An Automated Multimodal Biometric Recognition System Using Hybrid Deep Learning Models

Juhi Priyani ,Pankaj Nanglia,Paramjit Singh , Vikrant Shokeen,Anshu Sharma

Research Scholar,Maharaja Agrasen University,Baddi, Indiajuhipy@gmail.com
Associate Professor,Maharaja Agrasen University,Baddi, India,nanglia.pankaj@gmail.com
Assistant Professor,Maharaja Agrasen University,Baddi, India, ppparamjitsingh@gmail.com
Assistant Professor,MSIT,Delhi,India, shokeen18@gmail.com
Research Scholar, Maharaja Agrasen University, ahshukaushik10@gmail.com

Abstract

Most of the biometric developments have been extensively utilized for diverse security applications. A deep learning-based model serves as an end to end learning framework for multimodal biometric recognition, allowing for the proper learning of feature representation. Some of the major drawbacks faced in the existing approaches are dimensionality issues, convergence problems and poor network training ability. To overcome these issues, an automated multimodal biometric recognition system is proposed. After the collection of iris, finger print and palm print images from CASIA dataset, they are pre-processed using Quantized histogram equalization (QuaHE) for enhancing the image contrast and bilateral gaussian filter (BiGF) for effective removal of noises. From the pre-processed images, the suitable features are extracted using Hybrid vision capsule transformer model (HVicT) whereas the optimal characteristics features are chosen by utilizing Tent coati optimization algorithm (TcoA). The selected global and local features are fused together to generate a feature vector using Weighted feature fusion (WFF) approach. Finally, the recognition outcomes are obtained using Dense Pelican Channel block weighted autoencoder model (DWAM). Here, the appropriate person can be recognized using dense autoencoder model with channel block attention whereas the hyper parameters can be tuned effectively using pelican optimization algorithm. The performance evaluation are carried out using PYTHON simulation platform whereas the accuracy of proposed work is obtained to be 99.02%, specificity as 98.94% and F1 score as 98.63%. The proposed work promotes improved outcomes on comparison with several existing methodologies.

Index terms: Quantized histogram, Bilateral gaussian, Hybrid vision capsule, weighted fusion, dense autoencoder, Channel block attention.

DOI: 10.48047/ecb/2023.12.Si8.686

Introduction

The biometric verification and recognition systems have become a developing trend for individual determination dependent upon the behavioural and physiological characteristics [1]. Diverse physiological characteristics like iris, fingerprint, face, ear, palm print, etc. and behavioural features like signature, voice, keystroke dynamics, gait, etc. have been considered into an account for recognition performance evaluation [2]. Over the past decades, the security and multimedia communities have generally implemented biometric systems to carry certain tasks like surveillance cameras, facial recognition, physical access controls, and etc [3-4]. Biometrics act as the science of accomplishing person identification using semi or fully automated methodologies [5]. The biometric traits can be distinguished into two groups including extrinsic and intrinsic biometrics.

The extrinsic biometrics like iris and finger print are visible and influenced by external factors whereas the intrinsic traits like palm vein, finger vein cannot be influenced by external factors [6-7]. The two forms of biometric recognition systems are unimodal and multi modal. Even though the unimodal systems are trustworthy, they are prone to sensed data noises, attack vulnerabilities, non-universality issues and class similarities.

Generally, the multimodal biometric systems need more than one trait for recognizing the appropriate user. To conquer the issues faced in unimodal biometric systems, the multimodal biometric system has been applied widely in most of the real world applications [8-9]. For recognition strategies, most of the biometric researchers have utilized machine learning (ML) algorithms [10-11]. But, some feature extraction methodologies do not promote effective learning outcomes. Also, there is a complexity in handling biometric data and they cannot rely on image transformations.

The emergence of deep learning (DL) methodologies [12-14] has created a greater impact in the recent years. The utilization of DL based methods can overcome most of the drawbacks faced in ML algorithms especially associated in feature learning and cooperation with biometric image transformations [15]. In biometric recognition, the DL based model promotes an end to end learning framework and can learn the features effectively for accurate recognition [16]. The DL methodologies including convolutional neural network (CNN), multi-layer neural network (MLNN) and deep neural network (DNN) are concentrated in most of the existing research works. However, decent results seldom occur and it frequently takes a lot of effort. On motivated by the existing issues, an advanced DL model is proposed for exact biometric recognition.

The suggested research work is organized into a variety of segments. Section 2 surveys multimodal biometric recognition works done by various researchers in learning methodologies. Section 3 deals with the working of proposed methodology using novel image processing approaches. In Section 4, the simulations used to evaluate the effectiveness of the anticipated work outcome are covered. The conclusion and future scope of the planned effort are then provided under Section 5 with the appropriate citations.

1. Related works

A decision fusion methodology was proposed by Mustafa et al. [17] with the integration of finger print and iris biometrics. Initially, the Gray-Level Co-occurrence Matrix (GLCM) was utilized for feature extraction and k-nearest neighbor (KNN) was employed for categorizing the fingerprint and iris samples. The feature extraction approach was employed for promoting comparison between the related features and feature database dependent upon the Euclidean distance. The data were collected from CASIA-Iris databases V1, V2 and FVC 2004 Fingerprint database. The accuracy rate of 95% was obtained through the presented decision fusion approach. But the recognition rate was declined due to ineffective feature consideration and learning ability.

Rajasekar et al. [18] presented an enhanced multimodal biometric approach called an optimized fuzzy genetic algorithm for a smart city that is dependent upon score level fusion. The significant section of the presented research was the inclusion of fuzzy policy incorporated with soft computing approaches. The iris images were pre-processed using Wildes iris localization, daugman’s rubber sheet iris normalization whereas the features were extracted using log Gabor filter. The finger print images were pre-processed using Gabor filter, ore point region of interest selection and line based feature extraction was carried out. The extracted features were fused together using a score level fusion and finally, a decision was made but the false rejection rate was very high.

A Hybrid approach to multimodal biometric recognition dependent upon feature-level fusion of Face, iris and thumbprints was presented by Safavipour et al. [19]. Initially, the face features were extracted through kernel based approaches, iris features were extracted through Daugman algorithm and Hough Transform whereas the thumb print features were extracted using Gabor filter. After the process of normalization, kernel methods were utilized for feature vector mapping. The dimensionality reduction procedures were employed for merging the features. The consumption of time in multimodal biometric recognition was very high because of model complexity in feature learning.

An adaptive multi-biometric fusion policy for the dynamic combination of gait and face biometric data through the adoption of keyless attention deep neural networks was presented by Prakash et al. [20]. Also, certain external factors like distance and viewpoint to the camera were examined in this study. The gait features and face features were extracted using a convolutional long short term memory (LSTM). Through the adoption of bilinear pooling approach, the face and gait features were merged effectively. In this research, CASIA gait dataset was employed to examine the performance of the proposed model. The proposed research obtained 90% of accuracy and 0.389 as log loss.

On conducting the survey over recent multimodal biometric recognition systems, there emerges certain drawbacks like ineffective feature consideration, learning ability, declined accuracy performance, higher false rejection rate and improper decision making. Also, the time consumption, model complexity, improper data consideration and less overall system efficiency are resulted. Hence to overcome these existing limitations, an efficient automated biometric recognition system using multimodal data is introduced in the proposed work.

