



## A Hybrid Graph Neural Network Framework Integrating Handcrafted Features for Real-Time Iris Recognition in Motion

Usman A. ABDURRAHMAN<sup>1\*</sup>, Abdulkadir A. BICHI<sup>2</sup>, Usman HARUNA<sup>3</sup>, Akibu M. ABDULLAHI<sup>4</sup>

<sup>1\*</sup>Department of Information and Communication Technology, Northwest University, Kano, Kano State, Nigeria

<sup>2,3</sup>Department of Software Engineering, Northwest University, Kano, Kano State, Nigeria

<sup>4</sup>School of Computing and Informatics, Albukhary International University, Alor Setar, Malaysia

<sup>1\*</sup>usmanalhaji79@gmail.com, <sup>2</sup>engraabubakar@gmail.com, <sup>3</sup>usmancyz01@gmail.com, <sup>4</sup>akibu1989@hotmail.com

### Abstract

*Iris recognition in dynamic environments, such as surveillance footage or real-time video streams, remains a significant challenge due to motion blur, occlusion, and the high computational cost of processing sequential frames. While traditional texture-based methods like Gabor filters struggle with motion deformations, modern deep learning approaches, particularly Graph Neural Networks (GNNs), offer superior spatial analysis but often at the expense of real-time performance. This paper proposes a novel hybrid framework that addresses these limitations by integrating handcrafted feature descriptors directly into a GNN architecture. Rather than relying solely on learned embeddings, the model initializes graph nodes using a fusion of traditional texture patterns and deep features, providing a richer and more resilient representation of the iris structure from the outset. Furthermore, we introduce a lightweight message-passing mechanism optimized for edge deployment, significantly reducing latency to meet the 25–30 frames-per-second requirement of real-time systems. By combining the interpretability and speed of traditional methods with the adaptive power of graph-based deep learning, the proposed approach enhances recognition accuracy under motion conditions while ensuring scalability for large databases. The proposed method achieves 95.1% recognition accuracy across three datasets, improving upon the original GNN framework by 2.3 percentage points and traditional methods by 9.2 points. It maintains robust performance under motion blur at 89.2% accuracy while operating at 35 frames per second on edge hardware. The system scales efficiently to one million users with query times of just 5.3 milliseconds. Experimental results demonstrate that this hybrid strategy outperforms both standalone deep networks and conventional algorithms, offering a viable path toward practical, real-world iris recognition in motion.*

**Keywords:** *Iris recognition, graph neural networks, feature fusion, moving image sequences, real-time biometrics.*

### 1.0 Introduction

Biometric identification has become a cornerstone of modern security systems, with iris recognition widely regarded as one of the most accurate and reliable methods available. The intricate and unique patterns of the human iris remain stable throughout an individual's lifetime, making them an ideal biometric marker. Since Daugman's foundational work in the early 1990s [1], iris recognition systems have evolved significantly, transitioning from constrained, cooperative environments to more dynamic and unconstrained settings. However, the proliferation of motion imagery in applications such as surveillance, border control, and mobile authentication has introduced new challenges that traditional systems are ill-equipped to handle [2].

When the subject's eye is in motion, captured images are often plagued by motion blur, which spreads edge information across multiple pixels and attenuates high-frequency components essential for texture analysis; occlusions from eyelids or eyelashes that intermittently obscure critical iris regions; and inconsistent illumination that alters local contrast and intensity distributions. These factors degrade the quality of the iris texture, rendering conventional feature extraction techniques, such as Gabor filtering and Local Binary Patterns, less effective [3]. Specifically, Gabor filters rely on precise orientation and frequency responses that become smeared under motion, while LBP's fixed neighborhood patterns lose discriminative power when spatial relationships are distorted by blur. As recent comprehensive reviews confirm, deep learning approaches have emerged as powerful alternatives, yet challenges persist in non-cooperative environments typical of surveillance applications [16, 26]. As noted by Bichi et al., motion images offer enhanced user experience and storytelling potential, but extracting high-fidelity biometric features from them requires a high degree of precision [2]. Furthermore, the sheer volume of data in video streams (25–30 frames per second) demands processing frameworks that are not only accurate but also computationally efficient enough for real-time deployment [2].

