Online-Journals.org

iJIM International Journal of Interactive Mobile Technologies

iJIM | eISSN: 1865-7923 | Vol. 18 No. 22 (2024) | OPEN ACCESS

https://doi.org/10.3991/ijim.v18i22.50865

PAPER

Design and Development of Multimodal Biometric System Using Finger Veins and Iris by CNN Integrated with Hybrid SIO and Whale Optimization Techniques

Naitik S.T.[1–3](✉), **J.V. Gorabal**[1,2]

[1]Department of CSE, Research Center – ATME College of Engineering, Mysore, Karnataka, India

[2]Visvesvaraya Technological University, Belagavi, Karnataka, India

[3]Department of CSE, Dayananda Sagar University, Bengaluru, Karnataka, India

naitik.st-cse@dsu.edu.in

ABSTRACT

Biometrics encompasses technological and scientific advancements in monitoring and interpreting biological data from the human body to enhance system security by offering precise and dependable patterns and techniques for person authentication and recognition. Its solutions are widely used by governments, the armed forces, and enterprises. Single sources of data in biometric systems, referred to as unimodal systems, are effective but frequently struggle with noisy data. Many of these challenges can be solved with multimodal biometric systems, which incorporate two or more biometric modalities. In this study, we utilized an optimized Convolutional Neural Network (CNN) for multimodal biometric recognition. For optimization, a hybrid of Swarm Intelligence (SI) and Whale Optimization (WO) algorithms was employed. The Finger Vein (FV) and iris modalities were chosen for biometric recognition. Data for both modalities were collected from the SDUMLA-HMT database and preprocessed before being fed into the CNN model for feature extraction and selection. Following the CNN modeling, both feature- and score-level fusion techniques were applied for individual recognition. The developed hybrid SI-WO-CNN model was evaluated against two other optimized models, namely the SI-CNN and WO-CNN. Experimental results show that the proposed hybrid CNN model achieves the highest accuracy, reaching 99% through the score-level fusion technique. Furthermore, the proposed model was compared with recent research works, demonstrating its effectiveness in biometric recognition.

KEYWORDS

biometric system, iris, Finger Vein (FV), deep learning, accuracy, optimization

1 INTRODUCTION

Automated biometric identification uses a person's physical characteristics and is a common feature of modern consumer goods [1]. Fields such as financial services,

Naitik, S.T., Gorabal, J.V. (2024). Design and Development of Multimodal Biometric System Using Finger Veins and Iris by CNN Integrated with Hybrid SIO and Whale Optimization Techniques. *International Journal of Interactive Mobile Technologies (iJIM)*, 18(22), pp. 97–114. https://doi.org/10.3991/ijim.v18i22.50865

Article submitted 2024-07-02. Revision uploaded 2024-08-19. Final acceptance 2024-08-30.

immigration, consumer electronics, electronic government, and online commerce are examples where it is applied. A biometric modality identifies a person based on a unique physical trait that meets specific standards. The main features include circumvention, collectability, performance, uniqueness, permanence, and universality [2]. A biometric human trait must possess at least one of these characteristics.

The property of being universal signifies that it is present in all human beings. The universality means all human beings possess the common biometric traits, and in capturing and preparing templates and using them for authentication and verification, there must be a universal measure to use these biometric traits.

Since each individual is unique, the characteristics must manifest in a slightly different manner to be deemed distinctive. The biometric trait may be behavioral or physiological and will be unique for each individual. Since the trait features are unique per person, the features vary from individual to individual. The uniqueness of the feature of the biometric trait helps us to authenticate and identify them effectively.

For a biometric characteristic to be deemed permanent, it must maintain its stability for a sufficient duration. We know that, as we have many seasons, the data collected from biometric traits also modifies. We must ensure that the collected data is stored very carefully in the form of templates and ensure the protection of templates against seasonal changes or any other accidental changes. The stability of the biometric data is enforced by storing them in the templates and using them against captured data. The integrity of the data is achieved via stabilizing the biometric data.

To qualify as collectible, another important characteristic is the collection of a biometric characteristic; the biometric data can be collected from images, sensors, datasets, etc. and must have the capability to be measured in a manner that enables its acquisition from human subjects. There are several methods to capture the human biometric data, but ensure that captured data must reflect the features of human characteristics.

The performance demonstrates an ability to precisely identify human individuals based on that attribute. The degree to which the biometric modality is convenient and inconvenient determines its level of acceptability. Circumvention denotes the level of complexity involved in executing a counterfeit attack to deceive a system based on biometric modalities.

Human characteristics are classified into physiological and behavioral features [3]. Physiological qualities concern the human body, whereas behavioral features concern the daily activities of humans. Most physiological qualities do not change much over time, while behavioral traits are more vulnerable to changes. Intrinsic and extrinsic modalities are two further classifications of physiological properties. Extrinsic physiological qualities occur outside the human body, whereas intrinsic modalities reside within it. Because of their transparency, extrinsic modalities are more vulnerable to planned impostor attacks. Parts of the biometric recognition system deal with sensors, feature extraction, matching, and decision-making. Both unimodal and multimodal biometric recognition systems exist [4]. In a unimodal system, only one biometric feature is used to identify the user. There are certain limitations to unimodal systems, despite their proven effectiveness compared to more conventional methods. Problems may arise due to sensed data noise, non-universality, susceptibility to spoofing attempts, and similarities within and between classes. To function properly, multimodal biometric systems must recognize users based on more than one attribute [5]. Because they address concerns with unimodal biometric systems, they have found widespread application in practical contexts. It is feasible to combine various characteristics in multimodal biometric systems by utilizing data stored in one of the modules. Fusion can occur

at four levels: sensor, score, feature, and decision. Multimodal biometric systems offer several advantages over unimodal systems and represent a promising strategy for secure recognition. In this study, we selected Finger Vein (FV) and iris data for a multimodal biometric system. For automatic recognition, the optimized Convolutional Neural Network (CNN) model is employed. The reason for choosing them is detailed below.

First, due to a high degree of privacy, it is preferred to employ the FV pattern as a biometric feature [6]. Vein patterns are difficult, if not impossible, to shoot in natural light due to their subcutaneous location. Because of this feature, biometric systems are significantly less vulnerable to spoofing attempts. The second benefit of using the FV is that it improves the integrity and endurance of the biometric system. This is due to the difficulty of surgically altering the human body's venous pattern. The acceptance of FVs as a biometric characteristic is the third element affecting their widespread use. FV pictures can be taken using either noncontact or weak-contact settings. As a result, the biometric data collection system is more user-friendly. FV data collection is more interference-resistant since it is less influenced by scars, moisture, and the presence of any unpleasant substance on the finger's surface. This capability makes FV biometrics the favored choice in biometric systems.