2. Proposed methodology

Multimodal biometric recognition acts as an extensive phase in the recent years specifically when huge data are provided as the input. Accurate recognition of multimodal biometrics using input data, lessening of recognition error and enhancing the overall accuracy tends to be a highly challenging scenario. Hence, an efficient DL based Multimodal biometric recognition system is proposed to obtain increased performance outcomes.

Initially, the data required to carry out the proposed multimodal biometric recognition model is collected from CASIA dataset of fingerprint, iris and palm print. To improve the contrast and get rid of the noise, images are pre-processed. Then the relevant features are extracted and most suitable features are selected to enhance the feature learning ability. The selected features from multimodal data are fused together and the person corresponding to the data can be recognized effectively. The overall schematic representation of proposed multimodal biometric recognition model is demonstrated in Figure 1.

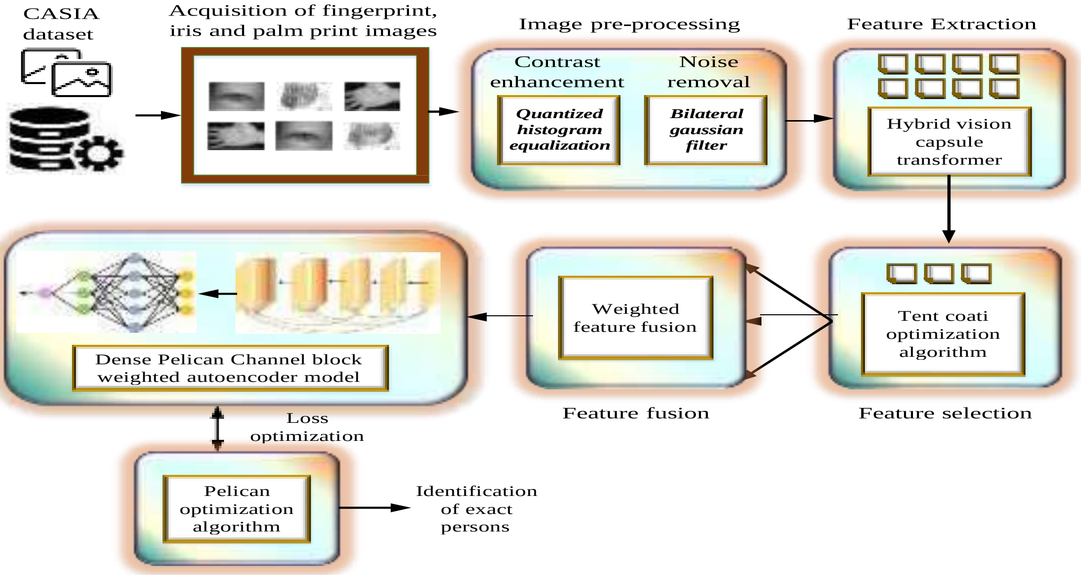

CASIA dataset
Acquisition of fingerprint, iris and palm print images
Image pre-processing
Contrast enhancement
Noise removal
Quantized histogram equalization
Bilateral gaussian filter
Feature Extraction
Hybrid vision capsule transformer
Dense Pelican Channel block weighted autoencoder model
Weighted feature fusion
Tent coati optimization algorithm
Feature fusion
Feature selection
Loss optimization
Pelican optimization algorithm
Identification of exact persons

Figure 1: Overall schematic architecture

2.1 Image pre-processing

Image pre-processing is a significant step needs to be carried out to obtain precise multimodal biometric recognition outcome. In the proposed work, the collected biometric images are pre-processed using QuaHE for contrast enhancement and BGF for noise eradication. The procedure of two processes are described in detail as follows.

2.1.1 Quantized histogram equalization

To enhance the contrast of an image, QuaHE technique is adopted and it involves four processes like histogram separation, clipping, allocation of grey level range and histogram equalization. Initially, the histogram of the original image is separated into two sub histograms. The medians of sub histograms are used for the further division and thus forming four sub histograms. The maximum and minimum intensity values of the input histogram are considered as the separating points. Using the below given equations, the separating points can be estimated as follows.

$$a_1 = 0.25 \times \{In_W \times In_H\} \tag{1}$$

$$a_2 = 0.50 \times \{In_W \times In_H\} \tag{2}$$

$$a_3 = 0.75 \times \{In_W \times In_H\} \tag{3}$$

From the above equations, a_1, a_2 and a_3 denotes the intensity set of values 0.25, 0.50 and 0.75. Next, clipping process is carried to regulate the enhancement rate of histogram equalization [21] to conquer over enhancement problems. To balance the enhancement space for every sub histograms, a new gray level dynamic range is allocated dependent upon the ratio overall pixels to the gray level spans for every sub histogram. The processes can be mathematically formulated as follows.

$$Span_k = S_{k+1} - S_k \tag{4}$$

$$Factor_k = Span_k \times (\log_{10} s_k)^{\alpha} \tag{5}$$

$$Range_k = (V-1) \times Factor_k \,/ \sum_{l=1}^{4} Factor_l \tag{6}$$

From the above equations, the dynamic gray level utilized by k^{th} sub histogram of an input image is represented as $Span_k$. The k^{th} separating point is signified as S_k and the total number of pixels in the k^{th} sub histogram can be denoted as s_k. The dynamic level range is denoted as $Range_k$ and α indicates the emphasis amount provided over s_k .To reduce the model complexity, the equation (6) can be rewritten as follows.

$$Range_k = (V-1) \times Span_k \,/ \sum_{l=1}^{4} Span_l \tag{7}$$

The new dynamic range is assigned from $[k_{Start}\ k_{End}]$ in the k^{th} sub histogram and it can be expressed in the below given equations.

$$k_{Start} = (k-1)_{End} + 1 \tag{8}$$

$$k_{End} = k_{Start} + Range_k \tag{9}$$

Once the new dynamic ranges are identified for every quadrant sub histograms, the final step is equalization of sub histograms independently. If the k^{th} histogram is assigned from

$[k_{Start}\ k_{End}]$, the histogram equalization outcome can be identified through the transfer mapping function $D(u)$ that can be given as follows.

$$D(u) = (k_{Start} - k_{End}) \times CDF\,(U_t) + k_{Start} \tag{10}$$

From the above expression, the cumulative density function can be indicated as $CDF\,(U_t)$. Instead of minimum and maximum intensities from the output dynamic range, k_{Start} and k_{End} are employed in the general histogram equalization expression.

