Multi-modal Biometric Authentication Based on Deep Adversarial Learning Utilizing ECG and Fingerprint Modality

Abdullah Alduhailan [1,*], Nazhatul Hafizah Kamarudin [1], Siti Norul Huda Sheikh Abdullah [1], and Aminu Dau [2]

[1] Centre for Cyber Security, Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia, Bangi, Malaysia
[2] Department of Computer Science, Hassan Usman Katsina Polytechnic, Katsina, Nigeria
Email: Mazd794_3@hotmail.com (A.A.); nazhatulhafizah@ukm.edu.my (N.H.K.); snshabdullah@ukm.edu.my (S.N.H.S.A.); dauaminu@gmail.com (A.D.)
*Corresponding author

Abstract—Biometric authentication systems have become essential for secure and reliable identity verification. Most of the existing biometric methods are unimodal systems. Unimodal methods typically focus on single method such as fingerprint, iris, face or Electrocardiogram (ECG). However, unimodal systems often suffer from limitations such as susceptibility to spoofing, noise sensitivity, and reduced accuracy. In this study, we propose a novel multimodal biometric authentication framework that integrates ECG and fingerprint data based on deep learning method. The proposed model used a Transformer-based Generative Adversarial Network (TGAN) for the dataset augmentation. An enhanced Visual Geometry Group (VGG-16) model is applied for deep feature extraction. To combine modality-specific representations, the model used a weighted feature fusion technique. To further improve classification performance, we employ a Multi-Support Vector Machine (Multi-SVM) classifier. The proposed model was evaluated using the Physikalisch-Technische Bundesanstalt (PTB-XL), Electrocardiogram Identification Database (ECG-ID), Fingerprint Verification Competition 2004 (FVC2004), and LivDet 2023 datasets. Results demonstrated that our proposed approach outperforms the baseline models. Specifically, the multimodal system achieved up to 98.4% accuracy with an Area Under the Curve (AUC) of 0.993, and an Equal Error Rate (EER) as low as 1.1%.

***Keywords*—biometric authentication, Convolutional Neural Network (CNN), deep learning, Fingerprint, Electrocardiogram (ECG), Support Vector Machine (SVM)**

I. Introduction

Biometric authentication has evolved as a solid way for identity verification which harnesses unique physiological and behavioral attributes such as finger prints, faces and Electrocardiogram (ECG) signals [1]. Traditional techniques such as passwords and PINs, suffer from weaknesses which include theft, forgetfulness, and brute-force assaults [2] which makes them less effective. In contrast, biometrics provide inherent advantages, which include uniqueness and resistance to forgery. However, unimodal biometric systems generally, face some challenges such as spoofing attacks, and environmental variations which can lead to poor performance [3]. To address these limitations, recently, multimodal biometric authentication methods has been introduced. These models incorporate various biometric modalities to boost performances.

Recent breakthroughs in deep learning based models have further transformed biometric authentication systems [4, 5]. These models aid in improving feature extraction and learning from complicated data patterns. Deep learning models have exhibited exceptional performance in biometric applications. Convolutional Neural Networks (CNNs) is the most popular deep learning model for biometrics [2]. These models can learn sophisticated representations of biometric data which makes them particularly efficient in addressing intra-class variances thus, boosting classification accuracy [6].

Recently, ECG-based authentication has gained attention due to its unique characteristics [7]. ECG-based methods are very powerful in liveness detection as well as resistance to spoofing. Unlike fingerprints and facial recognition [3], ECG signals are generated by the electrical activity of the heart. They usually differ based on individual physiological systems. This intrinsic uniqueness makes ECG a good alternative for biometric identification [8]. However, despite their benefits, ECG based systems are very susceptible to noise, and electrode placement [7]. This implies the requirement of improved feature extraction approaches to ensure trustworthy authentication.

On the other hand, Fingerprint recognition remains one of the most widely adopted biometric modalities [9]. This

Manuscript received August 21, 2025; revised November 3, 2025; accepted November 24, 2025; published March 26, 2026.

is because of their high accuracy, and widespread deployment in various authentication systems. Recently, fingerprint-based systems have been extensively studied in literature [10]. These approaches exploit various algorithms thereby enhancing their robustness against distortions, and spoofing attacks. However, despite their effectiveness, fingerprint authentication systems have been indicated to be vulnerable to presentation attacks [3]. This is particularly important, where synthetic fingerprints are used to deceive the system. This limitation among others underscores the need for multimodal approaches to improve performances.

Thus, motivated by the strengths and limitations of existing biometric authentication methods, in this research, we propose a multimodal based method for biometric authentication that uses an enhanced deep learning-based method. The model integrates ECG and fingerprint modalities based on the weighted fusion strategy. The proposed approach aims at addressing challenges such as noise sensitivity in ECG signals and spoofing vulnerabilities in fingerprint authentication.

Given the scarcity of large-scale biometric datasets, we essentially, leverage a Transformer-based to generate synthetic ECG and fingerprint data for better performances. To effectively improve feature extraction, we introduce an enhanced VGG 16 model. Further, we employ a weighted fusion approach that can help in adaptive feature fusion for better authentication. Finally, to improve classification accuracy, we incorporate a Multi-SVM classifier that adapts to varying feature distributions thereby reducing misclassification rates. The key contributions of this research are as follows:

1. We propose a Transformer-based GAN (TGAN) for data augmentation which is capable to generate synthetic ECG and fingerprint samples for enhancing the model training.
2. We design an enhanced VGG 16 for feature extraction which is capable to capture salient features in both ECG and fingerprint data for better representation learning.
3. We design a weighted fusion technique and Multi-SVM classifier for increasing the performance.
4. An extended series of experiments was done based on real-world datasets to evaluate the performance of the proposed model.

The remainder of this paper is organized as follows: Section II reviews related work. Section III discusses the recommended technique. Section IV includes experimental findings and performance evaluations. Section V concludes the study and discusses future research directions.

II. Related Work

Biometric authentication has been widely investigated in recent years [11]. Generally, biometric authentication can utilize either unimodal and multimodal systems [1]. Majority of the Traditional unimodal systems focus on fingerprints, ECG, iris, or face recognition. These approaches have demonstrated high accuracy but experience several shortcomings such as vulnerability to spoofing attacks, environmental noise, etc. To address the issues with the unimodal approaches, multimodal biometric systems have been introduced. These models combine two or more biometric traits by leveraging the strengths of different modalities. In this section, we review existing works on both unimodal and multimodal authentication systems.

A. Unimodal Biometric Approaches

Unimodal biometric approaches involve only one aspect of the biometric data for authentication [1]. Over many years, several unimodal approaches have been proposed which include fingerprint, ECG, iris, face etc. In this research we focus on the deep learning based fingerprint and ECG methods for authentication. Recently, several deep learning based fingerprint recognition methods have been investigated. For example, Madhusudhan *et al.* [3] proposed a deep learning-based finger vein model for biometric authentication. The authors incorporate intelligent learning techniques to improve performances. Their study highlights the potential of vein-based biometrics as an alternative to fingerprint authentication. Natarajan *et al.* [10] introduced an improved method that use hybrid optimization-tuned for fingerprint authentication. This method achieves improved performances compared to the prior approaches. However, the reliance on hybrid optimization techniques lead to computational overhead which makes it impractical in real-time applications. Shinde and Kayte [12] proposed a fingerprint recognition system based on pre-trained CNN model. Their study demonstrated that transfer learning could significantly improve accuracy. However, despite its success, this approach heavily depends on the quality of the pre-trained model, which may not generalize well across diverse datasets. Similarly, Ametefe *et al.* [13] conducted a study to explore deep transfer learning and data augmentation for fingerprint pattern classification. The authors showed that augmentation techniques can enhance model robustness against variations in fingerprint patterns. However, excessive augmentation may lead to synthetic data biases. This could affect real-world applicability. A similar study was conducted by Jain *et al.* [14], they employed deep learning model to design a decision-based median filter to preprocess fingerprint images. Their method improved feature extraction and reduced noise interference.

