Jordanian Journal of Computers and Information Technology (JJCIT), Vol. 09, No. 04, December 2023.

CAN THE COMBINATION OF FACIAL FEATURES ENHANCE THE PERFORMANCE OF FACE RECOGNITION?

Djellab Issam[1], Laimeche Lakhdar[2] and Redjimi Mohamed[3]

(Received: 18-Jul.-2023, Revised: 7-Sep.-2023 and 3-Oct.-2023, Accepted: 3-Oct.-2023)

ABSTRACT

In recent years, researchers have investigated into various approaches of data combination for face recognition, opening up a novel path of exploration aimed at enhancing recognition reliability by capitalizing on the synergy inherent in diverse data sources. This paper implements a comprehensive comparison between two combination methods based on the score-level and feature-level combination, to determine which method highly improves the overall system performance. In the initial method called Fusion-based Classifier Combination (FCC), we introduce a new fusion rule based on score-level combination. This novel model comprises three classifiers; each trained utilizing well-established feature extraction techniques: Local Binary Patterns (LBP), Histogram of Oriented Gradients (HOG) and Compact Binary Facial Descriptors (CBFD). Instead of adhering to conventional combination rules, such as majority vote or maximum scores, the derived scores from each classifier are merged and then trained using a Multi-Layer Perceptron (MLP) classifier to reach the final decision. In the subsequent method, named Sequential CNN deep learning-based face recognition (S-CNN), we extract high-level features from multiple image regions considered as sequential data, employing an ensemble of Convolutional Neural Networks (CNNs). In this scheme, the fully connected layers of each CNN-based image region are combined and fed into a Deep Neural Network (DNN) tailored for facial recognition. The experimental results obtained from well-known face datasets, including Labeled Faces in the Wild (LFW), Olivetti Research Laboratory (ORL) and IARPA Janus Benchmark-C (IJB-C) highlight the competitive performance of both the proposed multi-classifier combination model and the S-CNN deep-learning model when compared to state-of-the-art methods.

KEYWORDS

Classifier combination, Deep learning, Ensemble CNN, Face recognition, Machine learning.

1. INTRODUCTION

In recent years, there has been significant progress in the development of biometric systems for person recognition, which utilize physiological or behavioral biometric modalities. Among these modalities, facial technology has gained widespread popularity in biometric systems due to its ability to achieve a balance between user acceptability and system accuracy. Facial recognition offers valuable information for identifying individuals, including their identity, gender, ethnicity, age and emotional expressions.

In our daily lives, we have a natural ability to swiftly and easily recognize individuals based on their facial features. This remarkable capability extends to recognizing people from photographs, as it remains robust against variations in facial characteristics, viewing angles, lighting conditions and poses. Humans can effortlessly handle challenges, such as occlusions, changes in facial expressions, hairstyles or the effects of aging. However, training a computer system to perform face recognition, a task that humans excel at, presents a significant challenge. Researchers have conducted studies to investigate the human perception of face recognition from facial images, drawing insights from psychological findings. Furthermore, the existence of biological relationships and observable similarities in traits within the same family motivated researchers to leverage this phenomenon and develop face-recognition systems [1].

A face-recognition system is a biometric system that consists of two primary processes: enrolment and testing, which involve verification or identification. During the enrolment process, it is crucial to store a person's facial biometric features extracted from reliable samples in a dataset. These stored features

1. D. Issam is with Dep. of Computer Science, Badji Mokhtar Univ., Annaba, Algeria. Email: djellab_issam@univ-khenchela.dz
2. L. Laimeche is with Dep. of Computer Sci., Larbi Tebessi Univ., Tebessa, Algeria. Email: lakhdar.laimeche@univ-tebessa.dz
3. R. Mohamed is with Dep. of Computer Science, 20 Août 1955 Univ., Skikda, Algeria. Email: djellab_issam@univ-khenchela.dz

are later compared with the extracted features from the traits of the person whose face biometric features need to be verified or identified.

The testing process includes two modes: verification and identification. In the verification mode, the system verifies a person's facial biometric features by comparing the captured facial biometric features with the biometric template stored in the system dataset. This mode involves a one-to-one comparison to determine the authenticity of the biometric face relation. On the other hand, in the identification mode, the system aims to recognize a genuine user by searching for matching templates provided by the user within the dataset. The system performs one-to-several comparisons to identify an individual entity or fails to identify it if the subject is not enrolled in the system dataset [2].

In the existing literature, various techniques for facial recognition have been proposed, which can be broadly classified into two categories: handcrafted feature-based methods [3]-[6] and deep-learning data-based classification methods [7]-[19]. These techniques have achieved notable success in terms of facial identification or verification. However, despite their achievements, there are still several challenges that need to be addressed in the field of facial recognition. These challenges can be classified into two types. The first type is directly contested challenges, which are related to the relationships between faces. These challenges involve tasks, such as handling variations in pose, illumination, facial expressions and occlusions. The ability to accurately capture and represent these relationships is crucial for robust facial-recognition systems. The second type is indirectly contested challenges, which pertain to the dataset environment. These challenges include issues, such as limited data availability, imbalanced datasets and privacy concerns. Addressing these challenges is essential for developing reliable and efficient facial-recognition systems.

While previous methods, such as handcrafted feature extraction and deep learning, have shown potential in overcoming challenges in facial recognition, further research and innovation are needed to enhance the performance and reliability of these systems [20]. In this paper, we propose two novel facial recognition systems; namely, FCC and S-CNN systems, to address these challenges and achieve our goals.

The FCC system comprises three essential components. First, it employs a combination of handcrafted feature-extraction techniques, including LBP [21] and HOG [22] feature descriptors, along with a learned feature-extraction technique called CBFD [23]. This combination enables the extraction of relevant facial features. Second, a set of Support Vector Machines (SVMs) is utilized to generate face-recognition scores for each feature representation. Lastly, a Multi-Layer Perceptron (MLP) classifier serves as a combination model, integrating the SVM scores to determine the optimal fusion of information.

The S-CNN system consists of three key components. Firstly, it focuses on facial regions and treats them as sequential data by employing a series of CNNs. This means that the recognition of a facial region takes into account not only the current input, but also the knowledge acquired from previously processed facial regions. Secondly, the most effective fully connected layers obtained from the recognition of facial regions by each CNN are combined. Lastly, the extracted features are fed into a Deep Neural Network (DNN) for facial-recognition purposes.

The organization of the remaining sections in this paper is as follows: In Section 2, we offer a comprehensive review of recent studies that utilize machine-learning and deep-learning models for face recognition. Section 3 provides a detailed explanation of the proposed models. In Section 4, we present and analyze the experimental results obtained from our models. Furthermore, in Section 5, we conduct a comparative study between our proposed models and similar methods. Finally, in Section 6, we conclude the paper by summarizing the findings and presenting the final conclusions.

2. EARLY WORK

To ensure a comprehensive overview, we will divide this section into two parts, specifically addressing the utilization of different models. The initial part will encompass Machine Learning (ML) techniques, while the subsequent part will explore different methods rooted in Deep Learning (DL).

The study described in [3] focuses on the practical implementation of a sophisticated algorithm called FaceNet. The algorithm is employed within an access control system designed to effectively detect

faces and eyes, even under challenging lighting conditions. This detection capability is achieved through the utilization of face-encoding techniques. Additionally, facial feature extraction is performed using the HOG algorithm. The access control system includes a Compare Face function that incorporates a Support Vector Machine classifier to classify the face encodings and generate the desired output. To further enhance the system's functionality, RFID sensors and IR sensors are seamlessly integrated. Furthermore, a dedicated webpage is developed, offering access control management for the respective campus or organization. This webpage serves as a user-friendly interface, enabling manual control of the access system whenever necessary. Overall, the study presents a comprehensive solution for access control by effectively implementing the FaceNet algorithm in conjunction with various sensors and a user-friendly webpage interface.

Lakshmi and Ponnusamy in [4] introduced a novel feature descriptor for facial-expression recognition. The proposed approach combines the modified HOG and LBP feature descriptors. The methodology consists of several steps. Firstly, the Viola-Jones face-detection algorithm is employed to locate the facial region. Then, a Butterworth high-pass filter is applied to enhance the detected region, enabling the identification of the eye, nose and mouth regions using the Viola-Jones approach. In the next step, the proposed modified HOG and LBP feature descriptors are utilized to extract features from the detected eye, nose and mouth regions. These features are then concatenated and their dimensionality is reduced using Deep Stacked Auto Encoders. Finally, a multi-class Support Vector Machine classifier is employed for classification and recognition of facial expressions. The experimental results demonstrate the effectiveness of the proposed modified feature descriptors in accurately recognizing emotions on the CK+ dataset and JAFFE dataset.

