A Novel Approach Towards Biometric Recognition System Using Voice and Signature

Shoaib Ahmed[1,*], Aneela Pathan[2], Fareesa Khan Sohu[2], Syed Haseeb Shah[2], Rizwan Aziz Mangi[2], Tayyaba Abbasi[2],

[1]Oil & Gas Regulatory Authority (OGRA), Islamabad, Pakistan
[2]Department of Electronic Engineering, University of Larkana, Larkana, Pakistan
[*]Corresponding author: sshoaib@ogra.org.pk

Abstract

Recently, recognition systems have gained importance due to their vital role in security by identifying users while interacting with electronic devices to ensure reliability. The typical authentication systems based on PINs or passwords have revealed various issues of deceitful access. On the contrary, Biometrics can establish an apparent discrepancy between a genuine individual and a fraudulent imitator. Biometrics may be categorized as a Single Model or Multimodal System. Systems of Multimodal bio-metrics are used in Physical Access, Civil ID, Criminal ID, Network/PC Access, Kiosk/ATM, Retail/POS, Surveillance, e-commerce, and telephony. This paper proposes a novel and robust recognition system that uses voice and signatures as inputs to recognize the users. The proposed method uses (Mel Frequency Cepstral Coefficients) MFCC and (Vector Quantization) VQ as characteristic vectors to perform voice analysis, and Vertical Projection Profile (VPP), Horizontal Projection Profile (HPP), and Discrete Cosine Transform (DCT) features to design an offline signature identification system. In the design, a similarity score is obtained. False acceptance and rejection probabilities are measured based on the highest score for each uni-model system. Finally, both methods are merged to get an equal error rate, which is used to evaluate the effectiveness of the proposed approach. The results indicate that when compared with unimodal biometric systems, the proposed multi-model biometric system produces less EER, making the system more robust and reliable.

Keywords—Person Recognition, Voice Analysis, Multimodal system, Biometrics, Security

1 Introduction

Biometric recognition technology has made its way into many aspects of daily life in a relatively short period [1, 2]. Citizens of more than 50 nations carry machine-readable passports. They contain biometric data [3], facial images [4], and, in most cases, a digital representation of fingerprints to authenticate identity at the border [5]. Law enforcement agencies have built biometric databases containing fingerprints, voices, and DNA samples to make their work more efficient and controllable [6]. Biometrics are rapidly used in distant tele-biometric deployments, such as eCommerce and online banking, to supplement or replace traditional authentication techniques, such as PINs and passwords in commercial applications [7]. While biometrics-based authentication provides advantages over these methods, several drawbacks may still be observed in biometrics-based authentication systems. One of the potential solutions may be multi-model biometrics, where more than one biometric feature is incorporated.

Multimodal biometrics is associated with incorporating two or more biometric modalities in an exacting recognition method [8, 9].

Biometrics face recognition technology has gained a lot of attraction in the current practice of using human faces for verification drives [10]. Facial recognition captures the spatial geometry of various facial features [11].

Besides, personal identification is obtained by the use of one's voice. Voice biometrics is a natural and pervasive behavioral trait that current personal computers can collect [12].

ISSN: 2523-0379 (Online), ISSN: 1605-8607 (Print)
DOI: https://doi.org/10.52584/QRJ.2102.06

Multi-model biometric systems with different characteristic combinations are seen in various publications. A review of some of the existing multi-model biometric systems is reported in the preceding section.

One of the multi-model biometric systems based on finger-vein and finger-shape modalities is reported in [13]. The work is CNN-based, where to capture the figure print near-infrared camera sensor is used. The weighted sum, product, and perception methods are used to calculate the matching distance.

Based on deep learning temple matching, a biometric recognition system using three distinct features, face, fingerprint, and iris, is reported in [14]. In the proposed system, well-known methods of contourlet transform and lo-cal derivative ternary methods are used to extract the features. In design, weighted rank-level algorithms extracted the features.

A facial feature extraction using deep belief network (DBN) and left and right iris feature extraction and classification based on CNN is reported in [15]. The scores of both features are combined to provide rank-level fusion methods.

A CNN–based deep learning framework for face information-based recognition is given in [16]. It is observed that deep CNN models resulted in poor performance when insufficient training data was given.

