www.editada.org

International Journal of Combinatorial Optimization Problems and Informatics, 17(2), 2026, 89-101. ISSN: 2007-1558.
https://doi.org/10.61467/2007.1558.2026.v17i2.1261

An improved CNN-based methodology to prevent risky situations in people by recognizing emotions from facial features

José Félix Serrano-Talamantes[1], *Mauricio Olguín-Carbajal*[1], Gerardo Miramontes-de León[2], Héctor Durán-Muñoz[2], Claudia Sifuentes-Gallardo[2*], Carlos Avilés-Cruz[3], Gabriel de Jesús Celis-Escudero[3]

[1] Centro de Innovación y Desarrollo Tecnológico en Computación, Cómputo Inteligente-Instituto Politécnico Nacional (CIDETEC-IPN), Ciudad de México, México.
[2] Academic Unit of Electrical Engineering, Autonomous University of Zacatecas, Zacatecas, Jardin Juarez 147, 98160, Mexico
[3] Universidad Autónoma Metropolitana, Unidad Azcapotzalco. Departamento de Electrónica. Av. San Pablo 420 Col. Nueva el Rosario, C.P. 02128, Ciudad de México, México.

E-mails: *jfserrano@ipn.mx*, molguinc@ipn.mx *gmiram@ieee.org*, *hectorduranm@uaz.edu.mx*, *clausifuen@uaz.edu.mx*, *caviles@azc.uam.mx,* cgjcelis90@gmail.com

Abstract: This proposal shows a methodology designed for the process of detecting emotions through facial features. The process of facial features detection involves several stages among which the most important ones are: Acquisition of the training set of images, detection and segmentation of the face using face location techniques in images using Viola & Jones algorithm. We also make use of neural networks that are part of Deep Learning techniques. In this way we propose to recognize people's faces and also perceive their emotions through gestures that are captured by means of a camera, allowing us to obtain these in soft real time. The processes of this methodology were developed and programmed in the MATLAB and Python code. There was a significant improvement in the recognition results throughout the CNN.

Keywords: Deep Learning; Neural Networks; Google net; Computer Vision; MATLAB

Article Info
Received December 2, 2025
Accepted Jan 12, 2026

1 Introduction

To begin with, it should be noted that face detection and face recognition are different terms. Face detection is based on obtaining a series of squares that contain a face within their area, this process is repeated until no detectable face remains unframed. In simple terms, we can summarize face detection as the task prior to recognition, the task of obtaining a face to be recognized. Face recognition is limited to identifying the individual features of the person.

Face detection and face recognition are distinct processes: the first identifies and locates faces in an image, while the second analyzes those detected faces to determine the person's identity. In other words, detection is the preliminary step that provides the input for recognition.

The development of the proposed methodology involves several key stages designed to ensure accurate emotion recognition. These stages include face detection, feature extraction, and emotion classification through deep learning. For face detection, various algorithms have been explored, such as Hausdorff Distance, Local Binary Patterns, Neural Networks, Viola–Jones, and Convolutional Neural Networks (CNNs). Among these, the Viola & Jones algorithm is adopted in this work due to its efficiency, robustness, and relatively low false-positive rate.

The Hausdorff Distance method is edge-based and effective for grayscale images, using a similarity measure between a general face model and its possible instances within an image. (Jesorsky et al., 2001)

- Local Binary Patterns: The algorithm is based on dividing the images into grids, and these grids are used by defining the central pixel of each grid as the threshold boundary, all near neighbors below that boundary are marked as 0, values that are equal or higher are marked as 1, indicating that they belong to the object pixels set (Chang-Yeon, 2008).
- Neural Networks: Certainly, it is one of the technologies that since 2014 has emerged as one of the best solutions to various problems that had a somewhat complicated solution. The neural network algorithm works by decomposing the images into sections and extracting features from each section. This process is repeated several times and the number of times it is repeated depends on the total number of levels of the neural network (Hardesty L., 2017).
- Viola & Jones: The Viola Jones object detection method was proposed by Paul Viola and Michael (Viola, P., & Jones, M. J., 2001). The detection rate of this method is relatively high and very low in relation to false positives, which makes the algorithm so robust and processes images at a relatively high speed. This algorithm is composed of different classifiers, each located in different spaces of the image, these classifiers are known as weak classifiers, mainly because the false positive rate can be high, however, when combined in a single output these weak classifiers give us a fairly robust detection result. The output that represents the sum of these weak classifiers is called Strong Classifier ("Algorithm," 2019). See Fig 1.