In [5], Wanling and Shijun proposed an effective approach for face anti-spoofing using a combination of Discrete Wavelet Transform (DWT), LBP and Discrete Cosine Transform (DCT), along with an SVM classifier. The proposed strategy involves several steps. Initially, DWT features are generated by decomposing selected frames into various frequency components within 8x8 multi-resolution blocks. Next, DWT-LBP features are constructed to capture the spatial information of these blocks by horizontally connecting the LBP histograms of the corresponding DWT blocks in each frame. Subsequently, DWT-LBP-DCT features are obtained by vertically applying DCT operations on the DWT-LBP features, incorporating temporal information from the video file. This process enables the extracted DWT-LBP-DCT features to effectively represent the frequency-spatial-temporal characteristics of the video. Finally, an SVM classifier with a Radial Basis Function (RBF) kernel is trained for face anti-spoofing. Experimental evaluations conducted on two benchmark datasets; namely, REPLAY-ATTACK and CASIA-FASD, revealed that the proposed approach achieves high detection performance compared to existing methods.

In [6], a novel approach for three-dimensional face recognition is introduced, which combines the LBP feature descriptors and SVM classifier. The proposed method involves two main steps. Firstly, the LBP algorithm is utilized to extract relevant feature information from the three-dimensional face depth image. Subsequently, the SVM algorithm is employed to classify these extracted features. To evaluate the effectiveness of the proposed method, samples are selected from the Texas Three-dimensional Face Recognition (3DFRD) and a custom-built depth dataset. The experimental results demonstrate that the algorithm achieves a higher recognition rate while also reducing the computational time required for recognition.

In [7], a novel algorithm based on the Laplacian pyramid for deep 3D face recognition, which has practical applications in public settings, is proposed. The algorithm incorporates multi-mode fusion, dense 3D alignment and multi-scale residual fusion techniques. The approach begins by utilizing a 2D to 3D structure representation method to effectively capture information from key facial landmarks and perform dense alignment modeling. Subsequently, a five-layer Laplacian depth network is constructed using the 3D facial landmark model. During the training process, a multi-scale residual weight is integrated into the loss function to enhance the performance of the network. To ensure real-time performance, the proposed network is designed as an end-to-end cascade. This design allows for both accurate identification and efficient personnel screening, particularly in the context of epidemic control measures. The algorithm enables fast and high-precision face recognition, facilitating the establishment of a 3D face dataset. It demonstrates adaptability and robustness in challenging environments characterized by low light and noise, while also being capable of handling various skin colors and postures for face reconstruction and recognition.

Mamieva et al. [8] introduced a novel face-detection technique based on deep learning. The technique consists of two components: a region-offering network (RON) and a prediction network. The RON generates a list of area proposals that are likely to contain faces or Regions of Interest (RoIs). The prediction network is responsible for classifying these areas and refining the bounding boxes around the detected faces. Both components share common parameters with the feature-extraction convolution layers, allowing the architecture to achieve competitive performance in face-detection tasks. To train the model, the authors utilized the WIDER FACE dataset. The experimental results demonstrate that their method excels in face-identification tasks by achieving higher accuracy despite having a smaller model size and efficient computation.

In [9], the study introduces a ResNet-100-based feature embedding network combined with cutting-edge loss functions, including Center Loss, Marginal Loss, Angular Softmax Loss, Large Margin Cosine Loss and Additive Angular Margin Loss. They conduct a comprehensive evaluation involving face-verification and identification tasks, utilizing IJB-B and IJB-C datasets for assessing performance across pose, illumination and expression variations (PIE), FG-Net dataset for age-related analysis and SCface for low-resolution image scenarios. The MS-1MV2 dataset is used as the primary training dataset for system development. Following this, the study evaluates the performance of the network with the most suitable loss function for recognizing synthetic masked faces on the real masked face dataset, the cleaned RMFRD (c-RMFRD) dataset.

The rise of deep learning and its remarkable achievements across various domains have motivated numerous researchers in the energy-consumption field to adopt these techniques for electricity-consumption forecasting modeling. Thus, Sanchez-Moreno et al. in [10] proposed a novel face-recognition approach utilizing the YOLO-Face method for face detection. For the classification stage, they explored the concept of replacing the fully connected layer in a convolutional neural network (CNN) with a support vector machine (SVM) and analyzed the use of random forest (RF) and K-Nearest Neighbors (KNN). Their experimental results demonstrated that the FaceNet+SVM model achieved a high accuracy rate of 99.7% on the LFW dataset. Additionally, the FaceNet+KNN and FaceNet+RF models achieved accuracies of 99.5% and 89.1%, respectively, on the same dataset.

The authors in [11] introduced a novel face-recognition method that effectively tackles the difficulties associated with illumination and misalignment. Their proposed approach combines the LBP feature descriptors with the Improved Pairwise-constrained Multiple Metric Learning method (IPMML). Initially, LBP is utilized to extract texture features from the face images. Subsequently, Linear Discriminant Analysis (LDA) is employed to reduce the dimensionality of the features. The Fisher features are then partitioned into sub-blocks, treating each block as a column vector. By employing the IPMML classification metric, an optimal Mahalanobis matrix is derived. This matrix is used to compute the discriminative distance for face recognition. Finally, the Nearest Neighbor Classifier (NNC) is employed to classify the face images. Experimental results showed the effectiveness of the proposed method, achieving high recognition rates and displaying robustness against challenges, like illumination variations, facial-expression variation and misaligned face images.

A novel approach to face recognition that combines parallel ensemble learning of LBP feature descriptors and CNN is proposed by Tang et al. in [12]. By utilizing LBP for texture feature extraction and employing the extracted features as training data for the parallel CNN, the method effectively improves face-recognition accuracy by mitigating the adverse effects of illumination variations on facial features. The CNN architecture incorporates several crucial components to enhance its performance. The Inception module is employed to widen the network and improve its ability to represent complex features. Batch normalization is utilized to accelerate the training process and enhance convergence. Furthermore, skip connections are incorporated to facilitate information flow across different layers and boost recognition accuracy. The parallel ensemble learning strategy transforms the network structure from a single network into an ensemble, significantly augmenting the accuracy and generalization capabilities of the proposed approach.

To assess the performance of the proposed method, comprehensive experiments were conducted, comparing it with three other methods: Principal Component Analysis (PCA), HOG-CNN and CNN were used independently. The consistently superior results demonstrate the effectiveness of the proposed approach in face-recognition tasks, emphasizing its notable accuracy and efficacy in handling illumination challenges.

In their work presented in [13], the authors introduced an innovative loss function called the "Large Margin Cosine Loss" (LMCL). This loss function is developed by redefining the Softmax loss as a cosine loss, achieved through L2 normalization of both feature vectors and weight vectors to eliminate radial variations. Additionally, a cosine margin term is incorporated to enhance the decision margin within the angular space. Consequently, this approach leads to the minimization of intra-class variance and the maximization of inter-class variance, thanks to the normalization and the maximization of the cosine-based decision margin.

Zhao et al. in [14] introduced an innovative algorithm called iterative Multi-Output Random Forests (iMORF) for enhanced performance in multiple face-analysis tasks. The algorithm explicitly models the relationships among these tasks and iteratively leverages these relationships to improve overall performance. The iMORF algorithm adopts a hierarchical approach to face analysis, with a top-level forest dedicated to pose and expression classification and a bottom-level forest focused on regression of landmark positions. By estimating pose and expression, the algorithm incorporates a strong shape that constrains the variation of landmark positions. Additionally, the estimated landmark positions provide more discriminative shape-related features, further enhancing pose and expression predictions. This iterative exploitation of the interconnectedness between face-analysis tasks continues through cascaded hierarchical face-analysis forests until convergence is achieved. Through experiments conducted on publicly available real-world face datasets, the authors demonstrated that the proposed iMORF algorithm significantly improves the performance of each individual task involved in face analysis.

Muqeet and Holambe in [15] introduced a novel approach for extracting facial features that are robust to variations in expressions and poses. The method utilizes the Directional Wavelet Transform (DIWT)-based LBP histogram features and employs an efficient quadtree partitioning scheme to implement the DIWT. By utilizing the DIWT, the approach enables adaptive directional selection based on image characteristics and represents image edge manifolds. The combination of multi-region LBP histogram features from the top level sub-bands {LL, HL, LH} forms a highly efficient feature set. To evaluate the proposed method, various face datasets are used and the results demonstrate its superior discrimination ability. Compared to other methods, the proposed approach achieves the best rank-one recognition results. The experimental findings indicate that this work outperforms holistic approaches, like the texture feature LDA technique and Locality Preserving Projections (LPP), as well as local descriptors, such as LBP, Local Directional Patterns (LDP) and Weber local descriptors (WLD) methods when dealing with face images containing varying levels of expressions and pose variations. Moreover, this work exhibits better performance compared to non-adaptive LBP-based Multiresolution Analysis (MRA) methods, like Local Gabor Binary Patterns (LGBP), LSPBPS and CTLBP.

The work proposed in [16] introduced several modifications to enhance the performance of the network model for face recognition. These modifications include replacing the traditional convolutional layer with an MLP convolutional layer to improve feature extraction. Additionally, the MFM activation function is incorporated to effectively separate noise signals from information signals, thereby improving recognition. The inclusion of the Center Loss function reduces the distance between elements and improves generalization of learned features. Through extensive experiments, the network model demonstrates promising results. In large-scale face-prediction classification experiments, the model achieves a recognition rate of 82.3%. Furthermore, in face-verification experiments conducted on the LFW face dataset, the model achieves an accuracy rate of 84.5%, indicating high recognition performance. The experiments conducted on face images captured under different conditions showcase the robustness of the network model, except for slightly lower accuracy in face verification with side faces. Overall, the network model exhibits effective recognition.

