Advances in Artificial Intelligence and Machine Learning; Research 3 (3) 1407-1443 Received 23-06-2023; Accepted 02-09-2023; Published 09-09-2023

A Smart Face Recognition and Verification using Optimal Spatial and Spectral Feature Selection with Adaptive Multiscale Mobilenet

Santhosh Shivaprakash santhukit@gmail.com
Department of Electronics and Communication Engineering,
Kalpataru Institute of Technology,
Tiptur State, Karnataka- 572201, India
Visvesvaraya Technological University,
Belagavi, Karnataka State-590018, India

Sannangi Viswaradhya Rajashekararadhya svraradhya@gmail.com
Department of Electronics and Communication Engineering,
Kalpataru Institute of Technology,
Tiptur, Karnataka- 572201, India,
Visvesvaraya Technological University,
Belagavi, Karnataka State- 590018, India

Corresponding Author: Santhosh Shivaprakash

Abstract

Face recognition is complicated work, which is highly demanding while processing the inter-class similarities and intra-class differentiations in the acquired images in a wider range. However, the identification accuracy can be enhanced at some level through managing the system with non-matched templates. In the reality, face recognition is complicated owing to differentiations in the pose, background, illumination, and so on. On the other hand, the type of face recognition technique is recently dependent on machine learning-based facial features while avoiding the practical experiences in hand-craft features. Recent world, the face recognition technique using deep learning strategy is capable of learning efficient face features for getting a highly extraordinary efficiency. The face identification techniques use various conventional appearance methods, which are further applied for testing the efficiency via applying the benchmark facial images. The facial images gathered from the standard digital media are often noticed as problems regarding occlusion, lightning, and position conditions along with camera angle. The occluded images can be observed with the alignment, facial expressions, and human pose along with the camera axis. Thus, more attention must be taken care during the acquiring the facial images along with the coverage of the background. Hence, it is necessary for considering the precise pre-processing steps along with the necessary steps for recognizing the face. In this research, a new face recognition model is investigated utilizing deep learning techniques. Initially, the images are gathered from the standard resources. Next, they are pre-processed to increase the quality of images for further processes. Then, the spatial and spectral feature extraction is performed, where the spatial features are extracted by DeepLabV3, and the spectral features are extracted using Discrete Wavelet Transform (DWT) and the Discrete Cosine Transform (DCT). Next, the optimal spectral features and optimal spatial features are attained by using a new hybrid heuristic

Citation: Santhosh Shivaprakash and Sannangi Viswaradhya Rajashekararadhya. A Smart Face Recognition and Verification using Optimal Spatial and Spectral Feature Selection with Adaptive Multiscale Mobilenet Advances in Artificial Intelligence and Machine Learning. 2023;3(3):83.

algorithm known as Controlling Parameter-based Rock Hyraxes African Vultures Swarm Optimization (CP-RHAVSO) with the combination of Rock Hyraxes Swarm Optimization (RHSO) and African Vultures' Optimization Algorithm (AVOA) techniques. Then, these optimally selected features are combined and fed to the weighted fused feature extraction process. Finally, weighted fused features are given to the Adaptive Multiscale Mobilenet (AM-Mob) for face recognition, where the hyper parameters of M-Mob are tuned by the same CP-RHAVSO. Investigational results show that the proposed face recognition method on the standard database, according to experimental findings, yields a highly valuable accuracy rate.

Keywords: Face recognition and verification; DeepLabV3; Optimal spatial and spectral feature extraction; Weighted fused features extraction; Discrete wavelet transform; Discrete cosine transform; Controlling parameter-based rock hyraxes African vultures swarm optimization; Adaptive multiscale mobilenet;

1. INTRODUCTION

People utilize natural visualization systems for identifying human beings precisely. While designing the human visualization systems into face identification techniques and trained a computerized module for learning several human faces on the accessibility of huge-scale of digitally encoded faces [1]. Next, the computer module illustrates the furthest capability of recognizing and verifying other faces in several gathering platforms. Though, a human visual system cannot be visualized as a face identification system [2]. Thus, it varies a lot owing to varied acquisition platforms and differentiations among facial images of similar people. Even though face identification methodologies identify faces practically better, different problems still exist, and they are required to be handled for both uncontrolled and controlled facial images [3]. Some general problems presented in the acquired images are partially hidden faces, clutter, occlusion, head pose, the background of people, facial expressions, differentiations in illumination, and so on [4]. Moreover, a few of the additional issues are look-alike faces, plastic surgery, twin faces, makeup on the face, accessories, and ornaments [5]. These issues must be addressed via various unique algorithms with generalized face identification techniques. Among the unique techniques, they could not have the ability to solve the abovementioned issues [6]. Therefore, a reliable face identification technique is required to process the huge-scale training samples with considerable differences among faces for getting precise illustrations of lost facial properties in the facial images, which do not have the capacity of processing the larger samples for recognizing and creating more illustrations with precise face verification [7].

There are three solutions to the face recognition task comprising face matching, face identification, and face verification [8]. Face matching generally estimates whether two face instances are from the same people or not whereas face identification presents the process of identifying the unknown face by estimating it with entire facial images in the training dataset [9]. In addition, the face verification task illustrates the verification of declared identity by estimating a query face with a specific face in the dataset. This research work specifically formulates face recognition with a verification strategy for estimating efficiency [10]. Processing of several input images with differentiations is more helpful in reaching higher accuracy in a face identification system [11]. An effective face feature illustration is required for designing the facial recognition system and also needs to compute the distance among similar subjects [12].

The facial feature illustration approaches are generally categorized into two classes including deep feature learning techniques and hand-crafted features using manual designs [13]. While comparing with the recent historical approaches, the hand-crafted feature development methods are more classical ones as they are studied for a long duration whereas the recently emerging approaches are feature learning with the adoption of deep architectures. Some of the hand-crafted feature techniques are Scale-Invariant Feature Transform (SIFT), Gabor, Histogram of Oriented Gradients (HOG), Multi-Block Local Binary Pattern (MBLBP), and Local Binary Pattern (LBP), etc that are often used for recognizing the facial images [14]. These facial feature illustrators are known as local descriptors, which preserve the scale invariance or illumination robustness that is applicable for describing the facial images [15]. Practically, several features are fused together for enhancing the facial features for representing the facial images. The techniques adopted in the recent traditional methods have been designed by taking a single task or features when using the novel approaches integrating various features inclusive of edge, shape, color, and motion in higher levels [16]. Next, deep learning strategies are highly preferable for recognizing facial images. A few classifiers are deep neural networks (DNN), Support vector machines (SVM), Adaboost, Convolutional Neural Networks (CNN), deep CNN (DCNN), etc [17]. Finally, this research designs a novel facial image recognition approach with the exploration of facial images integrating deep learning strategies.

The research inventions implemented in this study are shown here.

- To explore a new innovative face recognition and verification framework using optimal spatial and spectral feature selection strategy along with enhanced transfer learning via heuristic mechanisms for aiding the real-time applications inclusive of healthcare, retail, education, immigration, access control, automobile security, safety purposes in the prevention of crimes, minimizing human interaction, etc to promote secured and automated identification.
- To offer the weighted fused feature extraction by recommending the optimal spatial and spectral feature extraction with the incorporation of acquired features from DeepLabv3, and DWT and DCT features along with the use of newly suggested heuristic strategy for gathering the most essential features.
- To recommend the adaptive transfer learning strategy known AM-Mob for performing face recognition by processing the weighted fused features for getting the recognized faces. Here, the heuristic improvement is incorporated to suggest the AM-Mob by tuning the parameters of multi-scale Mobilenet.
- To present the CP-RHAVSO strategy for acquiring the optimal spatial and spectral features, and optimizing the weight used for getting the weighted feature selection, along with the implementation of AM-Mob by tuning the momentum in multi-scale Mobilenet for promoting higher efficiency in face recognition with the promotion of superior convergence rate.

The forthcoming sections of this research paper are organized as follows. Section 2 presents the literature study. Section 3 offers the face recognition and verification model using advanced deep learning algorithm. Section 4 implements the development of CP-RHAVSO-based weighted fused feature selection for intelligent face recognition and verification. Section 5 offers the architecture of optimized AM-Mob for face recognition and verification. Section 6 evaluates the results, whereas Section 7 presents the conclusion

2. LITERATURE STUDY

2.1 Existing Works

In 2019, Garain *et al.* [18] have designed a new K-medoids Cohort Selection (KMCS) for choosing a reference group of non-matched templates regarding the considerable subjects. Generally, KMCS was used for clustering the entire scores of cohort subjects. Then, a cohort subset was chosen from their medoid of a cluster comprising of more scattered scores or members. The entire discriminative features were carried over with the cohorts for constituting the cluster while distinguishing it from others. Next, the facial features were attained, and then, the normalization of cohort score was done regarding the Max rule, Max-Min, and T-norm rules by estimating the matching scores among the query and probe images. It was carried out before performing the final choice of rejection or acceptance. The final outcomes were attained from the experimental results over other standard databases regarding other selection approaches.

In 2018, Li *et al.* [19] have offered a Deep Learning-aided Feature extraction for getting the Facial Textural Features (FTFA-DLF). This recommended FTFA-DLF has the ability to incorporate the features of hand-craft and deep learning features. Here, the texture features via hand-crafted feature techniques were acquired from the mouth, nose, and eyes areas. The features from deep learning with the hand-crafted features were included in the layer of functional values that have been modified adaptively for enhancing the efficiency of face recognition. The experimental outcomes were verified on the LFW face database for improving the accuracy rate.

In 2019, Zhuang *et al.* [20] have presented a new face recognition model by processing facial images with the framework of an automated estimation strategy. DCNN was used for training the face recognition module and estimated over several quality constraints like pose, occlusion, blurriness, contrast, brightness, etc. This model has sorted the input facial images by considering the trained face quality measure, and the recognition was done for choosing the good facial images. They have experimented with the KinectFace face and Color FERET databases. The efficient outcomes have recommended using the facial image quality measure and have differentiated bad images from good images, which were highly related to the final identification efficiency.

In 2022, Vasanthi *et al.* [21] have offered an approach for recognizing facial images based on biometric images along with the analysis of multivariate correlation that has acquired the lower-level visual and geometrical features. Here, the Active Shape Model (ASM) was used for gathering the geometric features whereas the color and texture of low-level features were collected from the chosen significant regions of the facial images. Then, the texture features were attained by autocorrelation scheme, and the YCbCrcolour strategy was utilized for collecting the color features. Finally, all these features were intended to build a matrix of feature tensors. The feature matrices of the target facial images were distinguished from the feature matrix of the significant facial images, in which the Canonical Correlation technique was used for saving the compared matrices in the database of the feature vectors. They were again tested for demonstrating whether it was highly important or not. Four standard datasets were used for the experimentation and also tested with various standard measures while dissimilating with classical techniques.

In 2019, Senthilkumar and Gnanamurthy [22] have offered a new technique of Wavelet Based face image Decomposition (WBFD). They have initially decomposed the acquired face images via two levels of the wavelet transform approach. Then, the attained images were again extinguished. The result analysis was estimated on the benchmark dataset and analyzed the performance of the designed model. Next, the statistical appearance features were acquired from the gathered images for reaching a higher level of accuracy rate. Moreover, this model was executed in a lower recognition time.

In 2022, Junaid *et al.* [23] have especially focused on recognizing the disguise invariant facial images along with deep learning strategies. Initially, the data augmentation was done on the acquired noisy images. Then, the Viola-Jones face detector was used for detecting the faces in the images, and then, pre-trained CNN was used for the classification with the fine-tuning of the disguise invariant facial images. The universal disguise-invariant features were learned from acquired face images of various subjects for properly identifying the faces under differing facial disguises with the assistance of pre-trained CNN. Various learning constraints were used for reducing the execution time and classification accuracy to select an appropriate model for recognition. The efficient balancing among efficiency and accuracy was carried out using Resnet-18, which has outperformed the conventional methods.

In 2018, Waisy *et al.* [24] have presented a new strategy for recognizing the face with the combination of Deep Belief Network (DBN) and hand-craft features for solving the issues of face identification in unconstrained circumstances. Fractal dimension and Curvelet transform were incorporated and known as Curvelet–Fractal method to get the features from the acquired images. The multidirectional and anisotropic transform were effectively applied for representing the major structure of the faces including curves and edges along with the usage of Curvelet transform whereas the texture features were gathered via the Fractal dimension. Moreover, face recognition was done in a multimodal and deeper way for adding the feature illustrations by a trained DBN. At last, the efficiency was analyzed over standard datasets over eminent techniques to showcase the superiority.

In 2019, Ameur *et al.* [25] have introduced a new descriptor known as Deep Hybrid Gabor Local Binarized Statistical Image Feature called (DeepGLBSIF) for gathering over-complete and effective acquisition of features in multi-level order. The CNN technique was used for recognizing the face from the training datasets. They have improved the utilization of interaction among the local and global features, which has helped in discriminating the illustrations of features and efficient performance. DeepGLBSIF framework was simply and efficiently built and understood the verification on faces. They have utilized Support Vector Machine (SVM) and Cosine distance metric for efficient classification. It was tested on standard facial datasets over traditional approaches.