On the other hand, ECG-based authentication has gained attention, recently, due to its resistance to spoofing [15]. To this end, several deep learning based methods have been proposed for ECG-based biometric authentication. Prakash *et al.* [7] proposed a deep learning-based authentication system based on ECG signals. Similar research has been conducted by Agrawal *et al.* [16], who applied deep learning algorithms to ECG-based user authentication. The study demonstrated that deep learning models could effectively learn ECG patterns. However, issues related to sensor variability and user adaptability remain open challenges. In similar manner, Prakash *et al.* [2] used deep learning method to propose ECG beat template matching. This

approach improved classification accuracy on various evaluation metrics. Mageshbabu and Mohana [17] introduced an enhanced ECG-based biometric authentication using a hybrid Convolutional Neural Network-Long Short Term Memory (CNN-LSTM) framework for single-heartbeat ECG authentication. The proposed model outperformed the previous method on same datasets and evaluation metrics. However, while the model is promising, the reliance on single-heartbeat samples may introduce inconsistencies due to heart rate variability. Zhang *et al.* [18] utilized CNN and transfer learning for ECG-based human identification. The proposed approach utilized multi-view feature representations which improve generalization.

B. Multimodal Biometric Authentication

Multimodal biometrics aim to enhance security through integration of two or more biometric traits [5]. Recent studies have explored different fusion techniques to achieve better performance. For example, Heidari and Chalechale [6] proposed a multimodal authentication approach by fusing finger knuckle print and finger nail features. The authors showed improved accuracy compared to existing unimodal methods. However, this method faces challenges in feature alignment and fusion complexity. Another study conducted by Ammour *et al.* [4] introduced a deep learning based method which integrate ECG, fingerprint, and finger knuckle print for better performances. This method improved accuracy against spoofing. Similarly, Salturk and Kahraman [19] investigated dynamic signatures and facial data fusion for biometric authentication. Yadav [20] developed a deep learning approach which integrate iris, fingerprint, and handwritten signature traits. However, one issue with handwritten signatures is variability due to handwriting inconsistencies.

In another study, Alshardan *et al.* [21] explored CNN-based fusion techniques to concatenate fingerprint and finger vein data for biometric authentication. Experimental results have shown that the approach demonstrated improved accuracy. Hammad and Wang [22] introduced a parallel score fusion technique for ECG and fingerprint authentication using CNN. Jomaa *et al.* [23] developed an end-to-end deep learning fusion model for fingerprint and ECG presentation attack detection. Li *et al.* [24] introduced adaptive deep feature fusion for continuous authentication. The proposed method leveraged data augmentation and improve performances compared to the previous approaches. Buriro *et al.* [25] introduced SWIPE Generative Adversarial Network (SWIPEGAN), a data augmentation technique using GANs for smartphone user authentication. Although the model has shown effective results, however, the traditional GAN-generated synthetic data may introduce biases that affect biometric recognition.

C. The Research Gap

The reviewed studies demonstrate significant advancements in biometric authentication using deep learning methods. However, several challenges remain. For example, deep learning-based biometric models, particularly multimodal approaches, introduce computational challenges that may hinder real-time deployment. While deep learning enhances recognition accuracy, adversarial attacks remain a critical challenge. Biometric systems must integrate adversarial defense mechanisms. Physiological biometrics such as ECG and finger vein exhibit high inter-subject variability, requiring more sophisticated normalization techniques. Multimodal biometrics improve security but face challenges in aligning and fusing heterogeneous features effectively.

III. Overview of the Proposed Model

This section presents an overview of the proposed deep learning–based multimodal biometric authentication framework and explains how its key components interact to achieve robust, secure, and efficient user authentication. Biometric authentication plays a vital role in modern security infrastructures, offering a reliable means of verifying an individual's identity based on intrinsic physiological signals. However, traditional unimodal biometric systems, which rely on a single modality such as the ECG or fingerprint, often suffer from challenges such as noise sensitivity, feature variability, spoofing attacks, and environmental distortion. These limitations can significantly reduce recognition accuracy and compromise system reliability in real-world scenarios. To address these challenges, this research proposes an integrated multimodal authentication framework that combines the complementary strengths of ECG and fingerprint modalities using a T-GAN for intelligent data augmentation. More specifically the model uses an enhanced VGG-16 architecture for discriminative feature extraction, a weighted feature fusion strategy for multi-level integration, and an optimized Multi-Support Vector Machine (Multi-SVM) classifier for final authentication. Together, these components form a unified, end-to-end learning system capable of improving recognition accuracy, reducing spoofing susceptibility, and enhancing robustness against data variability.

The workflow of the proposed framework operates in four sequential stages. First, the Transformer-GAN generates high-quality synthetic ECG and fingerprint samples to expand the dataset and improve model generalization. Second, an improved VGG-16 network is employed to extract deep, discriminative representations from both biometric modalities. The network is refined with batch normalization, attention enhancement, and adaptive pooling layers, enabling it to capture fine-grained features and temporal–spatial dependencies within ECG waveforms and fingerprint textures. Third, a weighted feature fusion mechanism integrates the extracted features at multiple abstraction levels. This mechanism adaptively assigns importance weights to each modality, ensuring that both ECG and fingerprint features contribute proportionally to the final representation based on their discriminative strength. The fusion process leverages both low-level and high-level feature correlations to capture complementary information effectively. Finally, the fused feature vectors are fed into a Multi-SVM classifier that

performs the final authentication decision. The Multi-SVM classifier improves decision boundaries between genuine and impostor classes by leveraging multiple kernel functions, ensuring optimal separation in high-dimensional feature space. Fig. 1 shows the general framework of the proposed model.

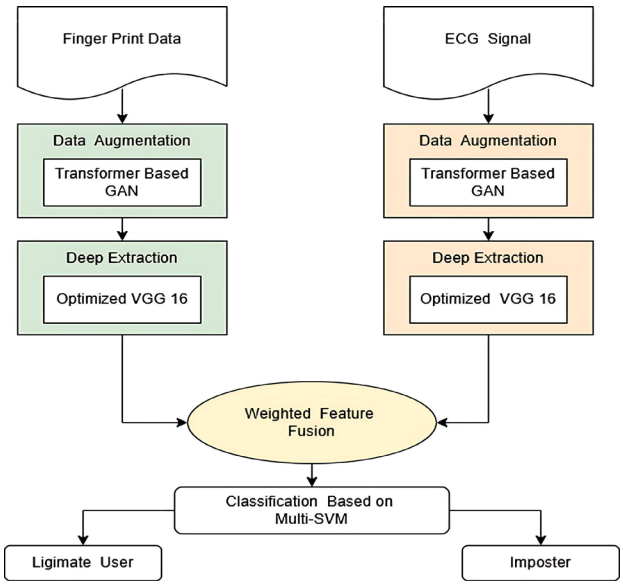


Fig. 1. General framework of the proposed model.

A. Adversarial Learning for Data Augmentation

Data scarcity is a major limitation in biometric authentication, especially for ECG signals, which vary across individuals and conditions. To address this issue, in this research we introduce an adversarial learning based data augmentation. Our proposed approach use a Transformer based Generative Adversarial Network (TGAN). A GAN consists of two networks, namely, Generator (G) and Discriminator (D) [26]. G takes a random noise vector $z{\sim}p_z(z)$) and maps it to a synthetic biometric sample $x^2 = G(z)$. D Takes an input biometric sample (either real or generated) and outputs a probability $D(x)$ which indicate whether it is real or fake. The objective of the GAN is formulated as a min-max optimization problem:

$$\min_G \max_D E_{x\sim p_{data}(z))}[\log D(x)] + E_{z\sim p_z(z))}[\log(1 - D(G(z)))]$$

where $p_{data}(x)$ represents the real data distribution of biometric samples (ECG and fingerprint). $p_z(z)$ is the noise distribution. $G(z)$ Generates synthetic samples from noise. $D(x)$ is the discriminator's probability. X is a real biometric sample. Fig. 2 illustrates the basic structure of the GAN model.