In [17], a comprehensive framework called 3DPalsyNet was introduced for detecting mouth motion and grading facial palsy. The framework utilizes a modified 3D CNN architecture with a ResNet backbone to capture the dynamic actions present in video data. The performance of the proposed architecture was assessed using two datasets, resulting in an F1-score of 82% for mouth-motion detection and an impressive F1-score of 88% for facial-palsy grading.

In [18], the authors introduced a new approach utilizing PCANet as the foundation, combined with linear SVM and NN classifiers. The PCANet model, as outlined in this study, consists of two stages

for feature extraction and a single nonlinear output stage. The extracted features are then separately utilized in the linear SVM and NN classifiers. To evaluate the proposed method, the authors conduct experiments comparing its results against well-established feature-extraction techniques, such as LBP, Gabor and Hierarchical Multiscale LBP. This evaluation is performed using multiple datasets, including XM2VTS orL, AR, Extended Yale B and LFW. The test results demonstrate that PCANet exhibits superior resilience to variations caused by occlusion, illumination, pose, noise and expression. Consequently, this method holds a significant promise for enhancing face recognition applications.

In [19], Zhou and Feng presented a novel decision-tree ensemble technique known as gcForest (multi-Grained Cascade Forest). This method constructs a deep-forest ensemble with a cascade structure, enabling effective representation learning. Through adaptive determination of the cascade levels, gcForest can automatically adjust the model complexity, resulting in exceptional results even with limited data. Notably, gcForest exhibits a substantial reduction in the number of hyper-parameters compared to deep neural networks. The experimental findings from their work illustrate that gcForest achieves highly competitive performance on par with deep neural networks.

3. METHODOLOGY AND PROPOSED APPROACHES

Combination methods are techniques used to merge the outputs of multiple models, classifiers or information sources, with the aim of improving overall performance, robustness or providing more reliable predictions. These methods find applications in various fields, such as machine learning, pattern recognition and data fusion. In order to enhance the performance, robustness and reliability of facial-recognition systems, the present study implements two combination methods based on score-level and feature-level combination. These methods are employed to determine which approach significantly enhances the overall system performance.

The contributions of this paper are summarized as follows:

- In our initial proposition, in contrast to traditional combination techniques, such as score-level, feature-level and image-level techniques, we introduce an inventive fusion rule based on MLP classifier. Operating at the score-level, this methodology entails concatenating scores derived from individual models and then training the MLP classifier to compute the fitting score.
- Facial recognition does not uniformly rely only on the complete facial structure; instead, it can be reliant on specific facial components under certain conditions. From this perspective, we explore a novel S-CNN model predicated on facial regions. The fundamental concept of our proposition is based on linking the recognition of a given facial region with the recognition of the preceding facial region. This principle draws inspiration from sequential data recognition paradigms, such as text generation. In other words, the fully connected layer of the CNN that achieves optimal recognition for the initial facial region is combined with the features of the subsequent facial region and this sequence continues. Ultimately, the fully connected layers of the composite CNNs are merged and inputted into a DNN classifier to evaluate the overall system performance.

3.1 Machine Learning-based Face-recognition Approach

The facial-recognition approach proposed based on machine learning, as depicted in Figure 1, involves the following crucial steps: 1) Image preprocessing: the initial step encompasses face detection and image cropping to isolate the faces within the input images. 2) Feature extraction: in the second step, a variety of features are extracted from the preprocessed facial images. Three distinct schemes; namely, LBP, CBFD and HOG, are employed to extract different sets of features. 3) Combination model using

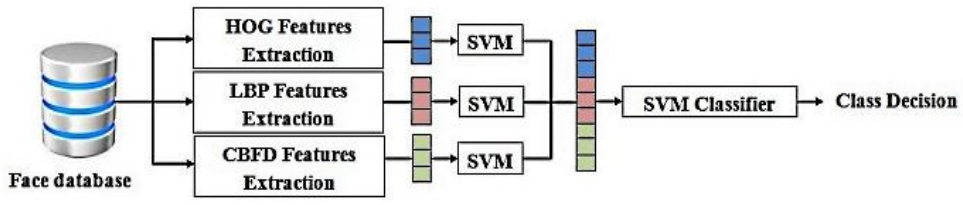

Figure 1. Fusion-based classifier combination (FCC) framework.

multiple classifiers: this step employs a combination model that integrates the outputs of three classifiers. The scores obtained from the three classifiers are combined and provided as input to the MLP classifier to determine the optimal combination for facial recognition.

3.1.1 Pre-processing Step

The preprocessing step plays a vital role in extracting valuable information from digital facial images, leading to a significant improvement in the accuracy of our facial-recognition schemes. Within our proposed schemes, this preprocessing step encompasses two primary stages: face-detection stage and cropping and resizing stage.

1) **Face detection:** In face-related applications, face detection plays a fundamental role. It involves the utilization of algorithms designed to detect and precisely locate essential points on a face, known as landmarks. The primary objective of this step is to accurately identify the facial region in order to extract features exclusively from the pertinent areas of the input image. For the purpose of facial region detection in each image, we employ a landmark-detection algorithm. Specifically, we utilize a 68-landmark shape detector that automatically identifies the facial landmark points [24]. In the first step, we focus on selecting two specific points: the top of the eyebrows and the cheeks. These points allow us to precisely localize the facial region, which is utilized in the proposed classifier combination scheme.
2) **Cropping and Resizing:** The 68-landmarks algorithm is applied to capture the distinctive characteristics associated with the facial-recognition process, such as the corners of the eyes, mouth and nose, as well as the cheeks, chin, top of the nose and forehead. These skin areas are known to be highly correlated with aging. In our study, we utilize two specific landmarks located at the top of the eyebrows and the mouth as reference points for determining the width and height of a rectangle. To ensure accurate localization of the facial region, we prefer defining slightly larger rectangles that overlap with each other. This approach enables us to cover a wider area of the face. Once the facial region is defined, the input image is cropped to include only the portion covered by the landmark-defined rectangles.

3.1.2 Feature Extraction

Feature extraction is an essential component of pattern-recognition applications, as the quality of classification results relies heavily on the distinctiveness and variability of the extracted features used to differentiate between different patterns. In our proposed methods, we employ various techniques for feature extraction, including the LBP handcraft technique, the learned handcraft technique CBFD and HOG in each method. While the LBP and HOG descriptors are explained in detail in references [21]-[22], we will provide a brief overview of the CBFD feature-extraction technique in this section.

CBFD (Compact Binary Facial Descriptors)

Jiwen et al. [23] introduced a novel feature-extraction method called the Compact Binary Face Descriptor (CBFD), which aims to enhance the performance of binary codes through a learning phase. This approach incorporates intelligence to overcome limitations and improve effectiveness. The training phase and image-feature extraction can be summarized as follows:

1) ***Training Phase***: This method focuses on the robustness of binary codes in relation to local changes in image texture. The compact binary codes are learned directly from raw pixels to represent the images. It is important to note that for better classification results, CBFD features should be constructed using a set of image samples that are provided within the same context. During CBFD feature learning, the training vectors are generated by considering the relationship between each pixel and its surrounding neighborhood. Specifically, an analysis is performed on the image using a rectangular window of size $(2R+1)\times(2R+1)$, where R is a positive integer. This window is centered on each pixel, allowing for the extraction of relevant information from the local context surrounding that pixel.

Let's define $X = [x_1, x_2, ..., x_n]$ as the set of training vectors, referred to as Pixel Difference Vectors (PDVs). These PDVs are obtained by measuring the difference between the central pixel and its neighboring pixels within a predefined window. The size of each vector is $(2R+1)\times(2R+1)-1$,

excluding the PDV between the central pixel and itself (PVD_0=0).

The goal of the CBFD feature extraction is to learn K hash functions $(w_k)_{K=1...n}$ that quantize each vector $x_n, (n=1,...,N)$ into a binary vector $b_n = [b_{n1}, b_{n2}, ..., b_{nk}]$. This quantization is achieved through the following formula:

$$b_{nk} = 0.5 \times (sgn(w_k^T \times x_n) + 1) \quad (1)$$

Here, $sgn(v) = 1 if\ v \succ \tau\ and\ -1$ otherwise, where τ denotes the threshold used for binary conversion of features.

To build the projection matrix w, which comprises all the hash functions w_k, we initialize it with the K first eigenvectors of the covariance matrix $(C = XX^T)$. Then, an optimization task is performed to minimize the objective function, $J(w_k)$ defined as:

$$min\, J(w_k) = J_1(w_k) + \lambda_1 \times J_2(w_k) + \lambda_2 \times J_3(w_k) \quad (2)$$

The parameters λ_1 and λ_2 are predefined and used to balance the effects of different terms. The terms J_1, J_2 and J_3 are selected to ensure that: (1) The variance of the learned binary codes is maximized; (2) The quantization loss between the original feature and the encoded binary codes is minimized; (3) The feature bins in the learned binary codes are evenly distributed as much as possible.

