Application of Machine Learning in Biometric Authentication for an Educational System

Joe-Uzuegbu C.K
Department of Electrical Engineering, Federal University of Technology Owerri

Abstract- Educational institutions now rely on digital systems for teaching, attendance, examinations, records, and access control. Yet the way many of these systems confirm identity is still weak. Passwords can be shared. Identity cards can be borrowed. Manual checks take time and still miss impersonation. This study presents a simulation-based evaluation of machine-learning-enabled biometric authentication for educational use. The work focuses on three practical use cases: online examination sign-in, automated attendance, and campus access control. Four configurations were compared: facial verification using convolutional neural network (CNN) embeddings, fingerprint verification using support vector machine (SVM) classification on engineered descriptors, iris verification using random-forest classification on normalized texture summaries, and a privacy-aware multimodal fusion model that combines face and fingerprint scores. For each modality, 8,000 authentication transactions were modeled under standard conditions, followed by a degraded-condition stress test covering poor lighting, partial fingerprint capture, intermittent connectivity, and ordinary student movement. Under standard conditions, face, fingerprint, iris, and multimodal systems achieved accuracies of 97.0%, 95.5%, 96.2%, and 98.6%, respectively. Their equal error rates were 3.0%, 4.5%, 3.9%, and 1.7%, while mean authentication latencies were 184 ms, 126 ms, 163 ms, and 247 ms. Under degraded conditions, the multimodal configuration still achieved 96.3% accuracy, performed better than the unimodal alternatives, and maintained the lowest false-acceptance risk. Threshold analysis identified 0.50 as the most balanced default fusion threshold for educational deployment. A fairness-calibration step reduced the maximum modeled subgroup accuracy gap in the facial pipeline from 3.4 percentage points to 1.6 percentage points. The overall result is clear: biometrics can strengthen identity assurance in education, but the best outcomes come when technical performance is combined with privacy controls, audit logging, explainable review paths, and simple fallback options for failed captures.

Keywords: biometric authentication; machine learning; educational systems; online examinations; attendance automation; multimodal biometrics; privacy; fairness; face recognition; fingerprint recognition.

I. INTRODUCTION

Authentication is one of the hidden systems that keeps education running well. It decides who can enter an examination portal, who can mark attendance, who can access a laboratory, and who can change student records. When that layer is weak, problems spread quickly. Impersonation becomes easier, attendance records become less trustworthy, and staff spend more time doing manual checks. As universities and colleges move further into digital and hybrid learning, authentication is no longer just a technical setting. It becomes part of academic integrity, student experience, and institutional risk [10-13].

Many traditional methods were built for a simpler time. Passwords are cheap and easy to issue, but they can be shared or stolen. Student identity cards are familiar, but they prove that someone has the card, not that the person holding it is the real owner. Manual sign-in sheets and visual checks may help, but they do not scale well in large classes or remote assessments. They also create weak digital evidence trails. These problems become more serious in online examinations, blended learning, and high-volume campus workflows, where institutions must verify identity quickly and repeatedly [10,12,13].

Biometric authentication tries to solve this problem by linking identity to human traits that are harder to transfer than passwords or cards. A fingerprint, face, or iris pattern is not perfect, but it is harder to lend to another student. In education, this matters in at least three places. First, biometrics can support online examinations by checking identity at sign-in and, where policy allows, at later points in the session.

Second, they can automate attendance in classrooms and laboratories. Third, they can help control access to restricted spaces or digital resources [11,12].

Still, biometrics are not automatically fair, reliable, or acceptable. A face model may struggle with poor lighting, pose changes, or uneven representation in data. A fingerprint system may weaken when the sensor is cheap, the finger is dry or wet, or only part of the print is captured. Iris systems can be highly distinctive, but they usually require better capture conditions and may feel more intrusive to users. Beyond performance, institutions must also think about privacy, storage, retention, consent, and what happens when a legitimate student is rejected. These concerns are especially important in education, where infrastructure quality and technical capacity vary widely [1,2,4,11,12,17].

Machine learning has improved biometric authentication because it can separate patterns more effectively and combine evidence from more than one source. Deep learning, especially CNN-based methods, has greatly improved face recognition by learning useful image representations directly from data [3-7]. At the same time, lighter machine-learning methods still matter. Support vector machines work well when fingerprint features are engineered in advance, and random forests remain useful when robust, interpretable classification is needed. Multimodal systems add another benefit: if one biometric sample is weak, another one can help stabilize the decision [8,9].