2.2 Problem Specification

Two necessary factors for determining the efficiency of various face recognition techniques suffer from recognition time and recognition accuracy. Similarly, they do not consider the background and occlusion images that exist in the gathered facial images. Moreover, those approaches gather the features without considering these two drawbacks, which leads to bad recognition accuracy. Some of the existing face recognition approaches are reviewed in TABLE 1. KMCS [18] achieves a higher convergence time and offers faster computation, and has less expensive. It is more sensitive to noise. FTFA-DLF [19] achieves better face recognition accuracy and performance with the hand-craft features and increases the efficiency regarding superior light illumination robustness. It reduces computational complexity. DCNN [20] efficiently distinguishes high-quality face images from poor-quality ones and has finally ensured recognition performance. It is not applicable to real-time large-scale datasets. ASM [21] minimizes the computational time and is more suitable for the biometric-based authentication system. It suffers from exploring large-scale datasets. WBFD [22] attains better recognition performance, lower memory, and less recognition time. It does not include details of the lips, nose, and eyes, after two levels of wavelet transform in Yale faces. CNN [23] reduces the average execution time and increases the average recognition performance and higher generalization ability. It does not solve illumination conditions in dark images. DBN [24] minimizes the running time and gives superior performance on constraints like high diversity in terms of noise, lighting conditions, facial expressions, and so on. It faces complications in processing various facial differentiations. SVM [25] is a less expensive approach and attains the lowest computational complexity. It is not suitable for video-based face recognition. Consequently, there is a requirement on offering a new face recognition approach with a deep learning strategy.

3. FACE RECOGNITION AND VERIFICATION MODEL USING ADVANCED DEEP LEARNING ALGORITHM

3.1 Face Recognition and Verification Framework

Generally, biometrics applications use face recognition or verification frameworks. The face is the major justifiable biometrics and it has been the most general approach of identification, where the visual interactions along with the emotions in facial images play a major role in estimating the recognition of faces. Authentication models generally use the iris, voice, and fingerprint patterns for biometric verification. However, these data acquisition methods require following adequate constraints of precise positioning of the speaker or microphone for gathering the voice signals, the finger must be placed in accurate orientation and position. On the other hand, in collecting facial images, the non-intrusive method must be followed so, it can be utilized as a biometric scenario for surreptitious activities. Moreover, the face is a unique feature of people. Thus, face identification is more significant owing to the ability of their solutions to address object recognition and classification tasks, and also the ability of their possible applications in research domains. The faces that exist in the images are identified in face recognition systems in an automatic way, which is categorized into verification of face (authentication) and face recognition (identification). In the authentication or verification of face process, there is a one-to-one matching for comparing query facial images as opposed to entire facial images in the dataset for determining the identity of the query images. Nowadays, face recognition has been progressed to a great extent, yet it suffers from several complications regarding accessories, occlusion, pose, expression, illumination, and viewpoint along with a few unconstrained tasks, which may differ accordingly. However, facial components inclusive of facial outline, mouth, nose, and eyes are located by taking the position points, and thus, there is a need of normalizing the facial images regarding geometrical characteristics like pose and size. Moreover, the face is often further normalized regarding photometrical characteristics like grayscale, and illumination. Thus, there is a need of normalizing the faces in terms of photo-metrically and geometrically with the incorporation of feature extraction methods for offering efficient details to vary among faces of various human beings and stability regarding the differentiations in a photometrical and geometrical way. Thus, machine learning strategies come into this field of face recognition. Consequently, these techniques suffer from processing

Table 1: Features and challenges of existing face recognition approaches

Author [citation]	**Methodology**	**Features**	**Challenges**
Garain *et al.* [18]	KMCS	• It achieves a higher convergence time. • This model offers faster computation and has less expensive.	• It is more sensitive to noise.
Li *et al.* [19]	FTFA-DLF	• It achieves better face recognition accuracy and performance with the hand-craft features. • It increases the efficiency regarding superior light illumination robustness.	• It reduces the computational complexity.
Zhuang *et al.* [20]	DCNN	• It efficiently distinguishes high-quality face images from poor-quality ones. • It has finally ensured the recognition performance.	• It is not applicable for real-time large-scale dataset.
Vasanthi *et al.* [21]	ASM	• It minimizes the computational time. • It is more suitable for biometric-based authentication system.	• It suffers from exploring large scale datasets.
Senthilkumar and Gnanamurthy [22]	WBFD	• It attains better recognition performance, lower memory, and less recognition time.	• It does not include details of the lips, nose and eyes, after two levels of wavelet transform in Yalefaces.
Junaid *et al.* [23]	CNN	• It reduces the average execution time and increases the average recognition performance. • It has higher generalization ability.	• It does not solve illumination conditions in dark images.
Waisy *et al.* [24]	DBN	• It minimizes the running time. • It gives superior performance on constraints like high diversity in terms of noise, lighting conditions, facial expressions, and so on.	• It faces complications in processing various facial differentiations.
Ameur*et al.* [25]	SVM	• It is less expensive approach. • It attains lowest computational complexity.	• It is not suitable for video-based face recognition.

the occluded images owing to the obstacles, mask, or scarf in the faces, absence, and presence of moustache, beard, spectacles, etc, profile vs frontal, scaling factors, changes in pose, aging, variations in illumination, and facial expression modification. Thus, this research tries to solve the issues in facial expression variation. The recommended model on the face recognition model is displayed in FIGURE 1.

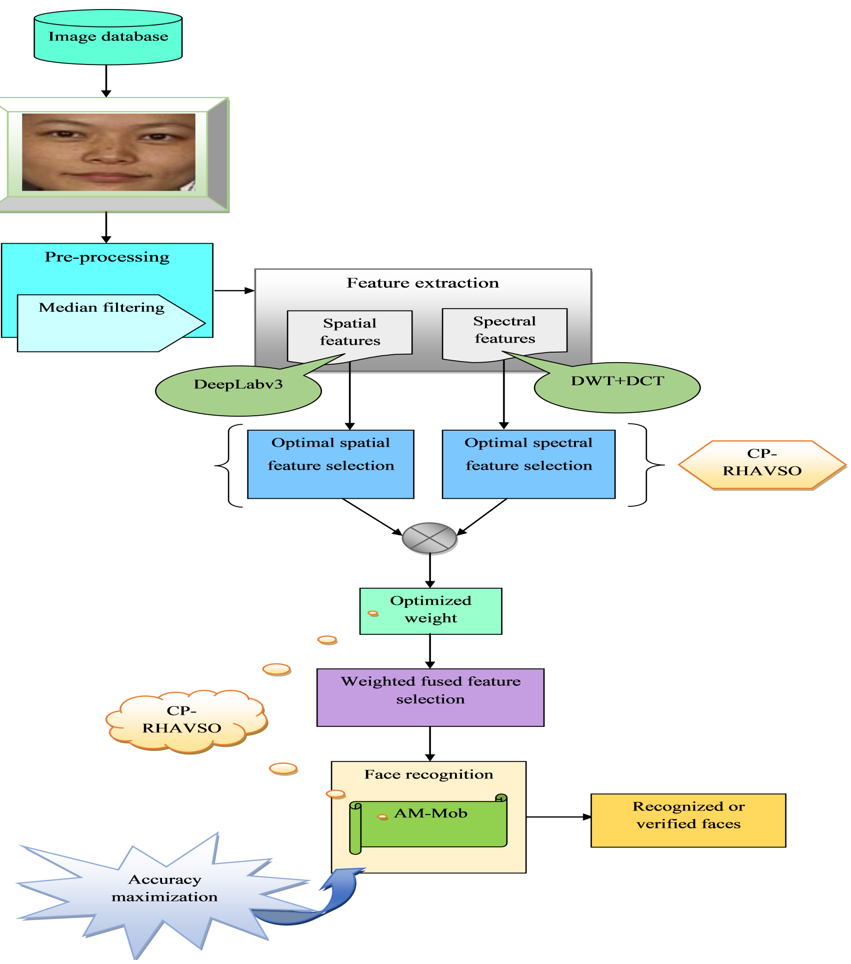


Figure 1: Transfer learning aided Face Recognition and Verification Framework

This research work acquires the standard facial images for recognizing faces. The proposed model primarily performs pre-processing as the essential step with the use of the median filtering technique for preserving the sharp edges and smoothing the spiky noise, and also avoiding the noise for further processing. Next, the spatial and spectral features are extracted, where DeepLabV3 is used to extract the spatial features, DWT and DCT are used to extract the spectral characteristics. The spectral and spatial feature extraction helps in getting the visual representation of images in better way. Then, the proposed CP-RHAVSO algorithm helps in getting the optimal spectral features and optimal spatial features. These chosen optimal spectral and spatial features are fed to the weighted fused feature extraction process, where the weights are tuned by the CP-RHAVSO algorithm. Finally, these acquired weighted fused features are fed to the AM-Mob for recognizing the faces, where the parameters in Multiscale Mobilenet are optimized by the newly suggested CP-RHAVSO technique.

This is named as AM-Mob and is designed for recognizing the facial images of persons. It aims to improve the facial recognition process' accuracy rate.

3.2 Face Recognition Dataset

In this proposed face recognition framework, the standard facial images are acquired for recognizing faces from two datasets.

Dataset 1: It is collected from "http://www.cfpw.io/. It is known as the CFPW dataset, which has a set of frontal-profile views of celebrities. The CFPW dataset has total 7000 images from 500 identities, with 10 frontal and 4 profile pictures each.

Dataset 2: It is named as Yale Dataset and gathered from "http://vision.ucsd.edu/content/yale-face-database. It has 165 images of 15 people with various facial expressions like a wink, surprised, sleepy, sad, right-light, normal, without or with glasses, left-light, happy, and center-light.

The, acquired images are known as ℓ_b, where $b = 1, 2, 3, \cdots, B$ and B specifies the total number of processed images. FIGURE 2 shows the examples acquired from the gathered datasets.

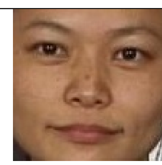
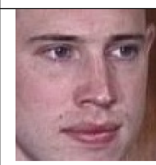
Description	1	2	3	4	5
Dataset 1					
Dataset 2					

Figure 2: Representative examples acquired from two datasets for face recognition process

4. DEVELOPMENT OF CP-RHAVSO-BASED WEIGHTED FUSED FEATURE SELECTION FOR INTELLIGENT FACE RECOGNITION AND VERIFICATION

4.1 Pre-processing of Images

The designed face recognition and verification model uses the median filtering for pre-processing the acquired images ℓ_b. It is a non-linear filtering method intended for eradicating the noise in images. It also focused on preserving the edges and avoiding errors in processing due to the noise. Median filter [26] works by operating over the window through the selection of median intensity in the window. It is an average and robust filter, which takes the averaging among the illustrative pixel in the adjacent. It processes the images regarding pixel by pixel and replaced them with the median value of the adjacent pixel. The window is named with the pattern of neighbours that performs the sliding of pixels by pixels. The quality of the images is enhanced by median filtering. The gathered images ℓ_b are randomly specified as a group of variables $\ell_b = (\ell_1, \ell_2, \cdots, \ell_B)$ and the median value is determined in Eq. (1).

$$med\,(\ell_b) = \begin{cases} \ell_{X+1} = \ell_r, & B = 2X + 1 \\ \frac{1}{2}\,(\ell_{X+1} + \ell_X) & B = 2X \end{cases} \tag{1}$$

Here, the median rank is specified as $r = \ell_{X+1}$. The 2-D median filter is measured for intensity values in Eq. (2).

$$yx_{ik,jk} = \underset{(rj,sj)\in\vartheta}{med} \left(\ell_{ik+rj,jk+sj}\right) \tag{2}$$

Here, window is termed as ϑ. At last, the filtered images ℓ_b^{MF} are acquired and fed to the next level of processing. The imaging results for the pre-processing step are illustrated in FIGURE 3.

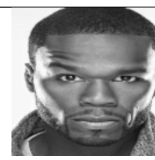
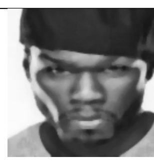
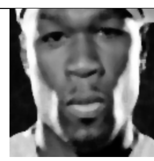
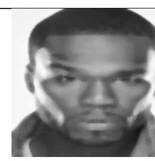
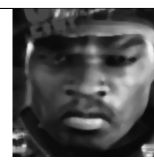
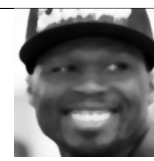
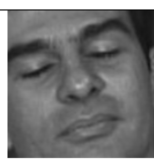
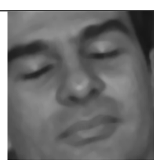
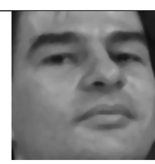
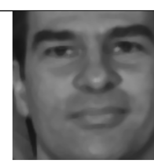
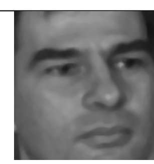
Description	1	2	3	4	5
Dataset 1					
Input images					
Pre-processed images					
Dataset 2					
Input images					
Pre-processed images					

Figure 3: Example of a figure caption. *(figure caption)*

4.2 Spatial and Spectral Feature Extraction Techniques

In this recommended face recognition model, from the pre-processed images ℓ_b^{MF}, the spatial and spectral features are gathered using the standard techniques as discussed here.

Spatial feature extraction: The spatial features are gathered from the images for preserving the essential spatial information of the images through maximizing the separability and reducing the dimensionality among the classes. The extracted spatial contextual information from the images is generally useful for classification purposes. DeepLabv3 [27] takes the input as the pre-processed images ℓ_b^{MF} that is used in this paper for getting the spatial features. This network uses various parallel atrous convolution in a parallel way with several rates. DeepLabv3 recovers the spatial information from the input images at gradually. It is a simple network with the adoption for getting the sharper details.

