Face Recognition approach via Deep and Machine Learning

Ola N. Kadhim
Department of Computer Science, Faculty of Computer Science & Mathematics, University of Kufa, Najaf, Iraq
Technical Institute of Al-Mussaib, Al-Furat Al-Awsat Technical University, Najaf, Iraq
ola.najah@atu.edu.iq

Mohammed Hasan Abdulameer
Department of Computer Science, Faculty of Education for Girls, University of Kufa, Najaf, Iraq
mohammed.almayali@uokufa.edu.iq

Abstract— Face recognition is a biometric technology that involves identifying and verifying individuals based on their facial features. It finds applications in security, surveillance, and user authentication systems. The extraction of facial image features and classifier selection are more challenging to identify with conventional facial recognition technologies, and the recognition rate is lower. The paper present proposed model combined between deep wavelet scattering transform network regarding the extraction of features and machine learning for classification purposes. The proposed model consists four stage: obtaining images, performing pre-processing, extracting features, and then applying classification techniques. using both SoftMax classifier (part of deep learning model) and Support Vector Machine classifier (SVM). We used property collected dataset called MULB dataset. The experimental result shows that SVM classifier provide better results than SoftMax classifier. The results from the experiments conducted on the MULB face database showcased the efficacy of the suggested face recognition approach. The proposed method achieved an outstanding recognition accuracy of 98.29% with SVM classifier and 97.87% with SoftMax classifier.

Keywords—Wavelet Scattering Network, Face recognition, Biometric, deep learning.

1 Introduction

Biometric recognition technology relies on physiological or behavioral attributes to identify individuals, has found extensive application across diverse sectors of society. Face recognition technology has emerged as a groundbreaking advancement in the field of biometric identification and surveillance [1]. With its ability to accurately identify individuals through their distinct facial characteristics, this technology has garnered considerable interest and extensive implementation in diverse fields. From enhancing security measures at airports and organizations to improving user authentication systems on smartphones, face recognition has revolutionized the way we interact with technology and ensure safety in different environments [2]. Face recognition, utilizing traditional techniques, has long been a fundamental approach in the field of computer vision and biometric identification. Before the advent of deep learning and advanced algorithms, traditional face recognition methods relied on the extraction of handcrafted features and the application of statistical classifiers. These techniques involved analyzing facial characteristics such as shape, texture, and spatial relationships to establish identity. Although traditional face recognition methods may not possess the same level of accuracy and robustness as deep learning approaches, they have laid the foundation for the development of modern face recognition systems [3]. With deep learning, face recognition has revolutionized the field of computer vision and biometric identification. With its ability to automatically extract and analyze intricate facial features, deep learning has meaningfully improved the accuracy and efficiency of face recognition systems. Through the utilization of deep neural networks, these systems these systems have the capacity to acquire knowledge and understand complex patterns, allowing for robust identification and verification of individuals [4]. The deep wavelet scattering transform, known for its ability to capture multi-scale and invariant representations, has emerged as a promising approach for extracting robust features from facial images. By decomposing facial data into different frequency bands and orientations, the deep wavelet scattering transform effectively captures both local and global information, enabling enhanced face recognition accuracy [5]. The upcoming sections of this paper follow this organization: Section 2 examines related literature, Section 3 outlines the research background, Section 4 presents the methodology and the proposed approach in detail, Section 5 analyzes the experimental results, and finally, Section 6 concludes the paper along with discussing future work.