A combination of iris and periocular region modalities-based biometric systems based on the deep transfer learning method is given in [17]. The proposed work used the VGG model to extract the feature and the swarm algorithm. The two modalities relied on matching score-level and feature-level fusion.

Several deep learning models with transfer learning for an automated plankton recognition system are reported in [18], and in [19], a method based on finger vein recognition is proposed.

A multi-model, iris, and left iris biometrics recognition is reported in [20, 21]. The work proposed a deep learning method based on a multi-class support vector machine (SVM) algorithm to extract and classify iris features.

A finger vein and finger knuckle print features using three CNNs, alexnet, VGG16, and resnet50, interacting with transfer learning, were proposed in [21]. Three of the methods were combined to produce the classifications.

An optimal human motion dynamic-based biometric system is reported in [22], which developed a CNN-based deep transfer learning for the human identification framework.

This paper proposes a novel and robust recognition system that uses voice and thumb signatures as inputs to recognize the users. The proposed method uses (Mel Frequency Cepstral Coefficients) MFCC and (Vector Quantization) VQ as characteristic vectors to perform voice analysis. Vertical Projection Profile (VPP), Horizontal Projection Profile (HPP), and Discrete Cosine Transform (DCT) features to design an offline signature identification system. In the design, a similarity score is obtained. False acceptance and rejection probabilities are measured based on the highest score for each uni-model system. Finally, both methods are merged to get an equal error rate, which is used to evaluate the effectiveness of the proposed approach.

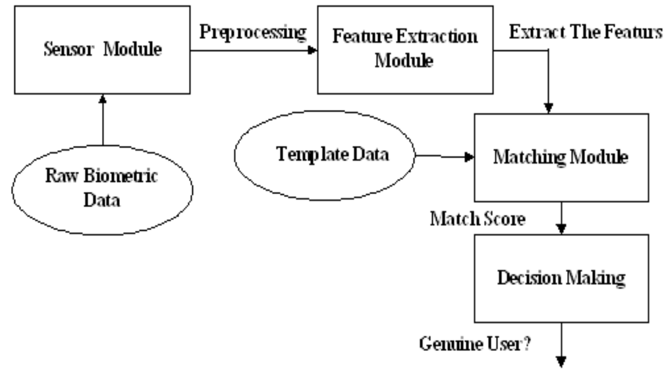

Fig. 1: Fundamental Modules in Biometric System

The envisioned biometric system (Fig.1) consists of four fundamental modules: Sensor-module, feature-extraction segment, matching segment, and decision segment [22].

- **Sensor module:** The sensor module contains the biometric sensor or scanner to measure or record the biometric data that needs to be transferred to the feature extraction module for further processing. A biometric system's overall cost depends most on the sensor module [23].
- **Feature extraction module:** The feature extraction module consists of various mathematical algorithms and techniques to process the raw data obtained by the sensors. The desired data is separated and sent to the matching module to perform further actions [24].
- **Matching Module:** The matching module provides a match score when the data from the future extracted module is compared with the stored template in the database. The score received from the matching module is sent to the decision-making module to see if the required threshold levels are generated [25].
- **Decision-making module:** As the name sug-

gests, the decision-making module tells whether the user is genuine. These are used to either validate a person's identity or provide a ranking of the enrolled identities for identifying an individual [26].

The paper further proceeds as follows: in section 2, some of the preliminaries related to the paper are given. In section 3, the proposed architecture of a multi-level biometric system is presented, and in section 4, results and discussions are provided.

2 Preliminaries

This section provides some elementary information, which this article will repeatedly use.

- **Biometric Functionalities:** The very initial function of any bio-metric system should be identification and verification. The enrollment and authentication stage tasks fall under verification and identification methods.
- **Enrollment:** Enrollment is assigning a user to the bio-metrics system.
  - **Verification (one-to-one correspondence):** In verification, an operator's distinctiveness is established by bio-metrics available measurements that are contrasted with the defined user's model. As a result, biometric information is gained, pre-processed, and turned into structured features, and the generated score is evaluated using a standard threshold.
  - **Identification (one-to-many matching):** As the name suggests, bio-metrics statistics are collected (pre-processed) through a systematic set of actions, re-formed into elements, and then post-processed before being matched with all appropriate and valuable models.
- **Voice Recognition:** Voice identification is identifying speakers using narrator voices.
- **Pre-processing and Silence Removal:** The noise is added in voice transmissions. As a response, in voice analysis, a subtle reduction approach must be used first to detect "clean" voice segments.
- **Feature extraction using MFCC:** The spectral component of the human acoustic system cannot be detected on a linear scale. So, the Mel-frequency scale is a nonlinear scale to produce spectral information.

