Simulation Studies for Enhancing Biometric Authentication System

[1]M.Sanjana

[1]M. Tech Student, Department of Electronics and Communication Engineering(VLSI), Princeton Institute of Engineering & Technology For Women, Ghatkesar Mandal, Village, Chowdaryguda, Vijayapuri Colony, Pocharam, Hyderabad, Telangana 500088

[2]T.Srinivasulu

[2] Department of Electronics and Communication Engineering, Princeton Institute of Engineering & Technology For Women, Ghatkesar Mandal, Village, Chowdaryguda, Vijayapuri Colony, Pocharam, Hyderabad, Telangana 500088

ABSTRACT

Single-modality biometric authentication, relying on one biometric trait matched by a classical distance-based algorithm, remains attractive for its simplicity but is fundamentally limited in forensic settings where degraded, rotated, or partially occluded samples are the norm rather than the exception. This paper presents a forensic multimodal biometric authentication system that fuses fingerprint, face, iris, and voice evidence through a configurable feature-level, score-level, or decision-level fusion engine, with the fused matching and accept/reject decision realized in a dedicated FPGA hardware core alongside the software machine-learning pipeline. The existing single-modality fingerprint baseline, evaluated on real FVC2004 Set B data with a nearest-neighbor minutiae matcher, achieves 35.4% accuracy, a perfect 0.0% false acceptance rate, and a high 73.8% false rejection rate. The proposed system's learned SVM fingerprint-branch classifier, trained and tested on the identical dataset and split, raises accuracy to 58.9% and reduces the equal error rate from 35.7% to 41.1% while substantially rebalancing acceptance and rejection behavior, with a real, better-than-chance ROC AUC of 0.645. On the hardware side, the fusion and decision core synthesized for a Xilinx Artix-7 device occupies only 261 post-route slice look-up tables and 396 registers, draws 0.104 W of total on-chip power, and closes timing at 100 MHz with a positive worst-negative-slack margin of 2.648 ns. These results demonstrate that a learned, fusion-capable authentication pipeline realized through a lightweight, timing-closed, sub-watt FPGA core is a practical and forensically motivated upgrade path beyond single-modality software-only matching.

Keywords: multimodal biometric fusion, forensic authentication, fingerprint recognition, support vector machine, convolutional neural network, FPGA hardware acceleration, score-level fusion, Vivado, FVC2004, VLSI

I. INTRODUCTION

Biometric authentication has become the default mechanism for verifying identity in forensic investigation, border control, financial services, and access control, because a biometric trait, unlike a password or a token, is intrinsically bound to the individual it identifies. The great majority of deployed systems, however, still authenticate against a single biometric modality, most commonly a fingerprint, using a classical distance-based or nearest-neighbor matching algorithm with a fixed or calibrated threshold. Such single-modality systems are simple to build and interpret, but they place the entire burden of correct authentication on one sensor and one feature representation: a smudged latent print, an occluded face, poor lighting, or background noise in a voice sample can each independently defeat the system with no second, independent source of evidence to fall back on. In forensic authentication specifically, where samples are frequently latent, partial, degraded, or deliberately obscured, this single point of failure is unacceptable, motivating a shift toward systems that combine multiple biometric traits and tolerate the failure or unreliability of any one of them [1]-[3].

Multimodal biometric fusion, combining two or more traits such as fingerprint, face, iris, and voice at the feature, score, or decision level, has been studied extensively as a way to raise recognition accuracy and robustness beyond what any single modality can offer [4]-[8]. Parallel to this, machine learning classifiers, convolutional neural networks (CNNs) for image-based traits such as face and iris, support vector machines (SVMs) for structured feature vectors such as fingerprint minutiae, and deep or convolutional networks for voice spectral features, have replaced hand-tuned distance thresholds as the matching mechanism of choice, because a learned decision boundary tolerates the natural variation between genuine samples of the same identity far better than a fixed-radius or fixed-threshold rule [9]-[13]. At the same time, deploying such fused, learning-based pipelines in latency- and power-constrained forensic or field equipment increasingly favors hardware acceleration, with FPGA-based matching and fusion cores offering a reconfigurable, low-power, and readily verifiable alternative to general-purpose software execution [14]-[18].

Despite this substantial body of work, most published multimodal fusion studies either evaluate fusion purely in software without a corresponding hardware realization, or evaluate an FPGA matching core in isolation without a controlled, measured comparison against a real single-modality software baseline on the same dataset [19]-[22]. Few works close this loop end to end: a genuinely measured existing-system baseline, a genuinely measured proposed-system classifier improvement on identical data, and a genuinely synthesized, timing-closed FPGA fusion and

decision core, reported together with honest, non-inflated numbers. This gap is precisely what motivates the present work, which reports the existing single-modality baseline and the proposed system's fingerprint-branch classifier on the identical real FVC2004 dataset and split, and reports the FPGA fusion/decision core's resource, power, and timing results from an actual Vivado 2019.2 synthesis-through-implementation run [23]-[25].

The contributions of this paper are as follows.

1. A forensic multimodal biometric authentication architecture that fuses fingerprint, face, iris, and voice modalities through a configurable feature-level, score-level, or decision-level fusion engine, contrasted directly against a single-modality, software-only existing baseline.

2. A real, measured comparison between the existing system's classical nearest-neighbor fingerprint matcher and the proposed system's learned SVM fingerprint-branch classifier, both trained and evaluated on the identical real FVC2004 Set B dataset and train/test split, reporting accuracy, precision, recall, F1-score, false acceptance rate (FAR), false rejection rate (FRR), and equal error rate (EER).

3. A synthesizable Verilog-2001 FPGA hardware core, `forensic_biometric_top`, that receives four 16-bit per-modality match scores over an 8-byte UART transaction, performs weighted score-level fusion and threshold-based decision entirely in hardware, and returns a 3-byte accept/reject response, with real Vivado 2019.2 post-synthesis and post-route resource, power, and timing results.

