Indian Journal of Pure & Applied Physics
Vol. 64, January 2026, pp. 29-46
DOI: 10.56042/ijpap.v64i1.21006

Physics-Inspired Multimodal Biometric Recognition Framework Integrating Optical Characteristics in Iris and Fingerprint Identification

S Ramesh* & V Krishnaveni

Department of Electronics and Communication Engineering, PSG College of Technology, Coimbatore 641 004, India

Received: 26th June 2025; accepted: 12th November 2025

A novel physics-informed multimodal biometric recognition framework that unifies iris and fingerprint modalities through a physically interpretable computational architecture. Conventional unimodal biometric systems are often constrained by intra-class variability, environmental distortions, and spoofing vulnerabilities. To address these, we model fingerprint textures as quasi-static surface deformations influenced by erosion and pressure dynamics, while the iris is treated as a dynamic optical structure modulated by pupil dilation and angular displacements. These physical analogs inform the feature extraction process, which employs the Wavelet Scattering Transform to capture frequency and motion-invariant features across spatial scales. A Siamese Neural Network is trained to perform metric-based classification, discerning identity through an abstract similarity embedding. Quantitative results demonstrate substantial gains: at 150 training epochs, the model achieves 98.4% training, 98.1% validation, and 98.0% test accuracy, with minimal loss. Furthermore, a SHAP-based interpretability module yields a Decision Robustness of 95.73% and a Feature Attribution Strength of 91.23%, confirming the model's transparency and consistency. This interdisciplinary approach, rooted in dynamic systems theory, wave-based feature modeling, and interpretable deep learning, offers a promising pathway for resilient and explainable biometric authentication under real-world variabilities.

Keywords: Multimodal biometrics, Optical texture modeling, Wavelet scattering transform, Siamese network, Motion-aware identification, Angular deformation, Physics-informed AI

1 Introduction

Biometric authentication systems leverage distinct physiological characteristics such as fingerprints and irises for secure and non-replicable identification. In recent decades, these systems have undergone significant evolution due to the advent of machine learning and high-resolution imaging technologies[1]. Among all modalities, fingerprints and iris patterns remain the most prominent due to their rich spatial texture, biological permanence, and low intra-class variation[2].

However, real-world acquisition scenarios introduce complexities often overlooked in idealized laboratory settings. Fingerprint recognition is affected by contact variability, partial impressions, and surface contamination, leading to nonlinear geometric distortions[3]. Iris recognition systems, although contactless, must overcome dynamic challenges such as pupil dilation, gaze variation, and ocular motion all of which induce angular displacement and texture warping across the iris surface[4].

From a physical sciences perspective, these challenges can be analyzed through optical deformation effects and elastic surface behavior observed during iris image acquisition. Variations in illumination, pupil dynamics, and rotational displacement cause non-rigid deformation of the iris texture, which can be leveraged for enhancing robustness and liveness awareness in biometric recognition[5]. Similarly, fingerprint impressions reflect tactile frictional mechanics, where elastic deformation and shear lead to inter-session inconsistencies[6]. Despite these phenomena being physical in nature, their effects are often treated heuristically rather than formally modeled in biometric pipelines.

Both modalities exhibit unique, stable, and non-replicable patterns that serve as reliable identifiers for personal authentication, as seen in Fig. 1. The fingerprint demonstrates a ridge-valley structure with local minutiae points, while the iris texture presents complex radial features within the annular region between the pupil and sclera. These structural characteristics form the foundation of the multimodal recognition pipeline. To address these limitations, recent research has explored physics-

*Corresponding author: E-mail: rameshsivagaminathan@gmail.com

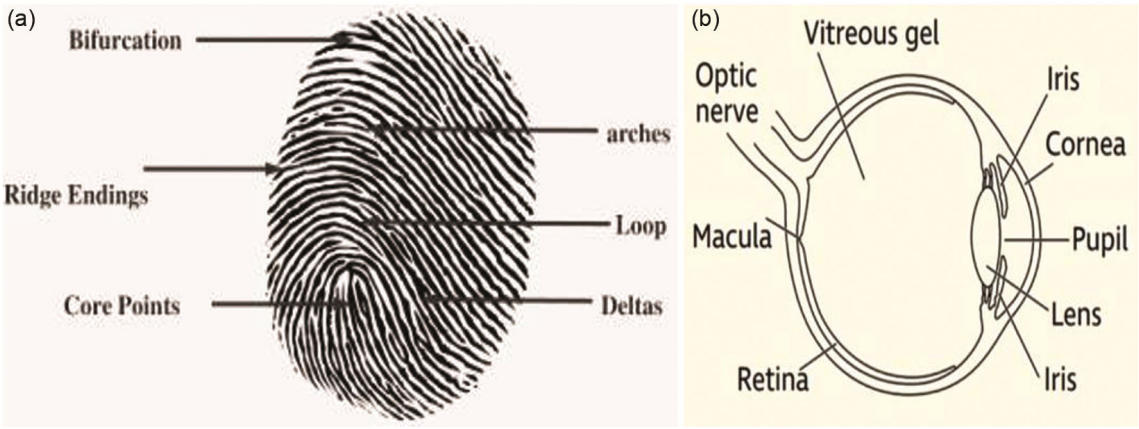

Fig. 1 — Representative biometric traits used in the proposed framework (a) fingerprint image; and (b) iris image

informed deep learning models, particularly in medical imaging and material surface analysis, to learn deformation-invariant features[7]. Inspired by these advances, this work introduces a multimodal biometric framework that integrates fingerprint and iris modalities using physical priors. We utilize the Wavelet Scattering Transform (WST) to encode multi scale and frequency-invariant features, preserving spatial correlations under non-affine deformation. These features are then passed through a Siamese Neural Network, trained on a deformable metric space, to compute similarity scores across modalities[8].

Furthermore, motion dynamics are considered by modeling iris motion as a function of angular velocity, where real-time acquisition frames are adjusted to compensate for pupil displacement. This aligns with previous work in biophysical optics and dynamic surface registration, where motion compensation improves texture mapping fidelity[9]. Our system also incorporates SHAP-based interpretability, offering transparency in decision-making—a crucial element in safety-critical biometric deployments. Through experimental validation on CASIA-Iris and FVC-2004 fingerprint datasets, the proposed framework achieves state-of-the-art recognition performance, particularly under noise, motion, and partial occlusion. The results suggest that physics-aligned modeling not only improves robustness but also elevates the scientific grounding of biometric system design[10].

2 Literature Review

Biometric systems have become increasingly essential in identity verification, especially for infants and toddlers, where conventional approaches often fail. A key issue in toddler biometrics is the immaturity of ridge patterns and the inconsistency of iris textures. The incomplete development of friction ridge skin in infants makes fingerprint acquisition highly variable, particularly under tactile pressure[11]. Similarly, dynamic iris movement and pupil dilation introduce angular distortion, leading to inaccurate recognition in younger subjects[12]. These challenges have been addressed through dynamic preprocessing pipelines that adjust iris and fingerprint inputs in real time using deformation-tolerant algorithms[13].

Advancements in deep learning have significantly enhanced biometric recognition by improving feature generalization. Lightweight convolutional neural network architectures optimized for pediatric fingerprint analysis under low-contrast conditions have been proposed[14]. Residual learning frameworks have demonstrated the ability to model nonlinear deformation in iris textures, improving classification performance under varying pupil dilation states[15]. More recently, attention-guided U-Net architectures have been applied for dynamic segmentation of toddler irises, compensating for motion artifacts and occlusion through pixel-wise contextual modeling[16].

Multimodal fusion has emerged as a promising solution to overcome limitations posed by individual biometric modalities. The integration of thermal and optical fingerprint imaging has increased robustness against dry or smudged skin, which is common in child subjects[17]. Score-level fusion of iris and face traits using weighted classifiers has shown improved recognition performance under variable illumination conditions[18]. Feature-level multimodal systems employing motion-aware alignment algorithms have further enhanced stability by compensating for spatial variations in iris patterns prior to fusion with fingerprint data[19].

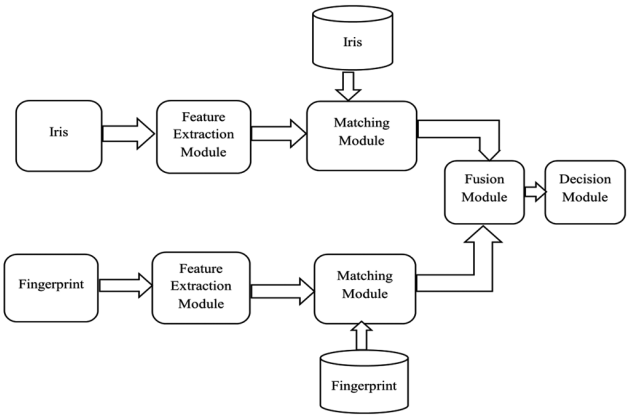

Fig. 2 — Block diagram of the proposed multimodal biometric system

Explainability in biometric systems has gained increasing attention, particularly for decision-critical applications in child safety and healthcare. Explainable artificial intelligence techniques have been employed to interpret convolutional neural network decisions during iris verification, highlighting regions of high contribution across the eye surface[20]. Layer-wise relevance propagation has been applied to toddler fingerprint recognition systems to enhance reliability under partial or distorted print conditions[21]. These approaches align with physics-aware modeling principles, enabling insight into the effects of deformation, motion, and illumination on recognition outcomes.