2 Related Literature

In 2018 Neamah H. Alskeini, et al. [6] To improve face recognition performance, two algorithms were proposed. The initial algorithm utilizes Sparse Representation-based Classification (SRC), Training Image Modification (TIM), Histograms of Oriented Gradients (HOG) descriptors, while also selecting the Maximum Number of Images (MNI) from sub-databases. In order to tackle the dimensionality problem, the entire database is partitioned into multiple sub-databases. The second algorithm involves the use of Convolutional Neural Networks (CNNs). In 2019 Soad Almabdy, et al. [7] employed various approaches using pre-trained Convolutional Neural Network (CNN) architectures. In the first approach, the researchers utilized pre-trained CNN models for example AlexNet and ResNet-50 for the extraction of features, followed by classification using SVM. In the second strategy, they utilized transfer learning based on the AlexNet model. to extract features and perform classification. The accuracy level ranges from 94% to 100%. %. In 2020 Arpita Gupta, et al. [8] proposed architecture for facial emotion recognition involves incorporating attention blocks, residual connections, and convolution networks into a deep self-attention network. Each block within the network serves a specific purpose - residual connections address the vanishing gradients issue, the convolution network extracts features, and attention significantly enhances the network's visual clarity. The suggested model demonstrated exceptional performance when compared to other CNN-based networks, achieving an impressive training accuracy of 85.76% and a validation accuracy of 64.40%. In 2020 Fahima Tabassum, et al. [9] The combination of the Discrete Wavelet Transform (DWT) coherence with four different algorithms, namely principal component analysis (PCA) error vector, PCA eigen vector, Linear Discriminant Analysis (LDA) eigen vector, and Convolutional Neural Network (CNN), was utilized. The results of these four approaches were fused using detection probability entropy and a Fuzzy system, resulting in an impressive recognition rate of 89.56% even in the worst-case scenario. and an impressive 93.34% for the best-case scenario. In 2022 Showkat Ahmad Dar, et al. [10] developed a novel Real Time Face Recognition (RTFR) system, comprising three primary stages. In the initial phase, the system captures real-time video footage and collects 1056 face images from 24 individuals using a camera with a resolution of 112*92. The subsequent stage employs an efficient RTFR algorithm to recognize faces by comparing them against a pre-existing database. This algorithm leverages two distinct deep learning approaches, namely CNN and VGG-16 with Transfer Learning, to improve the performance of the RTFR system. The proposed algorithm demonstrates impressive accuracy, yielding a result of 99.37.

3 Research Background

In this section, a brief explanation of the various tools (algorithms and techniques) employed in this proposal will be provided.

3.1 Multi-task Cascaded Convolutional Neural Networks (MTCNN)

MTCNN is a deep convolution neural network-based approach Regarding the task of detecting faces. and alignment that is capable of simultaneously doing both tasks. When compared to the traditional method, MTCNN demonstrates superior performance, locates the face precisely, moves at a faster rate, and has the ability to detect in real time. It consists of three cascaded neural networks, each designed to perform specific tasks in face detection. These networks work together to accurately detect and localize faces in an image. Before employing these networks, the original image should be scaled to various scales to create an image pyramid in order to achieve face recognition on a unified scale [11]. The overall structure of MTCNN shown in Figure 1.
P-Net (Proposal Network): The first stage of MTCNN is P-Net, which is responsible for generating candidate face regions (proposals) in the input image. It utilizes a fully convolutional network to propose potential face bounding boxes along with confidence scores for face detection. R-Net (Refine Network): The second stage, R-Net, further refines the candidate face regions proposed by P-Net. It filters out false positives and improves the accuracy of face detection by employing a more sophisticated network and fine-tuning the bounding box predictions. O-Net (Output Network): The final stage, O-Net, is responsible for accurately regressing landmarks on the face within the refined

face regions obtained from R-Net. It also performs additional face classification to determine if the detected regions indeed contain faces.