Different from the traditional GAN models [27], in this research, we introduce TGAN to improve the model performances. The TGAN uses Transformer-based architectures in the GAN generator. This advancement can help to improving feature extraction and generating high-resolution biometric data. The generator network of our proposed method consists of self-attention layers. This enables the model to learn long-range dependencies in the input features. Detail of the generator and discriminator is given in the following subsections.

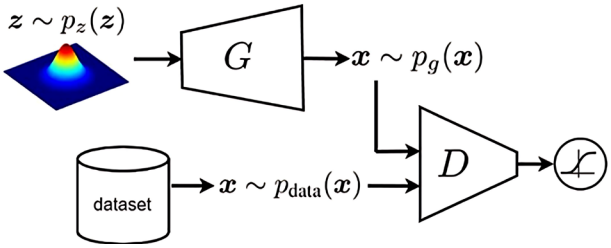


Fig. 2. An illustration of the GAN model [27].

1) Transformer based generator

Our transformer-based generator has two successive stages which include five stacked transformer encoder blocks equipped with dual-head multi-head attention mechanisms, as seen in Fig. 3. Positional encoding is implemented before each encoder block to improve the model's capacity for generating synthetic samples that closely mimic authentic biometric data. An upsampling block which consist reshaping and pixel-shuffle modules, is included between the two transformer stages. This will help to enhance parameter efficiency and improve adaptation to the biometric feature space. Subsequent to these steps, the model generates 1D sequential data with the embedding dimension diminished to one-fourth of its original size. To revert to the original embedding dimension of the produced samples, the 1D sequence is initially molded into a 2D feature map. It is then processed via a convolutional layer for dimensional reduction. This converted back into a 1D sequence format.

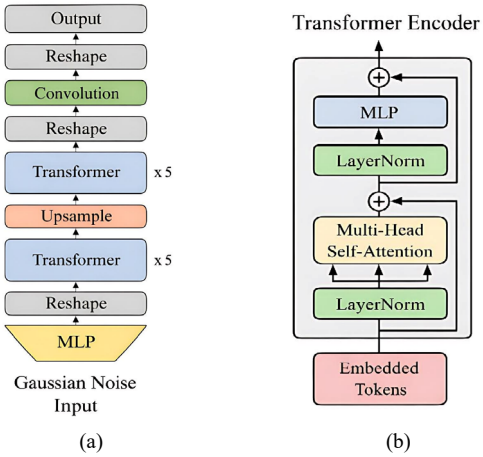


Fig. 3. Illustration of Transformer based generator: (a) the generator architecture; (b) conceptual diagram of the transformer.

The transformers in the generator consists of self-attention layers which can be given as follows:

$$Attention = (Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

where Q, K, V represents the query, key and value matrices computed from input features. d_k is the scaling

factor for preventing large dot product values from dominating the softmax function.

2) Discriminator

The discriminator serves as a classifier that detects if the incoming data is real or fabricated. Given that transformer-based discriminators generally underperform and require substantially more data [26], we choose for a CNN-based discriminator to solve this issue [28]. Accordingly, we build the discriminator using a convolutional architecture of three layers (Conv), a single average pooling layer (AvgPool), and two Fully Connected (FC) layers, as indicated in Table I.

TABLE I. THE CNN USED IN THE DISCRIMINATOR

Operator	Output
Conv (LeakyReLU)	32×100×3
Conv (LeakyReLU)	64×50×3
Conv (LeakyReLU)	128×25×3
AvgPool	128
FC (LeakyReLU)	512
FC	1

We used LeakyReLU, as the activation function for the three Conv levels and the first FC layer. Each Conv layer employs a fixed-size kernel. The number of filters doubles at each step to generate progressively compact and dense feature maps. Additionally, the second dimension of the kernel is set to 1 which permit the extraction of features along separate axes. Padding and stride are designed to minimize the data length by half after each convolution. The generated features are transmitted via the AvgPool layer to create 1D feature sequences. This is further processed by the two FC layers to identify the inputs as either genuine or synthetic.

B. Deep Feature Extraction

In this research, to further improve performance, we designed an improved VGG-16 architecture. VGG-16 is a deep CNN architecture with 16 layers, including 13 convolutional layers and 3 fully connected layers [29]. It follows a simple design pattern of using small 3×3 convolutional kernels stacked in multiple layers. The layers interspersed with max pooling for spatial downsampling. The general form of convolution in VGG-16 is:

$$Y = f(W * X + b)$$

where W represents the learnable convolutional filters. X is the input feature map. b is the bias term. $*$ denotes the convolution operation. $f(.)$ is the activation function. Fig. 4 illustrate the basic structure of the VGG 16 model.

However, unlike the basic VGG-16 model, our improved version integrates several aspects for performance improvement. This include Batch Normalization (BN) to stabilize training. It utilizes Global Average Pooling (GAP) instead of fully connected layers to reduce over fitting. An attention mechanism is also used to focus on the most relevant biometric features. To better capture multi scale features, a dilated convolutional structure is used. Detail of the enhancement can be given as follows:

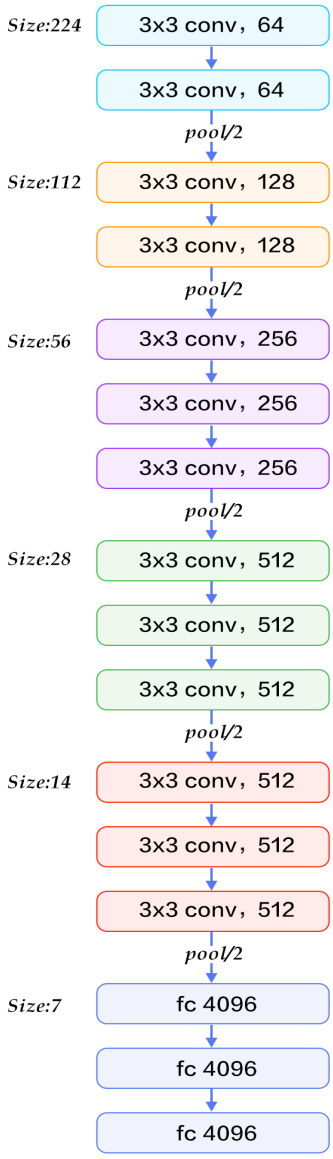


Fig. 4. Basic structure of the VGG16.

1) Batch normalization

To stabilize training and prevent vanishing/exploding gradients, we apply BN after every convolutional layer:

$$\hat{x} = \frac{x-\mu}{\sigma}, y = \gamma\hat{x} + \beta$$

where x is the input activation, μ and σ are the batch mean and standard deviation. γ and β are learnable scale and shift parameters.

2) Spatial attention mechanism

Since both ECG and fingerprint images contain critical regions that influence classification, to focus on these regions dynamically, we integrate a spatial attention mechanism. The spatial attention map is computed as:

$$S = \sigma(f_7(|\max(X), avg(X)|))$$

where X is the feature map. max (X) and $avg\ (X)$ are the max-pooled and average-pooled versions of X respectively. f_7 Represents a 7×7 convolution and σ is the sigmoid activation function.

3) Dilated convolutions

For better capturing of long-range dependencies in ECG signals and extract high-resolution fingerprint features, we replace the conventional convolutions with dilated convolutions:

$$y(i) = \sum_{k=0}^{K} w(k) \cdot x(i + r \cdot k)$$

where r is the dilation rate, which control the spacing between kernel elements. k is the filter size. $w(k)$ is the filter weight. After feature extraction, the model outputs a feature vector F for each biometric modality. This serves as input for feature fusion and subsequent classification:

$$F_{ECG} = VGG16(ECG), F_{FG} = (FingerPrint)$$

C. Weighted Feature Fusion And Classification

Since ECG and fingerprint biometrics provide complementary information, we employ a weighted fusion method to integrate their features at multiple levels. The fused feature representation F_{fusion} is computed as:

$$F_{fusion} = \alpha \cdot F_{ECG} + (1 - \alpha) \cdot F_{FP}$$

where α is the adaptive weight factor $(0 \leq \alpha \leq 1)$ which determines the contribution of each modality. Having obtained the fused features F_{fusion}, it is finally passed to the Multi-SVM classifier for authentication:

$$\hat{y} = Multi - SVM(F_{fusion})$$

where $\hat{y}$ is the predicted authentication label (genuine or impostor).