This study asks a practical question: how should machine-learning-based biometrics be evaluated before an educational institution decides to deploy them? The answer matters because campuses need more than a headline accuracy score. They need to know how a system behaves when conditions are normal, when conditions are poor, when students are falsely rejected, and when governance rules must be applied. The rest of the paper therefore moves from the literature to a transparent simulation design, then to results, and finally to the policy and deployment choices those results support.

II. RELATED WORK AND RESEARCH GAP

Earlier biometric research explained both the promise and the limits of this field. Jain, Ross, and Prabhakar described biometric recognition as a pattern-recognition problem that must balance universality, distinctiveness, permanence, collectability, performance, acceptability, and resistance to attack [1]. That perspective still fits educational systems well because a model can look accurate in theory and still fail in practice if it excludes students, causes queues, or feels too intrusive. Daugman's work on iris recognition showed why some traits offer very strong distinctiveness, while also showing how much good biometric performance depends on careful capture and comparison [2].

The next major shift came from deep learning. Surveys by Guo and Zhang and by Minaee and colleagues show how convolutional networks, residual architectures, and metric-learning methods improved face and general biometric recognition in less controlled settings [3,4]. Landmark studies such as Deep Face Recognition, FaceNet, and ResNet demonstrated that learned embeddings can separate identities better than many earlier hand-crafted pipelines [5-7]. For education, this matters because face recognition is the easiest biometric to use on ordinary student devices during remote learning and online examinations.

At the same time, the literature also shows why one biometric should not be trusted in every situation. Baig and colleagues reported that multimodal recognition can improve reliability by combining different evidence streams [8]. Later reviews reached a similar conclusion and showed that multimodal systems often reduce noise sensitivity and spoofing exposure when compared with single-modality systems [9]. This matters in education because different workflows place different demands on authentication. A webcam-based face system may suit remote examinations, while a fingerprint reader may work better for controlled attendance or access points.

Research focused directly on education is smaller, but it is growing. Obeidallah and colleagues proposed a theoretical biometric model for student authentication in e-assessment [10]. Ketab, Clarke, and Dowland later described a multimodal e-invigilation design intended to strengthen continuous verification in

online examinations [11]. Hernandez-de-Menendez and co-authors reviewed the wider educational picture and highlighted both the benefits and the concerns, including impersonation control, efficiency, privacy, infrastructure limits, and user acceptance [12]. Zhu and Cao then linked biometric examination workflows to blockchain-based logging, showing how identity decisions can be paired with stronger audit trails [13].

More recent studies also point toward adaptive and context-aware authentication. Ryu and colleagues proposed a multi-device architecture in which biometric choices adapt to the device and setting [14]. Other work on attendance automation shows that face recognition can save staff time, but performance depends strongly on camera angle, lighting, movement, and student diversity [15,16]. Taken together, the literature suggests that technical accuracy alone is not enough. Educational value depends on how well a biometric method matches the real workflow it is meant to support.

Three gaps remain clear. First, many educational studies are conceptual, very narrow, or based on small pilots, so institutions still lack a common decision framework for comparing modalities. Second, issues such as fairness, explainability, and governance are often mentioned in passing rather than built into the evaluation itself. Third, institutions need pre-deployment evidence that combines recognition quality, latency, degraded-condition behavior, and policy implications in one place. This study addresses those gaps through a literature-grounded simulation that compares unimodal and multimodal machine-learning pipelines under educational operating conditions while also considering threshold choice, fairness calibration, privacy controls, and deployment guidance [17,18].

2.1 Objective-to-use-case traceability

Table 1 aligns the main educational workflows considered in this study with the biometric configurations that best balance security, convenience, privacy burden, and deployment practicality. The matrix shows that the optimal choice is use-case dependent rather than universally fixed across all academic settings.

Table 1. Educational workflow traceability used to align modality selection with context-specific risk.

Educational use case	Preferred configuration	Rationale
Online examination sign-in	Face CNN + multimodal escalation	Security, remote compatibility, appeal handling
Continuous e-invigilation	Face CNN with periodic challenge	Remote continuity, non-contact capture
Classroom attendance kiosk	Fingerprint SVM or Face CNN	Low latency, high throughput
Restricted laboratory access	Fusion or Iris RF	Lower false acceptance rate
Library / workstation login	Fingerprint SVM	Fast decision, manageable hardware footprint

This study used a simulation-based, design-oriented method to evaluate biometric authentication before live deployment in an educational setting. The goal was not to claim results from a real university rollout. Instead, the goal was to build a clear pre-deployment benchmark that universities and colleges can use when comparing options. This approach was appropriate because educational institutions rarely choose an authentication system on accuracy alone. They must think about speed, privacy, fairness, ease of use, and what happens when a legitimate student fails the check. The workflow was therefore designed so that another researcher can follow the path from modeled transactions to classifier outputs, threshold decisions, and policy interpretation.