  Feature Extraction from Signature Information: DCT (Discrete Cosine Transform), HPP (Horizontal Projection Profile), and VPP (Vertical Projection Profile) are used to extract offline and online features in the signature identification method.
- **Vertical Projection Profile (VPP):** The number of columns in the signature image equals the size of the VPP. The intensities around which the picture pixels are focused are represented by the VPP histogram.
- **Horizontal Projection Profile (HPP):** An array containing the total pixels in each signature image's row. The HPP is the same size as the signature image's number of rows. HPP is a type of histogram as well.
- **Voice Corpus:** Total number of voice samples used in the system.
- **Signature Corpus:** Total number of signature samples used in the system.
- **False Reject Rate or False Non-Match Rate (FRR or FNMR):** The prospect discovers a structurally failed match between the input prototype and a matching model in the database (a record).
- **FAR (The False Acceptance Rate) or FMR (False Match Rate):** A false acceptance rate is the possibility of the system incorrectly matching the input prototype to a non-matching model in the database (a record).
- **Equal Error Rate or Cross over Error Rate (EER or CER):** The rate at which both accepted and rejected mistakes become equal is called the EER or CER (Equal Error Rate or Crossover Error Rate).

3 Proposed Architecture of Multi-level Biometric System

The proposed architecture of the multi-biometric system is given in Fig. 2. We need to perform four mandatory steps in speech or signature biometrics. The first step is enrollment. The second is feature extraction, the third is modeling, and the fourth is normalization.

The proposed system is a multi-model biometric system consisting of speech and signature recognition. In both cases, first, the user is enrolled in the database. Once enrollment is done, the features from the registered data are extracted to perform necessary operations and to produce biometric individuality.

The MFCC technique for voice and VPP-HPP-DCT technique for signature are used in the proposed design for feature extraction.

After features are extracted, modeling is performed by comparing the feature data with online available

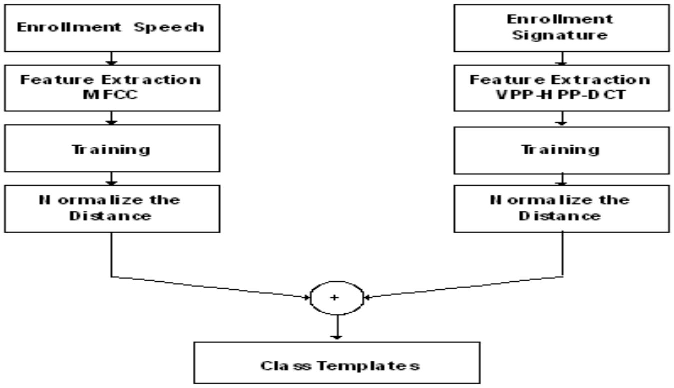

Fig. 2: Proposed Architecture of Multi-model Biometric System

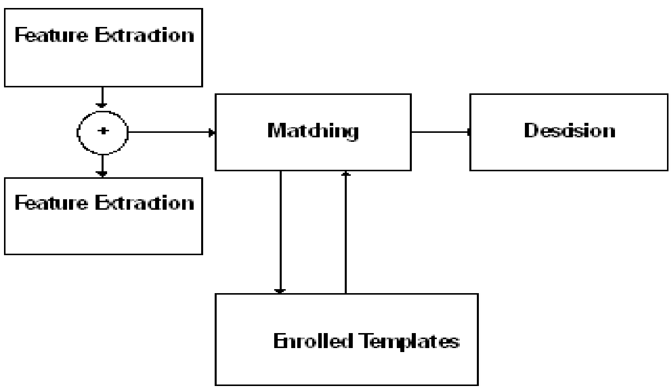

Fig. 3: Elaborated Architecture of Multi-model Biometric System

data-based. The comparison produces the closest match if one is available.