Iris recognition is a cutting-edge biometric technology that confirms a person's identity by analyzing their distinct eye patterns [7]. This cutting-edge technology has aroused broad interest due to its exceptional precision and longevity. The iris is a part of the eye that has unique patterns that differ from person to person. These patterns remain persistent throughout a person's life, indicating that they are not subject to change. Iris recognition is exceptionally accurate and dependable due to its stability and distinctiveness, which greatly decreases the potential of misidentification. One of the primary benefits of iris recognition is its amazing accuracy. Biometric processes such as iris identification differ from others such as fingerprint or face recognition since two irises are unlikely to be similar. This makes it a reliable identification mechanism for a wide range of applications by considerably reducing the possibility of incorrect acceptance or rejection. Iris recognition is a biometric technology that does not involve physical contact or trespass. Using a basic camera to capture iris patterns eliminates the requirement for direct contact with the device. As a result, patients do not need to worry about maintaining personal hygiene or being ill during therapy. Iris recognition is a simple and quick approach to verifying identity. In a couple of seconds, the system can identify a person by capturing their iris image and comparing it to templates already in its database. Iris recognition is ideal for mission-critical applications that demand quick verification due to its speed and user-friendliness.

2 LITERATURE SURVEY

Experts in multimodal biometric security have devised a variety of methods. In this section, we will briefly review a few significant additions to the existing literature. The research [8] suggests efficient multimodal biometric techniques that combine left and right palm prints using a matching score combination. We demonstrate and train multimodal biometric identification systems using K-Nearest Neighbors (KNN) and CNN to identify and recognize individuals based on the scores from several biometric data sources. The biometric systems are trained using well-known biometric benchmark databases to produce a reliable and secure verification/identification system. The proposed model is tested in harsh scenarios utilizing

noisy datasets. Computer simulations indicate that the CNN and KNN approaches outperform other prominent current biometric verification methods. [9] suggests a score-level fusion-based multimodal biometric approach for smart city applications. To be more precise, the suggested approach combines enhanced performance multimodal fusion technologies with a modified fuzzy genetic algorithm to eradicate existing problems. Experiments in diverse biometric contexts demonstrate significant improvements over baseline approaches. According to the findings, the proposed method beats the alternatives in terms of accuracy, precision, recall, false acceptance/rejection rates, and Equal Error Rate (EER). The proposed technique achieves a lower EER while increasing accuracy.

A study [10] suggests a coding-based method known as the bit-transition code to address the lack of research into the development of a biometric identification system that combines iris and palmprint modalities. At each pixel location, the binary transitions of the Gabor-filtered images are encoded. Through score-level fusion, the performances of the palmprint and iris modalities of the individual are combined. Our study is conducted using three benchmark iris/palmprint databases: the IITD palmprint database, the PolyU palmprint database, and an additional benchmark database. To evaluate the efficacy, Receiver Operator Characteristic (ROC) curves and other metrics are utilized. To establish the practicality of the proposed strategy, it is comprehensively contrasted with several cutting-edge approaches. The paper [11] describes a multimodal biometric authentication system based on the deep integration of Electrocardiogram (ECG) and FV. The proposed system consists of three major components: authentication, pre-processing, and feature extraction. The preprocessing phases include normalization and filtering techniques. The feature extraction approach includes the use of a deep CNN model. We then verify the collected features using five popular machine-learning classifiers. We also use Multi-Canonical Correlation Analysis (MCCA) to compress the massive data set into a lower-dimensional feature space and speed up the authentication procedure. Two biometric systems, one based on ECG patterns and the other on-Field Potential (FP) patterns, are fused into one multimodal system through the use of feature and score fusion algorithms. The suggested multimodal system also uses MCCA feature fusion in conjunction with a KNN method to increase authentication accuracy.

According to the research [12], a CNN and a Bimodal Fusion Feature (BFF) layer of facial features and FVs make up the biometric method. It is in the feature layer that the fusion occurs. The two biometrics' relative importance is decided by applying the self-attention method. After that, the RESNET model incorporates the self-attention weight feature along with the BFF channel. The experimental part employed models such as AlexNet and VGG-19 to extract features from FV and face pictures, proving that the BFF layer is quite effective. These features were then input into the BFF module. According to rigorous testing, both models achieve excellent identification accuracy levels, demonstrating the effectiveness of BFF. Research [13] presents and tests a multimodal ultrasonic recognition system that combines 3D hand and palmprint data. For both functions, the system creates volumetric images of the entire hand by gathering several 2D images at different depths. A 3D template is built by properly merging the 2D information gathered from each image. Verification and identification trials are used to evaluate recognition performance using a handcrafted database. Starting with two unimodal biometrics, we try merging them at the score level to see what happens. The findings show that fusion considerably improves the recognition performance of individual biometrics, with a 100% identification rate and an EER of 0.08%. The findings given in the research publications under consideration are lacking in terms of the simplicity of the technique used. The FAR and FRR, among other assessment indicators, and overall accuracy

have room for improvement. The researches tries to overcome the issues with the proposed hybrid SI-WO-CNN model.

3 METHODOLOGY

The methodology employed for developing a multimodal biometric recognition system using optimized CNN is elaborately detailed in this section. For this study, the SDUMLA-HMT dataset is utilized, which facilitates the multimodal recognition of both iris and FV. Initially, separate preprocessing steps are implemented for iris and FV images to prepare them for effective feature extraction. The architecture and hyperparameters of these CNNs are meticulously optimized using a hybrid approach that integrates Swarm Intelligence (SI) and Whale Optimization (WO) algorithms. This optimization aims to achieve the best possible recognition performance by fine-tuning the network parameters to the specific traits of iris and FV biometrics. Following the feature extraction phase, fusion techniques are applied at both feature and score levels to integrate the biometric data effectively. The performance of the biometric recognition system is evaluated using a comprehensive set of metrics, including accuracy, True Positive Rate (TPR), True Negative Rate (TNR), False Acceptance Rate (FAR), and False Rejection Rate (FRR). The entire process, from data preprocessing to the evaluation of the system, is illustrated in Figure 1.