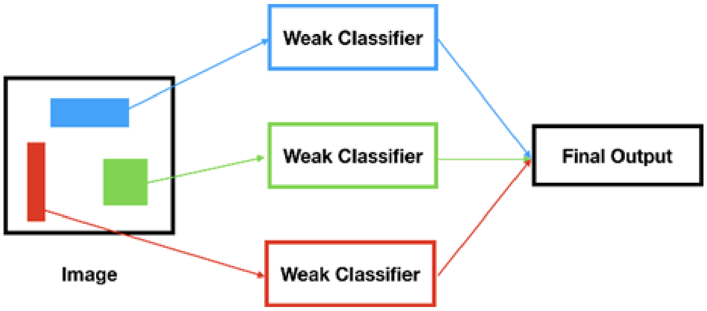


Fig. 1. Viola & Jones Detection Algorithm.

This will be the algorithm that will be used for the development of this proposal, mainly because of its advantages:

- Efficient feature selection.
- Scale and location invariant.
- The type of training can be used to detect different objects.
- Detection using multiple classifiers strengthens the success rate.

The remainder of this paper is structured as follows: Section II reviews related works and theoretical background; Section III presents the proposed approach; Section IV details the topology-based theory; Section V explains the system methodology; Section VI discusses experimental results; and Section VII provides conclusions and future work.

2. Theoretical Background and State of the Art

There are a large number of algorithms that allow us to perform the task of face recognition, for example:

- Eigen Faces.
- Fisher Faces.
- Local Binary Patterns Histogram.
- Convolutional Neural Network (CNN).

These are the techniques mainly used for face recognition, next, it is mentioned the performance and the main advantages of each of these algorithms.

2.1. Eigen Faces

Eigen Faces is one of the oldest face recognition algorithms that exist, and its performance is quite reliable as long as the training set for face recognition of this algorithm is large and with a controlled light environment, this is because it performs a "Principal Component Analysis" (PCA) for face recognition. For the training of this algorithm, we must first obtain a set of images that will be used, these images must be taken as frontal and cropped images of the face, in addition, it is suggested that the images of the faces are taken in a controlled light environment (Trivedi S., 2009).

- Advantages: Simplicity and speed of classification.
- Disadvantages: Slow scalability when adding new faces and low accuracy as it alone does not achieve good classification.

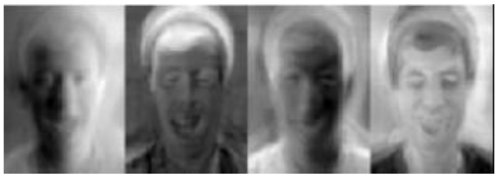


Fig. 2. Face recognition using Eigen Faces (Trivedi S,2009).

2.2. Fisher Faces

This algorithm is similar to Eigen Faces, it does a per pixel component analysis for its operation and classification.The improvement in this algorithm is that instead of doing a Principal Component Analysis (PCA), it performs a Linear Discriminant Analysis (LDA). Being a similar algorithm to the previous one, but with improvements, the training process is similar, first we must obtain a set of images that will be used, these images must be frontal and cropped shots of the face, as an additional point it is suggested that the images of the faces are taken in a controlled light environment (Martinez A, 2011).

- Advantages: Higher accuracy,
- Disadvantages: It has a higher computational cost.

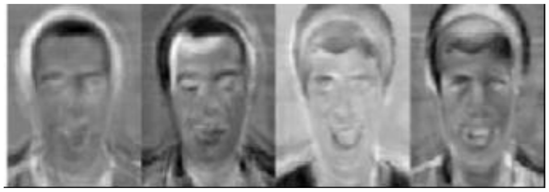
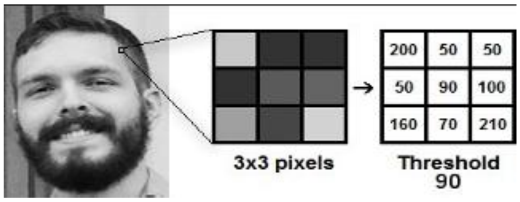


Fig. 3. Facial recognition using Fisher faces. (Martínez, 2011).

2.3. Local Binary Patterns Histogram

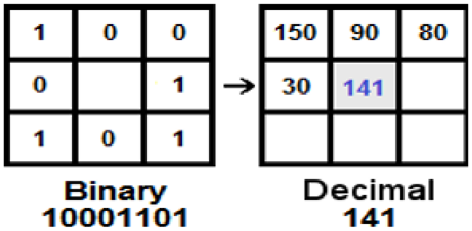
It is a texture operator that works with images at the pixel level, by labeling pixels using a threshold or thresholding that allows us to delimit the pixels that belong to a certain region of the image, the calculation of the limiting threshold can be performed with cardinality 4 or 8, and the center will help us determine whether or not neighboring pixels belong to the same region of the image (Chang-Yeon,2008).



200	50	50
50	90	100
160	70	210

Fig. 4. Face region recognition using Local Binary Patterns Histogram (LBPH) part 1 (Chang-Yeon, 2008).