IV. Experimental Datasets

We conducted experiments using a combination of ECG datasets, fingerprint datasets, and a composite multimodal dataset (created by pairing samples from both domains). We choose the datasets because of their relevance.

1) ECG datasets

Two different databases are used in this paper for ECG authentication. The first database is PTB-XL ECG Dataset [30] which is a large-scale 12-lead clinical ECG dataset containing over 21,000 recordings from 18,885 patients, annotated with detailed diagnostic labels. The datasets have Sampling Rate of 500 Hz and 100 Hz versions. It provides high-resolution ECG waveforms with labeled cardiac abnormalities. This makes it ideal for evaluation of authentication under clinical and clean conditions. We extracted Lead I and II for biometric feature learning and segmented signals into fixed windows for input into the CNN model. Fig. 5 shows an example of one record from PTB database.

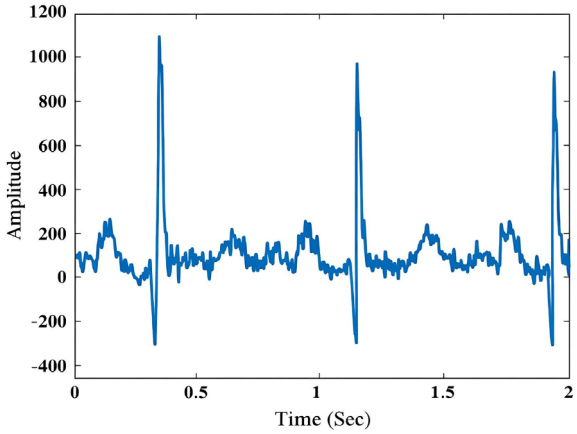


Fig. 5. Example of one record from PTB database.

The second datasets is ECG-ID Database [31]. ECG-ID is a publicly available dataset specifically designed for biometric identification using Electrocardiogram (ECG) signals. It is one of the few datasets focused on person identification using ECG patterns. It has a total Subjects of 90 individuals. The Signals per Subject involves two recordings, each lasting about 20 s. The Sampling Rate is 500 Hz. We used Lead I (single-lead ECG). Fig. 6 shows an example sample of the ECG-ID signal.

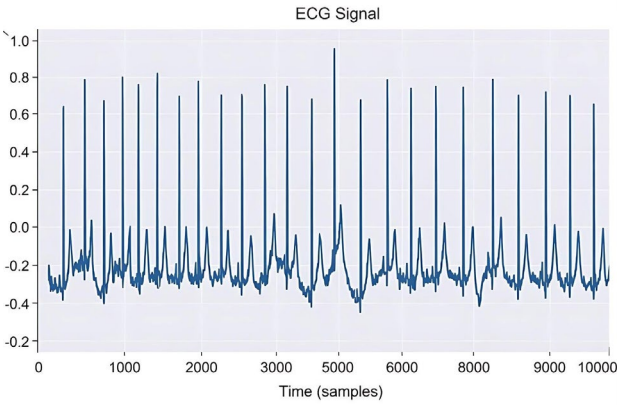


Fig. 6. Sample of the ECG signal.

2) Fingerprint datasets

For the fingerprint, two different datasets were used to evaluate the model performance based on the fingerprint recognition. The first dataset is FVC2004 database [32]. There are four distinct datasets in all (DB1, DB2, DB3, and DB4). There are 110 fingers and 880 impressions (8 impressions per finger) in each dataset. These fingerprints were gathered using three separate scanners. Thermal Sweeping Sensor for DB3, Synthetic Generator for DB4, and Optical Sensor for DB1 and DB2. The pictures are all 8-Bit grayscale TIF files. 500 dpi is the picture resolution. As seen in Fig. 7, we chose 63 patients at random and only took into account two samples per person.

The second dataset is LivDet (Liveness Detection) [33]. It contains live and spoofed fingerprints captured. Each of the four datasets, namely, CrossMatch, DigitalPersona, GreenBit, and Biometrika represents a distinct scanner and contains around 2000 images of both

actual and false fingerprints. There are two sets in this database: training and testing. Six distinct spoof materials Ecoflex, Gelatine, Latex, WoodGlue, Liquid Ecoflex, and Room-Temperature Vulcanizing (RTV) were used to create the fake samples. We only utilized two samples per person, and we choose 100 people at random. Typical samples of authentic and fraudulent fingerprint pictures from one dataset of the LivDet2023 database utilized in the studies are displayed in Fig. 8.

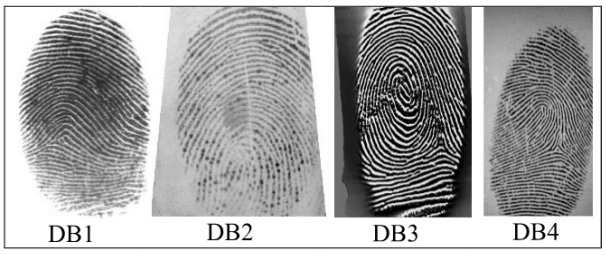


Fig. 7. Sample image from each datasets on FVC2004.

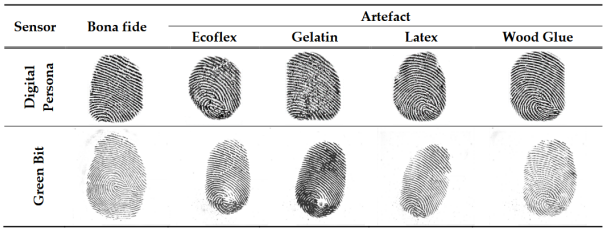


Fig. 8. Bona fide and artefact samples from LivDet datasets captured using Green Bit and Digital Persona sensors.

3) *Multimodal dataset (ECG + fingerprint fusion)*

A multimodal dataset was created by combination of synchronized ECG signals and fingerprint images from the above datasets [34]. In order to demonstrate that our approach is not limited to a single database, we combined the fingerprint and ECG datasets to create two multimodal databases, which we then utilized for assessment. One subject from the PTB XL database was paired at random with a subject from the FC2004 fingerprint database in the first multimodal database (CMD1). Lastly, the CMD1 is obtained from 100 participants, each of whom has two fingerprint and ECG samples. We paired a subject from the LivDet fingerprint database with a subject from the ECG-ID database at random in the second multimodal database (CMD2). Lastly, 63 participants had their CMD2 obtained; each contains two fingerprint and ECG samples.

V. Results and Discussion

This section present the results and discussion of the proposed mode. We first provide the experimental results and discussion of the ECG-based authentication, the fingerprint authentication and the proposed multimodal authentication.

A. *ECG-Based Authentication Result*

To objectively assess the effectiveness of the proposed framework, its performance is first compared with the popular pretrained CNNs. This include the original VGG-16, ResNet-50, InceptionV3, and MobileNetV2. Additionally, authentication time is evaluated for all models. To this end, two publicly available ECG datasets were used in this experiment: PTB-XL and ECG-ID. To ensure fair and consistent evaluation, both the baseline and proposed models were trained and tested under the same experimental configuration. The experimental results obtained on the PTB-XL and ECG-ID dataset are presented in Tables II and III, respectively. The performance was assessed based on popular evaluation metrics which include Accuracy, Precision, Recall, F1-Score, Equal Error Rate (EER), and Average Authentication Time.