3.1.2 Bilateral Gaussian filter

To eradicate the noise and retain original image information, BiGF technique is utilized in this research. For smoothening the image, input and guidance image are made to be non-similar. The guidance image is attained through Gaussian blur procedure and it can be signified as follows.

$$Z(y) = \sum\nolimits_{yz} F^{G}_{p,q}(G)\, R_q \tag{11}$$

From the above equation, the input image is specified as R whereas the guidance image is indicated as G .The midpoint position of input image is denoted as y in which the filter kernel $F^{G}_{p,q}$ can be formulated as follows.

$$F^{G}_{p,q} = \frac{1}{X_f} \exp\left(- \frac{\|p-q\|^2}{F_m^2} \right) \tag{12}$$

From the above equation, X_f denotes the normalizing factor and Gaussian spatial kernel is signified as $-\frac{\|p-q\|^2}{F_m^2}$. The enhanced Gaussian bilateral kernel is mathematically formulated as below.

$$F^{BGF}_{p,q}(R, G^-) = \frac{1}{X_f} \exp\left(- \frac{\|p-q\|^2}{\sigma_p^2} \right) \exp\left(- \frac{\|R_p - G^-\|^2}{\sigma_q^2} \right) \tag{13}$$

The range kernel can be defined as $\left(- \frac{\|R_p - G^-\|^2}{\sigma_q^2} \right)$ and the filtered outcome of BGF can be expressed as follows.

$$Z(y) = \sum\nolimits_{yz} F^{BGF}_{p,q}(R, \overline{G})\, R_q \tag{14}$$

The general BF [27] can process the local information only whereas the textures cannot be processed effectively. The BGF algorithm eradicates the textures and conserve the edges which includes two sections including local regularization and global optimization. For gradient smoothening, local regularization is carried out using cross scale relative value L that regulates the scale variance.

$$L = \left| \frac{B^{\sigma 1}_{BGF} * \delta O_{p,q}}{B^{\sigma 2}_{BGF} * \delta O_{p,q}} \right| \tag{15}$$

From the above equation, O indicates the smoothing image, $\delta O_{p,q}$ represents the gradient of O in p or q direction. Also, an adaptable parameter δ utilized where $\delta_1 < \delta_2$, $\left| B^{\sigma 1}_{BGF} * \delta O_{p,q} \right|$ and $\left| B^{\sigma 2}_{BGF} * \delta O_{p,q} \right|$ indicates the small and large scale feature. To enhance the outcomes of image smoothing, global optimization function can be defined as follows.

$$Arg_O \min \left\| O - R \right\|_2^2 + m\left(\left\| \frac{B^{\sigma 1}_{BGF} * \delta O_{p,q}}{B^{\sigma 2}_{BGF} * \delta O_{p,q}} \right\|_1 + \left\| \frac{B^{\sigma 1}_{BGF} * \delta O_{p,q}}{B^{\sigma 2}_{BGF} * \delta O_{p,q}} \right\|_1 \right) \tag{16}$$

From the above equation, the input image is signified as R and p specifies the parameter greater than zero. L2-norm fidelity term, $\left\| O - R \right\|_2^2$ is used to minimize the distance between relativity outcome of BGF and input image. The normalized outcome of BGF can be reorganized as follows.

$$\left\| \frac{B^{\sigma 1}_{BGF} * \delta O_{p,q}}{B^{\sigma 2}_{BGF} * \delta O_{p,q}} \right\|_1 = \left\| \frac{\left(B^{\sigma 1}_{BGF} * \delta O_p \right)^2}{\left(B^{\sigma 2}_{BGF} * \delta O_p \right)\left(D^{\sigma 1}_{GABF} * \delta O_p \right)} \right\|_1 \tag{17}$$

$$\approx B^{\sigma 1/2}_{BGF} * \frac{1}{B^{\sigma 1}_{BGF} * Max\left(\left(\delta O_p \right), \varphi \right)} \left\| \delta_p O \right\|_2^2 \tag{18}$$

The BGF regularization using L1 normalization is complex and hence it is estimated using equations (17) and (18). Through BGF, effective noise removal results can be attained with enhanced quality.

3.2 Feature extraction

The relevant features that are required to promote the overall recognition accuracy can be extracted using HVicT model. In HVicT model, certain functionalities of capsule network [28] are combined in vision transformer model [29] like capsule embedding, squashing and token embedding. Along with it, transformer encoder block, average pooling layer and multilayer perceptron block are also utilized for extracting the relevant features. To attain capsule feature encodings, every image is divided into three patches comprising different context information which are then directed towards capsule embedding block. For each stream, the equivalent patches are handled using capsule embedding process followed by token embedding process. Figure 2 indicates the architecture of HVicT model.

The patches with similar intensity values are transformed into vectorial capsules using capsule embedding module. It comprises of convolutional layer with suitable kernel size, padding and stride for representing the capsules. To normalize the length, squashing is applied over capsules that can be represented as follows.

$$NorC_W = \frac{\|C_W\|^2}{1+\|C_W\|^2}\,\frac{C_W}{1+\|C_W\|} \tag{19}$$

From the above equation, $NorC_W$ signifies the normalized capsule and C_W specifies the original capsule. The weight and height of capsules get shrinked over half sizes compared to the original ones. It can be indicated as follows.