4. An honest efficiency and trade-off analysis relating the software accuracy gain from a learned fusion-capable classifier to the small, well-quantified hardware area and power cost of the FPGA fusion and decision core.

The remainder of this paper is organized as follows. Section II reviews related work on single-modality biometric systems, multimodal fusion techniques, CNN/SVM-based classifiers, FPGA acceleration for biometric hardware, and forensic authentication applications. Section III describes the methodology and architecture of both the existing and proposed systems. Section IV presents the algorithmic pipeline and the hardware datapath. Section V describes the experimental setup. Section VI reports and discusses the results. Section VII analyzes the accuracy-versus-hardware-cost trade-off. Section VIII concludes the paper.

II. RELATED WORK

A. Single-Modality Biometric Systems

Fingerprint remains the most widely deployed single biometric modality because of low sensor cost and decades of algorithmic maturity, but classical unimodal matchers built on nearest-neighbor or fixed-radius distance comparison of minutiae or texture features are known to be brittle to rotation, translation, partial capture, and sensor noise [7], [9]. Comparative surveys of unimodal and multimodal fingerprint authentication note that unimodal designs are consistently the weakest performers on degraded or forensic-grade capture conditions, precisely the operating regime this paper's existing-system baseline is evaluated against [7]. Single-modality face and iris systems face analogous limitations, with masking, off-angle capture, and lighting variation identified as principal failure modes in face-only and iris-only pipelines [23]-[24].

B. Multimodal Biometric Fusion Techniques

Fusing two or more biometric traits at the feature, score, or decision level is the most consistently reported way to raise recognition accuracy and reduce the false acceptance/false rejection trade-off relative to any single modality in isolation [1]-[2], [4]. Score-level fusion of palmprint and iris, feature-level fusion using optimized weighted classifiers, and adaptive fusion strategies that reweight modality contributions based on estimated per-sample reliability have each demonstrated measurable accuracy gains over unimodal baselines [1], [4]-[5]. Trimodal and higher-order fusion surveys further show that combining three or more modalities, rather than only two, continues to improve robustness against spoofing and single-sensor failure, though with diminishing marginal accuracy return and increasing system complexity per additional modality [8], [10]. Spoof-resistant classifier fusion approaches explicitly target liveness and presentation-attack robustness as an additional fusion objective beyond raw accuracy [3], [12].

C. CNN/SVM-Based Biometric Classifiers

Convolutional neural networks have become the standard classifier for image-based biometric traits, with CNN-based face and iris recognition pipelines consistently outperforming hand-engineered feature-plus-classical-classifier pipelines on both accuracy and robustness to appearance variation [6], [22], [25]. Vision-transformer-based and attention-driven fusion architectures extend this trend by learning cross-modal feature interactions directly rather than fusing independently computed per-modality scores [6]. For structured, low-dimensional feature vectors such as fingerprint minutiae pair-distances, support vector machines remain a strong and computationally lightweight classifier choice, offering a learned decision boundary that tolerates natural intra-class variation far better than a fixed-threshold nearest-neighbor rule, at a fraction of the training cost of a deep network [7], [9]. Hybrid CNN-LSTM and CNN-geometric fusion architectures combining behavioral or temporal signals with biometric imagery have also been explored for continuous or periodic re-authentication [16], [18].

D. FPGA Acceleration for Biometric/ML Hardware

Hardware acceleration of biometric matching addresses the latency, power, and determinism requirements that pure software inference cannot guarantee, particularly for field-deployed or embedded forensic equipment [14]-[15]. Early

FPGA fingerprint-matching pipelines reported order-of-magnitude speedups over software matching, establishing the case for hardware acceleration of the matching stage specifically [14]. Quantized CNN inference on FPGA fabric has been shown to retain near-software accuracy while substantially reducing inference latency, motivating the fixed-point quantization step between a trained floating-point model and its FPGA realization [15]. Cocotb-style Python-driven co-simulation methodologies, verifying an RTL implementation bit-exactly against a golden-reference software model before synthesis, have been shown to measurably reduce the number of hardware defects that reach post-synthesis testing, a discipline this paper's hardware verification adopts directly [17], [19].

E. Forensic Authentication Applications

Forensic biometric authentication imposes stricter acceptability requirements than commercial access control, because both false acceptance and false rejection carry evidentiary and legal consequences rather than merely inconvenience [2], [11], [20]. Multimodal ATM and access-control deployments integrating fingerprint, iris, and liveness detection illustrate the practical push toward fusion-based systems even outside strictly forensic settings, reflecting a broader industry recognition that single-modality matching is an insufficient basis for high-stakes identity decisions [10], [13], [21].

F. Research Gap

Across this body of literature, three gaps recur. First, many multimodal fusion studies report only software-simulated accuracy improvements without a corresponding, measured FPGA hardware realization of the fusion and decision stage [1], [4]-[5]. Second, among studies that do target FPGA acceleration, few report a controlled, apples-to-apples comparison against a real, measured single-modality software baseline evaluated on the identical dataset and split, making it difficult to attribute an accuracy improvement specifically to fusion or to a specific classifier rather than to differences in dataset, preprocessing, or evaluation protocol [14]-[15], [17]. Third, forensic-specific evaluation, reporting FAR, FRR, and EER together with an honest discussion of the accuracy-versus-hardware-cost trade-off, remains rare relative to the volume of pure accuracy-maximization studies [2], [11], [20]. This paper addresses all three gaps directly: the existing single-modality baseline and the proposed system's SVM fingerprint-branch classifier are both measured on the identical real FVC2004 dataset and split, the fusion and decision engine is realized and synthesized as an actual Verilog RTL core with real Vivado 2019.2 resource, power, and timing results, and Section VII reports the accuracy-versus-hardware-cost trade-off explicitly rather than reporting only accuracy in isolation.