In the last, identical scores that everyone's bio-metrics attribute has produced are being normalized. Obtain a standard matching score for all enrollments. The goal is to achieve the maximum average score.

A further elaboration of the proposed multi-model biometric system is provided in Fig 3.

The Pseudo code for the proposed multi-model biometric system is given below.

Input: Signature Image, Sample Voice Output: Except or Reject

Algorithm 1

```
Input: Signature Image, Voice Sample
Except or Reject
  Begin
  Input: Signature Image
  Output: Signature match Score
    Begin
    Input Signature Image
    Image Processing
    Vertical Projection Profile (VPP)
    Horizontal Projection Profile (HPP)
    Feature Matching For Signature
    Feature Extraction from Signature
      Perfect Match
      False Reject Rate
      False Acceptance Rate
      Equal Error Rate
    End
  Input: Voice Sample
  Output: Voice match score
    Begin
    Input Voice Sample
    Voice Processing
    Mel Frequency Cepstral Coefficients
    Feature Extraction from Voice
    Feature Matching for voice
      Perfect Match
      False Reject Rate
      False Acceptance Rate
      Equal Error Rate
    End
Normalize (Signature match score, Voice match score)
  End
```

3.1 Modeling in Speech Biometrics System

After enrollment, the next step is feature extraction. To extract them, we need Frequency Cepstral Coefficients (FCC). The spectral coefficients obtained from FCC do not regulate energy. As a result, we raise the energy factor by including Double Delta acceleration. The feature extraction method is depicted in the figure below.

Two well-known models in speech recognition systems are the generator network model and the discriminative model. The generative model aims to encapsulate the underlying distribution of training data, such as class centroids and variation around the centroid. Stochastic models, such as Gaussian Mixture Models (GMM), Hidden Markov Models (HMM), and others, are the most prevalent generative models in speech recognition. On the other hand, exclusionary models

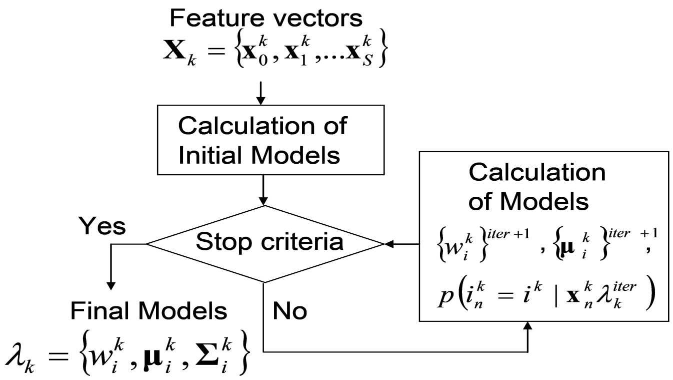

Fig. 4: Feature Extraction Method Based on FCC

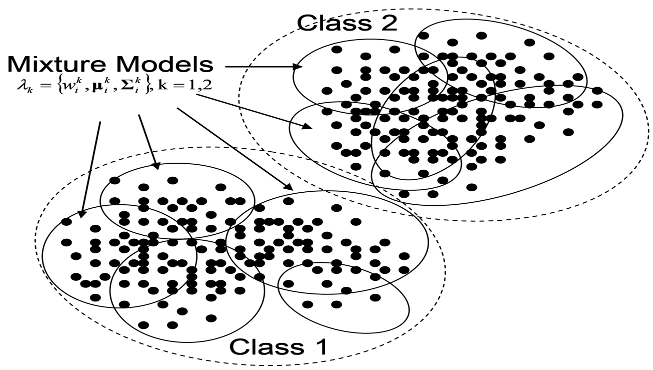

Fig. 5: EM Clustering Showing Characteristics Vector

(EM) do not always model the entire circulation but instead focus on differentiating among distribution regions. Our proposed design uses an EM model in which initial feature vectors produce the EM clusters. Figure 5 depicts EM clustering, where the (black dots) indicate the characteristic vector.