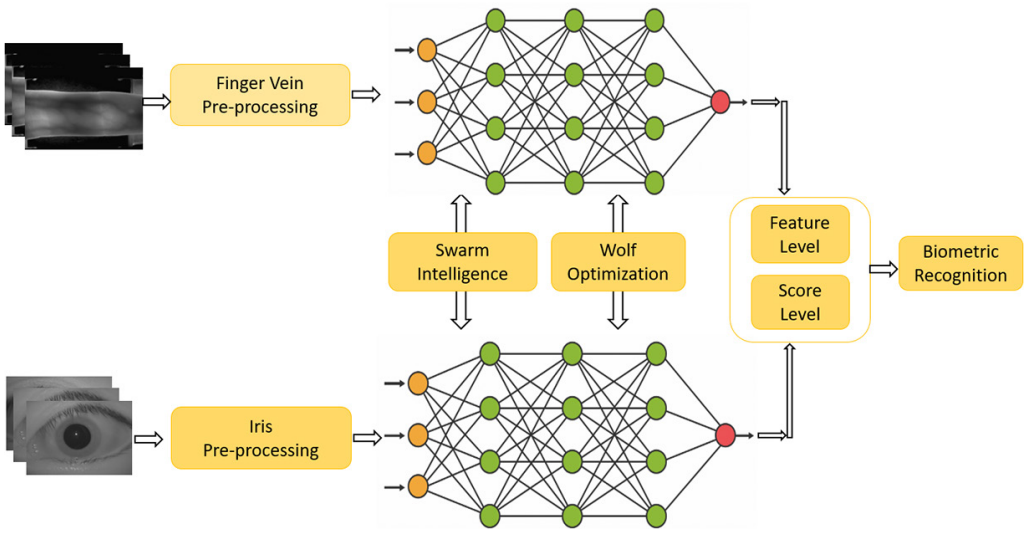

Fig. 1. Proposed system architecture

3.1 Data

In this study, we used FV and iris data from SDUMLA-HMT [14].

1. The data came from Shandong University in Jinan, China, in the summer of 2010. The data collection method involved 106 subjects.
2. 61 of whom were male and 45 females, ranging in age from 17 to 31.
3. Each person had their fingerprint, face, FV, gait, and iris biometric information recorded.
4. SDUMLA-HMT comprises five sub-databases: iris, face, gait, vein, and fingerprint database.

Wuhan University's lab developed a device that captures images of veins in the fingers.

1. To obtain FV images, the capturing technique involved asking each participant to submit images of the middle, index, and ring fingers on the right and left hands. This process was performed six times for each of the six fingers.
2. Figure 2.a displays a few sample images of FV. There are 3,816 images in the FV database, calculated as $6 \times 6 \times 106$.
3. Each image measures 320×240 pixels and is saved in the "bmp" format. Our FV database is a little over 0.85 gigabytes in size.

Under near-infrared light, we took iris scans using a smart iris-sensing device made by a Chinese University.

a) The subjects were instructed to remove their glasses and maintain an eye-to-device distance ranging from 6 to 32 cm to prevent reflection.
b) Each participant submitted a total of ten iris photographs, five from each eye.
c) Figure 2.b displays sample iris images from our database.
d) There are 1,060 images in the iris database, totaling $2 \times 5 \times 106$.
e) All iris images are saved in the "bmp" format, which has 256 gray levels and dimensions of 768×576 pixels.
f) Our iris database occupies about half a GigaByte (GB) of space.

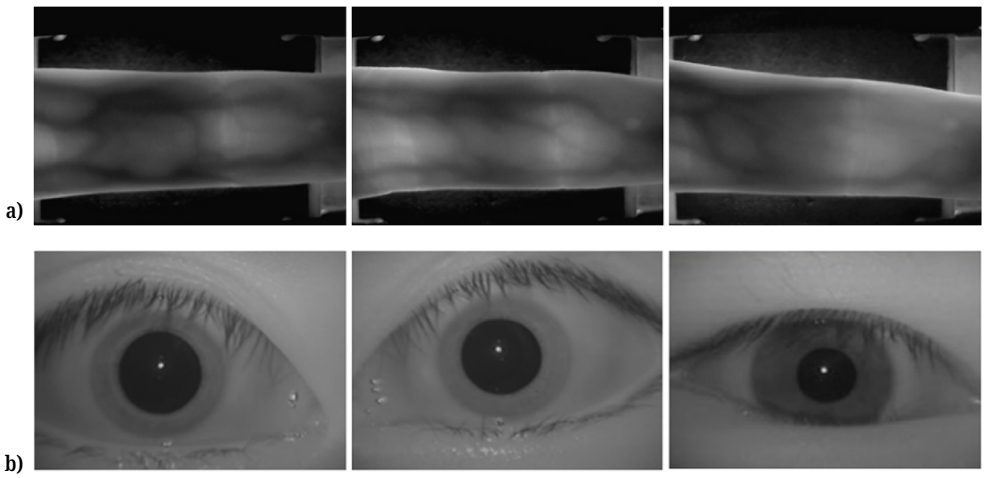

Fig. 2. Biometric data samples

3.2 Data pre-processing

The preprocessing is done on both iris and FV images. For FV authentication, a variety of preprocessing techniques are used. The suggested algorithm includes the following:

1. To improve accuracy, we first darken the image's background by setting the pixels to zero.
2. Horizontal finger alignment requires inserting the finger into a structural device that resembles a tube. Aligning the images can also be accomplished using software approaches. Rotating the image to reveal the horizontal edges is the next step after identifying the edges of the finger image.

3. Limited adaptive histogram equalization is used to enhance the contrast, brightness, noise reduction, and detail clarification capabilities of images [15]. The intensity image values are adjusted to make the images more contrasty. The output's histogram primarily shows gray levels that are mostly black. If the output window's median score is 50% gray, all pixels are turned to black. In cases where one pixel in the input set has a lower gray level than every other pixel in the surrounding window, the cumulative total is utilized to adjust that pixel's gray level in the output.
4. By default, image normalization sets the ideal scale factor for venous databases at 0.6, reducing the original image size to a quarter of its original size.
5. Region of Interest (ROI) extraction increases precision by excluding irrelevant parts of the FV image. Adjust the height of the edge points to a range of 30–60% of the image height. Vertically crop the image at the specified spots. Horizontally crop the image within a range of 5% to 10% from the left and right borders.

Iris recognition requires preprocessing processes such as localization and normalization [16, 17]. Localization smoothies the image with the canny edge method and removes background noise with the Gaussian filter.

1. Gradients can help you locate vertical, horizontal, and diagonal edges in an image.
2. Non-maximum suppression helps to conceal all gradient values, except local maxima, to some extent.
3. Using the difference between the two threshold values, also known as "double thresholding," we can identify pixels with a weak edge.
4. Edge tracking using hysteresis—this technique removes weak edge pixels caused by noise or color changes.
5. Iris localization is the preprocessing step for an iris image using a sophisticated edge detection method that employs a Laplacian of Gaussian filter and normalizes through pupil dilation. During preprocessing, the color image is transformed to grayscale to facilitate localization, and the image's binarization is utilized to determine the center of the pupil. The procedure for converting from polar to Cartesian coordinates is as follows:

$$X = r\cos\theta \tag{1}$$

$$Y = r\sin\theta \tag{2}$$

Noise reduction is aided by this conversion. Matrix values in the grayscale serve as training data. This method uses the pupil's radius as its input. The fact that every iris has its own distinct pupil-to-iris boundary distance further demonstrates that the two boundaries are not concentric. In most cases, the size of a pupil will vary by 10–80% relative to the diameter of its iris. Thus, iris recognition proves to be an extremely reliable recognition algorithm. An iris template is comprised of between 256 and 512 bytes.