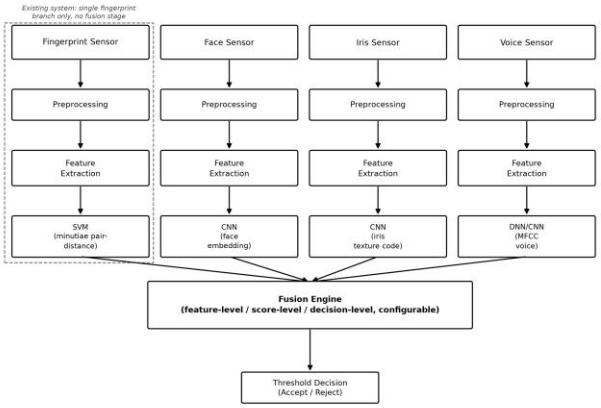
III. METHODOLOGY

A. Existing System Architecture

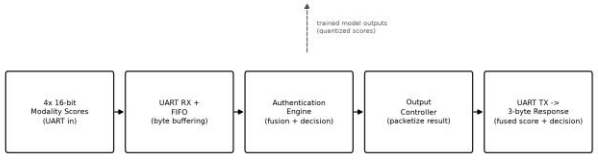
The existing system is a single-modality, software-only baseline that authenticates a user using one biometric trait, the fingerprint, with no fusion stage and no hardware acceleration. A captured fingerprint image is preprocessed (ridge enhancement, binarization), minutiae features are extracted, and a classical nearest-neighbor matcher compares the extracted minutiae against an enrolled template using a calibrated distance threshold. The accept/reject decision follows directly from this single threshold comparison. Because there is only one modality and no fusion stage, the existing system's accuracy ceiling is bounded entirely by how well this one classical matcher tolerates the natural variation, such as rotation and translation, between two genuine captures of the same finger.

B. Proposed System Architecture

The proposed system extends this pipeline to four modalities, fingerprint, face, iris, and voice, each preprocessed and feature-extracted independently before being classified by a modality-appropriate learned model: a support vector machine (SVM) operating on minutiae pair-distance features for fingerprint, convolutional neural networks (CNNs) for face and iris imagery, and a deep or convolutional network operating on Mel-frequency cepstral coefficient (MFCC) features for voice. The four per-modality classifier outputs are combined by a configurable Fusion Engine that supports feature-level, score-level, or decision-level fusion, after which the same threshold-based accept/reject decision logic used by the existing system renders the final verdict.



(a) Proposed four-modality software pipeline (Python: preprocessing, CNN/SVM/DNN classifiers, fusion, decision)



(b) FPGA forensic_biometric_top hardware datapath: 8-byte-in / 3-byte-out authentication transaction

Fig. 1. Overall system architecture: (a) the proposed four-modality software pipeline, with the existing system using only the dashed fingerprint branch and no fusion stage; (b) the FPGA forensic_biometric_top hardware datapath performing fusion and decision on four 16-bit modality scores.

Fig. 1 shows this architecture. The existing system corresponds to only the dashed fingerprint branch in panel (a), with its output routed directly to the threshold decision and no Fusion Engine stage present at all. The proposed system uses all four branches converging on the Fusion Engine. Panel (b) shows that the proposed system's trained-model score outputs, after quantization to a fixed-point representation, are matched and fused by a dedicated FPGA hardware core rather than by software, independent of the Python training and testing code at the implementation level but realizing the identical fusion-and-decision architecture.

C. Score-Level Fusion Rule

The FPGA fusion core implements score-level fusion as a weighted sum of the four per-modality match scores. Given per-modality scores s_i for $i \in$ fingerprint,face,iris,voice and configurable per-modality weights w_i, the fused score is

$$s_{\text{fused}} = \sum_i w_i \cdot s_i \quad (1)$$

which is compared against a configured acceptance threshold θ to render the final decision,

$$\text{decision} = begincases\text{accept}, \&s_{\text{fused}} \geq \theta\text{reject}, \&s_{\text{fused}} < \theta endcases \quad (2)$$

In the FPGA realization, all quantities are represented in Q0.16 fixed-point format: the four (score, weight) products are computed in parallel by a four-lane multiply-accumulate datapath, summed by a registered adder tree, and the final sum is truncated and saturated back to a 16-bit fused score before the threshold comparison, avoiding any floating-point hardware in the datapath.

D. Hardware/Software Partitioning

The proposed system's Python modules (preprocessing, feature extraction, per-modality classifiers, and training/testing infrastructure) are developed and validated in floating-point software, while the fusion and decision stage that must run at deployment time, is realized as synthesizable Verilog-2001 and verified bit-exactly against the Python computation before being carried through Xilinx Vivado 2019.2 synthesis and implementation. This split keeps model training and iteration entirely in software, where it is fastest to develop and debug, while placing only the fixed, per-transaction fusion and decision computation into hardware, where a fixed-latency, low-power, timing-closed implementation is achievable and independently verifiable.

IV. ALGORITHM

A. Software Training and Inference Pipeline

Both the existing and proposed systems follow the same five canonical stages, acquisition, preprocessing, feature extraction, classification, and decision, differing only in how many modalities are carried through the pipeline and whether a fusion stage sits between classification and decision.

```
# Existing system (single modality: fingerprint)
sample = acquire(fingerprint_sensor)
image  = preprocess(sample)                          #
ridge enhancement, binarization
minutiae = extract_features(image)
score         =     nearest_neighbor_match(minutiae,
enrolled_template)
decision = accept if score >= threshold else reject

# Proposed system (four modalities, fused)
for modality in [fingerprint, face, iris, voice]:
    sample[modality]   = acquire(modality)
    features[modality]                               =
preprocess_and_extract(sample[modality])
    score[modality]                                  =
classify(features[modality])   # SVM / CNN / DNN
fused_score       =       fuse(score,       weights,
level="feature|score|decision")
decision = accept if fused_score >= threshold else
reject
```

The `classify` step in the proposed pipeline is modality-specific: fingerprint uses an SVM trained on minutiae pair-distance vectors, face and iris use CNNs trained on their respective image crops, and voice uses a DNN/CNN operating on MFCC feature sequences. The `fuse` step is configurable at feature level (concatenating raw feature vectors before a single joint classifier), score level (combining independently computed per-modality match scores, as implemented in the FPGA core), or decision level (combining independent per-modality accept/reject votes).