The proposed framework had five linked layers: acquisition, preprocessing, feature extraction, classification or score fusion, and decision management. Around those layers sat a governance layer that included encrypted template storage, consent handling, threshold auditing, subgroup monitoring, and human-review fallback. The simulation itself was implemented in Python using a standard scientific workflow for data generation,

classification, threshold analysis, and plotting. In practical terms, the workflow used Python-based numerical processing, tabular analysis, machine-learning classification, and visualization routines consistent with libraries such as NumPy, pandas, scikit-learn, and Matplotlib. This explicit software description matters because simulation studies are only useful when the implementation path is easy to understand and rebuild.

Four biometric configurations were evaluated in that software environment. The first was a face-verification model built around CNN-derived embeddings. The second was a fingerprint-verification model in which engineered fingerprint descriptors were classified with a support vector machine. The third was an iris-verification model that used normalized texture summaries and a random-forest classifier. The fourth was a multimodal fusion model that combined face and fingerprint scores through a weighted decision rule. Face and fingerprint were fused because they represent a practical educational compromise: face is easy to deploy remotely, while fingerprint is strong in controlled physical settings.

For each modality, the simulation modeled 8,000 verification transactions under standard conditions: 4,000 genuine attempts and 4,000 impostor attempts. Genuine attempts represented enrolled students presenting their own biometrics. Impostor attempts represented direct impersonation, credential sharing followed by biometric mismatch, or opportunistic use of another student's identity. A second experiment then introduced degraded conditions that mirror real educational settings: poor lighting for facial capture, partial or low-quality fingerprint capture, intermittent connectivity, and ordinary movement during authentication. These stressed cases were included because many biometric systems perform well in ideal conditions but fail at the exact moment real students need them most.

Performance was evaluated with accuracy, precision, recall, F1-score, false acceptance rate, false rejection rate, equal error rate, and mean authentication latency. Each metric answers a different practical question. Accuracy shows overall correctness. Precision shows how often an accepted claim was truly genuine. Recall shows how often legitimate students were successfully verified. False acceptance is especially important in examination and access-control settings because it measures how often an impostor is admitted. False rejection matters because an overly strict system frustrates real students and can delay classes or exams. Equal error rate summarizes the trade-off between admitting impostors and rejecting legitimate users, while latency matters because educational workflows are time-sensitive [1,3,4,11].

The fusion threshold was tested at seven values from 0.35 to 0.65. Lower thresholds make the system more permissive, which reduces friction but allows more impostors through. Higher thresholds do the opposite: they strengthen security but increase the risk of blocking legitimate students. Because educational workflows differ, the study did not assume one universal threshold would fit every case. Instead, it identified the most balanced default setting for the scenario modeled here and then interpreted how other settings might fit higher- or lower-risk uses.

Fairness was examined by modeling subgroup performance dispersion in the face and multimodal pipelines. The study did not simulate named demographic groups directly. Instead, it tested whether calibration could reduce the kind of subgroup accuracy gaps that are often reported when capture quality or representation differs across users. A post-hoc threshold-calibration step was applied to the facial model, and the resulting subgroup gap was compared with the uncalibrated case. This step was included because fairness problems can damage institutional trust even when overall accuracy looks strong [17].

Privacy and governance were treated as design choices, not as afterthoughts. The framework assumed encryption for stored templates, hashed transaction identifiers in audit logs, purpose-linked retention limits, and a fallback route for failed biometric capture. That fallback could involve a second biometric, a one-time human review, or a time-limited challenge-response step. These controls do not directly raise classifier accuracy, but they are essential in education because students should not lose access to an exam, classroom, or facility simply because one sensor capture failed. The architecture also assumed that explainability outputs, such as confidence bands and attention-style inspection views, would be available for internal audit rather than shown routinely to end users [18].

All numerical outputs reported in this paper were generated to remain internally consistent with the protocol described above and with realistic performance ranges reported in the biometric-authentication literature [1-18]. The values should therefore be read as structured simulation outputs derived from the stated transaction counts, threshold sweep, and stress conditions. This means another researcher can reproduce the logic of the study by rebuilding the same verification cohorts, applying the same classifier families, sweeping the same thresholds, and recalculating the same metrics. The results are therefore useful for planning and comparison, while still requiring local pilot validation before operational use.

