Journal of Robotics and Control (JRC)
Volume 6, Issue 3, 2025
ISSN: 2715-5072, DOI: 10.18196/jrc.v6i3.26490

Efficient Multimodal Biometric Identification via Gabor-Enhanced Attention Networks

Phan Minh Than [1], Hoanh Nguyen [2*]
[1, 2] Faculty of Electrical Engineering Technology, Industrial University of Ho Chi Minh City, Ho Chi Minh City, Vietnam
Email: [1] phanminhthan@iuh.edu.vn, [2] nguyenhoanh@iuh.edu.vn
*Corresponding Author

***Abstract*—Achieving robust multimodal biometric identification requires advanced feature extraction strategies and effective integration of diverse data modalities. Conventional methods often encounter limitations such as computational complexity and degradation of critical information during feature transformation. Although deep learning models address feature extraction challenges, their heavy architectures hinder real-world deployment. Moreover, traditional fusion strategies, based mainly on simple concatenation, overlook critical intermodal correlations, leading to suboptimal recognition accuracy. In this study, we propose a lightweight Gabor Attention Network framework designed for efficient multimodal biometric recognition. Our approach utilizes learnable Gabor filters to capture detailed local and directional features with enhanced precision and reduced computational burden compared to standard convolutions. We further introduce a convolutional attention mechanism that adaptively refines intermediate feature representations, and a novel attention-driven fusion architecture that dynamically models and exploits intermodal dependencies. Extensive experiments on two multimodal datasets demonstrate that our model achieves superior performance compared to several state-of-the-art methods, attaining up to 99.49% accuracy and 0.35% Equal Error Rate, while maintaining high efficiency with only 10.6M parameters, 0.85 GFLOPs, and 60 FPS inference speed. These results highlight the effectiveness of our biologically inspired and attention-enhanced design for achieving high-accuracy, low-complexity multimodal biometric identification.**

Keywords—Gabor Attention Networks; Dynamic Attention Mechanism; Feature Fusion; Multimodal Biometrics; Biometric Recognition.

I. INTRODUCTION

The continuous evolution of information technology has significantly influenced personal identification mechanisms, making them critical components of contemporary security infrastructures [1]. Traditional identification approaches, such as passwords and physical tokens, inherently suffer from vulnerabilities including loss, theft, and easy compromise. Consequently, biometric recognition methods, which identify individuals based on unique physiological or behavioral characteristics, have attracted considerable interest due to their superior reliability and security. Prominent biometric methods include facial recognition, fingerprint scanning, finger-vein identification, palmprint verification, and iris recognition. Recently, extensive research efforts in unimodal biometric identification systems have demonstrated promising results, highlighting their remarkable effectiveness in various security-related scenarios [2]-[15]. Despite these advancements, unimodal biometric approaches have inherent limitations. Environmental noise can distort biometric images, reducing the signal-to-noise ratio (SNR) and degrading feature extraction accuracy. Variations in biometric traits due to aging or sensor inconsistencies further impair model generalization. Moreover, unimodal systems are particularly vulnerable to spoofing attacks, wherein artificial replicas of biometric traits (e.g., fake fingerprints, 3D face masks) can significantly deceive recognition systems, with attack success rates reported as high as 70% in some benchmarks. These factors collectively lead to unreliable authentication outcomes in practical deployments. To overcome these challenges, researchers have proposed multimodal biometric systems that integrate multiple biometric modalities [16]-[19]. By leveraging complementary features across different biometric traits, multimodal systems aim to enhance recognition accuracy, robustness, and anti-spoofing resistance. Consequently, multimodal biometrics has emerged as an essential area of study, offering enhanced performance in identity verification, security monitoring, and other critical applications. Extensive research has explored combinations such as face-fingerprint, face-palmprint, and palmprint-iris fusion. Among these, finger-based modalities, including finger-vein, fingerprint, palmprint, and knuckle print, have garnered particular attention due to user convenience, stable acquisition, and multimodal complementarity. However, effectively integrating multimodal biometric features remains challenging. Traditional fusion strategies often rely on naive feature concatenation, which ignores semantic inconsistencies and fails to model intrinsic intermodal correlations, resulting in redundant feature spaces and degraded classification performance. Additionally, existing deep learning approaches frequently suffer from excessive computational complexity, making real-time deployment on resource-constrained systems difficult. Thus, despite its promising potential, the full advantages of multimodal biometrics have yet to be fully realized.

Deep learning, particularly Convolutional Neural Networks (CNNs), has recently demonstrated extraordinary success across a wide array of applications, particularly in computer vision tasks such as image classification, object detection, and biometric identification. CNNs' ability to automatically learn hierarchical features from raw data without extensive manual preprocessing has led to significant performance enhancements. Beyond traditional vision tasks, deep learning has also been effectively applied in diverse domains such as dentistry (e.g., image segmentation and

treatment planning), voice and deepfake detection, UAV path planning, epileptic seizure detection, sign language recognition, diabetes prediction, intrusion detection in IoT, river water level estimation, pediatric psychological evaluation through children's drawings, and dengue disease prediction [20]-[40]. These broad applications underscore the adaptability and robustness of CNNs and other deep learning models in solving complex, real-world problems. Importantly, the success of deep learning in these varied fields further motivates its application in the multimodal biometrics domain, where accurate, efficient, and robust feature learning across heterogeneous modalities remains a significant challenge requiring more specialized architectural innovations.

Based on this analysis, we propose a novel and lightweight Gabor attention framework specifically designed for multimodal biometric recognition. Unlike traditional convolutional methods, our approach utilizes learnable Gabor filters to accurately capture both local texture and directional edge features, thereby improving recognition precision while reducing computational cost. In addition, we integrate a convolutional attention mechanism to strengthen intermediate feature representations by adaptively emphasizing salient spatial and channel-wise information. Furthermore, we develop an advanced attention-driven fusion architecture capable of dynamically modeling intermodal correlations, enabling optimal synergy between distinct biometric modalities. Extensive evaluations performed on two publicly available multimodal datasets confirm that our proposed framework significantly outperforms conventional methods, achieving superior accuracy, lower Equal Error Rates (EER), and reduced model complexity, thereby offering a highly effective and efficient solution for practical multimodal biometric identification.

The remainder of this paper is organized as follows. Section II reviews recent developments in multimodal biometric systems, highlighting relevant advances and existing challenges. Section III details the proposed lightweight attention-based feature fusion framework, including the Gabor layer, feature enhancement, and feature fusion mechanisms. Section IV presents the experimental setup, performance evaluation, and critical discussions. Finally, Section V concludes the paper and outlines future research directions.

II. LITERATURE REVIEW

Multimodal biometrics, combining two or more biometric traits (e.g., face, fingerprint, iris, voice), have attracted growing attention in the past five years due to their ability to address the shortcomings of unimodal systems [41]. By consolidating multiple sources of evidence, these approaches typically yield higher accuracy and stronger anti-spoofing resistance than any single modality [42]. An illustrative example is an Android-based authentication mechanism that integrates face and voice, which significantly reduced error rates on resource-constrained devices [43]. In other cases, researchers have combined different pairs of physiological and behavioral biometrics (e.g., heart signals, iris) to enhance robustness and combat noise [44]. Several works also underscore how kernel-based fusion or correlation methods can successfully align ear, face, and other traits to improve recognition performance [45]. Moreover, the design of novel architectures for multimodal authentication continues to evolve, leading to deeper exploration of convolutional strategies [46].

Recently, deep learning has emerged as the primary driver of multimodal biometric integration, automating the extraction of discriminative features for diverse traits [47]. One of the earliest deep learning-based multimodal prototypes fused face, iris, and finger vein data, achieving greater reliability than unimodal counterparts [41]. Subsequent efforts employed deep score-level fusion, such as combining face and palmprint classifiers to outperform older late-fusion approaches [48]. Alongside these achievements, comprehensive surveys emphasize that modern CNN- and DNN-based architectures can effectively handle the complexity of multi-biometric input data [49]. Furthermore, advanced systems have successfully incorporated voice and face traits in tandem, yielding higher matching accuracy for user authentication tasks [50].

In terms of fusion strategies, researchers have investigated feature-level, score-level, and decision-level integration [51]. Weighting the match scores from fingerprint and iris proved remarkably robust under fluctuating image quality, outperforming simpler rule-based combinations [52]. Similarly, some studies used correlation-based approaches to align face, iris, and fingerprint features in a unified space, while others combined parallel pipelines through SVM- or RF-based fusion for high reliability [53]. A classic example is Garg et al.'s method that performed a novel face-fingerprint feature integration to handle noisy data [54]. Score-level strategies have likewise attracted attention for their simplicity in combining unimodal classifiers [55]. Beyond these, additional work has explored real-time iris recognition pipelines, highlighting the necessity of stable feature extraction methods for effective multimodal synergy [56]. Some researchers also have studied cross-spectral matching, bridging data from differing sensors (e.g., visible and near-infrared) [57].