3.2 Modeling in Signature Biometrics System

The proposed signature biometric system employed Vertical Projection Profile (VPP), Horizontal Projection Pro-file (HPP), and Discrete Cosine Transform (DCT) features to design an online and offline signature identification system. Offline signature verification heavily relies on feature extraction. There are two kinds of services in offline signature recognition: static and pseudo-dynamic features in one group and global and local features in the other. The VPP and HPP are static signature features, whereas the DCT possesses global signature image characteristics. The VPP range is the same as the number of columns in the signature image. The intensities around which the image pixels are concentrated are indicated by VPP, which is also a type of histogram. VPP decided on the horizontal

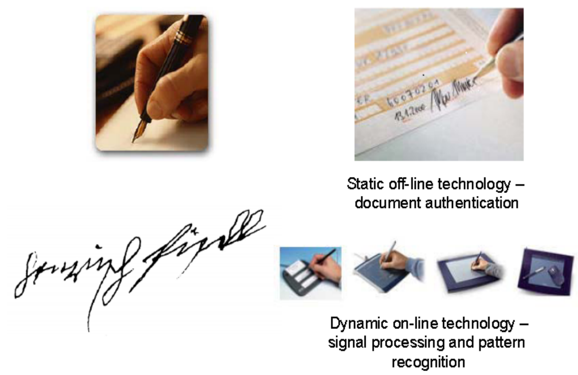

Fig. 6: Online and Offline Signature Recognition System

starting and ending areas of the image. As a result, this might be used as a distinguishing attribute of a signatory. Although signature zones vary in size even among users, the average value of the vertical projection profile is used as a feature in this task. In a signature image, a horizontal Projection Profile (HPP) is an array containing the sum of pixels in each row. The number of rows in the signature image determines the HPP size. HPP is a type of histogram as well. The intensities around which the picture pixels are concentrated are called histogram points. HPP gives the vertical starting and ending points of the image. As an outcome, this could be used as a distinguishing attribute of a signatory.

4 Results and Discussions

The statistical analysis of the data and experimental setup includes three types of experiments: voice recognition, signature recognition, and a multimodal system that provides both voice and signature recognition.
In the signature recognition system, 100 participants were selected under the age limit of 17 to 40. Their signatures were scanned with an HP Scan Jet 5300C scanner with indoor and outdoor lighting. The picture resolution to store the signatures was 300 dpi (digits per inch). An 8-bit grayscale picture was produced as a result. The saved format was BMP (bits mapping). Each participant was to give 25 example signatures. Five signatures from all the participants were required for the training session.
Like the signature recognition system, the total number for speech recognition was also set to 100 participants with 20 speakers. The sampling frequency for the voice was set to 8000 Hz, and the tests were conducted in various environments. For each participant, there were four sentences to examine. Those sentences were

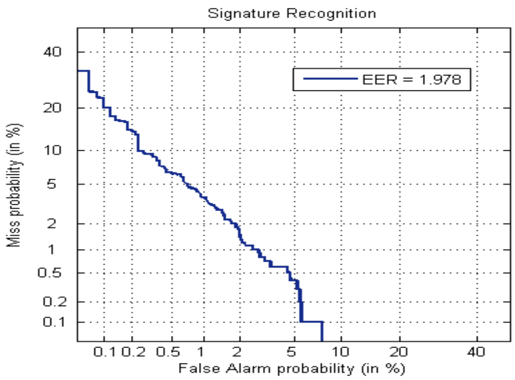

Fig. 7: Signature Recognition System showing EER

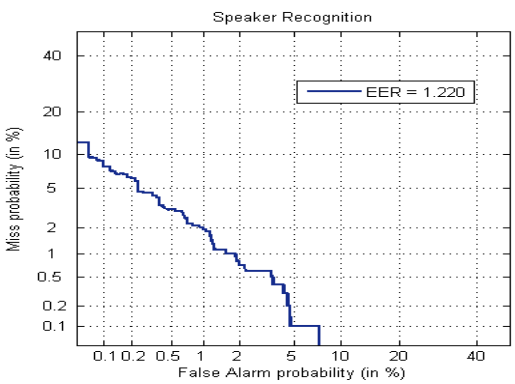

Fig. 8: Speaker Recognition System showing EER

provided to the training session for matching to the database.