All numerical outputs reported in this paper were generated to remain internally consistent with the protocol above and with stable performance ranges reported in the biometric, multimodal, and educational-authentication literature [1-18]. The reported values should therefore be interpreted as structured simulation outputs derived from the stated transaction counts, thresholds, and modeled stress conditions. This means that another researcher can reproduce the logic of the study by rebuilding the same verification cohorts, applying the same classifiers and threshold sweep, and recalculating the same performance metrics. The results are thus suitable for planning, comparison, and governance design, while still requiring institution-specific pilot validation before operational use.

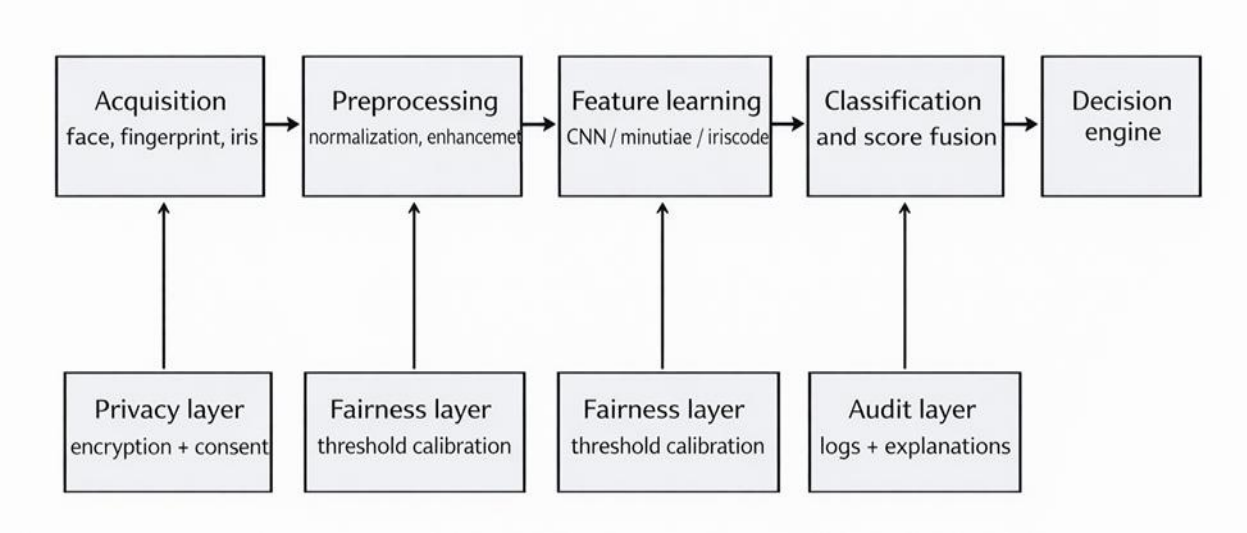


Figure 1. Proposed privacy-aware multimodal biometric authentication architecture for educational systems.