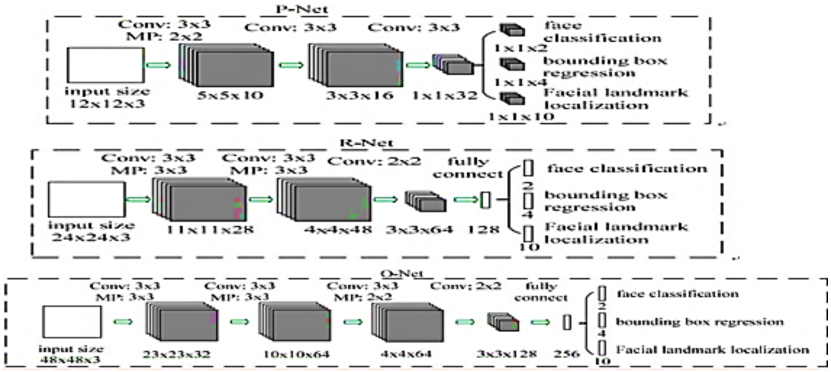

Fig. 1. Overall structure of MTCNN

3.2 Wavelet Scattering Transform Network

Wavelet techniques prove to be powerful instruments for creating meaningful data representations and extracting features, making them compatible with various classification algorithms. The wavelet scattering transform, specifically, enables the generation of dependable features that exhibit local stability against minor deformations. These features can be effectively combined with a deep neural network, enhancing the network's capabilities for tasks like pattern recognition and classification. The Scattering Wavelet Network (ScatNet) was first developed by Mallat [12] Using a series of wavelet transforms in conjunction with a modulus operator. A deep wavelet consists of many layers, where the input of one layer is used as the output of the following layer. Each layer consists of three operations [13], as shown in Figure 2.

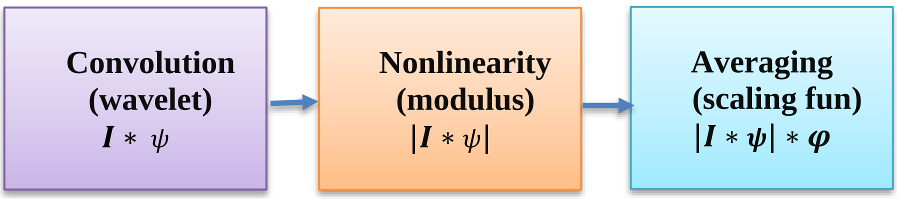

Fig. 2. Operations of Wavelet Scatter where I represents the input data, ψ denotes a wavelet function and φ represents an averaging low-pass filter.

Suppose we have an image (I). The first scattering coefficient is the average of the image and can be obtained by convolving the image with scaling filter (low pass filter) φ_j, as Equation 1 [14]:

$$S_{0,j}(I) = I * \varphi_J \qquad (1)$$

Where j a certain scale.

First-layer scattering coefficients are obtained by applying convolution to the input image (I) using a wavelet filter ($\psi_{\lambda 1}$) at a certain scale (j) and taking the modulus of the resulting coefficients, followed by low-pass filtering with a scaling filter (φ_j). The coefficient is determined by Equation 2 [14]:

$$S_{1,j}(\lambda_1, I) = |I * \psi_{\lambda 1}| * \varphi_j(I) \qquad (2)$$

Second-layer scattering coefficients are obtained by convolving the image (I) by a wavelet filter ($\psi_{\lambda 1}$) at a certain scale (j), taking the modulus of the resulting coefficients, followed by convolving again with another wavelet filter ($\psi_{\lambda 2}$) at a different scale and taking the modulus, and finally low-pass filtering with a scaling filter (φ_j). The coefficient is determined by Equation 3 [14]:

$$S_{2,j}((\lambda_1, \lambda_2), I) = ||I * \psi_{\lambda 1}| * \psi_{\lambda 2}| * \varphi_j(I) \qquad (3)$$

Note that the length of local translation invariance is determined by the parameter (j), which is the breadth of the low-pass filter. In addition to the amount of scales the transform can produce, φ signifies a low-pass filter, ψ denoted wavelet, λ is the rotation operations, since S_1, j and S_2, j are the outputs of low-pass filters, they can be down-sampled according to the filter width 2 power j. The last layer of the Wavelet Scattering network is the average pooling layer, which computes the average of each scattering coefficient over the spatial domain. The network's final output is a vector of all the calculated averages, which captures the low-level and high-level texture features of the input signal or image. Equation 4 determines the final layer's output [14]

$$S_{k,j}\big(((\lambda_1, \lambda_2, \dots, \lambda_k), I\big) = |||I * \psi_{\lambda 1}| * \dots |I * \psi_{\lambda k}| * \varphi_j(I) \qquad (4)$$

where S represents the final output vector, and k represents the network's layer count. The values of $\lambda_1, \lambda_2, \dots, \lambda_k$ represent the possible values of the wavelet filter orientation angles at each layer. The values of j for each layer determine a measure of how many scales the transform can produce. The final output vector S contains information about the local spatial arrangement of texture features, which can be used for various image analysis and classification tasks [14] [15]. Figure 3 show wavelet scattering transform network.

