1470 images	7	78.37
In this study	**14,000 images**	**7**	**83.14**

As depicted in Table 4, Piparsaniyan et al. [12] worked on resized images and 210 images for the seven emotions using 30 images in each class. Burkert et al. [13] studied two datasets: the MMI dataset had anger, disgust, fear, happiness, sadness, and surprise emotions, accounting for 2900 videos, while the CKP dataset had only 210 images depicting anger, disgust, fear, happiness, sadness, surprise, and contempt. Comparison of the results obtained with these methods to those of the present study confirms that the model developed here contributes to the literature by making a difference in recognition rate, dataset size, and feature extraction methods. The aim of this study was to detect emotions of anger, disgust, fear, happiness, sadness, surprise, and neutrality from facial expressions in images preprocessed with Gabor, HOG, and Haar wavelet methods to capture specific details about emotions from facial expressions.

4.Conclusion

In modeling the features obtained using the Gabor filter, Haar wavelet transform, and HOG feature extraction methods with CNN, the standard feature extraction for all three classes indicated confusion between anger and disgust during the test process. Emotion detection from facial expressions was successfully achieved using Gabor Filter, HOG, and Haar Wavelet methods for the fearful, happy, angry, and sad classes. The most successful feature extraction methods for surprise, neutral, and disgust emotions were the Haar wavelet, Gabor Filter, and HOG, respectively.

These three methods were chosen for this study's preprocessing steps for an emotion recognition system mainly because each has different advantages. For example, the Gabor filter effectively captures fine details in facial images and localizes the images in the frequency and orientation domains. Haar wavelet filters effectively capture basic shape information in face images and detect simple patterns, such as edges and corners. Based on the test performances for the Gabor filter, Haar wavelet, and HOG methods, Ekman's psychological results [2] can be calculated mathematically using the Gabor filter.

Previous studies [13, 33] have conducted six-class modeling, whereas seven-class problems were addressed in the present study. Other studies [12, 19] have processed seven emotions; however, the dataset size in the present study was approximately 7 times larger than the datasets used previously [12, 19]. Larger datasets and more emotion classes are needed to obtain better models in the future. Our group has planned further studies on this subject.

Haar wavelet filters have been chosen as they effectively capture basic shape information in face images and detect simple patterns such as edges and corners. The test performances were for the Gabor filter, Haar wavelet, and HOG. According to the results obtained, Ekman's psychological results [2] can be calculated mathematically using the Gabor filter.

In this study, a custom CNN was employed for emotion recognition from facial expressions by explicitly targeting and detecting facial landmarks and expressions. The network presented here was designed with layers that progressively focused on more minor, finer-grained features as the depth of the model increased. This custom CNN is a flexible, task-specific convolutional network that is optimized for the specific requirements of this type of emotion detection research and offers greater control over the model's architecture compared to pre-trained models. Comparison of the present results with those derived using other methods reveals that our methodology improves the recognition rate, dataset size, and feature extraction. These results further motivate us to apply our methodology for representation purposes the next studies. In the future, our aim is to compare various architecture models, such as AlexNet, DenseNet, Resnet, and Xception.

Ethics in Publishing

There are no ethical issues regarding the publication of this study.

Author Contributions

Şengül Bayrak was responsible for all the experimental studies. Fatima Amiry modeled the HOG experimental phase with CNN. Anisah Kaso was responsible for the Haar wavelet method and CNN modeling. Mina Çakır modeled CNN with the Gabor filtering method.

References

[1] Ralph Adolphs, Leonard Mlodinow, and Lisa Feldman Barrett What is an emotion?, Current Biology Magazine, 2019.

[2] Saraa Clemente Paul Ekman'a Gore Mikro-ifadeler, 2022.

[3] Turetsky Goossens, B. I., Kohler, C. G., Indersmitten, T., Bhati, M. T., Charbonnier, D., Gur, R. C. Facial emotion recognition in schizophrenia: when and why does it go awry Schizophrenia research, 2007, pp. 94(1- 3), 253-263.

[4] Madeline B. Harms, Alex Martin Gregory L. Wallace "Facial Emotion Recognition in Autism Spectrum Disorders: A Review of Behavioral and Neuroimaging Studies", 2010.

[5] Ian M. Anderson, Clare Shippen,Gabriella Juhasz, Diana Chase, Emma Thomas, Darragh Downey, Zoltan G. Toth,Kathryn Lloyd-Williams, Rebecca Elliott and J. F. William Deakin "State-dependent alteration in face emotion recognition in depression.", 2018.

[6] Neha Jain, Shishir Kumar, Amit Kumar, Pourya Shamsolmoal and iMa- soumeh Zareapoor "Hybrid deep neural networks for face emotion recogni- tion.", 2018.

[7] Deepak Kumar,Jain Pourya, Shamsolmoali Paramjit and Sehdev "Extended deep neural network for facial emotion recognition.", 2019.

[8] Hongli Zhang, Alireza Jolfaei, And Mamoun Alazab "A Face Emotion Recognition Method Using Convolutional Neural Network and Image Edge Computing.", 2019.