Codebook Learning: The purpose of the codebook is to reduce the number of binary vectors associated with each image. The training vectors (PDVs) are projected onto the matrix w and then, the k-means clustering algorithm is applied to obtain the centroids of the resulting binary vectors. These centroids form the codebook, which represents the classes.

2) Image-feature Extraction Phase: The feature-extraction process relies on the projection matrix w (CBFD feature) and the codebook obtained during the training phase. After obtaining all the PDV vectors $X = [x_1, x_2, ..., x_n]$ for the image, their binary counterparts are determined by projecting them onto the matrix w:

$$V_b = 0.5 * (sgn(w^T * X) + 1) \quad (3)$$

Each binary vector is then replaced with the closest vector coordinate in the codebook (bin). Subsequently, a histogram is constructed using the different coordinates, representing the entire image feature. To extract discriminative feature vectors, the raw image is segmented into multiple regions, treating each region as an individual image with its own CBFD features (w) and codebook. For each region, a histogram (H_s) is created. Finally, concatenating all the histograms results in a comprehensive vector (v) that represents the entire image:

$$v = [H_1, H_2, ..., H_M] \quad (4)$$

Where, M represents the number of regions. In our experimental results, our primary objective is to determine the number of regions that yields the highest accuracy for the facial age-estimation system. We systematically vary the number of regions to evaluate its impact on the performance of the system, aiming to identify the optimal configuration that maximizes accuracy in estimating facial age.

3.1.3 Fusion-based Classifier Combination (FCC)

Combining the decisions of multiple classifiers is an effective approach for improving classification rates, particularly in challenging pattern recognition problems. Extensive research has shown that, in many applications, fusing the outputs of multiple simpler classifiers tends to yield better recognition rates compared to relying on a single, more complex classifier. This fusion of multiple classifiers leverages their individual strengths and can lead to enhanced performance in recognizing and categorizing patterns accurately [25].

This research introduces a novel model for combining classifiers: Fusion-based Classifier Combination (FCC). The FCC method assumes that all classifiers are trained using the entire feature space and are both competitive and complementary to each other. It combines the output scores of all classifiers to make a final decision, leveraging their collective knowledge and capabilities. The proposed combination model can be summarized as follows:

Jordanian Journal of Computers and Information Technology (JJCIT), Vol. 09, No. 04, December 2023.

- First, we have a training sample set consisting of pairs (x_i, y_j), where *i* ranges from 1 to *n*. Each sample x_i is described by d-dimensional features in a feature space $(x_i \in R^d)$, while y_i represents the corresponding category label of the sample, taking values from the set {1, n}. The number of dimensions in the feature space is denoted by *d*.
- Next, each basic classifier *j* receives a set of input data and makes predictions for each input, resulting in a score vector of size *n* representing the probabilities assigned to each class:

$$[S_j^1, S_j^2, ..., S_j^n]^T \in [0;1] \tag{5}$$

Considering the scores assigned to class *i* by base classifier *j* as S_j^i, a Multilayer Perceptron (MLP) network is employed. This MLP network consists of an input layer that takes in the obtained scores, a single hidden layer and utilizes the sigmoid activation function. The combination model makes a decision for class *i* based on the output layer of the MLP classifier, which is determined by the following formula:

$$C = \sum_{i}^{n} w_{ij} \times S_j^i + b_j \tag{6}$$

Here, w_{ij} represents the weights and b_j is the bias value of classifier *j*.

- Additionally, let $v = (v_1, v_1, ..., v_m)$ denote the actual output vector of the model, where the components $v_i \ (for\ i = 1,...,M)$ represent the combination classifier's final determination of the probability of the input samples belonging to class C_i. To update the weights and bias value, it is necessary to compute the prediction error of the model. This can be achieved by using Formula (7) to calculate the error of the j^{th} node in the output layer. The prediction error ε of the j^{th} node in the output layer is given by:

$$\varepsilon = v_j(1 - v_j) - (t_j - v_j) \tag{7}$$

Here, t_j represents the desired output value of the model.

3.2 S-CNN Deep Learning-based Face-recognition Approach

When a subject is asked to confirm the relationship between two face images, it is likely that his/her attention will be focused on specific facial features, such as the eyes, mouth and nose. We believe that these facial key-points are crucial for facial-recognition analysis. Furthermore, geometrically, there exists a high relationship between the different regions within face images. Let's consider the baseline formed by connecting the centers of the two eyes. Assuming the distance between the eye centers is represented by d, the vertical distances from the nose, eyebrows and mouth to this baseline offer valuable information for distinguishing between faces.

Our proposal introduces an innovative approach called "sequential facial region-based face recognition" aimed at improving the performance of facial-recognition systems. This novel approach treats facial images as a sequence of data, drawing inspiration from the progress made in tasks involving sequences, such as text and video recognition. Our methodology involves the use of multiple individual Convolutional Neural Networks (CNNs), as visually depicted in Figure 2. Each of these CNNs is purposefully designed to handle input data from a specific facial region, facilitating a thorough analysis of various facial components to enhance recognition accuracy.

Consider the representation of a facial image as a sequence of facial regions, denoted as *x*. At each discrete time step *t*, we identify a specific facial region, denoted as $x^{(t)}$, which serves as the input to the corresponding basic CNN. For each of these time steps, we compute a hidden state, $FC^{(t)}$, which plays a crucial role as the network's "memory." This hidden state is determined by combining information from the current input $x^{(t)}$ and the hidden state from the previous time step $FC^{(t-1)}$. Mathematically, this combination is achieved through concatenation and it can be expressed as shown in Equation (8).

$$CNN^{(t)} = x^{(t)} \oplus FC^{(t-1)} \tag{8}$$

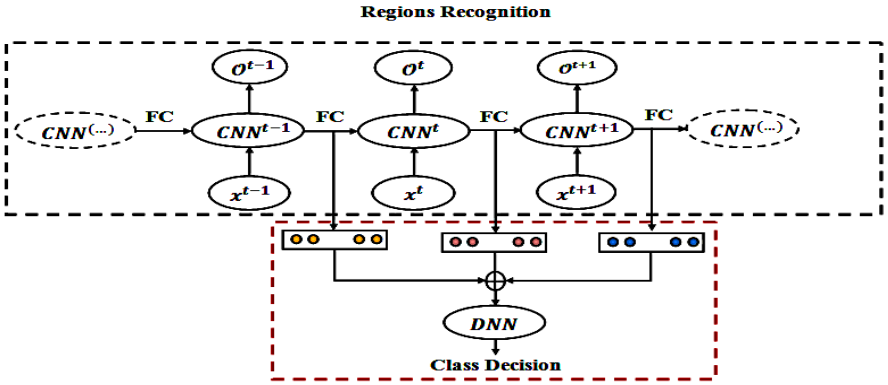

Figure 2. S-CNN deep learning-based face-recognition framework.

The result of this computation, represented as $O^{(t)}$, encapsulates the information and features that the CNN has extracted from the specific facial region $x^{(t)}$.

Moving to the second stage of our proposed approach, we consolidate the most effective fully connected layers from each of the individual CNNs. These layers have proven to be adept at identifying and processing the input facial regions optimally. This combination of fully connected layers generates a feature vector, a comprehensive representation of information from the concatenated feature vectors. This feature vector is subsequently employed as input for a Deep Neural Network (DNN) classifier. Leveraging the richness of information contained in this combined feature vector, the DNN classifier ensures efficient and effective face recognition.

4. Experiments and Discussion

A series of experiments were conducted to validate the effectiveness of the proposed methods. Each experiment focused on evaluating and comparing the performance of these methods using three well-known datasets: LFW [26] orL [27] and IJB-C [28]. These datasets were chosen, as they provide a diverse range of face images and serve as common benchmarks in the field of facial recognition. The experiments aimed to assess the accuracy and robustness of the proposed methods on these datasets, providing empirical evidence of their effectiveness in real-world scenarios.

4.1 Datasets' Description

To assess the effectiveness of the proposed methods, we conducted evaluations using three facial datasets: LFW, ORL and IJB-C.

4.1.1 LFW Dataset

The LFW dataset contains 5,749 unique individuals. Among these individuals, 1,680 have multiple images stored in the dataset, while the remaining 4,069 have only a single image. These images are saved as JPEG files and have dimensions of 250 by 250 pixels. The majority of the images are in color, although a small portion of them are grayscale. To obtain these images, the 68-landmark shape detector [24] is utilized, which accurately identifies the location of 68 facial landmarks. Subsequently, the detected faces undergo a process of resizing and cropping to achieve a consistent and fixed size.

4.1.2 ORL Dataset

On the other hand, the ORL dataset is a well-established dataset extensively employed in face-recognition research. It comprises a set of grayscale face images obtained from 40 distinct individuals. Each individual contributes ten images to the dataset. The images in the ORL dataset have a resolution of 92 by 112 pixels and are stored in a standard JPEG format. The images are captured under controlled conditions, incorporating variations in facial expressions, lighting conditions and slight pose changes. The subjects in this dataset encompass diverse genders, ages and ethnicities, making it a suitable resource for assessing the performance of face-recognition algorithms across a broad range of individuals. The dataset is commonly used for tasks, such as face detection, face recognition and facial-expression analysis.

