IAES International Journal of Artificial Intelligence (IJ-AI)
Vol. 15, No. 6, December 2026, pp. 5592~5604
ISSN: 2252-8938, DOI: 10.11591/ijai.v15.i6.pp5592-5604

A hybrid deep learning approach for robust facial recognition

Hicham Benradi, Adil Ammar, Hamza Natek, Abdelali Lasfar
Laboratory of Systems Analysis, Information Processing and Industrial Management (LASTIMI), Mohammadia School of Engineering, Higher School of Technology Salé, Mohammed V University, Rabat, Morocco

Article Info

Article history:

Received Nov 11, 2025
Revised Aug 8, 2026
Accepted Aug 31, 2026

Keywords:

Convolutional neural network
Facial recognition
Histogram equalization
Histogram-oriented gradient
Principal component analysis

ABSTRACT

Computer vision aims to enable machines to understand and interact in a visual environment. One of its major applications is facial recognition. Its goal is to identify individuals from a facial image, but this operation remains complicated due to variations in lighting, changes in facial expressions, and partial occlusions. This article proposes a hybrid method dedicated to high-performance facial recognition in the presence of variance or occlusion. To do this, preprocessing is applied to the set of images to improve their quality using the histogram equalization technique. Then, reduced in size to eliminate the least relevant features using principal component analysis (PCA). After that, using the histogram-oriented gradient (HOG) descriptor, the features were extracted from the images obtained in the previous step and submitted to an architecture created using a convolutional neural network (CNN) designed to exploit the extracted descriptors to train the model. The results of our simulations using the Olivetti Research Laboratory (ORL) and University of Manchester Institute of Science and Technology (UMIST) databases show high performance, with accuracy rates reaching 99.75% for the ORL dataset and 99.13% for the UMIST dataset, demonstrating the effectiveness of the approach and confirming the value of combining manual descriptors with deep learning architectures.

This is an open access article under the CC BY-SA license.

Corresponding Author:

Hicham Benradi
Laboratory of Systems Analysis, Information Processing and Industrial Management (LASTIMI)
Mohammadia School of Engineering, Higher School of Technology Salé, Mohammed V University
Rabat, Morocco
Email: benradi.hicham@gmail.com

1. INTRODUCTION

Facial recognition is one of the all-too-dynamic techniques in computer vision, adopted in many areas such as security [1], [2], biometrics [3], [4], video surveillance [5], [6], disease detection [7]–[9], presence systems [10], [11], and personal identification [10], [11]. This technique relies on an all-too-precise process from image acquisition and face extraction to individual identification. However, the performance of the latter can be impacted by the presence of several types of variations, such as lighting, contrast, image acquisition angle, or the presence of occlusion, such as wearing caps, glasses, and scarves during image acquisition. In this context, convolutional neural networks (CNNs) are considered to be one of the most robust and suitable methods for this situation, and stand out for their high capacity to extract the most relevant and discriminating features from images. This operation is performed using several types of filters, divided into two groups of layers: the first group is responsible for extracting features using several convolution layers (convolution filters, maxpool, and dense), and the second group is responsible for classifying input images into a corresponding class called the fully connected layer. Well-known for their robustness, CNNs are used in many fields, including healthcare [12]–[14], agriculture [15]–[17], and

industry [18]–[20]. On the other hand, there are also other techniques capable of extracting features from images, such as the histogram of oriented gradients (HOG) descriptor [21], which encodes the local structure of contours by capturing the distribution of gradient orientations in an image, making it particularly effective for object detection [22]. Also, the scale-invariant feature transform (SIFT) method [23], whose main objective is to extract scale- and rotation-invariant points of interest by describing their neighborhood with robust descriptors, makes it a powerful technique for image matching [24].

Several research studies have addressed the problem of face recognition in the presence of occlusion and variations of different types, to improve individual identification. Anil and Suresh [25] designed an effective hybrid method for facial recognition called HOG-CKLEM, which combines the effectiveness of CNN with kernel extreme learning machines (KELM) and also HOG, which are used as descriptors to extract the most discriminating facial features. Simulation results for this method using the Olivetti Research Laboratory (ORL) database yielded satisfactory results with a remarkable accuracy rate of 90.62%. Abdullah *et al.* [11] developed a new robust approach for occlusion-resistant facial recognition by leveraging transfer learning from the pre-trained VGG-16 model, used to extract discriminative facial features. These representations are then organized and classified using the k-nearest neighbors (KNN) algorithm. The experiments carried out showed high performance with an accuracy rate of up to 95%. Another study focused on evaluating advanced preprocessing techniques integrated into a pipeline using contour detection with the Canny algorithm, histogram lighting normalization techniques, and gamma correction has been reported [26]. Simulation results showed a very significant improvement either in the case of low-light conditions, where pre-processing improved recognition from 64.4% accuracy to 84.2%, or in the case of occlusion, where accuracy improved from 62.5% to 80.3%. Another method was proposed [27], based on adaptive gamma correction to improve the brightness of facial images. This step is followed by contour enhancement using the Laplace operator, then weighted linear fusion between the gamma-corrected image and the contour-enhanced image to generate an optimized version of the facial image. The experimental results show an improvement in the MobileFaceNet model, achieving an accuracy of 99.27% using the labeled faces in wild (LFW) database and 90.18% using the AgeDB database. Darma *et al.* [28] presented a method combining a multilayer CNN with a support vector machine (SVM) classifier. Experimental results show that the proposed CNN + SVM model achieves a remarkable accuracy of 99.87%.

This article proposes a mixed approach that combines two feature extraction techniques: those based on conventional methods and those derived from CNN filters, to optimize facial recognition performance and also strengthen the feature extraction phase. The method consists of two main phases: image pre-processing and a second phase using a CNN architecture to train the model. The first stage is carried out using image processing methods capable of firstly improving image quality by correcting contrasts and homogenizing illumination using the histogram equalization technique, then using principal component analysis (PCA) [29] to reduce the dimensionality of the images used by eliminating all irrelevant information while retaining discriminating information, and finally by applying filtering using the HOG descriptor to improve local geometric characteristics. The result of this phase will be divided into two sets: the first train will be used in the learning phase and consists of 80 ù of the images used, and the second test consists of the remaining 20% and will be used in the testing and validation phase of the model. Finally, a CNN architecture consisting of several convolution filters and other fully connected filters will be used to form a robust, high-performance model validated by several metrics capable of recognizing faces effectively even under conditions of occlusion or variation. Simulations of the method were carried out using two databases of ORL faces [30] and University of Manchester Institute of Science and Technology (UMIST) faces [31], which contain facial images taken under conditions representing several variations (lighting, facial expression, and different angles of image capture) and several cases of occlusion (wearing corrective glasses, sunglasses, and beard). The results obtained demonstrated high performance, with accuracy rates up to 99 %.

