Effective Face Recognition using Adaptive Multi-scale Transformer-based Resnet with Optimal Pattern Extraction

Santhosh Shivaprakash[1], Dr. Sannangi Viswaradhya Rajashekararadhya[2]

Research Scholar, Department of Electronics and Communication Engineering
Kalpataru Institute of Technology, Tiptur, Karnataka, 572201, India[1]
Professor and Former Principal, Department of Electronics and Communication Engineering
Kalpataru Institute of Technology, Tiptur, Karnataka, 572201, India[2]

Abstract—Human face is the major characteristic for identifying a person and it helps to differentiate each person. Face recognition methods are mainly useful for determining a person's identity with the help of biometric techniques. Face recognition methods are used in many practical applications like criminal identification, the phone unlocks systems and home security systems. It does not need any key and card, and it only requires facial images to provide high security over several applications. The interdependencies of the encryption methods are highly reduced in the deep learning-enabled face recognition models. Conventional methods did not satisfy the present demand due to poor recognition accuracy. Therefore, an advanced deep learning-based face recognition framework is implemented to authenticate the identity of individuals with high accuracy by using facial images. The required facial images are taken from the standard databases. The collected images are preprocessed using median filtering. The preprocessed facial images are subjected to spatial feature extraction, where the Local Binary patterns (LBP) and Local Vector Patterns (LVP) are utilized to extract the relevant optimal patterns from the facial images. Here, optimal pattern extraction is done with the Improved Rat Swarm Optimization Algorithm (IRSO). Then, the facial recognition is done over the extracted optimal features with the usage of the implemented Adaptive Multi-scale transformer-based Resnet (AMT-ResNet), where the parameters in the recognition network are optimized by using the IRSO. The efficiency of the developed deep learning adopted face recognition model is validated through different heuristics algorithms, and baseline face recognition approaches.

Keywords—*Face recognition; facial images; optimal pattern extraction rate; local binary patterns; local vector patterns; improved rat swarm optimization algorithm; adaptive multi-scale transformer-based Resnet*

I. INTRODUCTION

The individuals have been recognized with the help of emerging biometric-based methodologies in the current situation. These biometric-based techniques compute the behavioral and individual's physiological characteristics to ascertain and determine their identity [17]. It exactly identifies the individuals instead of performing people authentication and giving permission for those particular individuals to access the physical domains with the support of utilizing smart cards, plastic cards, PINs, keys, passwords, and tokens [15]. But, it is

hard to remember the PINs and passwords, and also, it is a possibility to guess and steal tokens, keys, and cards very easily [7]. In most cases, the magnetic cards of the individuals are missed, and then it can be misplaced, duplicated, purloined, and forgotten by intruders; hence, it becomes unreadable and corrupted [24]. Using biometric-based techniques, an individual's biological traits cannot be stolen, forgotten, forged, or misplaced [2]. When compared to other biometric approaches, face detection is one of the fastest biometric approaches and the least intrusive technique that includes iris recognition, fingerprint recognition, and face recognition [16]. Consider an example, instead of placing an individual's hands in a reader in surveillance systems, the people are requested to place their eyes in front of a scanner, which technique is generally known as iris recognition [30]. But, the face recognition systems unobtrusively capture images of individuals' faces when they are reached in a predefined area. In these face recognition systems, there is no capture delay and intrusion, and the subjects in the face recognition systems are entirely unaware of the process [18].

Face recognition approaches are accomplished several substantial attentions from the market and research communities that are the capability of dealing with real-world problems related to facial images [33]. In general, the face recognition approach is formulated as follows: from a given image or video of a scene, a person is verified and identified from a stored database of faces [29]. Before the detection process, the verification and identification were made by taking the similarity measure among two face images and calculating non-match or match pixel coefficients [6]. Moreover, during the identification process, the similarity measure between the collected face image and face images in the original database is examined to reduce the misclassification to determine the person's identity [21]. Face recognition is natural, convenient, and nonintrusive then it is helpful in a wide range of applications. But, several face recognition approaches encounter great challenges when recognizing the person's identity from spatial images due to the variations in illumination, background clutter, facial pose, image resolutions, and expressions [25]. Then, several artificial intelligence-based face recognition approaches are introduced that have accomplished great success over face recognition.

Deep learning techniques are assumed as the greatest breakthrough over face image recognition because of their powerful learning capacity. The hand-crafted features and the discriminative features are effectively learned through deep learning techniques [13]. The deep face-related face recognition techniques improved the recognition accuracy, but it slightly suffered from degradation in performance when encountering the large-scale dataset. To support effective face recognition in real-world scenarios, fine-tuning the pre-trained models are important [38]. Moreover, the labeling of large-scale datasets is difficult while using these deep learning-based face recognition approaches [37]. Furthermore, the existing deep learning-based techniques are well-suitable for limited modalities, but the complementary features present in the much more modalities are not well descriptive [9]. Therefore, an efficient face recognition model with the support of deep learning techniques has been designed to provide extensive results over face recognition. The main contribution of the developed deep intelligence-based face recognition framework is described below.

- To design an intelligent deep learning-based face recognition framework with the help of a heuristic strategy to identify the person's identity using their face and boosting security.

- To implement an IRSO in the developed face recognition model for the selection of optimal patterns and optimizing the hidden neuron count in adaptive multi-scale transformer-based ResNet for increasing the recognition accuracy of the implemented framework.

- To introduce an effective optimal pattern extraction approach from the LBP and LVP patterns by using the developed IRSO in the developed face recognition model to enhance the performance of the developed framework.

- To present an adaptive multi-scale transformer-based ResNet structure with parameter optimization via the developed IRSO for recognizing the faces very sensitively that is helpful to attain a higher recognition rate under varying illumination conditions, expressions, pose, and background.

- To validate the efficacy of the newly suggested face recognition model by comparing the experimental results among several optimization strategies and baseline face recognition approaches following different positive and negative measures.

The remaining sections used to develop the new deep learning-based face recognition approach are explained as follows. Section II gives the recently implemented deep learning-based face recognition approaches with their merits as well as demerits. Section III illustrates the structural description of the newly implemented face recognition model, the function of median filtering in the developed model, and the description of the database. Section IV enumerates the optimal solution obtained using the developed IRSO and the LBP and LVP pattern extraction process. Section V gives a detailed explanation of the developed multi-scale transformer-based ResNet for the detection of faces and its objective

function. The comparative analysis and the experimental results are given in Section VI, and the conclusion of the face recognition approach is summarized in Section VII.

II. LITERATURE SURVEY

A. Related Works

In 2021, Michael *et al.* [12] facial expression identification approach with the help of deep learning algorithm, which utilized Convolutional Neural Network (CNN) for learning the parameters from the face images. Here, the FER2013 dataset has been used to obtain the sample face images. Various patterns from the face images were retrieved from the original images, and the experimental results were compared over various approaches. The computational cost of this method has been highly reduced by using this approach. The recognition accuracy of the implemented model has been highly improved to baseline approaches.

In 2019, Ranjan *et al.* [27] have introduced a simultaneous face detection algorithm for localizing landmarks, recognizing genders, and estimating poses using deep CNN. Here, the multi-task learning algorithm has been implemented to fuse the intermediary layers to provide the fused features in deep CNN from the face images. Hence, the synergy among the tasks has been greatly exploited, and it improved the performance individually. Moreover, the HyperFace-ResNet and Fast-HyperFace network has been demonstrated to enhance the speed of the face identification process. The implementation results of the developed model have the ability to capture the local and global information in faces, and it provided very effective results than the other approaches.

In 2018, Abhijay *et al.* [1] have introduced a robust and efficient method to recognize individual faces in real-time. The filters have been applied to the collected sample images to eliminate unwanted features and noises. From the filtered images, the binary patterns were retrieved, and the resultant patterns were given to the Multilayer Perceptron (MLP) for recognition purposes. The extensive test results were compared over the benchmark datasets like FACES96, FACES95, and FACES94 in terms of expression, rotation, pose, illumination, and head scale. The test results gave better results and showed the developed face recognition model to be more highly efficient than the benchmark datasets.

In 2022, Junaid *et al.* [23] have studied the performance of the recently implemented deep learning-based Disguise Invariant Face Recognition (DIFR) method, which integrated the data augmentation method for removing noises. The individual faces were effectively recognized by the Viola Jones face detector and the classification was done with the support of CNN. The disguise-invariant features have been effectively learned from the facial images for identifying the person from the face images. Finally, the comparative analysis and the comprehensive experiments have been conducted over six different datasets, from that the Resnet-18 has given a good trade-off than the other face recognition approaches in terms of efficiency, accuracy and average execution time.

In 2019, Zhao *et al.* [19] have introduced a face identification method via the Generative Adversarial Network (GAN) with a dual agent system, which identified the

unlabeled patterns from the face images, and the fine details of the images were highly preserved by using this model during the realism refinement. The real and face identities were effectively determined by using this dual-agent system. Moreover, the profile face images have been generated with the support of off-the-shelf 3D face models with varying poses. By using the fully convolution approach high-resolution images have been generated. The poses were sensitively preserved, and the stability of the mechanism has been improved by analyzing the experimental results over conventional face recognition datasets.

In 2018, Al-Waisy *et al.* [3] have developed a Deep Belief Network (DBN) for extracting local handcrafted features to address face recognition problems, which has been highly suitable for unconstrained conditions. Initially, the Fractal dimension was merged with the Curvelet transform to represent the main structure of the face that included both curves and edges. Then, the pixel intensity representations were replaced with the local feature representations with the help of the DBN structure for the multi-modal recognition of facial images. At last, the effectiveness of the developed DBN-based facial recognition approach was evaluated by performing several extensive experiments over different datasets such as LFW databases, CAS-PEAL-R1, FERET, and SDUMLA-HMT. The developed DBN-based face recognition approach outperformed well than the other baseline frameworks with respect to recognition accuracy and time complexity.