TABLE II. ECG Authentication Results on PTB-XL Dataset

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	EER (%)	Time (ms)
VGG-16	95.4	94.7	94.9	94.8	2.1	88.6
ResNet-50	94.8	93.9	94.1	94.0	2.4	94.1
InceptionV3	94.3	93.2	93.6	93.4	2.7	106.3
MobileNetV2	93.1	91.8	92.3	92.0	3.3	54.2
Proposed Model	97.6	96.8	97.2	97.0	1.2	86.4

TABLE III. ECG Authentication Results on ECG-ID

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	EER (%)	Time (ms)
VGG-16	91.5	90.3	90.7	90.5	3.6	85.9
ResNet-50	90.2	89.1	89.4	89.2	4.2	92.4
InceptionV3	89.7	88.5	89.0	88.7	4.5	105.8
MobileNetV2	88.1	86.8	87.2	87.0	5.1	52.6
Proposed Model	94.5	93.2	93.7	93.4	2.8	84.3

As shown from the tables, the proposed model significantly outperforms all baseline CNNs. It experiences the highest accuracy of 97.6% and 94.5% in term of PTB-XL Dataset and ECG-ID dataset, respectively. In terms of the average authentication time, it shows the lowest EER of 1.2% and 2.8% for the of PTB-XL Dataset and ECG-ID dataset, respectively. This demonstrates its superior capability in discriminating between individuals based on ECG signals. For example, MobileNetV2 offers the fastest inference time, it does so at the cost of reduced accuracy and reliability. The proposed model maintains a good trade-off with an authentication time which is acceptable for real-time applications. Although coarser signals and sensor variability result in much worse performance than PTB-XL, the suggested framework shows significant generalizability to actual ECG authentication scenarios. The results from both datasets provide strong support for the usefulness of the suggested method.

B. *Fingerprint-Based Authentication Results*

To validate fingerprint based authentication results, comparative analysis is conducted using LivDet 2023, and FVC2004. Pretrained CNNs were used for the comparison. This include VGG-16, ResNet-50, InceptionV3, and MobileNetV2. Tables IV and V show the results on LIVEDET and FVC2004 datasets, respectively.

TABLE IV. PERFORMANCE METRICS ON LIVDET 2023 DATASET

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	EER (%)	Time (ms)
VGG-16	93.2	92.5	91.8	92.1	4.3	87.3
ResNet-50	92.7	91.4	91.1	91.2	4.6	91.5
InceptionV3	91.6	90.8	89.7	90.2	5.1	103.7
MobileNetV2	90.3	89.0	88.5	88.7	5.8	53.6
Proposed Model	96.4	95.2	95.6	95.4	2.1	1.1

TABLE V. PERFORMANCE METRICS ON FVC2004 DATASET

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	EER (%)	Time (ms)
VGG-16	94.6	93.8	93.5	93.6	3.2	88.9
ResNet-50	94.1	92.9	93.0	92.9	3.6	93.2
InceptionV3	93.4	91.6	92.0	91.8	4.1	105.2
MobileNetV2	91.7	90.5	90.1	90.3	4.7	52.1
Proposed Model	97.1	96.4	96.7	96.5	1.7	84.6

The suggested model showed improved accuracy across both datasets. For instance, the suggested model outperformed all baselines with an accuracy of 96.4% and 97.1 on the LivDet 2023 and FVC2004, respectively. The model shows a significant drop in EER in terms of security, going as low as 2.1% and 1.7% on LivDet 2023 and FVC2004, respectively. With an average inference time of around 52 ms, MobileNetV2 continues to be the fastest model for authentication speed, but accuracy and security are compromised. With a balanced runtime of about 85 ms, our suggested approach is incredibly effective without sacrificing accuracy. The performance of LivDet 2023 demonstrates that the model can discriminate between live and fake fingerprints across aspects. Equally, on FVC2004, it successfully manages variances owing to skin dryness, pressure discrepancies, and sensor conditions.

C. Multimodal Authentication (ECG + Fingerprint) Results

To test the efficacy of the proposed model, we also performed experiments by merging ECG and fingerprint modalities. Two datasets were merged for the multimodal study. To this end, we construct a Combined Datasets (CMD). This include CMD1 (PTB-XL + FVC2004) and CMD2 (ECG-ID + LivDet 2023). Tables VI and VII shows the experimental results of the CMD1 and CMD2, respectively.

From the results, it can be observed that the multimodal approach using the proposed model significantly outperformed all single-modality and baseline multimodal configurations. For instance, on the CMD1 dataset, the model achieved an outstanding 98.4% accuracy. Similarly, on the CMD 2 dataset, 97.9% accuracy was recorded, surpassing baseline VGG-16 and other pretrained models by a margin of 2–5%. ERR were drastically reduced to 1.1% and 2.3%, respectively. This demonstrates the complementarity of ECG and fingerprint features when processed through adversarial trained networks.

TABLE VI. PERFORMANCE OF CMD1

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	EER (%)	Time (ms)
VGG-16	96.5	95.7	95.8	95.7	2.5	88.5
ResNet-50	96.2	95.1	95.4	95.2	2.9	93.7
InceptionV3	95.4	94.0	94.3	94.1	3.4	106.8
MobileNetV2	93.9	92.6	92.9	92.7	4.1	53.6
Proposed Model	98.4	97.6	97.9	97.7	1.1	5.3

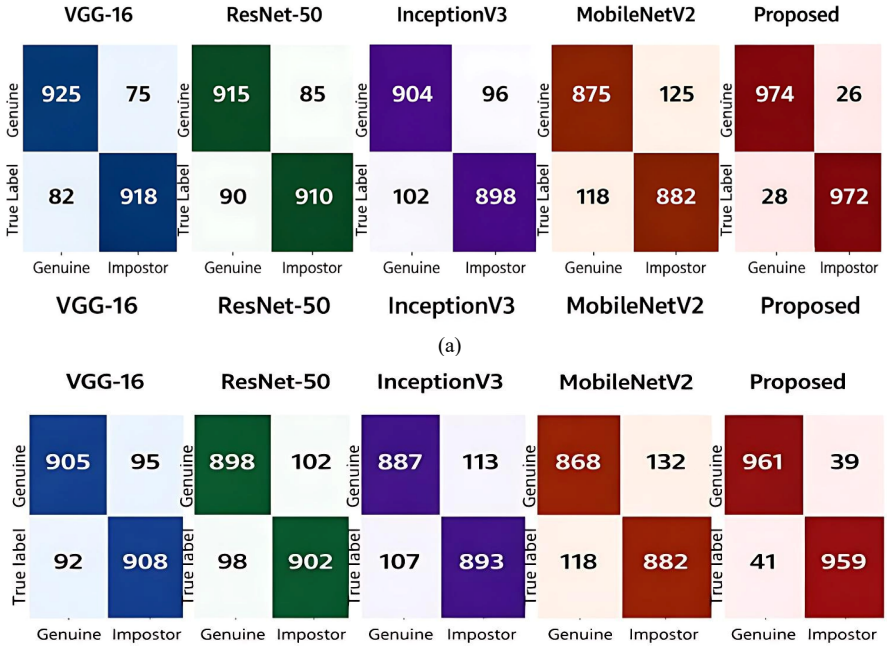
TABLE VII. PERFORMANCE ON CMD 2

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	EER (%)	Time (ms)
VGG-16	94.2	93.0	93.1	93.0	3.1	85.7
ResNet-50	93.6	92.3	92.4	92.3	3.6	92.1
InceptionV3	92.9	91.4	91.7	91.5	4.0	104.6
MobileNetV2	91.2	89.9	90.3	90.1	4.9	51.4
Proposed Model	97.9	95.5	95.9	95.7	2.3	83.9

While MobileNetV2 retained the lowest latency (~51–53 ms), its performance was consistently the weakest. The proposed model delivered near-real-time performance, which maintains average authentication times below 88 ms. The results confirm that the proposed model, offers state-of-the-art multimodal authentication performance. Fig. 9 shows the comparison of confusion matrices with different CNN models. The metrics are calculated based on a test set containing balanced genuine and impostor attempts. As can be seen from the figures, the proposed model clearly shows superior performance, with minimal misclassification of both genuine and impostor attempts. More specifically, the proposed model consistently achieves the lowest False Acceptance Rate (FAR) and False Rejection Rate (FRR) across both datasets, reflected in higher true positive and true negative values. Fig. 10 shows the Receiver Operating Characteristic (ROC) curves for all the models. It can be observed from the figures that the proposed model consistently achieves the highest AUC value. This confirms superior classification performance and better separation between genuine and impostor attempts.