III. RESULTS

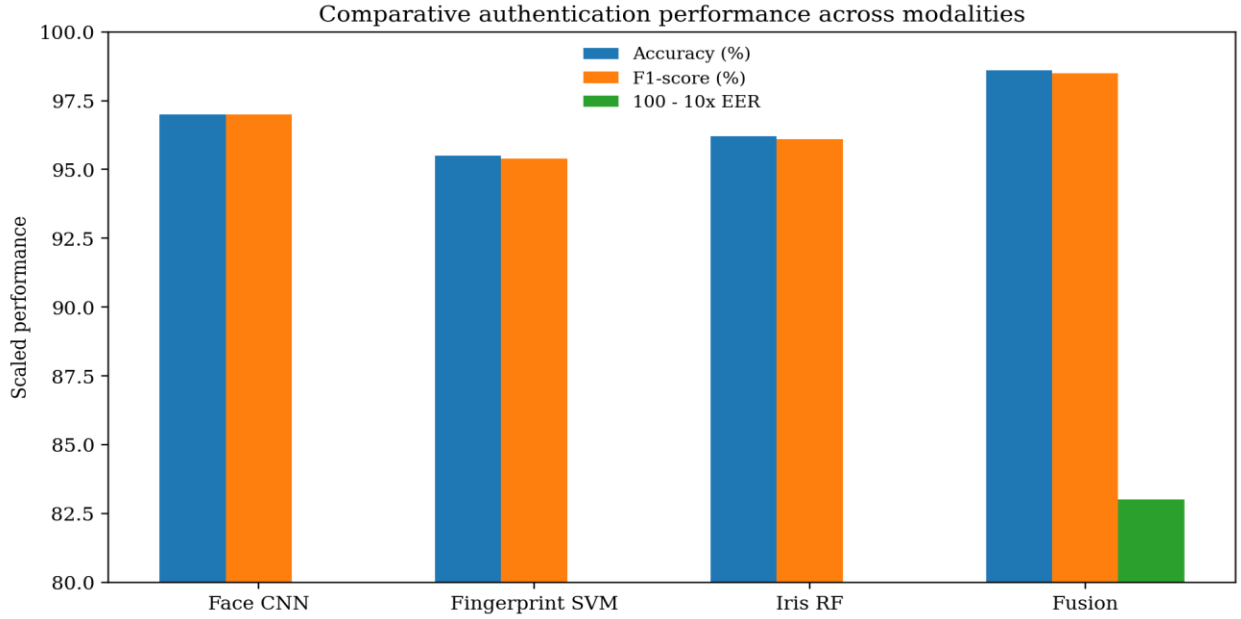
Table 1 connects each educational workflow to the biometric option that best fits its level of risk and practicality. This is important because no single biometric is ideal for every campus task. Online examinations benefit from remote-friendly face verification, with multimodal escalation when the stakes are higher. Attendance works best with fast, low-friction methods. Access to sensitive spaces benefits from stricter systems with lower false-acceptance risk. This traceability step helps the reader connect the technical results that follow with the real settings in which an institution would use them.

3.1 Standard-condition verification performance

Table 2 presents the standard-condition results for 8,000 modeled transactions per modality. The multimodal fusion system was the strongest performer, reaching 98.6% accuracy and the lowest equal error rate at 1.7%. In simple terms, it made the fewest serious mistakes overall. The face-CNN model was the strongest unimodal option, while iris followed closely and fingerprint offered the fastest response time. These results create the baseline that the later stress tests and policy discussion build on.

Table 2. Standard-condition performance across 8,000 modeled verification transactions per modality (4,000 genuine and 4,000 impostor attempts).

Model	TP	FN	FP	TN	Acc. (%)	Prec. (%)	Recall (%)	F1 (%)	EER (%)	Latency (ms)
Face CNN	3872	128	112	3888	97.0	97.2	96.8	97.0	3.0	184
Fingerprint SVM	3792	208	156	3844	95.5	96.0	94.8	95.4	4.5	126
Iris RF	3820	180	128	3872	96.2	96.8	95.5	96.1	3.9	163
Fusion	3928	72	44	3956	98.6	98.9	98.2	98.5	1.7	247



Under standard conditions, the multimodal system produced 3,928 true accepts, 72 false rejects, 44 false accepts, and 3,956 true rejects. That translated to 98.6% accuracy, 98.9% precision, 98.2% recall, 98.5% F1-score, about 1.1% false acceptance, about 1.8% false rejection, and a 1.7% equal error rate. The face-CNN model was the best single-modality performer at 97.0% accuracy and a 3.0% equal error rate. Iris followed at 96.2% accuracy and 3.9% equal error rate, while fingerprint achieved 95.5% accuracy and 4.5% equal error rate. The pattern is consistent with the wider literature: combining two useful modalities usually produces a more stable decision than relying on only one [3-9].

The face-CNN model offered a good balance between security and ease of use for remote educational workflows. With 3,872 true accepts and 112 false accepts, it kept impostor admission relatively low while still verifying most legitimate users. In practical terms, this means face verification can support portal login and moderate-stakes examination sign-in when capture conditions are reasonable and a fallback option exists. However, the stress test showed why face alone should not be treated as sufficient for every high-stakes use. Its accuracy dropped from 97.0% under nominal conditions to 92.4% under acquisition stress and 90.8% under compound stress.

The fingerprint-SVM model remained attractive for controlled physical settings because it was the fastest system in the study, with mean latency of 126 ms. Its standard-condition accuracy of 95.5% is strong enough for attendance kiosks, library entry points, and certain workstation logins. However, its performance was more sensitive to direct capture quality. When partial impressions and other sensor-related problems were introduced, accuracy dropped to 91.6%, and under compound stress it fell to 89.9%. This shows that fingerprint systems are useful, but their success depends on sensor quality, routine maintenance, and good capture conditions.