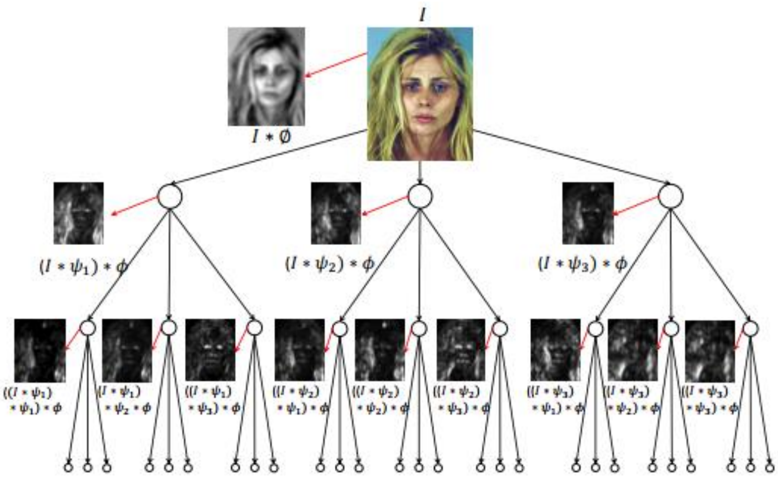

Fig. 3. Wavelet Scattering Network [16]

3.3 SoftMax Classifier

SoftMax classifier is a type of classification algorithm commonly used in machine learning for multi-class classification tasks. It evaluates the probabilities of the input belonging to various classes and allocates it to the class exhibiting the highest probability. Mathematically, given an input vector z = [z_1, z_2, ..., z_k] of k real numbers, the SoftMax function calculates the probability distribution as Equation 5:

$$P(y = i|z) = \exp(z_i)\,/\sum_j^N(\exp\,(z_j)) \quad for\ i = 1,2,3,\dots,k \tag{5}$$

Here, $P(y = i \mid z)$ represents the likelihood of the input belonging to class i. z_i stands for the i-th element of the input vector z. Σ denotes the summation over all elements j in the vector z. N indicates the total number of classes. The SoftMax classifier is often used as the output layer in deep neural networks for multi-class classification tasks, and it plays a crucial role in training and optimizing the network's parameters to minimize the cross-entropy loss between the predicted probabilities and the actual target labels [17].

3.4 Support Vector Machine (SVM) Classifier

SVM is a widely used supervised machine learning algorithm utilized for classification and regression purposes. It is particularly renowned for binary classification tasks, though it can be adapted for multi-class scenarios too. SVM is especially valued for its proficiency in dealing with high-dimensional data and its capability to handle complex datasets with well-defined margin separation. It operates by locating the optimal hyperplane in a high-dimensional space that divides the data into various classes [18].

4 Methodology

Through the use of a wavelet scattering transform network, we are investigating the performance of face recognition in this study. This section explains in detail the proposed model, which consists of four stages: 1) Image Acquisition,2) Pre-processing stage, 3) Features Extraction stage, 4) Classification stage. Figure 4 shows the steps of a proposed model.

4.1 Image Acquisition Stage

In the implementation we start with image Acquisition from face sub-dataset in MULB dataset, then spilt the dataset to 70% for training, 10% for validation, 20% for testing.

4.2 Pre-processing Stage

In this stage, first we detected face in image by using Multi-task Cascaded Convolutional Neural Networks (MTCNN) that presented in section 4.1. Second, we cropping the face images region, third we resize all images to the suitable dimensions. (200×200) for the next stage, finally we transform all input images into grayscale representations.