4.1.3 IJB-C Dataset

The IJB-C (IARPA Janus Benchmark-C) dataset is widely used in the field of face recognition. It is designed to evaluate and advance the performance of face-recognition algorithms under challenging real-world conditions. The dataset contains a total of 31,334 still images, with 21,294 images featuring human faces and 10,040 images containing non-face content. On average, there are approximately 6 images available for each subject in the dataset. These images capture various facial expressions, poses and lighting conditions, making it a diverse and challenging dataset for face-recognition tasks. In addition to still images, IJB-C includes 117,542 frames extracted from 11,779 full-motion videos. Each video typically contains multiple frames of the same subjects, contributing to a more comprehensive evaluation of face-recognition algorithms. IJB-C is accompanied by a well-defined evaluation protocol that specifies how to split the dataset into training and testing sets, as well as the performance metrics used to assess face-recognition algorithms.

4.2 Protocol Description

In our experimental setup, we partitioned the face images from the ORL dataset into two distinct sets. The training samples for face-recognition systems consisted of 240 face images, comprising 6 images from each subject. The remaining 160 face images from the ORL dataset were reserved for testing purposes. Additionally, for the LFW dataset, we utilized 3300 face images (equivalent to 80% of the dataset) as the training samples for face recognition systems, while the remaining 769 face images from the LFW dataset were allocated for testing.

IJB-C introduces a comprehensive evaluation framework comprising eight distinct protocols for assessing the performance of face detection, verification, recognition and clustering across different scales and scenarios. In our study, we have specifically focused on the 1: N mixed recognition protocol, which assesses algorithms' capabilities in identification scenarios. Within this framework, there are two separate galleries; namely, Gallery 1 (referred to as G1) and Gallery 2 (referred to as G2). Each gallery contains one template per subject, which is generated by randomly selecting a half of the subject's still images. The remaining media instances are allocated to the probe set. G1 encompasses 1,772 subjects, accompanied by 5,588 still images, while G2 comprises 1,759 subjects and 6,011 still images. It's important to note that these galleries are entirely distinct from each other, facilitating open-set identification scenarios.

4.3 Evaluation Metrics

In order to assess the effectiveness of our newly proposed face-identification system, we performed thorough evaluations on a range of datasets, which encompassed ORL, LFW and IJB-C. These evaluations were carried out using the accuracy evaluation metric. Accuracy serves as a fundamental and extensively employed measure in classification systems, particularly in facial recognition. It determines the overall precision of the model by quantifying the ratio of correct predictions to the total predictions made. Mathematically, accuracy is defined as follows:

$$Accuracy = \frac{Number\ of\ Correct\ Predictions}{Total\ Number\ of\ Predictions} \times 100\% \tag{9}$$

In the context of facial-recognition systems, a prediction is considered accurate if the system successfully identifies or verifies the individual in the image. Conversely, an inaccurate prediction occurs when the system fails to recognize the individual or incorrectly identifies them as someone else. Accuracy serves as an easily comprehensible metric that offers a broad assessment of the system's effectiveness.

4.4 Evaluation of Performance

The experimental results in this section are divided into two main parts. The first part examines the outcomes of the FCC approach that was proposed. The second part presents the experimental results of the S-CNN approach. Following these analyses, a comparison is conducted among the leading systems to determine the most effective one.

4.4.1 Parameters' Setting

A parameter-setting phase is conducted for the feature-extraction algorithms used in the proposed FCC approach; namely, LBP, CBFD and HOG, before evaluating their performance.

1) LBP Parameters' Setting

In order to enhance the performance of the LBP feature descriptors, two key parameters are considered: the radius r of the pattern surrounding the central pixel and the number of points along the outer radius p [21]. To determine the optimal values for these parameters, a series of experiments were conducted. The LBP algorithm was tested with different combinations of samples and radius values, such as (12, 2), (12, 4), (12, 6), (12, 8), (16, 2), (16, 4), (16, 6) and (16, 8). The objective behind this parameter variation was to achieve improved results and enhance precision in the LBP algorithm when used in conjunction with an SVM (Support Vector Machine) classifier. Table 1 demonstrates the performance of the proposed system, with variations in both the number of samples and the radius *(p,r)*, while employing the SVM classifier.

In the context of the ORL dataset, the table reveals that the LBP descriptor achieves the highest level of accuracy for face recognition. More precisely, when employing **16** samples and a radius of **4**, the LBP descriptor achieves an impressive accuracy precision of **85.1%**. This underscores the exceptional performance of the LBP descriptor when it comes to recognizing faces in the ORL dataset. Turning our attention to the LFW dataset, the table indicates a respectable face-recognition accuracy of **76.8%**, even if lower than what was achieved in the ORL dataset. Nevertheless, the LBP descriptor remains effective in recognizing faces within the LFW Dataset, albeit with a slightly reduced level of accuracy compared to its performance in the ORL dataset. In summary, within the IJB-C dataset, we observe a slight reduction in face-recognition accuracy, specifically reaching **73.4%**, when using a configuration of **12** samples and a radius of **2**. This decline can be attributed to the presence of lower-quality images within the validation protocol of IJB-C.

Table 1. Improving accuracy of SVM using LBP-based features.

Datasets	LBP Parameters (*p, r*)	Accuracy [%]
ORL	(16, 4)	**85.1**
LFW	(12, 8)	**76.8**
IJB-C	(12, 2)	**73.4**

2) HOG Parameters' Setting

Similarly, a parameter-setting phase is conducted for the HOG algorithm. This phase focuses on two important parameters: the size of the blocks and the percentage of overlap between adjacent blocks. Specifically, we are interested in determining the optimal block size, while studies suggest that a 50% overlap between blocks is sufficient for effective algorithm performance [22]. To obtain better parameters, we conducted experiments by testing the HOG algorithm with blocks of different sizes. For example, we examined block sizes ranging from 10×10, 12×12 and so on, up to 32×32, while maintaining the same percentage of overlap. Varying the block size enables us to assess and identify the optimal configuration that yields improved results in terms of precision. Table 2 presents the best results obtained with the ORL, LFW and IJB-C datasets using different block sizes. This table highlights the impact of varying block sizes on the performance of the HOG algorithm and provides insights into the effectiveness of different configurations for face recognition.

From the results presented in Table 2, it is evident that face recognition achieves higher levels of accuracy in the context of the ORL dataset. To provide more detail, when utilizing a block size of **16×16**, the SVM algorithm, driven by the HOG technique, attains an impressive accuracy precision of **87.0%**. This result underscores the HOG algorithm's effectiveness in accurately identifying faces within the ORL dataset. Similarly, in the case of the LFW dataset, the recognition rates obtained exhibit competitive performance when juxtaposed with those observed in the ORL dataset. An accuracy rate of **79.8%** is achieved with a block size of **20×20**, indicating the robust performance of the HOG feature-extraction method in recognizing relationships within the LFW dataset. These outcomes closely mirror the results obtained in the ORL dataset. However, when turning our attention to the IJB-C dataset, we note a slight decrease in face-recognition accuracy, particularly at **76.2%**, when employing a block size of **12×12**. This decrement can be attributed to the dataset's inclusive nature, encompassing various subject categories and factors, such as facial hair, skin color and substantial pose variations.

Table 2. Improving accuracy of SVM using **HOG**-based features.

Datasets	HOG Parameters (*w*×*w*)	Accuracy [%]
ORL	16×16	**87.0**
LFW	20×20	**79.8**
IJB-C	12×12	**76.2**

3) CBFD Parameters' Setting

The CBFD feature-learning technique employs a predefined set of parameters tailored to our specific requirements. However, certain parameters require testing and fine-tuning to optimize the performance of our age-estimation system using this feature-extraction method. Through a series of experiments, we aimed to identify the most suitable parameters for CBFD in order to enhance our system's performance.

The CBFD algorithm relies on key parameters, including the window size, binary threshold, quantization method and projection matrix size. In our experimentation, we specifically focused on investigating the impact of the window size parameter. Our goal was to determine the optimal window size that would result in enhanced performance based on the metrics that we considered. To accomplish this, we conducted experiments using various combinations of region sizes including (3, 3), (5, 5), (7, 7), (9, 9), (11, 11), (13, 13), (15, 15) and (17, 17). By varying this parameter, we aimed to find the window size that yielded the best performance based on the metrics that we considered. Additionally, for the quantization parameter, we utilized the adaptive-quantization method with a defined threshold of **0.9**. We chose this specific approach to discretize the continuous-valued features in the CBFD algorithm. Moreover, for the feature normalization parameter, we employed Z-score normalization, which helps standardize the input data. This normalization technique ensures that the features are invariant to variations in image appearance and illumination. Furthermore, we utilized L1 regularization with a lambda value of **0.01** as a parameter to prevent overfitting and promote generalization in the CBFD algorithm.