2. METHOD

This method comprises five essential steps. The first step involves improving image quality by applying histogram equalization. Next, the size of the images is reduced by removing irrelevant data using PCA. The third step involves extracting features enhanced by the HOG descriptor. This is followed by the use of a CNN architecture to train our facial recognition model to be resistant to variations and occlusion. Finally, the model's performance is validated using several metrics. Figure 1 shows all these steps.

2.1. Face database

To evaluate the effectiveness of the proposed facial recognition framework, two publicly available benchmark datasets, namely ORL [30] and UMIST [31], were employed. These datasets were selected because they contain facial images exhibiting variations in facial expressions, head poses, and appearance.

Their characteristics provide a suitable basis for evaluating recognition performance under different acquisition conditions. A brief description of each dataset is provided as follows.

i) ORL dataset [30]: is a facial database comprising 400 images of 40 individuals. All images captured under conditions representing different types of variation (different facial expressions and wearing glasses). Figure 2 shows a sample of the images in this database.

ii) UMIST dataset [31]: is a face database consisting of 564 images of 20 individuals of different genders (male and female) representing a diversity of ethnicities and ages. All the images represent a variation in pose and are taken from different angles, ranging from full profile to frontal view. Figure 3 shows a sample of the images in this database.

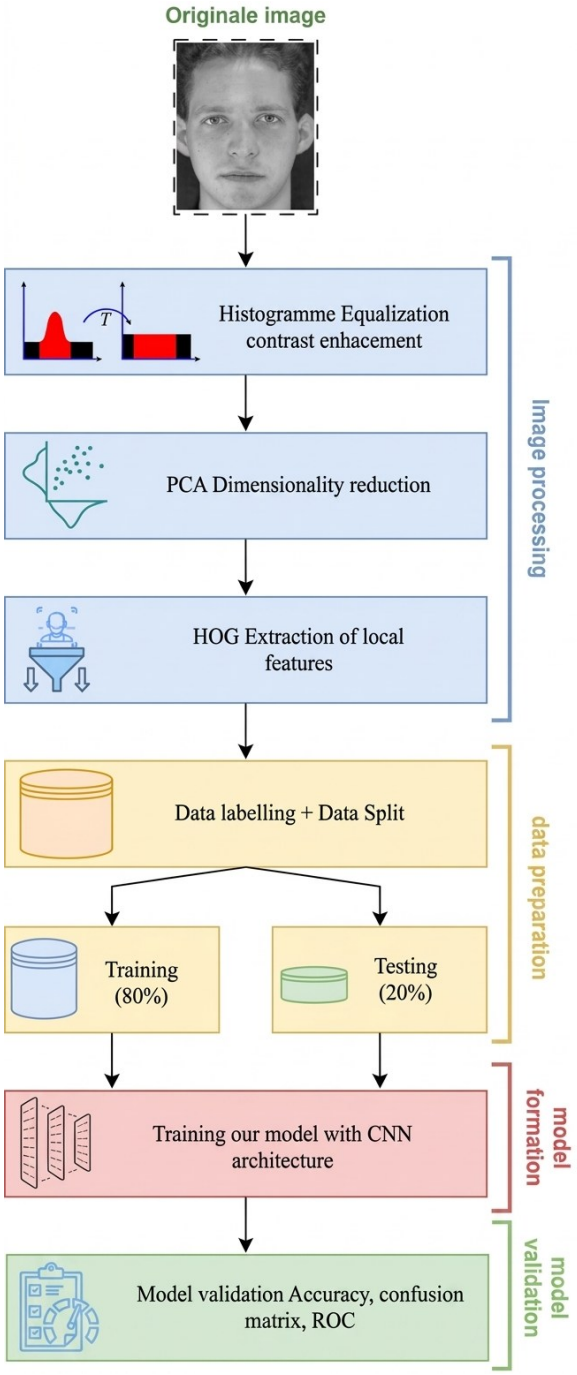


Figure 1. Steps of the proposed method

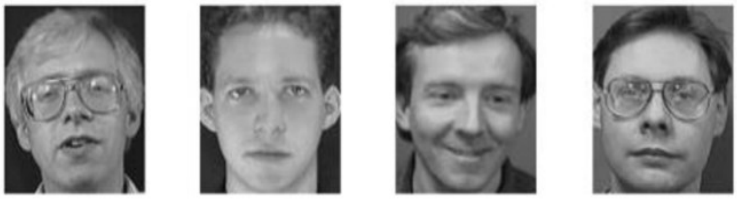


Figure 2. ORL images illustrating variations present in the ORL database



Figure 3. Example image from the UMIST database profile layout from the front view

2.2. Histogram equalization

The first step in the method is the application of histogram equalization, a technique designed by Gonzalez and Woods [32] to improve overall image contrast. This is achieved by distributing the gray levels more evenly. This is beneficial for increasing the quality of detail in underexposed or overexposed areas. This technique is based on transforming the intensity r of a pixel using a cumulative distribution function $T(r)$, as defined in (1).

$$T(r) = (L-1)\sum_{k=0}^{r}\frac{n_k}{N} \quad (1)$$

Where r is the original gray level, n_k is the number of pixels with intensity k, N is the total number of pixels in the image $T(r)$ is the transformed gray level, and L is the total number of available gray levels. Note that the result of the transformation is an image in which gray levels are redistributed to cover the entire image surface, making the output image's histogram more uniform and improving image quality. Figure 4 illustrates the application of this method on one of the images used in this method, which belongs to the ORL database. Figure 4(a) shows the original image and its histograms, Figure 4(b) presents the equalized image and its histograms, and Figure 4(c) compares the original and equalized histograms.

2.3. Principal component analysis

After applying histogram equalization, image dimension reduction was performed using PCA [29], in order to eliminate the irrelevant data contained in each image, while retaining the other discriminating data, and also to reduce noise in order to improve the feature extraction operation. This unsupervised statistical method projects the data into a new, lower-dimensional orthogonal space, the space of principal components. This operation begins with image vectorization, where each 2D image of size $h \times w$ is transformed into a column vector $x_i \epsilon R^d$, where $d = h \times w$, then assembled into a data matrix $x_i \epsilon R^{N \times d}$ which contains the row image vectors where N refers to the total number of images and d the dimension of an image vector, equivalent to the number of pixels. The data are then centered by subtracting the mean of each pixel over the entire image set, to ensure that the variance around the mean is captured. Subsequently, the covariance matrix is calculated using the formula $C = \frac{1}{N} X_c^T X_c$ from the centered data X_c, then decomposed into eigenvalues and eigenvectors.