Recent literature has explored various avenues to address these challenges. Li and Zhang proposed motion-based feature extraction techniques for video-based iris recognition, focusing on tracking iris features across frames [4]. Similarly, Zhang et al. developed a dynamic framework that integrates feature tracking with motion compensation to mitigate the effects of blur and occlusion [5]. Deep learning approaches, particularly Convolutional Neural Networks (CNNs), have demonstrated remarkable success in image classification tasks and

have been applied to iris recognition with promising results [6]. More recently, Bichi et al. introduced a novel framework that utilizes Graph Neural Networks (GNNs) to model the spatial relationships between iris features, offering a sophisticated method for capturing both local textures and global topological structures [2]. This aligns with broader trends in computer vision where graph-based approaches have demonstrated significant potential for modeling structured spatial data [17, 18]. Despite these advances, existing methods often struggle to balance robustness against motion artifacts with the computational speed required for real-time applications. Many deep learning models remain too heavy for edge deployment, while traditional methods lack the adaptability to handle significant iris deformation [7].

This paper proposes a solution that bridges this gap through a hybrid approach, fusing traditional handcrafted texture features with deep learning capabilities within an optimized GNN architecture. Building upon the framework established by Bichi et al. [2], our method initializes graph nodes with a combination of conventional texture descriptors, such as Gabor responses, and deep feature embeddings extracted from local iris patches. This fusion provides a richer and more resilient representation of the iris structure, enabling the model to maintain accuracy even in the presence of motion blur. To address the computational demands of real-time processing, we introduce a lightweight message-passing mechanism specifically designed for efficient inference on resource-constrained hardware. By integrating the interpretability and speed of traditional methods with the adaptive power of graph-based deep learning, the proposed approach aims to enhance recognition accuracy in dynamic environments while ensuring scalability for large-scale biometric databases.

## 2.0 Methodology

### 2.1 Dataset and Acquisition

Experiments were conducted using three publicly available iris video databases, selected to provide complementary evaluation dimensions. The MBGC Visible Iris Video Database (640×480 resolution, 30 fps, varying illumination) was chosen because it contains visible-light imagery with challenging illumination variations, simulating uncontrolled outdoor surveillance conditions. The CASIA-Iris-Motion database provides sequences with intentional head and eye movement under near-infrared illumination, representing the more controlled but motion-prone conditions typical of border control scenarios. For controlled evaluation with known ground-truth motion parameters, a supplementary dataset was acquired using a Basler acA1300-60gm camera (Basler AG, Germany) with near-infrared illumination at 850nm. Videos were captured at 25 fps with 800×600 resolution at a distance of 30-50cm from the subject. This Kano-Iris-Motion dataset allowed us to systematically vary motion velocity and direction to quantify performance degradation under controlled conditions. The combination of datasets ensures our evaluation spans visible and NIR spectra, controlled and uncontrolled environments, and naturalistic versus systematic motion patterns.

### 2.2 Preprocessing and Iris Segmentation

Preprocessing comprised three steps. Boundary detection employed the Circular Hough Transform (CHT) implemented in OpenCV 4.8.0 (OpenCV team, USA). Images were convolved with a Gaussian filter ( $\sigma = 2.5$ ) and edges detected using the Canny operator (thresholds 50 and 150). The Gaussian sigma was selected based on empirical optimization on a validation set of 500 frames from the CASIA-Motion database, where  $\sigma = 2.5$  minimized segmentation errors by providing sufficient smoothing to reduce motion-induced noise while preserving essential iris boundary structure. Canny thresholds were set at 50 (low) and 150 (high) following the automated hysteresis thresholding approach suggested by OpenCV's implementation, which maintains the ratio of approximately 1:3 recommended in the literature for edge detection in biometric applications. CHT was applied with radius ranges of 40-80 pixels for the pupil and 100-160 pixels for the limbus, determined by analyzing the distribution of annotated iris boundaries in the training portions of all three datasets. These ranges encompass 95% of observed pupil and limbus radii across our combined datasets while excluding implausible values that would trigger false detections. When circular detection failed due to motion blur, the elastic contour refinement method of Zhang et al. [5] was applied, allowing a deformable spline to evolve under gradient constraints for up to 50 iterations. Recent advances in segmentation for motion-degraded imagery informed our parameter selection, particularly uncertainty-aware approaches for handling boundary ambiguity [15] and event-based methods for tracking rapid eye movements [21]. Eyelid occlusion was addressed using parabolic curve fitting to horizontal gradient maps, while eyelashes were detected by intensity thresholding (pixels below 30) and masked following Li and Zhang [4]. The detected iris region was normalized from Cartesian to polar coordinates using Daugman's rubber sheet model [1], producing a rectangular strip of 64×512 pixels.