This network aims to use cascaded atrous convolution for capturing multi-scale information by incorporating several atrous rates integrated with an improved Atrous Spatial Pyramid Pooling (ASPP) stage with the encoding of image-level features. It increases the efficiency even without using the Dense Conditional Random Field (CRF) to post-process the images. The dilated or atrous convolutions utilize a dilation or atrous rate for making the larger field-of-view filters. The dense feature maps are extracted by the convolutional networks to eradicate the downsampling processes and then, upsampled the equivalent filter kernels. The multi-scale contextual information is obtained using ASPP and attained the features maps from a backbone network. The object features are extracted by using several levels of atrous rates from four atrous parallel convolutions. Next, the global average pooling is applied to the final feature map of the backbone network by incorporating the global contextual information along with the integration of image-level features at several scales. In the last few blocks of the backbone, an atrous convolution is used. Finally, the outcomes of every operation with the channel are integrated and then, the final result is attained by the 1*1 convolution.

Finally, the spatial features $SF_b^{DeepLabv3}$ are acquired from the DeepLabv3.

Spectral feature extraction: Every image pixel is indicated as a pattern and its spectrum (it specifies a vector of various pixel values in several spectral channels), which is known as the initial group of features. This group of features is also specified as spectral features. Spectral features suffer from processing the spatial dependencies of the neighbouring pixels, and also redundant features. Thus, there is a need of minimizing the dimensionality of the input data. Thus, this research work uses the DCT and DWT techniques for gathering the spectral features from the pre-processed images.

DWT [28]: It is vastly applied in large-scale of applications like computer science, mathematics, engineering, science, etc. It acquires a superior compression ratio without avoiding essential information from the images. It does not suffer from the artifacts in the images. Thus, DWT is chosen here for gathering the spectral features.

The DWT of input pre-processed images $\ell_b^{MF}(ik, jk)$ is derived in Eq. (3).

$$\xi_\phi(a_0, e, f) = \frac{1}{\sqrt{EF}} \sum_{ik=0}^{E-1} \sum_{jk=0}^{F-1} \ell_b^{MF}(ik, jk)\, \phi_{(a_0,e,f)}(ik, jk) \tag{3}$$

$$\xi_\phi^h(a, e, f) = \frac{1}{\sqrt{EF}} \sum_{ik=0}^{E-1} \sum_{jk=0}^{F-1} \ell_b^{MF}(ik, jk)\, \chi_{(a,e,f)}^h(ik, jk) \tag{4}$$

Here, the translated and scaled basis functions are denoted by ϕ and χ, where the diagonal, vertical, and the horizontal information for scales are represented by h. These basis functions are derived as follows.

$$\phi_{(a,e,f)}(ik, jk) = 2^{\frac{a}{2}} \phi(2^a ik - e,\ 2^a jk - f) \tag{5}$$

$$\chi_{(a,e,f)}^{h}(ik, jk) = 2^{\frac{a}{2}} \phi^{h}(2^{a} ik - e,\, 2^{a} jk - f) \tag{6}$$

$$\begin{aligned} \ell_{b}^{MF}(ik, jk) = {} & \frac{1}{\sqrt{EF}} \sum_{ik=0}^{E-1} \sum_{jk=0}^{F-1} \xi_{\phi}(a_0, e, f)\, \phi_{(a_0,e,f)}(ik, jk) \\ & + \frac{1}{\sqrt{EF}} \sum_{h=I,V,D} \sum_{a=a_0}^{\infty} \xi_{\phi}(a_0, e, f)\, \phi_{(a_0,e,f)}(ik, jk) \\ & \times \sum_{e} \sum_{f} \phi_{\chi}^{h}(a_0, e, f)\, \chi_{(a,e,f)}^{h}(ik, jk) \end{aligned} \tag{7}$$

Finally, the spectral features are acquired from pre-processed images ℓ_{b}^{MF} and obtained the spectral features SP_{b}^{DWT} are acquired from the DWT.

DCT [29]: It is described as a method used for representing a restricted sequence of input data as the sum of cosine functions hovers at varied frequencies. The DCT performs the initially encoding procedure to encode the acquired pre-processed images. DCT generally enhances the contrast, effective for illumination differentiations while estimating with DWT, simpler and solved the complexity.

The DCT of input pre-processed images $\ell_{b}^{MF}(ik, jk)$ is derived in Eq. (8).

$$\begin{aligned} & \ell_{b}^{MF}(ik, jk) = \beta(ik)\, \beta(jk) \sum_{e=0}^{E-1} \sum_{f=0}^{F-1} \ell_{b}^{MF}(e, f) \\ & \cos\left(\frac{\pi(2ik+1)e}{2E}\right) \cos\left(\frac{\pi(2ik+1)f}{2F}\right), \forall 0 \le ik \le E-1 \\ & \forall 0 \le jk \le F-1 \end{aligned} \tag{8}$$

$$\beta(ik, jk) = \begin{cases} \sqrt{\frac{1}{(E,F)}} & (ik, jk) = 0 \\ \sqrt{\frac{2}{(E,F)}} & (ik, jk) \neq 0 \end{cases} \tag{9}$$

Similarly, the inverse DCT is derived in Eq. (10).

$$\begin{aligned} & \ell_{b}^{MF}(e, f) = \sum_{e=0}^{E-1} \sum_{f=0}^{F-1} \ell_{b}^{MF}(ik, jk) \cos\left(\frac{\pi(2ik+1)e}{2E}\right) \\ & \cos\left(\frac{\pi(2ik+1)f}{2F}\right), \\ & \forall 0 \le e \le E-1 \\ & \forall 0 \le f \le F-1 \end{aligned} \tag{10}$$

Then, the spectral features are attained from pre-processed images ℓ_{b}^{MF} and obtained the spectral features from the DWT is termed as SP_{b}^{DCT}.

The final set of spectral features is collectively known as $SP_{b}^{spectral} = \left\{SP_{b}^{DCT}, SP_{b}^{DWT}\right\}$, which are forwarded to the next level of process.

4.3 Optimal Spatial and Spectral Feature Selection

The proposed face recognition and verification model uses CP-RHAVSO algorithm for gathering the most essential features. The optimal features are chosen from the acquired features spatial features $SF_b^{DeepLabv3}$ and spectral features $SP_b^{spectral}$ using CP-RHAVSO algorithm. Here, the extracted optimal features are correspondingly specified as $OSF_{b*}^{DeepLabv3}$, and $OSP_{b*}^{spectral}$. Generally, the optimal feature selection helps in promoting the performance regarding minimization of training time, enhancement of accuracy and minimization of the over-fitting issues, eradication of redundant data, acquisition of best features from the total number of gathered data, etc. Thus, this proposed work explores the better effectiveness in face recognition model. The spatial features are selected as 5 using CP-RHAVSO algorithm whereas the algorithm selects the spectral features as 5 by the CP-RHAVSO algorithm. The process of optimal spatial and spectral feature selection from the feature extraction phase with the incorporation of CP-RHAVSO algorithm is diagrammatically illustrated in FIGURE 4.

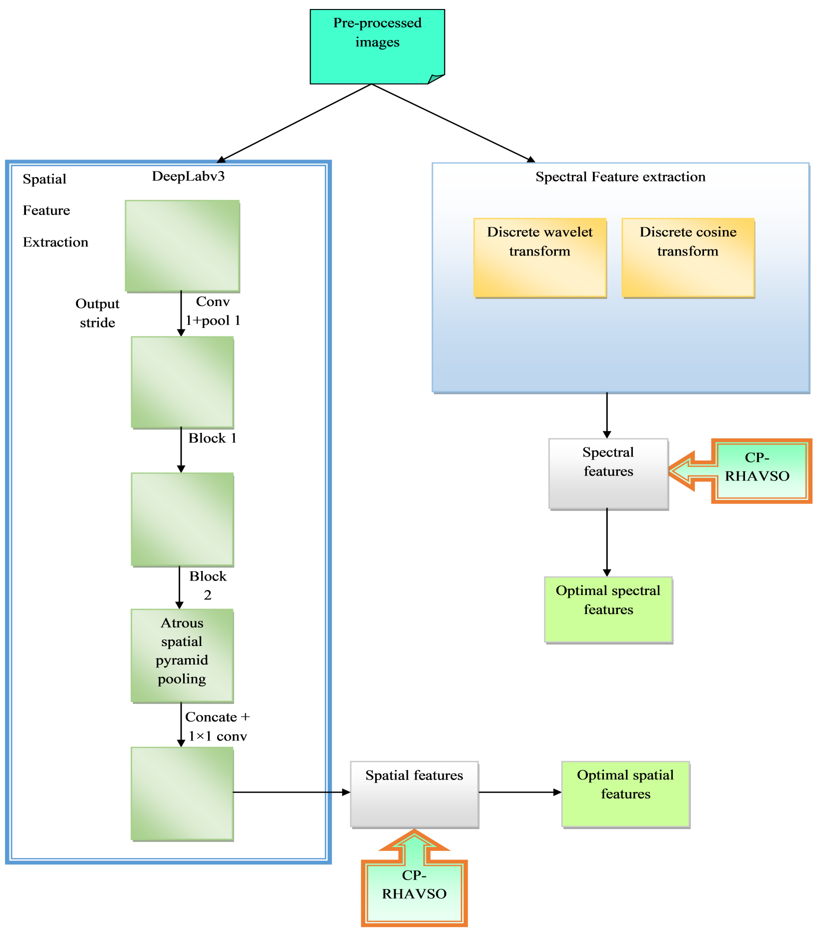


Figure 4: Optimal Spatial and Spectral Feature Selection using CP-RHAVSO algorithm

4.4 Optimal Weighted Fused Feature Selection

The proposed face recognition and verification model recommends the optimal weighted fused feature selection for optimally choosing the weighted features and forwarded to the face recognition framework. Here, the selection of optimal weighted fused features from the optimally chosen spatial and spectral features are done by CP-RHAVSO algorithm. The process of this optimal weighted fused feature selection using CP-RHAVSO algorithm is given in Eq. (11).

$$OWF_{ob*} = \left(Wg \times OSF_{b*}^{DeepLabv3}\right) + \left((1 - Wg) \times OSP_{b*}^{spectral}\right) \tag{11}$$

Here, the chosen optimal weighted fused features are known as OWF_{ob*}, the optimized weight is denoted as Wg that is tuned by CP-RHAVSO algorithm, the extracted optimal spatial and spectral features are termed as $OSF_{b*}^{DeepLabv3}$, and $OSP_{b*}^{spectral}$, which are chosen by CP-RHAVSO algorithm. Here, $ob* = 1, 2, 3, \cdots, OB*$ and the total amount of optimally selected features are equivalently derived as $OB*$. Here, the weight Wg is allocated in the range of [0.01 to 0.99] using CP-RHAVSO algorithm. The optimal weighted fused features are helpful in acquiring the noteworthy features and validating the efficiency regarding computational time, robustness, applicability, and reliability. This feature selection with a weighted strategy promotes the performance owing to the execution time and increasing the predictive efficiency. The recommended model enhances the accuracy and precision of face recognition along with the reduction of computational complexity. It finally reduces the redundancy in the gathered data rather than processing the raw data.

The representation of optimal weighted fused feature selection is illustrated in FIGURE 5.

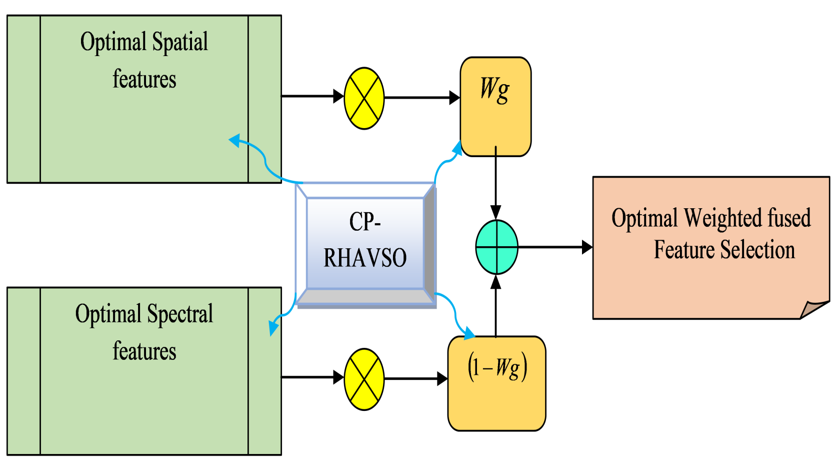


Figure 5: Optimal weighted fused feature selection using CP-RHAVSO algorithm

5. ARCHITECTURE OPTIMIZED ADAPTIVE MULTI-SCALE MOBILENET FOR FACE RECOGNITION AND VERIFICATION

5.1 Proposed CP-RHAVSO

The presented face recognition and verification framework adopts a new CP-RHAVSO technique for aiding various applications in real-time cases. This strategy helps in promoting the face recognition from the facial images. This strategy is useful for acquiring the optimal spatial and spectral features, and optimizing the weight used for getting the optimal weighted fused features, along with the implementation of AM-Mob by tuning the momentum in M-Mob for promoting higher efficiency in face recognition and verification with the promotion of superior convergence rate.

AVOA [30] has the ability to acquire optimal solutions, estimated and tested on the standard benchmark functions, the ability to avoid local optimization, premature convergence can be solved, lower computational complexity, and lower execution time, and thus, it is a more well-known and powerful optimization technique. However, it affects the global search of individuals in the solution area, falls into local optimal solutions in a later phase, and slower convergence speed using the own information of individuals. Lack of exploitation and exploration capability in the later stages is observed in AVOA. Therefore, RHSO [31] is adopted into this framework for getting the higher convergence ability towards offering the global optimal solutions along with the perfect balancing among the exploration and exploitation phases. It has only a few variables for achieving optimal solutions.