Each eigenvalue λ_j reflects the amount of variance explained by the associated eigenvector v_j. The principal components correspond to the directions maximizing the variance of the data, and are ordered according to the importance of their eigenvalue. By retaining only, the first k components, a reduced representation of each image is obtained in the form of a vector projected into a lower-dimensional space. The effectiveness of this reduction in size was evaluated using the root mean square error (RMSE), comparing the original images with their reconstruction using the selected principal components. The RMSE score obtained was 5.20, proving the effectiveness of PCA in preserving the essential characteristics of the images during compression while significantly reducing the data size. This is illustrated in Figure 5, which shows the application of PCA to an image resulting from the previous operation, histogram equalization.

Figure 5(a) presents the original image, Figure 5(b) shows the absolute difference between the original and reconstructed images, and Figure 5(c) presents the image reconstructed using PCA.

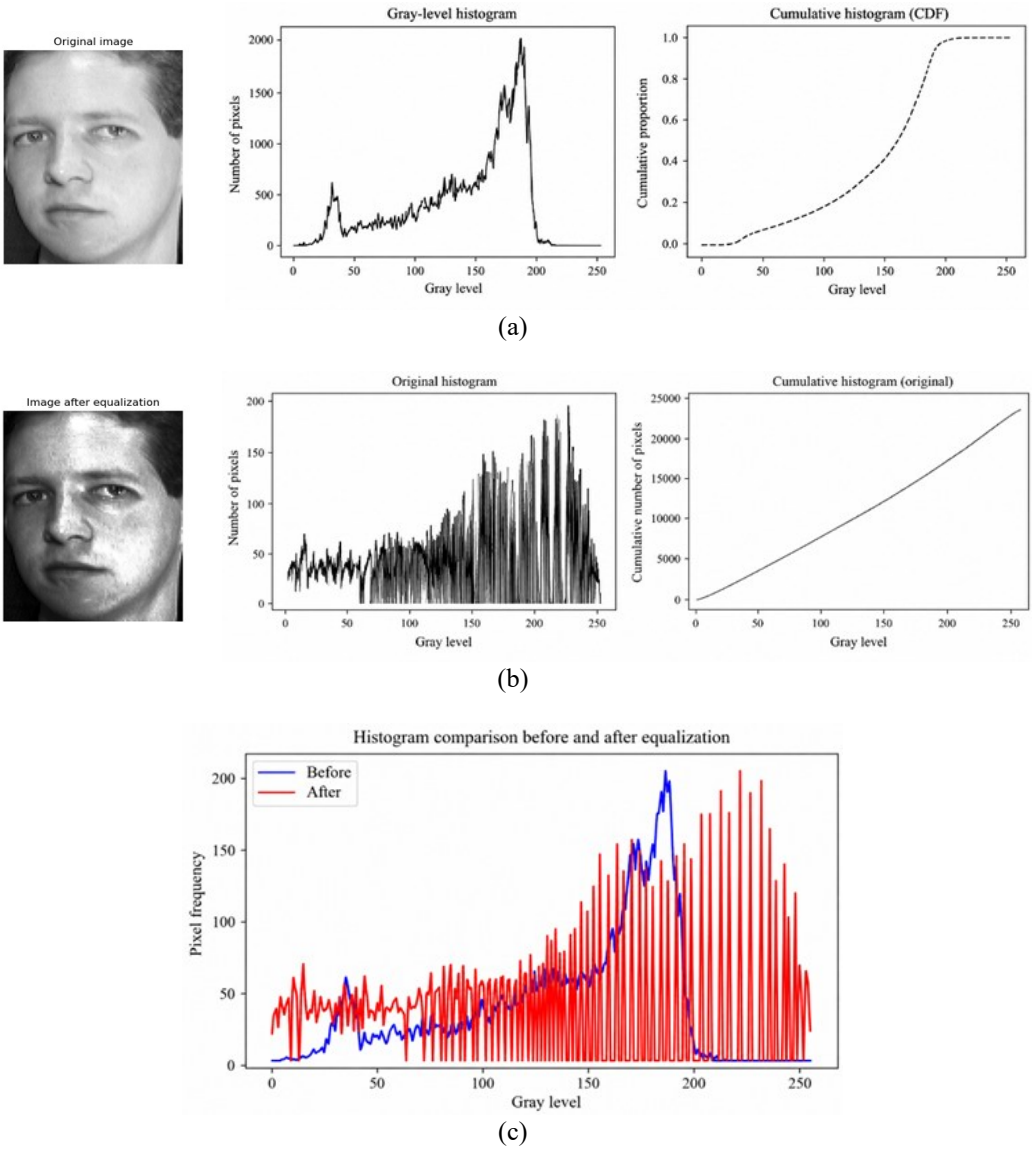


Figure 4. The effect of histogram equalization applied to an image: (a) original image and histograms, (b) equalized image and histograms, and (c) histogram comparison

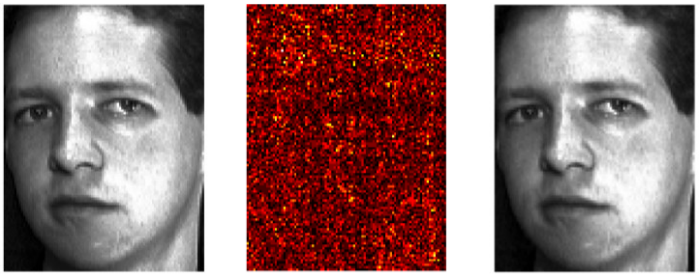


Figure 5. Example of PCA-based image reconstruction: (a) original image, (b) absolute difference, and (c) reconstructed image using PCA

2.4. Feature extraction using the histogram of oriented gradients descriptor

The next step consists of applying the HOG descriptor [21] to the reconstructed images compressed in the previous step using PCA. The objective of this operation is to extract local features that are resistant to variations in lighting and exposure. The process begins by calculating the horizontal and vertical intensity gradients for each pixel, as shown in (2) and (3).

$$G_x(x,y) = I(x+1,y) - I(x-1,y) \tag{2}$$

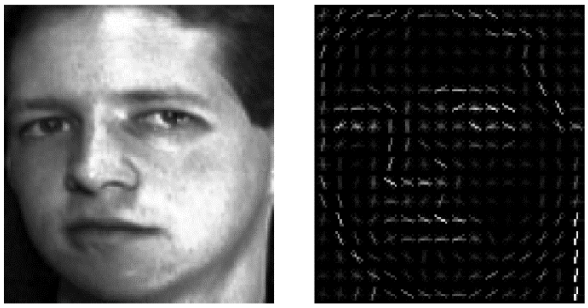
$$G_y(x,y) = I(x,y+1) - I(x,y-1) \tag{3}$$

Where $I(x,y)$ corresponds to the intensity of a pixel at position coordinates (x,y).