$$C = \sum_{w} A_w B_w C_w \tag{20}$$

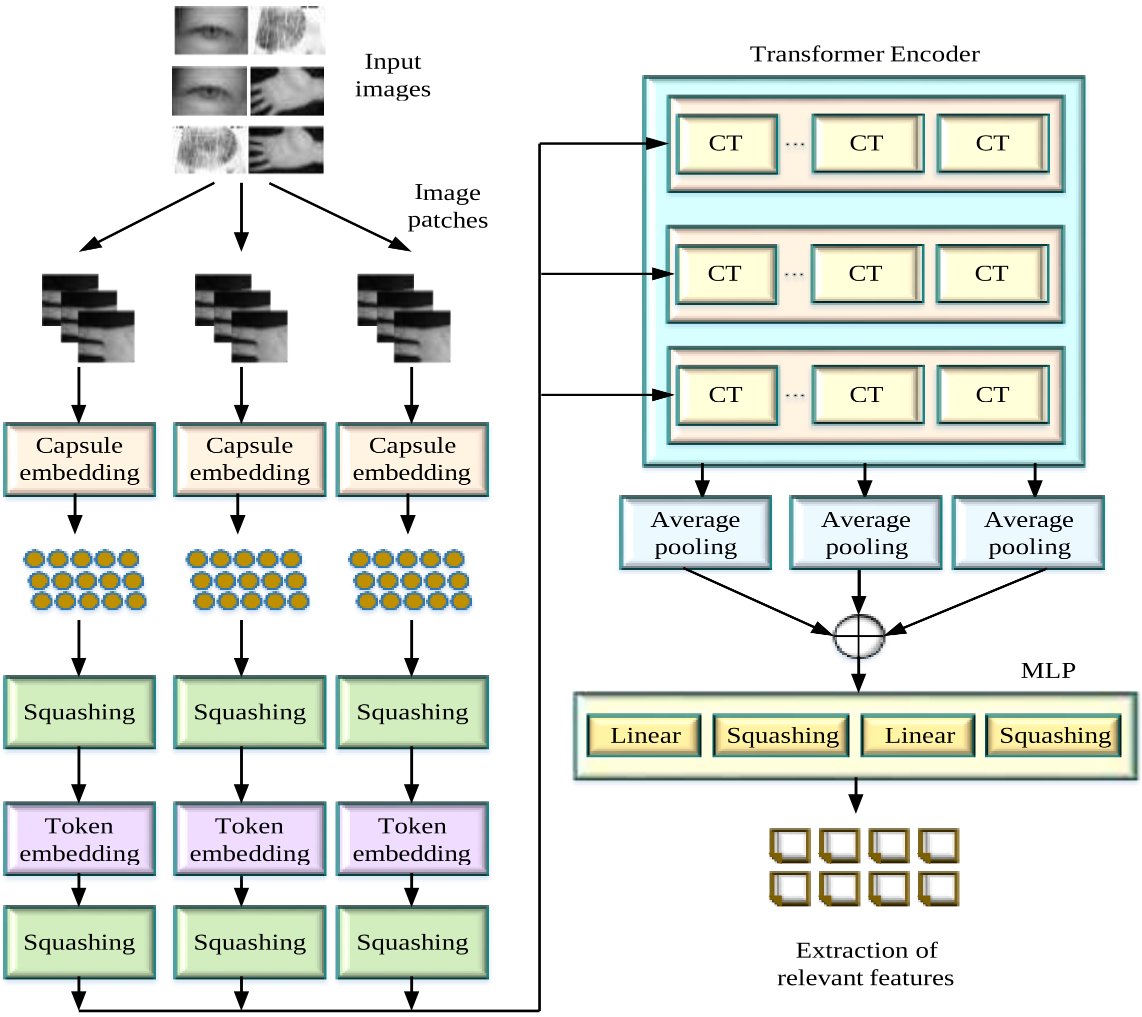


Figure 2: Schematic representation of HVicT model

From the above equation, the capsule inside the kernel is denoted as C_w, unnormalized capsule is meant as C, feature mapping matrix is symbolized as A_w and coupling coefficient is indicated as B_w. The squashing function is employed to compress the capsules and they are fed over transformer encoder for generating long range global features. The transformer encoder is loaded by a capsule transformer (CT) set composed of two parallel multi-head self-attention modules for attaining the global features and multilayer perceptron model to attain local searching of features. The multilayer perceptron is comprised of two linear layers and two squashing layers. The linear layer possesses its own weights for generating query, value and key. The multi-head self-attention module can be mathematically formulated as follows.

$$Attention\,(D,E,F) = Soft\max\left(\frac{DE^T}{\sqrt{z}}\right)F \tag{21}$$

The outcomes are produced with E query, key D and value F using linear input transformation. The scaling factors dependent upon the network depth is represented as Z. The outcome attained from CT set is given over average pooling layer and then to multi-layer perceptron layer to produce a set of multimodal biometric features.

3.3 Feature Selection

The computational complexity can be maximized because of processing higher dimensionality features. To attain accurate outcomes, the most appropriate features are selected using TcoA approach. The Coati optimization algorithm [30] is developed based on the analysis of coati behaviors. The position of coatis or the features can be randomly initialized as follows.

$$J_p : j_{p,q} = L_q + Rand\,(U_q - L_q),\ p = 1,2,3, \ldots ,U, \quad q = 1,2,3, \ldots ,V \tag{22}$$

From the above equation, J_p denotes the p^{th} feature position in search space, $j_{p,q}$ indicates the value of q^{th} variable and the total number of features are specified as $j_{p,q}$. L_q and U_q specifies the lower and upper bound of q^{th} decision variable. *Rand* represents the random real number of range$[0,1]$.

Attacking and Hunting phase over iguana

The exploration or initial phase updates the features in the search space on contemplating minimized error rate as the fitness function. The coati's update position at each iteration can be implied as follows.

$$J_p^{k+1} : j_{p,q}^{k+1} = j_{p,q} + Rand\,(Ig_q - \lambda . j_{p,q}),\ p = 1,2,\ldots.\left\lfloor \frac{U}{2} \right\rfloor, q = 1,2,\ldots V \tag{23}$$

where, λ implies the randomly chosen integer which equals the value to be 1 or 2 and *Ig* indicates the iguana position. When the iguana falls over ground, it is placed in a random search space. The coatis move based on that and is simulated using the below equation.

$$Ig_G : Ig_G^q = L_q + Rand\,(U_q - L_q), q = 1,2,3, \ldots ,V \tag{24}$$

$$J_p^{K1} : j_{p,q}^{K1} = \begin{cases} j_{p,q} + Rand\,(Ig_G^q - \lambda . j_{p,q}), & Fit_{Ig_G} < Fit_p \\ j_{p,q} + Rand\,(j_{p,q} - Ig_G^q), & else, \end{cases}$$
$$for\ p = \left\lfloor \frac{U}{2} \right\rfloor + 1, \left\lfloor \frac{U}{2} \right\rfloor + 2, \ldots.U\ and\ q = 1,2,\ldots V \tag{25}$$

The updated coati position is reasonable if it satisfies the fitness function or else preceding position is continued. The update plan for $p = 1,2,\ldots.U$ is replicated using the below given expression.

$$J_p = \begin{cases} J_p^{K1}, & F_p^{K1} < F_p \\ J_p & else \end{cases} \tag{26}$$

From the above equations, J_p^{K1} indicates the new position assessed for p^{th} coati, $J_{p,q}^{K1}$ represents both dimension and Ig_G indicates the iguana position on ground. Fit_{Ig_G} signifies the objective function value and the greatest integer function is denoted as $\lfloor . \rfloor$.

Escaping process from predators

When the predator attacks coati, the animal escapes from its current position during exploitation phase. To pretend the updation behavior, a random position is generated near present coati that is expressed below.

$$L_q^{Local} = \frac{L_q}{K}, U_q^{Local} = \frac{U_q}{K} \quad K = 1,2,3,.....,k \tag{27}$$

$$J_p^{K2} : j_p^{K2} = j_{p,q} + (1 - 2rand).\left(L_p^{Local} + Rand\left(U_p^{Local} - L_p^{Local}\right)\right) \\ p = 1,2,...U, q = 1,2,...V \tag{28}$$

The newly assessed position is acceptable if it improves the objective function value and the condition pretends using the below equation.

$$J_p = \begin{cases} J_p^{K2}, & F_p^{K2} < F_p \\ J_p & else \end{cases} \tag{29}$$

The new position assessed for p^{th} coati dependent upon exploitation phase is meant as J_p^{K2}. The q^{th} dimension is signified as $J_{p,q}^{K2}$, F_p^{K2} indicates the value of objective function value and K means the iteration counter. L_q^{Local} and U_q^{Local} indicates the lower and upper bound of q^{th} decision variable. To improve the selection performance efficiency, TcoA approach is employed. During initialization strategy, tent chaotic map is utilized to exchange random generation and the equation (22) can be rearranged as follows.

$$J_p : j_{p,q} = L_q + Tent_C\,(U_q - L_q),\ p = 1,2,3,....,U, \quad q = 1,2,3,.....,V \tag{30}$$

$$Tent_C^{k+1} = \begin{cases} \dfrac{Tent_C^k}{t}, & Tent^k \in (0,t) \\ \dfrac{1 - Tent_C^k}{1-t}, & Tent^k \in (t,1) \end{cases} \tag{31}$$

The position of coati is modified using $Tent_C$, tent chaotic map for global searching performance improvement. In attacking phase, dynamic weight factor β is employed to update the coati position. Equation (25) can be reframed as follows.

$$J_p^{K1} : j_{p,q}^{K1} = \begin{cases} \beta = \dfrac{e^{2(1-t/T)} - e^{-2(1-t/T)}}{e^{2(1-t/T)} + e_q^{-2(1-t/T)}} \\ j_{p,q} + Rand\,(Ig_G - \lambda .\, j_{p,q}), & Fit_{Ig_G} < Fit_p \\ j_{p,q} + Rand\,(j_{p,q} - Ig_G^q), & else \end{cases} \qquad (32)$$

From the above equation, t specifies the iteration counter and T signifies the maximum number of iterations. The optimal features can be selected using TcoA approach whereas the dimensionality issues can be overcome on prioritizing the most suitable features.