Recently, attention mechanisms and transformers have become highly influential in multimodal biometrics, enabling adaptive weighting of each modality's contribution [58]. For instance, a framework combining fingerprint images with heart signals leveraged a vision-transformer-based module that drastically improved detection of spoofing attacks [59]. Another study devised a dual-branch CNN with branch attention to emphasize the most salient aspects of each modality's feature map. In a federated learning context, an attention-based fusion aggregator was introduced for IoT-based biometric systems, demonstrating robust performance with a decentralized approach [60]. Building on these ideas, AuthFormer was proposed for elderly user authentication, dynamically adjusting the relative importance of each modality at inference time [61]. For security-critical scenarios, hashing-based methods can be layered onto attention-based fusion to safeguard stored templates [62]. Complementary investigations have examined advanced feature-level integrations, highlighting the value of synergy in bridging multiple biometric data streams [63]. Meanwhile, high-level reviews call for even broader information fusion

research to tackle emerging challenges in real-world deployments [64].

Moreover, an increasing number of works emphasize lightweight and IoT-friendly models that allow for on-device or edge-based deployment [65]-[68]. One representative system uses a mobile face-voice approach with carefully designed CNN blocks to limit the computational footprint [69]. Another focuses on finger multimodal fusion but employs depthwise separable convolutions to keep resource usage low [70]. Practical solutions also appear in federated learning setups, reducing the need to centralize potentially sensitive data. Additional efforts present narrower network topologies or specialized compression schemes to support high-speed processing while maintaining accuracy [71]. Some authors have also studied sensor-level concerns, outlining the challenge of inconsistent data quality in real-world settings [72]. Current progress in contactless fingerprint recognition and iris analysis further indicates that efficient deep learning components can preserve speed and performance [73]. Despite significant advances, tackling noise, variability, and scalability remains an ongoing pursuit.

Overall, recent years have seen transformative progress in multimodal biometrics, evolving from CNN-based feature extractors to attention-enabled transformers and lightweight architectures. These developments underscore how consolidating multiple biometric signals can enhance both accuracy and resilience, thereby overcoming fundamental hurdles of single-modality systems. Nonetheless, unresolved challenges persist in aligning heterogeneous data streams, improving interpretability, and maintaining user privacy as large-scale multimodal deployments become increasingly common. Addressing these dimensions will be crucial to further refining multimodal biometric technology in the years ahead.

III. METHODOLOGY

A. Overall Structure

Our proposed model, as shown in Fig. 1, is a lightweight attention-based feature fusion framework specifically designed for multimodal biometric identification using finger-vein and palmprint images. It comprises four primary stages: Gabor-based feature extraction, attention-driven feature enhancement, attention-driven feature fusion, and a final classification step. By appropriately leveraging the complementary information provided by these two biometric modalities, the model achieves robust and accurate recognition results while preserving a low computational footprint.

The network begins with two parallel branches, each dedicated to processing one of the modalities. For each branch, the input data first passes through a Gabor layer, which captures orientation- and frequency-sensitive features significant for identifying subtle edge and texture patterns. This layer is followed by a series of lightweight convolutional blocks involving operations like 3×3 convolutions, batch normalization, and ReLU activations that refine and expand the extracted features into more discriminative representations. Next, each modality-specific feature map proceeds to an attention mechanism for feature enhancement (AMFE). This component applies both channel-wise and spatial attention to adaptively emphasize the most relevant biometric cues within each feature map. By focusing on critical channels and spatial locations, the AMFE effectively diminishes noise and highlights essential information, ensuring that each modality retains its unique advantages before the fusion process.

Once the feature maps have been enhanced, the network merges them using an attention mechanism for feature fusion, referred to as AMFF. This module addresses potential inconsistencies in semantic features by adaptively aligning and combining the modality-specific representations, consolidating complementary details from finger-vein and palmprint images. Through careful weighting of these multimodal features, the AMFF produces a cohesive representation with heightened discriminatory power. Finally, the fused feature map undergoes a classification step to finalize the identification or verification outcome.

This streamlined approach ensures that the model remains computationally efficient while still capitalizing on the rich information present in both finger-vein and palmprint modalities. By integrating biologically inspired Gabor filtering, attention-enhanced feature extraction, and attention-based fusion, this lightweight framework delivers robust multimodal biometric performance suitable for real-world applications.

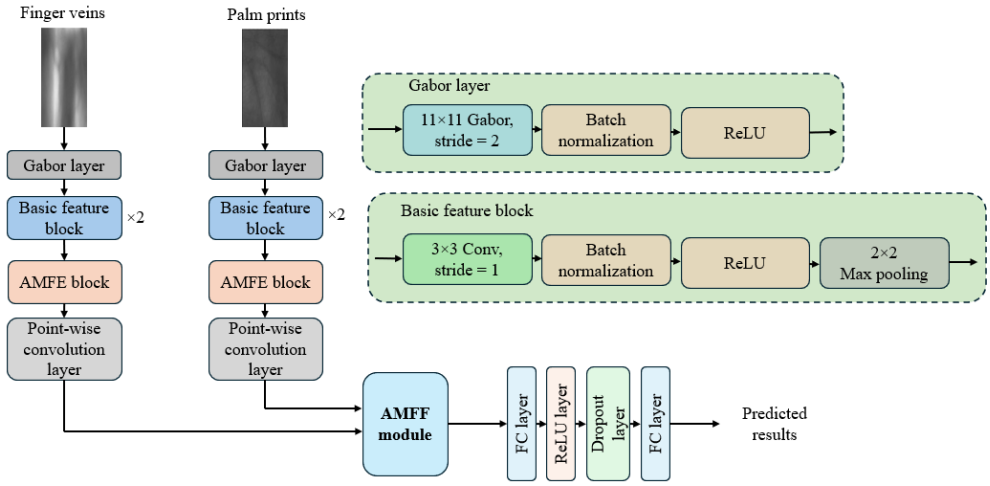

Fig. 1. Overview of the proposed lightweight attention-based feature fusion framework for multimodal biometric identification

B. Gabor Layer

Motivated by recent successes in deploying Gabor filters for efficient training convergence and effective feature extraction, our multimodal biometric recognition backbone integrates a specially designed Gabor layer as its initial feature extraction stage. This choice is influenced by the strong ability of Gabor filters to highlight image textures and edges, properties analogous to the early stages of human visual perception. Incorporating such biologically inspired characteristics ensures a more natural and robust extraction of features from multimodal biometric data, thus promoting enhanced recognition performance.

Specifically, the Gabor layer leverages a Gabor filter derived from a sinusoidal function modulated by a Gaussian envelope. This carefully designed mathematical construct fulfills key requirements, including differentiability, smoothness, and localized spatial-frequency characteristics. To mathematically represent the real part of the Gabor filter, the following formulation is used:

$$h(u,v,f,\alpha,\beta,\gamma) = \exp\left(-\frac{u'^2+\gamma^2 v'^2}{2\beta^2}\right)\cos(2\pi f u'+\alpha) \quad (1)$$

with rotated coordinates defined as:

$$u' = u\cos(\Phi) + v\sin(\Phi) \quad (2)$$

$$v' = -u\sin(\Phi) + v\cos(\Phi) \quad (3)$$

Here, f denotes the spatial frequency of the sinusoidal factor, controlling the granularity of features captured; β represents the scale or standard deviation of the Gaussian envelope, defining the size of the receptive field; Φ indicates the orientation angle, governing the directional sensitivity; α specifies the phase offset; and γ represents the spatial aspect ratio, determining the ellipticity of the Gabor kernel. The parameters f, α, β, Φ, and γ are all learnable during network training to ensure optimal adaptability to the input biometric data. To effectively initialize these parameters and facilitate training stability, frequency and orientation parameters are defined as follows:

$$f_m = \frac{\pi}{2}.2^{-\frac{(m-1)}{2}}, m \in [1,5] \quad (4)$$

$$\Phi_n = \frac{\pi}{8}(n-1), n \in [1,8] \quad (5)$$

The phase offset α is initialized uniformly from a distribution $U[0,\pi)$, ensuring diverse initial feature extraction capabilities.