### 2.3 Graph Construction from Iris Features

The normalized iris strip was transformed into a graph  $G = (V, E)$ . Nodes were positioned at key feature points detected using a modified SIFT detector [8] in OpenCV 4.8.0 (contrast threshold = 0.04, edge threshold = 10). Each node was associated with a fused 140-dimensional feature vector combining handcrafted texture features

(64-dimensional LBP histogram from a  $32 \times 32$ -pixel neighborhood plus 12-dimensional Gabor responses at 4 orientations and 3 scales) and deep features (64-dimensional embedding from a lightweight CNN patch encoder pre-trained on ND-IRIS-0405 [9] and fine-tuned using triplet loss). This fusion constituted the primary modification to the original Bichi *et al.* framework [2]. Edges were established using Delaunay triangulation of node coordinates in CGAL 5.5 (CGAL Project, France), with edge weights assigned based on spatial distance and angular alignment.

## 2.4 Graph Construction from Iris Features

The GNN consisted of three message-passing layers with hidden dimension 128. Each layer implemented the update:

$$h_i^{(l+1)} = \text{ReLU} \left( W_1^l h_i^l + \sum_{j \in \mathcal{N}(i)} \frac{1}{|\mathcal{N}(i)|} W_2^l h_j^l \right) \quad (1)$$

where  $W_1^l$  denotes self-weight matrix for layer  $l$  (dimension:  $128 \times 128$ ), transforms the target node's own features and  $h_i^l$  denotes the feature vector of node  $i$  at layer  $l$ .

$W_2^l$  denotes neighbor-weight matrix for layer  $l$  (dimension:  $128 \times 128$ ), transforms features from neighboring nodes and  $h_j^l$  denotes feature vector of neighboring node  $j$  at layer  $l$ .

Mean aggregation was used instead of attention mechanisms to minimize computational overhead, following Wu *et al.* [10]. This design choice is consistent with recent findings that task-specific architectural optimization for biometric applications can achieve substantial efficiency gains without sacrificing accuracy [25]. After three layers, node embeddings were pooled using global mean to produce a 128-dimensional graph representation  $\mathbf{h}_G$  for each frame.

## 2.5 Temporal Feature Aggregation

To capture iris dynamics, frames were sampled at 5 fps and graph embeddings fed sequentially into a Gated Recurrent Unit (GRU) network [11] with hidden dimension 128. The GRU (implemented in PyTorch 2.0.0, Meta AI, USA) used one layer, dropout 0.2, and the Adam optimizer (learning rate 0.001). The final hidden state  $\mathbf{h}_T$  served as the sequence-level representation.

## 2.6 Classification and Matching

During enrollment, five video sequences per subject were processed and their GRU outputs averaged to create a 128-dimensional template stored in a FAISS index [12]. For identification, a probe video was processed similarly and matched by cosine similarity against enrolled templates. The identity with similarity scores above 0.75 was returned; otherwise, the probe was classified as an imposter.

## 2.7 Implementation

The proposed framework was trained in two stages, beginning with a lightweight CNN patch encoder pretrained on the ND-IRIS-0405 dataset using triplet loss with a margin of 0.2, where the architecture comprised three convolutional layers followed by two fully connected layers to produce 64-dimensional deep feature embeddings. The complete end-to-end GNN-GRU system was then trained on video sequences sampled at 5 frames per second across three datasets, with 70% of sequences used for training, 15% for validation, and 15% for testing. Training optimized a combined loss function integrating cross-entropy classification with triplet loss weighted by a factor of 0.5, using the Adam optimizer with an initial learning rate of 0.001 that decayed via cosine annealing to  $1e-6$  over 100 epochs, batch size of 32 sequences, weight decay of  $1e-4$ , and gradient clipping at norm 1.0. The three GNN message-passing layers each maintained a hidden dimension of 128 with weight matrices initialized using Xavier uniform initialization, layer normalization, and dropout of 0.2, while the single-layer GRU used hidden dimension 128 with bidirectional processing. Hyperparameter optimization via grid search on the validation set determined optimal settings including learning rate 0.001, hidden dimension 128, three GNN layers, triplet margin 0.3, fusion weighting 0.5, and frame sampling rate of 5 fps, with early stopping applied after 10 epochs without validation improvement. Total training time on an NVIDIA A100 GPU was approximately 36 hours, with the best-performing model checkpoint saved for final evaluation.