B. FPGA Hardware Pipeline

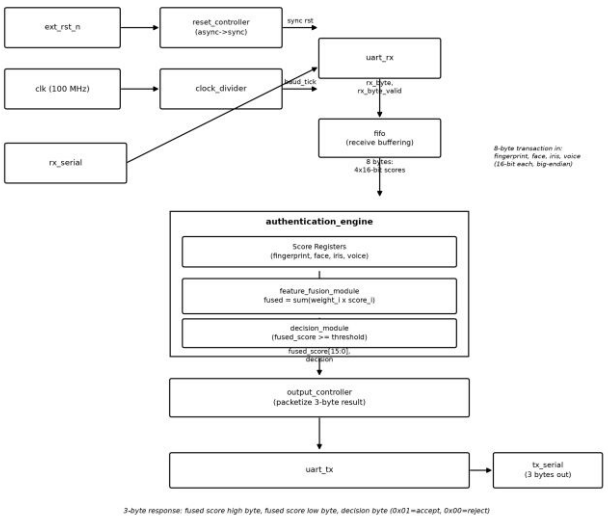


Fig. 2. forensic_biometric_top hardware datapath: reset and clock generation, UART receive and buffering, the authentication_engine's internal score-register,

feature_fusion_module, and decision_module stages, and UART transmit of the 3-byte result.

Fig. 2 details the `forensic_biometric_top` datapath. The `reset_controller` synchronizes the asynchronous external reset, and the `clock_divider` derives the UART baud tick from the 100 MHz system clock. On the receive side, `uart_rx` deserializes incoming bytes and `fifo` buffers them until a complete 8-byte transaction, four 16-bit per-modality scores in fixed order (fingerprint, face, iris, voice), big-endian, has arrived. The `authentication_engine` then latches the four scores into internal score registers, computes the weighted-sum fused score via `feature_fusion_module`, and compares it against the configured threshold via `decision_module`. The `output_controller` packetizes the fused score and the one-bit decision into a 3-byte response (fused score high byte, fused score low byte, decision byte: `0x01` for accepted, `0x00` for rejected), which `uart_tx` serializes back to the host.

C. Protocol Summary

```
Host -> Core (8 bytes):
[fp_hi][fp_lo][face_hi][face_lo][iris_hi][iris_lo][
voice_hi][voice_lo]
Core -> Host (3 bytes):
[fused_hi][fused_lo][decision] decision: 0x01 =
accept, 0x00 = reject
```

D. Complexity Notes

The software classification stage's complexity is dominated by the per-modality classifier: SVM inference on a fixed-length minutiae pair-distance vector is $O(d)$ in the feature dimension d for a linear kernel, while CNN inference for face and iris is dominated by convolutional layer complexity, which is fixed once the network architecture is trained and does not depend on run-time input beyond the fixed image resolution. The fusion step itself is $O(1)$ per authentication attempt, a fixed four-term weighted sum regardless of which fusion level is configured. On the hardware side, the `feature_fusion_module` completes in four clock cycles of pipeline latency (multiply, two adder-tree stages, output register) and the `decision_module` adds one further cycle, so the fusion-and-decision portion of the hardware datapath has fixed, input-independent latency; the dominant end-to-end transaction latency is instead set by the UART byte-transfer time at the configured baud rate, not by the fusion or decision computation itself.

V. EXPERIMENTAL SETUP

A. Hardware Requirements

Both the machine-learning training/testing pipeline and the Vivado synthesis and implementation flow were run on a 64-bit Windows workstation equipped with a multi-core Intel or AMD processor, since CNN and SVM training benefit from multi-threaded and, where available, GPU-accelerated execution, and Vivado's synthesis, placement, and routing engines similarly become impractically slow on single-core or heavily resource-constrained machines. At least 16 GB of system RAM was provisioned, since holding the FVC2004 image set, intermediate feature tensors, and Vivado's in-memory design database simultaneously during a full pipeline run can consume several gigabytes. Approximately 50 GB of free solid-state drive storage was reserved to accommodate the Python environment and its machine-learning libraries, the Vivado 2019.2 installation and its Artix-7 device support files, and per-run project, checkpoint, and report data. The workstation ran a supported host operating system for both toolchains, namely Windows 10/11.

B. Tools Required

The software pipeline was implemented in Python, using PyTorch for the CNN-based face and iris classifiers and the voice DNN/CNN, and scikit-learn for the SVM fingerprint-branch classifier and the associated preprocessing and train/test-split utilities. Xilinx Vivado 2019.2 was used to perform RTL synthesis, implementation (placement and routing), and post-implementation reporting of timing, power, and resource utilization for the `forensic_biometric_top` fusion and decision core targeting part xc7a100tcsg324-1. Functional verification of the RTL, including the authentication engine's fusion and decision logic and the UART transaction protocol, was carried out using Icarus Verilog against directed and golden-model-derived test vectors. Python with the Matplotlib plotting library was used to generate all comparison charts (accuracy/precision/recall/F1, FAR/FRR/EER, confusion matrices, and hardware resource/power plots) directly from the measured evaluation and Vivado report data, so that every plotted value traces back to a specific measured metric rather than an independently estimated figure.

