Contents lists available at ScienceDirect

Measurement: Sensors

journal homepage: www.sciencedirect.com/journal/measurement-sensors

Reliable human authentication using AI-based multibiometric image sensor fusion: Assessment of performance in information security

Shambhu Bharadwaj [a,*], Parag Amin [b], D. Janet Ramya [c], Swapnil Parikh [d]

[a] Department of College of Computing Sciences and Information Technology, Teerthanker Mahaveer University, Moradabad, Uttar Pradesh, India
[b] ISME - School of Management & Entrepreneurship, ATLAS SkillTech University, Mumbai, Maharashtra, India
[c] Department of CSIT, School of CS and IT, Jain (deemed to be)University, Karnataka, 562112, India
[d] Department of Computer Science and Engineering, Faculty of Engineering and Technology, Parul Institute of Technology, Parul University, Vadodara, Gujarat, India

ARTICLE INFO

Keywords:
Human authentication
Shifted delta cepstral coefficient (SDCC)
Local binary pattern (LBP)
Fine tuned cuckoo search
Optimized convolutional neural network (FCSO-CNN)

ABSTRACT

Reliable human authentication is essential for assuring the availability, confidentiality, and integrity of sensitive data and resources in the field of information security. Due to the vulnerabilities of conventional authentication techniques like passwords and PINs, interest in biometric-based authentication systems has developed. To get over this, we suggest an efficient voice and iris recognition-based multimodal biometric verification strategy for human authenticating tools. We first collect the voice and iris datasets from 150 people. Then, we used a median filter and high pass filter to preprocess the voice and iris data. The features of voice identification are extracted using the shifted delta cepstral coefficient (SDCC) and the Mel frequency discrete wavelet coefficient (MFCC), and these two coefficients are evaluated. The outcomes of the extraction of iris identification features using Local binary pattern (LBP) and Speeded robust features (SURF) are evaluated. The classifier Fine Tuned Cuckoo Search Optimized Convolutional Neural Network (FCSO-CNN) performs voice and iris recognition modalities. Voice and iris biometric systems may be combined into a single multimodal biometric system by fusing their respective feature sets and scoring algorithms. The results of the computer simulation demonstrate that for speech recognition, using the SDC and MFC coefficients produces better results, while for iris recognition, using the LBP and FCSO-CNN experiment produces superior outcomes. Additionally, the scores fusion works superior to other scenarios in the proposed multimodal biometrics system. The suggested system provides a dependable and strong solution for human identification, assisting in the improvement of security measures across a variety of industries, including financial institutions, governmental agencies, and the defense of key infrastructure.

1. Introduction

Human authentication, in the context of information security, is the practice of assessing a human's credentials before allowing them access to a protected system or data. The accessibility, confidentiality, and secrecy of data and resources may be protected through this procedure by limiting access to only people have been granted permission to use them. The combination of multiple biometric features captured by image sensors to improve the efficacy and dependability of human authentication and recognition systems which is referred to as multibiometric image sensor fusion, and it is an advanced approach used for safeguarding data. Human authentication is a topic that security professionals are thinking about more and more because of how important internet communication and commerce are becoming. The need for trustworthy methods to identify users and create confidence in digital interactions is expanding as cyber-threats, and the chance of unauthorized access to sensitive data become more likely. When it comes to preserving sensitive information, securing online accounts, and thwarting fraudulent activity, strong human authentication methods are crucial [1]. The internet has completely altered the ways in which we interact with one another and run our businesses. However, this growth has also given rise to a variety of challenges, such as the challenge of fostering trust in virtual environments where there are no outward signs of dependability. Human authentication that can be relied on creates a link between the digital and physical worlds, allowing businesses to verify customers' identities with complete confidence [2].

Passwords are widely used as a method of authentication. In order to protect their data and identities, users generate a random string of

[*] Corresponding author.
E-mail address: shambhu012345@outlook.com (S. Bharadwaj).

https://doi.org/10.1016/j.measen.2024.101140
Received 1 August 2023; Received in revised form 22 October 2023; Accepted 25 March 2024
Available online 4 April 2024

characters. However, passwords alone might be risky because they can be broken into if they are weak or easily guessed. Two-factor authentication (2FA) has become popular as a solution to this problem. With two-factor authentication (2FA), users are asked to submit additional proof of identity in the form of a verification code delivered to their registered mobile device. Two-factor authentication (2FA) increases security by using a combination of the user's knowledge (password) and possession (mobile device). Biometric authentication has evolved as a robust and practical approach to establishing an individual's identification [3]. Fingerprints, facial recognition, speech patterns, and iris scans are all examples of biometric identifiers that work in this way. Because these characteristics are so difficult to fabricate or copy, biometrics provides a very high level of security. Biometric authentication systems provide a dependable user experience by precisely matching an individual's biometric data against pre-registered templates using sophisticated algorithms and machine learning approaches. Another robust method to fortify human authentication is multi-factor authentication (MFA). Passwords, biometrics, hardware tokens, and smart cards are just a few examples of what can be used in an MFA setup to create a foolproof authentication system. Numerous-factor authentication (MFA) increases security by requiring users to present numerous pieces of proof before gaining access [4].

An emerging kind of authentication, behavioral analysis looks for consistent actions to verify a user's identity. It is now feasible to construct personalized behavioral profiles by evaluating user inputs, including typing speed, mouse movements, touchscreen gestures, and even the timing of operations. Anomalies and out-of-the-ordinary actions can be spotted with the aid of behavioral analysis, which can alert authorities to potential security breaches. Asking users questions about things they should know, such as their mother's maiden name or their first school, is an example of Knowledge-Based Authentication (KBA). Account recovery and two-factor authentication rely heavily on KBA. However, it has limitations owing to the ease with which sensitive data can be obtained through social engineering or data breaches. So that the authentication process can be trusted, it is essential that KBA systems be carefully designed [5]. In this study, we present a powerful multimodal biometric identification system that integrates iris and speech recognition technologies to authenticate human users. The accuracy and security of human authentication are improved by our suggested method, which makes use of the combination of these two modalities.