The two individual systems and a combined system are compared based on error trade-off curves showing how a system performs when all error rates are equal. A lower EER score indicates higher recognition.

The miss probability increases when a fact enrollment is identified as an impostor. When an imitator is authorized as total enrollment, the deceitful apprehension likelihood increases; for a signature recognition system, the EER is 1.978 as shown in Fig. 7. Figure 8 depicts a speaker recognition system with an equal error rate (EER) of 1.2. In contrast, Figure 9 depicts a multimodal biometric system with signature and speaker recognition systems. The Equal Error Rate EER values decreased to 0.2 compared to 1.9 and 1.2 for signature and speaker recognition EERs, indicating that the multi-modal biometric system performs better than a uni-modal biometric system.

Table 1 shows the EER values for three recognition systems.

The characteristics and parameters of individually simulated and multi-model-based simulation of biometric systems indicate that the multi-model

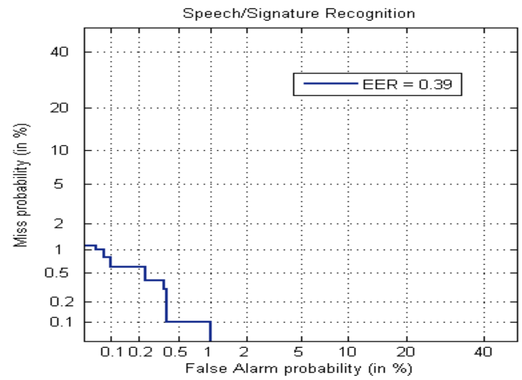

Fig. 9: Speech/Signature Recognition System showing EER

TABLE 1: Voice recognition EERs, signature recognition EERs, and voice/signature recognition EERs

S#	Category	EER values
01	Signature data	1.978
02	Voice data	1.220
03	Signature/ voice data	0.39

system is more accurate and reliable than every individual system. Hence the issues seen in uni-modal biometric systems may quickly be overcome by deploying the proposed multimodal biometric system.

In the future, the proposed model will be deployed in real-time for biometric identification, student verification, and employment in the academic environment.

References

[1] S. K. Choudhary and A. K. Naik, ”Multimodal biometric authentication with secured templates—A review,” in 2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI), 2019, pp. 1062-1069.

[2] S. A. Abdulrahman and B. Alhayani, ”A comprehensive survey on the biometric systems based on physiological and behavioural characteristics,” Materials Today: Proceedings, vol. 80, pp. 2642-2646, 2023.

[3] N. Najafzadeh, H. Kashiani, M. S. E. Saadabadi, N. A. Talemi, S. R. Malakshan, and N. M. Nasrabadi, ”Face image quality vector assessment for biometrics applications,” in Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2023, pp. 511-520.

[4] J. Taborecka-Petrovicova, M. Budinsky, and M. Sipulova, ”Implementation of Facial Biometric Technologies: A Business Perspective,” in International Conference on Applied Economics, 2022, pp. 357-374.

[5] S. Noori, ”Suspicious infrastructures: automating border control and the multiplication of mistrust through biometric e-gates,” Geopolitics, vol. 27, pp. 1117-1139, 2022.

[6] F. Jansen, J. Sánchez-Monedero, and L. Dencik, ”Biometric identity systems in law enforcement and the politics of (voice) recognition: The case of SiiP,” Big Data Society, vol. 8, p. 20539517211063604, 2021.

[7] D. Mahto and D. K. Yadav, ”Cloud-based secure telemedicine information system using crypto-biometric techniques,” EAI Endorsed Transactions on Pervasive Health and Technology, vol. 5, pp. e3-e3, 2019.

[8] Q. Yang, X. Chen, Z. He, and L. Chang, ”Survey on Deep Learning Based Fusion Recognition of Multimodal Biometrics,” in Biometric Recognition: 16th Chinese Conference, CCBR 2022, Beijing, China, November 11–13, 2022, Proceedings, 2022, pp. 511-518.

[9] A. S. Abdulbaqi, N. A. Turki, A. J. Obaid, S. Dutta, and I. Y. Panessai, ”Spoof Attacks Detection Based on Authentication of Multimodal Biometrics Face-ECG Signals,” in Artificial Intelligence for Smart Healthcare, ed: Springer, 2023, pp. 507-526.