In 2022, Durga and Rajesh [5] have offered a deep micro-facial emotion recognition approach with the support of CNN. Here, the 2D-ResNet CNN model has been employed to extract the Multi-class features from the collected face images. Moreover, the developed approach efficiently detected the maskable images from the face images via the 2D-ResNet CNN. This developed approach has been evaluated in Python 3.7.0 software tool, and the 2D-ResNet CNN network has been trained using the public dataset. Furthermore, the jsonify python, fastai, mysql, and flask packages were highly feasible and compatible in the developed architecture. The extensive outcome was shown that the developed approach had proven its efficiency concerning the performance metrics like sensitivity, accuracy, F1-score, and recall.

In 2015, Gao *et al.* [28] have proposed a supervised auto encoder for the recognition of individuals from the collected face images. The auto encoder was a type of new deep architecture, and it effectively learned the hidden patterns from the raw images. Initially, the mapping operation was performed between the canonical face of the person and the face variants from the images. An example of this mapping operation was the integration of normal illumination into neutral expressions. Then, the auto encoder retrieved the features that were robust to the variances of the pose, occlusion, expression, and illumination. After, the deep architecture was obtained over the supervised auto encoders, which have been utilized for the extraction of relevant features in image representation. Several experiments have been conducted over the newly implemented auto encoder-based face recognition approach, which has considered the benchmark datasets like Multi-PIE, Extended Yale B, AR, and CMU-PIE to ensure the efficacy of the developed face identification approach. The test results have shown that the stacked supervised auto encoder-based face representation has outperformed well than the other benchmark datasets.

B. Problem Statement

The face recognition approaches are mainly adopted for numerous applications like internet communication, surveillance, and security. The conventional approaches over face recognition provide satisfactory performance only under controlled scenarios. But, in real-world scenarios, the performance of the model is highly degraded under several conditions like occlusion, expressions, pose variations, and illuminations. Therefore, several supervised learning-based face recognition models are employed to provide better results over face recognition. The features and drawbacks of the developed deep learning-enabled face detection approaches are given in Table I. CNN [12] provides better results over the interpretation of facial expressions and facial recognition. And also, it highly decreases the computational cost due to the development of facial detection. But, it slightly suffers with generalization capability. In addition, it increases the instability and internal covariance, and hence the overfitting problem cannot be avoided. DCNN [27] has the capability to capture both local information and global information in faces. Further, the multi-task information is efficiently reduced to enhance the effectiveness of the recognition. Yet, the topological changes due to the viewpoint variations are not captured effectively in this approach. Consequently, it does not identify the discriminative facial information at various illumination conditions. Viola Jones detector with CNN [1] gives high efficiency and robustness in face recognition. Moreover, it captures the most crucial information from the facial images, and hence, the accuracy of the recognition process is increased. Yet, it is computationally expensive and impractical in darkness condition. Nevertheless, it is a very time-consuming process. In 2D-CNN [23], the training loss that occurred in the systems is highly reduced. Furthermore, it is highly robust for noise reduction. But, it does not meet the real-time requirements. In addition, the availability and feasibility of the model are low. GAN [3] needs less training data without the need for any domain information. Moreover, the resolution of the images is getting increased. Yet, it does not solve the generalized Eigenvalue problems. Furthermore, it does not provide any authenticity over face recognition. ResNet [5] utilizes sparse symbolic representation to symbolize the gradients in the direction domain. Consequently, the artifacts present in the model are highly reduced. Nonetheless, it does not give additional information on convergence. For instance, it needs hardware equipment to recognize face patterns. Auto encoder [28] increases the generalization ability under occlusion conditions. Subsequently, it highly reduces the low-rank error during face recognition. However, it requires more training samples to improve the scalability and availability of the face recognition system. Moreover, it has the capability to hold less amount of information. These drawbacks that arise in the deep learning-enabled face recognition approaches are resolved by the newly introduced face recognition model.

TABLE I.   FEATURES AND CHALLENGES OF THE CONVENTIONAL DEEP LEARNING-BASED FACE RECOGNITION MODELS

Author [Citation]	Methodology	Features	Challenges
Michael *et al.* [12]	CNN	• It provides better results over the interpretation of facial expressions and facial recognition. • It highly decreases the computational cost due to the implementation of facial recognition.	• It slightly suffers with generalization capability. • It increases the instability and internal covariance, and hence the overfitting problem cannot be avoided.
Ranjan *et al.* [27]	DCNN	• It has the capability to capture both local and global information in faces. • Multi-task information is efficiently reduced to enhance the performance of the recognition.	• The topological changes due to the viewpoint variations are not captured effectively in this approach. • It does not identify the discriminative facial information at various illumination conditions.
Abhijay *et al.* [1]	Viola Jones detector with CNN	• It gives high efficiency and robustness in face recognition. • It captures the most crucial information from the facial images, and hence, the accuracy of the recognition process is increased.	• It is computationally expensive and impractical in darkness conditions. • It is a very time-consuming process.
Junaid *et al.* [23]	2D CNN	• The training loss that occurred in the systems is highly reduced. • It is highly robust for noise reduction.	• It does not meet the real-time requirements. • The availability and feasibility of the model are low.
Zhao *et al.* [19]	GAN	• It needs less training data without the need for any domain information. • The resolution of the images is getting increased.	• It does not provide better results over face recognition due to the factors like various illuminations, same face, aging, and pose variations.
Al-Waisy *et al.* [3]	DBN	• It gives more accuracy and scalability over face recognition. • It effectively captures the shape variations that are irrespective of the illumination variabilities.	• It does not solve the generalized eigenvalue problems. • It does not provide any authenticity over face recognition.
Durga and Rajesh [5]	ResNet	• It utilizes sparse symbolic representation to symbolize the gradients in the direction domain. • The artifacts present in the model are highly reduced.	• It does not give additional information on convergence. • It needs hardware types of equipment for recognizing face patterns.
Gao *et al.* [28]	Auto encoder	• Under occlusion conditions, it increases the generalization ability. • It highly reduces the low-rank error during face recognition.	• It requires more training samples to improve the scalability and availability of the face recognition system. • It has the capability to hold less amount of information.

III.   ARCHITECTURAL VIEW OF THE DEVELOPED FACE RECOGNITION MODEL

A. Dataset Description of Face Recognition

The face images required for the identification of persons are collected from standard online databases. The detailed descriptions of the databases are given as below.

Dataset 1 (CFPW Dataset): This CFPW dataset is obtained from the publically available online source of "http://www.cfpw.io/ access date: 2022-12-07". This dataset contains celebrity photos from around the world. The images are collected in profile view and frontal view. More images of celebrities are added to the database.

Dataset 2 (Yale Dataset): The Yale dataset is available in the online source of "http://vision.ucsd.edu/content/yale-face-database: access date: 2022-12-07". Total 15 people's face images are there in the Yale dataset. It contains 165 images with various facial expressions of persons that includes sad, surprised, normal, wink, with or without glasses, and center-light, left-light as well as right-light.

The collected face images are indicated as FI_g, where $g = 1,2,3,...,G$, and the term G denotes the total number of sample images. The collected sample face images from two different datasets are given in Fig. 1.

Face Image Description	1	2	3	4	5
Sample face images from Dataset 1					
Sample face images from Dataset 2					

Fig. 1.   Sample face images required for the face recognition process.

B. Proposed Face Recognition Model

Face detection is one of the most difficult tasks in image retrieval and computer vision. Because the face recognition approaches are mainly applicable in various domains like driving license systems, healthcare systems, railway reservation systems, ATMs, passport authentication, and surveillance operation. While processing through the large-scale dataset, the face recognition approaches lack in their performance for system reliability, scalability, and robustness. During face recognition, the collected face image correlates with the face image from the database. But, if the query image is unavailable in the database, then it evaluates the similarity between the two images. This leads to an increase in the false identification of persons. Therefore, deep learning-based approaches are adopted for face recognition, where the error rate is highly reduced, and the recognition accuracy is improved. But, it faces several difficulties while recognizing the images in motion images, pose variations, and twins. In addition, persons with different hair colors, makeup, light intensity condition, noises, different illumination condition, hair partitions, beards, and other occlusions are difficult to identify the exact person. Hence, a new deep learning-based face recognition model is developed with the optimization algorithm to solve these disadvantages, and the architectural representation of this newly developed model is given in Fig. 2.
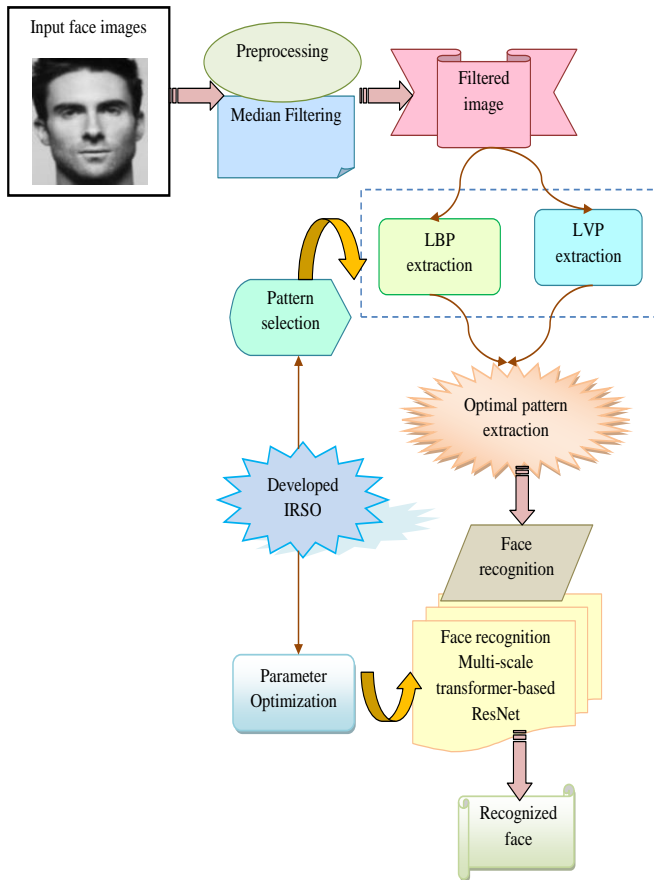

Fig. 2. Architectural illustration of the developed deep learning-based face recognition model.