[9] Liyanage C. De Silva, Tsutomu Miyasato, Ryohei Nakatsu "Facial Emotion Recognition Using Multi-modal Information.", 1997.

[10] Shinde, S., Pande, S. (2012). A survey on: Emotion recognition with respect to database and various recognition techniques. *International Journal of Computer Applications*, *58*(3), 9-12.

[11] Kaburlasos, V. G., Papadakis, S. E., Papakostas, G. A. (2013). Lattice computing extension of the FAM neural classifier for human facial expression recognition. *IEEE Transactions on Neural Networks and Learning Systems*, *24*(10), 1526-1538.

[12] Piparsaniyan, Y., Sharma, V. K., Mahapatra, K. K. (2014, April). Robust facial expression recognition using Gabor feature and Bayesian discriminating classifier. In *2014 International Conference on Communication and Signal Processing* (pp. 538-541). IEEE.

[13] Burkert, P., Trier, F., Afzal, M. Z., Dengel, A., Liwicki, M. (2015). Dexpression: Deep convolutional neural network for expression recognition. *arXiv preprint arXiv:1509.05371*.

[14] Yu, Z., Zhang, C. (2015, November). Image based static facial expression recognition with multiple deep network learning. In *Proceedings of the 2015 ACM on international conference on multimodal interaction* (pp. 435-442).

[15] Li, X., Yu, J., Zhan, S. (2016, November). Spontaneous facial micro-expression detection based on deep learning. In *2016 IEEE 13th International Conference on Signal Processing (ICSP)* (pp. 1130-1134). IEEE.

[16] Matlovic, T., Gaspar, P., Moro, R., Simko, J., Bielikova, M. (2016, October). Emotions detection using facial expressions recognition and EEG. In *2016 11th international workshop on semantic and social media adaptation and personalization (SMAP)* (pp. 18-23). IEEE.

[17] Xiang, J., Zhu, G. (2017, July). Joint face detection and facial expression recognition with MTCNN. In *2017 4th international conference on information science and control engineering (ICISCE)* (pp. 424-427). IEEE.

[18] Greche, L., Jazouli, M., Es-Sbai, N., Majda, A., Zarghili, A. (2017, April). Comparison between Euclidean and Manhattan distance measure for facial expressions classification. In *2017 International conference on wireless technologies, embedded and intelligent systems (WITS)* (pp. 1-4). IEEE.

[19] Kumar, S., Singh, S., Kumar, J. (2018). Automatic live facial expression detection using genetic algorithm with haar wavelet features and SVM. *Wireless Personal Communications*, *103*(3), 2435-2453. https://doi.org/10.1007/s11277-018-5923-y.

[20] Chang, F. J., Tran, A. T., Hassner, T., Masi, I., Nevatia, R., Medioni, G. (2018, May). Expnet: Landmark-free, deep, 3d facial expressions. In *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)* (pp. 122-129). IEEE.

[21] Jain, D. K., Shamsolmoali, P., Sehdev, P. (2019). Extended deep neural network for facial emotion recognition. *Pattern Recognition Letters*, *120*, 69-74.

[22] Xie, S., Hu, H., Wu, Y. (2019). Deep multi-path convolutional neural network joint with salient region attention for facial expression recognition. *Pattern recognition*, *92*, 177-191. https://doi.org/10.1016/j.patcog.2019.03.019

[23] Hammed, S. S., Sabanayagam, A., Ramakalaivani, E. (2020). A review on facial expression recognition systems. *Journal of critical reviews*, *7*(4), 903-905.

[24] Porcu, S., Floris, A., Atzori, L. (2020). Evaluation of data augmentation techniques for facial expression recognition systems. *Electronics*, *9*(11), 1892. https://doi.org/10.3390/electronics9111892

[25] Tsai, K. Y., Tsai, Y. W., Lee, Y. C., Ding, J. J., Chang, R. Y. (2021). Frontalization and adaptive exponential ensemble rule for deep-learning-based facial expression recognition system. *Signal Processing: Image Communication*, *96*, 116321.

[26] Almeida, J., Rodrigues, F. (2021, April). Facial Expression Recognition System for Stress Detection with Deep Learning. In *ICEIS (1)* (pp. 256-263).

[27] Shabbir, N., Rout, R. K. (2023). Variation of deep features analysis for facial expression recognition system. *Multimedia Tools and Applications*, *82*(8), 11507-11522. https://doi.org/10.1007/s11042-022-14054-w

[28] Kadakia, R., Kalkotwar, P., Jhaveri, P., Patanwadia, R., Srivastava, K. (2022, November). Analysis of Micro Expressions using XAI. In *2022 3rd International Conference on Computing, Analytics and Networks (ICAN)* (pp. 1-7). IEEE.

[29] Lee, K. W., Lee, H. J., Hu, H., Kim, H. J. (2022). Analysis of facial ultrasonography images based on deep learning. *Scientific reports*, *12*(1), 16480.

[30] Yaddaden, Y. (2023). An efficient facial expression recognition system with appearance-based fused descriptors. *Intelligent Systems with Applications*, *17*, 200166.