These are the sections of the paper: There is a summary of relevant work in Section 2, followed by a detailed discussion of the techniques used for voice recognition and iris identification in Section 3. Section 4 discusses multimodal biometric fusion. The findings of the computer simulation experiments are presented in section 5. The final segment concludes with some thoughts.

2. Related works

The study [6] proposed a safe multimodal biometric system that combines fingerprints and ECG data using a convolutional neural network (CNN). They were aware that it uses convolutional neural networks to combine ECG and fingerprint data for human authentication. The empirical outcomes showed that the suggested multimodal system performs better in terms of effectiveness, robustness, and reliability than cutting-edge multimodal authentication systems. The study [7] offered the "Secure Authentication Management human-centric Scheme (SAMS)" for verifying the identities of mobile phones on the blockchain so that users of MRM can confidently access and use their device-stored resources and data. Data fabrication was evaluated by simulating a malevolent person logging into the SAMS, and the findings proved that it was physically impossible. The study [8] presented a novel face-and-voice detection fusion technique as a powerful multimodal biometric detection solution for human authentication tools. According to simulated studies, using cepstral coefficients and statistical coefficients produces superior outcomes for sound identification, whereas using Eigen face and SVM produces better results for identifying faces. The study [9] proposed an electrooculogram (EOG) based architecture to describe and evaluate the human visual system (HVS) for virtual reality verification, where visual stimuli are created to elicit an HVS response. OcuLock has been shown to be secure against typical assaults, including impersonating and statistics assaults in tests involving 70 participants, with Equal Error Rates of 3.55% and 4.97%, respectively. The study [10] presented BioTouch, a secure re-authentication system that meets these criteria, in this work. The tests proved that BioTouch could detect 98% of abnormal behavior in as few as ten touches, with a maximum accuracy of 99.84%.The study [11] introduced a novel optimal neural network-based biometric image classification approach consisting of three distinct stages: preprocessing, feature extraction, and categorization. The metrics of effectiveness, such as FPR, FNR, sensitivity, specificity, and accuracy, are successfully estimated for the cutting-edge method, which proves highly effective at classifying images. The study [12] proposed an authenticating mechanism to ensure that only authorized users have access to implanted devices during times of crisis. They conducted extensive research into this plan to ensure it provides adequate safety for the patient. They investigated whether or not the device's privacy would be compromised during a wireless exchange of the key. The study [13] offered an innovative solution, integrating cloud support for multi-hop BANs with a lightweight Physical Unclonable Function (PUF)-based authentication mechanism. This novel approach improves data transmission reliability while people are moving around compared to conventional single-hop star networks. This authentication mechanism not only improves data transmission efficiency, but also drastically decreases wasted space and energy. The study [14] looked at the most pressing issues with EEG-based biometric verification. The expansion of online communities and the pursuit of novel approaches to user identification across various internet platforms have sparked a period of intense research and development in this area. The findings demonstrated that the α-rhythm pattern in the resting state with closed eyelids, with the lowest values of the rhythm's coefficients of variance but high within-group dispersion, was the most stable. The study [15] presented a new approach to developing biometric identification systems that use electrocardiogram (ECG) data. The suggested approach begins by using empirical mode decomposition (EMD) to clean up single-lead ECG data. SVM with a cubic kernel performs best in 10-fold cross-validation-based classification assessments, classifying fourteen individuals with an accuracy of 98.7%, specificity of 98.8%, and sensitivity of 100%.

3. Methods

The research presents an innovative multimodal biometric identification approach for human authentication. Fig. 1 depicts the Overall methodology. Integrating many biometric sensors utilizing AI methods results in more reliable and secure authentication and identification systems. Sensors as diverse as recognition of fingerprints, irises, faces, voices, and mobility analyzers might all are combined in this method. Using deep learning and pattern recognition, AI algorithms compile the data from these sensors into a detailed biometric profile that is both accurate and secure. We have proposed the fusion of AI techniques with various biometric sensors, with a specific focus on Voice Recognition and Iris Recognition.

3.1. *Voice recognition*

The use of vocal difference as a stand-alone biometric identification method has been advocated by several scholars. The speech biometric is a low-cost, user-friendly option for authentication because it uses the user's own voice, doesn't require the user to be physically present, and works remotely. Low security, poor accuracy, and difficulties with cross-channel situations are all issues that arise when relying on a person's voice as a biometric authentication method.