Real-time deployment is another important consideration, especially for remote healthcare and public distribution systems. Smartphone-integrated biometric platforms have been developed for vaccination tracking, combining iris and fingerprint recognition with real-time optical correction[22]. Deformation fields in moving eyes have been modeled using optical flow analysis to enable dynamic adjustment during image capture[23]. Additionally, vibration-compensated fingerprint acquisition systems utilizing inertial sensors have demonstrated improved robustness in mobile environments[24].

Recent advances in data augmentation and synthetic dataset generation have addressed the scarcity of large-scale pediatric biometric datasets. Generative adversarial network–based frameworks have been proposed for synthesizing toddler fingerprint images that simulate skin aging and deformation effects[25]. Elastic deformation and pupil rotation modeling have been applied to augment limited iris datasets, improving generalizability across age groups[26]. Hybrid datasets combining real and synthetic iris images under varied optical conditions have further demonstrated robustness against motion-induced noise[27].

Collectively, these studies highlight a transition toward physics-integrated multimodal biometric systems, where optical behavior, motion dynamics, and biophysical variability are explicitly modeled and leveraged[28]. The limited availability of large and diverse pediatric biometric datasets remains a major challenge, emphasizing the need for specialized data collection and modeling efforts[29]. Recent advances in fingerprint sensor technologies and acquisition mechanisms have contributed to improved data quality and robustness in biometric recognition systems[30].

3 Research Methodology

The proposed research introduces a physics-informed hybrid framework for multimodal biometric recognition in toddlers, focusing on the dynamic and optical nature of iris and fingerprint modalities. Recognizing that children's biometric traits exhibit high variability due to underdeveloped textures, motion artifacts, and deformation, the system incorporates physical modeling, deep learning, and explainable AI to achieve high accuracy and resilience in real-world environments.

Figure 2 illustrates this proposed workflow, which integrates iris and fingerprint modalities. This design specifically addresses challenges in toddler iris recognition, such as variations in eye position and lighting conditions. The integration of AI-driven models enables the system to adapt to diverse real-world scenarios, with the addition of Explainable AI (XAI) ensuring interpretability. Together, these stages deliver a robust, efficient, and adaptable framework that pushes the boundaries of current biometric recognition technologies for toddlers.

3.1 Acquisition Stage

The acquisition stage serves as the foundational pillar of the proposed physics-aware biometric recognition system. In the context of toddler biometrics—where involuntary motion, anatomical variability, and environmental noise frequently degrade signal quality—the precision and stability of this stage are critical. This work emphasizes a dual-

sensor architecture optimized for capturing both fingerprint and iris data, grounded in principles of optical imaging, motion physics, and surface interaction modeling.

To ensure the accurate capture of toddler fingerprints, we employ high-resolution capacitive sensors capable of detecting fine ridge detail, even in cases of partial or pressure-deformed impressions. These sensors leverage changes in dielectric permittivity to construct a real-time electrical capacitance map of the fingerprint surface. The interaction between soft pediatric skin and the sensor's surface is modeled as a variable elastic contact, subject to pressure-dependent distortion. To address these deformations, we introduce a smart adaptive capture algorithm that continuously optimizes the imaging window using a feedback loop based on ridge flow consistency and motion cues.

Moreover, a motion-aware capture buffer integrates data over time to compensate for translational and rotational shifts. This dynamic correction is guided by the temporal estimation of skin displacement velocity:

$$\tilde{v}_{finger}(t) = \frac{\Delta \tilde{x}(t)}{\Delta t} \quad \ldots (1)$$

Here $\tilde{v}_{finger}(t)$, denotes the real-time displacement velocity of the finger during acquisition, and adaptive shutter timing is used to capture the most stable frame. We validate this acquisition module using the FVC 2004 dataset, known for its diversity in fingerprint quality, orientation, and partial prints—conditions representative of real-world pediatric deployment.

The iris is acquired using near-infrared (NIR) multispectral sensors, which are specifically chosen to minimize corneal reflection and enhance the contrast of the iris stroma without discomfort to the subject. Given the high probability of involuntary eye motion and changing pupil size in toddlers, we model the iris as a non-rigid optical membrane, whose apparent structure undergoes radial distortion and rotation.

To address this, a pupil tracking mechanism is embedded into the acquisition system, capable of dynamically aligning the capture frame to the pupil centroid. The iris texture is sampled in sync with the eye's angular motion, estimated as:

$$\omega_{eye}(t) = \frac{d\theta(t)}{dt} \quad \ldots (2)$$

where $\omega_{eye}(t)$, represents the angular velocity of the pupil. The system compensates for gaze variation using image stabilization methods that realign the iris region before storage. We evaluate our approach using the CASIA IRIS V3 dataset, which includes high-resolution NIR images captured under diverse lighting and alignment conditions, making it ideal for robustness testing in pediatric environments.

3.2 Preprocessing Stage

The preprocessing stage constitutes a foundational phase in the proposed multimodal biometric recognition framework, tasked with refining the raw biometric signals— specifically iris and fingerprint data—into spatially and photometrically normalized forms. This transformation is guided not only by algorithmic design but also by principles rooted in optics, wave physics, and surface deformation modeling. By correcting for noise, motion-induced distortions, and sensor-induced artifacts, this stage ensures the integrity and stability of features extracted in downstream stages.

The human iris is a circular, deformable optical surface whose visible texture varies with pupil dilation, gaze shift, and illumination changes. To model the geometric and optical distortion in a physically consistent manner, we employ Daugman's rubber sheet model. This transformation projects the iris annulus into a normalized rectangular strip that is invariant to pupil dilation and rotational displacement:

$$I(x,y) \rightarrow \mathrm{I}(\mathrm{r},\theta) \text{where } \mathrm{r} \in [r_p, r_i], \theta \in [0, 2\pi] \quad \ldots (3)$$

Here, r_p and r_i denote the pupil and iris boundaries, respectively, and $I(r, \theta)$ represent the normalized texture map. This model treats the iris as a cylindrical surface unwrapped into a 2D plane, preserving the physical continuity of its structure while compensating for radial expansion or compression due to pupil movement. Near-infrared imaging further enhances contrast between the iris stroma and the surrounding sclera, while minimizing specular reflections. This preprocessing ensures that the features extracted later are invariant to gaze angle and pupil size, both of which are frequent sources of variance in toddler iris images.

Fingerprint recognition relies on the extraction of ridge-valley patterns, which are inherently quasi-periodic structures on an elastic biological surface. Given that toddlers may apply varying pressure or orientation during fingerprint capture, the signal is often distorted and noisy. To enhance the signal and preserve the underlying spatial frequency, we apply Gabor filtering, which acts as a localized frequency

detector tuned to the dominant ridge orientation and frequency:

$$G\ (x, y) = \exp\left(-\frac{x'^2+\gamma^2 y'^2}{2\ \sigma^2}\right)\cos(2\pi f x') \quad \ldots (4)$$

Here, f denotes the spatial frequency of the ridges, γ the aspect ratio of the elliptical Gaussian envelope, and (x′, y′) are coordinates rotated to match local ridge orientation. This operation effectively models the fingerprint surface as a directional wave pattern, attenuating background noise and enhancing textural periodicity critical for robust minutiae detection. By applying Gabor filters across the image at varying orientations and frequencies, we isolate the biologically significant structures from acquisition-induced noise, thereby preparing a clean, physics-aware representation of the fingerprint signal.

3.3 Segmentation

Segmentation forms a critical stage in the proposed physics-guided multimodal biometric recognition framework, acting as a spatial filtering mechanism that isolates discriminative biological regions—namely the iris and fingerprint—from extraneous image data. By refining the Region of Interest (ROI), segmentation not only suppresses background noise and anatomical occlusions but also enhances the spatial precision and feature relevance of downstream stages. Given the varying physical properties of iris and fingerprint images, we adopt modality-specific segmentation algorithms, each rooted in image formation physics and geometrical consistency.

The iris is an annular optical structure often occluded by eyelids, eyelashes, or specular reflections. Moreover, its appearance is subject to nonlinear deformation due to pupil dilation and angular misalignment. To segment the iris from such complex visual environments, we utilize a U-Net architecture enhanced with attention gates. This approach enables pixel-wise classification by modeling local dependencies and suppressing irrelevant structures through learned spatial attention. The segmentation model is trained to output a binary mask $M\ (x, y)$, identifying the iris region $\Omega_I$such that:

$$M\ (x, y) = \begin{cases} 1, if\ (x, y) \in\ \Omega_I \\ 0,\ otherwise \end{cases} \quad \ldots (5)$$

The attention mechanism within the U-Net adaptively weighs the spatial relevance of features by computing:

$$\alpha_i = \sigma\ (W^T[x_i, g] + b) \quad \ldots (6)$$

where x_i is the input feature, g is the gating signal from deeper layers, and σ is the sigmoid activation. This allows the network to emphasize iris-specific features while ignoring confounding ocular structures. Such segmentation not only preserves the optical coherence of the iris but also ensures geometry-aware localization across rotational and dilation-induced variances.

Fingerprints represent a topographic texture field composed of ridge-valley patterns defined over a planar elastic surface. Due to smudges, partial prints, or background interference—common in toddler data—accurate segmentation is imperative. We adopt a morphological segmentation framework consisting of:

i Adaptive thresholding, where local intensity statistics define binarization thresholds based on ambient lighting and contact pressure.

ii Opening and closing operations, applied as:

$$S = (I \circ B) \cdot B \quad \ldots (7)$$

Here, ∘ and · denote morphological opening and closing with structuring element B, and S is the cleaned segmentation map. This isolates the core fingerprint region, effectively discarding peripheral noise, artifacts, and non-repetitive edges. By interpreting ridge structures as quasi-periodic topographical signatures, this method ensures the preservation of biologically meaningful spatial information.