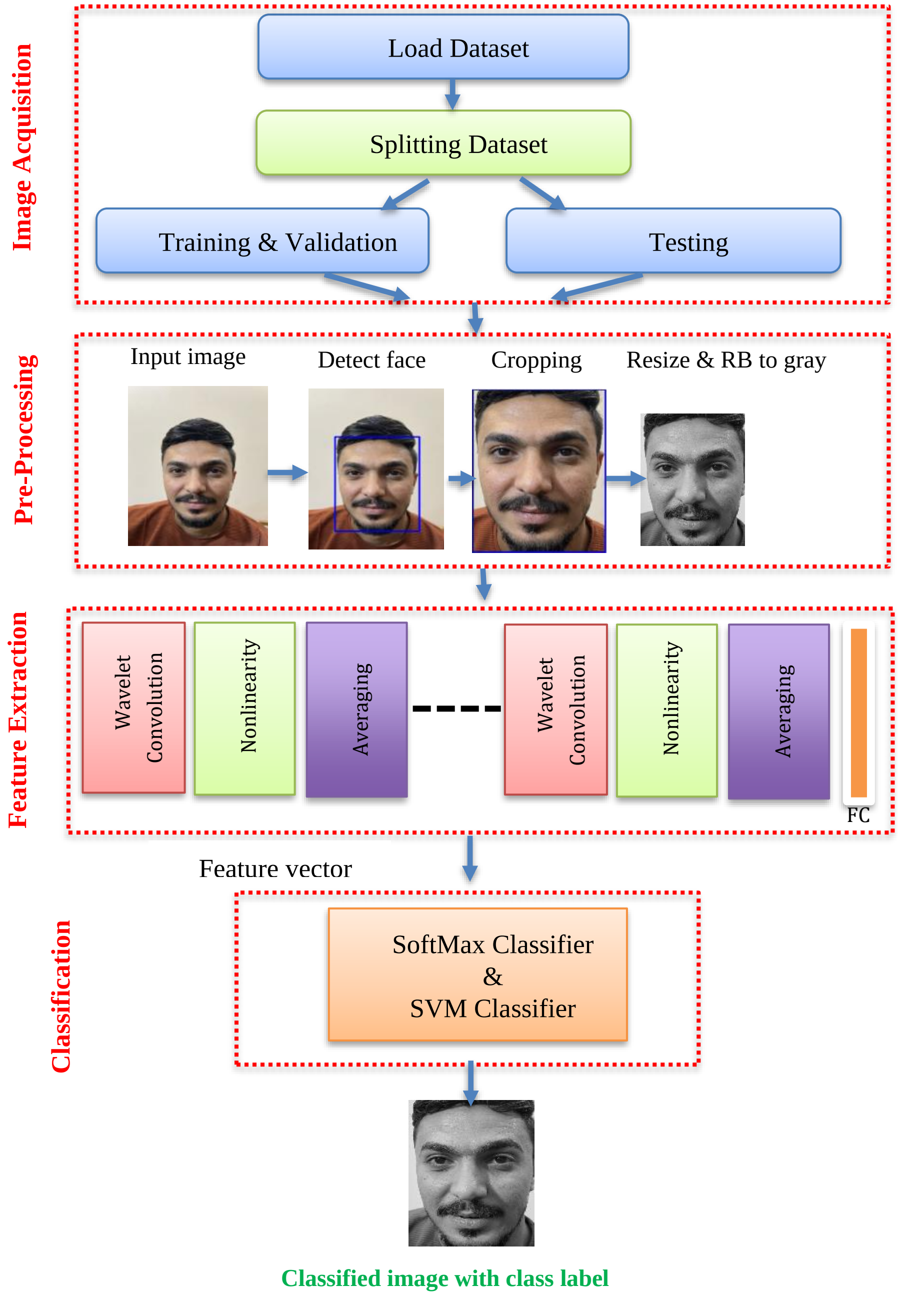

Fig. 4. General Structure of The Proposed Model

4.3 Features Extraction Stage

During this phase, we applied the Wavelet Scattering Transform Network to extract features from the face image, as discussed in section 4.2. The images underwent pre-processing before being passed through a network of wavelet filters, which comprised 3 levels and 10 nodes for each level. At each layer, the image was processed through a group of filters, resembling convolution filters used in a convolutional neural network. After that we use fully connected layer with activation function RELU, this layer enables the network to learn about nonlinear feature combinations by linking Every neuron in the preceding layer to each neuron in the subsequent layer. The outcome at this stage is a set of feature vectors that proves valuable for the classification process.