3.4 Weighted feature fusion

General concatenation of features cannot stimulate the actual implication of different modalities**.** To learn the multimodal local and global features effectively, WFF strategy with the utilization of weighted soft attention concatenation layers. A soft attention layer is employed over the obtained features from multimodal images that helps the system to focus on informative modality. By providing a weighted score vector, effective knowledge can be gathered with the indication of individual feature importance. WFF comprises three major sections such as soft attention module, weighted operation and decision making module. The weighting score W can be assessed using the following expression.

$$W = Soft\max\,(\tanh\,(C_k\,[Z_p^*, Z_q^*] + D_k)) \qquad (33)$$

From the above equation, C_k and D_k indicates the fusion attention parameters and W specifies the dimensional vector. The soft attention is evaluated over the multimodality features which are then concatenated together. The dense layers are utilized to learn the relationship among the features. Using this WFF strategy, the effectively concatenated feature vectors are attained. Figure 3 depicts the procedure of WFF strategy in feature concatenation.

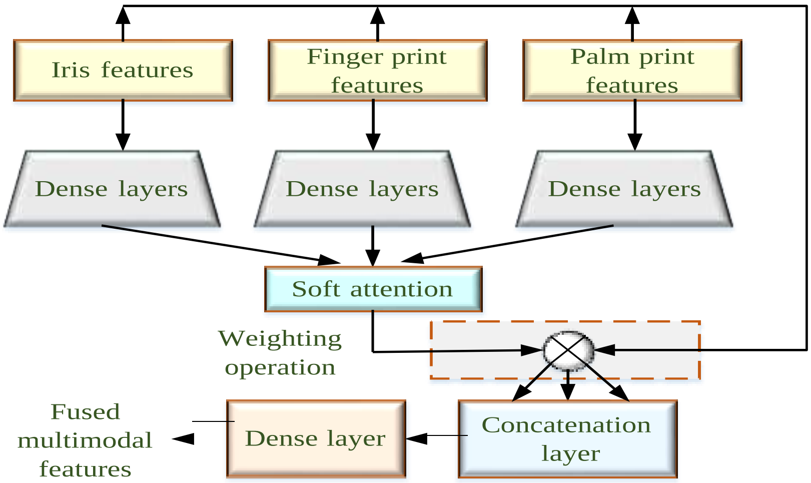


Figure 3: Procedure of WFF strategy

3.5 Multimodal Biometric recognition

The multimodal biometric data of fingerprint, iris and palm print can be effectively recognized using DWAM approach. The DenseNet-201 construction along with channel block attention module (CBAM) learns the features with the utilization of learnable weights. Because of feature reusing probability, it is parametrically effective through different layers. CBAM is an effective attention module and it assumes the attentional map along channel and spatial dimensions for better feature consideration. At first, the features are traversed over channel attention module to attain weighted outcomes. Then the features are passed towards spatial attention module whereas the assessment formula for channel and spatial attention modules are given below.

$$CHA(M) = \alpha\left(MLP(Avg_{p}(M) + MLP(Max_{p}(M))\right) \tag{34}$$

$$SPA(M) = \alpha\left(M(Concat(Avg_{p}(M),(Max_{p}(M)))\right) \tag{35}$$

From the above equations, M indicates the feature map, $CHA(M)$ means channel attention module and $SPA(M)$ indicates spatial attention module, α signifies the sigmoid function, Avg_{p} represents average pooling function and Max_{p} implies the maximum pooling function. The concatenation of features can be mathematically formulated as follows.

$$K^{I} = NL_{I}\left(\left[K^{0}, K^{1}, \ldots\ldots K^{I-1}\right]\right) \tag{36}$$

From the above equation, the non-linear transformation is indicated as NL_{I} and is characterized as a composite function containing ReLU, Convolution, batch normalization (BN) and CBAM. The incorporation of feature maps corresponding to layer 0 to $I-1$ can be implied as $\left[K^{0}, K^{1}, \ldots\ldots K^{I-1}\right]$. For down sampling, dense blocks including ReLU, Convolution, BN, CBAM and average pooling layers are utilized. The pooling layer gradually decreases the feature size for parameter and recognition complexity minimization. The feature map grouping from dense block is undertaken where the dimensions can be reduced using transition layer. The utilization of autoencoders has expanded in diverse sectors and in this research work, weighted autoencoders (WAE) are concentrated. Figure 4 indicates the schematic architecture of DWAM approach for multimodal data recognition.

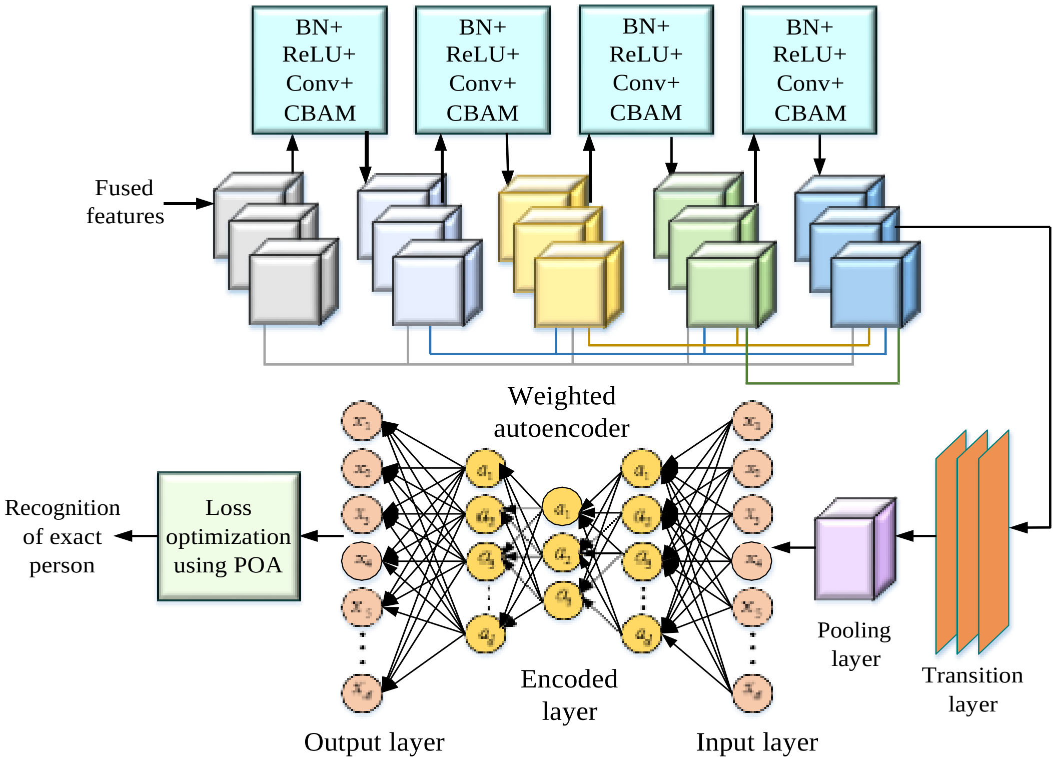


Figure 4: Schematic representation of DWAM approach

The AE acts as an unsupervised neural network which is used to regenerate a vector from its latent feature space. At first, the vector is utilized as the input $s^{(q)} \in [0,1]^z$ corresponding to the latent space as $r^{(q)} \in [0,1]^z$. The encoding function parameterized by $\varphi = \{W,u\}$ is denoted as,