A new deep learning-based face recognition model is designed with the help of a heuristic algorithm for replenishing the need for security and avoiding present-day crime. The wanted data is gathered from the CFPW and Yale databases. The collected face images are filtered through median filtering to remove the unwanted noises, then subjected to LBP and LVP pattern extraction. The optimal patterns are selected from the retrieved LBP and LVP patterns and placed in a new window. This selection process is carried out with the support of the developed IRSO, increasing the recognition accuracy during face recognition. These extracted optimal patterns are given to the efficient to perform the face recognition. Here, the hidden neuron count in the ResNet is optimized with the utilization of the implemented IRSO for improving the accuracy of the developed face recognition model. The test outcomes obtained from the newly implemented face recognition model are guaranteed with the conventional deep learning-based face recognition model and various heuristic algorithms. The ablation accuracy and convergence analysis in terms of the cost function are carried out in the developed face recognition model for validating the effectiveness of the developed IRSO--AMT-ResNet-based face recognition model.

C. Median Filtering of Input Images

The collected face images FI_g are given to the preprocessing stage for the eradication of noises and unwanted blurs present in the images. In the preprocessing stage, median filtering is used to remove the noises in the images. The median filter is generally a non-linear filtering approach, and it mainly focuses on edge preservation and noise removal over the collected images. Moreover, the errors while processing the images are highly reduced by removing noise present in the images. Fine details preservation may improve the quality of images, and hence the recognition rates are increased. The median filter [20] functioned through the window, and the selection of median intensity was carried out with the help of this window. It takes the average value of the neighborhood pixel values, and then it is named the average filer. The images are processed through pixel-by-pixel values, and the variations in the pixel values are changed with the averaged value obtained over the nearest pixels. This pixel changing related to the nearest neighbor through the window mechanism is named as sliding window mechanism. Through this sliding window, the quality of the images is getting highly increased. The acquired images FI_g are randomly represented as a group of variables $FI_g = FI_1, FI_2, FI_3, ..., FI_G$, and the median value of this image is evaluated using Eq. (1).

$$Med\left(FI_g\right) = \begin{cases} FI_{T+1} = FI_b, & G = 2T+1 \\ \dfrac{1}{2}\left(FI_{T+1} + FI_T\right), & G = 2T \end{cases} \tag{1}$$

Here, $b = FI_{T+1}$ represents the median rank and the intensity value of the face images are determined by using the 2-D median filter that is given in Eq. (2).

$$Zt_{mh,nh} = \underset{(bn,dn)\in\zeta}{Med}\left(FI_{nh+bn,nh+dn}\right) \tag{2}$$

Here, the sliding window in this approach is indicated as ζ. The finally obtained filtered facial images are represented by the term FI_g^{Mf}, and this high quality with noise-removed images is given to the input of the next level of face recognition approach.

IV. ENHANCED GANNET OPTIMIZATION ALGORITHM-BASED OPTIMAL PATTERN EXTRACTION FOR FACE RECOGNITION

A. Improved RSO

The suggested IRSO algorithm is used in the newly implemented deep intelligence-based face recognition approach, where the optimal pixels from the LVP and LBP pattern extraction approach are selected using the designed IRSO for maximizing the accuracy of the feature extraction process. Moreover, the hidden neuron count in ResNet is optimized with the support of developed IRSO that helps to improve the recognition accuracy of the designed IRSO-AMT-ResNet-based face recognition model. The RSO algorithm is used in the developed face recognition model because it avoids the local optima problem very effectively, and it highly balances the exploration as well as exploitation phases. Moreover, the computational complexity of this RSO is very less than other algorithms. But, real-world problems are not effectively handled in the RSO. Hence, to solve real-world problems and get best optimal solution in the search space, an improved RSO algorithm is developed. The developed IRSO algorithm has functioned through the parameter M. In the conventional algorithm, the parameter M is evaluated using Eq. (5). But, in the developed IRSO, the parameter M is evaluated using the newly introduced concept that is defined in Eq. (3).

$$M = 2 * \sqrt{\frac{((k*h)*l)}{IT_{MX}}} \quad (3)$$

Here, the terms k and h represent the random parameters, in which the values are selected at $k=1$ and $h=5$ to get best optimal solution. Furthermore, the term l denotes the iteration count and the term IT_{MX} indicates the maximum number of iterations. By using this modified concept, the best optimal solution is achieved in search space.

RSO: The RSO is inspired by the attacking behaviour as well as the chasing behaviour of the prey. Totally, two different species like, brown and black are there. The female rats and the male rats are named does and bucks. Generally, rats are socially intelligent based on nature. They are involved with various activities like boxing, tumbling, chasing, and jumping. The rats are characterized as the territorial animal that is live with a group of both female rats and male rats. In many cases, the character of rats is very argumentative, which results in the death of some animals. Hence, the RSO algorithm is designed based on the aggressive behaviour of the prey chasing and fighting, and it performs optimization in search space.

Rats chase the prey in a group based on social agnostic behaviour because rats are social animals. The best search agent in the group has knowledge about the position of the target prey. The new position of the search agents is upgraded in regards to the best search agent in the search space. The updated position of the best search agent is given in below Eq. (4).

$$\vec{S} = M \cdot \vec{S}_t(r) + N \cdot (\vec{S}_b(r) - \vec{S}_t(r)) \quad (4)$$

The position of the rat is denoted by the term $\vec{S}_t(r)$ and $\vec{S}_b(r)$ denotes the best optimal solution of the search agent. The formula utilized to estimate the parameters M and N is represented in Eq. (5) and Eq. (6), respectively.

$$M = L - r \times \left(\frac{L}{IT_{MX}}\right) \quad where, \ r = 0,1,2,...,IT_{MX} \quad (5)$$

$$N = 2.rd(\ ) \quad (6)$$

The parameters N and L are responsible for providing better exploitation and exploration in overall course of iterations, where the parameter L is chosen in the interval of $[1,5]$ and the term $rd(\ )$ denotes the random number and it is selected in the interval between $[0,2]$.

The prey fighting behavior of the rats in search space is given in Eq. (7).

$$\vec{S}_t(r+1) = |\vec{S}_b(r) - \vec{S}| \quad (7)$$

Here, the updated next position of the rats is given in $\vec{S}_t(r+1)$. The best solution is saved, and the best position of all other search agents is upgraded. The best exploration and exploitation are ensured by using this RSO. Algorithm 1 represents the pseudocode of the developed IRSO.

Algorithm 1: Designed IRSO
Initialize the population of RSO
Set the parameters N, L and M
Estimate the fitness value of the search agent
For $m = 1 \ to \ S_{pp}$
For $n = 1 \ to \ IT_{MX}$
Assume $S_r \leftarrow$ the best search agent
While $(r < IT_{MX})$
For all search agents
Update the parameter M with a newly developed concept in Eq. (3)
Update the current search agent position by Eq. (7)
End For
Update the parameters N, and L
Find any of the search agent goes beyond the search space
End For
End For
$r = r+1$
End while
Go back to the best optimal solution
End procedure

B. LBP and LVP Pattern

The input to be given for obtaining LBP and LVP from the preprocessed image is FI_g^{Mf} . The facial images are generally comprised of a composition of micro-patterns that are highly invariant regarding the monotonic transformations in grayscale. Hence, the LBP and LVP pattern retrieval from the facial images are required for recognizing the person with high efficiency.

The LBP [24] operator uses the thresholding approach to label the pixels of an image, where the thresholding operation is performed over the 3×3 nearest of all pixels in accordance with the center value, and the resultant patterns are obtained from the face images are binary numbers. Then, the binary numbers are described as a label presented in the histogram, which can be utilized as a texture descriptor. After completing the thresholding over 3×3 the window, then the window size is extended to use the neighborhood pixels. The bilinear interpolation among circular neighborhoods is used to allow any number of pixels and radius among the neighborhood. The neighborhood pixels are represented as (U, V) , where U defines the sampling points, and the term V indicates the radius of the circle. The LBP is known as uniform pattern, because it contains two most transition from 0 to 1 while taking the binary string in the form of a circular. The example of uniform patterns is the 00000000, 00010010 and 10000011. The histogram of the labeled image $FI_g^{Mf}(m, n)$ can be defined in Eq. (8).

$$HS_p = \sum_{m,n} IN\left\{FI_g^{Mf}(m,n) = p\right\}, p = 0,\ldots, l-1 \tag{8}$$

Here, the term l denotes the number of different labels which can be created by the LBP operator. The distribution of the local micro pattern information is present in the histogram that includes flat areas, spots, and edges. Hence, the whole image is divided into more number of regions $Rn_0, Rn_1, \ldots, Rn_{k-1}$. The spatially enhanced histogram is defined below in Eq. (9).

$$HS_{p,q} = \sum_{m,n} IN\left\{FI_g^{Mf}(m,n)IN(m,n) \in Rn_q\right\},$$
$$p = 0,1,\ldots, l-1, q = 0,1,\ldots, k-1 \tag{9}$$

Three different levels of locality can be obtained from this enhanced histogram. Moreover, the pixel-level pattern information is contained in the labels of the histogram. The regional-level information is obtained by summing the small regions, and the global description of the face is obtained by combining the regional information from the histogram.