4 CNN MODEL

The core architecture of a CNN consists of fully connected, convolutional, pooling, and nonlinear activation layers. In a typical network design, the input layer receives the preprocessed input image, which is then processed by an alternating sequence of Convolutional Layers (CL) and Pooling Layers (PL). Finally, the Fully Connected Layer (FCL) is utilized for categorization [18].

4.1 Convolutional layer

The convolution of many layers enables CNNs with a certain depth to extract varied input information. In most cases, higher-layer convolutions extract more abstract information, whereas bottom-layer convolutions often extract more fundamental features. The convolution kernels that make up the CL can be adjusted using parameters [19]. The weights in these kernels are learnable. It is standard practice to feed feature maps into the CL. The interconnected feature map's local area determines the weight matrix, which the convolution kernel uses to execute a series of convolution processes on the feature map region by sliding. Input feature maps typically have dimensions of $H \times W \times C$, where H represents height, W represents width, and C represents the number of channels. One method to summarize the data flow in the CL simply is:

$$feature_{surface_{out}} = f\left(\sum_{i=3}^{3} 1M_i * W_i + B\right) \tag{3}$$

The input feature map's feature surface is represented by M_i, the weight and bias matrix are denoted as W_i and M, the nonlinear activation function is denoted as $f(\cdot)$, and the output feature surface is denoted as $feature_{surface_{out}}$. The CL performs a specific computation called cross-correlation between the feature surfaces and the convolution kernel. Consider the following parameters when describing the size of a feature surface: a 2-Dimensional (2-D) input matrix i, strides s, a convolution size k, and padding p.

$$o = \left[\frac{i + 2p - k}{s}\right] + 1 \tag{4}$$

This method streamlines the process, speeds it up, and can handle massive datasets—all while drastically reducing training parameters and overfitting.

4.2 Pooling layer

Typically, the CL precedes the pooling layer. The PL is primarily used for the following reasons: reducing the network's computational load by processing the input image through down sampling and dimensionality reduction to decrease the number of connections between CLs; ensuring that the input image is invariant to scale, translation, and rotation; and increasing the output feature map's tolerance to distortion and error in individual neurons [20]. The two most common methods for pooling data are average and maximum pooling. Several solutions have been proposed to overcome the problem of overfitting in CNNs. In the pooling process, the general relationship between the sizes of the input and output matrices is as follows:

$$o = \left[\frac{i - k}{s}\right] + 1 \tag{5}$$

4.3 Nonlinear activation function

Using a suitable nonlinear activation function, one can create a functional link between the input and output of the network, which greatly improves progress.

Two examples of saturating nonlinearities are the sigmoid and the Tanh. When the input is very big or little, the sigmoid function can take on a value of Zero or one, while the tanh takes the value of –1 or one. Non-saturating nonlinearities have been proposed as remedies to the problems caused by saturated nonlinearities. These include ReLU, Leaky ReLU, RReLU, PReLU, and ELU.

4.4 Fully Connected (FC) layer

Typically, the FCL is established after the pooling and subsequent CLs, and its numerous layers consist of entirely connected neurons. By utilizing its multiple hidden layers, it extracts high-level features from the previous network in a more sophisticated manner. The category of an image is determined by the output vector, and the number of categories corresponds to the count of neurons at the output terminal. The FCL can be thought of as the CNN's classification layer. Gradient back-propagation is used to update the parameters of the FCL. The FCL often uses L2 regularization and dropout while training a big model with numerous parameters on a tiny dataset [21]. The main goal of utilizing them is to avoid model overfitting.

4.5 Loss function

The last layer of the FCL, known as the output layer, is responsible for achieving the final classification. Various visual tasks make use of different loss functions, which in turn impact the performance of the CNN architecture. All things considered, the loss function of choice for CNN models is now Softmax combined with Cross-Entropy. It has numerous improved variants that are useful for various visual tasks; these include L-Softmax, center loss, AM-Softmax, A-Softmax, etc.

4.6 Optimizer

It is widely acknowledged that gradient updates hold paramount significance in network training. Specifically, this process necessitates the use of partial differential calculations to transmit the gradient of the objective function (loss function) to the preceding network layer, this updates the learning configurations of each layer.

5 OPTIMIZATION TECHNIQUES

The hyper-parameters of the CNN model are tuned by the various optimization techniques, and a detailed explanation of the algorithm is given in this section.

5.1 Swarm Intelligence

The swarm optimization technique, inspired by bird hunting behavior, was created by Eberhart and Kennedy [22]. Finding the optimum solution for particles is quite similar to how birds find food. All that is needed to determine a particle's speed is the optimal global and local fitness value acquired in that iteration. Each particle has two specifications: position $X_i^t = \left[X_{i1}^t, X_{i2}^t, X_{i3}^t, \ldots, X_{id}^t \right]$ and velocity

$v_i^t = \left[v_{i1}^t, v_{i2}^t, v_{i3}^t, \ldots, v_{id}^t \right]$. During each iteration, the following formula is employed to update the velocity and position of all particles:

$$v_{id}^{t+1} = \omega v_{id}^t + c_1 r_1 \left(pbest_{id} - x_{id}^t \right) + c_2 r_2 \left(gbest_d - x_{id}^t \right) \tag{6}$$

$$x_{id}^{t+1} = x_{id}^t + v_{id}^{t+1} \tag{7}$$

In this context, x_{id}^t and v_{id}^t signify the initial velocities and positions of particle i; x_{id}^{t+1} and v_{id}^{t+1} denote the positions and velocities of particle i throughout iteration $t+1$, which indicates the current iteration. c_2 and c_1 are acceleration coefficients, $pbest_{id}$ is the optimal position of the i-th particle during this iteration, w is the inertial weight, $gbest_{id}$ is the optimal global position of the population, and the random parameters r_1 and r_2 are distributed between zero and one. With such a significant inertial mass, particles might easily diverge from the optimal local value at the time and fail to converge. When the inertial weight is properly balanced, particles gravitate toward the best local value. As a result, the inertial weight must be adjusted periodically. These qualities served as the foundation for the APSO algorithm, which calculates dynamic inertial weight. The inertial weight is adjusted according to the fitness score. Here is the formula for adaptive inertial weight:

$$\omega = \{\omega_{min} - \left(\omega_{max} - \omega_{min} \right) * \frac{\left(f_{cur} - f_{min} \right)}{\left(f_{avg} - f_{min} \right)}, f_{cur} \leq f_{avg}\ \omega_{max}, f_{cur} > f_{avg} \tag{8}$$

A particle's fitness value at any given time is represented by f_{cur}. In this population, f_{avg} represents the mean fitness value. The minimum fitness value in this population, denoted as f_{min}, is the lowest value among all particles.