Table 3 shows the results obtained from facial-recognition experiments conducted using the CBFD technique, wherein different window size parameters were employed. The findings demonstrate that when a window size of 7×7 is used, an impressive accuracy of **88.9%** is attained when applied to the ORL dataset. Likewise, with the LFW dataset, employing a window size of 13×13 yields a recognition rate of approximately **81.6%**. However, upon examining the IJB-C dataset, a minor decline in face-recognition accuracy is observed, specifically registering **79.54%** when a window size of 17×17 is utilized.

Table 3. Improving accuracy of SVM using **CBFD**-based features.

Datasets	Window size (*n* ,*n*)	Accuracy [%]
ORL	7×7	**88,9**
LFW	13×13	**81,6**
IJB-C	17×17	**79.54**

4.4.2 Fusion-based Classifier Combination (FCC) Performance

In this step, the research study utilizes the optimal parameters obtained from each feature-extraction technique to generate recognition scores. These scores are then concatenated and utilized as inputs for the multilayer perceptron (MLP) classifier, which serves as the combination model. Initially, support vector machine (SVM) classifiers are trained using two different kernel methods for both the LFW and ORL datasets. The selected parameters for SVM training are n = 3 for the polynomial kernel and σ = 0.125 for the RBF kernel. A value of C = 0.2 is employed during SVM training. Subsequently, the MLP classifier, acting as the combination model, is trained using the ReLU activation function in the hidden layer to introduce nonlinearity and the Softmax activation function is utilized in the output layer.

Table 4 provides a comprehensive overview of the performance of the MLP fusion technique when applied to three distinct datasets: LFW, ORL and IJB-C, utilizing different kernel methods. Notably, the RBF kernel emerges as the top performer across all three datasets, achieving the highest levels of accuracy. Specifically, when employing the RBF kernel, it achieves remarkable accuracy rates of

98.48% for the ORL dataset, **82.43%** for the LFW dataset and **81.84%** for the IJB-C dataset. These results underscore the robust accuracy levels that each kernel method can achieve when tailored to the specific characteristics of the respective datasets.

Table 4. Analyzing the statistical properties of SVM fusion with kernel methods.

Faces datasets	Kernel method	Accuracy
LFW	Polynomial kernel	78.67%
	RBF kernel	**82.43%**
ORL	Polynomial kernel	93.72%
	RBF kernel	**98.48%**
IJB-C	Polynomial kernel	75.45%
	RBF kernel	**81.84%**

The RBF kernel, in particular, stands out as the preeminent choice, demonstrating the highest recognition accuracy among the two kernel methods examined. Its superior performance makes it a widely preferred approach in fusion problems. Furthermore, it offers the practical advantage of requiring fewer parameters and encountering fewer numerical challenges compared to the polynomial kernel, enhancing its appeal in real-world applications.

4.4.3 S-CNN Deep learning-based Face-recognition Performance

The proposed S-CNN architecture for facial recognition incorporates a total of seven CNNs dedicated to recognizing specific facial regions (eyes, nose, mouth, top-left corner, top-right corner, bottom-left corner and bottom-right corner), as depicted in Figure 3.

Figure 3. The seven face regions used in the proposed S-CNN: the cropped image and its local regions, including the mouth, nose, eyes, top-left Corner, top-right corner, bottom-left corner and bottom-right corner.

In order to achieve superior accuracy in facial recognition, a series of experiments were conducted for each model, focusing on each individual facial region. These experiments aimed to identify the optimal parameters for the CNNs. The parameters considered include the choice of filter sizes (3×3, 5×5, 7×7 and 11×11) and the number of filters (8, 16, 32, 64 and 128) in the initial layer. This parameter-selection process was replicated for the subsequent layers within the architecture. Due to the comprehensive nature of the results obtained (resulting in 7×4×4=112 possibilities), Table 5 provides a concise summary of the highest accuracies achieved using the LFW, ORL and IJB-C datasets, with the selection of filter sizes and numbers serving as the key determining factors.

Table 5. Effectiveness of CNNs for facial-regions recognition.

	Datasets	LFW	ORL	IJB-C
Facial Regions	**Eyes**	73.74 %	94.23 %	**82.85%**
	Nose	71.80 %	92.33%	79.54%
	Mouth	72.10%	96.37%	77.60%
	Top-left corner	74.10 %	95.28%	75.76%
	Top-right corner	83.90%	**97.95%**	74.88%
	Bottom-left corner	84.20 %	96.10%	77.11%
	Bottom-right corner	**85.60 %**	97.49%	79.19%

The results presented in Table 5 clearly demonstrate the effectiveness of our proposed facial-recognition method in identifying facial regions across different datasets. In the case of the LFW

Jordanian Journal of Computers and Information Technology (JJCIT), Vol. 09, No. 04, December 2023.

dataset, we observe satisfactory performance, with the bottom-right corner particularly noteworthy, achieving an impressive accuracy **of 85.60%**. When applied to the ORL dataset, our method excels even further, achieving higher accuracy rates. Specifically, the top-right corner stands out with exceptional accuracy, reaching an impressive accuracy of **97.95%**. In the context of the IJB-C results, we witness significant improvements in performance compared to the FCC approach, with our method achieving an impressive accuracy of **82.85%**.

In order to achieve accurate face recognition and effectively handle variations, the fully connected layers of the basic CNN are concatenated to form the final feature vector. This feature vector is then utilized as input for the DNN classifier, enabling robust and precise face recognition.

In our experiments, we illustrate our process of determining the ideal number of neurons for the hidden layer as shown in Figure 4. We conducted a series of tests using MLP classifier with varying numbers of neurons in the hidden layer, ranging from 10 to 100. We maintained a consistent number of maximum iterations (2000 to 5000) and employed mean squared error (MSE) training. The transfer functions utilized were sigmoid functions.

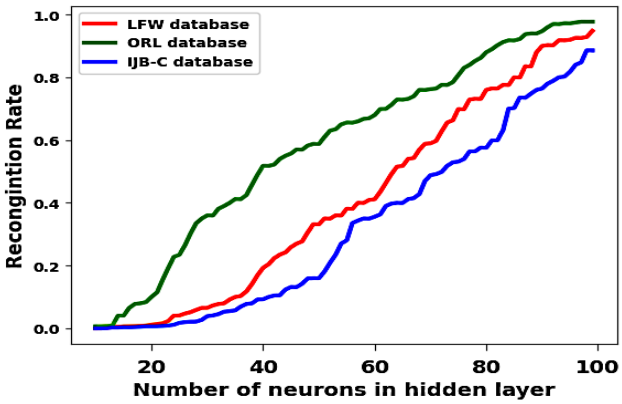

Figure 4. Recognition rates with the proposed S-CNN for LFW ORL and IJB-C dataset.

Figure 4 illustrates the performance evaluation conducted on the proposed facial-recognition system. The key observations and insights derived from this figure analysis can be outlined as follows:

- **High Recognition Rates with Over 90 Neurons:** One of the significant findings is the substantial improvement in recognition rates when employing more than 90 neurons in the hidden layer of the DNN classifier. In this configuration, the recognition rates consistently reached impressively high levels, ranging from **95.54%** to **97.75%**. This result underscores the effectiveness of the DNN architecture when coupled with CNNs for facial-recognition tasks.

- **Dataset-specific Variation:** The figure underscores the significance of dataset selection in influencing recognition performance. Notably, the recognition results for the ORL dataset outperformed those for the LFW and IJB-C datasets. Specifically, the ORL dataset achieved a recognition rate of **97.75%**, while the LFW and IJB-C datasets achieved recognition rates of **92.90%** and **88.59%**, respectively.

- **Challenges with 20-60 Neurons:** An intriguing observation pertains to the use of a relatively small number of neurons, specifically in the range of **20** to **60** neurons, within the hidden layer. During this range, the MLP algorithm, a component of the DNN, encountered convergence issues when applied to the LFW and IJB-C datasets. This issue can be attributed to the insufficient capacity of the hidden layer to effectively train the MLP classifier, highlighting the sensitivity of model architecture to dataset characteristics.

- **Dataset-specific Challenges:** The figure elucidates the specific challenges posed by the LFW and IJB-C datasets. The LFW dataset's difficulties are attributed to the considerable variations in facial orientation and expressions, which can complicate the recognition process. In contrast, the IJB-C dataset exhibits variations in both height and low image quality of the facial data, further complicating accurate recognition.

4.5 Evaluation of the Proposed S-CNN Model on SoTA Loss Functions

The central aim of face recognition, which includes both face verification and identification, is centered on the differentiation of facial features. However, the conventional Softmax loss function utilized in deep Convolutional Neural Networks (CNNs) often proves inadequate in terms of its discriminative capacity. To address this limitation, a variety of novel loss functions emerged in recent times, including Large Margin Cosine loss (CosFace) [29], Additive Angular Margin Loss (ArcFace) [30] and SphereFace Loss [31].

These advanced loss functions are designed to enhance the discriminative power of neural network feature embeddings by promoting a specific relationship between feature vectors and class centroids.

In this section, we assess the performance of the proposed S-CNN model by integrating, separately, two loss functions: CosFace and ArcFace into our proposed S-CNN model.