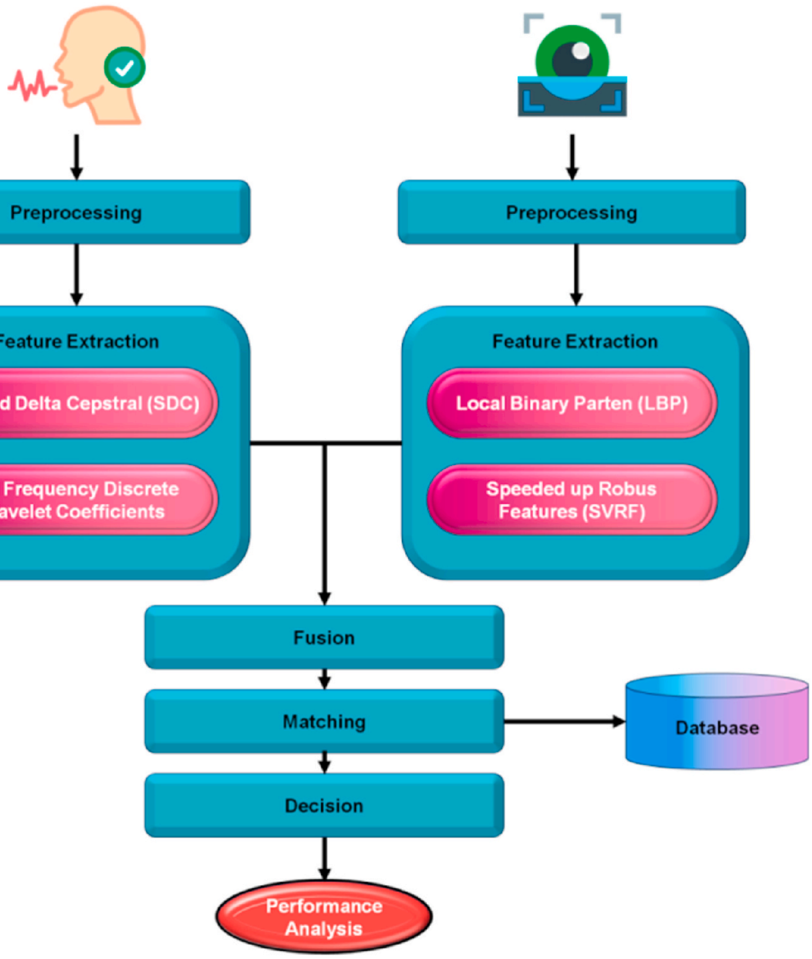
Preprocessing
Preprocessing
Feature Extraction
Shifted Delta Cepstral (SDC)
Mel Frequency Discrete Wavelet Coefficients
Feature Extraction
Local Binary Parten (LBP)
Speeded up Robus Features (SVRF)
Fusion
Matching
Database
Decision
Performance Analysis

Fig. 1. Overall methodology.

As the initial phase in the voice recognition preprocessing, the data using a high-Pass Filter. The second phase consists of coefficients, such as "Mel Frequency Discrete Wavelet Coefficients (MFCCs)" and Shifted Delta Cepstral Coefficients (SDCCs), which provide brief descriptions of a speaker's voice.

3.1.1. High-pass filter preprocessing

Commonly employed in voice recognition systems, high-pass filter preprocessing improves speech signals by filtering out low-frequency elements that aren't necessary for recognition. It functions by passing only signals above a predetermined cutoff frequency while dampening those below it. To improve the quality of the audio input for subsequent voice identification processing, a high-pass filter is applied to filter out low-frequency noise sources such as hums, rumbles, and ambient noise. The filter boosts the signal-to-noise ratio by isolating the vocal range of human speech from the background noise. An essential first step, preprocessing, improves the intelligibility and clarity of the speech signal, making it easier for subsequent speech recognition algorithms to correctly detect and interpret spoken words. Digital signal processing techniques allow for the implementation of a high-pass filter, which is used in conjunction with other preprocessing techniques, such as noise reduction and normalization, to improve voice recognition accuracy.

3.2. Iris recognition

The iris is the colorful part of the eye, and there is biometric technology called iris recognition that can tell people apart based on their iris patterns. In order to do this, high-resolution images of the iris must be captured, and then sophisticated algorithms must be used to isolate the iris's unique characteristics, such as its furrows, ridges, and freckles. Each person's iris pattern is then digitally template to act as their own identification number.

Compared to other biometric approaches, this technology has many benefits. Iris recognition has several potential uses, including in security, immigration, and policing. "Local Binary Pattern (LBP) and Speeded Up Robust Features (SURF)" are employed as iris discrimination algorithms in this work.

3.2.1. Median filter preprocessing

The median filter is an order statistics-based nonlinear spatial filter that greatly reduces salt-and-pepper noise. It swaps out a pixel's value with the middle grayscale value in its immediate vicinity. Consider a noisy input image, *h*, and an *m*» : *n* sub-image, S *wz*. It has its origin at the given coordinates (w, z). The filter's output at a given x and y coordinate is denoted by $e(w,z)$. Then, the expression gives us the 2-dimensional median filter.

$$e(w, z) = \underset{(t,s)\in_{t_{yx}}}{median}\{h(t, s)\} \tag{1}$$

3.2.2. Local Binary Pattern (LBP) feature extraction

The Local Binary Pattern (LBP) technique is used for texture analysis because of its speed and accuracy in capturing pixel intensity interactions. Our technique is interesting since we use LBP for iris recognition. The input iris image is preprocessed and normalized to accommodate for size and location changes before feature extraction and comparison. To account for the possibility of occlusions, we partition the top part of our normalized iris image into 32 blocks with a resolution of 90 × 720 pixels. Each of these blocks is handled as a node in a labeled graph that stands in for the iris pattern, and an LBP histogram with a radius of 2 (59 bins) is calculated for it. The spatial layout of these image blocks is used to establish links between the nodes, resulting in a hierarchical representation, as shows in Fig. 2.

3.2.3. Feature extraction using speeded up robust features (SURF)

The image of the eye is extracted from the face by first finding the pupil's circle and then identifying the region of interest (ROI). Furthermore, the iris-eyelid junction, reflections, and eyelashes are ignored in this model. It is generally agreed that SURF is one of the most effective and reliable face and eye detection algorithms currently available. Our data suggest that the SURF approach is not only three times faster than the Scale-Invariant Feature Transform (SIFT) but also achieves high levels of recall and accuracy. Furthermore, SURF is significantly more useful than multi-poses and rotation in addressing the difficulties associated with face identification. The SURF structure is shown in Fig. 3.