1) Evaluation of data augmentation quality

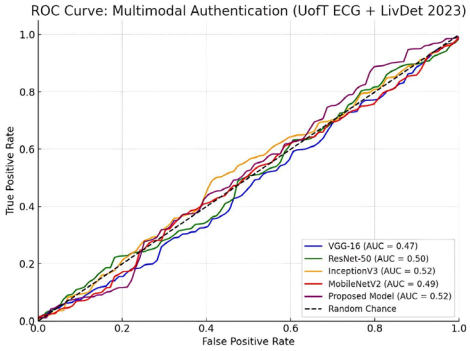
To assess the effectiveness of the T-GAN used in this study, a quality evaluation was performed focusing on statistical realism of the augmented data. The goal was to verify that the generated ECG and fingerprint samples faithfully represent real biometric distributions while enhancing the model's ability to generalize for unseen conditions. To evaluate the visual and structural fidelity of the generated data, dimensionality reduction technique such as Principal Component Analysis (PCA) was applied to both real and synthetic data to assess distributional alignment. The resulting visualizations (Fig. 11) show strong overlap between real and synthetic feature clusters, with synthetic data extending the margins of the real data distribution. This indicates that the Transformer-based GAN successfully captured underlying structural and temporal correlations while introducing controlled variability.



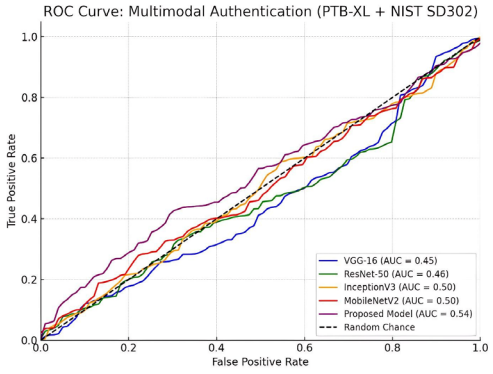
(a)

(b)

Fig. 9. Confusion matrices on (a) CMD1 and (b) CMD2.

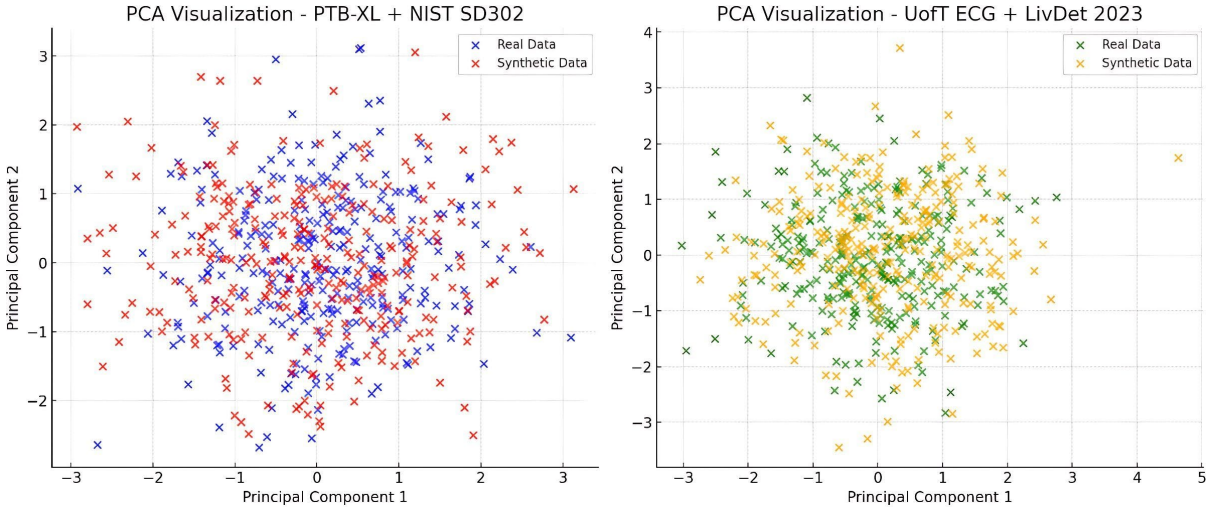


(a)



(b)

Fig. 10. The ROC curves for the multimodal authentication on (a) CMD1 and (b) CMD2.



(a) (b)

Fig. 11. PCA visualization of real and synthetic data distributions: (a) CMD1 and (b) CMD2.

The figure presents PCA-based visualizations of feature distributions for real and Transformer-GAN-generated (synthetic) biometric data across two multimodal authentication datasets: (a) CMD1 and (b) CMD2. In both cases, the blue/green clusters represent real data while the red/orange clusters correspond to synthetic data generated by the Transformer-based GAN. These results validate the statistical realism and diversity of the augmented dataset. The proximity between real and synthetic clusters implies that the augmented data not only enhances training diversity but also improves generalization and robustness of the multimodal authentication model against variations in acquisition conditions, noise, and spoofing attempts.

2) Computational cost and efficiency analysis

In addition to accuracy and security performance, evaluating the computational efficiency of the proposed multimodal authentication system is essential for determining its feasibility in real-world deployment. Computational cost was assessed in terms of training time, model size, and average inference latency across all evaluated models and dataset combinations. Tables VIII and IX shows the computation costs on CMD1 and CMD2 respectively.

TABLE VIII. COMPUTATIONAL COST ON CMD1

Model	Training Time (hours)	Model Size (MB)	Avg. Authentication Time (ms)	Memory Usage (GB)
VGG-16	4.2	138	88.6	5.8
ResNet-50	4.7	155	94.1	6.1
InceptionV3	5.1	173	106.3	6.5
MobileNetV2	2.6	62	54.2	3.3
Proposed Model	9.2	285	86.4	7.4

TABLE IX. COMPUTATIONAL COST ON CMD2

Model	Training Time (hours)	Model Size (MB)	Avg. Authentication Time (ms)	Memory Usage (GB)
VGG-16	3.9	138	85.9	5.4
ResNet-50	4.3	155	92.4	5.9
InceptionV3	4.8	173	105.8	6.3
MobileNetV2	2.3	62	52.6	3.1
Proposed Model	8.5	285	84.3	7.2

The proposed Enhanced VGG-16 with Transformer-GAN and Multi-SVM demonstrated superior authentication performance but at a higher computational cost compared to baseline models. Training the full multimodal framework including the GAN augmentation and weighted fusion mechanism took approximately 9.2 h on the CMD1 dataset and 8.5 h on the CMD2 dataset, nearly double the training duration of standard CNNs such as VGG-16 or ResNet-50. The model size also increased to 285 MB, primarily due to the inclusion of Transformer-based attention layers and multi-level fusion blocks. However, the average authentication (inference) time remained efficient, at around 84–86 ms, which is within the acceptable range for real-time biometric verification. Despite the increased training overhead, inference performance is competitive, ensuring feasibility for real-world deployment in security systems and access control environments. In comparison, MobileNetV2 achieved the lowest computational footprint (2.3–2.6 h training time, 62 MB model size, and ~52 ms authentication latency) but at the expense of significantly lower accuracy and robustness. The proposed system, while computationally heavier, delivers a favorable trade-off between performance and efficiency, offering improved resilience against spoofing and environmental noise while maintaining operational responsiveness. To enhance scalability, the framework can be optimized using model compression techniques such as pruning, quantization, or knowledge distillation, which would reduce size and training time while preserving accuracy. Such optimizations will make the proposed system more adaptable to edge devices and resource-constrained environments like wearable or mobile authentication platforms.