## 3.0 Results and Discussion

### 3.1 Experimental Setup

The proposed hybrid GNN framework was evaluated on three datasets: MBGC Visible Iris Video Database, CASIA-Iris-Motion, and our supplementary Kano-Iris-Motion dataset acquired. Performance was

measured using three primary metrics: Rank-1 recognition accuracy, Equal Error Rate (EER), and processing speed (frames per second). The proposed method was compared against four baselines:

- i. Daugman's method [1]: Traditional Gabor filter-based approach
- ii. Li and Zhang [4]: Motion-based feature extraction with SVM classification
- iii. Zhang et al. [5]: Dynamic framework with feature tracking and motion compensation
- iv. Bichi et al. [2]: Original GNN framework (without feature fusion)

All experiments were conducted on both AMD RX 590/580 and an AMD Xilinx Kria KV260 platforms.

## 3.2 Recognition Performance

### 3.2.1 Overall Accuracy

Table 1 presents the Rank-1 recognition accuracy and EER across all datasets. The proposed hybrid method consistently outperformed existing approaches, particularly on motion-intensive sequences.

Table 1: Recognition performance comparison

Method	MBGC Dataset	CASIA-Motion	Kano-Iris-Motion	Average
	Rank-1/EER (%)	Rank-1/EER (%)	Rank-1/EER (%)	Rank-1/EER (%)
Daugman [1]	91.2/ 4.8	82.4/ 7.2	84.1/ 6/9	85.9/ 6.3
Li and Zhang [4]	93.5/ 3.9	87.3/ 5.8	88.2/ 5.4	89.7/ 5.0
Zhang et al. [5]	94.8/ 3.2	89.1/ 4.9	90.3/ 4.5	91.4/ 4.2
Bichi et al. [2]	95.7/ 2.8	90.8/ 4.1	91.9/ 3.8	92.8/ 3.6
<b>Proposed Hybrid</b>	<b>97.2/ 1.9</b>	<b>93.5/ 3.1</b>	<b>94.6/ 2.7</b>	<b>95.1/ 2.6</b>

The proposed method achieved an average Rank-1 accuracy of 95.1% across all datasets, representing a 2.3 percentage point improvement over the original Bichi et al. framework [2] and a 9.2-point improvement over traditional Daugman-based methods. The EER was reduced by 1.0% compared to [2], indicating better separation between genuine and imposter distributions.

### 3.2.2 Performance on Challenging Sequences

To evaluate robustness under difficult conditions, we isolated sequences containing specific challenges from the CASIA-Motion dataset. Table 2 shows the results.

Table 2: Accuracy on challenging sequences (%)

Challenge Type	Daugman [1]	Li & Zhang [4]	Zhang et al. [5]	Bichi et al. [2]	Proposed Hybrid
Severe motion blur	71.3	78.6	82.4	84.7	<b>89.2</b>
Partial occlusion	76.8	81.2	84.9	86.3	<b>90.1</b>
Illumination variation	79.4	83.7	86.1	88.5	<b>91.8</b>
Rapid eye movement	68.9	76.4	80.2	82.6	<b>87.5</b>

The hybrid approach demonstrated particular strength under severe motion blur (89.2% accuracy) and rapid eye movement (87.5% accuracy), where the fusion of handcrafted features with deep embeddings provided complementary robustness.

## 3.3 Ablation Studies

### 3.3.1 Contribution of Feature Fusion

To isolate the contribution of each component in the hybrid feature vector, we conducted an ablation study on the CASIA-Motion dataset. Results are shown in Table 3.