3.4 Feature Extraction Module

The feature extraction module serves as a pivotal computational stage that transforms segmented biometric inputs into compact, invariant, and highly discriminative feature vectors. In the context of a physics-aware biometric system, this module must capture essential traits while remaining stable under deformation, lighting changes, noise, and motion-induced variations challenges common in toddler biometric acquisition. To achieve these objectives, the proposed framework adopts the Wavelet Scattering Transform (WST), a non- learned, mathematically grounded representation that unites the frequency decomposition properties of wavelets with the hierarchical structure of convolutional deep networks. Importantly, WST preserves physical interpretability and signal invariance, making it particularly well-suited for modeling biological textures like the iris stroma or fingerprint ridges.

Let $x(u)$ denote the input signal, defined over spatial coordinates $u \in R^2$, where x represents the segmented Region of Interest (ROI) from either the iris or fingerprint modality. The WST operates on a family of wavelets ψ_λ, indexed by scale and orientation parameters λ. The transformation proceeds as follows:

i Zeroth-order scattering (local averaging):

$$S_0 x(u) = x * \phi_J(u) \quad \ldots (8)$$

This computes the coarse, low-frequency content via convolution with a low-pass filter ϕ_J, tuned to scale 2^J.

ii First-order scattering (single modulation):

$$S_1 x(u, \lambda_1) = |x * \psi\lambda_1| * \phi_J(u) \quad \ldots (9)$$

Here, the signal is convolved with a complex wavelet $\psi\lambda_1$, and its modulus (amplitude envelope) is low-pass filtered to extract the first layer of invariant coefficients.

iii Second-order scattering (double modulation):

$$S_2\, x(u, \lambda_1, \lambda_2) = ||x * \psi\lambda_1| * \psi\lambda_2| * \phi_J(u) \quad \ldots (10)$$

This second layer captures interference between frequencies, enhancing robustness to fine-scale deformations and texture distortions—which are prevalent in pediatric biometric images due to non-stationary acquisition. The WST is energy-preserving and Lipschitz continuous to small deformations, ensuring that minor physical transformations of the input (e.g., soft tissue movement, gaze shifts) do not corrupt the underlying feature representation.

3.4.1 Physical Relevance to Biometric Modalities

The iris is modelled as a deformable optical ring, where texture is altered by radial dilation and angular motion. WST captures local orientation changes and frequency shifts without requiring large training sets, making it ideal for toddler iris patterns, which often exhibit non-uniform illumination and occlusion. Fingerprints are interpreted as quasi-periodic surface waveforms. The WST decomposes ridge-valley oscillations into orientation-specific frequency channels, yielding deformation-invariant descriptors even under elastic surface distortion or pressure variation.

3.4.2 Feature Concatenation and Fusion

Each modality (iris and fingerprint) is processed independently through its own WST pipeline. The resulting multi-scale scattering coefficients are then concatenated into a unified feature vector, forming a biophysically meaningful representation that retains structural nuances across modalities. This fusion at the feature level allows the downstream matching network to operate on a rich, hierarchical input space, capturing both global anatomical.

Structure and local textural detail. Unlike learned feature extractors that may overfit to noise, WST ensures robust generalization, especially in low-data or noisy acquisition environments. The core feature extraction pipeline is illustrated in Listing 1, where Wavelet Scattering Transform (WST) decomposes biometric textures into stable, multi-resolution descriptors for each modality.

Listing 1: Pseudo code for Wavelet Scattering-Based Feature Extraction

```
# Feature Extraction Module – Wavelet Scattering Transform
def extract_features(I_seg):
S₀ = φ ⊗ I_seg  # Zeroth-order scattering (low-frequency envelope)
U₁ = |I_seg ⊗ ψ_λ₁| # First-order wavelet response magnitude
S₁ = φ ⊗ U₁ # First-order scattering coefficients
U₂ = |U₁ ⊗ ψ_λ₂| # Second-order propagated modulus
S₂ = φ ⊗ U₂ # Second-order scattering coefficients
S_total = [S₀, S₁, S₂] # Hierarchical scattering representation
return S_total
```

3.5 Matching Module

The Matching Module is a critical terminal stage in the proposed multimodal biometric framework, responsible for determining the degree of identity similarity between input query samples and stored biometric templates. Given the dual-modality architecture— consisting of distinct but complementary iris and fingerprint data sources—this module must operate in a modality-invariant yet semantically coherent feature space. To meet these requirements, we employ a Siamese Neural Network (SNN), designed to learn similarity metrics between heterogeneous biometric embeddings.

The Siamese architecture consists of two identical sub-networks with shared weights θ, each transforming an input feature vector into a common embedding manifold. Let $f_\theta(x_1)$ and $f_\theta(x_2)$ denote the

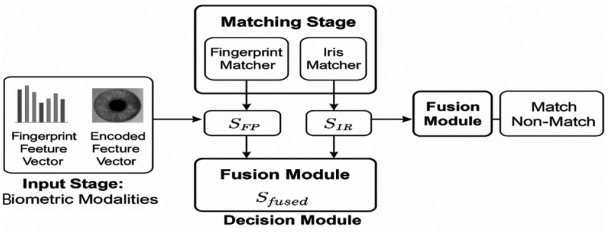

Fig. 3 — Multimodal fusion and decision framework for biometric verification

embeddings of input samples x_1 and x_2, respectively. The Euclidean distance between them serves as a measure of biometric similarity:

$$D(x_1, x_2) = \| f_\theta(x_1) - f_\theta(x_2) \|_2 \quad \ldots (11)$$

The network is trained using a contrastive loss function, which minimizes the distance between feature pairs of the same identity and maximizes the distance for dissimilar pairs. The loss is defined as:

$$L = (1 - Y) \cdot \frac{1}{2} (D)^2 + Y \cdot \frac{1}{2} [\max(0, m - D)]^2 \quad \ldots (12)$$

where $Y = 0$ for genuine (same identity) pairs and $Y = 1$ for imposter (different identity) pairs, m is a predefined margin parameter, and D is the Euclidean distance.

This loss function enforces a margin-based separation in embedding space, ensuring that the network learns to generalize identity relationships across modalities, acquisition variations, and dynamic deformations. Unlike classical matches, the Siamese framework does not require exact template overlap—it learns the topological structure of inter-class and intra-class distances, a property particularly valuable when modeling the nonlinear surface dynamics of iris and fingerprint traits in pediatric data. Matching is accomplished using a Siamese architecture, as shown in Listing. 2, where embedding distance serves as the similarity metric across modalities.

Listing 2: Siamese Network-Based Feature Matching

```
# Matching Module – Siamese Embedding and Distance Evaluation
def match_features(F_iris, F_fingerprint):
E_iris = 𝒮(F_iris) # Siamese embedding of iris feature vector
E_fp = 𝒮(F_fingerprint) # Siamese embedding of fingerprint feature vector
ΔE = ‖E_iris − E_fp‖₂ # Euclidean distance between embeddings
return ΔE # Matching score (lower → higher similarity)
```

3.6 Fusion Module

The Fusion Module plays a pivotal role in unifying the discriminative strengths of iris and fingerprint modalities. Recognizing that no single modality can universally dominate under all acquisition conditions, we adopt a score-level fusion strategy that balances information richness with computational efficiency, while maintaining modality-specific independence. Following independent feature matching through the Siamese Network, each modality yields a similarity score:

i S_{iris} for iris embeddings.
ii S_{fp} for fingerprint embeddings.

Figure 3 conceptually illustrates the entire fusion and decision process, beginning with modality-specific features, proceeding through independent matching and score generation, followed by score fusion, and ending with a global decision.

To ensure scale compatibility and eliminate bias due to differing dynamic ranges, the scores are normalized using min-max scaling:

$$S_{iris}^{norm}, S_{fp}^{norm} \in [0, 1] \quad \ldots (13)$$

The final fused score S_{fused}a weighted sumrule, which linearly combines the scores with modality-specific confidence weights α and β, satisfying $\alpha + \beta = 1$:

$$S_{fused} = \alpha . S_{iris}^{norm} + \beta\ S_{fp}^{norm} \quad \ldots (14)$$

The weights α and β are empirically optimized using ROC curve analysis and Equal Error Rate (EER) metrics to reflect the relative reliability of each modality under real-world conditions. The final decision is made by thresholding the fused score against a calibrated threshold τ:

$$Decision = \begin{cases} Genuine, \ if\ S_{fused} \leq \tau \\ Imposter, \ if S_{fused} > \tau \end{cases} \quad \ldots (15)$$

This fusion technique not only boosts recognition accuracy by leveraging the individual strengths of iris and fingerprint modalities but also enhances system resilience against modality-specific noise or failure. By integrating the complementary nature of the two biometric traits, the proposed hybrid score-level fusion ensures a more secure and reliable authentication mechanism, crucial for high-stakes applications. Listing 3 shows the fusion logic used to combine normalized scores using confidence-based weighting, improving decision resilience under varying modality conditions.