Thus, the integration of both RHSO and AVOA is carried out by adopting the controlling parameter CP_1 for determining the probability of choosing the strategies in the exploration phase of AVOA. If $CP_1 > 5$ is satisfied, then, the solutions are determined via RHSO strategy or else the AVOA updates the solutions. It promotes the superior performance for observing the higher efficiency regarding convergence rate. The AVOA algorithm is described as follows.

AVOA is designed by taking inspiration from a vulture (a hunting bird). These animals live in a group and followed a simple lifestyle for acquiring food resources. It lives in Africa and plays an individual characteristic and thus, it is categorized into three classes inclusive of stronger vultures known as the Lappet-face vulture, the weaker vulture, and the white-backed vulture. The stronger and healthy vultures have the ability to encircle weak vultures and received food. The feeding and finding of several vultures in Africa is modelled in AVOA. AVOA is formulated with the four strategies with the determination of the n-number of vultures in a search area, determination of fitness function of entire solutions for getting the categories of vultures including best vultures and second-best vultures. Then, the natural function is derived for discovering the food, where every cluster has an incapable vulture for getting the food. Therefore, from the hungry trap, the vultures are escaped by getting food. Here, the worst solutions are assumed in the $NP-$population by taking the hungriest and weakest vultures and attaining the final best solutions.

In the initial step, after forming the initial population, the fitness among entire solutions is determined and chosen the best vulture of the initial group is taken as the optimal solution BV_1. Next, the second-best solution is assumed as the second group BV_2. Then, Eq. (12) derives other vulture solutions with the movement of getting finer solutions for the initial and second groups. The whole population is re-determined in every iteration.

$$Q(j) = \begin{cases} BV_1 & if\ cp_j = S_1 \\ BV_2 & if\ cp_j = S_2 \end{cases} \tag{12}$$

Here, the search operation is determined by estimating the parameters S_1 and S_2, which are assigned in the range of $[0, 1]$, the Roulette wheel is used for gaining the best solutions with the probability rate for every group and computed in Eq. (13).

$$cp_j = \frac{S_j}{\sum_{j=1}^{np} S_j} \tag{13}$$

Here, rate of starvation of vultures is known as S and derived in Eq. (14).

$$S = (2 \times rt_1 + 1) \times a \times \left(1 - \frac{t_j}{t_{\max}}\right) + ts \tag{14}$$

$$ts = e \times \left(\sin^g\left(\frac{\pi}{2} \times \frac{t_j}{t_{\max}}\right) + \cos\left(\frac{\pi}{2} \times \frac{t_j}{t_{\max}}\right) - 1\right) \tag{15}$$

In the aforementioned equations, the rate of starvation of vultures is given as S, the arbitrary parameter is known as g, the arbitrary parameter in the boundary of $[0, 1]$ is specified as rt_1, the arbitrary parameter in the boundary of $[-2, 2]$ is indicated as e, the arbitrary parameter in the boundary of $[-1, 1]$ is derived as a, and recent iteration is given as t_j and the maximum number of iteration is noted as $t_{\max}$.

Then, the exploration phase is carried out with the help of RHSO algorithm.

RHSO algorithm is motivated from the Procavia capensis (Rock hyrax) mammals. It withstands several ambient temperatures and offers essential feed and water, which takes feeding every data in a circular motion. It also tries to predict the predators for getting food. If any threats are observed in the food source, then the alarm will be passed to their group. Moreover, a female is assumed as the leader of the group. The swarm of rock hyrax is formulated by assuming the leaders and their members. The leader FL selects the best and higher place for observing the residual of the group. The locations of rock hyrax are updated via Eq. (16).

$$V(j, i) = V(j, i) - (CR \times V(j, i) + FL) \tag{16}$$

$$FL = l_1 \times V(FL_V, i) \tag{17}$$

$$CR = \sqrt{\left(np_1^2 + np_2^2\right)} \tag{18}$$

$$np_1 = l_2 \times \cos(\phi) \tag{19}$$

$$np_2 = l_2 \times \sin(\phi) \tag{20}$$

Here, the radius in the boundary of $[0, 1]$ is referred to as l_2, a arbitrary number among $[0, 360]$ is represented by ϕ used for illustrating the angle of the move and executed in each generation, CR illustrates the circular motion for mimicking the circular system considered for rock hyrax, every diminution is denoted by i, the previous position of leader is given as FL_V, the previous location of search solution is mentioned as V and a arbitrary number among $[0, 1]$ is represented by l_1.

The exploitation phase of AVOA is designed with the incorporation of CP_2, and estimated along with the incorporating of rt_{CP_2}. Moreover, the competence for food is done with the factor of $|S| \geq 0.5$. Here, severe competition is noticed for the food collection and distributes the food with other vultures. However, the weak vultures focus on tiring and getting food from the strong vultures to gather food and create conflicts among them. This procedure is derived as follows.

$$V(j+1) = Df(j) \times (S + rt_4) - df(j) \tag{21}$$

$$df(j) = Q(j) - V(j) \tag{22}$$

$$Df(j) = |J \times Q(j) - V(j)| \tag{23}$$

Here, the recent position vector of vulture is given as $V(j)$, $Q(j)$ derives the optimal vultures of the two sets chosen, J specifies the coefficient vector utilized for enhancing the motion, and the arbitrary parameter in the boundary of [0, 1] is specified as rt_4.

Algorithm 1: CP-RHAVSO

```
Formulate the population and devise the parameters
Execute the fitness among the solutions
While (t < t_max)
    Derive the best and second best solutions
    For (every solution)
        Devise Q(j) by Eq. (12)
        Update S by Eq. (14)
        If |S| ≥ 1
            If CP_1 > 5
                Update the solutions by RHSO using Eq. (16)
            End if
        Else
            Update the solutions by AVOA
            If |S| < 1
            If |S| ≥ 0.5
                If CP_2 ≥ rt_CP_2
                    Execute the solutions via Eq. (21)
                Else
                    Execute the solutions via Eq. (24)
                End if
            Else
                If CP_3 ≥ rt_CP_3
                    Execute the solutions via Eq. (27)
                Else
                    Execute the solutions via Eq. (30)
                End if
            End if
            End if
        End if
    End for
End while
Return the optimal solutions
```

The flowchart of the implemented CP-RHAVSO technique is shown in FIGURE 6.

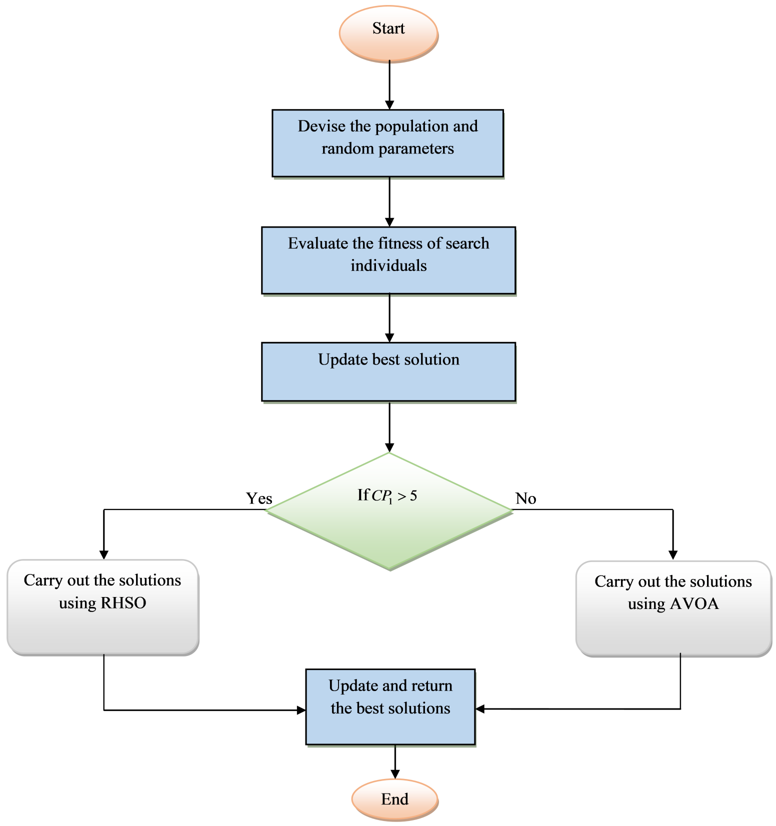


Figure 6: Flowchart of CP-RHAVSO technique for enhancing the face recognition and verification framework

The next step is to update the solutions via rotational flight as given in Eq. (24).

$$V\,(j+1) = Q\,(j) - (R_1 + R_2) \tag{24}$$

$$R_1 = Q\,(j) \times \left(\frac{rt_5 \times V\,(j)}{2\pi}\right) \times \cos\,(V\,(j)) \tag{25}$$

$$R_2 = Q\,(j) \times \left(\frac{rt_6 \times V\,(j)}{2\pi}\right) \times \sin\,(V\,(j)) \tag{26}$$

The arbitrary parameters in the boundary of [0, 1] is specified as rt_5 and rt_6, and the cosine and sine are correspondingly specified as $\cos$ and $\sin$. Then, the second phase of exploitation is determined based on the aggressive siege-fight scheme with the incorporation of CP_3.

Then, the movement of entire vultures to the food source is estimated and accumulated various species of vultures are on one food resource. The updating of the position of the recent vector is designed by Eq. (27).

$$V\left(j+1\right)=\frac{U_1+U_2}{2} \tag{27}$$

$$U_1=BV_1\left(j\right)-\frac{BV_1\left(j\right)\times V\left(j\right)}{BV_1\left(j\right)-V\left(j\right)^2}\times S \tag{28}$$

$$U_2=BV_2\left(j\right)-\frac{BV_2\left(j\right)\times V\left(j\right)}{BV_2\left(j\right)-V\left(j\right)^2}\times S \tag{29}$$

The vulture position in the upcoming iteration is denoted as $V\left(j+1\right)$.The aggressive competitiveness towards food is determined while taking $|S|<0.5$. The weaker vulture fights aggressively for food and derived this motion as given in Eq. (30).

$$V\left(j+1\right)=Q\left(j\right)-|df\left(j\right)|\times S\times levy\left(df\right) \tag{30}$$

Here, the distance of the vulture to one of the optimal vultures of two sets is determined by $df\left(j\right)$.

Finally, the solution updating is done until reaching the final set of iterations. The pseudo code of the designed CP-RHAVSO technique is depicted in Algorithm 1.

5.2 Multi-Scale Mobilenet

In this proposed model, M-Mob is selected for recognizing faces to identify the persons from the acquired facial images. It is the extended version of Mobilenet. M-Mob [32] is particularly designed based on the adoption of multi-scale convolution kernels into the expansion of channels in the block of MobileNet. It uses the sigmoid function instead of the hard-swish function for improving the multi-scalar feature extraction and enhancing the ability of non-linear mapping. This network comprises three layers. Initially, a kernel size of 3×3 is applied with the standard convolution for extending the input size to 16. Next, the output is normalized using the sigmoid function, and then, it enhances the non-linear mapping capability. Further, the higher dimensional features are extracted in the second layer with the M-Mob, and further, the dimension of the output feature is expanded up to 32 for increasing the elements of details and contents. Next, a kernel size of 1×1 is utilized with the standard convolution in the third layer for integrating the features of the amplified images in a linear way for getting the final outcomes. Generally, Mobilenet is designed by taking the earlier depth separable convolution and the residual framework is inverted with the linear bottleneck. The convolution is reduced to depth separable convolution for factorizing the common convolution when managing a superior representation capability. The dimension of feature maps in the input is extended and filtered the expanded feature maps with a depthwise lightweight convolution. The residual connection of Mobilenet is similar to the Residual network while using a similar shape of the input and output. Next, the gradient vanishing issue of deep networks is dealt with by the residual connection in the network. The attention strategy is enhanced in the M-Mob. Moreover, the squeeze-and-excitation module is adopted into the M-Mob for performing the linear integration of the expanded feature maps. Then, on every channel, the average pooling is carried out via the same module. The weights are further attained via the fully connected layers and then, the scaling of expanded feature maps is performed for getting the salient feature maps. In addition, the input channels are extended by adopting the varied sizes of convolutional kernels that are taken from the Inception module. Next, the fine scale is used for extracting the details on the acquired optimal weighted fused features. Therefore, it is more useful in recognizing the salient regions on the coarse scale. The multi-scale features are produced with varied sizes of convolutional kernels. The channel-extended convolution is adopted with the kernels of 1×1 and 3×3 convolutions. here, the kernels of 3×3 convolutions are focused on the regional ones based on the coarse scale nature whereas the kernels of 1×1 are adopted on the fine scale with the examination of pixel relations. Therefore, it acquires the regional and details on salient features for getting the final outcomes. The channel-wise concatenation of convoluted feature maps is performed on the output channel number of every kernel. The shallow network is endowed with the squeeze-and-excitation module to carry out the average pooling operation to maximize the global saliency with the higher recognition capability.