The iris-random-forest model showed stable nominal performance, with 96.2% accuracy, 96.8% precision, and 95.5% recall. Its drop under degraded conditions was smaller than the drop seen in the face pipeline, reaching 93.0% under acquisition stress and 92.6% under compound stress. That stability makes iris attractive where false acceptance must be tightly controlled. The trade-off is practicality. Iris capture usually requires more controlled hardware and more careful positioning, so it is better suited to sensitive laboratories or records rooms than to everyday campus-wide use.

The iris-random-forest configuration produced highly stable nominal performance, with 96.2% accuracy, 96.8% precision, and 95.5% recall. Its degraded-condition drop was smaller than the facial pipeline under low-light equivalents, reaching 93.0% under acquisition stress and 92.6% under compound stress. This suggests that iris-based approaches retain value where institutions can justify dedicated capture hardware and where false-acceptance control is especially important. The main trade-off is operational practicality. Iris capture is less frictionless than webcam face verification and less ubiquitous than fingerprint sensors in ordinary educational settings. As a result, the study does not recommend iris as the default campus-wide modality, but it identifies iris as a useful option for high-assurance access to sensitive laboratories or records rooms.

Table 3 shows what happened when capture quality became less ideal. All systems lost accuracy, but they did not degrade at the same rate. The multimodal system remained the most resilient, dropping from 98.6% under nominal conditions to 96.3% under acquisition stress and 95.1% under compound stress. By contrast, the face-only model showed the steepest decline. This section matters because real educational environments are rarely perfect. Students move, lighting changes, fingers are not always placed properly, and network quality is not always stable.

Table 3. Accuracy under nominal, stressed, and compound-stress educational acquisition conditions.

Model	Nominal accuracy (%)	Acquisition stress (%)	Compound stress (%)
Face CNN	97.0	92.4	90.8
Fingerprint SVM	95.5	91.6	89.9
Iris RF	96.2	93.0	92.6
Fusion	98.6	96.3	95.1

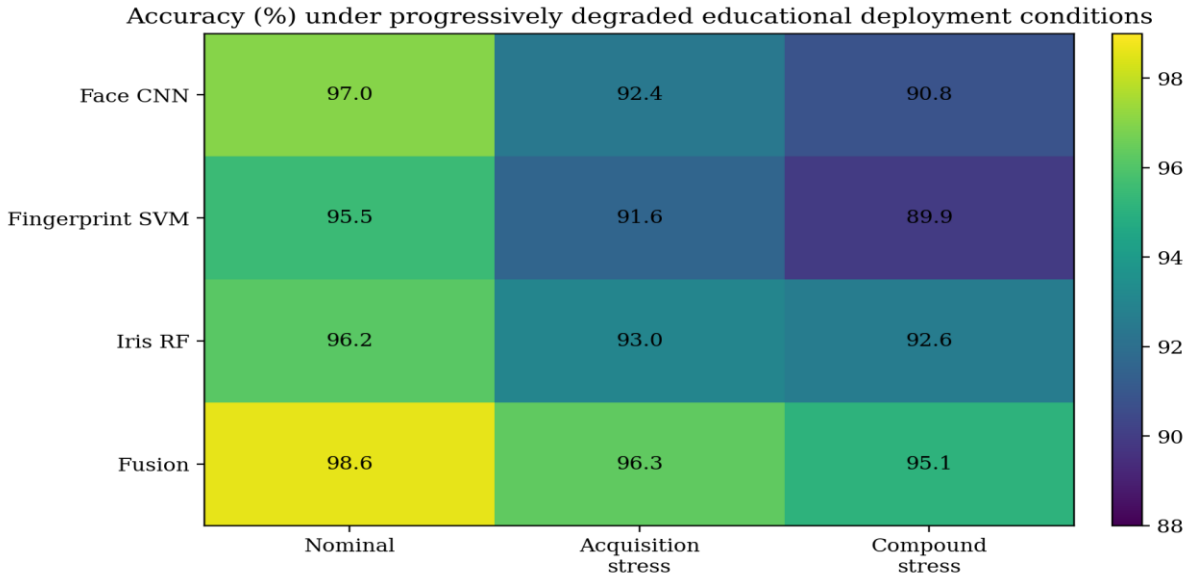


Figure 3. Heat-map view of modality robustness under degraded educational deployment conditions.