D. Ablation Study Results

To validate the effectiveness of individual components within the proposed multimodal biometric authentication framework, we conducted a detailed ablation study. To this end, we focus on three major architectural aspects which include T-GAN, Enhanced VGG-16 and Weighted Feature Fusion (WFF) of ECG and fingerprint modalities. Each component was selectively replaced to observe its isolated impact on system's performance. Experiments were conducted based on the combined CMD1 and CMD2 datasets. The model configurations used for the experiments is as follows:

a. Full Model: This involves T-GAN + Enhanced VGG-16 + WFF.
b. w/o T-GAN: This is the full model with no augmentation.
c. w/o Enhanced VGG-16: This is the model replaced with traditional VGG-16
d. ECG only: This is T-GAN + Enhanced VGG-16
e. Fingerprint only: This is T-GAN + Enhanced VGG-16

Fig. 12(a) and (b) shows the ablation results. From the results it can be shown that removing the TGAN led to the most significant drop in performance. As can be seen, the accuracy dropped by ~3–4%, EER increased by over 1.5%. This confirms the role of TGAN in enhancing the model. Replacing the enhanced feature extractor with a standard VGG-16 degraded performance by about 2.5%. This suggests that the upgraded VGG 16 led to better features extraction. The weighted fusion technique significantly beat basic feature concatenation which enhance the F1-Score and minimizing EER. This highlights the need of modality-aware fusion. When employed independently, ECG and fingerprint modalities performed pretty well, but fell short of the whole model. ECG-only setups notably suffered due to increased variability and susceptibility to signal noise. The combination of two modalities not only boosted accuracy but also dramatically lowered EER. This verify the complimentary nature of the two biometric features.

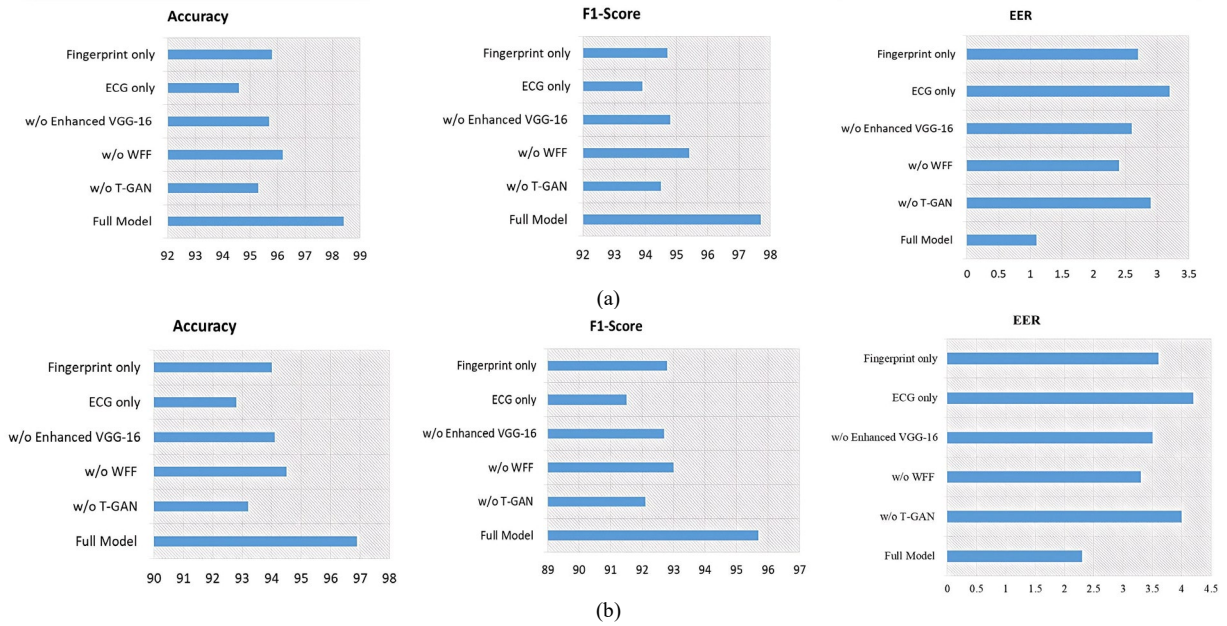


Fig. 12. Ablation results on (a) CMD1 and (b) CMD2.

E. Comparison with Existing Model

To further assess the performance of our proposed multimodal system, we also compare it with existing state of the art methods [23, 34–36]. The comparison results is given in Table X. From the results, it can be observed that, our proposed model outperforms all the compared models. Based on the results, it can be shown that our model obtains 97.9% accuracy. This shows a significant gain of (~2%) over others. This suggests that our proposed model provides strong complementary features. Essentially, the ECG + Iris model proposed in Ref. [36] shows a competitive performance of 95.65% which is slightly higher than most combinations. However, our proposed model outperforms it. In other hand, Multimodal models including knuckle or face features [34, 35] seem to perform worse. This is possibly due to lower discriminative power in those traits. The detail comparison results are shown Table X.

TABLE X. COMPARISON RESULTS

Ref	Modality	Dataset	Accuracy/%
[37]	spatial and temporal streams	UCF-101 +HMDB-51 (Multimodal)	94.2
[35]	Fingerprint, Iris, Knucle	CASIA, POLYU (Multimodal)	95
[34]	Fingerprint and ECG	SDUMLA-FV + CASIA-WebFace (Multimodal)	90
[23]	Fingerprint and ECG	livdet + ECG dataset (Multimodal)	95.32
[36]	ECG and IRIS	MMU Iris Dataset+ ECG-ID (Multimodal)	95.65
Our Proposed Model	ECG and Fingerprint	ECG-ID + LivDet (Multimodal)	97.9

The best results of our proposed model demonstrates a significant advancement in multimodal biometric systems using ECG and fingerprint modalities. This reflects potential improvements in feature extraction and fusion strategies.

VI. CONCLUSION

In this paper, we proposed a novel multimodal biometric authentication framework based on deep learning methods. The proposed model leverages the strengths of TGAN data augmentation, an enhanced VGG-16 architecture for feature extraction, and a weighted feature fusion strategy to integrate ECG and fingerprint biometrics. The proposed system addresses key limitations of unimodal authentication methods by effectively combination of physiological and anatomical signals for improving performances. Through extensive experiments on benchmark datasets PTB-XL, ECG-ID, FVC2004, and LivDet 2023 we demonstrated that our approach significantly outperforms several state-of-the-art baseline CNNs. To investigate the impact of the model components, we also conducted an ablation study. The results of the ablation confirmed that each component of the proposed framework plays a critical role in boosting overall performance. The proposed framework incorporates several defensive strategies at both the data and model levels to safeguard against spoofing, adversarial perturbations, replay attacks, and synthetic data injection. These countermeasures collectively ensure that the system maintains integrity, liveness assurance, and resilience under hostile conditions. Although the proposed framework's Transformer-based GAN augmentation and attention-driven feature extraction are theoretically expected to improve generalization and reduce vulnerability to adversarial noise, empirical validation through adversarial testing (e.g., Fast Gradient Sign Method or Projected Gradient Descent attacks) would provide stronger evidence. Future work should therefore incorporate controlled adversarial simulations to assess the model's defensive capabilities, quantify

robustness metrics such as attack success rate and perturbation tolerance, and explore adaptive countermeasures like adversarial training or feature smoothing to reinforce the system's practical security posture.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

A. A. conducted the research; N. H. K. and S.N.H.S.A. analyzed the data; A. D. wrote the paper; all authors had approved the final version.

ACKNOWLEDGMENT

The authors wish to thank the effort of reviewers for their comments and suggestions. We also appreciate the editor's efforts in facilitating the review process.

REFERENCES

[1] U. Sumalatha, K. K. Prakasha, S. Prabhu, and V. C. Nayak, "A comprehensive review of unimodal and multimodal fingerprint biometric authentication systems: Fusion, attacks, and template protection," *IEEE Access*, vol. 24, 64300–64334, 2024.