When the whole image is divided into more number of regions, it may await that some region contains more information to differentiate peoples from the extracted patterns. Hence, similarity measure between the patterns are estimated by using the below mentioned formula in Eq. (10).

$$\vartheta^2(A, B) = \sum_p \frac{(A_p - B_p)^2}{A_p + B_p} \tag{10}$$

This is the formula to calculate the similarity measure and it can be extended to the spatially enhanced histogram by adding over the patterns p and q . Finally obtained LBP patterns from the face images are indicated by LBP_h^{Hs} .

LVP [35] is used to define the local texture in structural information and one dimensional direction by estimating the patterns between the adjacent pixels and referenced pixels in accordance with varying distance and in different directions. The input facial images to be given for extracting the LVP patterns are FI_g^{Mf} . The local sub region of the FI_g^{Mf} images are indicated by the term LS , the direction value of a vector is denoted by $R_{\alpha,Dt}(E_z)$, where the term E_z represents the referenced pixel in the local sub region LS , the distance between the adjacent pixel and the referenced pixel are denoted by the term Dt and the index angle for varying direction is represented by α . The direction value of the vector is estimated using Eq. (11).

$$R_{\alpha,Dt}(E_z) = \left(LS(E_{\alpha,Dt}) - LS(E_z)\right) \tag{11}$$

The LVP is considered as in the α direction of a vector at E_z and the encoding operation of $LVP_\alpha(E_z)$ are given in Eq. (12).

$$LVP_\alpha(E_z) = \sum_{k=1}^{K} g\left(R_{\varpi,Dt}(E_k), R_{\varpi,Dt}(E_z)\right)$$
$$\times 2^{k-1} \big| \varpi \in \left\{\alpha, \alpha + 45^\circ\right\}, K = 8 \tag{12}$$

The mathematical formulation of $g(\cdot, \cdot)$ is indicated in Eq. (13).

$$g\left(R_{\varpi,Dt}(E_k), R_{\varpi,Dt}(E_z)\right) \times 2^{k-1}$$

$$\big| \varpi \in \left\{\alpha, \alpha + 45^\circ\right\}, K = 8 = \begin{cases} 1, & \begin{array}{c} if\ R_{\alpha+45^\circ,Dt}(E_k, H) \\ -\left(\dfrac{R_{\alpha+45^\circ,Dt}(E_z)}{R_{\alpha,Dt}(E_z)}\right. \\ \left.\times R_{\alpha,Dt}(E_{k,H})\right) \end{array} \geq 0 \\ 0, & else \end{cases} \tag{13}$$

Finally, the LVP is obtained by integrating the four 8-bit binary patterns of LVPs that is illustrated in Eq. (14).

$$LVP(E_z) = \left\{LVP_\alpha(E_z) \big| \alpha = 0^\circ, 45^\circ, 90^\circ, 135^\circ\right\} \tag{14}$$

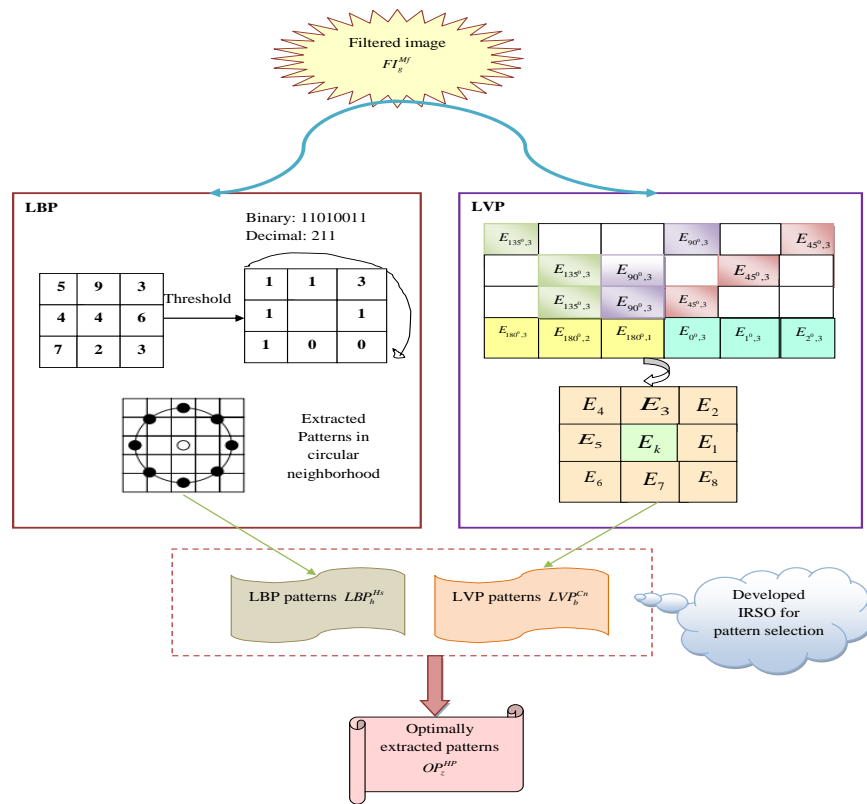The resultant LVP patterns from the preprocessed facial images are indicated by the term LVP_b^{Cn} .

Fig. 3. Optimal pattern selection using developed IRSO from LBP and LVP.

C. Optimal Pattern Extraction

The LBP pattern LBP_h^{Hs} and the LVP pattern LVP_b^{Cn} are extracted from the preprocessed facial images. The size of the LBP pattern image is considered as 128×128, and also the size of the LVP pattern image is considered in the same size of 128×128. Hence, the total pixel of the image should be 16384. From these LBP and LVP patterns, the optimal spatial features are chosen with the help of developed IRSO. The best spatial values are arranged in the new matrix to get a hybrid pattern. The selection of optimal pixel is made with new criteria that is, if the selected pixel number is 1, then get the pixels from LBP output or else, get pixels from the LVP output. At last, a new set of patterns are obtained from the LBP and LVP patterns. The optimally selected patterns, with the help of the developed IRSO are indicated by the term OP_z^{HP}, which is given to the input of the face recognition process. The optimal pattern extraction from LBP and LVP by utilizing the developed IRSO is depicted in Fig. 3.

V. AN ADVANCED FACE RECOGNITION USING ADAPTIVE MULTI-SCALE TRANSFORMER-BASED RESNET

A. Multi-scale Transformer-based Resnet

The AMT-ResNet is used in the developed face recognition framework to extract powerful discriminative features from the face images. The learning capacity of the CNN networks is high, and then the ResNet structure is utilized in the developed face recognition model. The input to be passed through the AMT-ResNet model is the resultant optimally extracted patterns OP_z^{HP}.

In AMT-ResNet [8], the encoder and decoder functions are used to detect the changes in the pixels and provide accurate recognition of face images. The encoding and decoding operation may also provide a good trade-off between the effectiveness and efficiency of the developed model. Then the feature-level fusion is also carried out over the query image and the database image. The local receptive field that arises in the network is effectively resolved by the ResNet. The multi-scale transformer divides the whole image into more number of non-overlapping regions that are represented as $P_1, P_2, ..., P_n$ and then classifies the pixel points in the facial images very precisely. The loss of information is effectively reduced by using this AMT-ResNet. Moreover, the transformer can find the local receptive fields after extracting features from the face images by using the ResNet model. The ResNet model used in the multi-scale transformer may improve the scalability of the developed face recognition framework. The extracted patterns OP_z^{HP} are given to the transformer, and then the learnable parameter encoding is performed with absolute or relative information about the images.

The attention fusion of the face images are represented by the term Y, and it is expressed in Eq. (15), Eq. (16), Eq. (17), and Eq. (18), respectively.

$$Y = [h^z(P^z) \| h^x(P^x)$$

(15)

$$Ot = Y + MS(Ln(Y)) \qquad (16)$$

$$Ot = \left[Ot^z \big\| Ot^x \right] \qquad (17)$$

$$J^z = k^z \left(Ot^z \right) \qquad (18)$$

Here, the term P^z denotes the pixel points in the given image. Moreover, the terms $h^z(\cdot)$ and $k^z(\cdot)$ represents the projection as well as the back projection functions that is used to align the dimension of the image.

The pairwise fusion of facial images is represented in Eq. (19).

$$S^z = \left[k^z \left(\sum_{m \in \{z,x\}} k^z \left(P_{CL}^m \right) \Big\| k^z \left(\sum_{m \in \{z,x\}} k^z \left(P_{Patch}^m \right) \right) \right) \right] \qquad (19)$$

Here, the term P_{Patch}^m is the patch information and the term P_{CL}^m represents the class information of the face images. The cross attention function is given in Eq. (20).

$$CrAt = At \cdot \vartheta \qquad (20)$$

Here, the term At is the attention map and it is estimated using Eq. (21).

$$At = soft\max \left(QL^t \Big/ \sqrt{\frac{Em}{D}} \right) \qquad (21)$$

Here, the term Em denotes the embedding dimension and D gives the number of heads parameter. The values of Q, ϑ and L are estimated using Eq. (22).

$$\vartheta = P_{CL}^{'z} \psi_\vartheta, \ L = P_{CL}^{'z} \psi_L, \ Q = P_{CL}^{'z} \psi_Q \qquad (22)$$

The cross attention of the normalization layer and the residual shortcut are represented in Eq. (23) and Eq. (24), respectively.

$$Y_{CL}^z = h^z P_{CL}^z + MS \left(LN \left(h^z \left(P_{CL}^z \right) \Big\| P_{Patch}^z \right) \right) \qquad (23)$$

$$J^z = k^z \left(P_{CL}^z \right) \Big\| P_{Patch}^z \qquad (24)$$

Here, the term MS defines the multiple heads in the transformer network. The AMT-ResNet structure produces better results over the face recognition, and the accuracy is highly improved. Fig. 4 depicts the basic structure of AMT-ResNet model.