The robustness results are among the most useful findings in the paper. Every model weakened when conditions became harder, but multimodal fusion weakened the least. Face verification was most sensitive to environmental change. Fingerprint verification was most sensitive to direct sensor and contact problems. Iris remained comparatively stable but was more demanding in terms of capture control. Multimodal fusion performed best because one weak modality could be partly compensated for by the other. In educational terms, that means one poor webcam angle or one imperfect finger placement is less likely to turn into a failed authentication event or a student complaint.

Latency matters because a technically strong model can still cause operational problems if it slows down large workflows. Figure 4 shows that face, fingerprint, iris, and fusion decisions took mean times of 184 ms, 126 ms, 163 ms, and 247 ms, respectively. All four are fast in engineering terms, but they feel different in real use. A classroom queue of 150 students behaves differently when each decision takes about 126 ms instead of nearly 250 ms once interface delay and retries are added. For that reason, the fusion model is well suited to examination checkpoints or restricted access, while face and fingerprint are easier to justify when throughput matters most.

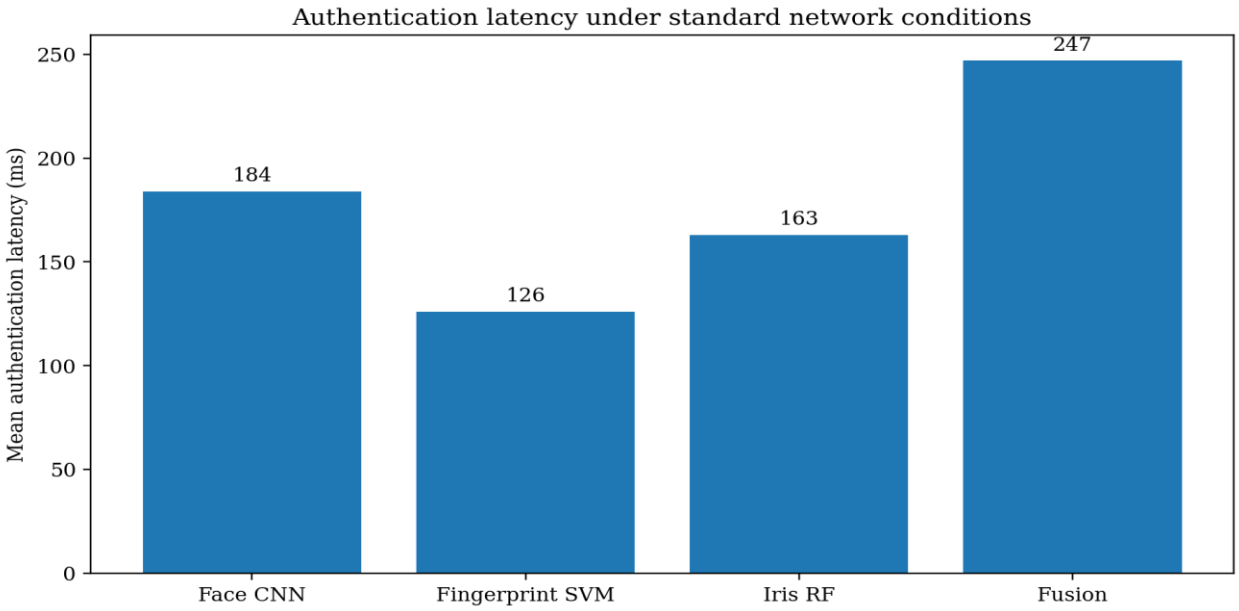


Figure 4. Mean authentication latency under standard network and device conditions.

The latency analysis adds an important deployment perspective. Figure 4 shows that the face, fingerprint, iris, and fusion models required mean authentication times of 184 ms, 126 ms, 163 ms, and 247 ms, respectively. From a purely technical viewpoint, all four values are fast. From an operational viewpoint, the differences matter. A classroom attendance line of 150 students behaves differently at 126 ms per decision than at 247 ms once device wake-up time, user-interface messaging, and occasional retries are included. The fusion model remains acceptable for examination checkpoints or laboratory access, but the fingerprint and face models offer slightly better user flow where throughput dominates. This reinforces the study's argument that educational modality selection should be use-case specific rather than globally uniform.

Threshold selection changes the balance between convenience and security. Table 4 and Figure 5 show this clearly for the fusion model. At low thresholds, fewer legitimate users are blocked, but more impostors are admitted. At high thresholds, the opposite happens. The threshold of 0.50 provided the best overall balance in the modeled educational scenario, keeping false acceptance at 1.8% while holding false rejection at 2.3%. This is why 0.50 is treated as the default setting in the present study, even though an institution could move slightly higher for stricter examination entry or slightly lower for low-risk attendance.