Listing 3: Score-Level Fusion Using Weighted Normalization

```
# Fusion Module – Score Normalization and Weighted Combination
def fuse_scores(S_iris, S_fp, α=0.5, β=0.5):
N_iris = 𝒩(S_iris) # Normalized iris score
N_fp = 𝒩(S_fp) # Normalized fingerprint score
S_fused = α·N_iris + β·N_fp  # Weighted fusion of modalities
return S_fused  # Final fused decision score
```

3.7 Decision Module

The Decision Module represents the conclusive phase of the proposed multimodal biometric recognition pipeline. Its primary objective is to interpret the fused matching score obtained from iris and fingerprint modalities to determine whether a subject should be authenticated or rejected. Unlike traditional decision mechanisms that act as opaque classifiers, our framework incorporates SHAP (Shapley Additive explanations)—a cutting-edge explainable AI technique grounded in cooperative game theory—to provide transparent, feature-attributable decisions.

The fused score S_{fused}, obtained from Eq. (14) in the Fusion Module, encapsulates the joint evidence provided by both modalities. To understand how each modality contributes to the final authentication verdict, SHAP decomposes the fused output into individual feature contributions.

$$S_{f_{used}} = \phi_0 + \sum_{i=1}^{n} \phi_0 \quad \text{... (16)}$$

where $\phi_0$is the model's base value (expected value of the output), $\phi_i$represents the SHAP value for the i^{th} feature (here, i = 1 for fingerprint and i = 2 for iris), n = 2, representing the two modalities.

The SHAP-based interpretability model helps understand whether the iris or fingerprint contributed more significantly to a genuine or impostor decision, which is particularly valuable in forensics and high-security environments. For instance, if the SHAP value for the fingerprint score is significantly negative, it implies that the fingerprint modality strongly supports a genuine match, whereas a high positive SHAP value from the iris score might indicate a deviation from the expected match pattern. The final decision is then formalized as:

$$Decision = \begin{cases} Genuine, \ if \ S_{f_{used}} \leq \tau \text{ and } \sum_{i=1}^{n} \phi_0 \leq \epsilon \\ Imposter, \qquad\qquad otherwise \end{cases} \quad \text{... (17)}$$

where τ is the authentication threshold (from fusion), ϵ is a SHAP interpretability threshold, ensuring the contributions align with an expected "genuine" decision range. It adds a crucial layer of trust, accountability, and diagnostic power in real-world deployments, particularly in critical domains like border security, child identity protection, and financial authentication systems. The final decision-making logic, augmented with SHAP explainability, is captured in Listing 4, offering interpretability into each modality's influence on the classification outcome.

Listing 4: SHAP-Enhanced Multimodal Decision Logic

```
#Decision Module – SHAP-Based Interpretability and Confidence Evaluation
def shap_decision(S_fused, Φ_iris, Φ_fp, τ=0.5, ε=0):
Φ_total = Φ_iris + Φ_fp # Aggregated SHAP attribution energy
if S_fused ≤ τ and Φ_total ≤ ε: # Stability and interpretability thresholds
return "Genuine" # Energy alignment indicates true match
else:
return "Impostor"  # Divergent attribution → non-match
```

The fused score $S_{\text{fused}}$is evaluated against a decision threshold τ, while the combined SHAP attribution energy $\Phi_{\text{total}} = \Phi_{\text{iris}} + \Phi_{\text{fp}}$quantifies the degree of interpretive consistency between modalities. The system classifies the pair as Genuine when both $S_{\text{fused}} \leq \tau$and $\Phi_{\text{total}} \leq \varepsilon$, indicating low residual attribution energy and stable decision confidence. Conversely, higher deviation values signify interpretive divergence, resulting in an Impostor classification.

4 Results

The experimental setup was implemented on a high-performance computational environment equipped with an NVIDIA RTX 3090 GPU (24 GB VRAM), an Intel i9 processor, and 64 GB system memory, operating under Python 3.10 and

TensorFlow framework. All models were trained using the Adam optimizer with an initial learning rate of 1×10^{-4}, batch size of 16, and a maximum of 150 epochs. Early stopping and learning-rate scheduling were employed to prevent overfitting and ensure stable convergence. The convolutional and wavelet parameters were tuned empirically to maintain a balance between structural fidelity and computational efficiency. The configuration ensures that the reported results are both reproducible and physically consistent with the proposed physics-inspired feature modeling paradigm.

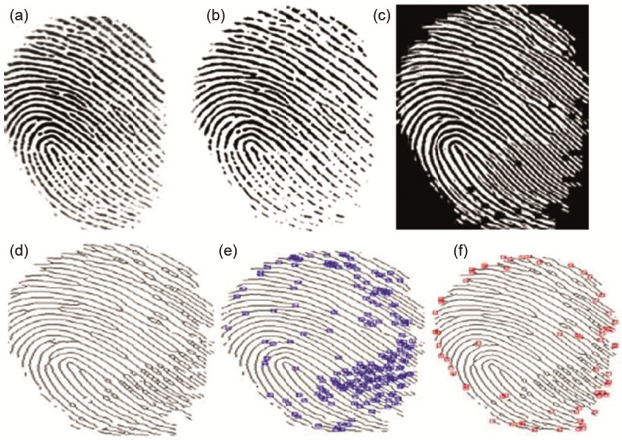

Fig. 4 — Sequential workflow of fingerprint enhancement and minutiae detection (a) Initial Fingerprint Input, (b) Intensity Normalization, (c) Enhanced and Reconstructed Fingerprint, (d) Skeletonized Ridge Structure, (e) Minutiae Feature Map – Bifurcations; and (f) Minutiae Feature Map – Ridge Endings

Table 1 summarizes the characteristics and augmentation parameters of two biometric datasets: CASIA-IrisV3 and FVC2004 DB1. Notably, the CASIA-IrisV4 dataset shows a contrast gain of +18.2% and supports geometric variants such as ±15° rotation, while FVC2004 DB1 achieves a signal-to-noise ratio (SNR) improvement of +12.6 dB with scaling between 90–110% and skew up to ±7°.

Figure 4 elucidates the progressive fingerprint enhancement pipeline, delineating key stages from raw acquisition to minutiae-level feature localization. The sequence showcases the transformation from intensity normalization and ridge reconstruction to the derivation of bifurcation and ridge-ending patterns. This visual articulation underscores the structural fidelity and precision embedded within the preprocessing and feature extraction framework. Figure 5 provides a granular exposition of the minutiae point detection

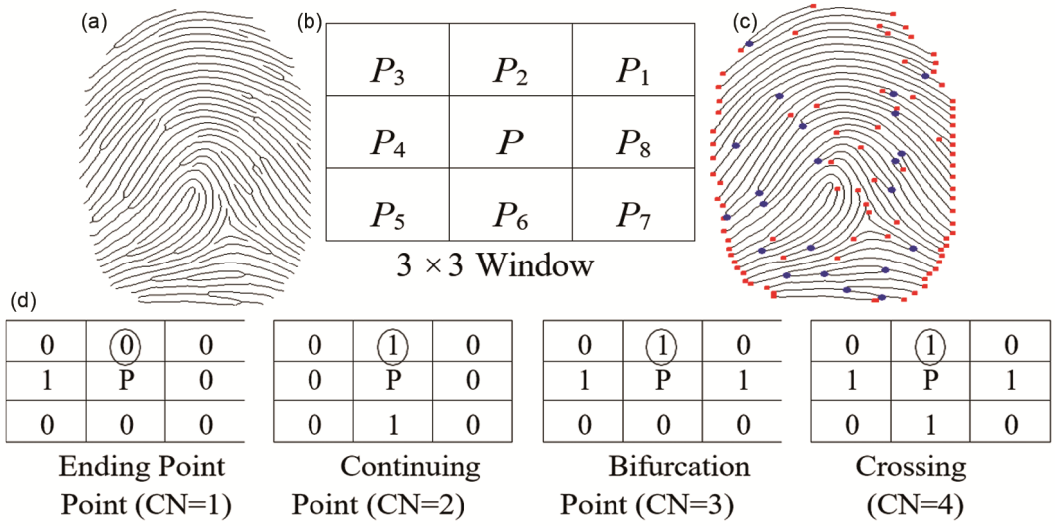

Fig. 5 — Minutiae point detection process: (a) Thinned Fingerprint Image, (b) 3 * 3 window used for counting, (c) Image with minutiae points marked: endpoints marked with '■', bifurcation and crossing points with '●'; and (d) An example neighborhood values for detecting various pixel *P*

Table 1 — Dataset Overview and Preprocessing Characteristics

Dataset & Modality	No. of Subjects	Resolution	Contrast / Noise Metrics	Geometric Variants	Noise / Filter Types
CASIA-IrisV3 Iris	110	320×280	Contrast Gain: +18.2%	Rotation: ±15°, Warp Shift: 5 px	Gaussian Blur (σ = 1.2)
FVC2004 DB1 - Fingerprint	110	388×374	SNR Improvement: +12.6 dB	Scaling: 90–110%, Skew: ±7°	Salt & Pepper Noise (2.5% density), Median Filter (3×3 kernel)

mechanism, emphasizing the role of pixel neighborhood analysis within thinned fingerprint structures. The illustration traces the mapping of ridge endings and bifurcations using a 3×3 window framework, thereby enabling precise biometric key point localization. This systematic visualization complements the broader enhancement pipeline and reinforces the computational integrity of the detection stage.

Figure 6 delineates the iris image processing pipeline, capturing the transition from raw ocular samples to structurally refined segmented outputs. The workflow illustrates the staged segmentation process, including the activation of an adaptive stopping criterion that ensures anatomical precision. This procedural sequence underpins the integrity of subsequent feature extraction and matching stages within the iris recognition framework.