B. Adaptive Multi-scale Transformer-based Resnet

The newly developed AMT-ResNet-based face recognition approach is useful for detecting the person's face from the collected face images under variations in pose, expression, occlusion, and illumination as well as light intensity conditions. The hidden neuron count in the AMT-ResNet is optimized through the developed IRSO for improving the recognition

accuracy rate during face recognition. The Multiscale Transformer network is used in the developed face recognition model because it effectively learns the discriminative variations in the face images. It provides better-recognized results over face recognition with highly reduced recognition time and computational complexity. In order to improve the robustness and reliability of face recognition, the adaptive multi-scale transformer-based ResNet is introduced, where the hidden neuron count in the ResNet is optimized with the help of the newly implemented IRSO. The main motive of the parameter optimization in the developed AMT-ResNet is to obtain higher recognition accuracy. The objective function of the developed IRSO-AMT-ResNet-based face recognition approach is given below in Eq. (25).

$$OB = \operatorname*{arg\,min}_{\left\{ Sp_n^{LBVP}, NH_m^{ResNet} \right\}} \left(\frac{1}{AR} \right) \qquad (25)$$
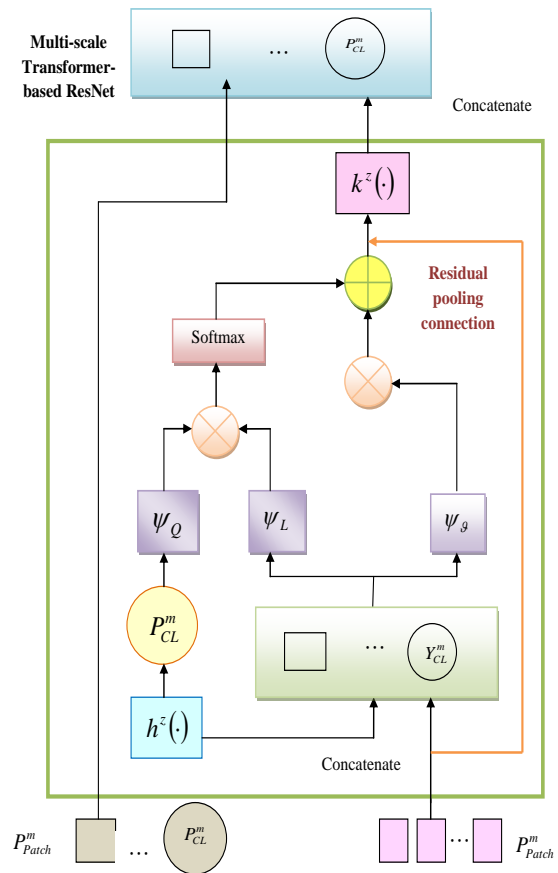

Fig. 4. Basic structure of AMT-ResNet model.

Here, the term Sp_n^{LBVP} gives the extracted patterns from LBP and LVP, and the term NH_m^{ResNet} indicates the hidden neuron count in the ResNet structure. The extracted patterns Sp_n^{LBVP} are optimized in the range of $[1,2]$ and the hidden neuron count is tuned in the range between $[50,100]$. Furthermore, the term AR represents accuracy and it is evaluated by using true and false positive as well as negative observations. The accuracy formula is given in below Eq. (26).

$$AR = \frac{R_P + R_N}{R_P + R_N + L_P + L_N} \tag{26}$$

Here, the term R_P denotes the true positive, L_P is the false positive, R_N is the true negative and L_N is the false negative observation value. The face recognition process using developed AMT-ResNet with parameter optimization is given in Fig. 5.
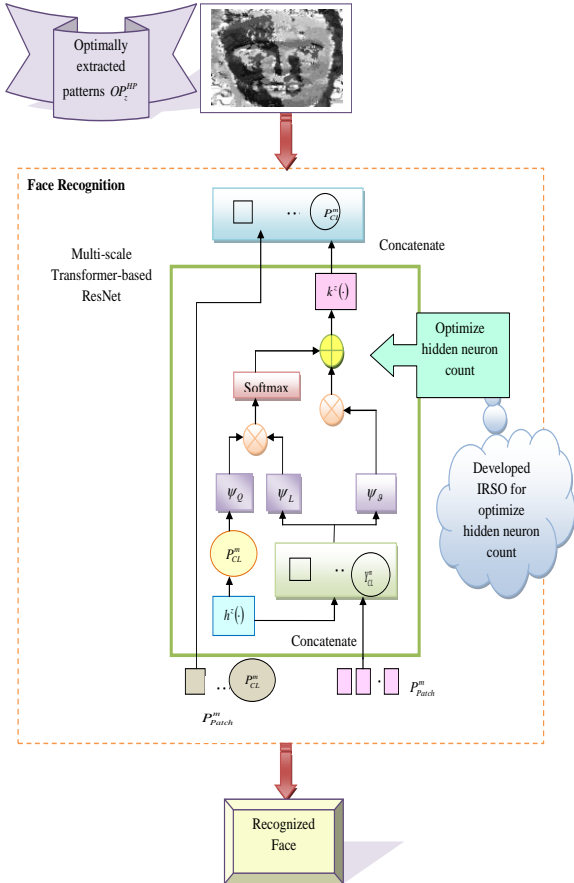

Fig. 5. AMT-ResNet-based face recognition model with parameter optimization.

VI. RESULTS AND DISCUSSIONS

A. Experimental Setup

The newly implemented IRSO-AMT-ResNet-based face recognition approach was designed in Python tool. Moreover, the implementation results of the implemented approach have been analyzed among various heuristic algorithms and conventional face recognition methodologies for validating the efficacy of the developed model. The population count and the maximum number of iterations that should be taken for performing the comparative analysis were 10 and 25, respectively.

The heuristic algorithms to be taken for performing the comparative analysis on the developed IRSO-AMT-ResNet-based face recognition approach were Cat Swarm Optimization (CSO) [32], Cuckoo Optimization Algorithm (COA) [10], Moth Flame Optimization (MFO) [31] and Rat Swarm

Optimization (RSO) [14] and the conventional face recognition approaches to be considered for analyzing the performance on the developed model was CNN [12], LSTM [11], VGG16 [4] and ResNet [5]. Various positive, as well as negative measures were used for analyzing the efficiency of the implemented face recognition model.

B. Validation Measures

The performance metrics used to validate the effectiveness of the developed face recognition model are several positive measures such as sensitivity, accuracy, specificity, precision, NPV, MCC, and F1-score, and also negative measures like FDR, FNR, and FPR. The formula used to calculate the positive and negative measures is summarized as below.

$$\text{Pr}\,ecision = \frac{R_P}{R_P + L_P} \tag{27}$$

$$F1 - score = \frac{2R_P}{2R_P + L_P + L_N} \tag{28}$$

$$Sensitivity = \frac{R_P}{R_P + L_N} \tag{29}$$

$$Specificity = \frac{R_N}{R_N + L_P} \tag{30}$$

$$FPR = \frac{L_P}{R_N + L_P} \tag{31}$$

$$FNR = \frac{L_N}{R_P + L_N} \tag{32}$$

$$FDR = \frac{R_P}{R_N + L_N} \tag{33}$$

$$MCC = \frac{R_P + R_N - L_P + L_N}{\sqrt{(R_P + L_P)(R_N + L_P)(R_P + L_N)(R_N + L_N)}} \tag{34}$$

C. Experimental Results

The resultant facial images obtained after the preprocessing, extracting LBP and LVP. The optimal pattern extraction of dataset 1 and dataset 2 are depicted in Fig. 6.

D. Ablation Study on Developed Face Recognition Model Based on the Accuracy

The accuracy of the suggested IRSO-AMT-ResNet-based face recognition approach over the extension of convolutional networks is illustrated in below Fig. 7. This comparative analysis is taken over the two datasets 1 and 2. From the analysis, the developed IRSO-AMT-ResNet-based face recognition approach gained improved accuracy of 8.235% than AlexNet, 6.97% than GoogleNet, 5.74% than ResNet 150, and 4.54 % than ResNet 152 for the learning percentage value of 65 while considering the dataset of 2. The accuracy of the developed IRSO-AMT-ResNet-based face recognition approach is slightly increased than the convolutional networks.

Dataset Description	Image Description	Original Images	Pre-processed Images	LBP Images	LVP Images	Optimal Pattern Extracted Images
Dataset 1	Image set 1					
	Image set 2					
	Image set 3					
Dataset 2	Image set 1					
	Image set 2					
	Image set 3					

Fig. 6.   The resultant facial optimal pattern extracted images from the developed face recognition model.

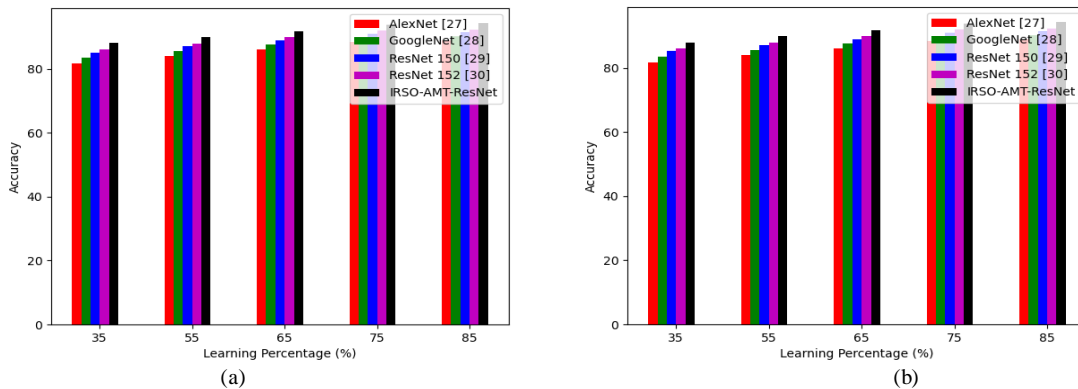(a)   (b)

Fig. 7.   Ablation study on suggested face recognition model over existing networks in regards with (a) Dataset 1 and (b) Dataset 2.