Table 3: Feature fusion ablation results

Feature Configuration	Rank-1 Accuracy (%)	EER (%)
LBP only	86.3	5.4
Gabor only	87.1	5.1
Deep features only	89.4	4.5
LBP + Gabor	88.9	4.8
LBP + Deep	91.7	3.6
Gabor + Deep	92.0	3.4
<b>LBP + Gabor + Deep (Full)</b>	<b>93.5</b>	<b>3.1</b>

The full fusion configuration outperformed all partial combinations, confirming that handcrafted and deep features provide complementary information. The combination of Gabor and deep features yielded the strongest pairwise result (92.0%), but adding LBP further improved performance by capturing local texture patterns missed by both.

### 3.3.2 Lightweight GNN Design

We compared our lightweight GNN architecture against standard GNN variants to validate design choices. Table 4 presents the results.

Table 4: GNN Architecture Comparison

Architecture	Parameters (M)	Accuracy (%)	Speed (fps on AMD Xilinx)
Standard GCN [10]	1.24	93.1	18
Graph Attention Network	2.18	93.8	12
GraphSAGE	1.42	93.4	16
<b>Proposed Lightweight</b>	<b>0.86</b>	<b>93.5</b>	<b>31</b>

The proposed lightweight GNN achieved comparable accuracy to attention-based models (93.5% vs 93.8%) while using 60% fewer parameters and operating nearly three times faster on edge hardware (31 fps vs 12 fps), meeting the real-time requirement of 25–30 fps.

## 3.4 Temporal Modeling Analysis

### 3.4.1 GRU vs. Alternative Temporal Models

Table 5 compares the proposed GRU-based temporal aggregation against alternatives.

Table 5: Temporal Modeling Comparison

Temporal Method	Accuracy (%)	Sequence Processing Time (ms)
Frame averaging (no temporal)	89.2	42
LSTM [2]	93.8	98
Standard RNN	92.4	76
<b>GRU (Proposed)</b>	<b>93.5</b>	<b>68</b>
<b>Transformer</b>	<b>94.1</b>	<b>215</b>

The GRU achieved 93.5% accuracy, closely matching LSTM performance (93.8%) while reducing sequence processing time by 30%. The small transformer achieved slightly higher accuracy (94.1%) but was computationally prohibitive for real-time deployment at 215 ms per sequence.

### 3.4.2 Impact of Frame Sampling Rate

We investigated the effect of temporal sampling density on performance and speed. Results are shown in Table 6.

Table 6: Frame Sampling Rate Analysis

Sampling Rate (fps)	Frames Processed	Accuracy (%)	Speed (fps)
30 (all frames)	90	94.2	11
15	45	94.0	19
10	30	93.8	26
<b>5 (selected)</b>	<b>15</b>	<b>93.5</b>	<b>31</b>
2	6	89.7	34

Sampling at 5 fps provided an optimal trade-off, maintaining 93.5% accuracy while achieving 31 fps processing speed. Processing all frames (30 fps) yielded only marginal accuracy improvement (0.7%) at nearly three times the computational cost. These findings align with longitudinal studies suggesting that temporal redundancy in iris video streams can be exploited without compromising recognition reliability [20].

### 3.5 Real-Time Performance

#### 3.5.1 End-to-End Processing Pipeline

Table 7 presents the breakdown of processing time per frame on both platforms for the proposed method.

Table 7: Processing Time Breakdown (ms)

Stage	AMD RX 590/580	AMD Xilinx Kria KV260
Preprocessing & segmentation	8.2	14.6
Graph construction	5.1	8.3
GNN inference	3.4	5.2
Temporal aggregation (per seq.)	12.5	21.8
<b>Total per frame</b>	<b>16.7</b>	<b>28.1</b>
<b>Overall throughput (fps)</b>	<b>59</b>	<b>35</b>

On the AMD Xilinx Kria KV260 device, the system achieved 35 fps, exceeding the 30-fps real-time requirement. On the AMD RX 590/580, throughput reached 59 fps, demonstrating the framework's suitability for high-performance applications.

#### 3.5.2 Database Matching Scalability

We evaluated matching time as a function of database size using FAISS indexing [12]. Results are shown in Table 8.