$$r^{(q)} = f\varphi(s^{(q)}) = Sig\,(W\,s^{(q)} + B) \qquad (37)$$

From the above equation, the nonlinear activation function is indicated as Sig, bias vector is indicated as B and $q \times q$ weight matrix is represented as W. The latent representation $r^{(q)} \in [0,1]^z$ are then mapped over reconstructed vector $s^{\wedge(q)} \in [0,1]^q$ in the input space.

$$s^{\wedge(q)} = g\varphi(v^{(q)}) = Sig\,(W\,v^{(q)} + B) \qquad (38)$$

The weight W of reverse mapping gets constrained by $W = WV$ whereas the AE is said to possess tied weights. The chief objective of training is for parameter learning $\varphi = \{W,u\}$ and $\varphi' = \{W,u\}$ is to minimize the average reconstruction error over input vector set $s^{(1)},.....s^{(t)}$.

$$(\varphi.\varphi') = Argq\underbrace{\max}_{\varphi.\varphi'}\frac{1}{t}\sum_{q=1}^{t}\overset{(\varphi.\varphi')}{}CL\,(s^{(q)}.si^{(q)}) \qquad (39)$$

$$(\varphi.\varphi') = Argq \underset{\varphi.\varphi}{\min} \frac{1}{t} \sum_{q=0}^{t} CL\left(s^{(q)}.g\varphi'\left(f_{\varphi}\left(s^{(q)}\right)\right)\right) \tag{40}$$

$$CL = \sum_{a=1}^{R} \sum_{b=1}^{S} - Q_a^{(b)} \log \left(b\right)_a^{(b)} \tag{41}$$

The cross entropy loss function is represented as CL and the hyper parameters can be optimized using pelican optimization algorithm (POA). The obtained output vector is represented as Q_a and the demanded output vector is meant as $Q_a = (0,...,0,1,...1,0,...,0)$. Due to CL generation, the recognition accuracy gets worsened and hence the losses obtained in the neural network model are optimized using POA. The outcome of first autoencoder is fed as the input over second autoencoder to obtain the most related information based on the multimodal biometric data. Through the implementation of proposed models, effective recognition of persons corresponding to the multimodal data can be obtained with improved accuracy and less error rates. The performance evaluation of the proposed work are given in detail as follows.

3. Results and discussion

This section deals with the detailed simulation outcome analysis on comparison with various state of the art approaches. The recognition performances of proposed work are evaluated through the conduction of experiments with respect to same dataset. The simulation set up of proposed multimodal biometric recognition is executed in PYTHON platform. The dataset details, performance metrics evaluation procedure, analysis and baseline model comparison are provided in the following sub-sections. Table 1 indicates the hyperparameter details of the proposed work.

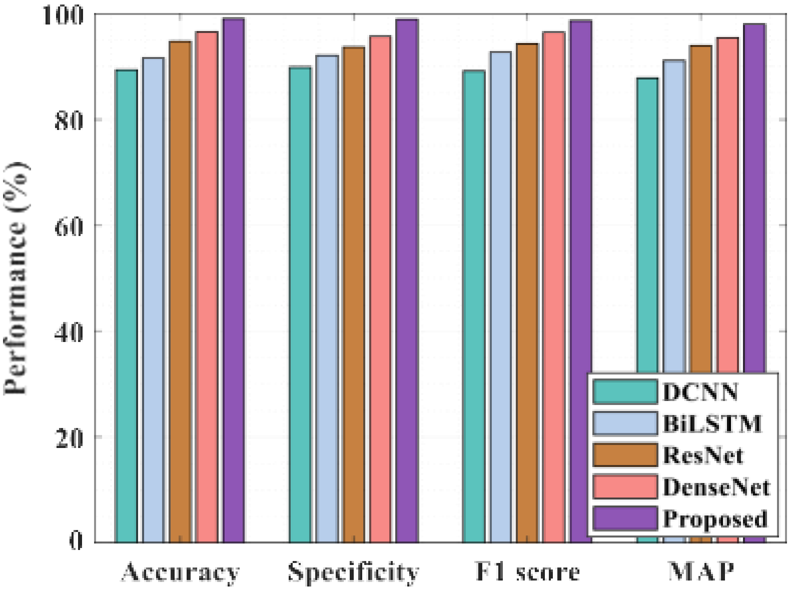
Table 1: Hyper parameter settings

Sl. No	Hyper-parameters	Proposed model
1.	Batch size	32
2.	Initial learning rate	0.0001
3.	Maximum epoch size	300
4.	Activation function	ReLU
5.	Maximum iteration	100
7.	Dropout rate	0.1
8.	Loss Function	Cross entropy
9.	Optimizer	Pelican

The experiments are conducted using CASIA dataset of fingerprint, iris and palm print. CASIA-Iris V3 comprises a total of 22,035 iris images from more than 700 subjects. All iris images possess 8 bit gray-level JPEG file format, obtained under near infrared illumination. CASIA-FingerprintV5 comprises of 20,000 fingerprint images from 500 subjects. The fingerprint images of CASIA-fingerprintV5 were obtained using URU4000 fingerprint sensor in one session. CASIA multi-spectral palm print image dataset comprises of 7,200 palm images gathered from 100 different people using a self-designed multiple spectral image device. For evaluation purposes, 80% of data are utilized for training and 20% are used for testing.

The existing methodologies like deep convolutional neural network (DCNN), bidirectional long short term memory (BiLSTM), Residual network (ResNet) and DenseNet are considered for comparison. The performance results in terms of planned and actual performance are illustrated in figure 5 using a variety of metrics, including Accuracy, Specificity, F1 score, and MAP.

Figure 5: Performance comparison of proposed and existing models



The accompanying graph clearly demonstrates that the suggested person identification model has achieved better results while taking into account features as accuracy, specificity, F1 score, and MAP. The proposed approach has accomplished better accuracy rate of 99.02% whereas the existing methods like DCNN, BiLSTM, ResNet and DenseNet has obtained 89.42%, 91.54%, 94.75% and 96.52% of accuracy rates. Due to increasing complexity and inadequate training capabilities, the suggested model achieves more accuracy than the previous techniques. The specificity of 98.94%, F1 score of 98.63% and MAP of 97.98% are obtained by the proposed model.