5.2 Whale optimization

Lewis and Mirjalili developed the WOA algorithm, which mimics humpback whale behavior [23]. Whales are not just the largest mammals on Earth, but they also possess extraordinary intelligence and evolved social habits, making them truly amazing beings. The WOA algorithm uses three principles to modify and enhance the candidate solution's position, as well as a collection of randomly chosen solutions. The three steps of the WOA will be discussed in the following sections: identification, exploitation, and exploration. When a humpback whale spots its prey, it circles it. Because determining the optimal location to hunt for prey in the search space is not yet evident, target hunting is regarded as the best option for the WO algorithm. As demonstrated, the remaining agents move to take advantage of the best search prey position:

$$\vec{D} = \left| \vec{C}\, \vec{X} * (t) - \vec{X}(t) \right| \tag{9}$$

$$\vec{X}(t+1) = \vec{X} * (t) - \vec{A}\, \vec{D} \tag{10}$$

Where t denotes the current iteration number, X^* signifies the location vector of the optimal solution, X signifies the position vector, $\vec{X}(t+1)$ signifies the whale's position at each iteration, and C and A signify the coefficient vectors.

$$\vec{A} = 2\vec{a}\,\vec{r} - \vec{a} \quad [11]$$

$$\vec{C} = 2\vec{r} \quad [12]$$

Here, r is a random vector and its scores range from 0 to 1, and a parameter's value decreases linearly from 2 to 0 during the exploration and exploitation phases. Humpback whales' bubble-net activity can be described using two mathematical models: Reducing the value of a in Equation (11) initiates the shrinking encircling process. It is important to note that the parameter a has a major effect on A's range. The parameter a decreases from 2 to 0 over iterations, while A is a random number ranging from $-a$ to a. The revised search agent position can be anywhere between the primary and best agent positions by selecting random values for A ranging from –1 to 1. The distance between the whale's X^* and Y^* coordinates and the prey's coordinates can be determined using the spiral updating position. The helical motion of a humpback whale can be reproduced by modeling the positions of the whale and its prey using spiral equations:

$$\vec{D'} = \left|\vec{X^*}(t) - \vec{X}(t)\right| \quad (13)$$

$$\vec{X}(t+1) = \vec{D'}.e^{bl}.cos(2\pi l) + \vec{X^*}(t) \quad (14)$$

Where L is a randomly chosen integer between –1 and 1, and b is the constant representing the spiral-shaped route path. Based on Eq. (15), the WOA assigns a 50% chance to each of the two conceivable whale-moving mechanisms: spiral and encirclement. A value between 0 and 1 is produced at random for the P parameter.

$$\vec{X}(t+1) = \{\vec{X^*}(t) - \vec{A}.\vec{D}\ if\ p < 0.5\ \vec{D'}.e^{bi}.cos(2\pi l) + \vec{X^*}(t) if\ p \geq 0.5 \quad (15)$$

Humpback whale search patterns are completely random and depend on each whale's relative position in the pod. Equations (16) and (17) are used to determine the whales' random motions.

$$\vec{D} = \left|\vec{C}.\vec{X_{rand}} - \vec{X}\right| \quad (16)$$

$$\vec{X}(t+1) = \vec{X_{rand}} - \vec{A}.\vec{D} \quad (17)$$

In the equation given above, a random position vector of the population is denoted by X_{rand}. A collection of solutions selected at random is utilized to initiate the WOA algorithm. The optimal present solution or an arbitrary choice of agents determines how the search agents' placements are altered in each cycle. To demarcate the exploration phase from the exploitation phase, we reduce the parameter a from two to zero. The agent chosen at random for the search is the one whose absolute value is less than or equal to one. The search agent's position is modified to discover the optimal solution if $\vec{A}$ is greater than one. The algorithm chooses a spiral or circular motion depending on the value of p.

5.3 Hybrid SI-WO

Both the SI and WO algorithms operate separately in a hybrid SI-WO setup. Because its logarithmic spiral function wraps a larger portion of a less-defined

search space, WO is used during the exploration. The algorithm tries to propose multiple solutions during the exploration phase. Changing the particle's position to match the whale's position simplifies a complex nonlinear problem by effectively shifting the solution closer to the optimal one; otherwise, the two positions are identical. With WOA, particles can be propelled faster toward the optimal value with less computational effort. In a search space where the optimal solution is unknown, SI is known to leverage this fact. By merging the best features of exploration with WO and exploitation with SI, we can ensure that we will achieve the best possible optimal solution to the problem without becoming stuck in local optima or experiencing stagnation. As we move towards the best possible solutions, we can consider hybrid SI-WO, which combines the strengths of both WO and SI in the exploration phase.

$$V_{id}^{t+1} = \omega v_{id}^{t} + c_1 r_1 \left(Whale_{Pos_{id}} - x_{id}^{t} \right) + c_2 r_2 \left(gbest_d - x_{id}^{t} \right) \tag{18}$$

In this study, SI-WO algorithms are used to train the parameters of a 2D CNN. To accomplish the identification task with minimal human intervention, it is necessary to determine the hyperparameters that will yield the best results. The flowchart of SI-WO-CNN is given in Figure 3.

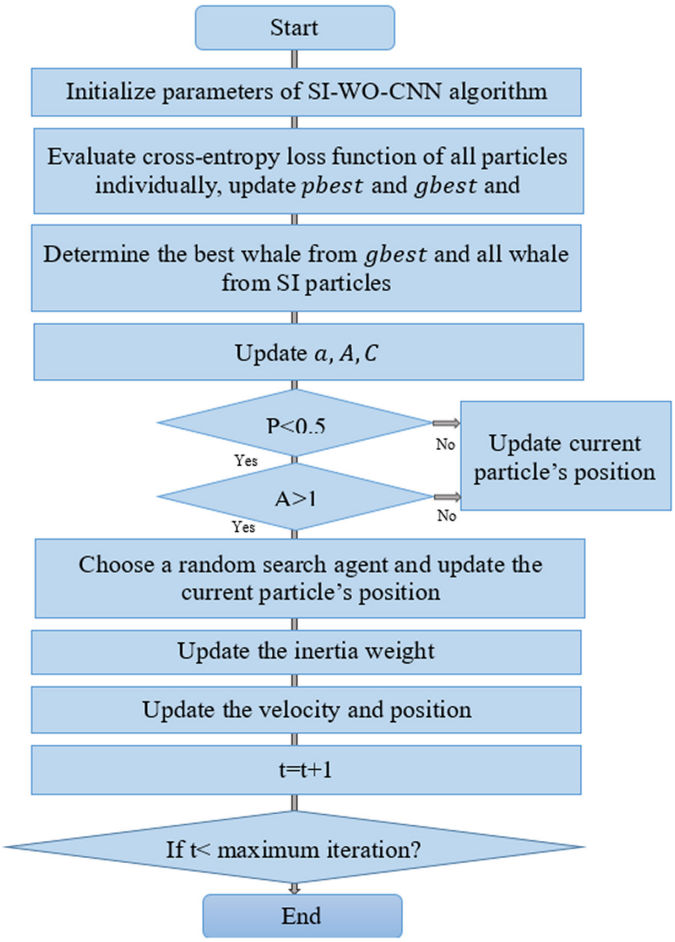

Fig. 3. Hybrid SI-WO-CNN model

6 FUSION STRATEGY

6.1 Feature level

The combination of characteristics that correspond to various attributes is an example of fusion at the feature level [24]. By combining the data retrieved from the two variables, additional user-representative characteristics were constructed. During the training phase of this fusion process, the model develops to identify the combined features. The second fully connected layer of the CNN models used to create the iris and FV models is utilized to fuse the models. The result is a single vector that combines the feature vectors from the second FCL of both CNN models; this vector is called:

$$X = x_r \mid x_v \tag{19}$$

x_r represent the iris features, and x_v the FV features that were recovered from the images. The SoftMax classifier takes the resulting vector (X) and uses the similarity score to categorize the image, ultimately allowing it to identify the individual.

6.2 Score level

Score-level fusion calculates the similarity score for the iris and FV using the output from the CNN's second FCL [25]. This information is subsequently provided to the CNN model's SoftMax classifier. The score fusion method consists of two stages. After normalizing each model's output, the scores from the CNNs were combined using a score fusion. Finally, the model identifies the receiver of the highest-scoring fused score. Two different methods of score fusion were used: the arithmetic and the product fusion. Using the arithmetic mean approach, a composite score is obtained by adding the scores for all the features and dividing it by the total traits. This equation defines the arithmetic mean rule.

$$S = \sum_{t=1}^{j} 1 \frac{s_t}{j} \tag{20}$$

where s_t represents the score vector of the trait t, and j indicates the total traits.

7 RESULTS AND DISCUSSION

In our study, we developed a multimodal biometric recognition system that uses both FV and iris data from the SDUMLA-HMT database. The system was created using a CNN model, which was optimized using a hybrid method that included SI and WO. This modification was critical in improving the model's performance by honing its capacity to automate feature extraction and biometric identification activities efficiently. In our trials, we used both feature-level and score-level fusion methodologies to assess the performance of the biometric system. The results show a significant improvement in biometric verification metrics compared to existing methodologies.

The SI-WO-CNN model produced a TPR of 98.02%, a TNR of 98.99%, and an accuracy of 98.5% through feature-level fusion. The FRR and FAR were 1.98% and 1.01%, respectively. This displays our model's ability to accurately identify and authenticate individuals based on their biometric features. The score level fusion technique improved the model's performance even further, resulting in an accuracy of 99%, a TNR of 98.02%, and a flawless TPR of 100%. The FRR was lowered to 0%, indicating no incorrect rejections, while the FAR rose slightly to 1.98%. Compared to other optimization strategies such as SI or WO, our hybrid approach produced better results. The SI-CNN model had the lowest performance at feature-level fusion, with an accuracy of 95.5%, TNR of 96.94%, and TPR of 94.12%. For a comprehensive view, all metrics of the optimized CNN model under various fusion techniques are tabulated in Table 1.

Table 1. Performance evaluation of optimized CNN models on multimodal biometric systems

DL Model	Hyperparameter Tuning	Fusion Approach	Accuracy	TNR	TPR	FRR	FAR
CNN	SI	Feature level	95.50	96.94	94.12	5.88	3.06
		Score level	96.50	95.10	97.96	2.04	4.90
	WO	Feature level	97.50	96.08	98.98	1.02	3.92
		Score level	98.00	97.03	98.99	1.01	2.97
	SI & WO	Feature level	98.50	98.99	98.02	1.98	1.01
		Score level	99.00	98.02	100.00	0.00	1.98

Figures 4 and 5 visually compare the true and false recognition rates across different optimized CNN models, illustrating the superiority of our approach. The implementation of feature and score-level fusion techniques also shows that strategic data integration can significantly enhance biometric system performance.

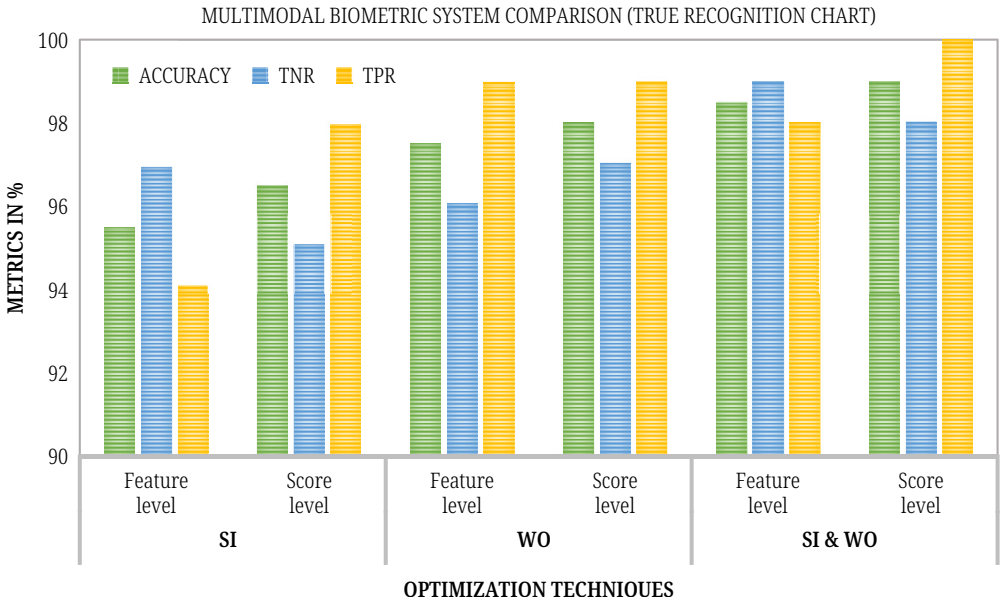

Fig. 4. Performance comparison of true recognition by optimized CNN model people's