Finally, values that are equal to or greater than the threshold are assigned a '1' and those that are not are assigned a '0'. This is in order to concatenate all the 0's and 1's in the neighborhood and thus obtain a decimal value that represents the pixel of the original image.



1	0	0
0		1
1	0	1

150	90	80
30	141	

Fig. 5. Face region recognition using Local Binary Patterns Histogram (LBPH) part 2 (Chang-Yeon, 2008).

As a final process we obtain a new image, but with more detailed features than the original one. It only remains to divide the images in MxN regions to extract the histogram of each region.

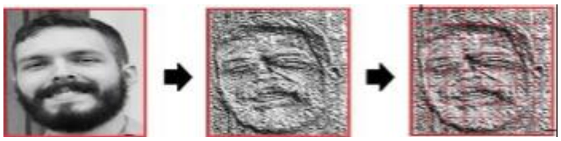


Fig. 6. Image with features and segmenting in MxN Regions (Chang-Yeon,2008).

Once the histogram of each region has been extracted, the extracted histograms are concatenated in order to have a single overall histogram, so as to be able to compare the extracted histogram against the histogram of the person to be identified (Chang-Yeon, 2008).

- Advantages: invariant to illumination, low computational cost, effective texture representation, no complex training required, flexible and scalable, and good interpretability (Ahonen et al, 2006).
- Disadvantages: sensitivity to rotation, dependence on scale, lower robustness against occlusions or noise, and limited performance compared to deep learning methods such as CNNs. (Ahonen et al, 2006).

2.3. Convolutional Neural Network (CNN):

A convolutional neural network (CNN) is a deep learning architecture (Xiao, M.,2024; Ye, X.,2024) that learns directly from raw data, eliminating the need for manual feature extraction. It can consist of dozens or even hundreds of layers, each designed to identify different characteristics within an image (Wang, A.,2025). During training, filters are applied to input images at multiple resolutions, and the output from each convolutional layer serves as the input for the next. These filters range from detecting simple patterns like edges and brightness to more complex features that uniquely identify objects. Similar to other neural networks, a CNN includes an input layer, an output layer, and multiple hidden layers in between.

- Advantages: Provides high accuracy levels, training process is simple, high malleability range to increase accuracy.
- Disadvantages: Neural network setup can become somewhat complex, training of the network can take days in cases of fairly large Dataset.

In this case this will be the development method to be used for the face recognition process.

Figure 7 shows the proposed deep learning model for identifying the six moods. The process begins with an image database containing images with dimensions of 224 × 224 × 3. Next, attributes are extracted using a varying number of filters at each stage. Next is vectorization, where 784 attributes are obtained. The final classification task uses four fully connected layers. The network output consists of the seven possible moods. Details of the CNN network are provided below.

The training process involves minimizing the cross/entropy loss functions with the backpropagation algorithm, while parameter optimization is carried out using stochastic gradient descent (Zhou, 2018; Boué, 2018). The joint entropy loss function is given by Equation (1).

$$£_{ALL} = (\theta_1, \theta_2) = £_{FE}(\theta_1) + £_{CLA}(\theta_2) \tag{1}$$

where $£_{FE}(\theta_1)$denotes the loss function for feature extraction tasks, and $£_{CLA}(\theta_2)$denotes the loss function associated with the classification layer (dropout ans soft-max layers). The overall loss function $£_{ALL}$ is expressed in Equation (2).

$$£_{ALL}(\theta_1, \theta_2) = -\sum_{i=1}^{7} \log\,[\hat{p}_{FE}\,(Moode_1/\theta_1)] - \log\,[\hat{p}_{CLA}(Moode_i/\theta_2)] \tag{2}$$