4.4 Classification Stage

Once the feature vectors are ready, the authentication process utilizes a classifier. During this phase, the test data is classified using the feature vectors stored in the feature database, leading to the identity recognition procedure. This study employs both the SoftMax and SVM classifiers for classification purposes. The architecture of the proposed wavelet combined with SoftMax classifier for Classification is shown in table 1.

Table 1. A concise overview of the proposed wavelet Scattering Model

Layer (type)	Shape of Output	Parameter's number
Input Layer	(200,200)	0
Scattering 2D	(331,25,25)	0
Flatten	(206875)	0
Dense	(512)	105920512
Dense	(176)	90288
Total params: 106,010,800 Trainable params: 106,010,800 Non-trainable parameters: 0		

5 Experimental Results and Analysis

Within this specific section, the proposed model's efficiency is evaluated using face sub-dataset in MULB dataset showed in section 3.

3520 frontal-face images with a bit depth of 24 are included in the MLUB and are divided among 176 participants. Images in (.jpg) format type were taken under various bright lighting situations using a range of expressions, positions, and accessories. We divided it into 70% for training (2464 images), 10% for validation (352 images), and 20% for testing (704).

5.1 MULB dataset

The MULB is a property multimodal biometric dataset, which contains homologous biometric traits. It comprises 20 images of each person's face in various poses, facial expressions, and accessories, 20 images of their right hand from various angles, and 20 images of their right iris from various lighting positions. The database contains real multimodal biometrics from 176 people, and all biometrics were accurately collected using the micro camera of the iPhone 14 Pro Max. The MULB dataset was put together at Al-Furat Al-Awsat Technical University in Kufa, Iraq, during the winter of 2023. A group of 176 individuals, consisting of 118 males and 58 females, with ages ranging from 17 to 54, took part in the data gathering procedure. Each participant had their face, hand, and iris biometric features gathered, resulting in the creation of three sub-dataset in MULB. This dataset will soon be available online and can be accessed without any cost for research and academic purposes. In this paper we use face sub-dataset for face recognition technique. It comprises of 3520 color images of 176 people’s faces.

5.2 Evaluation Metrics

To evaluate the model, this study uses a variety of metrics, including accuracy, loss function, F1 Score, precision, and recall.

Accuracy: It is the predominant metric utilized for assessing classification models. It represents the proportion of accurate predictions out of all the model's predictions. To calculate accuracy, the number of correct predictions is divided by the total number of predictions made by the model. The accuracy can be determined as Equation 6 [7]:

$$ACC = \frac{(T_{pos}+T_{neg})}{(T_{pos}+T_{neg}+F_{pos}+F_{neg})} \qquad (6)$$

Loss Function: It is a mathematical tool employed to measure the alignment between the model's predictions and the actual target values during the training process. It represents the discrepancy or error between the predicted values and the true values. The primary objective of the model during training is to minimize the loss function, aiming for predictions that closely match the actual target values.

Precision: It is an evaluation metric that gauges the proportion of true positive predictions (accurately predicted positive samples) out of all positive predictions made by the model. It serves as an indicator of the model's capability to minimize false positives, indicating how well it identifies positive instances correctly. It can be determined as Equation 7 [7]:

$$Pre = \frac{T_{pos}}{(T_{pos}+F_{pos})} \qquad (7)$$

Recall: It quantifies the fraction of true positive predictions among all actual positive samples present in the dataset. It evaluates the ability of model to avoid false negatives. The recall can be determined as Equation 8 [7]:

$$Rec = \frac{T_{pos}}{(T_{pos}+F_{neg})} \qquad (8)$$

F1-Score: It is determined by taking the harmonic mean of precision and recall. A high F1 Score indicates a balanced trade-off between precision and recall. The F1 score can be determined as Equation 9 [7]:

$$F1_{score} = 2 * \frac{(Pre*Rec)}{(Pre+Rec)} \quad (9)$$

Where, True Positive (T_{pos}): The number of samples that fall under the positive category and that the model correctly identifies as positive. True Negative (T_{neg}): The proportion of samples that the model properly predicts are negative and fall into the negative category. False Positive (F_{pos}): The number of samples that should be classified as negative but are instead the model mistakenly predicts as positive. False Negative (F_{neg}): The proportion of samples that fall into the positive category but that the model mistakenly predicts as negative.