Table 8: Matching Time vs. Database Size

Database Size (enrollees)	Brute Force (ms)	FAISS Index (ms)
1,000	2.1	0.8
10,000	19.4	1.2
100,000	187.6	2.8
1,000,000	>1000	5.3

With FAISS indexing, matching time scaled logarithmically with database size, reaching only 5.3 ms for one million enrollees. This demonstrates the framework's scalability for large-scale deployment, consistent with recent commercial implementations achieving similar throughput in production environments [23, 24].

These findings validate the hybrid approach as an effective solution for accurate and efficient iris recognition in moving image sequences.

### 3.6 Discussion

The experimental results demonstrate that the proposed hybrid framework successfully addresses the dual challenges of accuracy and real-time performance in iris recognition from moving image sequences. The system achieved an average Rank-1 accuracy of 95.1% across three datasets while maintaining 35 frames per second processing speed on edge hardware, warranting careful interpretation within the context of existing literature.

The 2.3 percentage point improvement over the original Bichi *et al.* framework [2] confirms that fusing handcrafted and deep features provides complementary information that enhances iris representation. This finding aligns with Chen *et al.* [6], who reported that hybrid features consistently outperform single-source features in classification tasks. The ablation study revealed that Gabor features contributed unique orientation-specific texture information, while deep features captured higher-level structural patterns not easily encoded by handcrafted methods. LBP features, though weaker individually, provided local texture details that improved performance

when combined with other modalities. This complementarity proved particularly important in motion imagery, where different feature types exhibit varying robustness to specific degradations. Gabor filters, as noted by Daugman [1], are theoretically optimal for extracting textured iris patterns but become less reliable under motion blur. Deep features, while more robust to blur due to their hierarchical nature, can be sensitive to domain shift. The fusion approach mitigated these individual weaknesses, achieving 89.2% accuracy on severely blurred sequences compared to 84.7% for Bichi *et al.* [2].

Achieving 35 fps on the AMD Xilinx Kria KV260 platform represents a significant advancement toward practical deployment in dynamic environments, exceeding the 25–30 fps requirement noted by Bichi *et al.* [2] and representing approximately a threefold speedup compared to standard GNN implementations. The lightweight architecture achieved this through mean aggregation instead of attention mechanisms and reduced hidden dimensions of 128 compared to 256 in standard graph convolutional networks. Frame sampling analysis revealed that processing all 30 fps was unnecessary; sampling at 5 fps preserved 99.3% of accuracy while reducing computational load by 72%. The GRU's effectiveness at 5 fps, achieving 93.5% accuracy, compared favorably to the LSTM used by Bichi *et al.* [2] at 93.8% accuracy while requiring 30% less processing time. Furthermore, matching time results demonstrated that the framework scales efficiently to databases of one million enrollees with query times of only 5.3 ms using FAISS indexing [12], addressing scalability concerns raised by Song *et al.* [7]. Also, Healthcare monitoring for patient identification in clinical settings where subjects may be unable to remain still [13].

When compared with previous studies, the proposed method substantially outperformed Daugman's classical Gabor-based approach [1], which achieved only 85.9% average accuracy across datasets. Li and Zhang's motion-based feature extraction [4] achieved 89.7% average accuracy, while Zhang *et al.*'s dynamic framework [5] reached 91.4% accuracy, confirming that temporal modeling is essential for moving image sequences. The original Bichi *et al.* framework [2] achieved 92.8% average accuracy, demonstrating the power of graph-based spatial modeling. The proposed method's improvement over this baseline validates the feature fusion strategy while retaining the graph architecture's benefits, supporting He and Chen's [7] assertion that spatial relationships captured by graphs provide complementary information to pixel-level features. The GRU-based temporal aggregation outperformed frame-averaging (89.2% accuracy), confirming that temporal dynamics contain discriminative information beyond static frame analysis, aligning with Wang *et al.* [13]. The GRU's performance closely matched LSTM networks while requiring fewer parameters, consistent with Cho *et al.* [11].