Implementation Details

The key steps to integrate each loss function (CosFace and ArcFace) into a basic CNN are:

1. In the last fully connected layer of each CNN in our model, we incorporated an additional layer designed for the computation of the integrated loss function. This layer accepts the feature vectors produced by the preceding layers as input and computes the specified loss function.

The employed loss functions can be expressed in the following manner:

- **CosFace Los Function**

$$L_A = -\frac{1}{N}\sum_{i=1}^{N} log \frac{e^{(cos(\theta_{yi})-m)}}{e^{(cos(\theta_{yi})-m)} + \sum_{j=1, j\neq y_i}^{N} e^{cos\,\theta_j}} \tag{10}$$

The key parameters used in the CosFace loss are: 1) the parameter *s* which controls the scaling of the cosine similarity scores. It determines how much we want to magnify or shrink the angular margin applied to the cosine similarity values and 2) the parameter *m* which specifies the angular margin added to the cosine similarity between the features and the weight vectors associated with the correct classes.

- **ArcFace Los Function**

$$L = -\frac{1}{N}\sum_{i=1}^{N} \log \frac{e^{\cos(\theta_{yi}+m)}}{e^{\cos(\theta_{yi}+m)} + \sum_{j=1, j\neq y_i}^{N} e^{\cos\,\theta_j}} \tag{11}$$

In this context, N denotes the batch size, y_i corresponds to the true label of the i^{th} example and *m* stands for the angular margin. The primary objective of the loss function is to optimize and increase the angular separation between the correct class and all other classes.

2. The output produced by the specified loss layer served as the ultimate output of our CNN model. This output was subsequently employed for both the training and evaluation or test stages of our experiments.
3. Finally, we evaluate our trained model on test datasets, taking into account the specified loss for feature embedding and classification.

In our experimental setup, we selected the margin values of *m = 0.35* for CosFace and *m = 0.50* for ArcFace. These choices were made based on their proven effectiveness in achieving strong performance, especially on low datasets, as demonstrated in previous research [30].

Table 6. Performance analysis of the proposed S-CNN method on CosFace and ArcFace loss functions.

S-CNN-based loss function	ORL	LFW	IJB-C
S-CNN based Softmax	**97.75%**	**92.20%**	**88.59%**
S-CNN based CosFace	**97.96%**	**93.35%**	**91.60%**
S-CNN based ArcFace	**98.95%**	**96.80%**	**92.25%**

As depicted in Table 6, when evaluating the performance of the proposed S-CNN method on the ORL, LFW and IJB-C datasets, the S-CNN model employing the ArcFace loss function achieved the highest level of accuracy, outperforming both the Softmax and CosFace variants. Specifically, it attained recognition rates of **98.95%** on the ORL dataset, **96.80%** on the LFW dataset and **92.25%** on the IJB-C dataset. The S-CNN approach using the CosFace loss also demonstrated strong performance, surpassing the Softmax variant on both datasets, with recognition rates of **97.96%** for ORL, **93.35%** for LFW and **91.60%** for IJB-C. In contrast, the S-CNN method employing the Softmax loss achieved the lowest recognition rates among the three methodologies, with rates of **97.75%** for ORL, **92.20%** for LFW and **88.59%** for IJB-C. Additionally, it is noteworthy that the ORL and LFW datasets consistently yielded higher recognition rates compared to the IJB-C dataset across all three loss functions, indicating variations in dataset characteristics and the effectiveness of the model.

5. COMPARISON STUDY

In this section, we conduct a comprehensive performance comparison of our proposed S-CNN method with recent state-of-the-art techniques [10]-[12], [16]-[18], including those based on various loss functions [9], [13]. Table 7 displays the accuracy results of these methods, covering both machine-learning and deep-learning approaches.

Regarding our machine learning-based method, our evaluation reveals that the fusion of classifiers achieves an impressive accuracy of **98.48%** on the ORL dataset, **82.43%** on the LFW dataset and **81.84%** on the IJB-C dataset. Notably, our proposed method outperforms the approach introduced by Muqeet et al. [11], which achieved an accuracy of 97.00%. Furthermore, when compared to deep learning-based methods, our approach surpasses Kong et al.'s [18] results using PCANet + KNN and PCANet + SVM, achieving accuracies of 91.50% and 97.50%, respectively.

Additionally, Table 7 provides a detailed comparison of the performance of our proposed deep learning-based method against recent deep-learning approaches. On the ORL dataset, our method attains an accuracy of **97.75%**; outperforming Kong et al.'s [18] results with accuracies of **91.50%** and **97.50%**. For the LFW dataset, our approach achieves an accuracy of approximately **92.20%**, surpassing the combination of FaceNet + RF [10] with an accuracy of **89.10%** and the combination of MLP + MFM with CNN [16] with an accuracy of 84.5%. Furthermore, our proposed method competes closely with Storey et al. [17] method, which achieved an accuracy of **93.60%**.

Table 7. Comparative analysis of the proposed methods with state-of-the-art (DL and ML techniques).

Machine learning

	Datasets		
Method	**High Quality**		**Mixed Quality**
	ORL	**LFW**	**IJB-C**
KNN (DWT+LBP) [11]	97.00%	-	-
Proposed method	**98.48%**	**82.43%**	**81.84%**

Deep learning

FaceNet + RF [10]	-	89.10%	-
LBP + Ensemble CNN [12]	100%		-
MLP + MFM in CNN [16]	-	84.50%	-
3D-CNN+ResNe [17]	-	93.60%	-
PCANet + KNN [18]	91.50%	-	-
PCANet + SVM[18]	-	97.50%	
ResNet-100 $_{CosFace}$ [9]	-	-	**92.20%**
ResNet-100 $_{ArcFace}$ [9]			**95.20%**
LMCL$_{CosFace}$ [13]	-	**92.69%**	-
LMCL$_{ArcFace}$ [13]		**93.30%**	
Comparison with state-of-the-art loss functions			
S-CNN based Softmax	**97.75%**	**92.20%**	**88.59%**
S-CNN based CosFace	**97.96%**	**93.35%**	**91.60%**
S-CNN based ArcFace	**98.95%**	**96.80%**	**92.25%**

Finally, we compare our proposed S-CNN model with state-of-the-art loss function methods [9], [13]. The comparison is presented in Table 7, where our model outperforms the approach introduced in [13], achieving **93.35%** and **96.80%** accuracy for the CosFace and ArcFace loss functions, respectively, on the LFW and IJB-C datasets. Compared to the method presented in [9], our proposed approach achieves competitive accuracies of **91.60%** for the CosFace loss function and **92.25%** for the ArcFace loss function, as opposed to **92.20%** for CosFace and **95.20%** for ArcFace obtained in [9].

6. CONCLUSION

This research paper addresses the challenges in facial recognition through the introduction of two innovative approaches: FCC and S-CNN. The effectiveness of three techniques; namely, LBP, HOG and CBFD, is evaluated in overcoming these challenges. The proposed solution involves the utilization of a novel multi-classifier combination model and a unique method for extracting high-level features from multiple image regions treated as sequential data using an ensemble of CNNs, followed by a DNN classifier for facial recognition.
The experimental results obtained from renowned facial datasets, including LFW, ORL and IJB-C, reveal the competitive performance of both the proposed multi-classifier combination model and the S-CNN deep-learning model when compared to state-of-the-art methods. Additionally, we have assessed the effectiveness of the proposed S-CNN model alongside state-of-the-art loss functions, such as CosFace and ArcFace. Based on the results that we have obtained from our experiments, we can illuminate specific strengths and weaknesses of our approach as follows:

- The experimental results show that FCC method based on combination at matching score level is likely to provide better recognition performance, as it contains more contented information which is both feasible and practical.
- This paper illustrates how to use CNN as a sequential model and we believe that it may open a door towards alternative to deep neural networks for many tasks. Traditional CNN will process an input and move onto the next one disregarding its sequence. In the proposed S-CNN, an image is considered as a series of sequential face regions that needs to be followed in order to understand. In other words, the first CNN receives a region of an image and passes it as a feature vector to the next CNN to predict the next face region based on the previous region and so on.
- Furthermore, we would like to mention that it is possible that the proposed S-CNN model could be used in other applications, such as age estimation, gender prediction or facial-emotion recognition.
- However, the proposed methods need to be accurate and robust enough to handle the variability and diversity of faces and datasets.
- In future research, we can explore the application of attention mechanisms to automatically identify distinguishing facial regions while effectively minimizing the impact of noisy areas.

REFERENCES

[1] I. Bendib, A. Meraoumia, M. Y. Haouam et al., "A New Cancelable Deep Biometric Feature Using Chaotic Maps," Pattern Recognition and Image Analysis, vol. 32, no. 1, pp. 109–128, 2022.

[2] M. Y. Haouam, A. Meraoumia, L. Laimeche and I. Bendib, "S-DCTNet: Security-oriented Biometric Feature Extraction Technique: An Effective Pathway to Secure and Reliable Biometric Systems," Multimedia Tools and Applications, vol. 80, pp. 36059–36091, 2021.