Then the amplitude will be deduced from the gradients $|G(x,y)| = \sqrt{G_x^2 + G_y^2}$ and the orientation $\theta(x,y) = \arctan(\frac{G_y}{G_x})$ of the gradient for each pixel. The image is then divided into cells, within which a histogram of amplitude-weighted orientations is constructed. To improve the descriptor's robustness to contrast variations, local normalization is applied to blocks of 2 × 2 cells using the L2 normalization represented by (4).

$$v' = \frac{v}{\sqrt{||v||_2^2 + \epsilon^2}} \tag{4}$$

Where v is the block's histogram vector and ϵ is a small constant value added to avoid division by zero. Finally, the normalized vectors from each block are then concatenated to generate the final feature vector $H = [v'_1, v'_2, \ldots, v'_n]$, effectively representing the geometric and textural structure of the face. Applying the HOG descriptor to PCA-compressed images thus makes it possible to take advantage of both dimension reduction and the robustness of local gradients, with a view to more reliable face recognition. Figure 6 shows an example of the result of feature extraction by the HOG descriptor from an image resulting from the previous steps (histogram equalization and PCA). Figure 6(a) shows the input image, while Figure 6(b) shows the HOG visualization.



(a) (b)

Figure 6. The application of the HOG descriptor to an image: (a) resized original image and (b) HOG visualization

2.5. Proposed convolutional neural network architecture

The final phase of the method is dedicated to the realization of a CNN architecture, which consists of two main components: the first is dedicated to feature extraction, which is the convolution layers based on a succession of convolutional, normalization, and subsampling blocks. The model starts with a first convolution block comprising two Conv2D layers with 32 filters of size 3 × 3 and a rectified linear unit (ReLU) activation. Each convolution is followed by BatchNormalization to stabilize and accelerate training, and 2 × 2 max-pooling to halve resolution. A 25% dropout is applied to limit overlearning. The second block follows the same structure, but with 64 filters, enabling more complex patterns to be captured at an intermediate scale. The third block uses 128 convolutional filters, consolidating high-level representations; this block also ends with a higher dropout (40%), due to the increased number of parameters. Next comes the second component represented by a fully connected layer responsible for the classification part, for which the

feature maps are flattened via a flatten layer, then passed to a dense layer of 256 neurons with ReLU activation function. Figure 7 illustrates the composition of our adopted CNN architecture.

The model was trained using the adaptive moment estimation (Adam) optimizer [33]. It is widely known for its enormous ability to combine the advantages of gradient descent algorithms with momentum and adaptive learning rate optimization. The Adam optimizer performs an update of the weights θ_t based on the first and second moment estimates of the gradient g_t, as presented in (5) and (6).

$$m_t = \beta_1 . m_{t-1} + (1 - \beta_1) . g_1 \quad (5)$$

$$v_t = \beta_2 . v_{t-1} + (1 - \beta_2) . g_t^2 \quad (6)$$

These moments are corrected for initial bias as presented in (7) and (8).

$$\hat{m}_t = \frac{m_t}{1-\beta_1^t} \quad (7)$$

$$\hat{v}_t = \frac{v_t}{1-\beta_2^t} \quad (8)$$

Weights are also updated as shown in (9).

$$\theta_{t+1} = \theta_t - \alpha . \frac{\hat{m}_t}{\sqrt{\hat{v}_t + \epsilon}} \quad (9)$$

Where α represents the learning coefficient, and ϵ represents a small constant ensuring numerical stability.

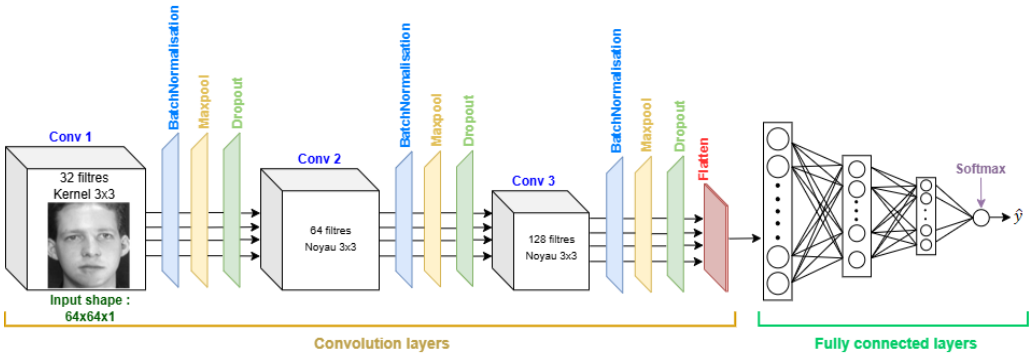


Figure 7. Proposed CNN architecture

2.6. Model evaluation

This model is evaluated using several evaluation metrics, which are:

i) The categorical cross-entropy loss function: a function dedicated to multi-class classification tasks as presented in (10).

$$l = -\sum_{i=1}^{c} y_i \log(\hat{y}_i) \quad (10)$$

Where c represents the number of classes, y_i represents the actual value (1 for the correct class and 0 in the opposite case), and $\hat{y}_i$ represents the probability predicted by the softmax layer.

ii) The accuracy metric for evaluating the accuracy of our model, which represents the proportion of labels correctly predicted on all examples, calculated as shown in (11).

$$Accuracy = \frac{number\ of\ correct\ prediction}{total\ number\ of\ examples} \quad (11)$$

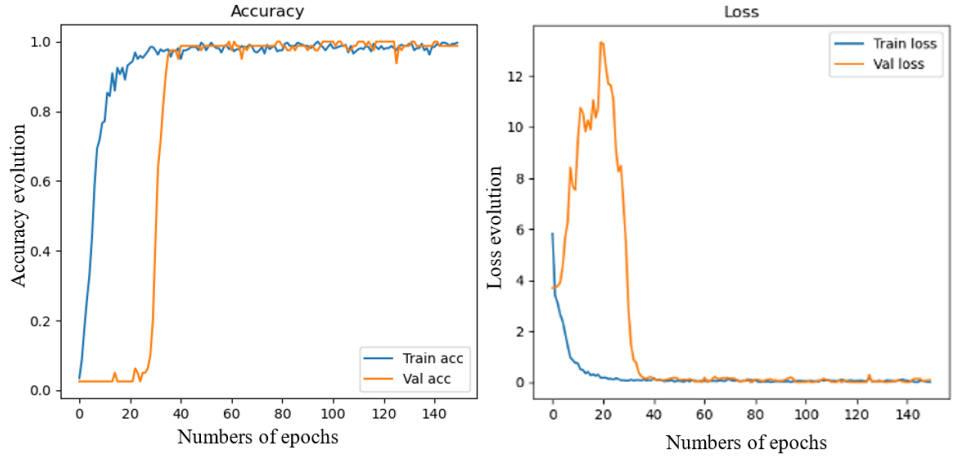
iii) The confusion matrix provides a detailed representation of a model's prediction results by indicating the correct and incorrect predictions for each class.

iv) To measure the discriminatory performance of a model, the receiver operating characteristic (ROC) curve is used. It analyses a model's ability to separate classes by representing the true positive rate as a function of the false positive rate.