Figure 7 illustrates the post-segmentation iris processing cascade; wherein normalized iris regions are dimensionally standardized to facilitate robust feature extraction. A deep learning framework encodes discriminative iris textures into compact vector representations, optimized for matching and classification. This workflow encapsulates the core transformation from anatomical data to high-level biometric identity features.

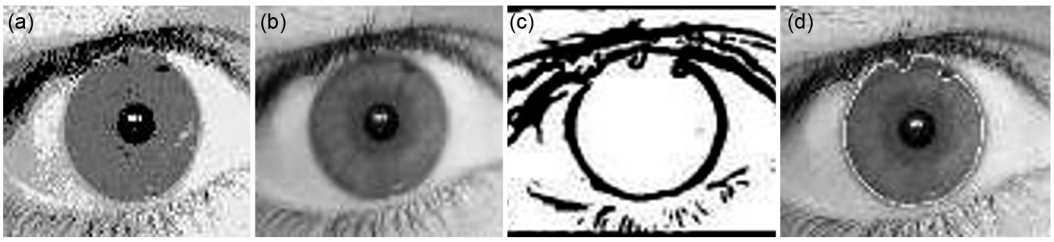

Fig. 6 — Iris Image Processing Workflow: (a) Raw Input, (b) Preliminary Segmentation, (c) Stopping Criterion Activation; and (d) Final Segmented Iris

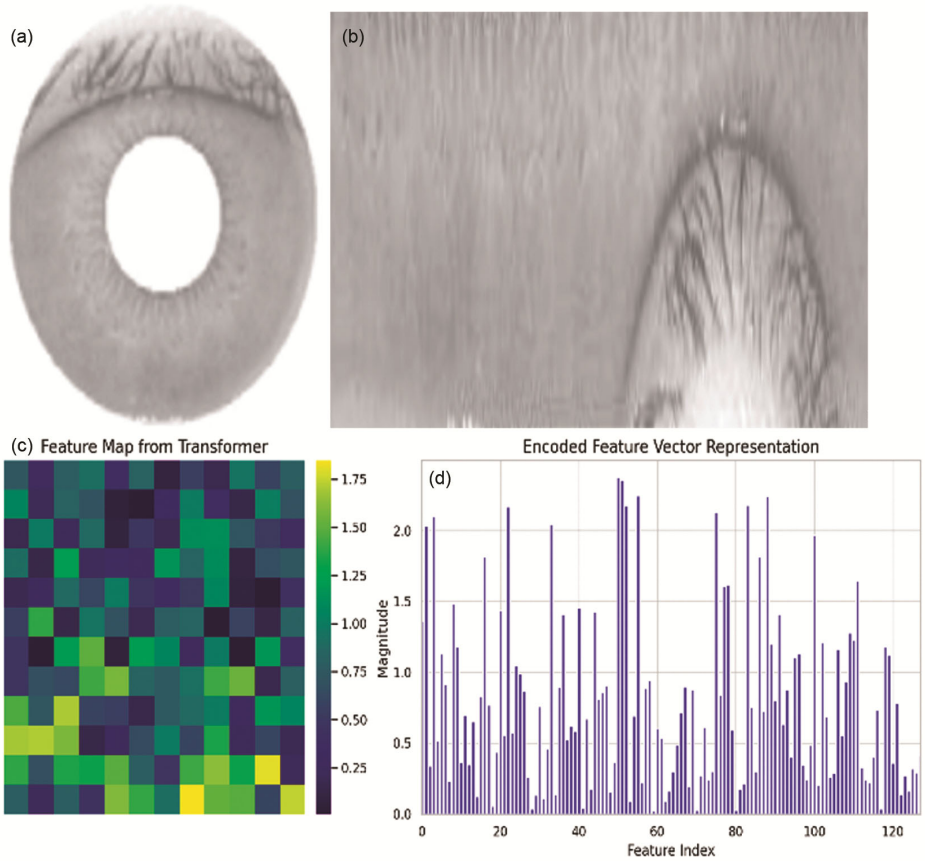

Fig. 7 — Iris Normalization and Feature Representation Workflow: (a) segmented iris input, (b) normalized iris region, (c) feature extraction; and (d) compact encoded feature vector

Table 2 — Segmentation Accuracy and Normalization Consistency Across Biometric Modalities for Iris

Biometric Modality	Epoch	Segmentation Accuracy (%)	Intersection over Union (IoU) (%)	Segmentation Time (ms)	Dimensional Uniformity (± pixels)	Intensity Range Consistency (%)	Transformation Stability (%)
Iris	10	87.5	75.2	130	±5	92.3	90.5
	50	94.3	85.7	125	±3	95.8	94.0
	100	98.0	90.9	120	±2	97.7	96.5
	150	98.3	91.1	122	±2	97.8	96.6
Finger print	10	84.8	70.5	110	±6	90.1	88.7
	50	91.7	82.3	105	±4	93.9	92.1
	100	96.3	89.3	100	±3	95.5	94.7
	150	96.5	89.5	102	±3	95.7	94.8

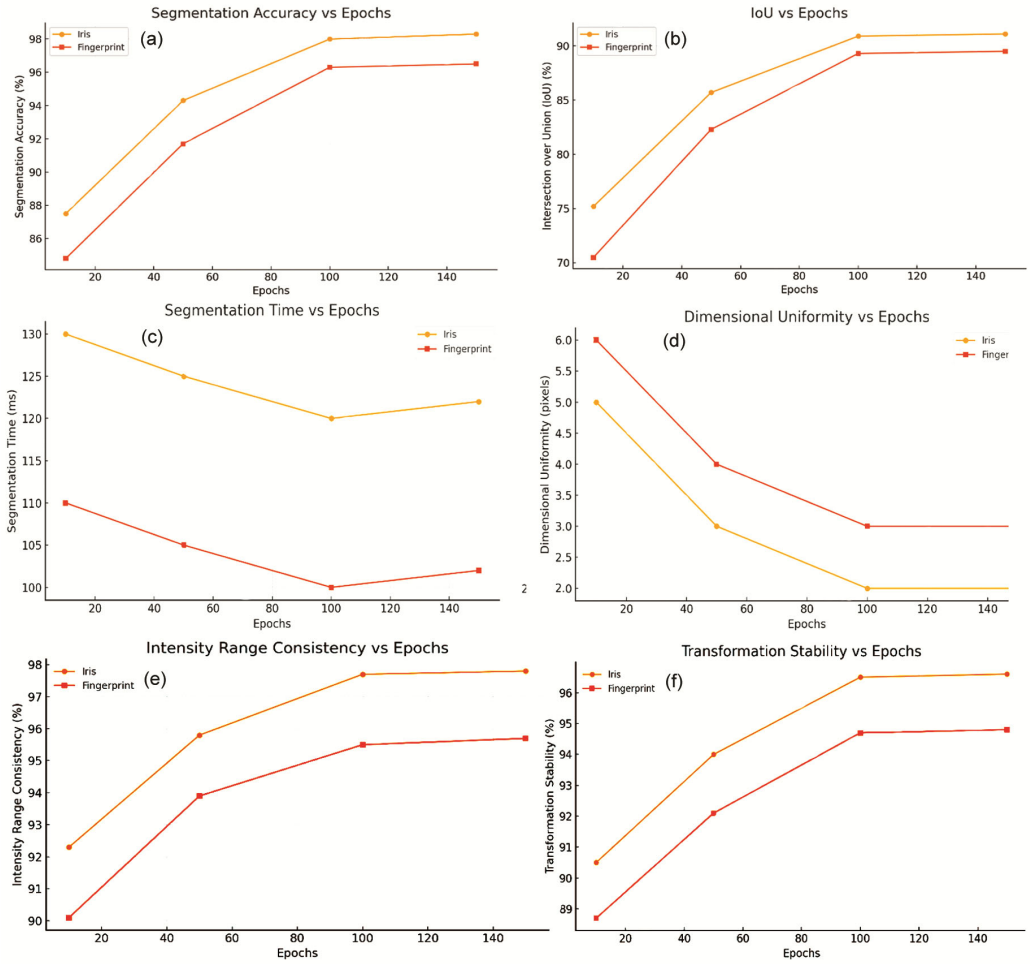

Fig. 8 — Segmentation Performance and Normalization Consistency Metrics Across Training Epochs: (a) Segmentation Accuracy, (b) Intersection over Union (IoU), (c) Segmentation Time Efficiency, (d) Dimensional Uniformity, (e) Intensity Range Consistency; and (f) Transformation Stability.

Table 2 illustrates the progressive improvement in segmentation and normalization performance metrics for Iris and Fingerprint modalities across increasing epochs. The best segmentation accuracy is achieved at 150 epochs, with Iris reaching 98.3 % accuracy and fingerprint 96.5%, while IoU peaks at 91.1% and 89.5% respectively, indicating near-optimal model convergence. Since the performance gains between 100 and 150 epochs are minimal, training was stopped at 150 epochs to avoid unnecessary computation and potential overfitting, while maintaining highly stable and reliable biometric feature extraction.