For an input biometric image X, the feature representation obtained by applying the Gabor layer for each output feature map F_i is computed by convolving the input channels with their respective Gabor filters and aggregating the results:

$$F_i = \sum_c X_c * G_{ic} \quad (6)$$

where X_c is the c-th input channel, the symbol $*$ denotes the convolution operation, and G_{ic} represents the learned Gabor convolution kernel weights corresponding to the c-th input channel and the i-th output feature channel. Each kernel G_{ic} is defined as:

$$G_{ic}(u,v) = \exp\left(-\frac{u'^2+\gamma_{ic}^2 v'^2}{2(\beta_{ic}^2+\epsilon)}\right)\cos(2\pi f_{ic}u' + \alpha_{ic}).\frac{1}{2\pi(\beta_{ic}^2+\epsilon)} \quad (7)$$

where ϵ is a small positive constant added to prevent division by zero, ensuring numerical stability. This normalization term ensures scale invariance, allowing consistent extraction of features irrespective of input image resolutions and scales. Such normalization results in smoother gradient distributions, improving training convergence and network stability.

Overall, the integration of this flexible and learnable Gabor layer significantly enhances the backbone's ability to accurately and efficiently extract discriminative biometric features across multiple modalities, paving the way for improved multimodal biometric recognition accuracy and robustness. Although Gabor convolutions are initially heavier than simple convolutions, their compact kernels (7×7 or smaller) and sparse feature selection reduce downstream computational loads, improving the overall efficiency of the pipeline.

C. Attention Mechanism for Feature Enhancement

To further improve the network's capability of selectively focusing on essential biometric features, we introduce a specialized attention mechanism for feature enhancement (AMFE), as shown in Fig. 2. This module strategically combines channel-wise and spatial attention maps to effectively highlight crucial regions and relevant channels in the biometric feature maps. By explicitly modeling interdependencies within and across feature channels and spatial dimensions, the proposed attention mechanism enables the network to adaptively prioritize informative features while suppressing irrelevant ones, significantly enhancing the accuracy of multimodal biometric identification.

In particular, channel attention focuses on identifying and emphasizing critical feature channels, serving as a selector that dynamically prioritizes channels containing the most discriminative biometric information. Traditional approaches typically rely solely on average pooling to condense spatial information, potentially missing significant discriminative details. Conversely, our approach leverages a combined pooling strategy, simultaneously applying average pooling and max pooling operations independently across spatial dimensions. The motivation behind this dual pooling operation arises from the observation that average pooling efficiently captures general contextual information, whereas max pooling excels at isolating prominent and distinctive feature responses. These complementary spatial descriptors are then individually processed through a shared network consisting of 1×1 convolutional layers, followed by an element-wise summation, producing the refined channel attention map M_{ch}. The channel attention map is computed as eq. (8).

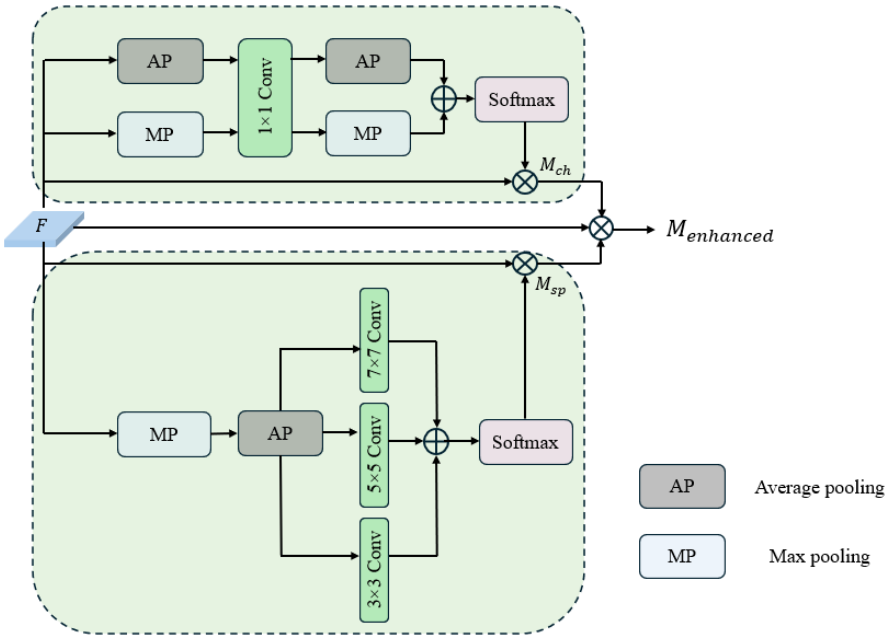

Fig. 2. Architecture of the attention mechanism for feature enhancement (AMFE), which incorporates both channel-wise and spatial attention modules to amplify the most discriminative biometric features from each modality

$$M_{ch} = \sigma(Conv_{1\times1}(AvgPool(F)) + Conv_{1\times1}(MaxPool(F))) \quad (8)$$

Here, σ represents the sigmoid activation function, $Conv_{1\times1}$ denotes a convolutional layer with kernel size 1×1, and F indicates the input feature map. Dual pooling captures both global context (via average pooling) and prominent features (via max pooling), leading to a more comprehensive attention descriptor. Compared to single pooling, dual pooling introduces minor computational cost but improves robustness, particularly for heterogeneous biometric data.

Additionally, to strengthen the spatial focus of the network, we introduce a spatial attention mechanism explicitly designed to capture significant spatial regions across various biometric modalities. Unlike channel attention, spatial attention targets local regions within feature maps, highlighting spatially relevant positions. We independently apply channel-wise average pooling and max pooling operations, concatenating these pooled features along the channel dimension to create a more informative descriptor. Subsequently, we utilize a multi-scale convolutional approach employing convolutional layers with kernel sizes of 3×3, 5×5, and 7×7 to robustly capture spatial information at multiple scales. The spatial attention map M_{sp} is computed using the following formula:

$$\begin{aligned} &M_{sp} \\ &= \sigma(Conv_{3\times3}[AvgPool(F);\ MaxPool(F)] \\ &+ Conv_{5\times5}[AvgPool(F);\ MaxPool(F)] \\ &+ Conv_{7\times7}[AvgPool(F);\ MaxPool(F)]) \end{aligned} \quad (9)$$

Here, $Conv_{3\times3}$, $Conv_{5\times5}$, and $Conv_{7\times7}$ represent convolutional operations with different kernel sizes, and the brackets $[.;.]$ indicate channel-wise concatenation. While multi-scale convolutions slightly increase FLOPs, they provide significantly stronger feature localization across different biometric scales, improving model generalization. All convolutions are lightweight (small number of channels), ensuring minimal overhead.

Building upon these two distinct yet complementary attention maps, the attention mechanism further integrates channel and spatial attention to achieve optimal feature enhancement. The channel attention map initially emphasizes important biometric feature channels, while the subsequent spatial attention map precisely locates these significant features within the spatial domain. Through sequential combination, the proposed module effectively highlights essential biometric information while filtering out redundant or noisy features. Consequently, the fusion of channel and spatial attention maps generates a refined and enhanced feature representation $M_{enhanced}$, defined as follows:

$$M_{enhanced} = M_{ch}.M_{sp}.F \quad (10)$$

In the above expression, . denotes element-wise multiplication, M_{ch} is the channel attention map, M_{sp} is the spatial attention map, and F is the original input feature map.

The AMFE block with specialized attention mechanisms significantly enhances multimodal biometric feature extraction by explicitly guiding the network to emphasize highly discriminative channels and spatial locations, thereby effectively improving both recognition accuracy and generalization capabilities.

D. Attention Mechanism for Feature Fusion

To effectively address the challenge of inconsistent semantic representations across multimodal biometric features, we propose a specialized attention mechanism explicitly designed for feature fusion (AMFF). This attention-based fusion module integrates diverse biometric modality features by adaptively aligning their semantic content, ensuring consistency and enhancing overall recognition accuracy. The structure of our fusion module, as illustrated in Fig. 3, comprises two distinct attention branches, each dedicated to the specific characteristics of its respective biometric modality.

For biometric modalities characterized by richer semantic content or superior quality (e.g., palm-print), we implement a global attention (GA) mechanism. This mechanism is particularly efficient in capturing comprehensive global context information within feature maps. The GA module first utilizes global average pooling to summarize the spatial dimensions, extracting a robust representation of the global feature context. Subsequently, these pooled representations are processed through a Multi-Layer Perceptron (MLP), effectively refining the global contextual details. The final global semantic vectors obtained through this process robustly preserve the comprehensive contextual information inherent to high-quality biometric modalities.

Conversely, for biometric modalities typically subject to lower quality or significant noise (e.g., finger-vein), we employ a local attention (LA) mechanism. This localized attention approach excels in isolating and highlighting essential fine-grained biometric details that may otherwise be obscured by noise or artifacts. Specifically, the LA mechanism leverages global max pooling, which focuses explicitly on the most prominent spatial features, capturing distinct and discriminative local patterns. The pooled results are then refined through an MLP network, ensuring the effective preservation and enhancement of local discriminative features critical to biometric identification.