5.3 Adaptive Multi-Scale Mobilenet

This recommended face recognition and verification framework suggests a novel AM-Mob for performing the recognition of faces with the introduction of CP-RHAVSO algorithm. It helps in maximization of accuracy to promote the recognition ability. This AM-Mob framework using CP-RHAVSO algorithm helps in the recognition of faces with the person by taking the input as optimal weighted fused features. If any new untrained data is given into this framework, then this model exhibits the output as unrecognized faces. Thus, this framework is highly recommendable for promoting the superior efficiency on face recognition. Here, the momentum coefficient in M-Mob is optimized using CP-RHAVSO algorithm. It promotes the higher recognition rate. The aim of this work is to enhance the accuracy rate of the recommended AM-Mob framework for face recognition as modelled in Eq. (31).

$$LK = \underset{\left\{Wg, OSF_{b*}^{DeepLabv3}, OSP_{b*}^{spectral}, MC\right\}}{\arg\min} \left(\frac{1}{Acry}\right) \tag{31}$$

In Eq. (31), the weight Wg is tuned in the range of [0.01 to 0.99] and momentum coefficient in M-Mob MC is tuned by CP-RHAVSO algorithm with the range of [0.01 to 0.99]. Thus, the accuracy $Acry$ is derived in Eq. (32).

$$Acry = \frac{(qrm + qrn)}{(qrm + qrn + qrp + qrq)} \tag{32}$$

Here, terms qrq, qrn, qrp, and qrm indicates the “false negatives, true negatives, false positives, and true positives”. This model finally achieves high quality outcomes to recognize the faces from the acquired datasets along with the optimization of momentum coefficient in M-Mob. The momentum coefficient utilizes the theory of ‘exponentially weighted average’ of the gradients for training along with the gradient descent technique. FIGURE 7 illustrates the diagrammatic representation of AM-Mob framework using CP-RHAVSO algorithm.

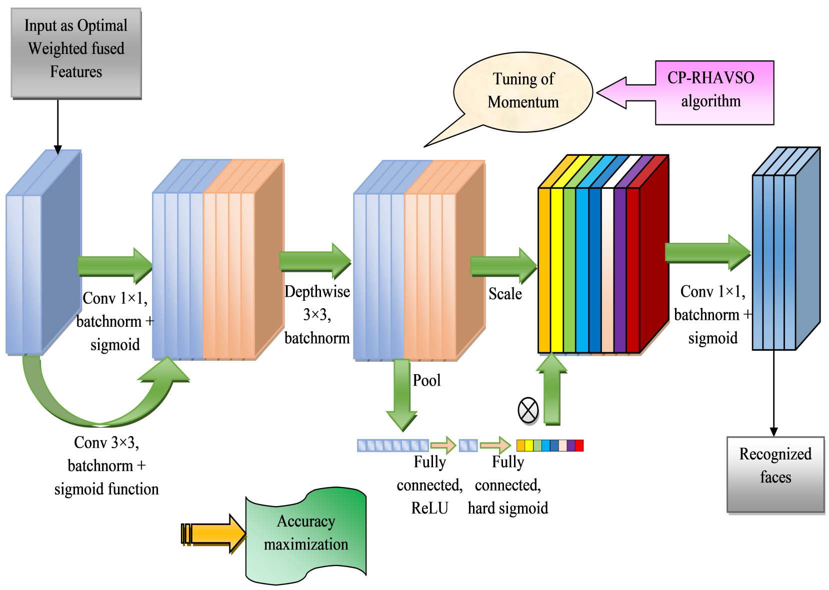


Figure 7: Recommended AM-Mob framework for face recognition using CP-RHAVSO algorithm

6. RESULTS AND EVALUATION

6.1 Simulation Setup

The proposed face detection model was evaluated in Python and estimated the performance over standard performance metrics along with factors including the maximum number of iterations was taken as 25, the number of population as 10, and the length of the solution was assumed as 12. The effectiveness was analyzed over techniques including Water Strider Algorithm (WSA) [33], Coronavirus Herd Immunity Optimization Algorithm (CHIA) [34], RHSO [31], and AVOA-M-Mob [30], CNN [23], DCNN [20], DBN [24], and M-Mob [32].

6.2 Performance Metrics

The standard measures listed for estimating the efficiency of this model are illustrated here.

(a) F1 score is equated in Eq. (33).

$$F1 - score = \frac{2qrm}{2qrm + qrp + qrq} \tag{33}$$

(b) Mathews correlation coefficient (MCC) is equated in Eq. (34).

$$MCC = \frac{qrm \times qrn - qrp \times qrn}{\sqrt{(qrm+qrp)(qrm+qrq)(qrn+qrp)(qrn+qrq)}} \tag{34}$$

(c) False Discovery Rate (FDR) is formulated in Eq. (35).

$$FDR = \frac{qrp}{qrp + qrm} \tag{35}$$

(d) False Negative Rate (FNR) is shown in Eq. (36).

$$FNR = \frac{qrq}{qrm + qrn} \tag{36}$$

(e) Negative Predictive Value (NPV) is derived in Eq. (37).

$$NPV = \frac{qrn}{qrq + qrn} \tag{37}$$

(f) False positive rate (FPR) is given in Eq. (38).

$$FPR = \frac{qrp}{qrp + qrn} \tag{38}$$

(g) Sensitivity is derived in Eq. (39).

$$Se = \frac{qrm}{qrm + qrq} \tag{39}$$

(h) Specificity is formulated in Eq. (40).

$$Spc = \frac{qrn}{qrn + qrp} \tag{40}$$

(i) Precision is indicated as pc, and formulated in Eq. (41).

$$pc = \frac{qrm}{qrm + qrp} \tag{41}$$

6.3 Estimation on Feature Extraction

The performance of the recommended face recognition model is verified in terms of optimal weighted fused feature selection as given in TABLE 2. The recommended optimal weighted fused feature using CP-RHAVSO algorithm acquires higher performance rate while differentiating it with other methods. Without optimization, the recommended CP-RHAVSO algorithm surpasses the performance and enhances the efficiency.

6.4 Analysis Over Varying the Learning Percentages on Heuristic Strategies

The competency of the suggested face recognition model is verified over standard algorithms for two diverse datasets as shown in FIGURE 8, and FIGURE 9. The CP-RHAVSO algorithm-based AM-Mob is implemented for face recognition framework that increases the overall enhancement of the designed strategy and achieved superiority over conventional methods. For example, the F1-score of the implemented AM-Mob using CP-RHAVSO algorithm at 55% is 53%, 50%, 33%, and 25% maximized in that order of WSA-M-Mob, CHOA-M-Mob, RHSO-M-Mob and AVOA-M-Mob techniques for the dataset 1. Consequently, the implemented model surpasses the superior

Table 2: Estimation on the optimal weighted fused feature selection with other feature selection approaches for the face recognition and verification framework

Terms	Spatial Features using Deeplabv3	Spectral Features using DWT and DCT	Fused spatial and spectral features	Optimal weighted fused feature using CP-RHAVSO algorithm
Dataset 1				
“Accuracy”	92.08986	91.57597	94.70474	96.27194
“Sensitivity”	92.1	91.6	94.77143	96.28571
“Specificity”	92.08984	91.57592	94.70461	96.27192
“Precision”	2.280114	2.132606	3.462386	4.921073
“FPR”	7.910163	8.424077	5.295391	3.728085
“FNR”	7.9	8.4	5.228571	3.714286
“NPV”	92.08984	91.57592	94.70461	96.27192
“FDR”	97.71989	97.86739	96.53761	95.07893
“F1-score”	4.450058	4.16817	6.680698	9.363582
“MCC”	0.138027	0.13261	0.17573	0.213253
Dataset 2				
“Accuracy”	92.13867	91.56533	94.704	96.27467
“Sensitivity”	92	91.6	94.53333	96.4
“Specificity”	92.1415	91.56463	94.70748	96.27211
“Precision”	19.28452	18.14101	26.71439	34.54372
“FPR”	7.858503	8.435374	5.292517	3.727891
“FNR”	8	8.4	5.466667	3.6
“NPV”	92.1415	91.56463	94.70748	96.27211
“FDR”	80.71548	81.85899	73.28561	65.45628
“F1-score”	31.8854	30.28433	41.65687	50.86177
“MCC”	0.400967	0.386413	0.487187	0.56517

efficiency over various heuristic strategies for dataset 2 also. Therefore, the recommended model is more applicable for detecting the faces in the acquired datasets.

6.5 Analysis Over Varying the Learning Percentages on Classifiers

The performance of the proposed face recognition framework is estimated over various classifiers for two datasets as shown in FIGURE 10, and FIGURE 11. The suggested CP-RHAVSO algorithm-based AM-Mob framework attains higher recognition performance while estimating with other strategies. Thus, the accuracy of the implemented CP-RHAVSO algorithm-based AM-Mob for dataset 2 at 85% is 15%, 12%, 8.8%, and 6.5% enhanced in that order of CNN, DCNN, DBN and M-Mob techniques. Therefore, the offered approach has clearly exhibited the superiority over other deep learners.

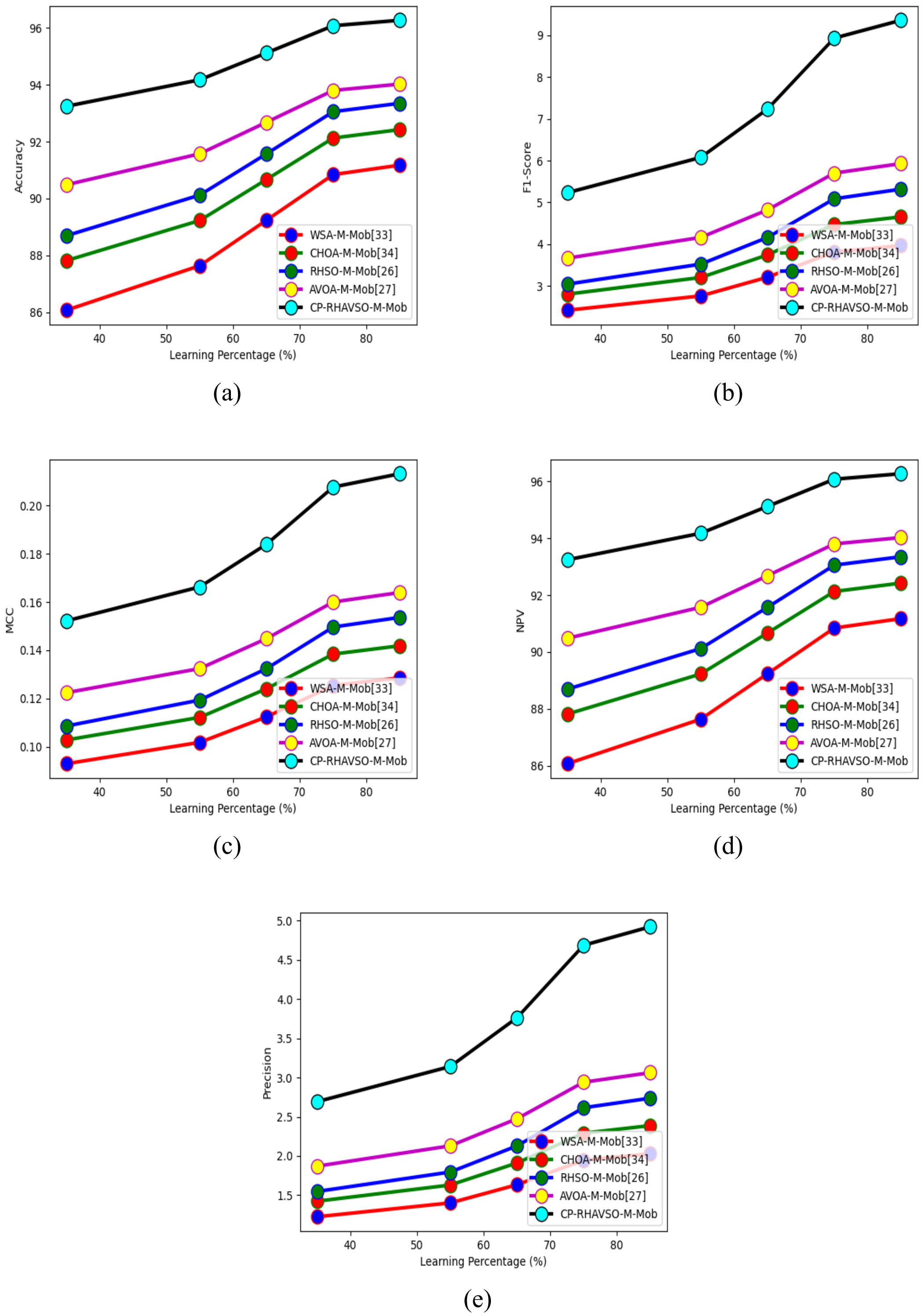
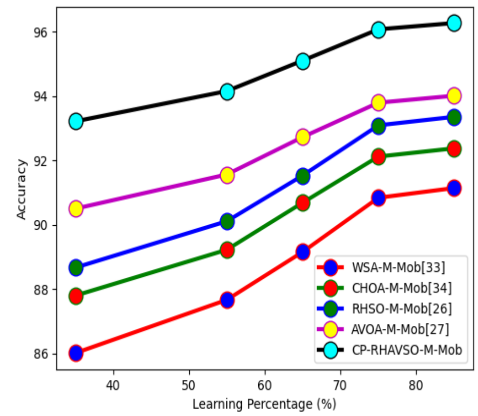
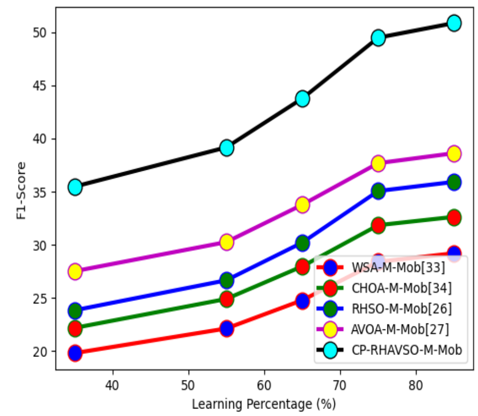