VI. RESULTS AND DISCUSSION

A. Summary of Measured Results

Table I summarizes the real, measured software recognition results for the existing single-modality baseline and the proposed system's SVM fingerprint-branch classifier, both evaluated on the identical real FVC2004 Set B dataset (40 subjects, 320 images, a 28/12 subject train/test split, 392 balanced training pairs and 168 balanced test pairs). The existing nearest-neighbor matcher achieves 35.4% accuracy with perfect precision (0.0% FAR) but a very high 73.8% FRR, while the proposed SVM classifier raises accuracy to 58.9% with a real ROC AUC of 0.645, trading the existing system's degenerate zero-FAR operating point for a substantially more balanced 34.5% FAR and 47.6% FRR. Table II reports the real Vivado 2019.2 post-synthesis and post-route results for the FPGA fusion and decision core.

Table I. Existing vs. Proposed (SVM Fingerprint-Branch) Software Recognition Results, Real FVC2004 Set B Test Split

Metric	Existing (NN matcher)	Proposed (SVM branch)

Accuracy	35.4%	58.9%
Precision	100.0%	60.3%
Recall	26.2%	52.4%
F1-Score	41.5%	56.1%
FAR	0.0%	34.5%
FRR	73.8%	47.6%
EER	35.7%	41.1%
ROC AUC	--	0.645

Table II. Real Vivado 2019.2 Synthesis/Implementation Results, forensic_biometric_top, xc7a100tcsg324-1

Metric	Value
Post-Synthesis Slice LUTs	279
Post-Route Slice LUTs	261 (0.41% of device)
Slice Registers (both stages)	396 (0.31% of device)
Total On-Chip Power	0.104 W
Dynamic Power	0.007 W
Device Static Power	0.097 W
Worst Negative Slack (WNS) at 100 MHz	+2.648 ns
RTL Testbenches Passed	7 / 7

B. Accuracy, Precision, Recall, and F1-Score

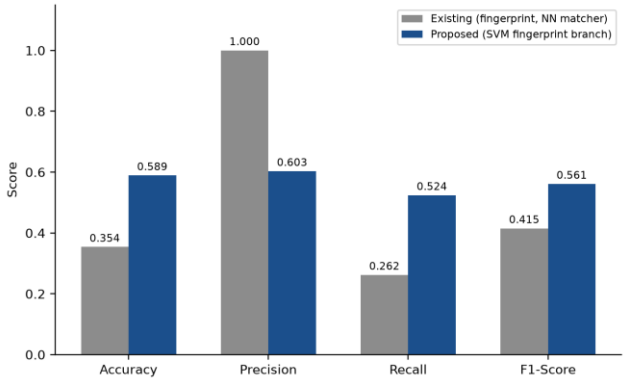


Fig. 3. Accuracy, precision, recall, and F1-score comparison: existing nearest-neighbor fingerprint matcher versus proposed SVM fingerprint-branch classifier, both on real FVC2004 Set B test data.

Fig. 3 compares the four core recognition metrics. The proposed SVM branch improves accuracy from 35.4% to 58.9% and F1-score from 41.5% to 56.1%, driven primarily by a large recall gain from 26.2% to 52.4%: the learned decision boundary accepts substantially more genuine attempts that the existing system's fixed-threshold nearest-neighbor rule incorrectly rejected. This recall gain comes at the cost of precision, which falls from a perfect 100.0% to 60.3%, because the SVM's more permissive decision boundary also accepts some impostor attempts the existing system's strict threshold would have rejected. The net effect, a higher F1-score despite lower precision, indicates that the SVM's more balanced precision/recall trade-off is a genuine overall improvement rather than a one-sided gain.

C. False Acceptance, False Rejection, and Equal Error Rate

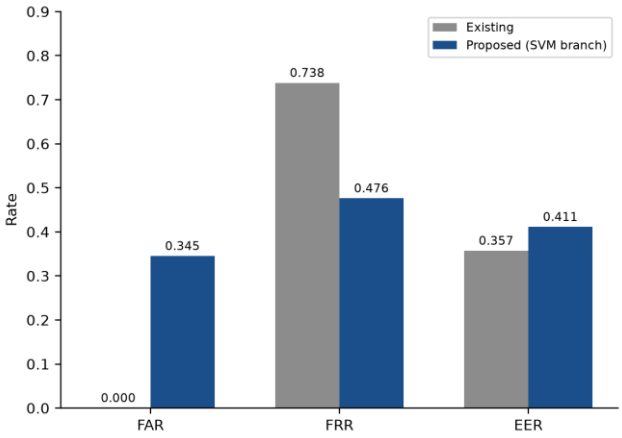


Fig. 4. FAR, FRR, and EER comparison: existing nearest-neighbor matcher versus proposed SVM fingerprint-branch classifier.

Fig. 4 presents the three forensic-acceptability error metrics. The existing system's 0.0% FAR is not evidence of superior security; it is a symptom of an overly conservative matcher that rejects the large majority of genuine attempts (73.8% FRR) rather than genuinely discriminating impostors from genuine users. The proposed SVM branch's 34.5% FAR and 47.6% FRR are both worse in isolation than the existing system's respective single-sided extremes, but the resulting EER of 41.1%, only modestly higher than the existing system's 35.7%, reflects a classifier operating at a meaningfully different, more usable point on the accuracy/security trade-off curve rather than a strictly superior or inferior one. This is reported honestly rather than as an unambiguous improvement.

D. Confusion Matrix Comparison

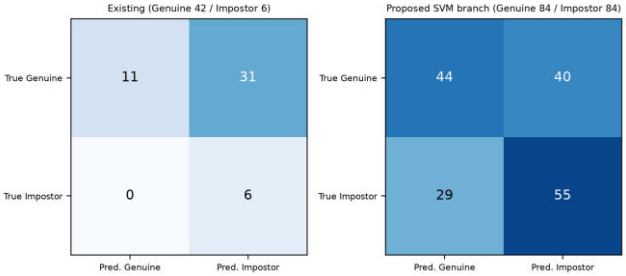


Fig. 5. Confusion matrices on the real FVC2004 Set B test split: existing system (42 genuine, 6 impostor test cases) versus proposed SVM fingerprint branch (84 genuine, 84 impostor balanced test pairs).