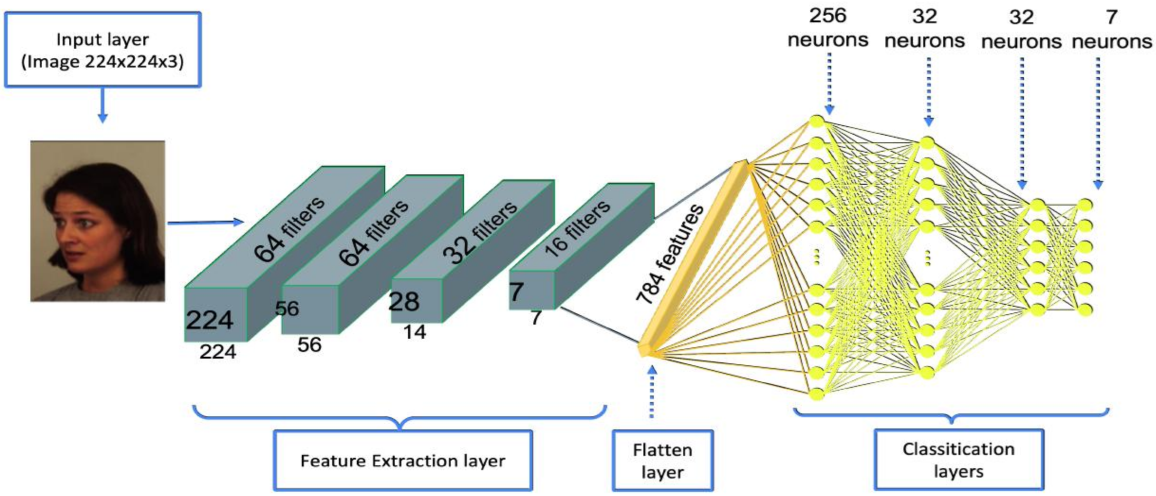


Fig. 7. Proposed Deep Learning model.

where $\theta_1$represents the parameter set of the CNN feature Extraction, and θ_2 represents the parameter set of the classification layer(dropout and Dense layers). $Moode_1$ denotes the given mode, where i = {Fear, Anger, Disgust, Happy, Neutral, Sad, Surprise}. The term $\hat{p}_{FE}$ $(Moode_1/\theta_1)$ denotes the conditional probability function of a given moode, conditions on the feature extraction parameter set θ_1, while $\hat{p}_{CLA}$ $(Moode_1/\theta_2)$ denotes the conditional probability function of a given moode, conditioned on the classification parameter set θ_2 .

The optimization of Eq. (2) is performed using the gradient taken over a minibatch of size N$_B$. Thus, the gradient over the entire objective function is described in the following.

$$\nabla_{(\theta_{1,}\theta_2)}(F_{ALL})[\frac{1}{N_R}\sum xlogF_{ALL}] = Gradient \tag{3}$$

The total cost function, expressed by Eq. (2) is optimized throughout the cross-entropy derivative (Mao A., 2023). The Gradient (See Eq (3)) must reach a minimum error close to zero. The parameters $(\theta_{1,}\theta_2)$ of the discriminator are optimized during the iterative training process (Yang et al., 2024).

As shown in Table 1, you can complete the task one layer at a time. As you can see, there are four Feature Extraction layers (from layer 1 to 22); a vectorization task (layer 23); and four classification layers (from 24 to 27). There are six tasks for each FE task: Convolution, Leaky ReLU, Max Pooling, Normalization, and Dropout (Gholamalinezhad et al.2020; Cai et al. 2019; Zhao et al.2024).

Table 1 Proposed CNN layers.

Task	Layer	Input dimension	Output dimension
FE 1	[1] Convolution1	[224 x 224 x 3]	[224 x 224 x 64]
	[2] Leaky_ReLU 1	[224 x 224 x 64]	[224 x 224 x 64]
	[3] Gaussian Noise 1	[224 x 224 x 64]	[224 x 224 x 64]
	[4] Max Pooling 1	[224 x 224 x 64]	[56 x 56 x 64]
	[5] Group Normalization 1	[56 x 56 x 64]	[56 x 56 x 64]
	[6] Dropout 1	[56 x 56 x 64]	[56 x 56 x 64]
FE 2	[7] Convolution 2	[56 x 56 x 64]	[56 x 56 x 64]
	[8] Leaky_ReLU 2	[56 x 56 x 64]	[56 x 56 x 64]
	[9] Gaussian Noise 2	[56 x 56 x 64]	[56 x 56 x 64]
	[10] Max Pooling 2	[56 x 56 x 64]	[56 x 56 x 64]
	[11] Group Normalization 2	[56 x 56 x 64]	[56 x 56 x 64]
	[12] Dropout 2	[28 x 28 x 64]	[28 x 28 x 64]
FE 3	[13] Convolution 3	[28 x 28 x 64]	[28 x 28 x 32]
	[14] Leaky_ReLU 3	[28 x 28 x 32]	[28 x 28 x 32]
	[15] Gaussian Noise 3	[28 x 28 x 32]	[28 x 28 x 32]
	[16] Max Pooling 3	[28 x 28 x 32]	[14 x 14 x 32]
	[17] Group Normalization 3	[14 x 14 x 32]	[14x 14 x 32]
FE 4	[18] Convolution 4	[14 x 14 x 32]	[14 x 14 x 16]
	[19] Leaky_ReLU 4	[14 x 14 x 16]	[14 x 14 x 16]
	[20] Gaussian Noise 4	[14 x 14 x 16]	[14 x 14 x 16]
	[21] Max Pooling 4	[14 x 14 x 16]	[7 x 7 x 16]
	[22] Group Normalization 4	[7 x 7 x 16]	[7 x 7 x 16]
Vectorization	[23] Flattening	[7 x 7 x 16]	784
Classification	[24] Dense 1	784	256
	[25] Dense 2	256	32
	[26] Dropout 2	32	32
	[27] Dense 3	32	7