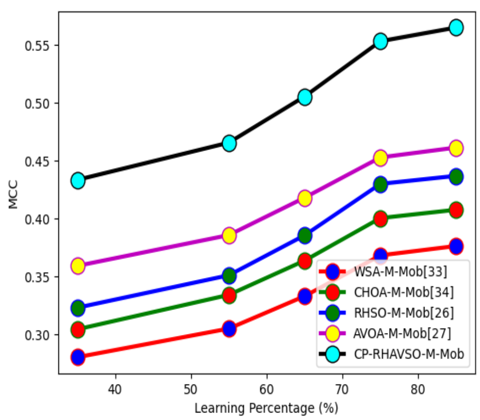
Figure 8: Inspection on the recommended face recognition model over heuristic algorithms by differing the learning percentages for dataset 1 respect to “(a) Accuracy, (b) F1-Score, (c) MCC, (d) NPV, and (e) Precision”



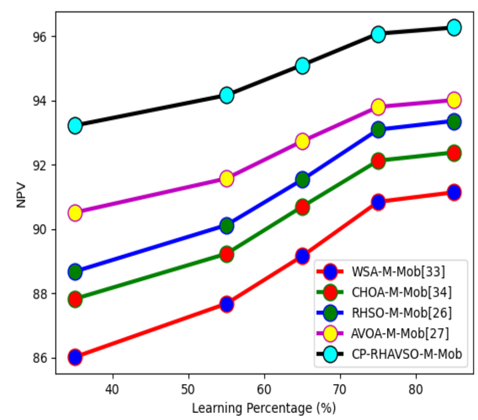
(a)



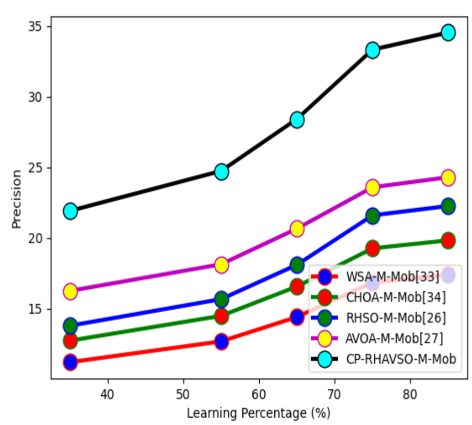
(b)



(c)

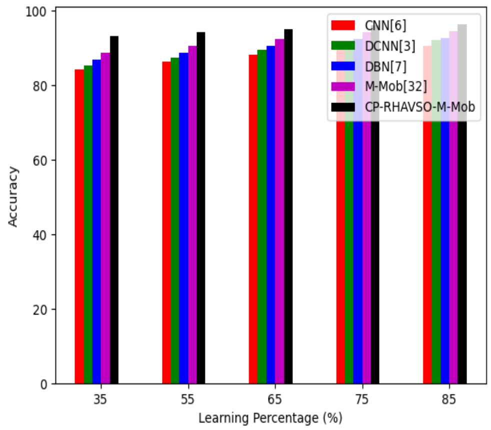


(d)

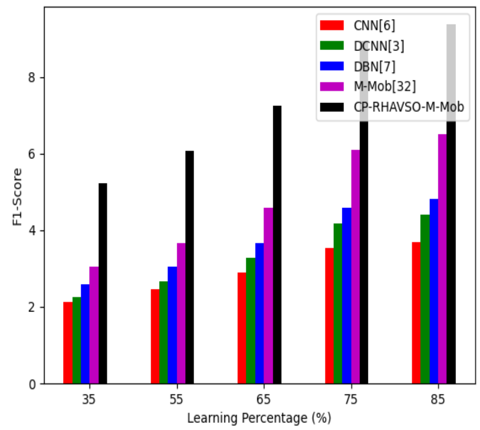


(e)

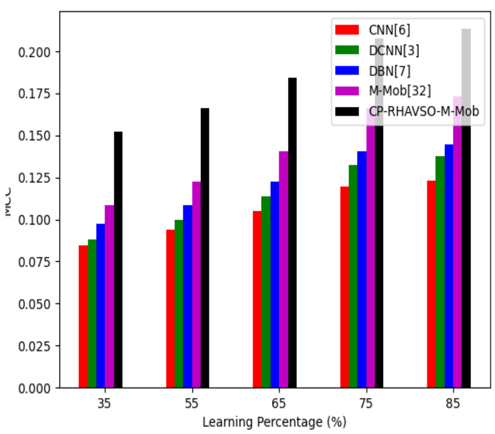
Figure 9: Inspection on the recommended face recognition model over heuristic algorithms by differing the learning percentages for dataset 2 respect to “(a) Accuracy, (b) F1-Score, (c) MCC, (d) NPV, and (e) Precision”



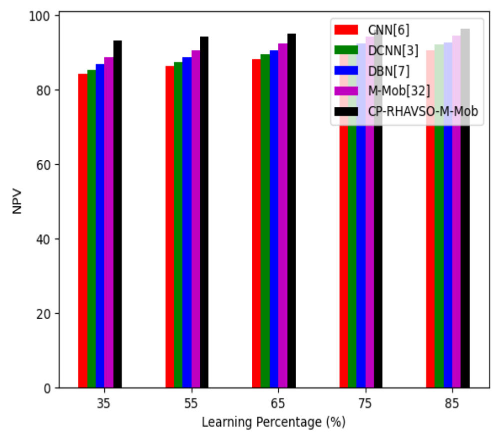
(a)



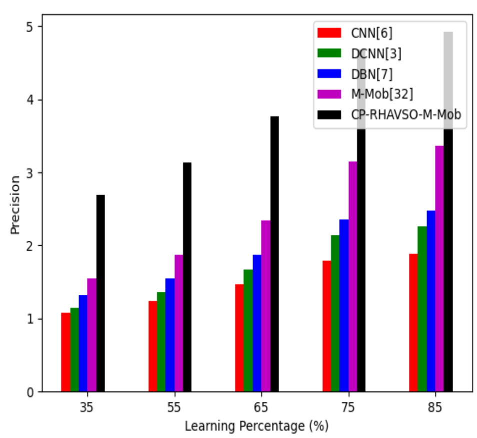
(b)



(c)



(d)



(e)

Figure 10: Inspection on the recommended face recognition model over deep learning algorithms by differing the learning percentages for dataset 1 respect to "(a) Accuracy, (b) F1-Score, (c) MCC, (d) NPV, and (e) Precision"

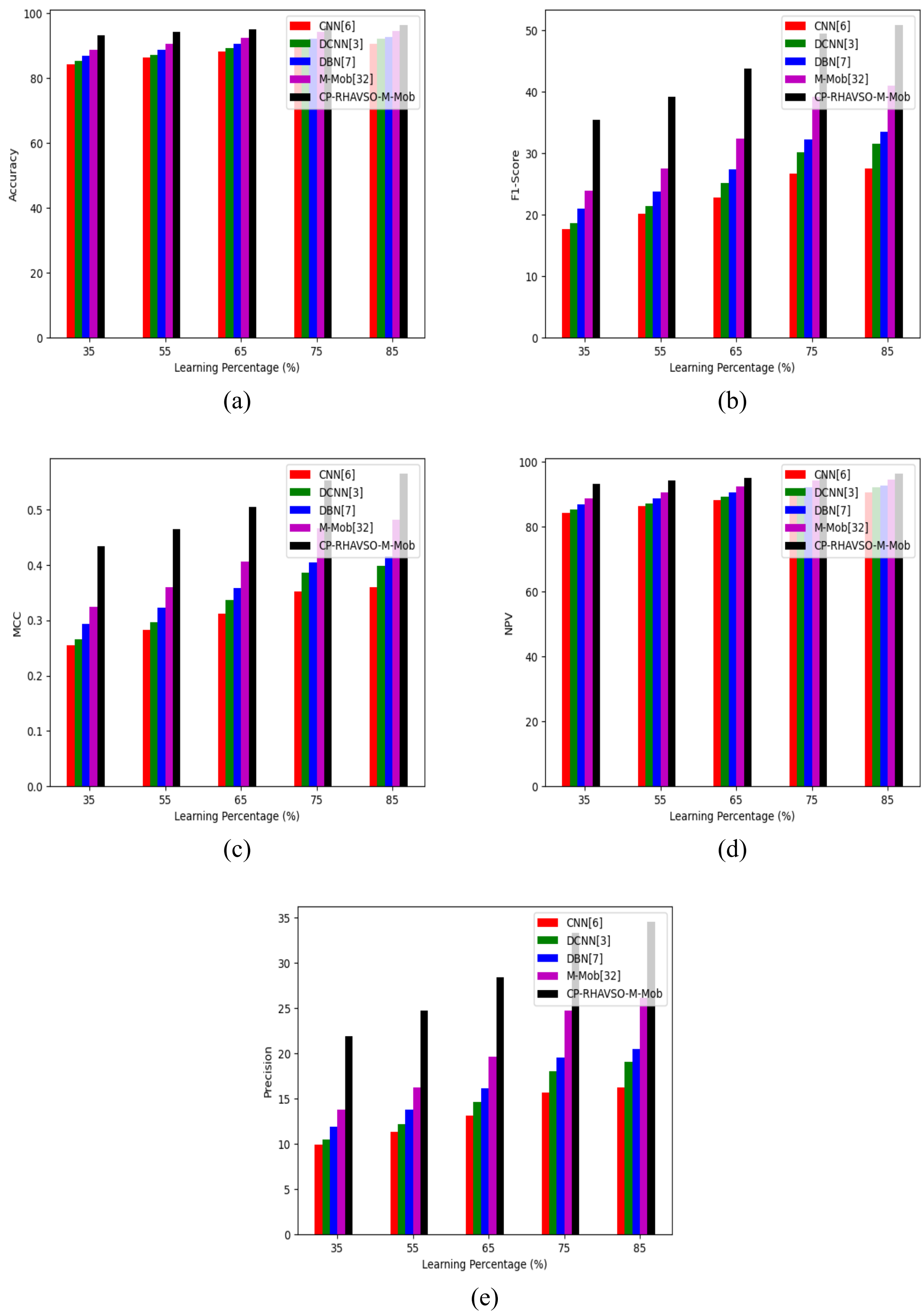
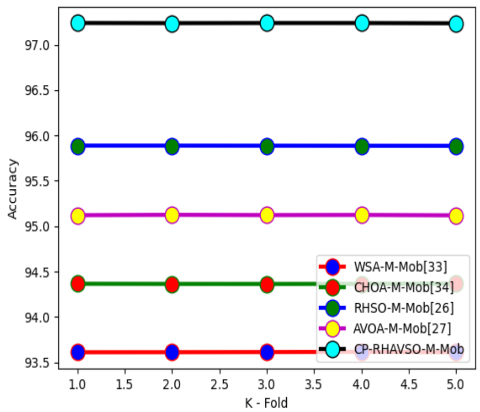
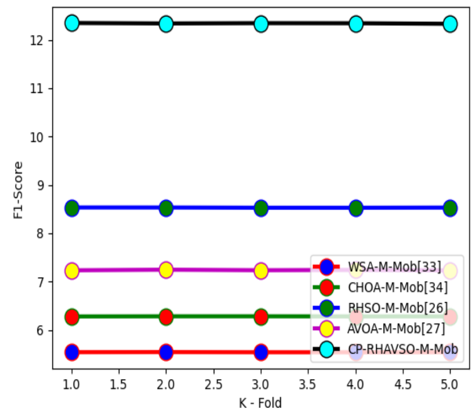


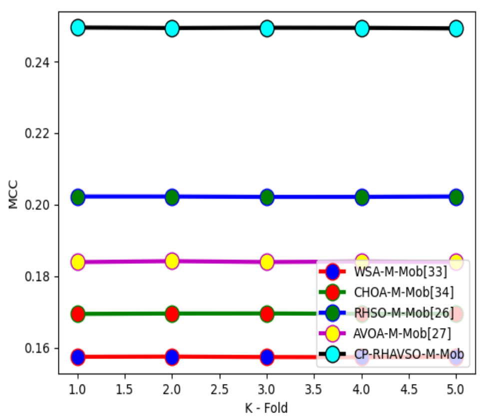
Figure 11: Inspection on the recommended face recognition model over deep learners by differing the learning percentages for dataset 2 respect to “(a) Accuracy, (b) F1-Score, (c) MCC, (d) NPV, and (e) Precision”



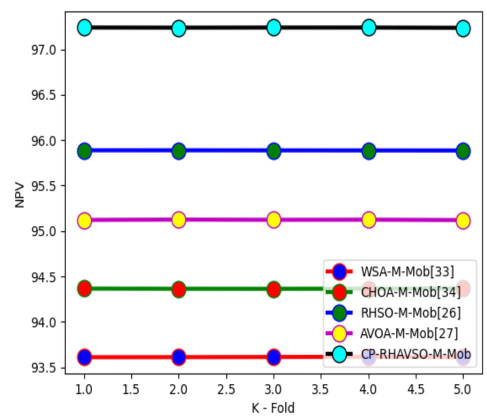
(a)



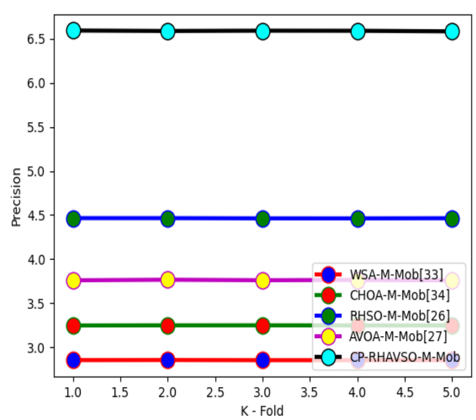
(b)



(c)