[31] Bartlett, M. S., Littlewort, G., Frank, M., Lainsc-sek, C., Fasel, I., Movellan, J. Recognizing facial expression: machine learning and application to spontaneous behavior. In 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05);2005. pp. 568-573.

[32] Sonmez, E., Albayrak, S. A facial component-based system for emotion classification. Turkish Journal of Electrical Engineering and Computer Sciences; 2016. 24(3): 1663-1673. https://doi.org/10.3906/elk-1401-18.

[33] Farajzadeh, N., Hashemzadeh, M. Exemplar-based facial expression recognition. Information Sciences;2018. 460: 318-330. https://doi.org/10.1016/j.ins.2018.05.057

[34] Baygin, M., Tuncer, I., Dogan, S., Barua, P. D., Tuncer, T., Cheong, K. H., Acharya, U. R. Automated facial expression recognition using exemplar hybrid deep feature generation technique. Soft Computing, 2023. 1- 17. 27:8721-8737. https://doi.org/10.1007/s00500-023-08230-9.

[35] Lu, S., Evans, F. Haar wavelet transform based facial emotion recognition. In 2017 7th international conference on education, management, computer and society, 2017. pp. 342-346. https://doi.org/10.2991/emcs17.2017.67

[36] Chowdhury, J. H., Liu, Q., Ramanna, S. (2024). Simple Histogram Equalization Technique Improves Performance of VGG Models on Facial Emotion Recognition Datasets. Algorithms, 17(6), 238.

[37] Haq, H. B. U., Akram, W., Irshad, M. N., Kosar, A., Abid, M. (2024). Enhanced real-time facial expression recognition using deep learning. Acadlore Trans. Mach. Learn, 3(1), 24-35.

[38] Talaat, F. M., Ali, Z. H., Mostafa, R. R., El-Rashidy, N. (2024). Real-time facial emotion recognition model based on kernel autoencoder and convolutional neural network for autism children. Soft Computing, 1-14.

[39] Meena, G., Mohbey, K. K., Indian, A., Khan, M. Z., Kumar, S. (2024). Identifying emotions from facial expressions using a deep convolutional neural network-based approach. Multimedia Tools and Applications, 83(6), 15711-15732.

[40] Pacal, I. MaxCerVixT: A novel lightweight vision transformer-based Approach for precise cervical cancer detection. Knowledge-Based Systems, 2024, 289, 111482. https://doi.org/10.1016/j.knosys.2024.111482

[41] Pacal, I. Enhancing crop productivity and sustainability through disease identification in maize leaves: Exploiting a large dataset with an advanced vision transformer model. Expert Systems with Applications, 2024, 238, 122099. https://doi.org/10.1016/j.eswa.2023.122099

[42] Pacal, I. A novel Swin transformer approach utilizing residual multi-layer perceptron for diagnosing brain tumors in MRI images. International Journal of Machine Learning and Cybernetics, 2024, 15:3579–3597 https://doi.org/10.1007/s13042-024-02110-w

[43]https://www.kaggle.com/datasets/jonathanoheix/face-expression-recognition-dataset (Access Time: 10 October 2023)

[44] Yamini Piparsaniyan, Vijay K. Sharma, K. Mahapatra Robust facial expression recognition using Gabor feature and Bayesian discriminating classifier, 2014

[45] Siddiqui, E. A., Chaurasia, V., Shandilya, M. (2023). Detection and classification of lung cancer computed to-mography images using a novel improved deep belief network with Gabor filters. Chemometrics and Intelligent Laboratory Systems, 235, 104763.

[46] Sandeep Kumar, Sukhwinder Singh Jagdish Kumar, Au-tomatic Live Facial Expression Detection Using Genetic Algorithm with Haar Wavelet Features and SVM, 2018

[47] Batziou, E., Ioannidis, K., Patras, I., Vrochidis, S., Kompatsiaris, I. (2023, January). Low-Light Image En-hancement Based on U-Net and Haar Wavelet Pooling. In International Conference on Multimedia Modeling (pp. 510-522). Cham: Springer Nature Switzerland.

[48] Dixit, U. D., Shirdhonkar, M. S., Sinha, G. R. (2023). Automatic logo detection from document image using HOG features. Multimedia Tools and Applications, 82(1), 863-878.

[49] Acar, Y. E., Saritas, I., Yaldiz, E. (2022). Com-parison of ML algorithms to distinguish between human or human-like targets using the HOG features of range-time and range-Doppler images in through-thewall applications. Turkish Journal of Electrical Engineering and Computer Sciences, 30(6), 2086-2096.

[50] Bayrak, S. (2024). Unveiling intrusions: explainable SVM approaches for addressing encrypted Wi-Fi traffic in UAV networks. Knowledge and Information Systems, 1-21. https://doi.org/10.1007/s10115-024-02181-9

[51] Phung, V. H., Rhee, E. J. (2019). A high-accuracy model average ensemble of convolutional neural networks for classification of cloud image patches on small datasets. Applied Sciences, 9(21), 4500. https://doi.org/10.3390/app9214500