E. Convergence Evaluation on Developed Model

The following Fig. 8 shows the convergence evaluation of the developed IRSO-AMT-ResNet-based face recognition model over dataset 1 and dataset 2. This convergence analysis is carried out over various heuristic algorithms and the developed IRSO-AMT-ResNet-based face recognition model secured with improved cost function rate of 0.29% than MFO-AMT-ResNet, 0.79% than COA-AMT-ResNet, 1.37% than CSO-AMT-ResNet and 2.43% than RSO-AMT-ResNet when considering the iteration number as 20 for dataset 2. Moreover, the cost function is highly decreased in the developed face recognition model rather than the other heuristic algorithms.

F. Accuracy Evaluation on Developed Face Recognition Model

The efficiency estimation of the developed IRSO-AMT-ResNet-based face recognition framework in terms of varying learning percentage value with respect to various heuristic algorithms is given in Fig. 9, and various conventional face recognition approaches is given in Fig. 10. This comparison among heuristic algorithms and conventional approaches are carried out over dataset 1 and dataset 2. The developed model obtained improved accuracy of 5.88%, 4.65%, 2.85%, and 2.27% than the heuristic algorithms CSO-AMT-ResNet, COA-AMT-ResNet, MFO-AMT-ResNet and RSO-AMT-ResNet for dataset 1 with reference to the learning percentage value of 55.

G. K-fold Comparison on Developed Face Recognition Model

The k-fold validation of the implemented IRSO-AMT-ResNet-based face recognition model when compared to various heuristic algorithms is given in Fig. 11, and various baseline face recognition approaches are shown in Fig. 12. The developed model accomplished with improved accuracy of 15.47%, 16.86%, 14.11% and 11.49% than the heuristic algorithms CNN, LSTM, VGG16 and ResNet for dataset 1 with respect to the K-fold value of 3. The developed model secured with high accuracy than the existing face recognition approaches.
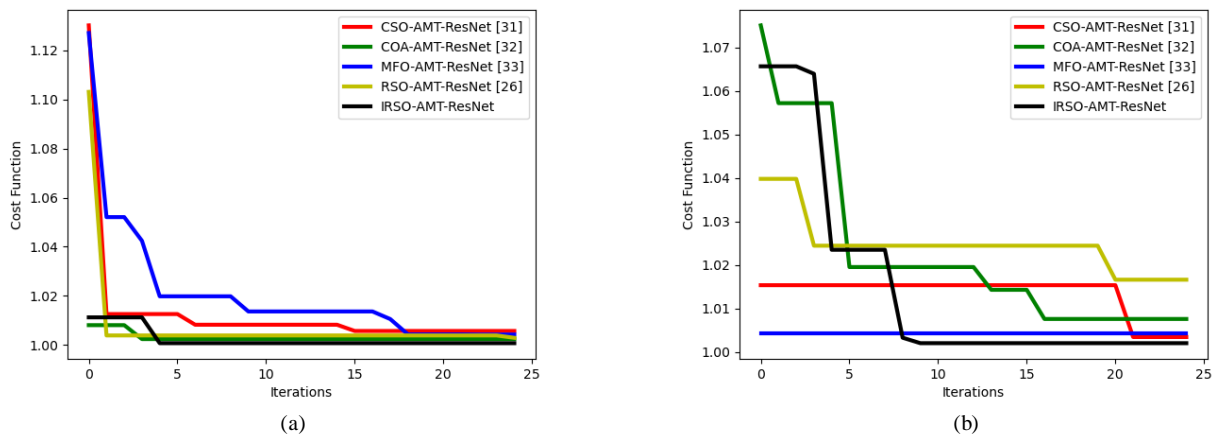

Fig. 8. Convergence evaluation of the suggested face recognition model over different heuristic algorithms in regards with (a) Dataset 1 and (b) Dataset 2.
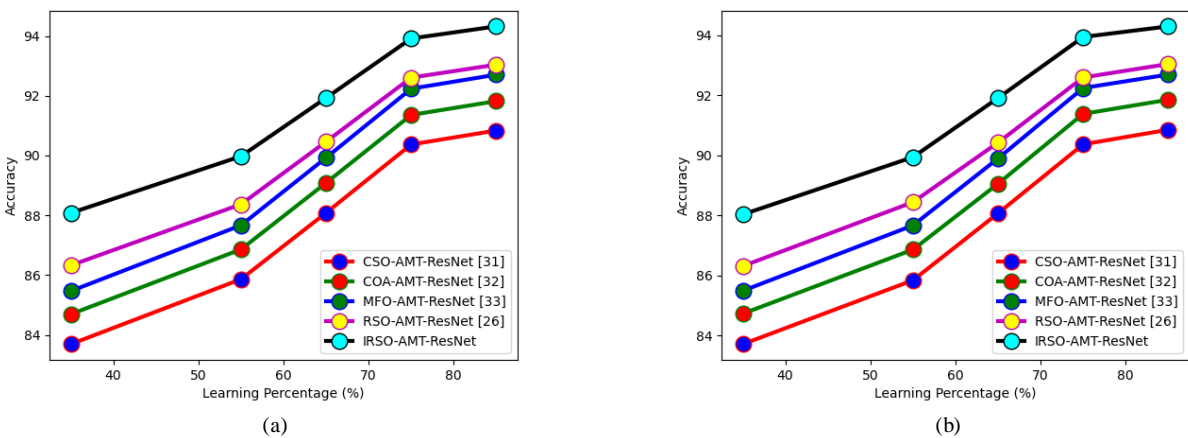

Fig. 9. Efficiency evaluation of the suggested face recognition model over different heuristic algorithms in regards with (a) Dataset 1 and (b) Dataset 2.
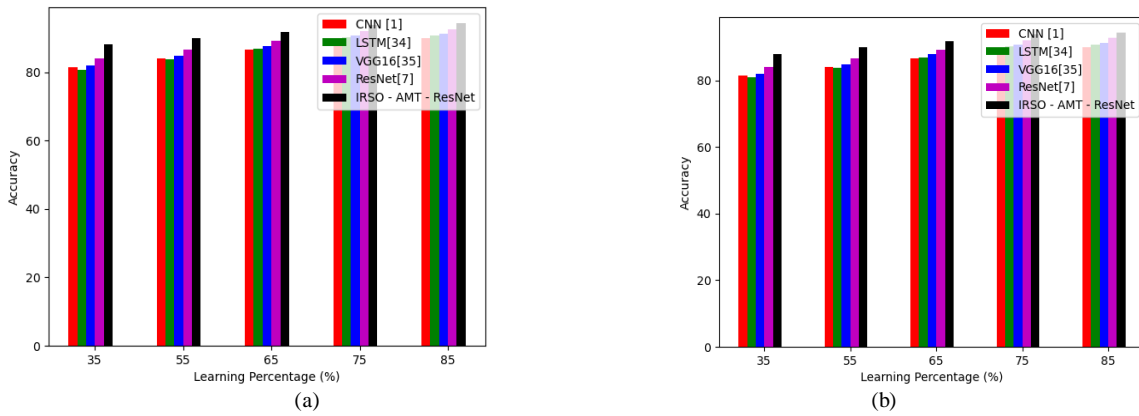
Fig. 10. Efficiency evaluation of the suggested face recognition framework over baseline models in regards with (a) Dataset 1 and (b) Dataset 2.
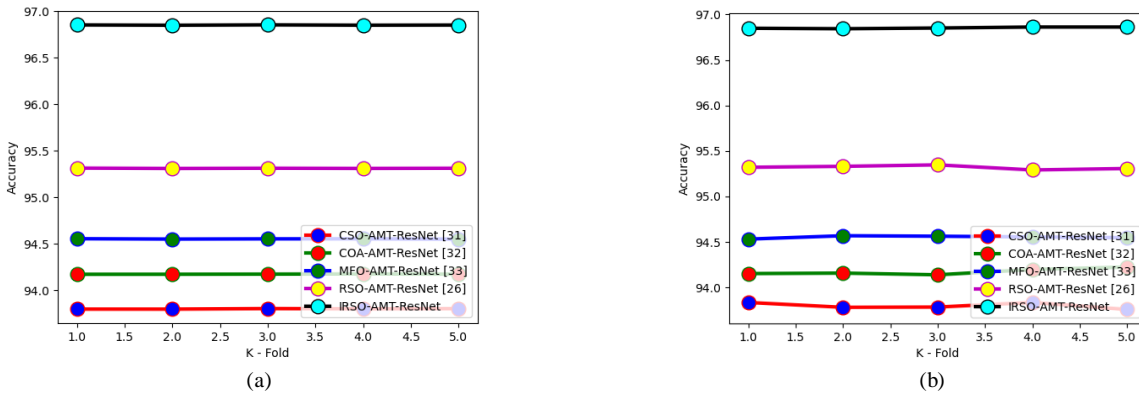

Fig. 11. K-fold estimation of the suggested face recognition framework over various heuristic algorithms in regards with (a) Dataset 1 and (b) Dataset 2.
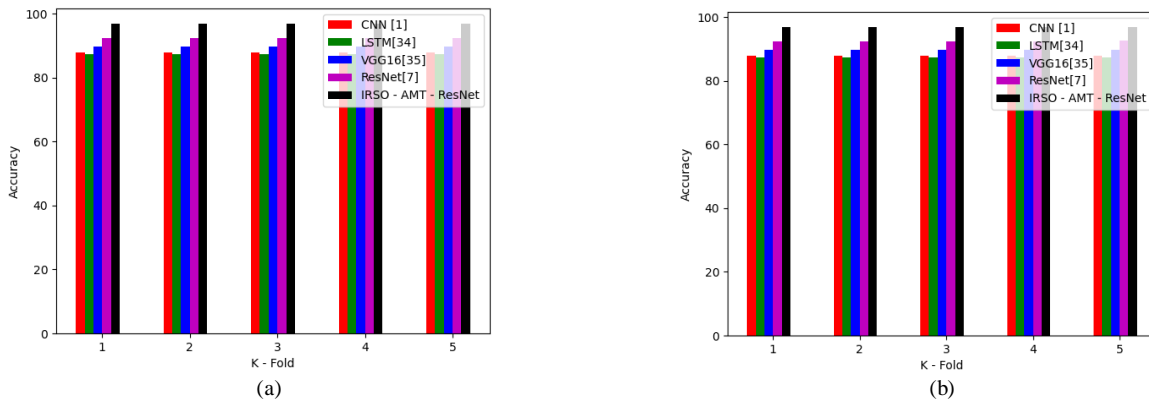

Fig. 12. K-fold estimation of the suggested face recognition framework over various baseline models in regards with (a) Dataset 1 and (b) Dataset 2.