FE= Feature Extraction

2.3.1 Dataset Characteristics

The following is a description of where the image data set is obtained from. In this section we would like to acknowledge the Dataset provided by the website (Lundqvist1 et al. 1998) & AKDEF (Lundqvist2 et al. 1998). It is a Dataset containing a total of 4900 images of facial expressions and a total of 7 different emotions: Fear, Anger, Disgust, Happiness, Neutral, Sadness and Surprise. Each emotion is captured from 5 different angles: 2 from the side, 2 from the side and 1 from the front. Each image has an identifier composed by a series of characters:

First Character – Photography Session:

- A = Session 1
- B = Session 2

Second Character – Gender:

- F = Female
- M = Male

Third and Fourth Characters – Identifier:

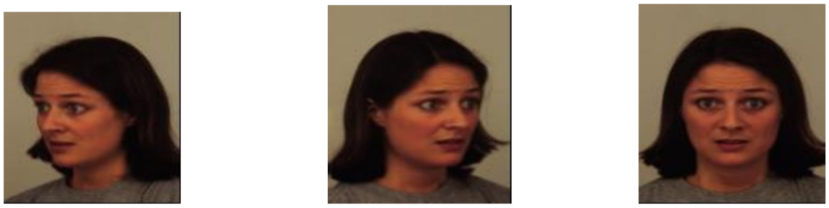
- Numeric Identifier: 01–35

Fifth and Sixth Characters – Expression:

- AF = Afraid (Fear)
- AN = Angry (Anger)
- DI = Disgusted (Disgust)
- HA = Happy (Happiness)
- NE = Neutral
- SA = Sad (Sadness)
- SU = Surprise (Surprise)

3. Methodology

Having defined that the use of the Viola & Jones algorithm is proposed for face detection and the use of an artificial neural network for the face recognition process, it remains to be noted that from the whole set of images provided from the Dataset only 3 of the 5 photos were used, mainly because we focused on recognizing faces at frontal angles as shown as follows: (See figures 8 a), b), and c), respectively) The implementation was carried out in Matlab (version 2024a) using the image processing tool (Matlab1, (2024), Matlab2, (2024), Matlab3, (2024)).



a) left side capture BF01AFHL. **b)** Right side capture BF01AFHR. **c)** front side capture BF01AFS.

Fig. 8. CNN training

Face Detection with Viola Jones

The process of face detection according to the Viola & Jones algorithm (Viola & Jones, 2001) is to import the set of images that are candidates for face extraction, as shown in figures 8, 9 and 10 respectively; the next step is to import the cascade object detector and define the threshold limit of the size of objects to be recognised, this step can be skipped and import the algorithm and use it as it is, but if no threshold limit is defined, false positives can be obtained. Once the threshold limit has been defined, the algorithm makes use of the weak classifiers to search the regions of the image for features that match that of the face and by joining the weak classifiers, we obtain the strong classifier, which generates a bounding box with the face within the region, this result is shown in the illustration as follows. (See fig.9).

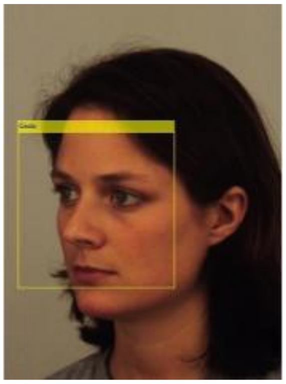


Fig. 9. Face detection through Viola & Jones algorithm.

The next step is to crop the region containing the face, this result is shown in fig. 10

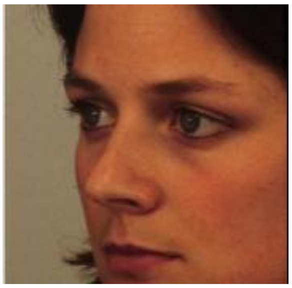


Fig. 10. identified facial crop.

Finally, once we have the segmented face, it only remains to save it in a new folder containing all these segmented faces of the corresponding emotion, for example, in fig. 11 it corresponds to the emotion "Neutral".

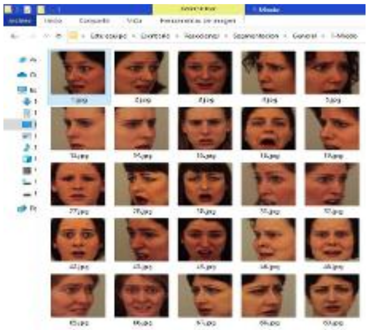


Fig. 11. Segmented faces folder.