[10] W. Si, J. Zhang, Y.-D. Li, W. Tan, Y.-F. Shao, and G.-L. Yang, ”Remote identity verification using gait analysis and face recognition,” Wireless Communications and Mobile Computing, vol. 2020, pp. 1-10, 2020.

[11] T. Bagautdinov, C. Wu, J. Saragih, P. Fua, and Y. Sheikh, ”Modeling facial geometry using compositional vaes,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 3877-3886.

[12] S. W. Shah and S. S. Kanhere, ”Recent trends in user authentication–a survey,” IEEE Access, vol. 7, pp. 112505-112519, 2019.

[13] Q. Yao, C. Chen, D. Song, X. Xu, and W. Li, ”A Novel Finger Vein Verification Framework Based on Siamese Network and Gabor Residual Block,” Mathematics, vol. 11, p. 3190, 2023.

[14] M. Khatri and A. Sharma, ”Deep Learning Approach based on Iris, Face, and Palmprint Fusion for Multimodal Biometric Recognition System,” International Journal of Performability Engineering, vol. 19, 2023.

[15] A. S. Al-Waisy, R. Qahwaji, S. Ipson, S. Al-Fahdawi, and T. A. Nagem, ”A multi-biometric iris recognition system based on a deep learning approach,” Pattern Analysis and Applications, vol. 21, pp. 783-802, 2018.

[16] C. Ding and D. Tao, ”A comprehensive survey on pose-invariant face recognition,” ACM Transactions on intelligent systems and technology (TIST), vol. 7, pp. 1-42, 2016.

[17] P. H. Silva, E. Luz, L. A. Zanlorensi, D. Menotti, and G. Moreira, ”Multimodal feature level fusion based on particle swarm optimization with deep transfer learning,” in 2018 IEEE Congress on Evolutionary Computation (CEC), 2018, pp. 1-8.

[18] A. Lumini and L. Nanni, ”Fair comparison of skin detection approaches on publicly available datasets,” Expert Systems with Applications, vol. 160, p. 113677, 2020.

[19] Z. Tao, X. Zhou, Z. Xu, S. Lin, Y. Hu, and T. Wei, ”Finger-vein recognition using bidirectional feature extraction and transfer learning,” Mathematical Problems in Engineering, vol. 2021, pp. 1-11, 2021.

[20] H. M. Therar and A. J. Ali, ”Personal Authentication System Based Iris Recognition with Digital Signature Technology,” Journal of Soft Computing and Data Mining, vol. 4, pp. 13-29, 2023.

[21] S. Daas, A. Yahi, T. Bakir, M. Sedhane, M. Boughazi, and E. B. Bourennane, ”Multimodal biometric recognition systems using deep learning based on the finger vein and finger knuckle print fusion,” IET Image Processing, vol. 14, pp. 3859-3868, 2020.

[22] Z. Qin, P. Zhao, T. Zhuang, F. Deng, Y. Ding, and D. Chen, ”A survey of identity recognition via data fusion and feature learning,” Information Fusion, vol. 91, pp. 694-712, 2023.

[23] Y. Gao, W. Wang, V. V. Phoha, W. Sun, and Z. Jin, ”EarEcho: Using ear canal echo for wearable authentication,” Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, vol. 3, pp. 1-24, 2019.

[24] B. Ammour, L. Boubchir, T. Bouden, and M. Ramdani, ”Face–iris multimodal biometric identification system,” Electronics, vol. 9, p. 85, 2020.

[25] K. M. Sudar, P. Deepalakshmi, K. Ponmozhi, and P. Nagaraj, ”Analysis of security threats and countermeasures for various biometric techniques,” in 2019 IEEE International Conference on Clean Energy and Energy Efficient Electronics Circuit for Sustainable Development (INCCES), 2019, pp. 1-6.

[26] M. S. Obaidat, S. P. Rana, T. Maitra, D. Giri, and S. Dutta, ”Biometric security and internet of things (IoT),” Biometric-based physical and cybersecurity systems, pp. 477-509, 2019.