Since the *GJO* matrix governs the localization procedure, it follows that the SURF detection method relies on the *GJO*, as given in Eq. (2).

$$GJO = \begin{bmatrix} K_{ww} & K_{wz} \\ K_{zw} & K_{zz} \end{bmatrix} \tag{2}$$

where Lxx stands for the Gaussian Laplacian of pictures captured by the eye. Convolution of a Gaussian with the derivative of the repaired eye templates at the second order is used to calculate the HIP. In Eq. (2), C_{ww}, C_{zz}, C_{zz}, and C_{wz} are used to estimate the values of the *GJO* parameters K_{ww}, K_{wz}, K_{zw}, and K_{zz}, respectively, such that:

$$Det(GJO_{apprax}) = C_{ww}C_{zz} - (xC_{wz})^2 \tag{3}$$

where x is the weight assigned to the rectangular area, low-quality pictures and noise removal using SURF and spatial filter analysis are prerequisites to the matching process. The results showed that when compared to the SIFT method, SURF is much quicker. Because the Trace T of the Hessian matrix contains information about the most salient ocular aspects of eye pictures, as in Eq. (3), SURF can speed up the detection process by using rapid indexing via the Laplacian parameters.

$$S = K_{ww} + K_{zz} \tag{4}$$

3.3. Fine-tuned cuckoo search optimization (FCSO)

The CS-introduced parameters O_b, λ, and α aid the algorithm in locating both globally and locally optimal solutions. When optimizing solution vectors, O_b and α played a crucial role and could be used to modify the algorithm's convergence speed. Both O_b and α have fixed values in the classic CS method. These parameters are fixed at the beginning of the process and will not be modified in subsequent generations. One major issue is the high number of iterations required to reach a satisfactory result. When O_b is little and is large, α the algorithm's performance suffers, and the number of iterations grows dramatically. Convergence occurs quickly but may not be optimal if O_b is set to a large number while α is set to a small one.

The method of modifying O_b and α distinguishes the ICS from the CS. In contrast to the CS algorithm, which employs hardcoded values for O_b and α, the ICS method allows for these parameters to be adjusted to optimize performance. In the first few generations, O_b and α need to be sufficiently large for the algorithm to compel a rise in the variety of solution vectors. For better fine-tuning of solution vectors, however, these values should be lowered in the final generations. In Eqs. (5)–(7), where *MJ* and *hm* Indicate both the overall iteration count and the

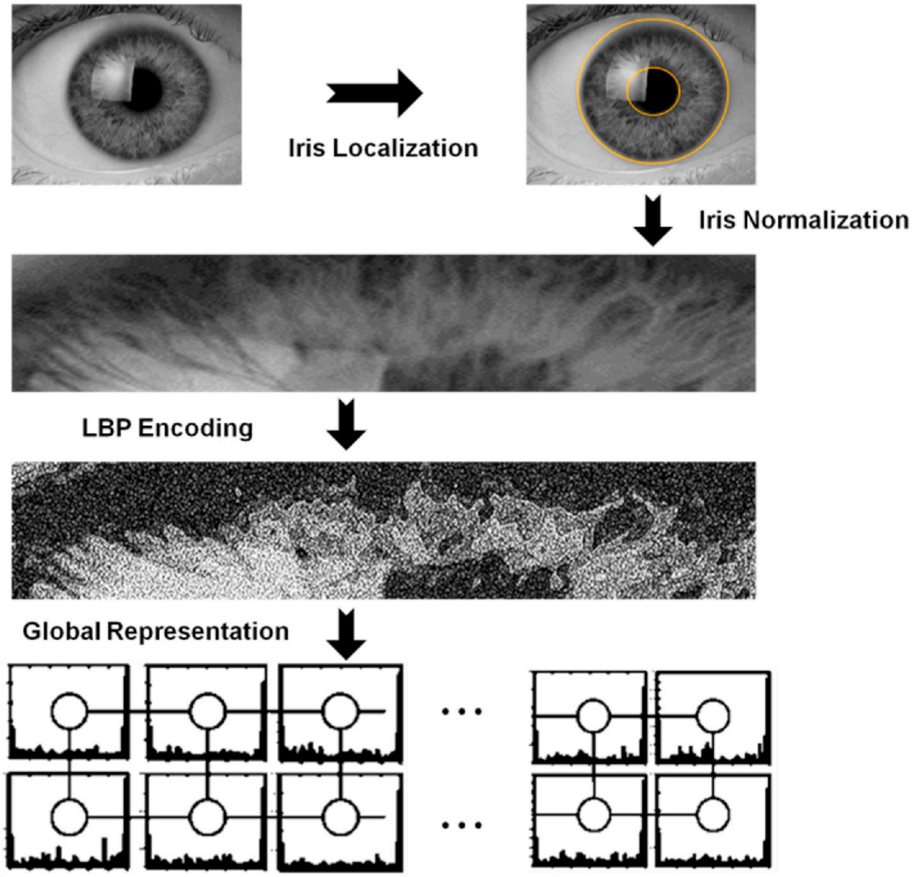

Fig. 2. The flowchart of the LBP.