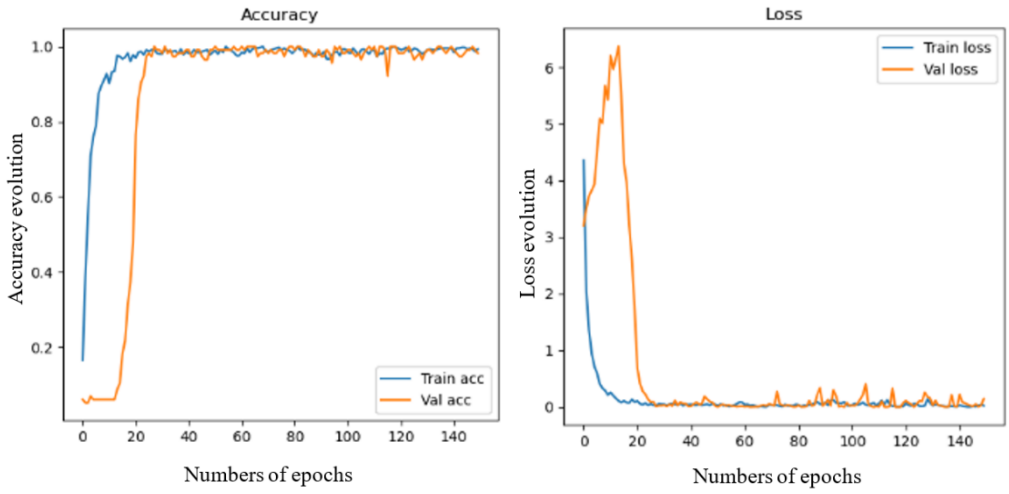
3. RESULTS AND DISCUSSION

This method was tested using the ORL and UMIST databases, each of which contains a variety of variations. The images were divided into two subsets: the first is the training set (train), comprising 80% of the images, and the second is the validation set (test), comprising the remaining 20%, used to validate the model. The training phase was carried out over 150 epochs. The performance evaluation metrics for the model were applied using the test subset.

An analytical reading of the two graphs (accuracy and loss) shown in Figure 8, derived from the experimental results obtained from the ORL database represented by Figure 8(a) and UMIST represented by Figure 8(b), shows an improvement in the accuracy values obtained by our model, achieving perfect accuracy of 99.75% for the ORL database and 99.13% for the UMIST database. Alongside this improvement, the loss value recorded in both cases also fell to extremely low levels. These results demonstrate that, although the learning process required a number of epochs to adapt to the variability of the databases (expressions, poses, and lighting conditions), the model was able to efficiently extract discriminating features thanks to its architecture, ensuring excellent generalization capability on both datasets.



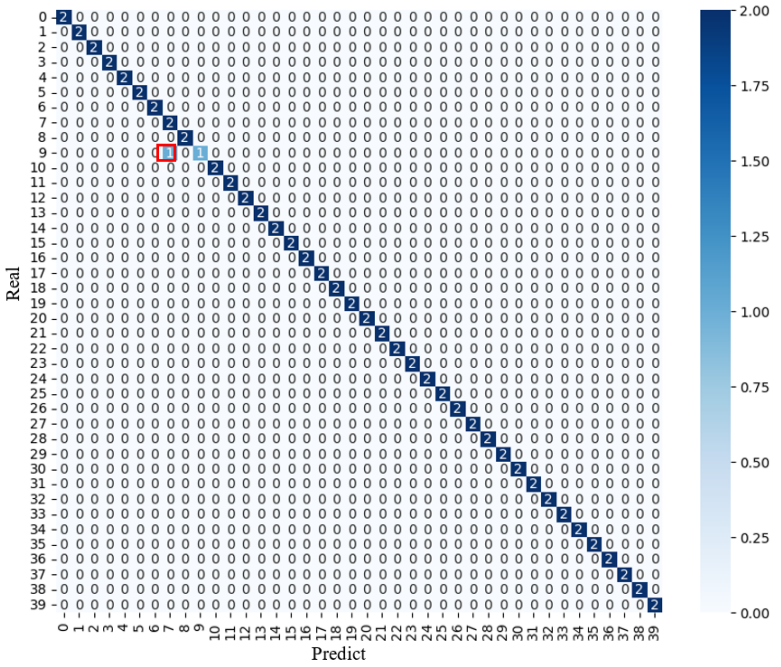
(a)



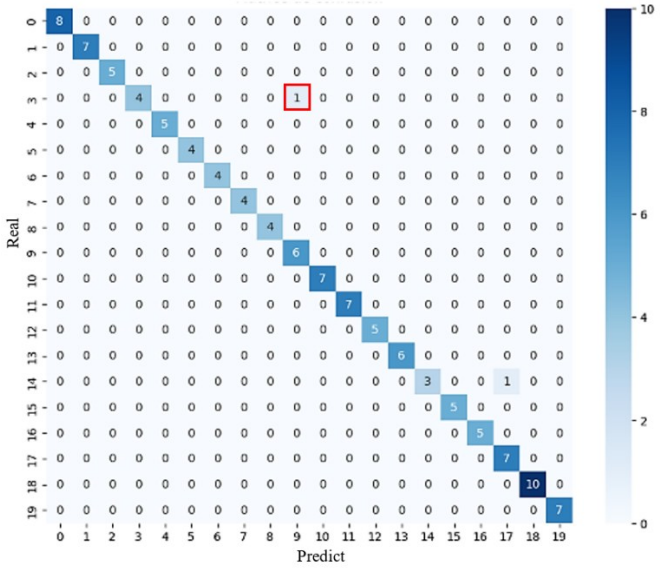
(b)

Figure 8. Plot of precision and loss rates as a function of the number of epochs: (a) ORL database and (b) UMIST database

The confusion matrix shown in Figure 9. This figure represents the results obtained from the ORL database, which corresponds to Figure 9(a), and the UMIST database, which corresponds to Figure 9(b), showing the remarkable performance achieved by our model. For the ORL database and the 80 images tested (two images per individual), 79 were correctly classified, with only one classification error between two visually similar individuals. The accumulation of predictions recorded on the main diagonal illustrates the system's effectiveness in discriminating between faces, resulting in an overall recognition rate of 99.75%. These results confirm the effectiveness and robustness of the model applied to the ORL database. In the case of the UMIST database, excellent results were also achieved, with the majority of images correctly classified. There were a few errors, notably in classes 3, 13, and 14, where one or two images were misclassified. Despite this, the model achieves a high overall accuracy (99.13%), demonstrating its ability to handle a more varied base, with different poses and facial expressions.