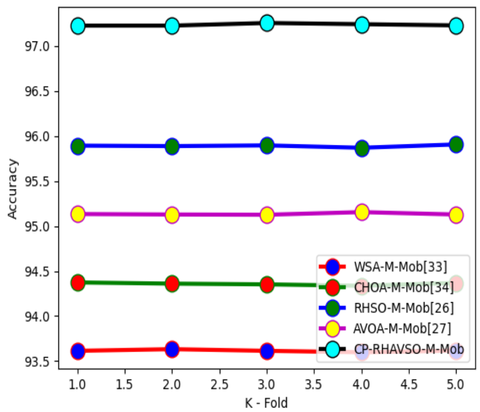


(d)

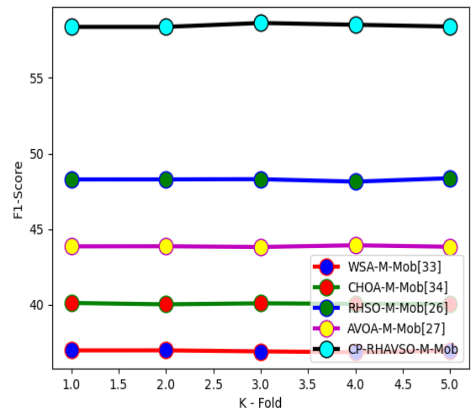


(e)

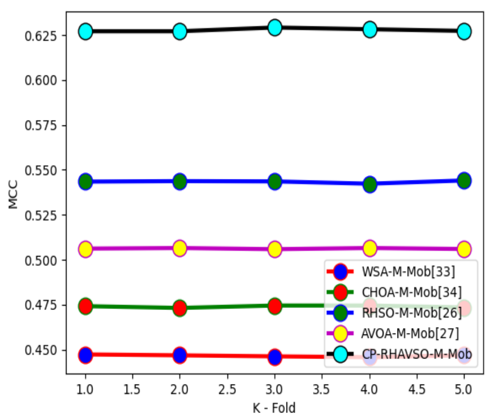
Figure 12: Inspection on the recommended face recognition model over heuristic algorithms by k-fold for dataset 1 respect to "(a) Accuracy, (b) F1-Score, (c) MCC, (d) NPV, and (e) Precision"



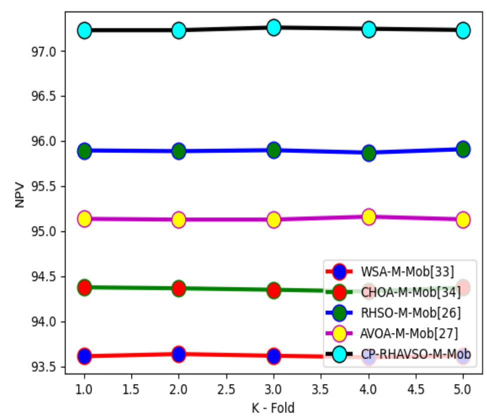
(a)



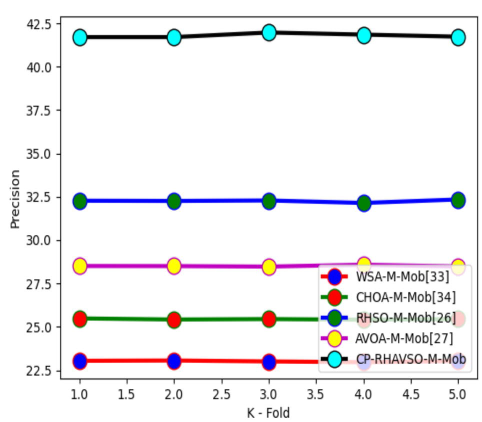
(b)



(c)

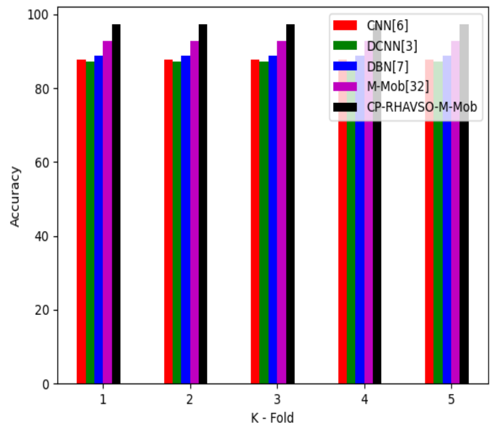


(d)

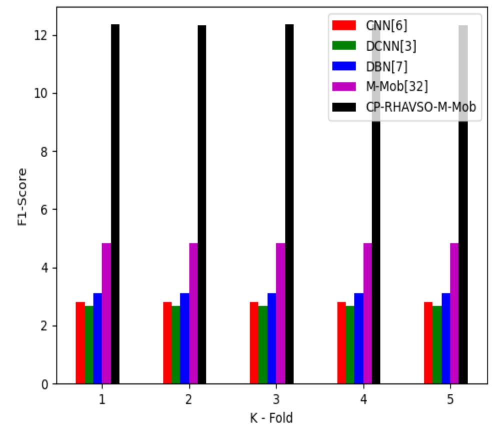


(e)

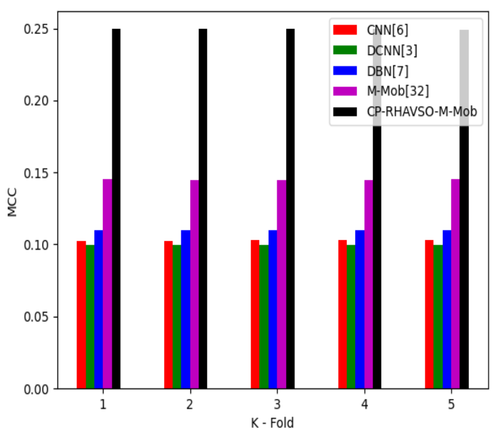
Figure 13: Inspection on the recommended face recognition model over heuristic algorithms by k-fold for dataset 2 respect to "(a) Accuracy, (b) F1-Score, (c) MCC, (d) NPV, and (e) Precision"



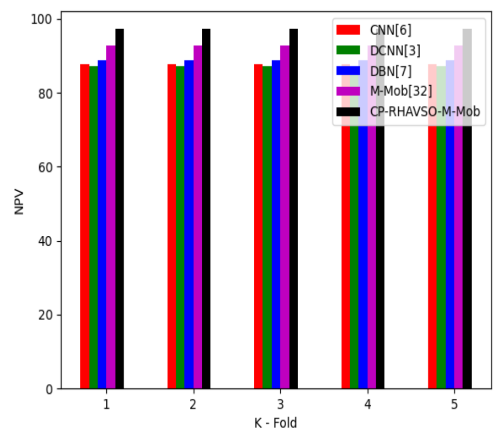
(a)



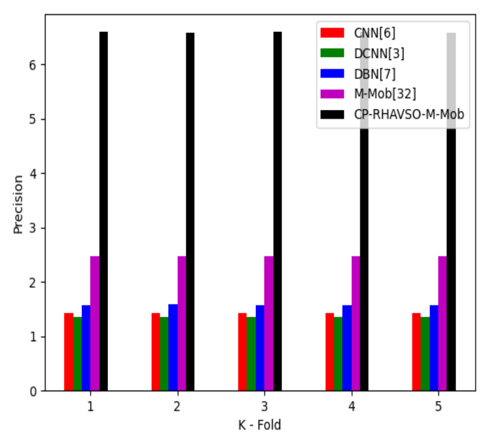
(b)



(c)

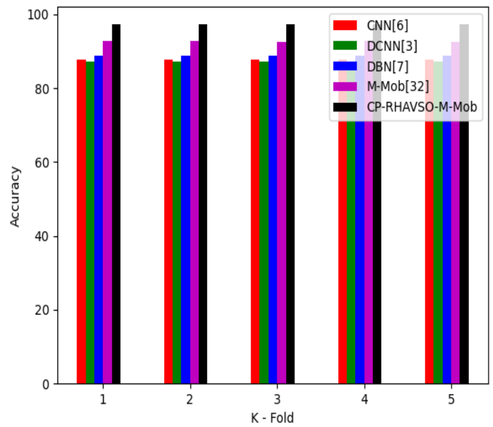


(d)

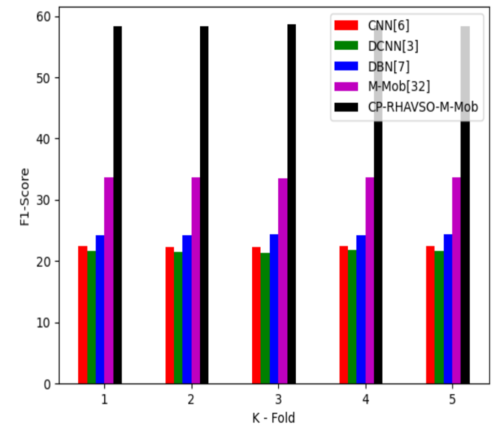


(e)

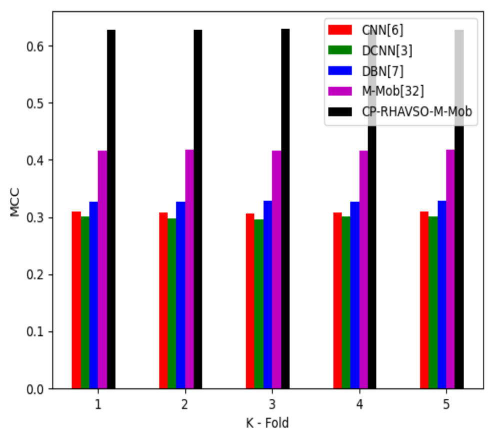
Figure 14: Examination on the designed anomaly frame classification model over deep learners for dataset 1 regarding “(a) Accuracy, (b) F1-Score, (c) FDR, (d) FNR, (e) FPR, (f) Precision, (g) MCC, (h) NPV, (i) Sensitivity and (j) Specificity”



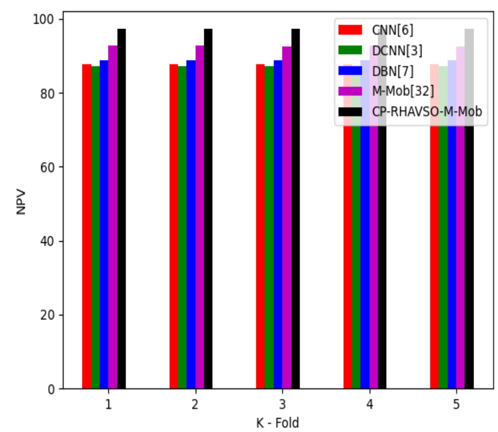
(a)



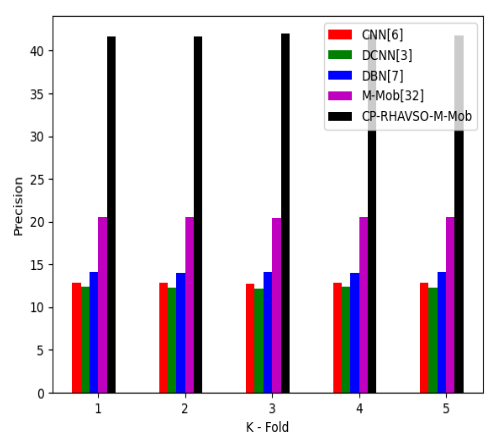
(b)



(c)

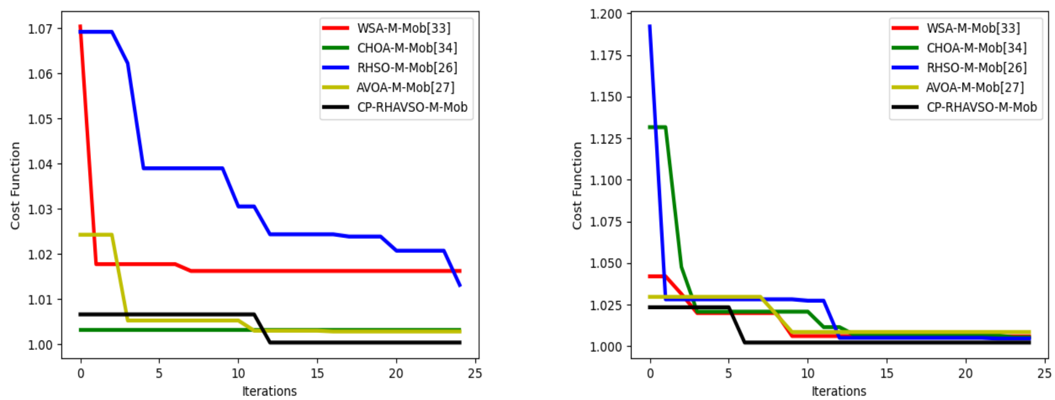


(d)



(e)

Figure 15: Inspection on the recommended face recognition model over deep learners by k-fold for dataset 2 respect to “(a) Accuracy, (b) F1-Score, (c) MCC, (d) NPV, and (e) Precision”



(a) (b)

Figure 16: Inspection on the convergence rate of the recommended face recognition model for "(a) Dataset 1, and (b) Dataset 2"

6.6 Analysis on K-Fold Over Heuristic Algorithms

The examination on the recommended face recognition framework is analyzed in terms of k-fold variations over heuristic strategies for two datasets as depicted in FIGURE 12, and FIGURE 13. The proposed CP-RHAVSO algorithm-based AM-Mob reaches higher efficiency in the face recognition framework over other deep learners. For example, the precision of the suggested CP-RHAVSO algorithm-based AM-Mob at 5 fold is 65%, 88%, 55%, and 4.6% improved in that order of WSA-M-Mob, CHOA-M-Mob, RHSO-M-Mob and AVOA-M-Mob techniques for the dataset 1. Finally, the designed strategy increases the superior performance over traditional methods.