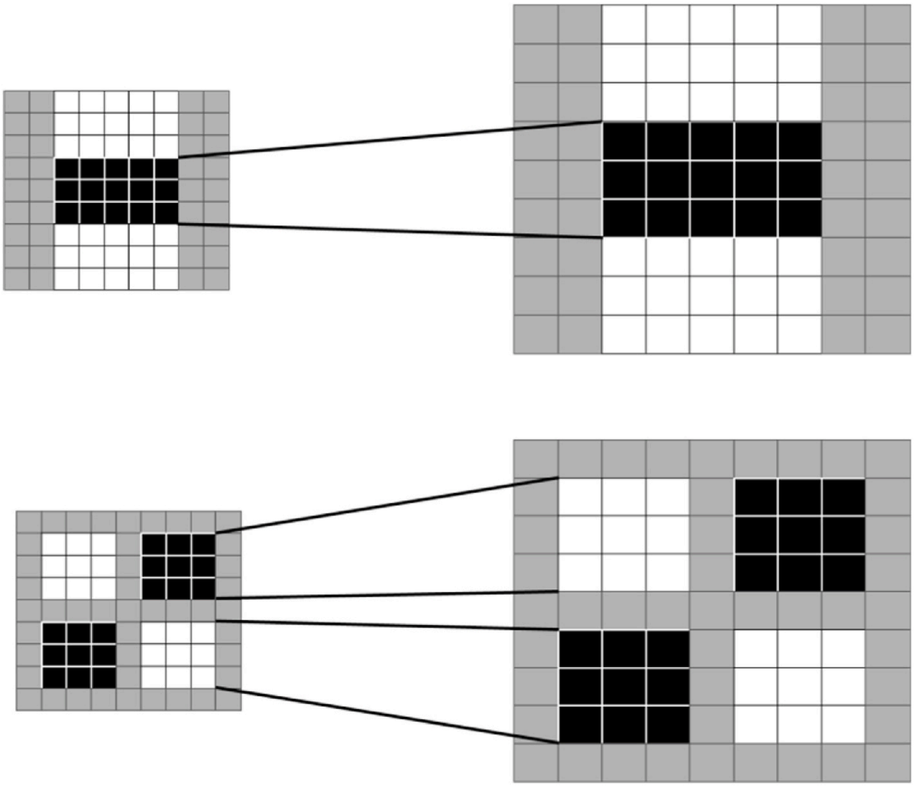

Fig. 3. Structure of speeded up Robust Features.

current iteration number, O_b and α undergo dynamic changes with the number of generations.

$$O_b(hm) = O_{b_{\max}} - \frac{hm}{MJ}\left(O_{b_{\max}} - O_{b_{\max}}\right) \tag{5}$$

$$\alpha(hm) = \alpha_{max}\exp(d.hm) \tag{6}$$

$$d = \frac{1}{MJ}Km\left(\frac{\alpha_{min}}{\alpha_{max}}\right) \tag{7}$$

3.4. Convolutional neural networks (CNN)

Pattern recognition is the main use of CNNs, commonly referred to as multi-layer neural networks. The algorithm's tolerance for slight input changes is increased using CNNs. Pattern recognition is a subfield of AI that attempts to discover and analyze characteristics and patterns in information. It includes the automatic identification of patterns and characteristics in datasets, which in turn enables the system to generate inferences, estimations, and categorization. The CNN deep learning technology is frequently utilized in image recognition and processing. They are modeled after the visual neural of people and are programmed to automatically and adaptively learn the structure of visual components in a specific imagine space. CNN is used to extract features and recognize patterns from visual and audio data for use in speech and iris recognition systems. By processing spectrograms or other time-frequency representations of audio signals, CNNs may be utilized for speech recognition to extract relevant characteristics for identifying and authenticating voices. CNNs are able to analyze the iris's intricate patterns and extract specific data for use in authentication, making iris recognition possible. Equations used in CNN often represent the calculations made at the convolutional and pooling layers. CNNs rely on a number of basic eqs. (8) and (9) are:

$$S(i,j) = (K*I)(i,j) = \sum\nolimits_m \sum\nolimits_n I(i-m, j-n)Km, n\Bigg) \tag{8}$$

The final feature map, denoted by S, Input picture is denoted by For convolution, K is the kernel. The point indices are denoted as i, j. The kernel's indices are m and n.

$$Y(i,j) = Pool(X, i, j) \tag{9}$$

Y represents the result of the pooling process. When X is the input, Both i and j represent geographical indices.

A minimal amount of pre-treatment is needed for operations that don't require a certain type of extractor. By connecting the convolution and aggregation layers, the recommended topology incorporates concepts from several types of DNN. Despite the fact that these DNNs are based on Hubel and Wiesel's studies of the cat's primary visual brain, they truly belong to a different class of models. The first component of the design, which is made up of a succession of convolution and aggregation layers, deals with automatic feature extraction, while the second section, which is composed of fully linked layers of neurons, deals with categorization.

3.4.1. Convolution layer

The number of convolution maps N in $\mathrm{N}_\mathrm{i}^\mathrm{j}(\mathrm{i}\epsilon\{1,\ldots,\mathrm{M}\})$, the size of the convolution cores $L_w \times L_z$ (usually square), and the connection method K^{j-1} are the parameters that define the convolutional layer $D_j(networklayerj)$. Each convolution map N_i^j is the result of multiplying the convolution core by the sum of the convolutions of the previous layer's maps, N_i^{j-1}. The outcome is fed into a bias a_i^j and a nonlinear transfer function $\Phi(w)$. The map of the fully linked layer can be derived using the expression shown below:

$$N_i^j = \Phi\left(a_i^j + \sum_{m=1}^{M} N_i^{j-1} * L_m^j\right) \tag{10}$$

3.4.2. Transfer learning

The goal of transfer learning is to improve the performance of a learning machine by using the expertise and results of another learning machine. Transfer learning, which is applying the information that one model learned while it was solving a specific problem to another problem, is more frequently used to take ConvNet models that are previously trained and rehabilitate them for the task. Learning that can be transferred comes in two flavors:

- Variable extraction from the ConvNet: in this scenario, the ConvNet acts as an extractor, meaning that a vector is taken from a certain layer of the model without changing its structure or weights and then applied to a new location.
- This is where the pre-trained model's structure and weights are applied to a freshly created ConvNet for fine-tuning. After making some tweaks to the pre-trained model's structure, a fresh model can be taught to do the new task.

3.4.3. Max-aggregation layer

Subsampling layers often come after the convolutional layers in traditional neural network topologies. Cards are shrunk using a subsampling layer, which also makes them invariant to any input rotations or translations. A max-pooling layer is a variation on a layer that has been proven to be superior in specific situations. The max-aggregation layer outputs the input layer's maximum activation value across non-overlapping sections of size $L_w \times L_z$.

3.4.4. Classic neural layer

The final layer's activation cards are set to 1, which parameterizes the convolution and pooling layers to produce a 1D attribute vector. The classification is then carried out by adding classical layers of neurons to the network, which are fully linked. In supervised learning, the classes may have any number of neurons that have been predetermined as being useful.