Creation of the Image Data Store: This is the process responsible for reading and storing the images in the programme space. Once we have stored our images, the next step is to extract the categories that are part of the Dataset, as mentioned above, we only focus on recognizing two emotions: neutral and scared. Once the images have been stored, the process of partitioning the set in two parts is pending: training and validation.

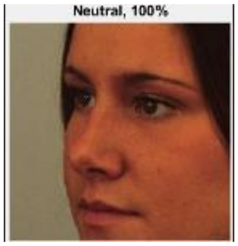
4. Experimental Results

Table 2 shows the results obtained using the Eigen Faces, Fisher Faces, and Local Binary Pattern Histograms (LBPH) methods. An average is presented for each of the aforementioned methods.

Table 2. Face segmentation results.

Emotion	Input Images	Detected Faces	Precision
Neutral	208	151	72.60%
Fear	208	164	78.85%
Final Precision			75.72%

In the following, we show a series of results obtained from the validation set. (See figs 12,13,14 and 15 respectively).



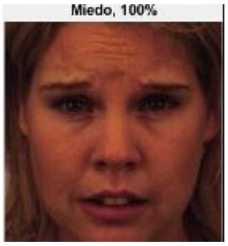
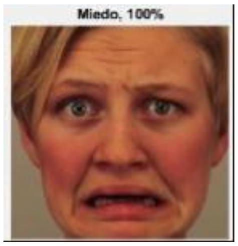


Fig 12.-Neutral image result. **Fig 13**.- Fear Image result.



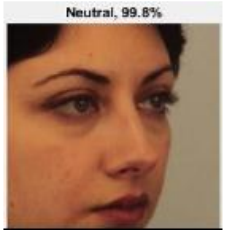
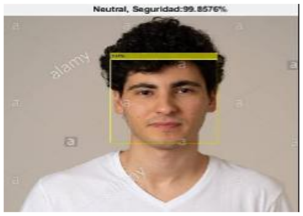


Fig. 14. Fear image result. **Fig 15.-** Neutral image result

We also include results from images that were not part of the network training set. (See figures 16 and 17 respectively).



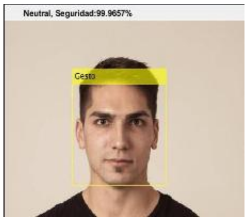


Fig 16.- Neutral image result **Fig 17.-** Neutral Image result.

If we want to improve the efficiency obtained in Table 1, then we propose: using a CNN neural network model starting from an initial stage, which means we will not need to use other types or models of learning to train this network, as shown in Figure 7. In addition, we will not only detect two emotional states, such as fear and neutrality, but also states of fear, anger, disgust, happy, neutral, sad and surprise Nor will we use flowcharts with algorithms to read, store, and partition faces, as initially proposed in this proposal. In addition, we will not only detect two emotional states such as fear or neutral, but also detect states of {Fear, Anger, Disgust, Happy, Neutral, Sad, Surprise.

4.1. Results for Deep Learning Model

In relation to the deep learning model, the algorithms were implemented on a work-station with the following characteristics. PC Alienware Aurora R14 Gaming 3 Intel © Core ^{TM} i9 - 12900KF, ubuntu 20.1, NVIDIA®GeForce RTX^{TM} 4090, 24 GB GDDR6X, 64 GB, DDR5, 4400 MHz, dual-channel. The implementation was carried out in Python (version 3.9) using the TensorFlow, Keras, Matplotlib, and NumPy libraries. The network training parameters are as follows.

a) loss='binary_crossentropy', b) optimizer=Adam(lr=0.0001, beta=0.4), c) metrics=['loss, accuracy'], d) batch_size = 128, e) epochs = 700, and f) sample_period = 50.

Figure 18 shows the accuracy and loss curve. As can be seen, the loss tends to zero and the accuracy tends to one.

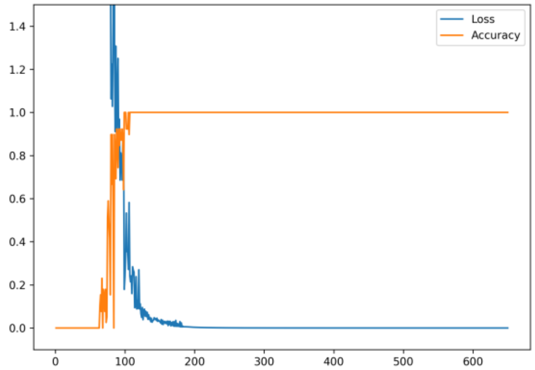


Fig. 18. Epoch-wise accuracy and loss trends of the proposed DL model.

The trained model was used to classify the 4,900 images that were used for training (resubstitution mode) obtaining 94.28% recognition for each mood (see Table 3).

Table 3. Confusion Matrix for the 7 mood classification.