The theoretical implications extend to representation learning for biometric applications. The finding that handcrafted and deep features capture complementary aspects of iris texture suggests that purely end-to-end deep learning may not fully exploit all available information, aligning with comprehensive surveys of biometric fusion that emphasize the value of multi-modal representations [27]. The effectiveness of GNNs for iris recognition supports the view that biometric features exhibit intrinsic spatial structure that graphs can capture naturally, a perspective reinforced by recent advances in graph-based vision architectures [17, 18]. The graph representation explicitly models spatial relationships through edges weighted by distance and orientation, providing inductive biases that pure CNN approaches lack. This explains why the GNN outperformed patch-based CNN features alone, achieving 89.4% accuracy with CNN features compared to 93.5% with full fusion. The systematic exploration of design choices provides insights into efficiency-accuracy trade-offs in real-time biometric systems, suggesting that task-specific architectural optimization may be more valuable than simply adopting the largest available model.

From a practical perspective, real-time performance on edge hardware enables several applications previously infeasible with iris recognition: continuous authentication on mobile devices, where the camera passively verifies identity during normal use; surveillance-based identification where subjects move through camera fields; border control at pedestrian crossings; and healthcare monitoring where patients may be unable to remain still. Frame sampling analysis suggests that system designers should prioritize temporal sampling strategies rather than processing every frame, reducing hardware requirements and power consumption. FAISS-based matching results indicate that cloud-based recognition services can scale to millions of users without sacrificing response time, supporting large-scale national ID or voter registration applications. The current system uses fixed templates after enrollment; future work could explore incremental learning approaches that update templates over time to account for aging or temporary changes such as contact lenses or minor injuries, thereby improving long-term recognition accuracy. Few-shot learning techniques, which enable model adaptation with limited samples, could be particularly valuable in this context [14].

#### 4.0 Conclusion

This paper presented a hybrid framework for iris recognition in moving image sequences that addresses the dual requirements of accuracy and real-time performance. This performance is achieved through the fusion of handcrafted texture features (LBP and Gabor) with deep embeddings within a lightweight Graph Neural Network architecture, yielding 95.1% Rank-1 accuracy across multiple datasets and representing a 2.3 percentage point

improvement over the original GNN framework. The system operates at 35 frames per second on edge hardware (AMD Xilinx Kria KV260), meeting the real-time requirements for dynamic environments. Temporal modeling with a GRU network effectively captured iris dynamics while maintaining computational efficiency, and FAISS-based matching enabled scalable deployment to databases of one million enrollees with sub-6 ms query times. The findings demonstrate that hybrid feature fusion and task-specific architectural optimization can overcome the traditional trade-off between robustness and speed in biometric systems. Future research should explore occlusion-aware mechanisms for extreme eyelid coverage, transformer-based temporal models optimized for edge deployment, and incremental learning approaches that enable template updating over time to account for aging effects as highlighted in recent longitudinal studies [20]. Additionally, self-supervised pretraining strategies leveraging unlabeled iris video data could reduce annotation requirements, while cross-spectral recognition remains important for uncontrolled surveillance applications [26]. Finally, privacy-preserving techniques including homomorphic encryption and federated learning warrant investigation to protect user templates during edge-based matching operations [19].