3.5. Fine tuned cuckoo search optimized convolutional neural network (FCSO-CNN)

Biometric authentication's unique ability to improve security while also improving efficiency has been the subject of intense research in recent years. The use of CNN for feature extraction and data classification from biometrics is an exciting and promising development in this field. CNN performance, however, is extremely sensitive to architecture and hyperparameter choices. To address this issue, we developed a novel method called Fine-tuned Cuckoo search optimization (FCSO) that is optimized for CNNs in the context of biometric authentication. FCSO is an optimization algorithm that takes its cue from the cuckoo bird, which is known for laying its eggs in the nests of other birds.

First, the FCSO-CNN framework randomizes CNN architecture and hyperparameters. The FCSO algorithm then evaluates each option to find the best arrangement. Cuckoos, symbolizing solutions, build nests in diverse search spaces and lay eggs, signifying new candidate configurations. In subsequent cycles, eggs with greater fitness values (better configurations) are more likely to be recognized and used. Levy flights, random walks, and Lévy flights plus random walks are used by the FCSO algorithm to survey the search space and avoid stalling at local maxima. The FCSO technique may repeatedly refine the CNN's architecture and hyperparameters to learn discriminative features from biometric data, improving its biometric authentication accuracy and durability. The FCSO-CNN approach works well for face recognition, fingerprint identification, and iris verification. Its ability to fine-tune CNN designs and optimize hyperparameters makes it outperform ordinary CNN models. The FCSO-CNN architecture is powerful and adaptable for biometric authentication systems.

3.6. AI-based multibiometric image sensor fusion

Multibiometric image sensor fusion systems that utilize AI require adaptability, which allows the system to dynamically respond to changing environment and user requirements. These systems must be prepared to respond for differences in sensor quality, external conditions, and individual user preferences. Such systems could continually develop and modify themselves to increase efficiency and endurance through the use of AI algorithms. For instance, in real-time applications, flexibility enables the system to modify identification thresholds, fusion weights, and feature extraction strategies in response to changes in the environment or the user's biometric features. In addition, it's essential in providing for future sensor technologies and maintaining compatibility, both of which contribute to the system's stability. In addition, adaptation is critical for overcoming security issues since it allows the system to quickly respond to potential dangers or assaults by changing its settings or providing extra levels of identification. As a whole, adaptability improves the range, confidence, and stability of AI-based Multibiometric image sensor fusion systems, enabling them to perform better in more types of real-world scenarios.

3.7. Multimodal biometric fusion

In multimodal biometric systems, prior to the matching process, fusion may entail the integration of data derived from multiple sensors or sets of features. After matching, fusion may involve combining the results of multiple matches. If the biometric system employs many sensors to measure the same feature, the sensors are fused. When doing feature fusion, numerous feature vectors taken from various biometric systems are combined into a single feature vector. The decision to combine scores or decisions after the matching procedure is open for debate. Using a formula like the Likelihood Ratio (LLR) or a rule like the sum, max, and min rule, we can fuse many matching scores into a single score. When results from various matching methods are at hand, the weakest fusion is used to make a final judgment.

In this study, we use and compare two fusion techniques: score fusion and feature fusion. The voice signal and iris image feature vectors are amalgamated into a unified feature vector, which is then compared against the enrollment template, resulting in a final matching score as an integrated biometric system. Fig. 4a shown as the result of the feature fusion. In Fig. 4b, we see the LLR formula for fusing scores, which yields the following equation (Eq. (11)):

$$T = \frac{o(T_{voice}|H) \bullet o(T_{iris}|G)}{o(T_{voice}|J) \bullet o(T_{iris}|J)} \tag{11}$$

In this study, $o(.|H)$ is the probability density function of matching scores for authentic persons, while $o(.|J)$ is the corresponding function for imposters. The T voice variable represents the voice recognition matching score, while the T iris variable reflects the iris identification matching score.

4. Result and discussion

To evaluate the performance of the proposed multimodal biometric system, data on the voices and irises of 150 people were collected. We gathered four iris photos from each participant, for a total of 600 irises, and had each person say a word four times, for a total of 600 voice signals. Iris images are scaled to 720 × 720 pixels in an RGB color model and voice signals are captured at 10 kHz over 5s. Xiaomi tabletsix cameras and a regular microphone used to collect the data.

The ROC curve and the "Equal Error Rate (EER)" have been used to assess the efficacy of the proposed method. A "receiver operating characteristic (ROC)" curve is generated by comparing the FAR and FRR. The "False Acceptance Rate (FAR) and False Rejection Rate (FRR)" both represent the percentage of legitimate studies that were incorrectly labeled as fake ones. When the FAR and the FRR is both equal, we have