Actual/Predicted	Fear	Anger	Disgust	Happy	Neutral	Sad	Surprise
Fear	**90.8**	0.4	0.7	2.0	0.4	3.2	2.5
Anger	0.3	**99**	0.33	0	0	0	0.33
Disgust	1.7	0.3	**92.7**	0.3	2.4	2.4	0.2
Happy	1.4	0.7	0.7	**93.2**	0	3.4	0.6
Neutral	0.4	0	0	0	**97.4**	2.2	0
Sad	0.8	1.5	2.3	1.5	2.2	**90.9**	0.8
Surprise	3.7	0	0	0	0.3	0	**96**

The results regarding the other statistical features, such as Average Accuracy, Precision, Recall, and F1 Score, are as follows: Accuracy= 94%, Precision= 94%, Recall=94%, F1 Score=94%.

4.2. Performance Analysis of CNN Model Accuracy

In this subsection, we present a comprehensive analysis of the performance of the proposed deep learning model for flow regime classification. The analysis begins by examining the spatial separability of the data, which is visualized by projecting it into a three-dimensional feature space. Following this, the significance of individual attributes is evaluated using gradient-based analysis at the output of the convolutional filters. Finally, data augmentation and performance evaluation are carried out to demonstrate the robustness of the proposed deep learning methodology (DLM).

To assess both robustness and accuracy, the following techniques were implemented: (i) Principal Component Analysis (PCA) was used to evaluate the separability of the data in the feature space; (ii) Grad-CAM (Gradient-weighted Class Activation Mapping) was used for visualization, enabling the identification of the most influential features in the model's decision-making process (Omae, 2025).

Figure 19 displays the PCA results after projecting the 784 features (obtained post-flattening) into three dimensions. The figure shows that the seven emotions—Afraid (blue), Angry (orange), Disgust (green), happy (red), Neutral (magenta), sad (brown), and Suprise (pink)—are clearly separated and at least linearly separable. This indicates the potential for high classification accuracy using the proposed model.

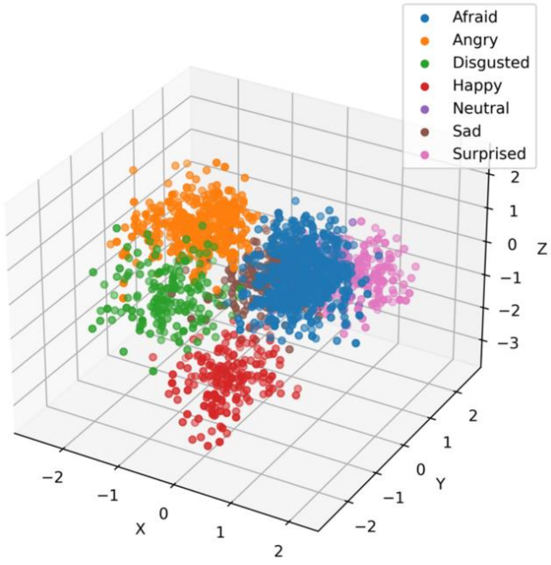


Fig. 19. Projection from high dimension (784) to dimension 3 using PCA.

The Grad-CAM technique (Omae, 2025) was made using the gradients from the already-trained model's output to the first layer of feature extraction, i.e., the first convolution task. Thirty- two filters are applied to the input image. Figure 20 shows the resulting Grad-CAM heat map. The original image is plotted in blue, and the gradients are plotted in red. As can be seen, the most significant gradients are around the face. Hence, it can be concluded that there are 4 segments of the face for its discrimination, Mouth, Nose, Eyes, and Beard.

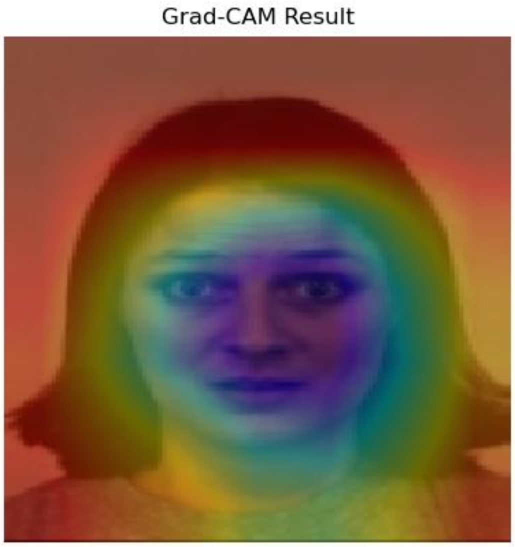


Fig. 20. Grad-CAM on image face after the first convolution layer.

4.3. Comparison versus pre-trained models

To evaluate the effectiveness of the proposed approach, its performance was compared against several widely used pre-trained models, including InceptionV3, VGG16, VGG19, Xception, and EfficientNetB0. These architectures, originally trained on large-scale datasets, were fine-tuned through transfer learning for facial emotion recognition on the KDEF/AKDEF databases, as reported in the respective reference papers.