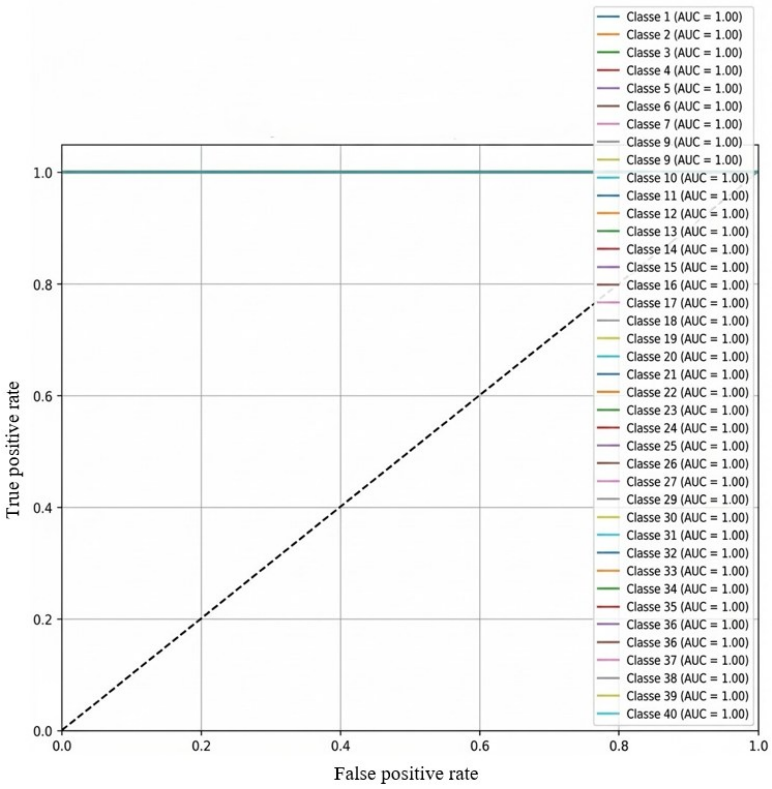
(a)



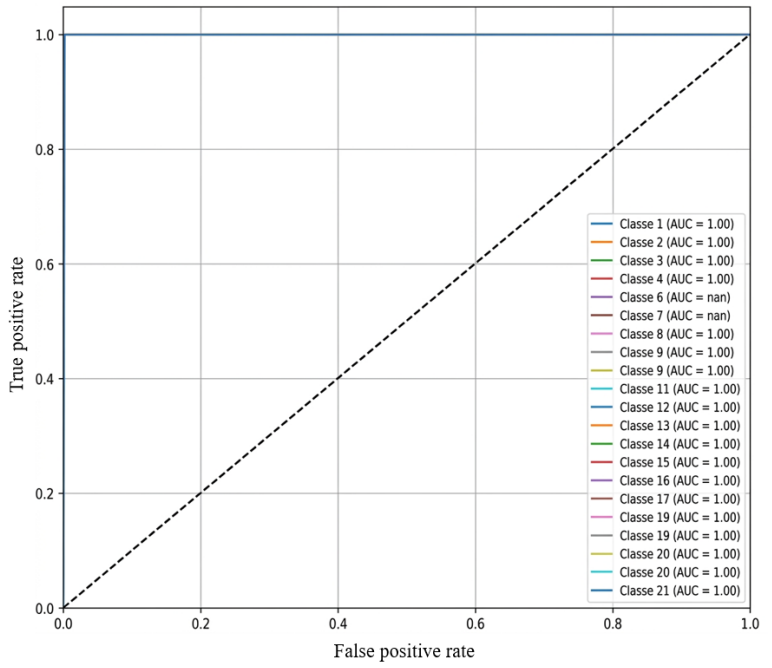
(b)

Figure 9. Evaluation of the performance of the model using the confusion matrix on (a) the ORL database and (b) the UMIST database

The ROC curves shown in Figure 10, obtained from the two databases: ORL represented by Figure 10(a) and UMIST represented by Figure 10(b), show high performance. In both cases, the class curves show an area under the curve equal to 1.00, meaning that the model can perfectly distinguish between images of each class. These results reflect optimal generalization and a very clear separation between the different classes of individuals.



(a)



(b)

Figure 10. Evaluation of the model performance using the ROC curve on (a) ORL and (b) UMIST bases

To further evaluate the effectiveness of the proposed method, a comparative analysis was conducted with existing facial recognition approaches using the same ORL and UMIST datasets, as presented in Table 1. The results demonstrate that the proposed approach achieves competitive recognition accuracy compared with previously published methods. These findings highlight the effectiveness and potential of the proposed hybrid framework for improving facial recognition performance.

Table 1. Comparing the accuracy rates of various works with those of our proposed method

Database	Method	Precision (%)
ORL	Local binary pattern + CNN [34]	100
	Hybrid-supervision learning [35]	96.67
	Particle swarm optimization algorithm	99.00
	Local ternary patterns - deep convolutional neural network [36]	98.75
	Proposed method	99.75
UMIST	Genetic algorithm [37]	100
	Sparse representation classifier & learning discriminative features [38]	99.13
	Cycle-reinforce hierarchical model (CRHM) [39]	99.51
	SIFT + SVM [39]	99.44
	Proposed method	99.13

4. CONCLUSION

This paper presents a new hybrid method for face recognition in adverse conditions with the presence of variations or occlusion by combining image processing techniques with a CNN architecture. The process began by applying histogram equalization to the set of images used to improve their quality, followed by a dimension reduction operation on the images from the previous stage to remove irrelevant data. Then, using the HOG descriptor, the most relevant local features were extracted and passed on to a customized CNN architecture consisting of three blocks in the convolution layer and another fully connected layer to train our model. Simulations of our method were carried out using two databases, ORL and UMIST, and the experimental results confirmed the performance of the approach, exceeding 99% accuracy on both databases. Prospects include the integration of attention mechanisms and the exploitation of other transformer-based architectures, with the aim of improving the adaptability and scalability of the model in more varied and dynamic environments.

FUNDING INFORMATION

Authors state no funding involved.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Hicham Benradi	✓	✓	✓	✓	✓	✓		✓	✓		✓		✓	
Adil Ammar		✓		✓	✓	✓		✓		✓	✓			
Hamza Natek	✓		✓		✓		✓			✓	✓		✓	
Abdelali Lasfar	✓	✓		✓		✓				✓		✓	✓	✓

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

DATA AVAILABILITY

The datasets used in this study, namely the ORL and UMIST face databases, are publicly available for academic research, subject to their respective access conditions. These datasets were used to evaluate the performance of the proposed hybrid deep learning approach for facial recognition. The experimental results

supporting the findings of this study are presented in the article. Additional experimental results and derived data are available from the corresponding author, [HB], upon reasonable request.