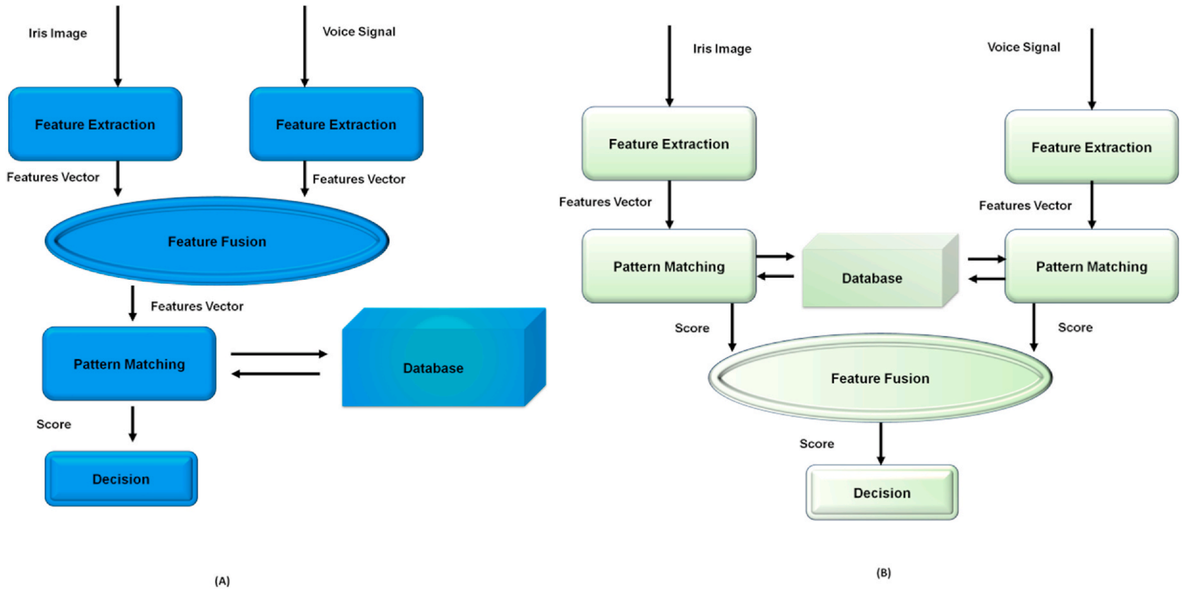

Fig. 4. The suggested multimodal biometric fusion technique block diagram (a) feature fusion; (b) LLR formula.

an EER, as shown in Eq. (12).

$$EER = \frac{FRR + FAR}{2}, whenFAR = FRR \tag{12}$$

Five hundred voice training signals have three people each. The SDC and MFCC characteristics optimize voice recognition. Voice recognition mode tests the remaining 100 voice signals (2 per participant). AWGN-degraded voice samples test the proposed approach. Fig. 5 shows several feature vector voice categorization ROC curves. EER values at the ROC-EEL junction are shown. Table 1 compares voice recognition feature vector EER values. MFCC features have the lowest EER (Fig. 5 and Table 1). The spectrum disentangles periodic temporal patterns in the waveform, and the MFCC's features assign those periodicities and frequent patterns to one or two distinct elements in the MFCC, therefore decomposing the harmonic series. This work uses SDCC and voice timbre parameters to teach and test the voice distinction technique, which increases performance.

In iris recognition computer experiments, 300 iris images (2 images/individual) are used for training, and 300 iris images (2 images/individual) are used for testing LBP and SURF to determine which method produces the best results. For testing, some irisimages are degraded using JPEG compression. Fig. 6 displays iris recognition ROCs curves. EER values are compared in Table 2. Table 2 and Fig. 6 display iris recognition results. These results show that the LBP with the FCSO-CNNiris recognition approach has a lower FAR and FRR than the other methods and a lower EER; hence it is used for iris recognition in this research. The LBP combined with FCSO-CNN is more robust since it optimizes class dispersion while minimizing it within class to determine the best projective direction. The two fusion process is shown in Fig. 7. This graphic displays the ROCs curves for speech recognition, iris identification, and after fusion utilizing features and scores.

Table 1
Different feature vectors are used by EER for voice recognition.

Voice Recognition Method	ERR (%)
MFC Coefficient + FCSO-CNN	3.89
SDC coefficient + FCSO-CNN	8.59

When comparing scores fusion and features fusion, the EER for the

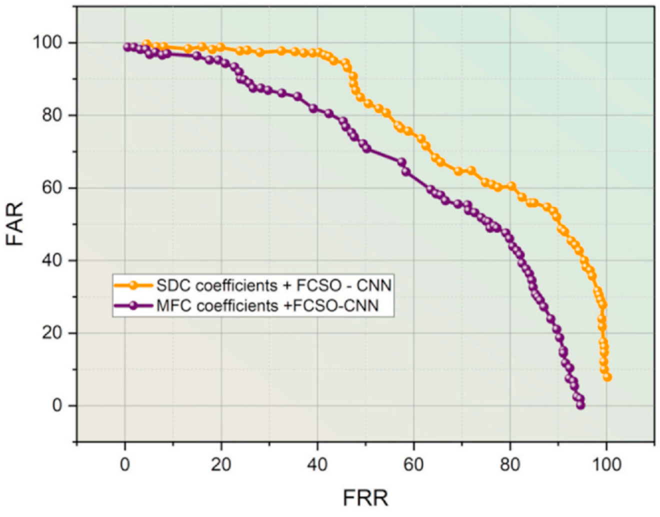

Fig. 5. Voice recognition ROC curves with several feature vectors.

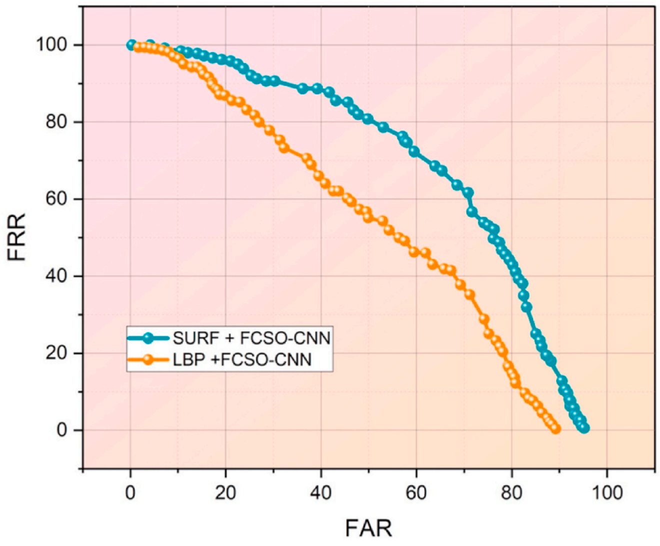

Fig. 6. Comparison of the algorithms' ROC curves for iris recognition.

Table 2
EER for the various iris-identification techniques.