[2] A. J. Prakash, K. K. Patro, S. Samantray *et al.*, "A deep learning technique for biometric authentication using ECG beat template matching," *Information*, vol. 14, no. 2, 65, 2023. doi: 10.3390/info14020065

[3] M. V Madhusudhan, V. U. Rani, and C. Hegde, "Finger vein recognition model for biometric authentication using intelligent deep learning," *Int. J. Image Graph.*, vol. 23, no. 03, 2240004, 2023.

[4] N. Ammour, Y. Bazi, and N. Alajlan, "Multimodal approach for enhancing biometric authentication," *Journal of Imaging*, vol. 9, no. 9, 168, 2023.

[5] U. Sumalatha, S. Prabhu, and V. C. Nayak, "Multimodal biometric authentication: A novel deep learning framework integrating ECG, fingerprint, and finger knuckle print for high-security applications," *Eng. Res. Express*, vol. 7, no. 1, 015207, 2025.

[6] H. Heidari and A. Chalechale, "Biometric authentication using a deep learning approach based on different level fusion of finger knuckle print and fingernail," *Expert Syst. Appl.*, vol. 191, 116278, 2022.

[7] A. J. Prakash, K. K. Patro, M. Hammad *et al.*, "BAED: A secured biometric authentication system using ECG signal based on deep learning techniques," *Biocybern. Biomed. Eng.*, vol. 42, no. 4, pp. 1081–1093, 2022.

[8] A. R. Yuniarti, S. Rizal, and K. M. Lim, "Single heartbeat ECG authentication: A 1D-CNN framework for robust and efficient human identification," *Front. Bioeng. Biotechnol.*, vol. 12, 1398888, 2024.

[9] M. Hammad, M. A. Wani, K. A. Shakil *et al.*, "Deep cancelable multibiometric finger vein and fingerprint authentication with non-negative matrix factorization," *IEEE Access*, 2024.

[10] S. K. Natarajan, R. Rathinasabapathy, J. Narayanasamy, and A. R. Aravind, "Biometric user authentication system via fingerprints using novel hybrid optimization tuned deep learning strategy," *Trait. du Signal*, vol. 40, no. 1, 375, 2023.

[11] K. Shaheed *et al.*, "A systematic review on physiological-based biometric recognition systems: Current and future trends," *Arch. Comput. Methods Eng.*, pp. 1–44, 2021.

[12] K. K. Shinde and C. N. Kayte, "Fingerprint recognition based on deep learning pre-train with our best CNN model for person identification," *ECS Trans.*, vol. 107, no. 1, 2209, 2022.

[13] D. S. Ametefe, S. S. Sarnin, D. M. Ali, and Z. Z. Muhammad, "Fingerprint pattern classification using deep transfer learning and data augmentation," *Vis. Comput.*, vol. 39, no. 4, pp. 1703–1716, 2023.

[14] D. K. Jain, S. Neelakandan, A. Vidyarthi, and D. Gupta, "Deep learning-based intelligent system for fingerprint identification using decision-based median filter," *Pattern Recognit. Lett.*, vol. 174, pp. 25–31, 2023.

[15] M. Gayathri, C. Malathy, and M. Prabhakaran, "A review on various biometric techniques, its features, methods, security issues and application areas," in *Proc. International Conference on Computer Vision and Bio Inspired Computing (ICCVBIC 2019)*, 2019, pp. 931–941.

[16] V. Agrawal, M. Hazratifard, H. Elmiligi, and F. Gebali, "ElectroCardioGram (ECG)-based user authentication using deep learning algorithms," *Diagnostics*, vol. 13, no. 3, 439, 2023.

[17] M. Mageshbabu and J. Mohana, "Enhanced ECG-based biometric authentication using a hybrid CNN-LSTM framework," in *Proc. 2024 First International Conference on Software, Systems and Information Technology (SSITCON)*, 2024, pp. 1–7.

[18] Y. Zhang, Z. Zhao, Y. Deng *et al.*, "Human identification driven by deep CNN and transfer learning based on multiview feature representations of ECG," *Biomed. Signal Process. Control*, vol. 68, 102689, 2021.

[19] S. Salturk and N. Kahraman, "Deep learning-powered multimodal biometric authentication: Integrating dynamic signatures and facial data for enhanced online security," *Neural Comput. Appl.*, vol. 36, no. 19, pp. 11311–11322, 2024.

[20] A. K. Yadav, "Deep learning approach for multimodal biometric recognition system based on fusion of iris, fingerprint and hand written signature traits," *Turkish J. Comput. Math. Educ.*, vol. 12, no. 11, pp. 1627–1640, 2021.

[21] A. Alshardan *et al.*, "Multimodal biometric identification: Leveraging Convolutional Neural Network (CNN) architectures and fusion techniques with fingerprint and finger vein data," *PeerJ Comput. Sci.*, vol. 10, e2440, 2024.

[22] M. Hammad and K. Wang, "Parallel score fusion of ECG and fingerprint for human authentication based on convolution neural network," *Comput. Secur.*, vol. 81, pp. 107–122, 2019.

[23] R. M. Jomaa, H. Mathkour, Y. Bazi, and M. S. Islam, "End-to-end deep learning fusion of fingerprint and electrocardiogram signals for presentation attack detection," *Sensors*, vol. 20, no. 7, 2085, 2020.

[24] Y. Li, L. Liu, H. Qin *et al.*, "Adaptive deep feature fusion for continuous authentication with data augmentation," *IEEE Trans. Mob. Comput.*, vol. 22, no. 10, pp. 5690–5705, 2022.

[25] A. Buriro, F. Ricci, and B. Crispo, "SWIPEGAN: Swiping data augmentation using generative adversarial networks for smartphone user authentication," in *Proc. the 3rd ACM Workshop on Wireless Security and Machine Learning*, 2021, pp. 85–90.

[26] P. Bhardwaj, K. Yadav, H. Alsharif, and R. A. Aboalela, "GAN-based unsupervised learning approach to generate and detect fake news," in *Proc. International Conference on Cyber Security, Privacy and Networking*, 2023, pp. 384–396. doi: 10.1007/978-3-031-22018-0_37

[27] I. J. Goodfellow, J. Pouget-abadie, M. Mirza *et al.*, "Generative adversarial nets," *Advances in Neural Information Processing Systems*, vol. 27, 2014.

[28] K. Sun, Q. Wen, and H. Zhou, "Ganster R-CNN: Occluded object detection network based on generative adversarial nets and faster R-CNN," *IEEE Access*, vol. 10, 2022.

[29] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv preprint, arXiv:1409.1556, 2014.

[30] P. Wagner *et al.*, "PTB-XL, a large publicly available electrocardiography dataset," *Sci. Data*, vol. 7, no. 1, pp. 1–15, 2020.

[31] T. S. Lugovaya. (2005). ECG-ID database. [Online]. Available: https://physionet.org/content/ecgiddb/

[32] D. Maio, D. Maltoni, R. Cappelli, A. K. Jain, and S. Prabhakar. (2004). FVC2004 fingerprint database. [Online]. Available: http://bias.csr.unibo.it/fvc2004/

[33] M. Micheletto *et al.*, "LivDet2023-fingerprint liveness detection competition: Advancing generalization," in *Proc. 2023 IEEE International Joint Conference on Biometrics (IJCB)*, 2023, pp. 1–8.

[34] 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, 6039, 2022.

[35] K. Aizi and M. Ouslim, "Score level fusion in multi-biometric identification based on zones of interest," *J. King Saud Univ. Inf. Sci.*, vol. 34, no. 1, pp. 1498–1509, 2022.
[36] K. Ashwini, G. N. K. Murthy, S. Raviraja, and G. A. Srinidhi, "A novel multimodal biometric person authentication system based on ecg and iris data," *Biomed Res. Int.*, vol. 2024, no. 1, 8112209, 2024.
[37] N. Yudistira and T. Kurita, "Correlation net: Spatiotemporal multimodal deep learning for action recognition," *Signal Process. Image Commun.*, vol. 82, 115731, 2020.