Formally, let the enhanced features of two biometric modalities be represented as F_A and F_B, respectively. These features are individually processed to derive modality-specific attention vectors A and B:

$$A = MLP(GAP(F_A)) \tag{11}$$

$$B = MLP(GMP(F_B)) \tag{12}$$

where GAP and GMP denote global average pooling and global max pooling operations, respectively. The vectors A and B capture essential modality-specific semantic characteristics, effectively resolving inconsistencies between the two modalities.

Subsequently, the modality-specific attention vectors undergo an element-wise interaction, producing a cross-modality attention vector. This vector serves as a dynamic fusion weight, ensuring optimal integration of modality-specific semantics:

$$F'_A = \sigma(A \otimes B) \otimes F_A \tag{13}$$

$$F'_B = \sigma(A \otimes B) \otimes F_B \tag{14}$$

where σ represents the Sigmoid activation function, and $\otimes$ denotes element-wise multiplication. This procedure selectively amplifies semantic features with complementary information across modalities, enhancing the semantic consistency of the fused representation.

To achieve precise fusion, we further employ convolutional operations with a kernel size of 1×1 to combine these attention-enhanced features, followed by a Softmax activation function. This combination allows the network to adaptively segregate and integrate multimodal features, optimizing their representations for biometric identification. The fusion of modality-specific feature maps is formulated as:

$$\tilde{F}_A, \tilde{F}_B = Split(Softmax(Concat(Conv_{1\times1}(F'_A), Conv_{1\times1}(F'_B)))) \tag{15}$$

The final fused feature map F_{fuse} leveraging complementary semantic strengths from both modalities is computed by integrating the original modality-enhanced features with attention-refined outputs:

$$F_{fuse} = \tilde{F}_A \otimes F'_A + \tilde{F}_B \otimes F'_B \tag{16}$$

This dual-branch fusion dynamically adapts to modality characteristics, substantially improving semantic consistency and achieving better multimodal integration compared to simple concatenation or static fusion. By explicitly addressing the issue of semantic inconsistency through this adaptive attention-based fusion approach, our method significantly enhances the semantic alignment between multimodal biometric features, thereby substantially improving biometric identification accuracy and robustness.

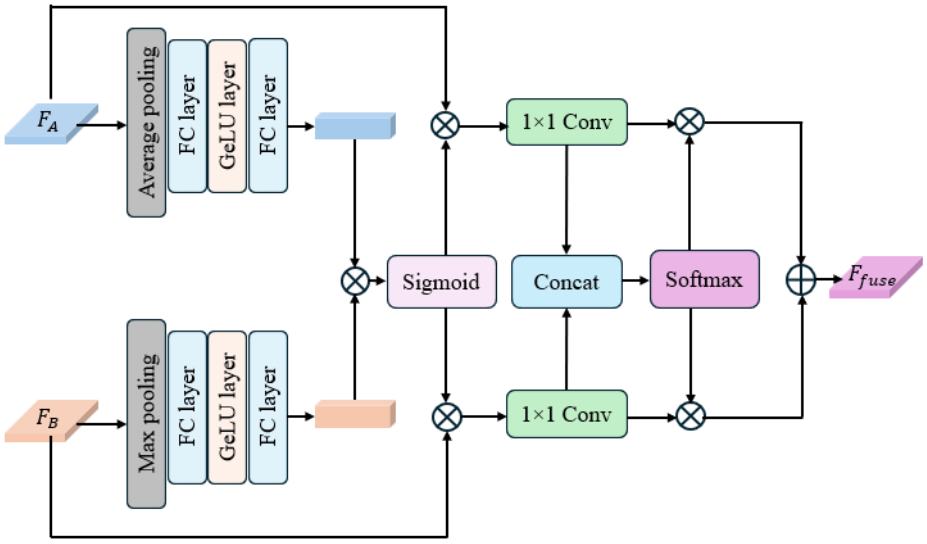

Fig. 3. Diagram of the attention mechanism for feature fusion (AMFF), showing modality-specific attention branches and their dynamic semantic alignment and integration process

IV. RESULTS AND DISCUSSIONS

A. Datasets

We evaluated our proposed model using one publicly available palmprint dataset (PALM [74]) and two open-access finger vein datasets (FV1 [75] and FV2 [76]). From these sources, we constructed two multimodal datasets, called PALM+FV1 and PALM+FV2, to investigate the effectiveness and robustness of our method under different scenarios.

In the PALM database, 300 participants contributed images across two separate sessions. For each session, every individual provided 10 palmprint images, resulting in 600 distinct classes, each containing 20 total samples and culminating in 12,000 images in total. We extracted the regions of interest (ROI) following the process described in [74], generating 200×100-pixel ROIs. The FV1 finger vein database includes samples from 156 individuals, each providing six images for both index and middle fingers, totaling 312 distinct fingers. After applying the extraction approach in [75], each ROI measured 81×333 pixels. Meanwhile, the FV2 finger vein database involves 123 participants, each providing six images per session for four different fingers, yielding 492 distinct finger classes in total. Across two sessions, 5904 images were collected, with each ROI sized at 300×100 pixels.

We combined the palmprint and finger vein data in two distinct ways. First, PALM+FV1 aligns the palmprints of 210 classes from PALM with the corresponding 210 classes from FV1. Since FV1 supplies 12 finger vein images per finger across two sessions, we took the first six images from both PALM sessions to match this structure. Second, PALM+FV2 pairs 492 classes from PALM with the 492 classes from FV2, which has 12 samples per class. We chose 12 images per class from both sessions of the PALM dataset to ensure consistency in class size and session alignment. These two new multimodal datasets, PALM+FV1 and PALM+FV2, provide a comprehensive basis for evaluating the performance of biometric recognition methods that fuse palmprint and finger vein information. By ensuring parallel class structures and session counts, we offer a robust framework for testing cross-modal and cross-session generalization in a single experimental setting.

B. Experimental Setups

All experiments were conducted on a workstation equipped with an Intel Core i7-14700K CPU and an NVIDIA RTX 4080 GPU. Our implementation relied on PyTorch as the deep learning framework. We ran each experiment for a total of 100 epochs, using the AdamW optimizer to update network parameters. An initial learning rate of 0.0001 proved effective in balancing convergence speed and stability, and we applied a weight decay of 0.00001 to mitigate overfitting. We used a batch size of 32 for training and a batch size of 64 for evaluation. During training, we randomly divided each dataset into training (70%), validation (10%), and test (20%) sets. To capture the natural variability of biometric data and improve generalization, we employed data augmentation strategies such as random horizontal flipping and random cropping on the palmprint images, as well as random rotation on the finger-vein images. We used cross-entropy loss as the primary objective function for all model variants.

We followed a 5-fold repetition strategy, reinitializing model parameters and data splits at each run. This repetition helped confirm that our performance metrics reflect consistent improvements rather than singular, luck-based results. At inference time, our method computes class probabilities for each test sample, returning the most probable class as the final prediction. Throughout these experiments, we measured performance using accuracy and Equal Error Rate (EER). Unless explicitly noted, we report the average and standard deviation of these metrics over the 5-fold trials. This combination of hardware capabilities, training protocols, and evaluation methods ensures a rigorous examination of the proposed model's behavior under various conditions.

C. Multimodal Identification Results

In Fig. 4, the training accuracy curves for the proposed approach steadily decrease over the course of 100 epochs, reflecting stable convergence without significant fluctuations or abrupt spikes. Concurrently, the training accuracy rises consistently and plateaus at a high level, indicating that the network effectively learns discriminative representations for both palmprint and finger-vein data. Notably, the validation curves exhibit a similar trend, suggesting minimal overfitting and a well-generalized model. A closer look at the curves reveals that the Gabor layer and attention modules help maintain a relatively smooth training process compared to standard CNN-based methods, with fewer oscillations in the loss. Overall, these observations underscore the effectiveness of the lightweight attention design and confirm that the training procedure progresses in a controlled and efficient manner.