REFERENCES

[1] J. Zhang and N. Hu, "Accuracy and robustness evaluation of deep learning algorithms in facial recognition systems," *Systems and Soft Computing*, vol. 7, Dec. 2025, doi: 10.1016/j.sasc.2025.200252.

[2] H. Choung, P. David, and T.-W. Ling, "Acceptance of AI-powered facial recognition technology in surveillance scenarios: role of trust, security, and privacy perceptions," *Technology in Society*, vol. 79, Dec. 2024, doi: 10.1016/j.techsoc.2024.102721.

[3] H. R. V. Kumar and M. Mathivanan, "A novel hybrid biometric software application for facial recognition considering uncontrollable environmental conditions," *Healthcare Analytics*, vol. 3, Nov. 2023, doi: 10.1016/j.health.2023.100156.

[4] X. Zhang, C. Xuan, Y. Ma, H. Su, and M. Zhang, "Biometric facial identification using attention module optimized YOLOv4 for sheep," *Computers and Electronics in Agriculture*, vol. 203, Dec. 2022, doi: 10.1016/j.compag.2022.107452.

[5] J. C. Musale, A. K. Singh, and S. Shirke, "Effective face recognition from video using enhanced social collie optimization-based deep convolutional neural network technique," *Journal of Visual Communication and Image Representation*, vol. 115, Jan. 2026, doi: 10.1016/j.jvcir.2025.104639.

[6] A. M. Rodriguez, Z. Geradts, M. Worring, and L. Unzueta, "Improved likelihood ratios for face recognition in surveillance video by multimodal feature pairing," *Forensic Science International: Synergy*, vol. 8, 2024, doi: 10.1016/j.fsisyn.2024.100458.

[7] L. Quenon *et al.*, "Unfamiliar facial identity discrimination and recognition impairment in prodromal Alzheimer's disease: a behavioral pattern separation and completion study," *Neuropsychologia*, vol. 215, Aug. 2025, doi: 10.1016/j.neuropsychologia.2025.109179.

[8] K. Rezaee, "Machine learning and facial recognition for down syndrome detection: a comprehensive review," *Computers in Human Behavior Reports*, vol. 17, Mar. 2025, doi: 10.1016/j.chbr.2025.100600.

[9] J. Qiang *et al.*, "Evaluation of an AI facial recognition system for Turner syndrome screening and facial complexity: a prospective cohort," *International Journal of Medical Informatics*, vol. 203, Nov. 2025, doi: 10.1016/j.ijmedinf.2025.105976.

[10] O. R. Shahin, Z. M. A. Azim, and A. I. Taloba, "Maneuvering digital watermarking in face recognition," *International Journal of Engineering Trends and Technology*, vol. 69, no. 11, pp. 104–112, Nov. 2021, doi: 10.14445/22315381/IJETT-V69I11P213.

[11] D. A. Abdullah *et al.*, "A novel facial recognition technique with focusing on masked faces," *Ain Shams Engineering Journal*, vol. 16, no. 5, Apr. 2025, doi: 10.1016/j.asej.2025.103350.

[12] S. Khattar, R. Kaur, and A. Kumar, "DeepHybrid-CNN: A hybrid approach for pre-processing of skin cancer images," *Computerized Medical Imaging and Graphics*, vol. 124, Sep. 2025, doi: 10.1016/j.compmedimag.2025.102611.

[13] P. Nanda, D. Rout, and S. Kumari, "Multi-class skin cancer detection using CNN-architecture based deep learning models," *Procedia Computer Science*, vol. 260, pp. 226–235, 2025, doi: 10.1016/j.procs.2025.03.197.

[14] G. Nirmala, P. P. Nayudu, A. R. Kumar, and R. Sagar, "Automatic cervical cancer classification using adaptive vision transformer encoder with CNN for medical application," *Pattern Recognition*, vol. 160, Apr. 2025, doi: 10.1016/j.patcog.2024.111201.

[15] A. T. Mengesha and M. A. Mengistie, "Applying transfer learning in CNN model architectures for detecting tomato leaf disease with explainable artificial intelligence," *Smart Agricultural Technology*, vol. 11, Aug. 2025, doi: 10.1016/j.atech.2025.101034.

[16] M. A. S. Nirob, P. Bishshash, and A. K. M. F. K. Siam, "COLD-12: A multi-level feature extraction hybrid CNN Model for accurate cotton disease diagnosis," *Franklin Open*, vol. 11, Jun. 2025, doi: 10.1016/j.fraope.2025.100263.

[17] J. Padhi, K. Mishra, A. K. Ratha, S. K. Behera, P. K. Sethy, and A. Nanthaamornphong, "Enhancing paddy leaf disease diagnosis -a hybrid CNN model using simulated thermal imaging," *Smart Agricultural Technology*, vol. 10, Mar. 2025, doi: 10.1016/j.atech.2025.100814.

[18] T. N. -Da, P. N. -Thanh, and M.-Y. Cho, "Real-time AIoT anomaly detection for industrial diesel generator based on an efficient deep learning CNN-LSTM in industry 4.0," *Internet of Things*, vol. 27, Oct. 2024, doi: 10.1016/j.iot.2024.101280.

[19] I. Priyadarshini, "An explainable autoencoder-based feature extraction combined with CNN-LSTM-PSO model for improved predictive maintenance," *Computers, Materials & Continua*, vol. 83, no. 1, pp. 635–659, 2025, doi: 10.32604/cmc.2025.061062.

[20] C. Mawardi, A. Buono, K. Priandana, and H. Herianto, "Hybrid convolutional vision transformer for extrusion-based 3D food-printing defect classification," *IAES International Journal of Artificial Intelligence*, vol. 14, no. 4, pp. 3311–3323, Aug. 2025, doi: 10.11591/ijai.v14.i4.pp3311-3323.

[21] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2005, pp. 886–893, doi: 10.1109/CVPR.2005.177.

[22] A. Vashaee, R. Jafari, D. Ziou, and M. M. Rashidi, "Rotation invariant HOG for object localization in web images," *Signal Processing*, vol. 125, pp. 304–314, Aug. 2016, doi: 10.1016/j.sigpro.2016.01.016.

[23] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, Nov. 2004, doi: 10.1023/B:VISI.0000029664.99615.94.

[24] M. Yacoub, M. Abdelwahab, K. Shiokawa, and A. Mahrous, "Estimating the drift velocity of plasma bubbles in airglow images using the scale invariant feature transform and the speeded up robust feature algorithms," *Advances in Space Research*, vol. 75, no. 2, pp. 2391–2402, Jan. 2025, doi: 10.1016/j.asr.2024.09.071.