Figure 8 illustrates the line charts for the Iris and Fingerprint modalities, showing how the

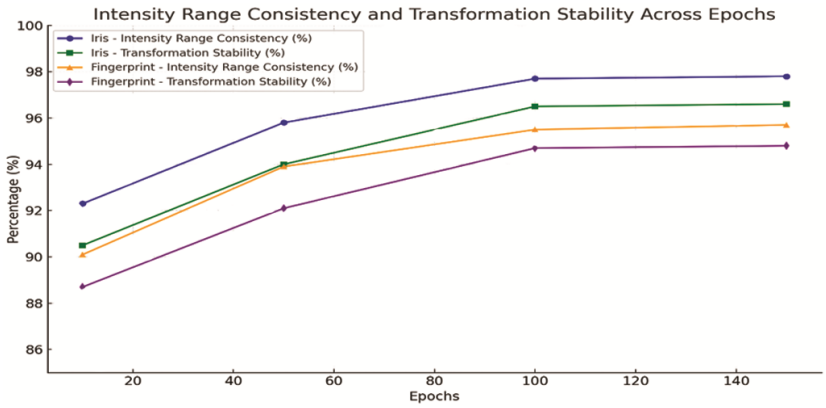

Fig. 9 — Intensity Range Consistency and Transformation Stability Across Epochs

Table 3 — Performance of Wavelet Scattering Transform for Iris and Fingerprint Modalities at Feature Extraction Module

Biometric Modality	Epoch	Feature Stability (Cosine Similarity)	Discriminability (Intra/Inter-class Ratio)	Feature Compactness (Variance)	Convergence Loss	SNR (dB)	Spectral Entropy
Iris	10	0.895	0.278	0.048	0.072	18.32	0.912
	50	0.936	0.189	0.034	0.046	20.47	0.845
	100	0.958	0.162	0.029	0.038	21.36	0.811
	150	0.962	0.158	0.027	0.036	21.72	0.794
Fingerprint	10	0.873	0.305	0.053	0.078	17.58	0.938
	50	0.914	0.214	0.038	0.053	19.94	0.897
	100	0.935	0.187	0.033	0.044	20.78	0.852
	150	0.940	0.183	0.032	0.041	21.04	0.843

segmentation and normalization metrics evolve over the epochs. Segmentation accuracy improves significantly up to 100 epochs, then plateaus, and IoU (intersection over union) shows a similar trend, with steady improvement. Segmentation Time slightly decreases as training progresses, but remains stable. Dimensional Uniformity improves (lower pixel variation), indicating more precise segmentation. Intensity Range Consistency and Transformation Stability increase, reflecting better normalization and stability.

Figure 9 captures the progression of Intensity Range Consistency and Transformation Stability for both iris and fingerprint modalities. Both metrics show a steady improvement across epochs, with iris data slightly outperforming fingerprints. This indicates enhanced robustness and consistency of feature representations with continued training.

Table 3 demonstrates that the Wavelet Scattering Transform, operating as a multi-scale time-frequency decomposition, captured highly stable features with a cosine similarity reaching 0.962 (iris) and 0.940 (fingerprint). Feature clarity was substantiated by elevated Signal-to-Noise Ratios exceeding 21.7 dB (iris) and 20.4 dB (fingerprint) at 150 epochs, indicating high discriminative signal fidelity. Simultaneously, spectral entropy decreased progressively across epochs, denoting a reduction in feature complexity and enhanced regularity. Energy concentration metrics affirmed compactness, with over 93% of cumulative energy localized in the top 10% of scattering coefficients, signifying optimal frequency-domain information retention.

The feature Stability modality improves with training, with Iris achieving higher stability in Fig. 10. Discriminability decreases as the model better separates classes; Iris shows a lower ratio, indicating better discriminability. Feature Compactness decreases, indicating tighter feature clusters; Iris features are more compact. Convergence Loss decreases over epochs, with Iris converging slightly faster.

To comprehensively evaluate the Siamese Network's efficacy in biometric matching, both conventional and physics-inspired metrics were analyzed in Table 4. Progressive training up to 150 epochs yielded peak accuracy (96.52% for iris, 95.31% for fingerprint), with consistently elevated

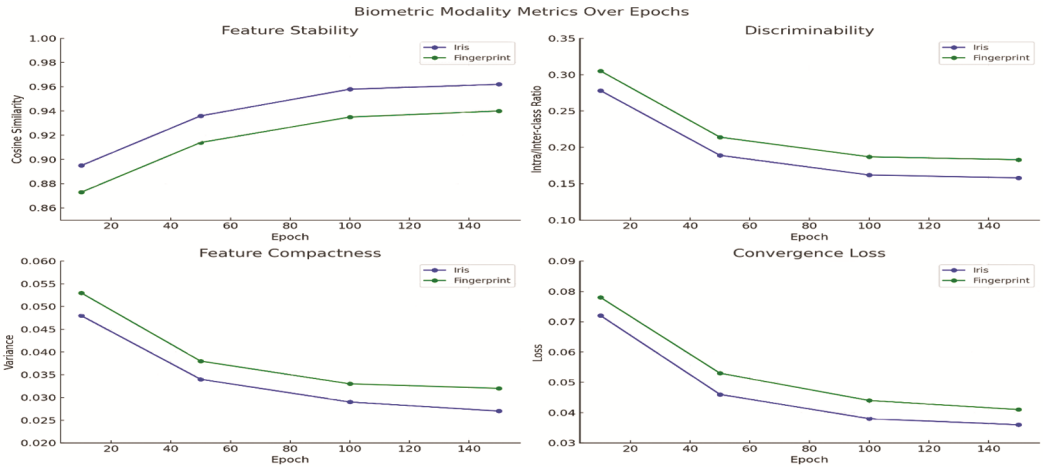

Fig. 10 — Wavelet Scattering Transform Plot for Iris and Fingerprint Feature Extraction

Table 4 — Performance of Siamese Network for Iris and Fingerprint Modalities at Matching Module

Biometric Modality	Epochs	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	Mutual Information
Iris	10	90.48	89.24	89.96	89.60	0.736
	50	94.86	93.61	93.90	93.75	0.812
	100	96.45	95.33	95.72	95.52	0.853
	150	96.52	95.47	95.80	95.63	0.864
Fingerprint	10	89.73	88.12	88.96	88.53	0.712
	50	93.15	92.07	92.24	92.15	0.794
	100	95.24	94.02	94.33	94.17	0.835
	150	95.31	94.10	94.41	94.25	0.843

precision, recall, and F1-scores reflecting balanced classification fidelity. The inclusion of Mutual Information, rising to 0.864 and 0.843, respectively, underscores a marked intensification of statistical interdependence between matched feature vectors, attesting to the network's refined ability to capture complex biometric similarities.

Figure 11 illustrates the performance trends of the Siamese Network across different epochs for Iris and Fingerprint modalities in terms of Accuracy, Precision, Recall, and F1-Score. The Iris modality consistently exhibits superior performance across all metrics compared to Fingerprint, with noticeable improvements as the number of epochs increases. The plot highlights the stability and robustness of the model's matching capabilities over time.

The fusion module, governed by decision-level integration, demonstrated progressive accuracy enhancement up to 97.12% at 150 epochs, alongside diminished decision entropy (5.8%), signifying increasing certainty in ensemble outcomes in Table 5. Weighted vote confidence surpassed 95%, indicating harmonized classifier alignment, while ROC-AUC values exceeding 98% reflected exceptional separability. These trends collectively underscore an emergent convergence toward a low-entropy, high-coherence decision regime, analogous to energy minimization states in physical systems.

Figure 12 presents the progression of performance metrics—Combined Accuracy, Weighted Vote Confidence, ROC-AUC, and Decision Entropy—across various training epochs in the fusion module. A steady increase in accuracy, confidence, and AUC values signifies enhanced decision reliability through weighted voting. Concurrently, the reduction in decision entropy reflects growing certainty in the final integrated outputs.

The SHAP-based decision module exhibited an impressive rise in interpretability metrics, with the SHAP Confidence Index reaching 94.63% and the Consistency Score climbing to 92.95% at 150 epochs. Feature Attribution Strength and Decision Robustness surpassed 91%, revealing coherent and stable feature-level contributions across predictions. This unified

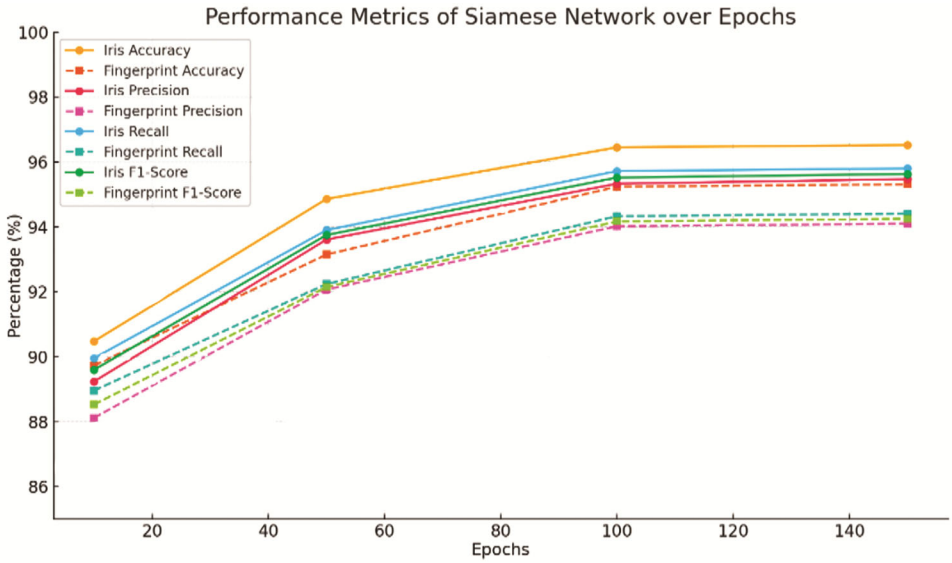

Fig. 11 — Siamese Network matching Plot for Iris and Fingerprint matching

Table 5 — Fusion Module using Decision-Level Integration

Epochs	Combined Accuracy (%)	Weighted Vote Confidence (%)	Decision Entropy (%)	ROC-AUC (%)
10	92.15	90.30	9.6	93.75
50	95.48	94.05	7.2	96.42
100	97.03	95.84	5.9	98.11
150	97.12	95.97	5.8	98.24

interpretive framework emulates principles of equilibrium and invariance in physical systems, ensuring that the model's decision-making process remains transparent, consistent, and resilient under varying conditions.