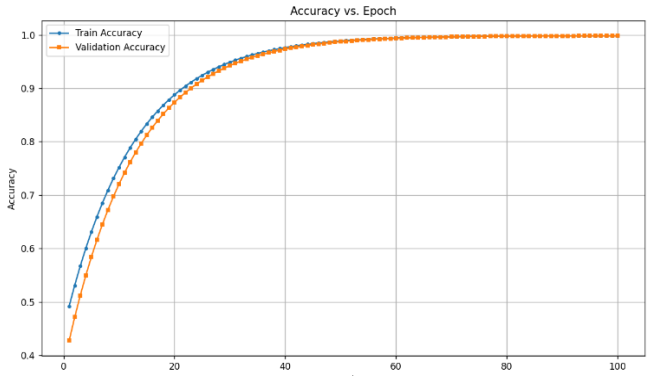

Fig. 4. Training and validation accuracy curves over 100 epochs for the proposed model

Table I demonstrates the superior classification performance of our method compared to various benchmark techniques across both PALM+FV1 and PALM+FV2 datasets. For PALM+FV1, our model achieves a remarkable 98.73% accuracy and a minimal EER of 0.89%, surpassing other methods such as NLNet and Enhanced DenseNet by noticeable margins in both metrics. In particular, the inclusion of Gabor-based feature extraction appears to capture essential edge and texture details that complement the more conventional convolutional representations. Attention-driven feature enhancement and fusion further amplify this

advantage by selectively focusing on critical channels and spatial regions, thereby mitigating the impact of noise and redundant cues. Similarly, on the PALM+FV2 dataset, our model attains a leading 99.49% accuracy with an EER of 0.35%, setting a new benchmark and reflecting the approach's robustness to variations in sample quality and texture patterns. The margin of improvement over other fusion-based strategies, such as Att-CNN and FAB-AEF, emphasizes the importance of adapting attention mechanisms to both global and local details across multiple modalities. Moreover, the consistently low EER values indicate that the proposed model maintains stable performance even under stricter false acceptance or false rejection constraints. These comprehensive gains across two distinct multimodal sets suggest that the network architecture, with its attention-driven and Gabor-oriented design, can capture nuanced biometric traits that remain elusive to standard CNNs. Overall, the results highlight the synergy of leveraging biologically inspired filters and attention-based fusion to deliver both high accuracy and robust error-tolerance for multimodal identification tasks.

TABLE I. COMPARISON OF BIOMETRIC IDENTIFICATION PERFORMANCE (ACCURACY AND EQUAL ERROR RATE) FOR VARIOUS BENCHMARK MODELS AND THE PROPOSED METHOD ACROSS PALM+FV1 AND PALM+FV2 DATASETS

Dataset	Model	Accuracy	EER
PALM+FV1	FPV [77]	92.16	3.42
	LC-CNN [78]	93.29	4.25
	NLNet [79]	96.45	1.88
	Enhanced DenseNet [80]	95.37	3.63
	Att-CNN [79]	91.55	5.61
	FAB-AEF [81]	91.19	5.68
	Our model	98.73	0.89
PALM+FV2	FPV [77]	97.69	0.76
	LC-CNN [78]	96.41	2.80
	NLNet [79]	98.64	0.84
	Enhanced DenseNet [80]	98.86	0.68
	Att-CNN [79]	95.42	3.10
	FAB-AEF [81]	95.19	3.15
	Our model	99.49	0.35

Turning to model complexity in Table II, our proposed framework balances a relatively modest parameter count (10.6 million) and FLOPs (0.85 G) while sustaining a high inference speed of 60 FPS. Notably, this parameter count is lower than several other architectures, including Att-CNN and FAB-AEF, which exhibit substantially higher complexity yet lower throughput. The reduced FLOPs highlight the computational efficiency of combining Gabor filters with streamlined attention modules, avoiding the heavy overhead often associated with deep and wide CNN layers. Despite being lightweight, the network's ability to process 60 frames per second suggests that it can handle real-time applications, a crucial advantage in security and authentication scenarios where latency must be minimized. Although some alternative models, like NLNet and Enhanced DenseNet, also achieve respectable FLOPs and FPS, our design consistently offers strong trade-offs between speed, memory usage, and accuracy. The Gabor-based convolution units account for a significant portion of the efficiency gains by implicitly encoding orientation-selective features that reduce the burden on subsequent layers. Additionally, the attention mechanisms are carefully tailored to apply only where necessary, thus avoiding substantial computational overhead. Collectively, these attributes demonstrate that the proposed architecture maintains high performance levels while remaining both resource-friendly and well-suited for deployment in real-world multimodal biometric systems.

Despite the overall strong performance, several limitations warrant discussion. First, while our model effectively handles moderate noise and variations, performance degradation is observed under extreme lighting conditions, severe occlusions, or excessive image noise. Such cases, common in real-world biometric applications, pose challenges due to disrupted textural and directional feature integrity, which Gabor-based filters heavily rely on. Second, the two datasets used in this study differ slightly in terms of image resolution, acquisition devices, and quality variability. Although the model achieves strong cross-dataset results, these differences could potentially bias its generalization capability. Future work should evaluate GANet on more diverse datasets, particularly those exhibiting larger intra-class variations or cross-sensor discrepancies. These limitations highlight the need for robustness enhancements under edge conditions.

TABLE II. MODEL COMPLEXITY ANALYSIS COMPARING PARAMETER COUNT, FLOPS, AND INFERENCE SPEED (FPS) FOR THE PROPOSED METHOD AND BASELINE ARCHITECTURES

Model	Parameters	FLOPs	FPS
FPV [77]	21.6 M	2.0 G	18
LC-CNN [78]	18.0 M	0.94 G	42
NLNet [79]	15.4 M	0.8 G	67
Enhanced DenseNet [80]	6.8 M	0.43 G	54
Att-CNN [79]	26.4 M	12.6 G	15
FAB-AEF [81]	71.0 M	9.4 G	12
Our model	10.6 M	0.85 G	60

D. Ablation Study

In this section, we investigate the impact of each primary component within our framework through an ablation study. We remove or replace modules from the final model configuration, measuring the resultant performance on both the PALM+FV1 and PALM+FV2 multimodal datasets. Specifically, we analyze (i) the effect of omitting the Gabor layer for initial feature extraction, (ii) the influence of excluding the attention mechanism for feature enhancement (AMFE), and (iii) the contribution of discarding the attention mechanism for feature fusion (AMFF). A baseline CNN without any specialized Gabor or attention layers is also included for comparison. Table III summarizes these results, reporting accuracy and EER values over five experimental runs.

TABLE III. ABLATION STUDY RESULTS FOR PALM+FV1 AND PALM+FV2

Method	PALM+FV1 (Acc / EER)	PALM+FV2 (Acc / EER)
Baseline CNN	91.20 / 5.84	95.60 / 3.42
+ Gabor layer only	94.50 / 3.98	97.40 / 1.70
+ Gabor + AMFE	95.70 / 2.85	98.20 / 1.08
+ Gabor + AMFF	96.40 / 2.15	98.70 / 0.94
+ Gabor + AMFE + AMFF (Full model)	98.73 / 0.89	99.49 / 0.35

First, the baseline CNN configuration shows the weakest performance across both datasets. By replacing its standard first convolutional block with the Gabor layer, accuracy improves notably, indicating that Gabor-based kernels capture valuable orientation and texture details early in the pipeline. Comparing the baseline to the "+Gabor layer only" variant, we observe an improvement of over 3% in accuracy for PALM+FV1 and over 1% for PALM+FV2, confirming the ability of Gabor filters to learn discriminative edge and ridge patterns that help distinguish subtle biometric features. Next, reintroducing the attention mechanism for feature enhancement (AMFE) while excluding the attention mechanism for fusion yields a further performance boost, underscoring the importance of selectively emphasizing salient channels and spatial regions within each modality-specific branch. The model's accuracy surpasses that of the "Gabor layer only" variant by 1–2% on both datasets, and the EER decreases more significantly on PALM+FV1, suggesting that channel and spatial attention jointly reduce errors related to noisy or low-contrast local features.

Finally, adding back the attention mechanism for feature fusion (AMFF) to yield the complete model leads to the highest recognition accuracy and the lowest EER across PALM+FV1 and PALM+FV2. This outcome demonstrates that the global and local attentions within AMFF serve to reconcile the remaining discrepancies between modality-specific representations, maximizing the synergy of complementary palmprint and finger-vein features. Notably, accuracy climbs to 98.73% for PALM+FV1 and 99.49% for PALM+FV2, matching or exceeding state-of-the-art results in Table I and further highlighting that fully leveraging both inter- and intra-modal attention produces the most robust and reliable multimodal feature embeddings.

While the proposed framework shows excellent potential, its adaptability to other multimodal systems (e.g., face-iris, face-fingerprint) should be further validated. Comparative studies with alternative lightweight architectures (e.g., MobileNetV3, EfficientNet-Lite with customized fusion strategies) are planned for future work to explore broader applicability. Additionally, future improvements may include: (i) developing robust pre-processing techniques to mitigate extreme noise and occlusion; (ii) implementing dynamic parameter tuning for Gabor filters based on environmental conditions; and (iii) evaluating performance under cross-sensor and cross-session biometric settings.