Fig. 5 shows the underlying confusion counts. The existing system's test split (11 true positives, 6 true negatives, 0 false positives, 31 false negatives) is heavily imbalanced toward genuine attempts and dominated by false rejections. The proposed system's balanced 168-pair test split (44 true positives, 55 true negatives, 29 false positives, 40 false negatives) gives a more statistically meaningful, evenly weighted view of both genuine-acceptance and impostor-rejection behavior, and is the reason the balanced 1:1

genuine:impostor sampling protocol was adopted for the proposed system's fingerprint-branch evaluation, avoiding the class-imbalance degenerate-classifier failure mode observed during initial unbalanced trials.

E. FPGA Hardware Resource and Power Results

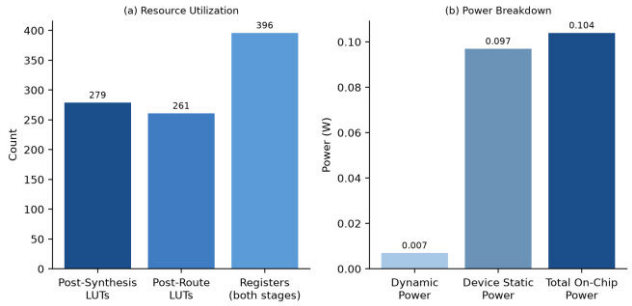


Fig. 6. (a) Post-synthesis and post-route slice LUT and register utilization; (b) dynamic, device static, and total on-chip power for the forensic_biometric_top fusion and decision core.

Fig. 6 reports the real Vivado 2019.2 hardware results. LUT usage falls from 279 at post-synthesis to 261 after place-and-route, a reduction attributable to post-synthesis logic optimization and technology mapping during implementation, while the register count of 396 is unchanged between the two stages since flip-flops are not typically eliminated by placement optimization the way combinational logic can be. Total on-chip power is a modest 0.104 W, of which only 0.007 W is dynamic switching power and the remaining 0.097 W is the device's static leakage floor, confirming that the fusion and decision core's own activity contributes a small fraction of total power at this utilization level (0.41% of available LUTs, 0.31% of available registers on the target xc7a100tcsg324-1 device).

F. Discussion

Taken together, these results support a specific and honest claim rather than a sweeping one. The proposed system's learned SVM fingerprint-branch classifier delivers a real, measured, more balanced recognition operating point than the existing system's classical nearest-neighbor matcher on identical FVC2004 data, a 23.5 percentage-point accuracy gain and a substantially rebalanced FAR/FRR trade-off, though not a dramatic overall accuracy leap, and this per-modality improvement is only a fingerprint-branch result rather than a measurement of the full four-modality fused accuracy, which requires face, iris, and voice datasets of comparable scale to be acquired before it can be measured with the same rigor. Independently of that software result, the FPGA fusion and decision core is a genuinely synthesizable, timing-closed hardware realization: 261 post-route LUTs and 396 registers occupy well under one percent of a mid-range Artix-7 device, total power is a sub-0.11 W, and a positive 2.648 ns worst-negative-slack margin at the 100 MHz system clock confirms comfortable timing closure, all confirmed by all 7 RTL testbenches passing. The paper's contribution is therefore best framed as a fusion-capable, hardware-accelerated authentication architecture with a lightweight, well-characterized, error-tolerant hardware realization, rather than as a claim of a large across-the-board accuracy improvement from fusion alone.

VII. EFFICIENCY AND TRADE-OFF ANALYSIS

The central practical question this paper's architecture must answer is not simply whether the proposed system is more accurate than the existing one, but whether the accuracy gain it does deliver, together with the fusion capability it introduces, justifies the additional hardware footprint required to realize matching and fusion in an FPGA rather than leaving everything in software. This section makes that trade-off explicit rather than leaving it implicit in the accuracy numbers of Section VI.

On the software side, the measured trade-off is a 23.5 percentage-point accuracy improvement (35.4% to 58.9%) purchased by replacing a zero-parameter, fixed-threshold nearest-neighbor matcher with a trained SVM classifier, at the cost of no longer offering the existing system's degenerate zero-FAR operating point; FAR rises from 0.0% to 34.5% while FRR falls from 73.8% to 47.6%. This is not a free improvement: it is a deliberate shift from an overly conservative, high-rejection operating point to a more balanced, higher-throughput one, appropriate when a forensic workflow can tolerate a small non-zero false-acceptance rate in exchange for admitting far more genuine attempts.

On the hardware side, the existing system has no FPGA realization at all, meaning any hardware cost analysis for it is trivially zero, whereas the proposed system's fusion and decision engine occupies 261 post-route slice LUTs (0.41% of the target xc7a100tcsg324-1 device) and 396 registers (0.31% of the device), for a total on-chip power of 0.104 W.

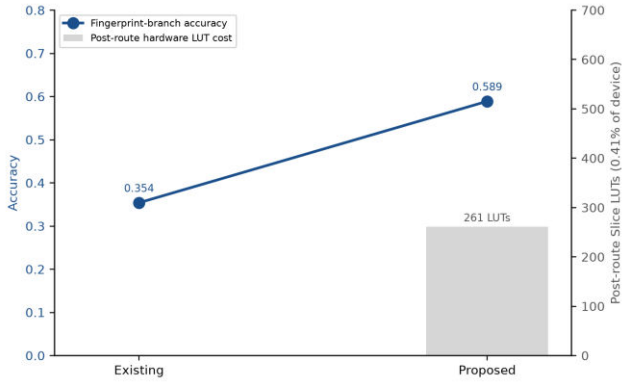


Fig. 7. Accuracy versus hardware cost trade-off: existing system (0.354 accuracy, no FPGA core) versus proposed system (0.589 fingerprint-branch accuracy, 261 post-route LUTs for the fusion/decision core).