H. Comparative Analysis of the Developed Face Recognition Model

The different baseline face recognition approaches are illustrated in Table II. The performance analysis of the implemented IRSO-AMT-ResNet-based face recognition model according to various optimization strategies is depicted in Table III. The developed model obtained with higher recognition sensitivity of 3.95%, 2.80%, 1.70%, and 1.38% than the heuristic algorithms like CSO-AMT-ResNet, COA-AMT-ResNet, MFO-AMT-ResNet and RSO-AMT-ResNet while taking the dataset 1. The recommended face recognition model achieved superior performance than the other

optimization strategies in terms of various positive and negative measures.

I. Ablation Study on Different Network Models for Face Recognition

The performance of the developed IRSO-AMT-ResNet-based face recognition model among various face recognition networks is given in below Table IV. The implemented model achieved with improved F1-score of 62.86% than AlexNet, 48.68% than GoogleNet, 33.22% than ResNet 150 and 22.49% than ResNet 152 for dataset 2. The overall performance of the developed face recognition model is greater than the convolutional networks for dataset 1 and dataset 2.

TABLE II.   EFFECTIVE EVALUATION ON DEVELOPED FACE RECOGNITION MODEL AMONG DIFFERENT OPTIMIZATION STRATEGIES

Terms	CSO-AMT-ResNet [32]	COA-AMT-ResNet [10]	MFO-AMT-ResNet [31]	RSO-AMT-ResNet [14]	IRSO-AMT-ResNet
Dataset 1					
Specificity	90.83739	91.8191	92.70444	93.03968	94.31935
Sensitivity	90.77143	91.78571	92.77143	93.07143	94.35714
FDR	98.05333	97.80104	97.515	97.39023	96.77852
Accuracy	90.83726	91.81903	92.70457	93.03974	94.31943
MCC	0.125381	0.135035	0.145283	0.149481	0.168703
FPR	9.162611	8.180905	7.295563	6.960321	5.680647
NPV	90.83739	91.8191	92.70444	93.03968	94.31935
Precision	1.946667	2.198957	2.485	2.609769	3.22148
F1-score	3.811592	4.295016	4.840345	5.077171	6.23025
FNR	9.228571	8.214286	7.228571	6.928571	5.642857
Dataset 2					
Specificity	90.85714	91.85578	92.70204	93.04762	94.29932
Sensitivity	90.8	91.86667	92.4	93.06667	94.4
FNR	9.2	8.133333	7.6	6.933333	5.6
MCC	0.368683	0.393899	0.416319	0.428333	0.472196
Accuracy	90.856	91.856	92.696	93.048	94.30133
FPR	9.142857	8.144218	7.297959	6.952381	5.70068
Precision	16.85226	18.71266	20.53333	21.45712	25.25865
NPV	90.85714	91.85578	92.70204	93.04762	94.29932
F1-score	28.4283	31.09206	33.6	34.87384	39.85364
FDR	83.14774	81.28734	79.46667	78.54288	74.74135

TABLE III.   EFFECTIVE EVALUATION ON DEVELOPED FACE RECOGNITION MODEL AMONG CONVENTIONAL APPROACHES

Terms	CNN [12]	LSTM [11]	VGG16 [4]	ResNet [5]	IRSO-AMT-ResNet
Dataset 1					
Specificity	89.9292	90.82539	91.38297	92.698	94.31935
Sensitivity	89.98571	90.74286	91.48571	92.6	94.35714
FNR	10.01429	9.257143	8.514286	7.4	5.642857
MCC	0.117813	0.125247	0.130803	0.144927	0.168703
Accuracy	89.92931	90.82523	91.38317	92.6978	94.31943
FPR	10.0708	9.174606	8.617034	7.302004	5.680647
Precision	1.759143	1.943572	2.083299	2.478388	3.22148
NPV	89.9292	90.82539	91.38297	92.698	94.31935
F1-score	3.450826	3.805632	4.07383	4.827568	6.23025
FDR	98.24086	98.05643	97.9167	97.52161	96.77852
Dataset 2					
Specificity	89.95374	90.85442	91.38776	92.6966	94.29932
Sensitivity	90.26667	90.66667	91.46667	92.93333	94.4
FNR	9.733333	9.333333	8.533333	7.066667	5.6
MCC	0.350055	0.368068	0.382122	0.418566	0.472196
Accuracy	89.96	90.85067	91.38933	92.70133	94.30133
FPR	10.04626	9.145578	8.612245	7.303401	5.70068
Precision	15.49554	16.82752	17.81355	20.6152	25.25865
NPV	89.95374	90.85442	91.38776	92.6966	94.29932
F1-score	26.45048	28.38656	29.8196	33.74486	39.85364
FDR	84.50446	83.17248	82.18645	79.3848	74.74135

TABLE IV.   ABLATION STUDY ON THE DEVELOPED FACE RECOGNITION MODEL AMONG DIFFERENT NETWORK MODELS

Terms	AlexNet [39]	GoogleNet [26]	ResNet-150[22]	ResNet-152[36]	IRSO-AMT-ResNet
Dataset 1					
Specificity	88.97924	90.19287	91.43052	92.34552	94.31935
Sensitivity	88.98571	90.38571	91.41429	92.31429	94.35714
FNR	11.01429	9.614286	8.585714	7.685714	5.642857
MCC	0.11055	0.120169	0.131087	0.140844	0.168703
Accuracy	88.97926	90.19326	91.43049	92.34546	94.31943
FPR	11.02076	9.807129	8.569482	7.65448	5.680647
Precision	1.592345	1.813466	2.093016	2.359832	3.22148
NPV	88.97924	90.19287	91.43052	92.34552	94.31935
F1-score	3.128704	3.555593	4.092335	4.602023	6.23025
FDR	98.40765	98.18653	97.90698	97.64017	96.77852
Dataset 2					
Specificity	88.98231	90.1551	91.41224	92.33469	94.29932
Sensitivity	89.2	90.13333	91.6	92.4	94.4
FNR	10.8	9.866667	8.4	7.6	5.6
Precision	14.17974	15.7429	17.87666	19.74359	25.25865
Accuracy	88.98667	90.15467	91.416	92.336	94.30133
FPR	11.01769	9.844898	8.587755	7.665306	5.70068
MCC	0.330043	0.352999	0.383203	0.407278	0.472196
NPV	88.98231	90.1551	91.41224	92.33469	94.29932
F1-score	24.46964	26.80412	29.91509	32.53521	39.85364
FDR	85.82026	84.2571	82.12334	80.25641	74.74135

VII.   CONCLUSION

A new deep intelligent-based face recognition approach has been presented to identify individuals with high recognition accuracy. This face recognition helped to identify the thieves, and it has been mainly used in forensic applications. The face images were collected from two distinct databases, and median filtering was used to filter the image to eliminate the noises. The LBP and the LVP patterns were extracted from the filtered face images, and the optimal patterns were extracted from this by utilizing the developed IRSO. The optimally selected patterns from the face images were given to the AMT-ResNet, where the hidden neuron count was optimized through the newly implemented IRSO for increasing recognition accuracy. The implementation results obtained from the developed IRSO-AMT-ResNet-based face recognition model have been validated by analyzing the results over the conventional face recognition models according to ablation accuracy and convergence analysis. The developed IRSO-AMT-ResNet-based face recognition model has attained with improved accuracy of 5.94%, 4.59%, 3.15%, and 2.12% than the convolutional networks like AlexNet, GoogleNet, ResNet 150, and ResNet 152. Furthermore, the accuracy of the developed deep intelligence-based face detection approach was highly enhanced than the conventional face recognition techniques and heuristic algorithms.

REFERENCES

[1] A Vinay, Abhijay Gupta, Aprameya Bharadwaj, Arvind SrinivasanK, N Balasubramanya Murthy, S Natarajan, "Deep Learning on Binary Patterns for Face Recognition," Procedia Computer Science, vol.132, pp. 76-83, 2018.

[2] A. R. Faizabadi, H. F. B. M. Zaki, Z. B. Z. Abidin, N. N. W. N. Hashim and M. A. B. Husman, "Efficient Region of Interest Based Metric Learning for Effective Open World Deep Face Recognition Applications," IEEE Access, vol. 10, pp. 76168-76184, 2022.

[3] Alaa S. Al-Waisy, Rami Qahwaji, Stanley Ipson & Shumoos Al-Fahdawi, "A multimodal deep learning framework using local feature representations for face recognition," Machine Vision and Applications, vol. 29, pp.35–54, 2018.

[4] Arun Kumar Dubey, Vanita Jain, "Automatic facial recognition using VGG16 based transfer learning model," Journal of Information and Optimization Sciences, vol. 41, 2020.

[5] B Kanaka Durga, V.Rajesh, "A ResNet deep learning based facial recognition design for future multimedia applications," Computers and Electrical Engineering, vol. 104, pp.108384, December 2022.

[6] B. Kocacinar, B. Tas, F. P. Akbulut, C. Catal and D. Mishra, "A Real-Time CNN-Based Lightweight Mobile Masked Face Recognition System," IEEE Access, vol. 10, pp. 63496-63507, 2022.