V. CONCLUSION

In this paper, we introduced a lightweight, attention-driven framework for multimodal biometric identification that systematically integrates palmprint and finger-vein data. Our design leverages a specialized Gabor filter layer to capture orientation-specific and directional edge features, complementing conventional convolutional extraction techniques. Furthermore, we integrated two tailored attention modules: the Attention Mechanism for Feature Enhancement (AMFE) to emphasize salient channels and spatial regions within each modality, and the Attention Mechanism for Feature Fusion (AMFF) to harmonize intermodal inconsistencies for robust and efficient feature fusion. Comprehensive evaluations on two publicly available multimodal biometric datasets demonstrated that our method achieves state-of-the-art performance, yielding notable gains in both recognition accuracy and error rates compared to conventional CNN-based fusion strategies. Additionally, the framework maintains a modest computational footprint, making it practical for real-time or resource-constrained deployments.

While these results are promising, several limitations merit consideration. The model's performance can degrade under extreme lighting variations, severe occlusions, or high noise levels, which are commonly encountered in real-world biometric applications. Moreover, although the framework demonstrates strong performance on palmprint and finger-vein data, its adaptability to other biometric modalities (e.g., face, iris, gait) remains an open question requiring further validation. The potential trade-off between lightweight design and the richness of extracted features in highly complex or cross-domain biometric environments also warrants further exploration.

Looking forward, several research directions remain. Future work will explore the model's scalability to larger datasets and evaluate its resilience against adversarial attacks, a critical consideration for security-sensitive applications. Investigating the model's performance in cross-sensor, cross-session, and cross-spectral biometric settings is another important avenue. Furthermore, enhancing interpretability and developing privacy-preserving mechanisms for multimodal biometric systems will be prioritized to address emerging ethical and regulatory challenges.

Overall, our findings highlight the effectiveness of combining biologically inspired feature extraction with targeted attention mechanisms in multimodal biometric recognition. The proposed framework offers a promising step toward bridging the gap between high discriminative performance and computational efficiency, while also opening pathways for further innovation in robust, adaptable, and secure multimodal biometric systems.

REFERENCES

[1] R. K. Mahmood *et al.*, "Optimizing Network Security with Machine Learning and Multi-Factor Authentication for Enhanced Intrusion Detection," *Journal of Robotics and Control (JRC)*, vol. 5, no. 5, pp. 1502–1524, 2024, doi: 10.18196/jrc.v5i5.22508.

[2] O. N. Kadhim and M. H. Abdulameer, "A multimodal biometric database and case study for face recognition based deep learning," *Bulletin of Electrical Engineering and Informatics*, vol. 13, no. 1, 2024, doi: 10.11591/eei.v13i1.6605.

[3] M. M. M. Nawawi, K. A. Sidek, and A. W. Azman, "ECG biometric in real-life settings: analysing different physiological conditions with wearable smart textiles shirts," *Bulletin of Electrical Engineering and Informatics*, vol. 12, no. 5, 2023, doi: 10.11591/eei.v12i5.5133.

[4] P. Assiroj, H. L. H. S. Warnars, E. Abdurachman, A. I. Kistijantoro, and A. Doucet, "The influence of data size on a high-performance computing memetic algorithm in fingerprint dataset," *Bulletin of Electrical Engineering and Informatics*, vol. 10, no. 4, 2021, doi: 10.11591/EEI.V10I4.2760.

[5] A. M. Aljuboori and M. H. Abed, "Finger knuckle pattern person identification system based on LDP-NPE and machine learning methods," *Bulletin of Electrical Engineering and Informatics*, vol. 11, no. 6, 2022, doi: 10.11591/eei.v11i6.4236.

[6] M. T. S. Al-Kaltakchi, S. A. M. Al-Sumaidaee, and R. R. O. Al-Nima, "Classifications of signatures by radial basis neural network," *Bulletin*

of Electrical Engineering and Informatics, vol. 11, no. 6, 2022, doi: 10.11591/eei.v11i6.3931.

[7] I. B. Mohammed, B. S. Mahdi, and M. S. Kadhm, "Handwritten signature identification based on MobileNets model and support vector machine classifier," *Bulletin of Electrical Engineering and Informatics*, vol. 12, no. 4, 2023, doi: 10.11591/eei.v12i4.4965.

[8] M. H. Hamd and R. A. Rasool, "Optimized multimodal biometric system based fusion technique for human identification," *Bulletin of Electrical Engineering and Informatics*, vol. 9, no. 6, 2020, doi: 10.11591/eei.v9i6.2632.

[9] H. A. Ismael, J. M. Abbas, S. A. Mostafa, and A. H. Fadel, "An enhanced fireworks algorithm to generate prime key for multiple users in fingerprinting domain," *Bulletin of Electrical Engineering and Informatics*, vol. 10, no. 1, 2020, doi: 10.11591/eei.v10i1.2521.

[10] M. F. Jassim, W. M. S. Hamzah, and A. F. Shimal, "Biometric iris templates security based on secret image sharing and chaotic maps," *International Journal of Electrical and Computer Engineering*, vol. 12, no. 1, 2022, doi: 10.11591/ijece.v12i1.pp339-348.

[11] H. Al-Mahafzah, T. AbuKhalil, M. Alksasbeh, and B. Alqaralleh, "Multi-modal palm-print and hand-vein biometric recognition at sensor level fusion," *International Journal of Electrical and Computer Engineering*, vol. 13, no. 2, 2023, doi: 10.11591/ijece.v13i2.pp1954-1963.

[12] Y. I. Ibrahim and E. A. J. Sultan, "Iris recognition based on 2D Gabor filter," *International Journal of Electrical and Computer Engineering*, vol. 13, no. 1, 2023, doi: 10.11591/ijece.v13i1.pp325-334.

[13] I. Riaz, A. N. Ali, H. Ibrahim, and I. A. Huqqani, "Biometric classification system for dorsal finger creases utilizing multi-block circular shift combination local binary pattern," *International Journal of Electrical and Computer Engineering*, vol. 14, no. 5, pp. 5234–5243, Oct. 2024, doi: 10.11591/ijece.v14i5.pp5234-5243.

[14] M. Kusban, A. Budiman, and B. H. Purwoto, "Image enhancement in palmprint recognition: a novel approach for improved biometric authentication," *International Journal of Electrical and Computer Engineering*, vol. 14, no. 2, 2024, doi: 10.11591/ijece.v14i2.pp1299-1307.

[15] H. H. M. al Karaawi, M. Q. Dhahir, and I. A. Alamer, "Development modeling methods of analysis and synthesis of fingerprint deformations images," *International Journal of Electrical and Computer Engineering*, vol. 10, no. 6, 2020, doi: 10.11591/ijece.v10i6.pp6053-6060.

[16] J. S. Mane and S. Bhosale, "Advancements in biometric authentication systems: A comprehensive survey on internal traits, multimodal systems, and vein pattern biometrics," *Revue d'Intelligence Artificielle*, vol. 37, no. 3, 2023, doi: 10.18280/ria.370319.

[17] F. Balcı, "DM-EEGID: EEG-Based Biometric Authentication System Using Hybrid Attention-Based LSTM and MLP Algorithm," *Traitement du Signal*, vol. 40, no. 1, 2023, doi: 10.18280/ts.400106.

[18] S. Wang and R. Zhang, "Multi-Grained Deep Cascade Learning for ECG Biometric Recognition," *Traitement du Signal*, vol. 40, no. 2, 2023, doi: 10.18280/ts.400227.

[19] M. K. Singh, S. Kumar, and D. Nandan, "Biometric Face Identification: Utilizing Soft Computing Methods for Feature-Based Recognition," *Traitement du Signal*, vol. 41, no. 5, pp. 2721–2728, Oct. 2024, doi: 10.18280/ts.410545.

[20] F. Carrillo-Perez *et al.*, "Applications of artificial intelligence in dentistry: A comprehensive review," *Journal of Esthetic and Restorative Dentistry*, vol. 34, no. 1. 2022. doi: 10.1111/jerd.12844.

[21] A. S. Ahmed and A. M. Khaleel, "Enhancing Voice Authentication with a Hybrid Deep Learning and Active Learning Approach for Deepfake Detection," *Journal of Robotics and Control (JRC)*, vol. 5, no. 6, pp. 2002–2014, 2024, doi: 10.18196/jrc.v5i6.23502.

[22] A. T. Hermawan, I. A. E. Zaeni, A. P. Wibawa, Gunawan, W. H. Hendrawan, and Y. Kristian, "A Multi Representation Deep Learning Approach for Epileptic Seizure Detection," *Journal of Robotics and Control (JRC)*, vol. 5, no. 1, 2024, doi: 10.18196/jrc.v5i1.20870.