A clear analysis can be made from the above figure that the EER of proposed work in person recognition has attained improved outcomes compared to the existing models. The EER of 3.86 is obtained by the proposed model whereas the existing methods like DCNN, BiLSTM, ResNet and DenseNet has obtained 5.95, 4.95, 4.32 and 4.08 respectively. Better error rate resistances are attained by the proposed model by concentrating over the most desirable features. The existing models like DCNN, BiLSTM, ResNet and DenseNet has accomplished less performance than the proposed model because of certain drawbacks like utilization of non-significant features and increased training time. On comparing the RMSE values between the proposed and existing models, the proposed model obtained only lesser rate of RMSE as 1.026 accordingly. Figure 6 (a)-(b) indicates the ROC analysis and PR curve performances.

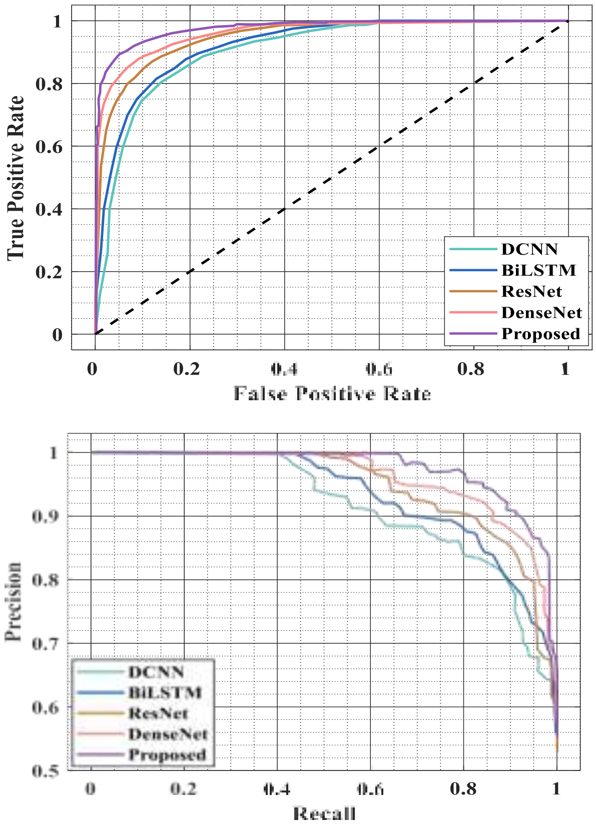


Figure 6: Evaluation outcomes (a) ROC curve (b) PR curve

The PR curve generally tends to explore the trade-off between precision and recall. Also, it is employed to evaluate the recognition effectiveness. The ROC analysis demonstrates better capability to distinguish the target classes. The ROC curve is plotted between FPR and TPR whereas the ability in multimodal biometric recognition is enhanced based on respective input parameters. The optimum cut-off displays the maximum sensitivity with minimum specificity. The ROC curves are often examined to expose the trade-off between TPR and FPR for every possible probability. It labels the efficiency of biometric recognition and also indicates the degree of capability. Higher the rate of ROC denotes better the performance of biometric recognition. In Figure 7, the suggested model's confusion matrix is shown

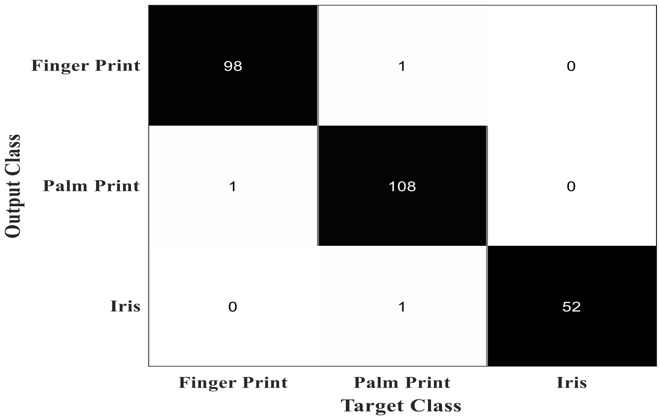
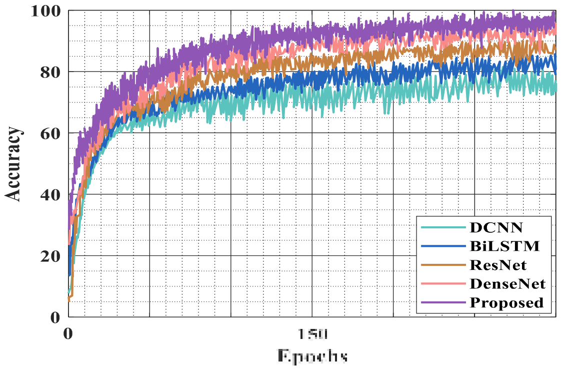
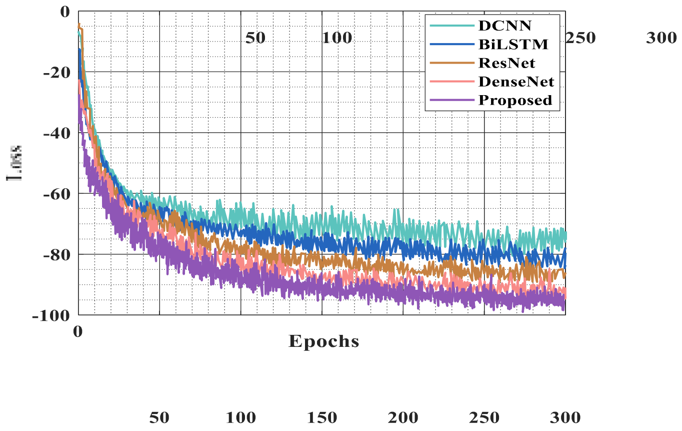


Figure 7: Confusion matrix

The reputation of proposed model in recognizing the persons corresponding to iris, palm print and finger are analysed. From the above figure, it can be obviously stated that the proposed model recognizes the persons with improved accuracy. On considering Iris images as the input, 52 images are accurately recognized whereas 1 image is wrongly predicted as palm print. With respect to Palm print images as the input, 108 images are accurately recognized whereas 1 image is wrongly recognized as finger print. In case of input finger print images, 98 images are exactly predicted whereas 1 image is wrongly predicted as palm print. The biometric multimodal data can be effectively recognized with minimal error possibilities and hence the accuracy of proposed work is widely enhanced.Figure 8 (a)-(b) represents the accuracy and loss curve performance in terms of proposed andexisting methodologies.





(a) (b)

Figure 8: Performance outcomes (a) Accuracy curve (b) Loss curve

By changing the epoch size from 0-300, accuracy and loss performance of the model are evaluated. The training and testing loss is examined and the network has been trained for 300 epochs. The accuracy increases with increased epoch size whereas the loss decreases with increased epoch size which when increased to 300, an accuracy between 97 to 100% is attained. And it is clear from the graphic that the suggested model achieves maximal accuracy. If the epoch size is increased to 300, the model obtains a loss in the range between 0.05-0.1. The proposed model gained minimal losses because of effective data training process using novel methodologies.