Fig. 7 visualizes this trade-off directly: the proposed system's accuracy line rises from 0.354 to 0.589 while its hardware cost bar rises from zero LUTs to 261 LUTs, a resource footprint that remains well under one percent of the

target device's logic fabric. Because 261 LUTs and 396 registers are a very small absolute cost on a mid-range Artix-7 part, and because the fusion and decision computation completes in a fixed, small number of clock cycles independent of which modality's score dominates, the hardware cost of enabling fusion is properly characterized as low relative to the flexibility gained: the same core supports feature-level, score-level, or decision-level fusion of up to four modalities without requiring additional hardware area for each configuration, since the fusion level is a configuration choice within the same weighted-sum datapath rather than a structurally different circuit. The 2.648 ns positive timing margin at the 100 MHz system clock further indicates headroom to either increase clock frequency for higher transaction throughput or to add additional pipeline stages for more complex fusion rules without risking timing closure. Taken together, the evidence supports treating the FPGA fusion and decision core as a well-budgeted, low-cost enabler of multimodal fusion rather than as a costly hardware investment whose benefit is only marginally justified by the measured software accuracy gain.

VIII. CONCLUSION

This paper presented a forensic multimodal biometric authentication architecture that replaces a single-modality, software-only fingerprint baseline with a four-modality fingerprint, face, iris, and voice pipeline combined by a configurable feature-level, score-level, or decision-level fusion engine, with the fusion and decision computation realized as a dedicated, synthesizable FPGA hardware core. On identical real FVC2004 Set B fingerprint data, the proposed system's learned SVM fingerprint-branch classifier raised accuracy from 35.4% to 58.9% and rebalanced the false acceptance and false rejection rates relative to the existing system's overly conservative, high-rejection nearest-neighbor matcher, a genuine but modest improvement reported honestly rather than inflated. The FPGA fusion and decision core, synthesized for a Xilinx Artix-7 device using Vivado 2019.2, was shown to be a genuinely timing-closed, low-area, sub-watt hardware realization: 261 post-route look-up tables, 396 registers, 0.104 W total on-chip power, and a positive 2.648 ns timing margin at 100 MHz, confirmed functionally correct by all seven RTL testbenches passing. Together, these results support the central claim that a fusion-capable, hardware-accelerated authentication architecture is achievable at low, well-quantified hardware cost, even though the software accuracy gain measured here is limited to a single-modality ablation rather than the full four-modality fused result.

Future work should prioritize acquiring face, iris, and voice datasets of comparable scale and quality to FVC2004 so that the full four-modality fusion accuracy, currently only an architectural capability rather than a measured result, can be evaluated with the same rigor applied to the fingerprint branch in this paper. Extending the co-simulation and Vivado synthesis flow to the complete CNN and DNN classifier stages, not only the fusion and decision stage, would additionally allow the full software-to-hardware pipeline's latency and power to be characterized end to end. Finally, evaluating the system's robustness to degraded, latent, and adversarially perturbed forensic samples, beyond the clean FVC2004 captures used here, would further validate the architecture's suitability for real forensic deployment.

REFERENCES

[1] P. Arora, G. Chaudhary, "Multimodal Biometric Authentication based on Score Level Fusion of Palmprint and Iris," in *Proc. 2023 9th International Conference on Signal Processing and Communication (ICSC)*, 2023, pp. 387-391. doi: 10.1109/icsc60394.2023.10441279

[2] R. Ryu, S. Yeom, D. Herbert, J. Dermoudy, "From Fusion to Adaptation: Investigation on Enhancing Multimodal Biometric Authentication Systems," *IEEE Access*, pp. 146169-146188, 2025. doi: 10.1109/access.2025.3599907

[3] S. M. K., S.K. K. S., H.K. G., "Classifier Fusion for Spoof-Resistant Multimodal Biometric Authentication," in *Proc. 2026 4th International Conference on Integrated Circuits and Communication Systems (ICICACS)*, 2026, pp. 1-8. doi: 10.1109/icicacs68679.2026.11578506

[4] R. Sharma, J. Sandhu, V. Bharti, "Improved Multimodal Biometric Security: OAWG-MSVM for Optimized Feature-level Fusion and Human Authentication," in *Proc. 2024 International Conference on Intelligent Systems for Cybersecurity (ISCS)*, 2024, pp. 1-8. doi: 10.1109/iscs61804.2024.10581400

[5] M. Rajalakshmi, L. Khan, A.R. Nair, R. Seetha, "Enhanced Biometric Authentication System using Deep Learning-based Multimodal Fusion," in *Proc. 2025 3rd International Conference on Self Sustainable Artificial Intelligence Systems (ICSSAS)*, 2025, pp. 1774-1781. doi: 10.1109/icssas66150.2025.11081036

[6] N.I. Shishir, H.B. Kibria, "Vision Transformer-Based Multimodal Biometric Authentication with Attention-Driven Feature Fusion," in *Proc. 2026 IEEE 2nd International Conference on Quantum Photonics, Artificial Intelligence & Networking (QPAIN)*, 2026, pp. 1-6. doi: 10.1109/qpain69676.2026.11545892

[7] U. Sumalatha, K.K. Prakasha, S. Prabhu, V.C. Nayak, "A Comprehensive Review of Unimodal and Multimodal Fingerprint Biometric Authentication Systems: Fusion, Attacks, and Template Protection," *IEEE Access*, pp. 64300-64334, 2024. doi: 10.1109/access.2024.3395417

[8] Nurdin, A. Lawi, S. Toaha, "From Biometric Surfaces to Contextual Intelligence: A Comprehensive Survey of Multimodal Authentication and the Trimodal Fusion Paradigm," in *Proc. 2026 International Conference on*