[25] J. Anil and L. P. Suresh, "A novel fast hybrid face recognition approach using convolutional Kernel extreme learning machine with HOG feature extractor," *Measurement: Sensors*, vol. 30, Dec. 2023, doi: 10.1016/j.measen.2023.100907.

[26] A. I. Awodeyi, O. A. Ibok, I. Omokaro, J. U. Ekwemuka, and M. O. Ighofiomoni, "Effective preprocessing techniques for improved facial recognition under variable conditions," *Franklin Open*, vol. 10, Mar. 2025, doi: 10.1016/j.fraope.2025.100225.

[27] J. Li, Y. Ding, Z. Shao, and W. Jiang, "Face recognition method based on fusion of improved MobileFaceNet and adaptive Gamma algorithm," *Journal of the Franklin Institute*, vol. 361, no. 17, Nov. 2024, doi: 10.1016/j.jfranklin.2024.107306.

[28] A. S. Darma, F. S. Mohamad, O. A. Diekola, and I. M. Sulaiman, "Deep learning approach for face recognition based on multi-layer CNN&SVM," *International Journal of Engineering Trends and Technology*, vol. 71, no. 8, pp. 388–409, Aug. 2023, doi: 10.14445/22315381/IJETT-V71I8P34.

[29] M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of Cognitive Neuroscience*, vol. 3, no. 1, pp. 71–86, Jan. 1991, doi: 10.1162/jocn.1991.3.1.71.

[30] F. S. Samaria and A. C. Harter, "Parameterisation of a stochastic model for human face identification," in *1994 IEEE Workshop on Applications of Computer Vision*, 1994, pp. 138–142, doi: 10.1109/ACV.1994.341300.

[31] D. B. Graham and N. M. Allinson, "Characterising virtual eigensignatures for general purpose face recognition," in *Face Recognition*, Berlin, Heidelberg: Springer, 1998, pp. 446–456, doi: 10.1007/978-3-642-72201-1_25.

[32] R. C. Gonzalez and R. E. Woods, *Digital image processing*. London, United Kingdom: Pearson Education Limited, 1992.

[33] D. P. Kingma and J. L. Ba, "Adam: a method for stochastic optimization," in *3rd International Conference on Learning Representations*, 2015, pp. 1–15.
[34] J. Tang, Q. Su, B. Su, S. Fong, W. Cao, and X. Gong, "Parallel ensemble learning of convolutional neural networks and local binary patterns for face recognition," *Computer Methods and Programs in Biomedicine*, vol. 197, Dec. 2020, doi: 10.1016/j.cmpb.2020.105622.
[35] S. Zhao, W. Liu, S. Liu, J. Ge, and X. Liang, "A hybrid-supervision learning algorithm for real-time uncompleted face recognition," *Computers and Electrical Engineering*, vol. 101, Jul. 2022, doi: 10.1016/j.compeleceng.2022.108090.
[36] A. Zeroual, M. Amroune, M. Derdour, and A. Bentahar, "Lightweight deep learning model to secure authentication in mobile cloud computing," *Journal of King Saud University - Computer and Information Sciences*, vol. 34, no. 9, pp. 6938–6948, Oct. 2022, doi: 10.1016/j.jksuci.2021.09.016.
[37] P. Sukhija, S. Behal, and P. Singh, "Face recognition system using genetic algorithm," *Procedia Computer Science*, vol. 85, pp. 410–417, 2016, doi: 10.1016/j.procs.2016.05.183.
[38] M. Liao and X. Gu, "Face recognition approach by subspace extended sparse representation and discriminative feature learning," *Neurocomputing*, vol. 373, pp. 35–49, Jan. 2020, doi: 10.1016/j.neucom.2019.09.025.
[39] B. Hicham, C. Ahmed, and L. Abdelali, "Face recognition method combining SVM machine learning and scale invariant feature transform," *E3S Web of Conferences*, vol. 351, May 2022, doi: 10.1051/e3sconf/202235101033.

BIOGRAPHIES OF AUTHORS



Hicham Benradi received his Ph.D. in Engineering Sciences, specializing in Artificial Intelligence, from the Mohammadia School of Engineers (EMI), Mohammed V University, Rabat, Morocco, in 2024. He earned his master's degree in Data Engineering and Software Development from the Faculty of Sciences, Rabat, in 2020, and his bachelor's degree in Mobile Application Engineering from the High School of Technology at Salé (EST Salé) in 2018. He is currently an assistant professor of Computer Science and Artificial Intelligence at EST Salé, Mohammed V University, Morocco. His research interests include artificial intelligence, machine learning, deep learning, computer vision, facial recognition, medical image analysis, and data engineering. He can be contacted at email: hicham.benradi@est.um5.ac.ma.



Adil Ammar is currently a senior lecturer (HDR) and serves as the vice president for Information Systems and Digitalization at Mohammed V University in Rabat, Morocco. He is also an associate member of the Laboratory for Systems Analysis, Information Processing, and Industrial Management (LASTIMI). His professional interests include digital transformation strategies in higher education, business process digitization, AI-driven automation, and integrated campus management platforms. He has extensive experience in IT project governance, leading inter-institutional digital initiatives, and deploying large-scale information systems, including electronic signature solutions, online academic services, and digital workflow platforms. He is actively committed to advancing innovation, AI adoption, and modern digital governance within universities through research initiatives, institutional programs, and collaborative projects. He can be contacted at email: adil.ammar@um5.ac.ma.



Hamza Natek is currently a lecturer of Software Engineering at the École Supérieure de Technologie de Salé (EST Salé), Mohammed V University, Rabat, Morocco. He holds a master's degree in Computer Engineering, with a specialization in software architecture and intelligent systems. He has professional experience in full-stack web development. His research interests include model-driven architecture (MDA), artificial intelligence, software engineering, and NoSQL database transformation. He has authored and co-authored several research papers published in Scopus-indexed journals and presented at international conferences. He can be contacted at email: hamzanatek@gmail.com.



Abdelali Lasfar is a professor and researcher in computer science at the Salé School of Technology (Mohammed V University of Rabat). He holds a Ph.D. in Computer Science and is primarily interested in artificial intelligence, information processing, and intelligent systems applied to IoT environments. On the scientific front, he is a member of the Laboratory for Systems Analysis, Information Processing, and Industrial Management (LASTIMI). His research focuses in particular on: i) image content search and indexing and ii) the integration of artificial intelligence for predictive analysis of industrial data. He is the author and co-author of several scientific publications in international journals and conferences. He can be contacted at email: abdelali.lasfar@est.um5.ac.ma.