[23] L. Zholshiyeva, T. Zhukabayeva, D. Baumuratova, and A. Serek, "Design of QazSL Sign Language Recognition System for Physically Impaired Individuals," *Journal of Robotics and Control (JRC)*, vol. 6, no. 1, pp. 191–201, 2025, doi: 10.18196/jrc.v6i1.23879.

[24] H. F. Mahdi and B. J. Khadhim, "Enhancing IoT Security: A Deep Learning and Active Learning Approach to Intrusion Detection," *Journal of Robotics and Control (JRC)*, vol. 5, no. 5, pp. 1525–1535, 2024, doi: 10.18196/jrc.v5i5.22292.

[25] N. Tawfeeq and J. Harbi, "Advanced Ensemble Deep Learning Framework for Enhanced River Water Level Detection: Integrating Transfer Learning," *Journal of Robotics and Control (JRC)*, vol. 5, no. 5, pp. 1422–1435, 2024, doi: 10.18196/jrc.v5i5.22291.

[26] A. I. Khlaif, M. S. Naceur, and M. Kherallahr, "AI-Driven Classification of Children's Drawings for Pediatric Psychological Evaluation: An Ensemble Deep Learning Approach," *Journal of Robotics and Control (JRC)*, vol. 6, no. 1, pp. 124–141, 2025, doi: 10.18196/jrc.v6i1.23302.

[27] M. A. Shihab, H. A. Marhoon, S. R. Ahmed, B. Al-Attar, M. T. Al-Sharify, and R. Sekhar, "Towards Resilient Machine Learning Models: Addressing Adversarial Attacks in Wireless Sensor Network," *Journal of Robotics and Control (JRC)*, vol. 5, no. 5, pp. 1582–1602, Jan. 2024, doi: 10.18196/jrc.v5i5.23214.

[28] N. Sutarna, C. Tjahyadi, P. Oktivasari, M. Dwiyaniti, and Tohazen, "Hyperparameter Tuning Impact on Deep Learning Bi-LSTM for Photovoltaic Power Forecasting," *Journal of Robotics and Control (JRC)*, vol. 5, no. 3, pp. 677–693, 2024, doi: 10.18196/jrc.v5i3.21120.

[29] Y. Pamungkas, M. R. N. Ramadani, and E. N. Njoto, "Effectiveness of CNN Architectures and SMOTE to Overcome Imbalanced X-Ray Data in Childhood Pneumonia Detection," *Journal of Robotics and Control (JRC)*, vol. 5, no. 3, pp. 775–785, 2024, doi: 10.18196/jrc.v5i3.21494.

[30] P. Ganapathi, R. Arumugam, and S. Dhathathri, "An intelligent obfuscated mobile malware detection using deep supervised learning algorithms," *Bulletin of Electrical Engineering and Informatics*, vol. 13, no. 4, pp. 2604–2612, Aug. 2024, doi: 10.11591/eei.v13i4.6974.

[31] H. Nguyen, T. A. Nguyen, and N. D. Toan, "Optimizing feature extraction and fusion for high-resolution defect detection in solar cells," *Intelligent Systems with Applications*, vol. 24, Dec. 2024, doi: 10.1016/j.iswa.2024.200443.

[32] H. Nguyen, T. Q. Ngo, H. T. T. Uyen, and M. K. Duong, "Enhanced object recognition from remote sensing images based on hybrid convolution and transformer structure," *Earth Science Informatics*, vol. 18, no. 2, Feb. 2025, doi: 10.1007/s12145-025-01751-x.

[33] N. D. Toan, L. H. Le, and H. Nguyen, "Adaptive Compression Techniques for Lightweight Object Detection in Edge Devices," *Mathematical Modelling of Engineering Problems*, vol. 11, no. 11, pp. 3071–3081, Nov. 2024, doi: 10.18280/mmep.111119.

[34] R. Jain, P. Singh, and A. Kaur, "An ensemble reinforcement learning-assisted deep learning framework for enhanced lung cancer diagnosis," *Swarm and Evolutionary Computation*, vol. 91, p. 101767, Dec. 2024, doi: 10.1016/J.SWEVO.2024.101767.

[35] T. Yan, J. Rashid, M. S. Saleem, S. Ahmad, and M. Faheem, "A Hybrid Deep Learning Approach for Green Energy Forecasting in Asian Countries," *Computers, Materials and Continua*, vol. 81, no. 2, pp. 2685–2708, Nov. 2024, doi: 10.32604/CMC.2024.058186.

[36] X. Zhang and T. Sun, "Deep reinforcement learning for collision avoidance of autonomous ships in inland rivers," *Proceedings of the Institution of Civil Engineers - Transport*, Dec. 2024, doi: 10.1680/JTRAN.24.00068.

[37] G. Erdős and Z. Dosztányi, "Deep learning for intrinsically disordered proteins: From improved predictions to deciphering conformational ensembles," *Current Opinion in Structural Biology*, vol. 89, p. 102950, Dec. 2024, doi: 10.1016/J.SBI.2024.102950.

[38] M. Zech, H. P. Tetens, and J. Ranalli, "Toward global rooftop PV detection with Deep Active Learning," *Advances in Applied Energy*, vol. 16, p. 100191, Dec. 2024, doi: 10.1016/J.ADAPEN.2024.100191.

[39] N. Khan, K. Kulkarni, Y. Mahale, S. Kolhar, and S. Mahajan, "Waste Objects Segregation Using Deep Reinforcement Learning with Deep Q Networks," *Ingenierie des Systemes d'Information*, vol. 29, no. 6, pp. 2219–2229, Dec. 2024, doi: 10.18280/isi.290612.

[40] M. Hasanat, W. Khan, N. Minallah, N. Aziz, and A. U. R. Durrani, "Performance evaluation of transfer learning based deep convolutional neural network with limited fused spectro-temporal data for land cover classification," *International Journal of Electrical and Computer Engineering*, vol. 13, no. 6, 2023, doi: 10.11591/ijece.v13i6.pp6882-6890.

[41] N. Alay and H. H. Al-Baity, "Deep learning approach for multimodal biometric recognition system based on fusion of iris, face, and finger

vein traits," *Sensors (Switzerland)*, vol. 20, no. 19, 2020, doi: 10.3390/s20195523.

[42] X. Zhang, D. Cheng, P. Jia, Y. Dai, and X. Xu, "An Efficient Android-Based Multimodal Biometric Authentication System with Face and Voice," *IEEE Access*, vol. 8, 2020, doi: 10.1109/ACCESS.2020.2999115.

[43] M. Abdul-Al, G. K. Kyeremeh, N. O. Parchin, R. A. Abd-Alhameed, R. Qahwaji, and J. Rodriguez, "Performance of Multimodal Biometric Systems Using Face and Fingerprints (Short Survey)," in *IEEE International Workshop on Computer Aided Modeling and Design of Communication Links and Networks, CAMAD*, 2021. doi: 10.1109/CAMAD52502.2021.9617766.

[44] R. M. Jomaa, M. S. Islam, and H. Mathkour, "Improved sequential fusion of heart-signal and fingerprint for anti-spoofing," in *2018 IEEE 4th International Conference on Identity, Security, and Behavior Analysis, ISBA 2018*, 2018. doi: 10.1109/ISBA.2018.8311476.

[45] X. Xu and Z. Mu, "Feature fusion method based on KCCA for ear and profile face based multimodal recognition," in *Proceedings of the IEEE International Conference on Automation and Logistics, ICAL 2007*, 2007. doi: 10.1109/ICAL.2007.4338638.

[46] S. S. Sengar, U. Hariharan, and K. Rajkumar, "Multimodal Biometric Authentication System using Deep Learning Method," in *2020 International Conference on Emerging Smart Computing and Informatics, ESCI 2020*, 2020. doi: 10.1109/ESCI48226.2020.9167512.

[47] Q. Yang, X. Chen, Z. He, and L. Chang, "Survey on Deep Learning Based Fusion Recognition of Multimodal Biometrics," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2022. doi: 10.1007/978-3-031-20233-9_52.

[48] C. Medjahed, A. Rahmoun, C. Charrier, and F. Mezzoudj, "A deep learning-based multimodal biometric system using score fusion," *IAES International Journal of Artificial Intelligence*, vol. 11, no. 1, 2022, doi: 10.11591/ijai.v11.i1.pp65-80.

[49] S. Vatchala *et al.*, "Multi-modal biometric authentication: Leveraging shared layer architectures for enhanced security," *IEEE Access*, 2025, doi: 10.1109/ACCESS.2025.3534223.

[50] B. Alharbi and H. S. Alshanbari, "Face-voice based multimodal biometric authentication system via FaceNet and GMM," *PeerJ Computer Science*, vol. 9, 2023, doi: 10.7717/peerj-cs.1468.