As summarized in Table 4, the results highlight the variation in accuracy among these models. Deeper and more optimized architectures such as VGG16 and Xception achieve higher recognition rates compared to the rest. This comparison provides a clear benchmark for assessing the competitiveness of the proposed CNN model for the KDEF.

Table 4. Comparison versus pre-trained models (Accuracy (%) and Reference).

Model	KDEF-dataset
This work (2025)	**94.28**
InceptionV3	95.10 [74]
VGG16	93.02 [74]
VGG19	95.29 [74]
Xception	95.10 [6]
EfficientNetB0	93.00 [34]

The results confirm that the proposed model surpasses existing pre-trained networks in terms of classification accuracy for the KDEF dataset, achieving recognition of 94.28%.

5. Conclusions and Future Work

This work confirms the power of deep learning in recognizing human emotions from facial features. While traditional methods such as Eigenfaces, Fisherfaces, and Local Binary Patterns provide useful baselines, their accuracy remains limited. In contrast, the proposed Convolutional Neural Network (CNN) showed outstanding results, reaching 94% in accuracy, precision, recall, and F1-score, and achieving perfect recognition in resubstitution testing. These results highlight the robustness of CNNs in capturing subtle emotional cues without the need for manual feature engineering.

The CNN demonstrated not only high reliability but also strong interpretability. Techniques such as PCA confirmed the separability of emotions in feature space, while Grad-CAM visualizations showed that the model naturally focused on critical regions of the face—the eyes, mouth, nose, and beard—when making decisions. This indicates that the network learned meaningful human-like patterns, reinforcing its applicability to real-world contexts.

Beyond numbers, the system demonstrates how advanced AI can be harnessed to improve safety and interaction. Applications may range from silent alarms in critical situations to adaptive interfaces in healthcare and education. Its adaptability and scalability open doors to future improvements with larger and more diverse datasets. Future work will explore real-time implementation and integration into intelligent environments, ensuring broader applicability and enhanced social impact.

References

Ahonen, T., Hadid, A., & Pietikäinen, M. (2006). Face description with local binary patterns: Application to face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 28*(12), 2037–2041. https://doi.org/10.1109/TPAMI.2006.244

Gholamalinezhad, H., & Khosravi, H. (2020). *Pooling methods in deep neural networks: A review* (arXiv:2009.07485). https://arxiv.org/abs/2009.07485

Hardesty, L. (2017, April 14). *Explained: Neural networks and deep learning*. MIT News. https://news.mit.edu/2017/explained-neural-networks-deep-learning-0414

Jesorsky, O., Kirchberg, K. J., & Frischholz, R. W. (2001). Robust face detection using the Hausdorff distance. In *Proceedings of the International Conference on Audio- and Video-Based Biometric Person Authentication* (pp. 90–95).

Jung, C.-Y. (2008). Face detection using LBP features. In *Proceedings of the International Conference on Convergence and Hybrid Information Technology*. IEEE.

Lundqvist, D., Flykt, A., & Öhman, A. (1998). *The Karolinska Directed Emotional Faces (KDEF)* [CD-ROM]. Karolinska Institutet. ISBN 91-630-7164-9

Martinez, A. M. (2009). Fisherfaces. *Scholarpedia, 4*(2), 5566. http://www.scholarpedia.org/article/Fisherfaces

MathWorks. (2024). *DagNetwork*. https://es.mathworks.com/help/deeplearning/ref/dagnetwork.html

Viola, P., & Jones, M. J. (2001). Rapid object detection using a boosted cascade of simple features. In *Proceedings of the IEEE CVPR* (Vol. 1, pp. I-511–I-518). https://doi.org/10.1109/CVPR.2001.990517

Viola, P., & Jones, M. J. (2001). Rapid object detection using a boosted cascade of simple features. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (Vol. 1, pp. I-511–I-518). IEEE. https://doi.org/10.1109/CVPR.2001.990517

Yang, Y., Feng, H., & Zhou, D.-X. (2024). *On the rates of convergence for learning with convolutional neural networks* (arXiv:2403.16459). https://arxiv.org/abs/2403.16459

Zhao, X., Wang, L., Zhang, Y., Han, X., Deveci, M., & Parmar, M. (2024). A review of convolutional neural networks in computer vision. *Artificial Intelligence Review, 57*(4), 99. https://doi.org/10.1007/s10462-024-10721-6

Zhou, X. (2018). Understanding the convolutional neural networks with gradient descent and backpropagation. *Journal of Physics: Conference Series, 1004*(1), 012028. https://doi.org/10.1088/1742-6596/1004/1/012028