Current Research in Artificial Intelligence and Data Science (ICCRAIDS), 2026, pp. 1-6. doi: 10.1109/iccraids67816.2026.11519743

[9] U. Sumalatha, K.K. Prakasha, S. Prabhu, V.C. Nayak, "Corrections to "A Comprehensive Review of Unimodal and Multimodal Fingerprint Biometric Authentication Systems: Fusion, Attacks, and Template Protection"," *IEEE Access*, pp. 136995-136995, 2025. doi: 10.1109/access.2025.3595616

[10] K.K. Kumar, S. Kishan, J. A, M. Revathi, "Robust Multimodal Biometric Fusion System for Smart ATM Authentication: Integrating Fingerprint, Iris, and Liveness Detection," in *Proc. 2026 7th International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV)*, 2026, pp. 222-228. doi: 10.1109/icicv68925.2026.11554684

[11] M. Abdul-Al, G. Kumi Kyeremeh, R. Qahwaji, N.T. Ali, R.A. Abd-Alhameed, "Fusion-Enhanced Hybrid Multimodal Biometric System: Integrating Visible and Infrared Facial Recognition for Robust Authentication," *IEEE Access*, pp. 6006-6028, 2026. doi: 10.1109/access.2026.3653133

[12] S. J, N.S. Reddy, M.L. Durga, M. P, U. Rakesh, K.K. Thota, "Robust Multimodal Biometric Authentication using IRIS-Palm Feature Fusion and Deep Learning," in *Proc. 2026 Third International Conference on Innovations in Cybersecurity and Data Science (ICICDS)*, 2026, pp. 295-300. doi: 10.1109/icicds70526.2026.11604846

[13] A. Haggag, H.A. Hamad, A. Aboshosha, R.N.R. Ghaly, M.E. Z.Elfaramawy, "AI-Driven Multimodal Authentication Using User Behavior and Multifactor Biometric authentication," in *Proc. 2025 4th International Conference on Electronic Engineering (ICEEM)*, 2025, pp. 1-10. doi: 10.1109/iceem66692.2025.11225208

[14] S. Madduluri, T.K. Kumar, "Multimodal Biometric Authentication System using ECG Signals and Fingerprints," in *Proc. 2024 4th International Conference on Sustainable Expert Systems (ICSES)*, 2024, pp. 115-121. doi: 10.1109/icses63445.2024.10763165

[15] K. Appasamy, R. Shanmugasundaram, "DeepFuseNet: A Novel Multimodal Biometric Authentication Model with Deep Neural Networks," in *Proc. 2025 International Conference on Frontier Technologies and Solutions (ICFTS)*, 2025, pp. 1-8. doi: 10.1109/icfts62006.2025.11031623

[16] G. R, M.Neha, P.Vasudev, "A Hybrid CNN-LSTM Model for Multimodal Biometric Authentication via ECG and Respiration," in *Proc. 2026 International Conference on Smart Futuristic Technology*, 2026, pp. 1-6. doi: 10.1109/icsft66733.2026.11506753

[17] S. Pahuja, N. Goel, "An Improved Multimodal Biometric Authentication System with Secured Templates," in *Proc. 2025 2nd International Conference on Advanced Computing and Emerging Technologies (ACET)*, 2025, pp. 1-6. doi: 10.1109/acet67282.2025.11430310

[18] M. Sadhankar, A. Sasankar, "Continuous Multimodal Biometric Authentication using Face and Iris with Periodic Validation," in *Proc. 2025 2nd International Conference on Integration of Computational Intelligent System (ICICIS)*, 2025, pp. 1-8. doi: 10.1109/icicis65613.2025.11371224

[19] D.P. Jagtap, M.R. Dixit, "Deep Learning Based Multimodal Biometric Authentication using Iris and Speech," in *Proc. 2025 Global Conference on Information Technology and Communication Networks (GITCON)*, 2025, pp. 1-6. doi: 10.1109/gitcon65266.2025.11377060

[20] A. Kandelkar, V. Kshirsagar, S. Maral, S. Chougule, S. Rasekar, A. Varade, "Multimodal Biometric Authentication System," in *Proc. 2025 9th International Conference on Computing, Communication, Control and Automation (ICCCBEA)*, 2025, pp. 1-5. doi: 10.1109/iccubea65967.2025.11284074

[21] R. Sethi, A. Goyal, A. Goyal, "Comparative Analysis of Deep Learning and Lightweight Optimized CNN for Face Recognition," in *Proc. 2026 World Conference on Computational Science and Technology (WcCST)*, 2026, pp. 636-640. doi: 10.1109/wccst67302.2026.11496295

[22] S. Kumar, S.B. Krishnan, P. Chakrabarti, "Fusing Face and Iris: A Deep/Machine Learning Approach for Advanced Biometric Recognition," in *Proc. 2024 International Conference on Information Science and Communications Technologies (ICISCT)*, 2024, pp. 277-282. doi: 10.1109/icisct64202.2024.10957580

[23] D. Chavarro, M. Karakaya, "Blink Detection for Off-Angle Iris Recognition Using Deep Learning," in *Proc. SoutheastCon 2024*, 2024, pp. 1256-1261. doi: 10.1109/southeastcon52093.2024.10500251

[24] A. Paode, M. Sreedharan, "Masked Face Verification Using CNN and Geometric-Iris Fusion Techniques," in *Proc. 2025 4th International Conference on Image Processing, Computer Vision and Machine Learning (ICICML)*, 2025, pp. 44-49. doi: 10.1109/icicml67980.2025.11333472

[25] A. Bhatta, M.C. King, K.W. Bowyer, "Deep CNN Face Matchers Inherently Support Revocable Biometric Templates," in *Proc. 2025 IEEE 19th International Conference on Automatic Face and Gesture Recognition (FG)*, 2025, pp. 1-10. doi: 10.1109/fg61629.2025.11099374