[51] R. Srivastava, "Score-Level Multimodal Biometric Authentication of Humans Using Retina, Fingerprint, and Fingervein," *International Journal of Applied Evolutionary Computation*, vol. 11, no. 3, 2020, doi: 10.4018/ijaec.2020070102.

[52] H. Abderrahmane, G. Noubeil, Z. Lahcene, Z. Akhtar, and D. Dasgupta, "Weighted quasi-arithmetic mean based score level fusion for multi-biometric systems," *IET Biometrics*, vol. 9, no. 3, 2020, doi: 10.1049/iet-bmt.2018.5265.

[53] Sonal, A. Singh, and C. Kant, "Optimized hybrid SVM-RF multi-biometric framework for enhanced authentication using fingerprint, iris, and face recognition," *PeerJ Computer Science*, vol. 11, p. e2699, Feb. 2025, doi: 10.7717/peerj-cs.2699.

[54] M. Garg, A. S. Arora, and S. Gupta, "A novel feature biometric fusion approach for iris, speech and signature," *Computer Methods in Material Science*, vol. 20, no. 2, 2020, doi: 10.7494/cmms.2020.2.0655.

[55] D. Maiti and M. Basak, "Multimodal biometric integration: Trends and insights from the past quinquennial," *World Journal of Advanced Research and Reviews*, vol. 23, no. 3, pp. 1590–1605, Sep. 2024, doi: 10.30574/wjarr.2024.23.3.2741.

[56] H. H. Rasheed, S. S. Shamini, M. A. Mahmoud, and M. A. Alomari, "Review of iris segmentation and recognition using deep learning to improve biometric application," *Journal of Intelligent Systems*, vol. 32, no. 1, 2023, doi: 10.1515/jisys-2023-0139.

[57] A. K. Sharma, S. Bhattacharya, and M. Reza, "Dual Channel Multi-Attention in ViT for Biometric Authentication using Forehead Subcutaneous Vein Pattern and Periocular Pattern," *arXiv preprint arXiv:2412.19160*, 2024.

[58] Liujun, C. Mingjin, Guoyong, and H. Zengxi, "Multimodal Biometric Recognition Neural Networks with Branch Attention Mechanisms," in *Proceedings - 2023 Asia Conference on Advanced Robotics, Automation, and Control Engineering, ARACE 2023*, 2023, doi: 10.1109/ARACE60380.2023.00021.

[59] N. Ammour, Y. Bazi, and N. Alajlan, "Multimodal Approach for Enhancing Biometric Authentication," *Journal of Imaging*, vol. 9, no. 9, 2023, doi: 10.3390/jimaging9090168.

[60] L. Lin, Y. Zhao, J. Meng, and Q. Zhao, "A Federated Attention-Based Multimodal Biometric Recognition Approach in IoT," *Sensors*, vol. 23, no. 13, 2023, doi: 10.3390/s23136006.

[61] R. Yang, Q. Zhang, and L. Meng, "AuthFormer: Adaptive Multimodal biometric authentication transformer for middle-aged and elderly people," *arXiv preprint arXiv: 2411.05395*, 2024.

[62] V. Talreja, M. C. Valenti, and N. M. Nasrabadi, "Deep Hashing for Secure Multimodal Biometrics," *IEEE Transactions on Information Forensics and Security*, vol. 16, 2021, doi: 10.1109/TIFS.2020.3033189.

[63] M. H. Safavipour, M. A. Doostari, and H. Sadjedi, "Deep Hybrid Multimodal Biometric Recognition System Based on Features-Level Deep Fusion of Five Biometric Traits," *Computational Intelligence and Neuroscience*, vol. 2023, no. 1, 2023, doi: 10.1155/2023/6443786.

[64] M. L. Gavrilova, "Information Fusion: A Decade of Innovations in Biometric Multimodal Research," *Computer*, vol. 58, no. 4, pp. 27–36, Apr. 2025, doi: 10.1109/MC.2025.3526135.

[65] Z. Wahid, A. H. Bari, F. Anzum, and M. L. Gavrilova, "Human Micro-Expression: A Novel Social Behavioral Biometric for Person Identification," *IEEE Access*, vol. 11, 2023, doi: 10.1109/ACCESS.2023.3283932.

[66] Z. Guo, H. Ma, and A. Li, "A lightweight finger multimodal recognition model based on detail optimization and perceptual compensation embedding," *Computer Standards & Interfaces*, vol. 92, p. 103937, Mar. 2025, doi: 10.1016/J.CSI.2024.103937.

[67] M. G. S. Murshed, S. K. Abbas, S. Purnapatra, D. Hou, and F. Hussain, "Deep Learning-Based Approaches for Contactless Fingerprints Segmentation and Extraction," *arXiv preprint arXiv:2311.15163*, 2023.

[68] R. Ryu, S. Yeom, S. H. Kim, and D. Herbert, "Continuous Multimodal Biometric Authentication Schemes: A Systematic Review," *IEEE Access*, vol. 9. 2021. doi: 10.1109/ACCESS.2021.3061589.

[69] Z. Y. Sheng, Y. Ai, Y. N. Chen, and Z. H. Ling, "Face-Driven Zero-Shot Voice Conversion with Memory-based Face-Voice Alignment," in *MM 2023 - Proceedings of the 31st ACM International Conference on Multimedia*, 2023. doi: 10.1145/3581783.3613825.

[70] G. Chen, D. Luo, F. Lian, F. Tian, X. Yang, and W. Kang, "A Multimodal Biometric Recognition Method Based on Federated Learning," *IET Biometrics*, vol. 2024, no. 1, 2024, doi: 10.1049/2024/5873909.

[71] P. Kaplesh, A. Gupta, D. Bansal, S. Sofat, and A. Mittal, "A systematic review of end-to-end framework for contactless fingerprint recognition: Techniques, challenges, and future directions," *Engineering Applications of Artificial Intelligence*, vol. 150, p. 110493, 2025.

[72] R. Halabi *et al.*, "Comparative Assessment of Multimodal Sensor Data Quality Collected Using Android and iOS Smartphones in Real-World Settings," *Sensors*, vol. 24, no. 19, p. 6246, Sep. 2024, doi: 10.3390/s24196246.

[73] D. Yuvasini, S. Jegadeesan, S. Selvarajan, and F. A. Mon, "Enhancing societal security: a multimodal deep learning approach for a public person identification and tracking system," *Scientific reports*, vol. 14, no. 1, p. 23952, Dec. 2024, doi: 10.1038/s41598-024-74560-9.

[74] L. Zhang, L. Li, A. Yang, Y. Shen, and M. Yang, "Towards contactless palmprint recognition: A novel device, a new benchmark, and a collaborative representation based identification approach," *Pattern Recognition*, vol. 69, 2017, doi: 10.1016/j.patcog.2017.04.016.

[75] A. Kumar and Y. Zhou, "Human identification using finger images," *IEEE Transactions on Image Processing*, vol. 21, no. 4, 2012, doi: 10.1109/TIP.2011.2171697.

[76] M. S. Mohd Asaari, S. A. Suandi, and B. A. Rosdi, "Fusion of Band Limited Phase only Correlation and Width Centroid Contour Distance for finger based biometrics," *Expert Systems with Applications*, vol. 41, no. 7, 2014, doi: 10.1016/j.eswa.2013.11.033.

[77] H. Ren, L. Sun, J. Guo, and C. Han, "A Dataset and Benchmark for Multimodal Biometric Recognition Based on Fingerprint and Finger

Vein," *IEEE Transactions on Information Forensics and Security*, vol. 17, 2022, doi: 10.1109/TIFS.2022.3175599.

[78] S. Li, B. Zhang, S. Zhao, and J. Yang, "Local discriminant coding based convolutional feature representation for multimodal finger recognition," *Information Sciences*, vol. 547, 2021, doi: 10.1016/j.ins.2020.09.045.

[79] Z. Guo, H. Ma, and J. Liu, "NLNet: A narrow-channel lightweight network for finger multimodal recognition," *Digital Signal Processing*, vol. 150, p. 104517, Jul. 2024, doi: 10.1016/J.DSP.2024.104517.

[80] W. Wu, Y. Zhang, Y. Li, C. Li, and Y. Hao, "A Hand Features Based Fusion Recognition Network with Enhancing Multi-Modal Correlation," *CMES - Computer Modeling in Engineering and Sciences*, vol. 140, no. 1, pp. 537–555, Apr. 2024, doi: 10.32604/CMES.2024.049174.

[81] Y. Huang, H. Ma, and M. Wang, "Multimodal Finger Recognition Based on Asymmetric Networks With Fused Similarity," *IEEE Access*, vol. 11, 2023, doi: 10.1109/ACCESS.2023.3242984.