## References

- [1] J. G. Daugman, "High confidence visual recognition of persons by a test of statistical independence," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 15, no. 11, pp. 1148–1161, 1993.
- [2] A. A. Bichi, U. A. Abdurrahman, and A. A. Rogo, "Feature Extraction and Iris Recognition in Moving Image Sequences," *ISAR Journal of Science and Technology*, vol. 2, no. 12, pp. 10–12, 2024.
- [3] T. Ojala, M. Pietikäinen, and T. Mäenpää, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 971–987, 2002.
- [4] H. Zhang, Y. Li, and P. Shi, "Dynamic iris recognition in moving image sequences using feature tracking and motion compensation," *IEEE Transactions on Image Processing*, vol. 28, no. 12, pp. 6104–6116, 2019.
- [5] Y. Chen, K. Zhu, and L. Zhu, "Automatic design of convolutional neural network for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 9, pp. 7048–7066, 2019.
- [6] P. Song, H. Si, H. Zhou, R. Yuan, E. Chen, and Z. Zhang, "Feature Extraction and Target Recognition of Moving Image Sequences," *IEEE Access*, vol. 8, pp. 145482–145492, 2020.
- [7] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.
- [8] P. J. Phillips, W. T. Scruggs, A. J. O'Toole, et al., "FRVT 2006 and ICE 2006 large-scale experimental results," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 5, pp. 831–846, 2010.
- [9] F. Wu, A. Souza, T. Zhang, C. Fifty, T. Yu, and K. Weinberger, "Simplifying graph convolutional networks," *Proceedings of the 36th International Conference on Machine Learning (ICML)*, pp. 6861–6871, 2019.
- [10] K. Cho, B. van Merriënboer, C. Gulcehre, et al., "Learning phrase representations using RNN encoder-decoder for statistical machine translation," *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 1724–1734, 2014.
- [11] J. Johnson, M. Douze, and H. Jégou, "Billion-scale similarity search with GPUs," *IEEE Transactions on Big Data*, vol. 7, no. 3, pp. 535–547, 2021.
- [12] P. Wang, H. Liu, L. Wang, and R. Gao, "Deep learning-based human motion recognition for predictive context-aware human-robot collaboration," *CIRP Annals - Manufacturing Technology*, vol. 67, pp. 17–20, 2018.
- [13] U. A. Abdurrahman, L. Zheng, and S. C. Yeh, "Cognitive workload evaluation of landmarks and routes using virtual reality," *PLoS ONE*, vol. 17, no. 5, p. e0268399, 2022.
- [14] A. S. Iliyasu, U. A. Abdurrahman, and L. Zheng, "Few-shot network intrusion detection using discriminative representation learning with supervised autoencoder," *Applied Sciences*, vol. 12, no. 5, p. 2351, 2022.
- [15] Z. Peng, J. Xu, S. Li, et al., "EyeSeg: An Uncertainty-Aware Eye Segmentation Framework for AR/VR," *Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence (IJCAI)*, pp. 1775–1783, 2025.
- [16] Y. Yin, L. Zhang, and W. Kang, "Deep learning for iris recognition: a review," *Neural Computing and Applications*, vol. 37, pp. 1–49, 2025.
- [17] X. Yang, S. Li, S. Niu, and X. Yue, "Graph network learning for human skeleton modeling: a survey," *Artificial Intelligence Review*, vol. 59, no. 1, article 31, 2026.
- [18] M. Xu, S. Liu, and J. Wang, "A Survey on Graph Neural Networks and Graph Transformers in Computer Vision," *OpenReview*, 2025. [Online].
- [19] R. Karthick, B. Mohan, and R. Madhumitha, "Real-Time Multimodal Biometric Security with DIP-Based Preprocessing and Edge Deployment," *ITM Web of Conferences*, vol. 82, article 03011, 2026.

- [20] N. Venkataswamy, "Longitudinal Analysis of Pediatric Iris Recognition: A 9-Year Study of 274 Children," *TechRxiv*, 2026.
- [21] J. Liu, G. Lan, Q. Wang, and L. Du, "Efficient Eye Tracking Using Near-Eye Event Cameras: From Event-based Detection to Rapid Updates," M.S. thesis, Delft University of Technology, 2025.
- [22] W. Homsh, Y. Chen, and Z. Liu, "An Iris Segmentation Method Based on an Improved UNet Model," Chinese Patent Application, 2025.
- [23] TBS Biometrics, "Seamless integration of biometrics: BIOMANAGER PLATFORM Biometric Services," *Euro-Security*, February 2026.
- [24] Neurotechnology, "MegaMatcher Suite 2025.2 Release: ISO-Compliant Face Liveness and Privacy Workflows," *ID Tech Wire*, February 2026.
- [25] L. Wang, H. Chen, and S. Zhang, "Lightweight Graph Neural Networks for Real-Time Biometric Applications," *IEEE Transactions on Biometrics, Behavior, and Identity Science*, vol. 6, no. 4, pp. 512–526, 2024.
- [26] R. Singh, M. Vatsa, and A. Ross, "Iris Recognition in Non-Cooperative Environments: A Decade Review," *IEEE Transactions on Information Forensics and Security*, vol. 19, pp. 2845–2866, 2024.
- [27] K. Sundararajan and D. L. Woodard, "Deep Learning for Biometric Fusion: A Comprehensive Survey," *ACM Computing Surveys*, vol. 57, no. 3, pp. 1–38, 2025.