4. Conclusion

In this research work, the persons corresponding to iris, finger print and palm print images from CASIA dataset can be effectively recognized. Through this research, the demerits of existing methodologies like dimensionality issues, convergence problems, decline accuracy rates and poor network training ability are effectively overcome. Better performances are obtained through by pre-processing the images at first using QuaHE and BiGF methods. The relevant features were extracted using HVicT model and the dimensionality issues were overcome by selecting the most optimal features using TcoA model. The obtained multimodal features are effectively fused together using WFF strategy. The person recognition using multimodal data were undertaken using DWAM approach. Also, the hyper parameters are effectively tuned to reduce the loss functions and by this the error rates are extensively minimized. The overall accuracy of 99.02%, specificity as 98.94% and F1 score as 98.63% are attained by the proposed model evaluated using PYTHON platform. The EER of proposed model is obtained to be 3.86 and the RMSE performance is evaluated as 1.026 that are far inferior to the existing models. The suggested study can be expanded in the future by utilising other datasets, feature matching strategies can be initiated and hybrid optimization methodologies can be used to concentrate on more effective features.

REFERENCES

[1] Alay, Nada, and Heyam H. Al-Baity. "Deep learning approach for multimodal biometric recognition system based on fusion of iris, face, and finger vein traits." Sensors 20, no. 19 (2020): 5523.

[2] Medjahed, Chahreddine, Abdellatif Rahmoun, Christophe Charrier, and Freha Mezzoudj. "A deep learning-based multimodal biometric system using score fusion." IAES International Journal of Artificial Intelligence 11, no. 1 (2022): 65.

[3] Gunasekaran, K., J. Raja, and R. Pitchai. "Deep multimodal biometric recognition using contourlet derivative weighted rank fusion with human face, fingerprint and iris images." Automatika: časopis za automatiku, mjerenje, elektroniku, računarstvo i komunikacije 60, no. 3 (2019): 253-265.

[4] Daas, Sara, Amira Yahi, Toufik Bakir, Mouna Sedhane, Mohamed Boughazi, and El-Bay Bourennane. "Multimodal biometric recognition systems using deep learning based on the finger vein and finger knuckle print fusion." IET Image Processing 14, no. 15 (2020): 3859-3868.

[5] Hamd, Muthana H., and Marwa Y. Mohammed. "Multimodal biometric system based face-iris feature level fusion." Int. J. Mod. Educ. Comput. Sci 11, no. 5 (2019): 1-9.

[6] Ammour, Basma, Larbi Boubchir, Toufik Bouden, and Messaoud Ramdani. "Face–iris multimodal biometric identification system." Electronics 9, no. 1 (2020): 85.

[7] El mehdi Cherrat, Rachid Alaoui, and Hassane Bouzahir. "A multimodal biometric identification system based on cascade advanced of fingerprint, fingervein and face images." Indonesian Journal of Electrical Engineering and Computer Science 18, no. 1 (2020): 1562-1570.

[8] Ammour, Basma, Larbi Boubchir, Toufik Bouden, and Messaoud Ramdani. "Face–iris multimodal biometric identification system." Electronics 9, no. 1 (2020): 85.

[9] Kusse, Gergito, and Tewoderos Demissie. "APPLICATIONS OF MULTIMODAL BIOMETRICS AUTHENTICATION FOR ENHANCING THE IOT SECURITY USING DEEP LEARNING." Ethiopian International Journal of Engineering and Technology 1, no. 1 (2023): 1-11.

[10] Vijay, M., and G. Indumathi. "Deep belief network-based hybrid model for multimodal biometric system for futuristic security applications." Journal of Information Security and Applications 58 (2021): 102707.

[11] Walia, Gurjit Singh, Tarandeep Singh, Kuldeep Singh, and Neelam Verma. "Robust multimodal biometric system based on optimal score level fusion model." Expert Systems with Applications 116 (2019): 364-376.

[12] Elhoseny, Mohamed, Ehab Essa, Ahmed Elkhateb, Aboul Ella Hassanien, and Ahmed Hamad. "Cascade multimodal biometric system using fingerprint and iris patterns." In Proceedings of the International Conference on Advanced Intelligent Systems and Informatics 2017, pp. 590-599. Springer International Publishing, 2018.

[13] Hammad, Mohamed, Yashu Liu, and Kuanquan Wang. "Multimodal biometric authentication systems using convolution neural network based on different level fusion of ECG and fingerprint." IEEE Access 7 (2018): 26527- 26542.

[14] Regouid, Meryem, Mohamed Touahria, Mohamed Benouis, and Nicholas Costen. "Multimodal biometric system for ECG, ear and iris recognition based on local descriptors." Multimedia Tools and Applications 78 (2019): 22509- 22535.

[15] Bhilare, Shruti, Gaurav Jaswal, Vivek Kanhangad, and Aditya Nigam. "Single-sensor hand-vein multimodal biometric recognition using multiscale deep pyramidal approach." Machine Vision and Applications 29 (2018): 1269- 1286.

[16] Soleymani, Sobhan, Ali Dabouei, Hadi Kazemi, Jeremy Dawson, and Nasser M. Nasrabadi. "Multi-level feature abstraction from convolutional neural networks for multimodal biometric identification." In 2018 24th International Conference on Pattern Recognition (ICPR), pp. 3469-3476. IEEE, 2018.

[17] Mustafa, Ahmed Shamil, Aymen Jalil Abdulelah, and Abdullah Khalid Ahmed. "Multimodal biometric system iris and fingerprint recognition based on fusion technique." International Journal of Advanced Science and Technology 29, no. 03 (2020): 7423-7432.

[18] Rajasekar, Vani, Bratislav Predić, Muzafer Saracevic, Mohamed Elhoseny, Darjan Karabasevic, Dragisa Stanujkic, and Premalatha Jayapaul. "Enhanced multimodal biometric recognition approach for smart cities based on an optimized fuzzy genetic algorithm." Scientific Reports 12, no. 1 (2022): 1-11.

[19] Safavipour, Mohammad H., Mohammad A. Doostari, and Hamed Sadjedi. "A hybrid approach to multimodal biometric recognition based on feature-level fusion of face, two irises, and both thumbprints." Journal of Medical Signals and Sensors 12, no. 3 (2022): 177-191.

[20] Prakash, Ashwin, Athira Nambiar, and Alexandre Bernardino. "Multimodal Adaptive Fusion of Face and Gait Features using Keyless attention based Deep Neural Networks for Human Identification." arXiv preprint arXiv:2303.13814 (2023).

[21] Veluchamy, Magudeeswaran, and Bharath Subramani. "Image contrast and color enhancement using adaptive gamma correction and histogram equalization." Optik 183 (2019): 329-337.

[22] Bhonsle, Devanand, Vivek Chandra, and G. R. Sinha. "Medical image denoising using bilateral filter." International Journal of Image, Graphics and Signal Processing 4, no. 6 (2012): 36.

[23] Pandian, A. Pasumpon. "Identification and classification of cancer cells using capsule network with pathological images." Journal of Artificial Intelligence 1, no. 01 (2019): 37-44.

[24] Wang, Libo, Shenghui Fang, Xiaoliang Meng, and Rui Li. "Building extraction with vision transformer." IEEE Transactions on Geoscience and Remote Sensing 60 (2022): 1-11.

[25] Dehghani, Mohammad, Zeinab Montazeri, Eva Trojovská, and Pavel Trojovský. "Coati Optimization Algorithm: A new bio-inspired metaheuristic algorithm for solving optimization problems." Knowledge-Based Systems 259 (2023): 11001