6.7 Analysis on K-Fold Over Classifiers

The k-fold analysis on the designed face recognition model over various classifiers for two datasets are correspondingly specified in FIGURE 14, and FIGURE 15. From the analysis, the suggested CP-RHAVSO algorithm-based AM-Mob exhibits the higher effectiveness over other deep learners and also promoted the real-time efficiency of the face recognition framework.

6.8 Convergence Rate Estimation

The convergence rate of the advanced face recognition model using suggested AM-Mob using CP-RHAVSO algorithm for two datasets is given in FIGURE 16. The effectiveness in terms of convergence rate using CP-RHAVSO is examined and shown the superior performance by reaching the minimum cost function over other heuristic strategies.

Table 3: Comparative examination on the proposed face recognition framework over two datasets in estimating with traditional heurisitc algorithms

Terms	WSA-M-Mob [33]	CHOA-M-Mob [34]	RHSO-M-Mob [31]	AVOA-M-Mob [30]	CP-RHAVSO-M-Mob
Dataset 1					
“Accuracy”	91.17243	92.42334	93.35	94.02534	96.27194
“Sensitivity”	91.17143	92.47143	93.37143	94.15714	96.28571
“Specificity”	91.17243	92.42325	93.34996	94.02508	96.27192
“Precision”	2.027776	2.387424	2.736766	3.061381	4.921073
“FPR”	8.827569	7.576754	6.650043	5.974921	3.728085
“FNR”	8.828571	7.528571	6.628571	5.842857	3.714286
“NPV”	91.17243	92.42325	93.34996	94.02508	96.27192
“FDR”	97.97222	97.61258	97.26323	96.93862	95.07893
“F1-score”	3.967314	4.654675	5.317669	5.929958	9.363582
“MCC”	0.128599	0.141878	0.153656	0.16397	0.213253
Dataset 2					
“Accuracy”	91.14667	92.38133	93.36	94.016	96.27467
“Sensitivity”	91.33333	92.26667	93.06667	94.13333	96.4
“Specificity”	91.14286	92.38367	93.36599	94.01361	96.27211
“Precision”	17.38579	19.8224	22.25765	24.29456	34.54372
“FPR”	8.857143	7.616327	6.634014	5.986395	3.727891
“FNR”	8.666667	7.733333	6.933333	5.866667	3.6
“NPV”	91.14286	92.38367	93.36599	94.01361	96.27211
“FDR”	82.61421	80.1776	77.74235	75.70544	65.45628
“F1-score”	29.21109	32.63381	35.92383	38.62144	50.86177
“MCC”	0.376555	0.407865	0.437117	0.46155	0.56517

6.9 Comparative Analysis

The comparative examination on the designed face recognition and verification framework is inspected for datasets over existing frameworks as given in TABLES 3 and 4. The sensitivity of the recommended CP-RHAVSO-M-Mob is 5.6%, 4%, 3%, and 2.2% maximized in that order of WSA-M-Mob, CHOA-M-Mob, RHSO-M-Mob, and AVOA-M-Mob techniques for dataset 1. Similarly, for the dataset 2, the precision of the recommended CP-RHAVSO-M-Mob is 61%, 54%, 99%, and 46% enhanced in that order of CNN, DCNN, DBN, and M-Mob. Consequently the designed model has finally ensured the higher efficiency over other methods.

Table 4: Comparative examination on the proposed face recognition framework over two datasets in estimating with traditional deep learners

Terms	CNN [23]	DCNN [20]	DBN [24]	M-Mob [32]	CP-RHAVSO-M-Mob
Dataset 1					
“Accuracy”	90.51366	92.02246	92.6834	94.55623	96.27194
“Sensitivity”	90.65714	92.1	92.57143	94.64286	96.28571
“Specificity”	90.51337	92.0223	92.68362	94.55605	96.27192
“Precision”	1.879105	2.261248	2.472895	3.36667	4.921073
“FPR”	9.48663	7.977698	7.316376	5.443945	3.728085
“FNR”	9.342857	7.9	7.428571	5.357143	3.714286
“NPV”	90.51337	92.0223	92.68362	94.55605	96.27192
“FDR”	98.1209	97.73875	97.5271	96.63333	95.07893
“F1-score”	3.681893	4.414121	4.81711	6.502046	9.363582
“MCC”	0.122821	0.137395	0.144727	0.173	0.213253
Dataset 2					
“Accuracy”	90.47733	92.016	92.64	94.55733	96.27467
“Sensitivity”	90.53333	92.13333	92.8	94.66667	96.4
“Specificity”	90.47619	92.01361	92.63673	94.5551	96.27211
“Precision”	16.24791	19.05681	20.45855	26.1896	34.54372
“FPR”	9.52381	7.986395	7.363265	5.444898	3.727891
“FNR”	9.466667	7.866667	7.2	5.333333	3.6
“NPV”	90.47619	92.01361	92.63673	94.5551	96.27211
“FDR”	83.75209	80.94319	79.54145	73.8104	65.45628
“F1-score”	27.55123	31.58135	33.52601	41.0286	50.86177
“MCC”	0.360413	0.398612	0.416459	0.48233	0.56517

7. CONCLUSION

This work has focused on recognizing and verifying the faces in the acquired facial images for promoting several applications. This model has performed the pre-processing as the essential step by the median filtering approach with the aim of eradicating the noise in these images. Next, the spatial features were extracted by DeepLabV3, and the spectral features were acquired using DWT and DCT methods. Then, these features were fed to the proposed CP-RHAVSO algorithm for gathering the optimal spectral features and optimal spatial features for further process. These features were multiplied with the optimized weight with the objective of getting the weighted fused features, where the weights were tuned by the CP-RHAVSO algorithm. At last, the obtained weighted fused features were fed to the AM-Mob for recognizing the faces, where the parameters in M-Mob were optimized by CP-RHAVSO algorithm for gaining the precise and optimal outcomes. Thus, this model has recognized the facial images of persons with the higher accuracy rate. With the performance evaluation, the accuracy of the prescribed CP-RHAVSO-M-Mob is 6%, 4.6, 3.9%, and 1.8% consequently higher in that order of CNN, DCNN, DBN, and M-Mob for dataset 2. Therefore, the recommended model has ensured the better face recognition and verification results over other approaches.

References

[1] He R, Cao J, Song L, Sun Z, Tan T, et al. Adversarial Cross-Spectral Face Completion for Nir-Vis Face Recognition. IEEE Trans Pattern Anal Mach Intell. 2020;42:1025-1037.

[2] Geng X, Zhou ZH, Smith-Miles K. Individual Stable Space: An Approach to Face Recognition Under Uncontrolled Conditions. IEEE Trans Neural Netw. 2008;19:1354-1368.

[3] He R, Wu X, Sun Z, Tan T. Wasserstein CNN: Learning Invariant Features for Nir-Vis Face Recognition. IEEE Trans Pattern Anal Mach Intell. 2019;41:1761-1773.

[4] Liu F, Zhao Q, Liu X, Zeng D. Joint Face Alignment and 3D Face Reconstruction With Application to Face Recognition. IEEE Trans Pattern Anal Mach Intell. 2020;42:664-678.

[5] Li SZ, Chu R, Liao S, Zhang L. Illumination Invariant Face Recognition Using Nearinfrared Images. IEEE Trans Pattern Anal Mach Intell. 2007;29:627-639.

[6] Lu J, Liong VE, Zhou X, Zhou J. Learning Compact Binary Face Descriptor for Face Recognition. IEEE Trans Pattern Anal Mach Intell. 2015;37:2041-2056.

[7] Galbally J, Marcel S, Fierrez J. Image Quality Assessment for Fake Biometric Detection: Application to Iris, Fingerprint, and Face Recognition. IEEE Trans Image Process. 2014;23:710-724.

[8] Fu C, Wu X, Hu Y, Huang H, He R, et al. Dvg-Face: Dual Variational Generation for Heterogeneous Face Recognition. IEEE Trans Pattern Anal Mach Intell. 2022;44:2938-2952.

[9] Ding C, Xu C, Tao D. Multi-Task Pose-Invariant Face Recognition. IEEE Trans Image Process. 2015;24:980-993.

[10] Huang Z, Shan S, Wang R, Zhang H, Lao S, et al. A Benchmark and Comparative Study of Video-Based Face Recognition on Cox Face Database. IEEE Trans Image Process. 2015;24:5967-5981.

[11] Park U, Tong Y, Jain AK. Age-Invariant Face Recognition. IEEE Trans Pattern Anal Mach Intell. 2010;32:947-954.

[12] Chen L. A Fair Comparison Should Be Based on the Same Protocol–Comments on ”Trainable Convolution Filters and Their Application to Face Recognition”,. IEEE Trans Pattern Anal Mach Intell. 2014;36:622-623.

[13] Luo M, Cao J, Ma X, Zhang X, He R, et al. Fa-Gan: Face Augmentation Gan for Deformation-Invariant Face Recognition. IEEE Trans Inf Forensics Sec. 2021;16:2341-2355.

[14] Hwang W, Wang H, Kim H, Kee SC, Kim J, et al. Face Recognition System Using Multiple Face Model of Hybrid Fourier Feature Under Uncontrolled Illumination Variation. IEEE Trans Image Process. 2011;20:1152-1165.

[15] Zhao J, Xiong L, Li J, Xing J, Yan S, et al. 3D-Aided Dual-Agent Gans for Unconstrained Face Recognition. IEEE Trans Pattern Anal Mach Intell. 2019;41:2380-2394.

[16] Moeini A, Moeini H. Real-World and Rapid Face Recognition Toward Pose and Expression Variations via Feature Library Matrix. IEEE Trans Inf Forensics Sec. 2015;10:969-984.

[17] An Z, Deng W, Hu J, Zhong Y, Zhao Y, et al. APA: Adaptive Pose Alignment for Poseinvariant Face Recognition. IEEE Access. 2019;7:14653-14670.

[18] Garain J, Kumar RK, Kisku DR, Sanyal G. Addressing Facial Dynamics Using K-Medoids Cohort Selection Algorithm for Face Recognition. Multimedia Tool Appl. 2019;78:18443-18474.

[19] Li Y, Lu Z, Li J, Deng Yanzi. Improving Deep Learning Feature With Facial Texture Feature for Face Recognition. Wirel Personal Commun. 2018;103:1195-1206.

[20] Zhuang N, Zhang Q, Pan Cenhui, Ni B, Xu Y, et al. Recognition Oriented Facial Image Quality Assessment via Deep Convolutional Neural Network. Neurocomputing. 2019;358:109-118.

[21] Vasanthi M, Seetharaman K. Facial Image Recognition for Biometric Authentication Systems Using Acombination of Geometrical Feature Points and Low-Level Visual Features. J King Saud Univ Comput Inf Sci. 2022;34:4109-4121.

[22] Senthilkumar R, Gnanamurthy RK. A Robust Wavelet Based Decomposition of Facial Images to Improve Recognition Accuracy in Standard Appearance Based Statistical Face Recognition Methods. Clust Comput. 2019;22:12785-12794.

[23] Khan MJ, Khan MJ, Siddiqui AM, Khurshid K. An Automated and Efficient Convolutional Architecture for Disguise-Invariant Face Recognition Using Noise-Based Data Augmentation and Deep Transfer Learning. Vis Comput. 2022;38:509-523.

[24] Al-Waisy AS, Qahwaji R, Ipson S, Al-Fahdawi S. A Multimodal Deep Learning Framework Using Local Feature Representations for Face Recognition. Mach Vis Appl. 2018;29:35-54.

[25] BilelAmeur MB, Belahcene M, Masmoudi S, Ben Hamida A. Efficient Hybrid Descriptor for Face Verification in the Wild Using the Deep Learning Approach. Opt Mem Neural Networks. 2019;28:151-164.

[26] Sharma R, Patterh MS. A New Pose Invariant Face Recognition System Using PCA and Anfis. Optik. 2015;126:3483-3487.

[27] Chen L-C, Papandreou G, Schroff F, Adam H. Rethinking Atrous Convolution for Semantic Image Segmentation.2017. ARXIV PREPRINT. https://arxiv.org/pdf/1706.05587.pdf

[28] Tabassum F, Imdadul Islam Md, Tasin Khan RTasin, Amin MR. Human Face Recognition With Combination of Dwt and Machine Learning. J King Saud Univ Comput Inf Sci. 2022;34:546-556.

[29] Aswis AE, Abo-Elsoud MA. Face Recognition Based on PCA and Dct Combination Technique. Int J Eng Res Technol (IJERT). 2015;4:1295-1298.

[30] Abdollahzadeh B, Gharehchopogh FS, Mirjalili S. African Vultures Optimization Algorithm: A New Nature-Inspired Metaheuristic Algorithm for Global Optimization Problems. Comput Ind Eng. 2021;158:107408.

[31] Al-Khateeb B, Ahmed K, Mahmood M, Le D-N. Rock Hyraxes Swarm Optimization: A New Nature-Inspiredmetaheuristic Optimization Algorithm. Comput Mater Continua. 2021;68:643-654.

[32] Wang Gaihua, Yuan G, Li T, Lv M. An Multi-Scale Learning Network With Depthwise Separable Convolutions. IPSJ Trans Comput Vis Appl. 2018;10:1-8.

[33] Kaveh A, Dadras Eslamlou A. Water Strider Algorithm: A New Metaheuristic and Applications. InStructures. 2020;25:520-541.

[34] Al-Betar MA, Alyasseri ZAA, Awadallah MA, Abu Doush IA. Coronavirus Herd Immunity Optimizer (Chio). Neural Comput Appl. 2021;33:5011-5042.