Iris Recognition Method	ERR (%)
LBP + FCSO-CNN	2.54
SURF + FCSO-CNN	5.56

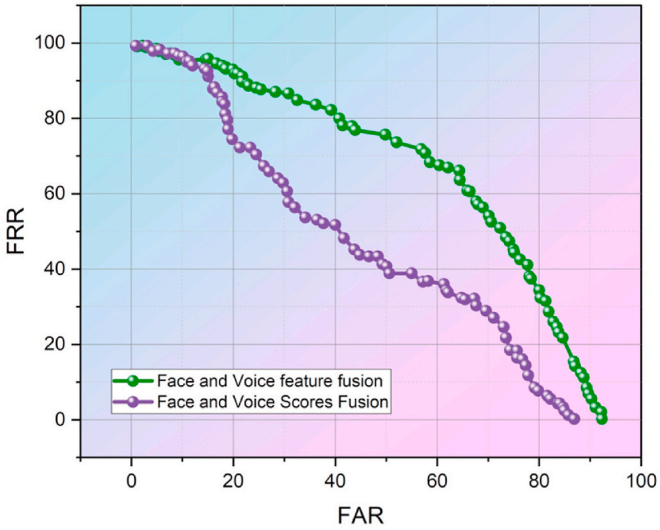

Fig. 7. Comparison of the suggested multimodal fusion method's ROC curves.

former is 0.71 while the latter is 1.70. Scores fusion is preferable to other methods since it can take use of the differences between biometric features to improve accuracy when identifying users. The genuine-impostor LLR minimizes mistake probability. Results show that the proposed multimodal fusion strategy is promising. Fig. 7 indicates that the multimodal system reduces EER.

5. Conclusion

We propose a fusion-based multimodal biometric identification system that uses both voice and iris recognition to verify a person's identity. The best features for voice and iris recognition are extracted using separate feature extraction software. The findings of the voice recognition process demonstrated that simulating the MFCC utilizing an FCSO-CNN classifier scenario produces the best results. According to the results of the iris identification technique, the LBP with FCSO-CNN classifier-based iris differentiation approach is the most effective iris identification methodology out of those that were looked at. The multimodal fusion technique is a promising one, even though it produces the lowest EER, according to the fusion results which were presented. In comparison to existing biometric techniques, the suggested scheme performs superior. The results of the computer simulation experiments show that the suggested modal is superior to the suggested modal for iris identification and the suggested fusion scenarios. Future research will examine triple multimodal employing the iris, face, and speech. One multimodal biometric technique will incorporate three different biometrics. One of the suggested system's limitations is that requires user participation to capture information about the user's voice and iris. Because of this issue, the system might not function also in situations when users are unable or unwilling to provide biometric information. To make this multimodal biometric system even more secure and appropriate for more types of applications, including IoT devices, further research may investigate the integration of new biometric modalities and complicated AI algorithms to boost safety and adaptability.

Code availability

Not applicable.

Ethics approval

Not applicable.

Consent to participate

Not applicable.

Consent for publication

Not applicable.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

Acknowledgments

There is no funding.

References

[1] M. Ahmed, M. Roushdy, A.B. Salem, Intelligent technique for human authentication using FUSION OF palm and finger veins, in: Computer Systems and Information Technologies, 2021, pp. 11–19, 2.
[2] E. Sujatha, A.C. Nil, Multimodal biometric authentication algorithm at score level fusion using hybrid optimization, Wireless Commun. Technol. 2 (1) (2018) 1–12.
[3] M. Hammad, Y. Liu, K. Wang, Multimodal biometric authentication systems using convolution neural networks based on different level fusion of ECG and fingerprint, IEEE Access 7 (2018) 26527–26542.
[4] B. Zhou, J. Lohokare, R. Gao, F. Ye, EchoPrint: two-factor authentication using acoustics and vision on smartphones, October, in: Proceedings of the 24th Annual International Conference on Mobile Computing and Networking, 2018, pp. 321–336.
[5] Z. Zhao, Y. Zhang, Y. Deng, X. Zhang, ECG authentication system design incorporating a convolutional neural network and generalized S-Transformation, Comput. Biol. Med. 102 (2018) 168–179.
[6] M. Hammad, K. Wang, Parallel score fusion of ECG and fingerprint for human authentication based on convolution neural network, Comput. Secur. 81 (2019) 107–122.
[7] H.W. Kim, Y.S. Jeong, Secure authentication-management human-centric scheme for trusting personal resource information on mobile cloud computing with blockchain, Human-Centric Comput. Inform. Sci. 8 (1) (2018) 1–13.
[8] A. Abozaid, A. Haggag, H. Kasban, M. Eltokhy, Multimodal biometric scheme for human authentication technique based on voice and face recognition fusion, Multimed. Tool. Appl. 78 (2019) 16345–16361.
[9] S. Luo, A. Nguyen, C. Song, F. Lin, W. Xu, Z. Yan, OcuLock: exploring human visual system for authentication in virtual reality head-mounted display, January, in: 2020 Network and Distributed System Security Symposium (NDSS), 2020.
[10] C. Zhang, S. Li, Y. Song, Q. Meng, L. Lu, M. Hou, BioTouch: reliable Re-authentication via finger bio-capacitance and touching behavior, Sensors 22 (9) (2022) 3583.
[11] P.S. Chanukya, T.K. Thivakaran, Multimodal biometric cryptosystem for human authentication using fingerprint and ear, Multimed. Tool. Appl. 79 (2020) 659–673.
[12] T. Belkhouja, X. Du, A. Mohamed, A.K. Al-Ali, M. Guizani, Biometric-based authentication scheme for implantable medical devices during emergency situations, Future Generat. Comput. Syst. 98 (2019) 109–119.
[13] X. Tan, J. Zhang, Y. Zhang, Z. Qin, Y. Ding, X. Wang, A PUF-based and cloud-assisted lightweight authentication for multi-hop body area network, Tsinghua Sci. Technol. 26 (1) (2020) 36–47.
[14] N.N. Lebedeva, E.D. Karimova, Stability of human EEG patterns in different tasks: the person authentication problem, Neurosci. Behav. Physiol. 50 (2020) 874–880.
[15] S. Aziz, M.U. Khan, Z.A. Choudhry, A. Aymin, A. Usman, ECG-based biometric authentication using empirical mode decomposition and support vector machines,

October, in: 2019 IEEE 10th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), IEEE, 2019, pp. 906–912.