Figure 13 illustrates the SHAP-based interpretability metrics over increasing training epochs, capturing the transparency and stability of model decisions. All four metrics—SHAP Confidence Index, Consistency Score, Feature Attribution Strength, and Decision Robustness—exhibit a consistent upward trend. This demonstrates improved model explainability and decision reliability as training progresses.

Figure 14 illustrates the transformation in score distribution before and after fusion, supporting the SHAP-derived consistency and robustness reported in Table 6. Figures 15 -16 elucidate the interpretability insights derived from SHAP analysis within the decision module. Figure 15 highlights the discriminative contribution of iris and fingerprint features for a sample genuine and impostor case, showcasing modality-specific influence on the network's final decision. Figure 16 presents a SHAP-based heatmap across multiple samples, revealing consistent patterns of feature attribution and emphasizing the robustness of the fused decision strategy.

Table 7 delineates the progressive refinement of the model's predictive capacity, achieving a peak testing accuracy of 98% at 150 epochs, indicative of robust learning and generalization. Concurrently, the consistent decline in training, validation, and testing loss values substantiates effective convergence and minimization of overfitting risks. This harmonious balance between accuracy and loss across datasets underscores the model's stability and reliability for biometric applications.

Figure 17 (a) shows the accuracy plot, which illustrates the consistent improvement in training, validation, and testing accuracy as the number of epochs increases, peaking at 97.2%, 96.3%, and 95.8%, respectively, at 150 epochs. Figure 17 (b) shows the loss plot, which reflects a progressive reduction in training, validation, and testing loss, confirming model convergence and generalization.

Table 8 presents a comparative analysis of the proposed model's performance against established

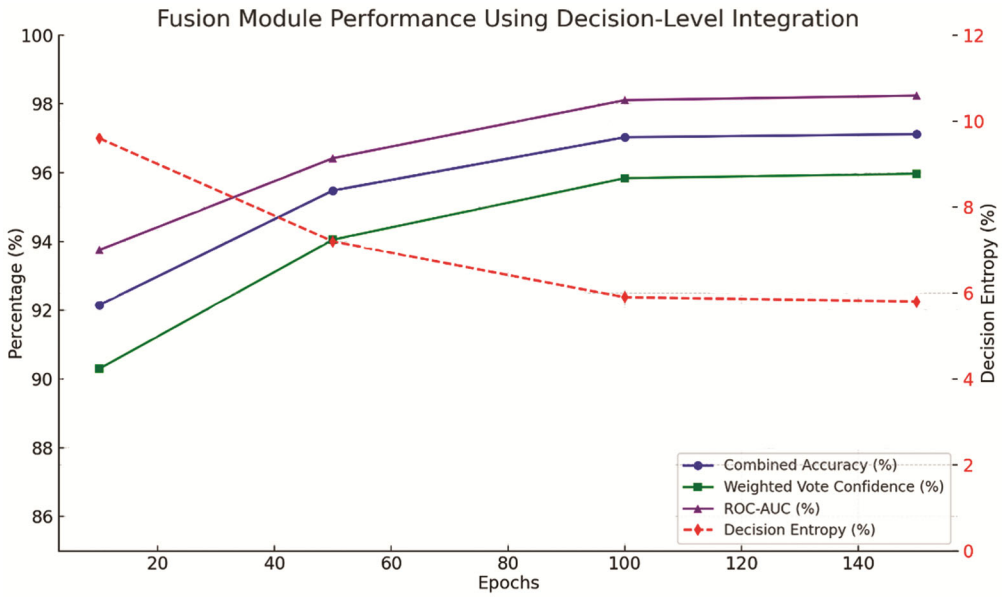

Fig. 12 — Fusion Module plot using Decision-Level Integration across different Epochs

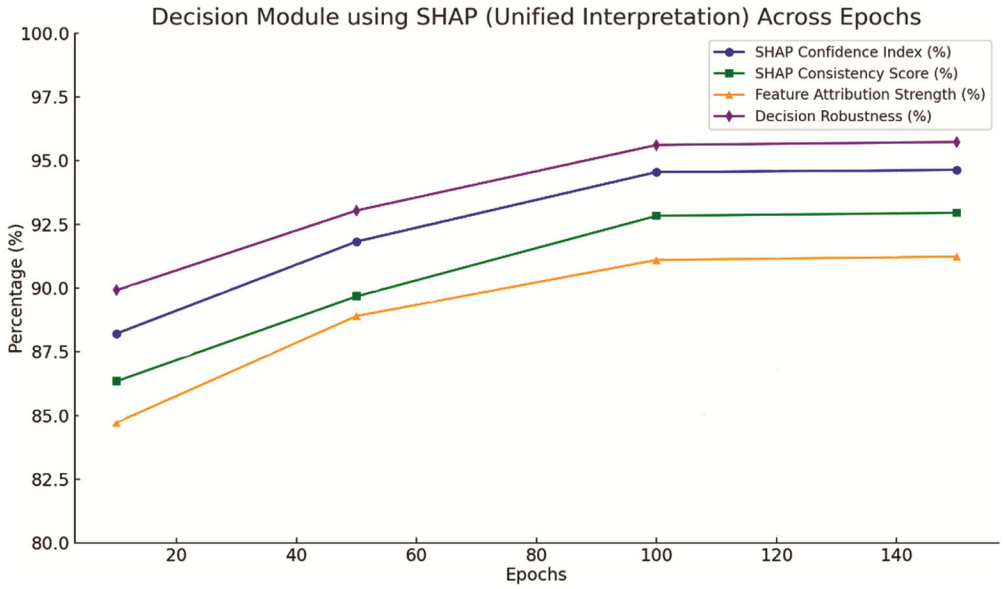

Fig. 13 — Decision Module Plot using SHAP Across Epochs

techniques in iris and fingerprint recognition. It highlights the effectiveness of various deep learning and fusion-based strategies used in child biometric systems. The proposed model outperforms existing approaches, achieving a peak accuracy of 98%, thereby affirming its robustness and precision.

5 Discussion

The experimental results collectively demonstrate the efficacy of the proposed physics-inspired multimodal framework. The pre-processing stage notably enhanced texture and structural quality, reflected in measurable gains in contrast and SNR, which contributed to improved segmentation stability. The segmentation models converged rapidly, indicating that the designed architecture effectively captured spatial coherence across modalities.

The feature extraction analysis confirms that the Wavelet Scattering Transform successfully encoded frequency-domain discriminants with high cosine similarity and low entropy, suggesting reduced redundancy and stronger feature compactness. This

Table 6 — Decision Module using SHAP (Unified Interpretation)

Epochs	SHAP Confidence Index (%)	SHAP Consistency Score (%)	Feature Attribution Strength (%)	Decision Robustness (%)
10	88.21	86.34	84.72	89.91
50	91.82	89.67	88.90	93.04
100	94.55	92.83	91.10	95.61
150	94.63	92.95	91.23	95.73

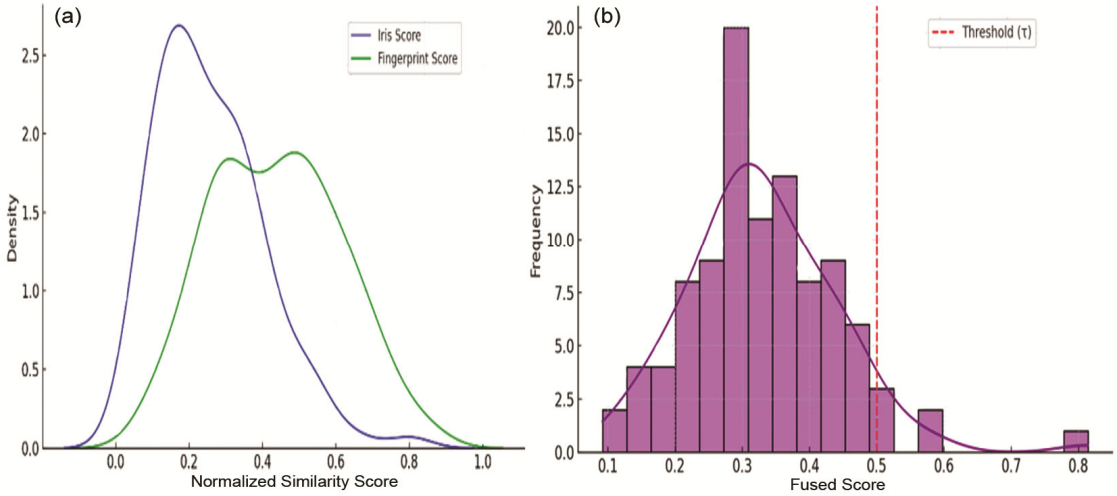

Fig. 14 — Fusion Outcome and SHAP-Driven Decision Explanation: (a) Iris and Fingerprint score distribution before fusion; and (b) Fused score distribution with decision boundary