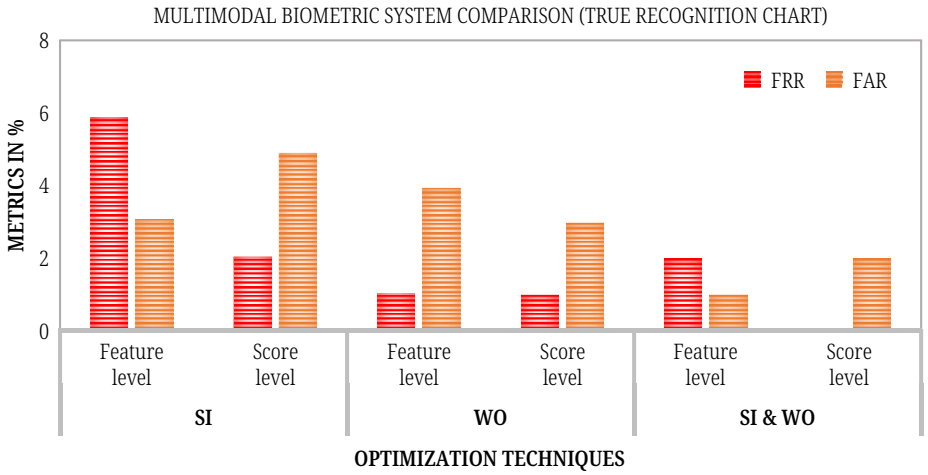

Fig. 5. Performance comparison of false recognition by optimized CNN model

In this study, the performance of our proposed multimodal biometric system, which combines iris and FV, was assessed against existing research to demonstrate its efficacy, and it is given in Table 2. Specifically, the system developed in Reference [21], which utilized finger texture combined with FV biometrics, achieved an accuracy of 92.95%. Reference [22] employed a multimodal approach integrating fingerprint and ECG, achieving a higher accuracy of 98.94%. Similarly, the system in reference [23] combined fingerprint with FV, reaching an accuracy of 98.78%. Our model, however, outperformed all these configurations by integrating iris recognition with FV, attaining an accuracy of 99%. This improvement highlights the effectiveness of combining iris patterns, which are highly unique and stable, with the complementary security features provided by FV recognition, thereby establishing a robust, secure, and more accurate biometric identification system.

Table 2. Comparison of proposed work with state-of-the-art methods

Sl. No.	Paper No	Biometric Recognition Techniques	Dataset	Optimization Method/Model	Accuracy
1	[8]	Face, left and right palm prints	FEI face dataset and IITD palm print database	KNN + CNN	99%
2	[9]	Iris + fingerprint	CASIA V3-Iris, FVC2006-Fingerprint	Optimized Fuzzy Genetic algorithm	99.88%
3	[10]	Iris + palmprint	MMDB1,2, IITD Iris DB, PolyU Palmprint DB	Not used	99.67%
4	[11]	ECG + fingervein	ECG-ID, FVPolyU, VeinECG	Multi-Canonical Correlation Analysis (MCCA) + KNN	99.88%
5	[12]	Face + fingervein	CASIA WebFace, FVUSM, SDUMLA-FV	CNN-Alexnet, VGG19	99.98%
6	[13]	3D Ultrasound palmprint hand geometry	Own dataset	NA	99.92%
7	[26]	Finger texture + FV	HKPU	CNN-AlexNet, VGG16, Score level fusion	92.5%
8	[27]	Fingerprint + ECG	PTB ECG and LivDet2015 database	CNN + QG-MSVM	98.94%
9	[28]	Fingerprint + FV	SDUMLA-HMT	Hybrid GA + PSO	98.78%
10	Proposed	Iris + FV	SDUMLA-HMT	CNN-DBN + HySIO + WIO	99%

8 CONCLUSION

Noise in the sensor data, non-universality of the biometric characteristic, limited degrees of freedom, and unacceptable error rates are some of the inherent issues with single biometric systems. A multi-biometric system attempts to circumvent these issues by offering several pieces of identification that all point to the same person. In this study, the multimodal biometric system is designed using iris and FV data. For automatic recognition of the multimodal biometrics, we developed a hybrid SI-WO-CNN model. The CNN model's hyperparameters are fine-tuned using the proposed optimized model. After the CNN model, fusion techniques like feature-level and score-level are utilized for individual recognition. The proposed hybrid SI-WO-CNN model achieves an accuracy of 99%, with FAR and FRR of 1.98% and 0%, respectively, by score-level fusion technique; and the model gives metrics of 98.02%, 1.01%, and 1.98% by feature-level fusion technique. The outcome explains the superior performance of the proposed model in the biometric recognition system. In the future, the developed multimodal biometric system will be deployed in real-time using FPGA technology. While implementing in hardware, we will analyze the accuracy and processing time.

9 ACKNOWLEDGEMENT

We thank and are grateful to Dr. D. Suresha, Professor and Head of the CSE, SITM, for his invaluable suggestions in making contributions to this paper. We also remember to thank Dr. Madhu, Professor and HOD of CSE VVIET, for his extendable support and guidance. Dr. Udaykumar Reddy KR, Dean-SoE, Dayananda Sagar University for his expertize and uncommendable guidance in the optimization techniques used in the research work.

10 REFERENCES

[1] G. Singh, G. Bhardwaj, S. V. Singh, and V. Garg, "Biometric identification system: Security and privacy concern," in *Artificial Intelligence for a Sustainable Industry 4.0*, S. Awasthi, C. M. Travieso-González, G. Sanyal, and D. Kumar Singh, Eds., 2011, pp. 245–264. https://doi.org/10.1007/978-3-030-77070-9_15

[2] Preeti Kumari and Ms. Pooja Ahlawat, "A comparative study of different biometric technologies," *International Journal of Enhanced Research in Science, Technology & Engineering*, vol. 6, no. 5, pp. 2319–7463, 2017.

[3] S. Dargan and M. Kumar, "A comprehensive survey on the biometric recognition systems based on physiological and behavioural modalities," *Expert Systems with Applications,* vol. 143, p. 113114, 2020. https://doi.org/10.1016/j.eswa.2019.113114

[4] O. N. Kadhim and M. H. Abdulameer, "Biometric identification advances: Unimodal to multimodal fusion of face, palm, and iris features," *Advances in Electrical & Computer Engineering,* vol. 24, no. 1, pp. 91–98, 2024. https://doi.org/10.4316/AECE.2024.01010

[5] N. Bala, R. Gupta, and A. Kumar, "Multimodal biometric system based on fusion techniques: A review," *Information Security Journal: A Global Perspective,* vol. 31, no. 3, pp. 289–337, 2022. https://doi.org/10.1080/19393555.2021.1974130

[6] A. H. Mohsin *et al.*, "Finger vein biometrics: Taxonomy analysis, open challenges, future directions, and recommended solution for decentralised network architectures," *IEEE Access,* vol. 8, pp. 9821–9845, 2020. https://doi.org/10.1109/ACCESS.2020.2964788

[7] Z. R. Sami, H. K. Tayyeh, and M. S. Mahdi, "Survey of iris recognition using deep learning techniques," *Journal of Al-Qadisiyah for Computer Science and Mathematics,* vol. 13, no. 3, pp. 47–56, 2021. https://doi.org/10.29304/jqcm.2021.13.3.826

[8] C. Medjahed, A. Rahmoun, C. Charrier, and F. Mezzoudj, "A deep learning-based multimodal biometric system using score fusion," *IAES Int. J. Artif. Intell. (IJ-AI),* vol. 11, no. 1, pp. 65–80, 2022. https://doi.org/10.11591/ijai.v11.i1.pp65-80