[7] C. Galea and R. A. Farrugia, "Matching Software-Generated Sketches to Face Photographs With a Very Deep CNN, Morphed Faces, and Transfer Learning," IEEE Transactions on Information Forensics and Security, vol. 13, no. 6, pp. 1421-1431, June 2018.

[8] Chun-Fu (Richard) Chen, Quanfu Fan, Rameswar Panda MIT-IBM Watson AI Lab, "CrossViT: Cross-Attention Multi-Scale Vision Transformer for Image Classification," 2021.

[9] D. B. Giap, T. Ngoc Le, J. -W. Wang and C. -N. Wang, "Adaptive Multiple Layer Retinex-Enabled Color Face Enhancement for Deep Learning-Based Recognition," IEEE Access, vol. 9, pp. 168216-168235, 2021.

[10] D. Chitara, K. R. Niazi, A. Swarnkar and N. Gupta, "Cuckoo Search Optimization algorithm for designing of multimachine Power System Stabilizer," IEEE 1st International Conference on Power Electronics, Intelligent Control and Energy Systems (ICPEICES), 2016.

[11] Débora C. Corrêa, Denis H. P. Salvadeo, Alexandre L. M. Levada, José H. Saito, Nelson D. A. Mascarenhas, Jander Moreira, "Using LSTM

Network in Face Classification Problems," The international jouranal, 2008.

[12] Edeh Michael Onyema, Piyush Kumar Shukla, Surjeet Dalal, Mayuri Neeraj Mathur, Mohammed Zakariah, and Basant Tiwari, "Enhancement of Patient Facial Recognition through Deep Learning Algorithm: ConvNet," Open Access, Article ID 5196000, 2021.

[13] G. Hermosilla, D. -I. H. Tapia, H. Allende-Cid, G. F. Castro and E. Vera, "Thermal Face Generation Using StyleGAN," IEEE Access, vol. 9, pp. 80511-80523, 2021.

[14] Gaurav Dhiman, Meenakshi Garg, Atulya Nagar, Vijay Kumar & Mohammad Dehghani, "A novel algorithm for global optimization: Rat Swarm Optimizer," Journal of Ambient Intelligence and Humanized Computing, vol. 12, pp.8457–8482, 2021.

[15] H. Roy, D. Bhattacharjee and O. Krejcar, "Interpretable Local Frequency Binary Pattern (LFrBP) Based Joint Continual Learning Network for Heterogeneous Face Recognition," IEEE Transactions on Information Forensics and Security, vol. 17, pp. 2125-2136, 2022.

[16] J. Deng, J. Guo, J. Yang, N. Xue, I. Kotsia and S. Zafeiriou, "ArcFace: Additive Angular Margin Loss for Deep Face Recognition," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 44, no. 10, pp. 5962-5979, 1 Oct. 2022.

[17] J. Lu, V. E. Liong, G. Wang and P. Moulin, "Joint Feature Learning for Face Recognition," IEEE Transactions on Information Forensics and Security, vol. 10, no. 7, pp. 1371-1383, July 2015.

[18] J. Y. Choi and B. Lee, "Ensemble of Deep Convolutional Neural Networks With Gabor Face Representations for Face Recognition," IEEE Transactions on Image Processing, vol. 29, pp. 3270-3281, 2020.

[19] J. Zhao, L. Xiong, J. Li, J. Xing, S. Yan and J. Feng, "3D-Aided Dual-Agent GANs for Unconstrained Face Recognition," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 41, no. 10, pp. 2380-2394, 1 Oct. 2019.

[20] Joungeun Bae and Hoon Yoo, "Fast Median Filtering by Use of Fast Localization of Median Value," International Journal of Applied Engineering Research, vol. 13, pp. 10882-10885, 2018.

[21] M. Abdelmaksoud, E. Nabil, I. Farag and H. A. Hameed, "A Novel Neural Network Method for Face Recognition With a Single Sample Per Person," IEEE Access, vol. 8, pp. 102212-102221, 2020.

[22] Mohamed Loey, Gunasekaran Manogaran, Mohamed Hamed N. Taha, and Nour Eldeen M. Khalifad, "Fighting against COVID-19: A novel deep learning model based on YOLO-v2 with ResNet-50 for medical face mask detection," vol.65, pp.102600, 2021.

[23] Muhammad Junaid Khan, Muhammad Jaleed Khan, Adil Masood Siddiqui & Khurram Khurshid," An automated and efficient convolutional architecture for disguise-invariant face recognition using noise-based data augmentation and deep transfer learning," The Visual Computer, vol. 38, pp.509–523, 2022.

[24] N. A. A. -H. And and M. Prince, "A Classification of Arab Ethnicity Based on Face Image Using Deep Learning Approach," IEEE Access, vol. 9, pp. 50755-50766, 2021.

[25] Neha Soni, Enakshi Khular Sharma, Amita Kapoor, "Hybrid meta-heuristic algorithm based deep neural network for face recognition," Journal of Computational Science, vol. 51, pp.101352, April 2021.

[26] R. Anand, T. Shanthi, M. S. Nithish & S. Lakshman, "Face Recognition and Classification Using GoogleNET Architecture," Soft Computing for Problem Solving, pp. 261–269, 2019.

[27] R. Ranjan, V. M. Patel and R. Chellappa, "HyperFace: A Deep Multi-Task Learning Framework for Face Detection, Landmark Localization, Pose Estimation, and Gender Recognition," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 41, no. 1, pp. 121-135, 1 Jan. 2019.

[28] S. Gao, Y. Zhang, K. Jia, J. Lu and Y. Zhang, "Single Sample Face Recognition via Learning Deep Supervised Autoencoders," IEEE Transactions on Information Forensics and Security, vol. 10, no. 10, pp. 2108-2118, Oct. 2015.

[29] S. Nagpal, M. Singh, R. Singh and M. Vatsa, "Regularized Deep Learning for Face Recognition With Weight Variations," IEEE Access, vol. 3, pp. 3010-3018, 2015.

[30] S. Zhang, X. Pan, Y. Cui, X. Zhao and L. Liu, "Learning Affective Video Features for Facial Expression Recognition via Hybrid Deep Learning," IEEE Access, vol. 7, pp. 32297-32304, 2019.

[31] Seyedali Mirjalili, "Moth-flame optimization algorithm: A novel nature-inspired heuristic paradigm," Knowledge-Based Systems, vol. 89, pp. 228-249, November 2015.

[32] Shu-Chuan Chu, Pei-wei Tsai, and Jeng-Shyang Pan, "Cat Swarm Optimization," 9th Pacific Rim International Conference on Artificial Intelligence, 2006.

[33] T. Zhang, W. Zheng, Z. Cui, Y. Zong and Y. Li, "Spatial–Temporal Recurrent Neural Network for Emotion Recognition," IEEE Transactions on Cybernetics, vol. 49, no. 3, pp. 839-847, March 2019.

[34] Timo Ahonen, Abdenour Hadid, and Matti Pietik¨ainen, "Face Recognition with Local Binary Patterns," 2004.

[35] Tsung-Yung Hung and Kuo-Chin Fan, "Local Vector Pattern In High-Order Derivative Space For Face Recognition," ICIP, 2014.

[36] Wenle Xu, Rayan S Cloutier, "A facial expression recognizer using modified ResNet-152,"EAI Endorsed Transactions on Internet of Things, vol. 7, No. 28, 2021.

[37] Y. Li, G. Cao and W. Cao, "LMDAPNet: A Novel Manifold-Based Deep Learning Network," IEEE Access, vol. 8, pp. 65938-65946, 2020.

[38] Y. Tang, X. Zhang, X. Hu, S. Wang and H. Wang, "Facial Expression Recognition Using Frequency Neural Network," IEEE Transactions on Image Processing, vol. 30, pp. 444-457, 2021.

[39] Yifeng Zhaoand Deyun Chen, "Expression Recognition Using Improved AlexNet Network in Robot Intelligent Interactive System," Internet of Robotic Things-Enabled Edge Intelligence Cognition for Humanoid Robots, Article ID 4969883, 2022.

AUTHORS' PROFILE

Mr. Santhosh S., received his B.E. degree from SJMIT, Chitradurga, affiliated to VTU, Karnataka, India in 2005 and M.Tech., degree from MCE, Hassan, affiliated to VTU, Karnataka, India in 2011. He has a total experience of more than 16 years in teaching at Kalpataru Institute of Technology, Tiptur. Currently, he is working as Associate Professor in the department of CSE at Kalpataru Institute of Technology, Tiptur, Karnataka, India affiliated to VTU, Karnataka, India. He is pursuing Ph.D. at Kalpataru Institute of Technology research centre, affiliated to VTU, Karnataka, India. His areas of interest include Face Recognition, Image Processing and Pattern Recognition.

Dr. S. V. Rajashekararadhya, working as Professor in the Department of Electronics and Communication Engineering, Kalpataru Institute of Technology, Tiptur, Tumkur District, Karnataka, India affiliated to Visvesvaraya Technological University, Belagavi, Karnataka, India. He has a total academic and research experience of more than 30 years at various capacities as Lecturer, Senior Lecturer, Assistant Professor, Professor, Principal, NBA Coordinator, Member BOE VTU, Member LIC VTU, BOS UVCE Bangalore. He received his B.E. degree from Mysore University, Karnataka, India, in 1991, M.E. degree from UVCE, Bangalore, Karnataka, India in 2000 and Ph.D. degree from Anna University, Chennai, India in 2010. He received FDP fund from VGST, Karnataka. His areas of interest include Image Processing, Pattern Recognition, Handwriting Analysis, Feature Extraction, Neural Networks, Face Recognition, Machine Learning and Support Vector Machines. He has published more than 20 research publications in International Journals and Conferences. He has authored 2 book chapters.