[3] R. Rameswari, K. S. Naveen, A. M. Abishek and C. Deepak, "Automated Access Control System Using Face Recognition," Materials Today: Proceedings, vol. 45, Part 2, pp. 1251-1256, 2021.

[4] N. Singhal, V.Ganganwar, M. Yadav, A. Chauhan, M. Jakhar and K. Sharma, "Comparative Study of Machine Learning and Deep Learning Algorithm for Face Recognition," Jordanian Journal of Computers and Information Technology (JJCIT), vol. 07, no. 03, pp. 313-325, September 2021.

[5] Z. Wanling and X .Shijun, "Face Anti-spoofing Detection Based on DWT-LBP-DCT Features," Signal Processing: Image Communication, vol. 89, p. 115990, DOI: 10.1016/j.image.2020.115990, 2020.

[6] S. Liangliang, W. Xia and S. Yongliang, "Research on 3D Face Recognition Method Based on LBP and SVM," Optik, vol. 220, p. 165157, DOI: 10.1016/j.ijleo.2020.165157, 2020.

[7] K. Weiyi., Y. Zhisheng and L. Xuebin, "3D Face Recognition Algorithm Based on Deep Laplacian Pyramid under the Normalization of Epidemic Control," Computer Communications, vol. 199, pp. 30-41, 2023.

[8] D. Mamieva, A.B. Abdusalomov, M. Mukhiddinov and T. K. Whangbo, "Improved Face Detection Method via Learning Small Faces on Hard Images Based on a Deep Learning Approach," Sensors, vol. 23, no. 1, p. 502, DOI: 10.3390/s23010502, 2023.

[9] G.-S. J. Hsu, H. -Y. Wu, C.-H. Tsai, S. Yanushkevich and M. L. Gavrilova, "Masked Face Recognition from Synthesis to Reality," IEEE Access, vol. 10, pp. 37938-37952, 2022.

[10] A. S. Sanchez-Moreno et al., "Efficient Face Recognition System for Operating in Unconstrained Environments," Journal of Imaging, vol. 7, no. 9, p. 161, 2021.

[11] L. Zhou, H. Wang, S. Lin. et al., "Face Recognition Based on Local Binary Pattern and Improved Pairwise-constrained Multiple Metric Learning," Multimedia Tools Application, vol. 79, pp. 675–691, DOI: 10.1007/s11042-019-08157-0, 2020.

[12] 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, p. 105622, DOI: 10.1016/j.cmpb.2020.105622, 2020.

[13] J. Deng, J. Guo, N. Xue and S. Zafeiriou, "ArcFace: Additive Angular Margin Loss for Deep Face Recognition," Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition, pp. 4690-4699, Long Beach, USA, 2019.

[14] Z. Mortezaie and H. Hassanpour, "A Survey on Age-invariant Face Recognition Methods," Jordanian Journal of Computers and Information Technology (JJCIT), vol. 05, no. 02, pp. 87 - 96, August 2019.

[15] M.A. Muqeet and R. S. Holambe, "Local Binary Patterns Based on Directional Wavelet Transform for Expression and Pose-invariant Face Recognition," Applied Computing and Informatics, vol. 15, no. 2, pp. 163-171, 2019.

[16] H. Yu, J. Zhao and Y. Zhu, "Research on Face Recognition Method Based on Deep Learning," Proc. of the 12th Int. Congress on Image and Signal Processing, Biomedical Engineering and Informatics (CISP-BMEI), Suzhou, China, 2019, pp. 1-5, 2019.

[17] G. Storey, R. Jiang, S. Keogh, A. Bouridane and C. Li, "DPalsyNet: A Facial Palsy Grading and Motion Recognition Framework Using Fully 3D Convolutional Neural Networks," IEEE Access, vol. 7, pp. 121655-121664, 2019.

[18] J. Kong, M. Chen, M. Jiang, J. Sun and J. Hou, "Face Recognition Based on CSGF (2D) 2PCANet," IEEE Access, vol. 6, pp. 45153-45165, 2018.

[19] Z. H. Zhou and J. Feng, "Deep Forest: Towards an Alternative to Deep Neural Networks," Proc. of the 26th Int. Joint Conf. on Artificial Intel. (IJCAI-17), pp. 3553-3559, DOI: 10.24963/ijcai.2017/497, 2017.

[20] M. Ramgopal et al., "Masked Facial Recognition in Security Systems Using Transfer Learning," SN Computer Science, vol. 4, Article no. 27, DOI: 10.1007/s42979-022-01400-w, 2023.

[21] D. Samai, A. Meraoumia, H. Bendjenna and L. Laimeche, "Oriented Local Binary Pattern (LBPθ): A New Scheme for an Efficient Feature Extraction Technique," Proc. of the IEEE Int. Conf. on Mathematics and Inf. Techn. (ICMIT), DOI: 10.1109/MATHIT.2017.8259710, Adrar, Algeria, 2017.

[22] B. Berkant, "Implementation of Hog Edge Detection Algorithm Onfpga's," Procedia - Social and Behavioral Sciences, Vol. 174, pp. 1567-1575, DOI: 10.1016/j.sbspro.2015.01.806, 2015.

[23] L. Jiwen et al., "Learning Compact Binary Face Descriptor for Face Recognition," IEEE Transactions on Pattern Analysis and Machine Intelligence (T-PAMI), vol. 37, no. 10, pp. 2041-2056, 2015.

[24] T. Devries, K. Biswaranjan and G. W. Taylor, "Age Estimation and Face Verification across Aging Using Landmarks," Proc. of the IEEE Canadian Conf. on Computer and Robot Vision, pp. 98–103, Montreal, Canada, 2014.

[25] S. Liu., X. Li, C. Hu et al., "Spammer Detection Using Multi-classifier Information Fusion Based on Evidential Reasoning Rule," Scientific Reports, vol. 12, Article no. 12458, DOI: 10.1038/s41598-022-16576-7, 2022.

[26] G. B Huang et al., "Labeled Faces in the Wild: A Dataset for Studying Face Recognition in Unconstrained Environments," Proc. of Workshop on Faces in 'Real-Life' Images: Detection, Alignment and Recognition, [Online], Available: http://vis-www.cs.umass.edu/lfw/, 2008.

[27] F. S. Samaria and A. C. Harter, "Parameterization of a Stochastic Model for Human Face Identification," Proc. of IEEE Workshop on Applications of Computer Vision, pp. 138-142, Sarasota, USA, 1994.

[28] B. Maze, J. Adams, J. A. Duncan et al., "IARPA Janus Benchmark-C: Face Dataset and Protocol," Proc. of the 2018 IEEE Int. Conf. on Biometrics (ICB), pp. 158-165, Gold Coast, Australia, 2018.

[29] H. Wang, Y. Wang, Z. Zhou, X. Ji, D. Gong, J. Zhou, Z. Li and W. Liu, "CosFace: Large Margin Cosine Loss for Deep Face Recognition," Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition, pp. 5265–5274, DOI 10.1109/CVPR.2018.00552, 2018.

[30] W. Liu, Y. Wen, Z. Yu, M. Li, B. Raj and L. Song, "SphereFace: Deep Hypersphere Embedding for Face Recognition," Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition, pp. 212–220, arXiv: 1704.08063, 2017.

[31] M. Kim, A. K. Jain and X. Liu, "AdaFace: Quality Adaptive Margin for Face Recognition," Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR2022), pp.18729-18738, [Online], Available: https://openaccess.thecvf.com/content/CVPR2022/papers/Kim_AdaFace_Quality_Adaptive_Margin_for_Face_Recognition_CVPR_2022_paper.pdf, 2022.

ملخص البحث:

تبحثُ هذه الورقة في مقارنةٍ شاملةٍ بين طريقتين من طُرق تشكيل تركيبات البيانات من أجل تحسين الموثوقية لأنظمة تمييز الوجوه. الطّريقةُ الأولى تسمّى تركيبة المصنِّفات القائمة على الاندماج (FCC)، ويتألّف النّموذج في هذه الطّريقة من (3) مصنِّفات يتمّ تدريب كلٍّ منها باستخدام إحدى تقنيات استخلاص السِّمات المعروفة. أما الطّريقة الثّانية فهي طريقة التّعلُّم العميق باستخدام الشّبكات العصبية الالتفافية المتعاقِبة (S-CNN)، وفيها يتمّ استخلاص عددٍ من السِّمات عالية المستوى من مناطق مختلفة في صُوَرٍ للوجوه وإدخالها على مجموعة متسلسلة من الشّبكات العصبية الالتفافية. بعد ذلك، تُجَمَّع مُخرجات تلك الشّبكات وتُدْخَل الى شبكة تعلُّم عميق عصبية (DNN) مخصَّصة لتمييز الوجوه.

وقد جرى اختبارُ النّموذجيْن المقترحَيْن في هذه الدّراسة على ثلاثٍ من مجموعات البيانات الخاصّة بصُوَر الوجوه. وقد تمّ الحصولُ على نتائج تؤكّد تنافسية الطّريقتين المستخدّمتين عند مقارنتهما بطرق تشكيل تركيبات البيانات الخاصّة بأنظمة تمييز الوجوه.

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