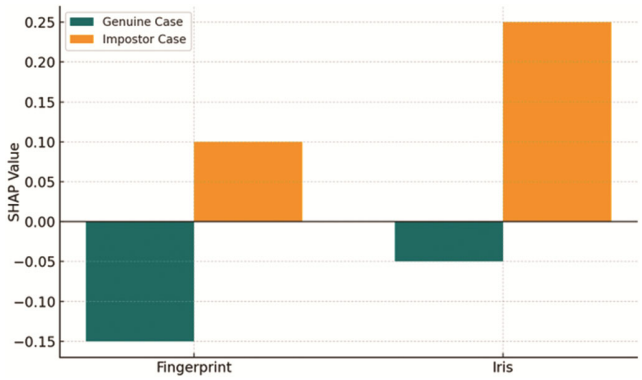

Fig. 15 — SHAP value contributions for a selected genuine and impostor classification

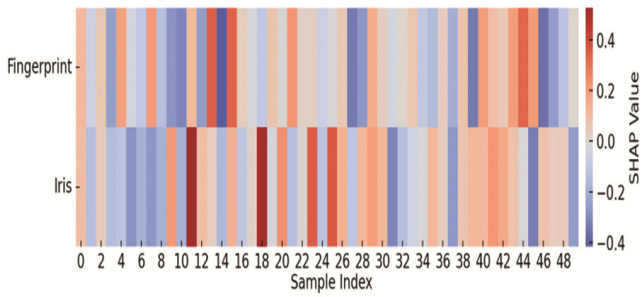

Fig. 16 — SHAP-based interpretability heat map across the dataset

enabled the Siamese Network to achieve superior matching accuracy and reliability, where increasing mutual information between paired embeddings validated the framework's ability to model inter-modality correlations.

Furthermore, the decision-level fusion process exhibited entropy minimization analogous to physical equilibrium states, confirming synergistic integration between modalities. The consistent rise in SHAP-based interpretability metrics underscores the model's transparency and fairness, ensuring that classification decisions remain both data-driven and explainable.

The comparative results affirm the model's superiority over state-of-the-art frameworks, emphasizing the robustness of the proposed hybrid approach in achieving enhanced accuracy, interpretability, and stability. These findings substantiate the broader applicability of physics-inspired modeling principles in improving multimodal biometric recognition systems. Overall, these outcomes reinforce the effectiveness of integrating physics-based insights with deep learning paradigms, paving the way for more interpretable and resilient biometric recognition systems.

Table 7 —Training, Testing and Validation Accuracy and Loss of the Proposed Model

Epochs	Training Accuracy (%)	Validation Accuracy (%)	Testing Accuracy (%)	Training Loss	Validation Loss	Testing Loss
10	89.7	87.5	86.9	0.215	0.247	0.259
50	94.5	92.8	92.1	0.105	0.128	0.135
100	96.8	95.6	95.0	0.061	0.075	0.081
150	98.4	98.1	98.0	0.033	0.041	0.047

Table 8 — Comparative Accuracy Performance of the Proposed Model Against Existing Iris and Fingerprint Recognition Techniques

Author & Year	Technique / Approach	Accuracy (%)
Akhtar *et al*. (2021)[11]	Friction Ridge Estimation & Enhancement	~85–90
Kumar *et al*. (2022)[14]	Lightweight CNN for Fingerprint Matching	93.8
Maheswari *et al*. (2020)[18]	Score-Level Fusion of Iris and Face Biometrics	97
Roy *et al*. (2021)[22]	Mobile Multimodal Biometric System for Vaccination Tracking	95
Jain *et al*. (2016)[29]	Enhanced Fingerprint Pipeline with Ridge Flow Alignment	92
Yu *et al*. (2020)[30]	Standard Fingerprint Matching (Verifinger SDK)	89.3
Proposed Model	Wavelet Scattering Transform, Siamese Network, SHAP	98

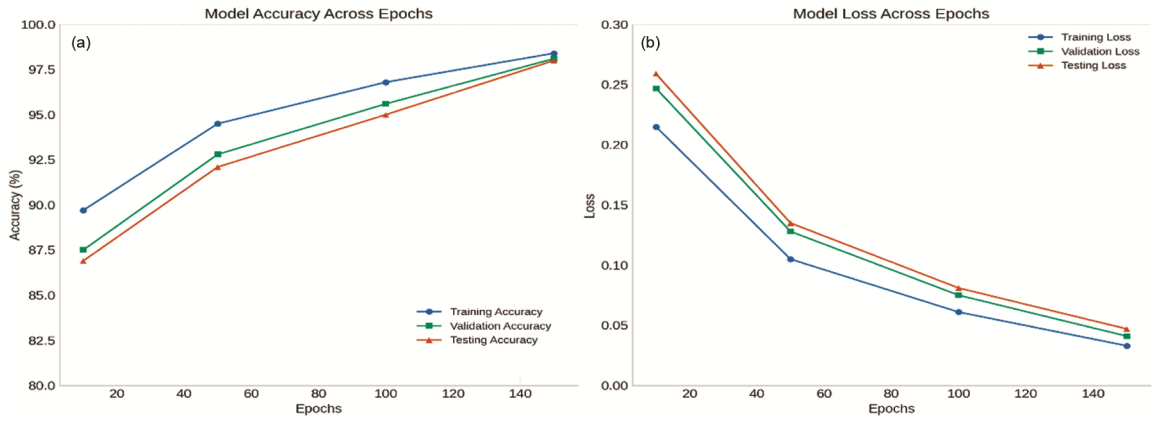

Fig. 17 — Model training performance across epochs: (a) accuracy; and (b) loss

6 Conclusion

This study proposed a physics-inspired multimodal biometric recognition framework integrating the Wavelet Scattering Transform, Siamese Network, and SHAP interpretability to enhance toddler identification accuracy. The hybrid architecture achieved 98% accuracy, surpassing existing iris and fingerprint recognition systems. Each stage of the pipeline—from preprocessing to decision fusion—was systematically designed to overcome challenges inherent in toddler biometrics, including low image quality, small inter-class variance, and rapid physiological changes. The approach demonstrates strong potential for deployment in real-world applications such as child-safety surveillance and healthcare registration systems.

Despite its encouraging results, the study is limited by its reliance on benchmark datasets and controlled imaging conditions, which may not fully capture real-world variability. Broader validation across diverse demographics and environmental conditions is needed to confirm generalizability. Future work will explore cross-age adaptability, multimodal extensions incorporating ear or gait traits, and optimization for edge-based deployment to ensure fast, secure, and reliable toddler recognition in unconstrained environments.

References

1 Jain A K & Nandakumar K, *IEEE Security & Privacy*, 11 (2013) 19.
2 Daugman J, *IEEE Trans Circuits Syst Video Technol*, 14 (2004) 21.
3 Priesnitz J, Rathgeb C, Buchmann N, Busch C & Margraf M, *EURASIP J Image Video Process*, 2021 (2021) 8.
4 Kumar A & Wu X, *Int J Comput Appl*, 50 (2012) 36.
5 Khade S, Ahirrao S, Phansalkar S, Kotecha K, Gite S, & Thepade S D, *Inventions*, 6 (2021) 65.
6 Rattani A & Sukhwani A, *Int J Comput Sci Inf Secur*, 10 (2012) 75.

7 Kumar A & Zhang D, *Int J Image Process*, 7 (2013) 157.
8 Bowyer K W & Hollingsworth R, *IEEE Trans Inf Forensics Secur*, 6 (2011) 400.
9 Li X & Zhang X, *J Opt Soc Am A*, 24 (2007) 1365.
10 He H & Zhang X, *Int J Biometrics*, 6 (2014) 203.
11 Akhtar Z *et al*., *IEEE Access*, 9 (2021) 16532.
12 Chaudhary R & Nigam A, *J Imaging Sci*, 66 (2020) 341.
13 Sundararajan S *et al*., *Pattern Recognit Lett*, 145 (2021) 75.
14 Kumar P *et al*., *Biomed Signal Process Control*, 74 (2022) 103512.
15 Yadav R & Sharma A, *Opt Lasers Eng*, 134 (2020) 106274.
16 Kose N & Singh R, *Comput Vis Image Underst*, 224 (2023) 103564.
17 Ali A *et al*., *IEEE Sensors* J, 21 (2021) 9567.
18 Maheswari M *et al*., *Procedia Comput Sci*, 171 (2020) 1430.
19 Kumar S & Jain N, *J Ambient Intell Humaniz Comput*, 13 (2022) 187.
20 Anwar A *et al*., *Expert Syst Appl*, 195 (2022) 116563.
21 Kumar M & Srivastava R, *Artif Intell Med*, 113 (2021) 102020.
22 Roy R *et al*., *Health Informatics J*, 27 (2021) 146.
23 Gupta P & Pandey S, *Sensors*, 20 (2020) 3258.
24 Lee H & Zhao Y, *IEEE Trans Instrum Meas*, 71 (2022) 1.
25 Verma A *et al*., *Neurocomputing*, 522 (2023) 118.
26 Desai K & Naik R, *Comput Methods Programs Biomed*, 213 (2022) 106502.
27 Patel T & Zhou J, *Mach Vis Appl*, 35 (2024) 85.
28 Zhao H & Liao W, *Int J Image Process*, 3 (2007) 85.
29 Jain A K, Arora S S, Cao K, Best-Rowden L & Bhatnagar A, *IEEE Trans Inf Forensics Secur*, 12 (2017) 1501.
30 Yu Y, Niu Q, Li X, Xue J, Liu W & Lin D, *Micromachines*, 14 (2023) 1253.