5.3 The Evaluation of the Proposed Model Using Various Evaluation Metrics

First, we evaluated the performance when extracting the image features from deep wavelet scattering transform network followed by SoftMax as a classifier. Second, Image features were extracted from the same network, and subsequently, SVM was employed as the classifier. After a number of training rounds, the best accuracy rate for the testing model was found to be 97.87%, and the loss function was 0.0948 with SoftMax classifier. During classification with the SVM classifier, the highest accuracy rate for the testing model was achieved, reaching 98.29%. Figure 6 and Figure 7 show the accuracy and loss function respectively for proposed model with SoftMax classifier. Table 2 show the evaluation metrics for proposed model. The chart in figure 8 show the accuracy of the model with both classifiers. From Table 2 and Figure 8 we notice the SVM classifier give better result for face recognition model.

Table 2. The Evaluation Metrics for Proposed Model

The Model	Accuracy	Precision	Recall	F1-score
Wavelet+SoftMax	0.9787	0.9899	0.9759	0.9785
Wavelet+SVM	0.9829	0.9865	0.9829	0.9827

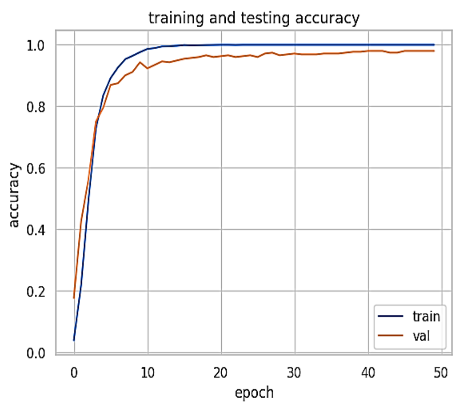

Fig. 6. The Proposed Model Accuracy

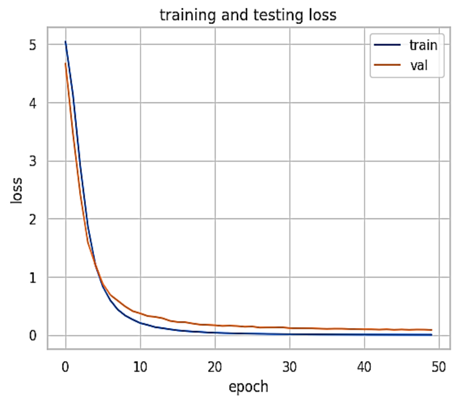

Fig. 7. The Proposed Model Loss function

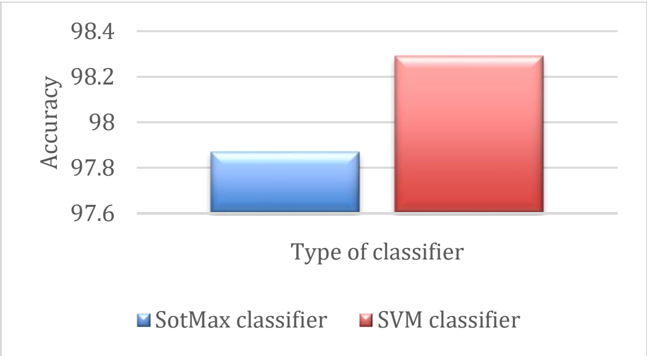

Fig. 8. The Relationship Between the Classifier and the Accuracy of the Model

6 Conclusion and Future Work

In this paper, deep wavelet scattering transform network were applied for feature extraction from face image and machine learning technique (SoftMax & SVM) for classification. According to the experimental results, the SVM classifier has demonstrated superior performance compared to the SoftMax classifier. The proposed face recognition method has been thoroughly evaluated on the MULB face database, showing its effectiveness. Specifically, the SVM classifier achieved an impressive recognition accuracy of 98.29%, while the SoftMax classifier achieved a slightly lower accuracy of 97.87%. In our future research, we plan to incorporate face recognition into multimodal biometric techniques, along with other types of biometric data that have demonstrated distinct and discriminative features in face images.