[9] V. Rajasekar *et al.*, "Enhanced multimodal biometric recognition approach for smart cities based on an optimized fuzzy genetic algorithm," *Scientific Reports,* vol. 12, 2022. https://doi.org/10.1038/s41598-021-04652-3

[10] R. Vyas, T. Kanumuri, G. Sheoran, and P. Dubey, "Accurate feature extraction for multimodal biometrics combining iris and palmprint," *Journal of Ambient Intelligence and Humanized Computing,* vol. 13, pp. 5581–5589, 2022. https://doi.org/10.1007/s12652-021-03190-0

[11] B. A. El-Rahiem, F. E. A. El-Samie, and M. Amin, "Multimodal biometric authentication based on deep fusion of electrocardiogram (ECG) and finger vein," *Multimedia Systems,* vol. 28, pp. 1325–1337, 2022. https://doi.org/10.1007/s00530-021-00810-9

[12] Y. Wang, D. Shi, and W. Zhou, "Convolutional neural network approach based on multimodal biometric system with fusion of face and finger vein features," *Sensors,* vol. 22, no. 16, p. 6039, 2022. https://doi.org/10.3390/s22166039

[13] A. Iula and M. Micucci, "Multimodal biometric recognition based on 3D ultrasound palmprint-hand geometry fusion," *IEEE Access,* vol. 10, pp. 7914–7925, 2022. https://doi.org/10.1109/ACCESS.2022.3143433

[14] Y. Yin, L. Liu, and X. Sun, "SDUMLA-HMT: A multimodal biometric database," in *Biometric Recognition, CCBR 2011*, in Lecture Notes in Computer Science, Z. Sun, J. Lai, X. Chen, and T. Tan, Eds., vol. 7098, Springer, Berlin, Heidelberg, 2011, pp. 260–268. https://doi.org/10.1007/978-3-642-25449-9_33

[15] B. S. Rao, "Dynamic histogram equalization for contrast enhancement for digital images," *Applied Soft Computing,* vol. 89, p. 106114, 2020. https://doi.org/10.1016/j.asoc.2020.106114

[16] F. Jan, N. Min-Allah, S. Agha, I. Usman, and I. Khan, "A robust iris localization scheme for the iris recognition," *Multimedia Tools and Applications,* vol. 80, pp. 4579–4605, 2021. https://doi.org/10.1007/s11042-020-09814-5

[17] Y. Jusman, S. C. Ng, and K. Hasikin, "Performances of proposed normalization algorithm for iris recognition," *International Journal of Advances in Intelligent Informatics*, vol. 6, no. 2, pp. 161–172, 2020. https://doi.org/10.26555/ijain.v6i2.397

[18] D. Bhatt *et al.*, "CNN variants for computer vision: History, architecture, application, challenges and future scope," *Electronics,* vol. 10, no. 20, p. 2470, 2021. https://doi.org/10.3390/electronics10202470

[19] P. Molchanov, S. Tyree, T. Karras, T. Aila, and J. Kautz, "Pruning convolutional neural networks for resource efficient inference," *arXiv preprint arXiv:1611.06440*, 2016. https://doi.org/10.48550/arXiv.1611.06440

[20] R. Nirthika, S. Manivannan, A. Ramanan, and R. Wang, "Pooling in convolutional neural networks for medical image analysis: A survey and an empirical study," *Neural Computing and Applications,* vol. 34, pp. 5321–5347, 2022. https://doi.org/10.1007/s00521-022-06953-8

[21] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *The Journal of Machine Learning Research,* vol. 15, no. 1, pp. 1929–1958, 2014.

[22] R. Eberhart and J. Kennedy, "A new optimizer using particle swarm theory," in *MHS'95. Proceedings of the Sixth International Symposium on Micro Machine and Human Science,* 1995, pp. 39–43. https://doi.org/10.1109/MHS.1995.494215

[23] S. Mirjalili and A. Lewis, "The whale optimization algorithm," *Advances in Engineering Software,* vol. 95, pp. 51–67, 2016. https://doi.org/10.1016/j.advengsoft.2016.01.008

[24] P. Sharma and M. Kaur, "Multimodal classification using feature level fusion and SVM," *International Journal of Computer Applications,* vol. 76, no. 4, pp. 26–32, 2013. https://doi.org/10.5120/13236-0670

[25] R. A. Rasool, "Feature-level vs. score-level fusion in the human identification system," *Applied Computational Intelligence and Soft Computing*, vol. 21, no. 1, pp. 1–10, 2021. https://doi.org/10.1155/2021/6621772

[26] S. A. Haider *et al.*, "An improved multimodal biometric identification system employing score-level fuzzification of finger texture and finger vein biometrics," *Sensors,* vol. 23, no. 24, p. 9706, 2023. https://doi.org/10.3390/s23249706

[27] M. Hammad, Y. Liu, and K. Wang, "Multimodal biometric authentication systems using convolution neural network based on different level fusion of ECG and fingerprint," *IEEE Access*, vol. 7, pp. 26527–26542, 2019. https://doi.org/10.1109/ACCESS.2018.2886573

[28] E. Sujatha, J. S. Jeba Sundar, P. Deivendran, and G. Indumathi, "Multimodal biometric algorithm using iris, finger vein, finger print with hybrid GA, PSO for authentication," in *Data Analytics and Management*, in Lecture Notes on Data Engineering and Communications Technologies, A. Khanna, D. Gupta, Z. Pólkowski, S. Bhattacharyya, and O. Castillo, Eds., vol. 54, Springer Singapore, 2021, pp. 267–283. https://doi.org/10.1007/978-981-15-8335-3_22

11 AUTHORS

Naitik S.T. is a research scholar and is working as an Asst. Professor at the Department of CSE, Dayananda Sagar University, Bengaluru, Karnataka 570028, India with 16 years of experience in Cyber Security. He is actively involved in research and design, innovation in Cyber Security domain. He has done many certifications, among them he is a certified "Cyber Hygiene Practitioner" from ISEA, Gov of India (GOI). He has written many books and book chapters in Cyber Security. He has also published many research articles in peer reviewed journals and conferences. He has got fundings from CySeck, GoK for the project "Deep learning for Deepfake videos detection" and from GoK for KSCST project, "Secure Data transfer using LiFi" (E-mail: naitik.st-cse@dsu.edu.in).

Dr. J.V. Gorabal is a Research Supervisor, with 25 years of teaching experience. He completed Ph.D. from JNTU Anantapur and is currently serving as Professor, Dept. of CSE, ATME, Mysore, Karnataka 570028, India. He has published many articles in peer reviewed journals. He has expertise in Biometric Security, Computer Networks etc., and has supervised many research scholars also (E-mail: dr.jvgorbal_cs@atme.edu.in).