7 Acknowledgment

Our research is not receiving support from any funding agency.

8 References

[1] M. Hasan Abdulameer and R. Adnan Kareem, "Face Identification Approach Using Legendre Moment and Singular Value Decomposition," International Journal Of Computing Digital System, vol. 13, no. 1, pp. 389-398, 2021.

[2] Y. Kortli, M. Jridi, A. Al Falou, and M. Atri, "Face recognition systems: A survey," Sensors, vol. 20, no. 2, pp. 3-36, 2020.

[3] Z. Xie, J. Li, and H. Shi, "A Face Recognition Method Based on CNN," in Journal of Physics: Conference Series, vol. 1395, no. 1, pp. 1-8: IOP Publisher, 2019.

[4] P. S. Prasad, R. Pathak, V. K. Gunjan, and H. Ramana Rao, "Deep learning based representation for face recognition," in Proceedings of the 2nd International Conference on Communications and Cyber Physical Engineering, pp. 419-424: Springer, 2020.

[5] A. Rehman, M. Harouni, M. Omidiravesh, S. M. Fati, and S. A. Bahaj, "Finger Vein Authentication Based on Wavelet Scattering Networks," Computers, Materials Continua, vol. 72, no. 2, pp. 3369-3383, 2022.

[6] N. H. Alskeini, K. N. Thanh, V. Chandran, and W. Boles, "Face recognition: Sparse representation vs. Deep learning," in Proceedings of the 2nd International Conference on Graphics and Signal Processing, pp. 31-37, 2018.

[7] S. Almabdy and L. Elrefaei, "Deep convolutional neural network-based approaches for face recognition," Applied Sciences, vol. 9, no. 20, pp. 1-21, 2019.

[8] A. Gupta, S. Arunachalam, and R. Balakrishnan, "Deep self-attention network for facial emotion recognition," Procedia Computer Science, vol. 171, pp. 1527-1534, 2020.

[9] F. Tabassum, M. I. Islam, R. T. Khan, and M. R. Amin, "Human face recognition with combination of DWT and machine learning," Journal of King Saud University-Computer Information Sciences, vol. 34, no. 3, pp. 546-556, 2022.

[10] S. A. Dar and S. Palanivel, "Performance Evaluation of Convolutional Neural Networks (CNNs) And VGG on Real Time Face Recognition System," Advances in Science, Technology and Engineering Systems Journal, vol. 6, no. 2, pp. 956-964, 2021.

[11] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao, "Joint face detection and alignment using multitask cascaded convolutional networks," IEEE signal processing letters, vol. 23, no. 10, pp. 1499-1503, 2016.

[12] S. Mallat, "Group invariant scattering," Communications on Pure Applied Mathematics, vol. 65, no. 10, pp. 1331-1398, 2012.

[13] B. Soro and C. Lee, "A wavelet scattering feature extraction approach for deep neural network based indoor fingerprinting localization," Sensors, vol. 19, no. 8, pp. 1-12, 2019.

[14] J. Andén and S. Mallat, "Deep scattering spectrum," IEEE Transactions on Signal Processing, vol. 62, no. 16, pp. 4114-4128, 2014.

[15] Z. Liu, G. Yao, Q. Zhang, J. Zhang, and X. Zeng, "Wavelet scattering transform for ECG beat classification," Computational mathematical methods in medicine, pp. 1-11, 2020.

[16] P. Pandey, R. Singh, and M. Vatsa, "Face recognition using scattering wavelet under Illicit Drug Abuse variations," in International Conference on Biometrics (ICB), pp. 1-6: IEEE, 2016.

[17] I. Kouretas and V. Paliouras, "Simplified hardware implementation of the softmax activation function," in 8th international conference on modern circuits and systems technologies (MOCAST), pp. 1-4: IEEE, 2019.

[18] C.-C. Chang and C.-J. Lin, "LIBSVM: A Library for Support Vector Machines," ACM Transactions on Intelligent Systems and Technology, vol. 2, no. 3, pp. 1-27, 2011.

Article submitted 13 July 2023. Accepted at 21 August 2023. Published at 30 September 2023.