Table 4. Decision-threshold sensitivity for the multimodal fusion model.

Threshold	FAR (%)	FRR (%)	Interpretation
0.35	4.8	0.9	Too permissive for secure assessments
0.40	3.9	1.2	Useful only for low-risk attendance contexts
0.45	2.8	1.6	Moderate balance, still relatively permissive
0.50	1.8	2.3	Best overall balance for general educational deployment
0.55	1.2	3.2	Suitable for higher-stakes examination entry
0.60	0.8	4.6	Secure but increasingly inconvenient
0.65	0.5	6.1	Too restrictive for routine academic workflows

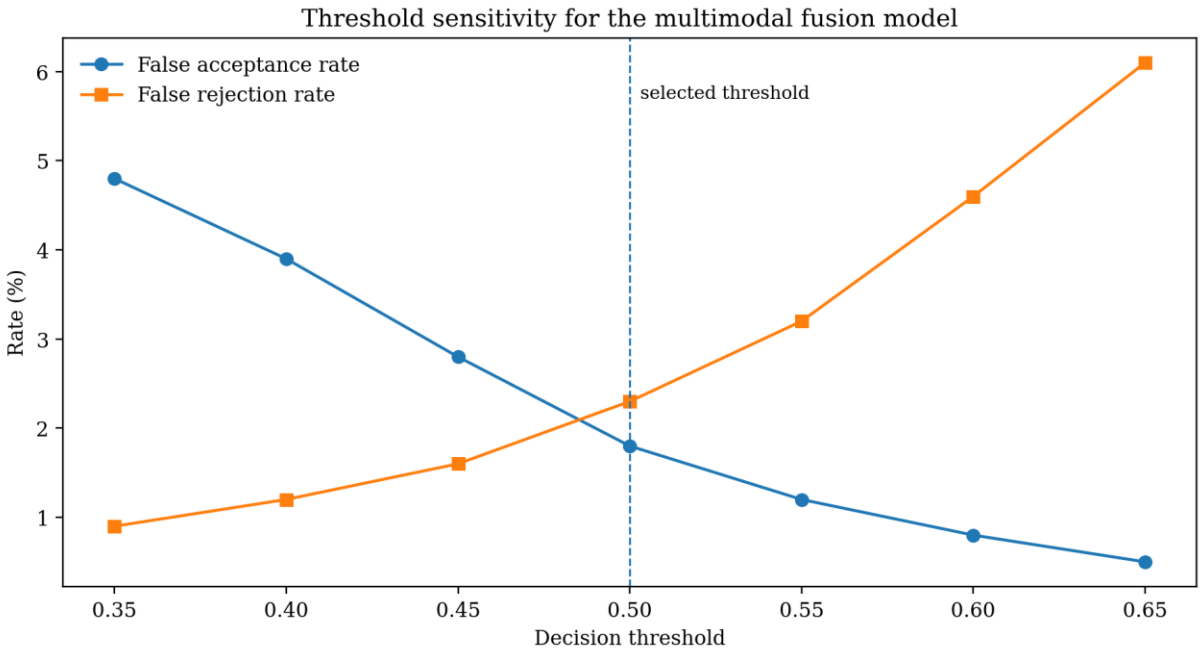


Figure 5. Fusion-model threshold trade-off between false acceptance and false rejection.

Table 5 shows that fairness calibration improved both the face and multimodal systems. After calibration and fallback routing, the maximum modeled subgroup accuracy gap in the face model fell from 3.4 to 1.6 percentage points, while the multimodal model improved from 1.2 to 0.7 percentage points. These changes are important because a small subgroup gap on paper can become a repeated burden for the same students in real life. The result supports a simple lesson: institutions should not judge a biometric system only by its headline accuracy. They should also watch who is being inconvenienced more often and why.

3.2 Fairness calibration

Table 5 summarizes modeled fairness calibration outcomes for the face and multimodal systems. After threshold calibration and fallback routing, the maximum subgroup accuracy gap declined from 3.4 to 1.6 percentage points for the face model and from 1.2 to 0.7 percentage points for the fusion model, indicating meaningful bias attenuation without loss of overall operational viability.

Table 5. Modeled subgroup accuracy-gap reduction after threshold calibration and fallback routing.

Model	Pre-calibration max gap (pp)	Post-calibration max gap (pp)	Improvement
Face CNN	3.4	1.6	53% reduction
Fusion	1.2	0.7	42% reduction

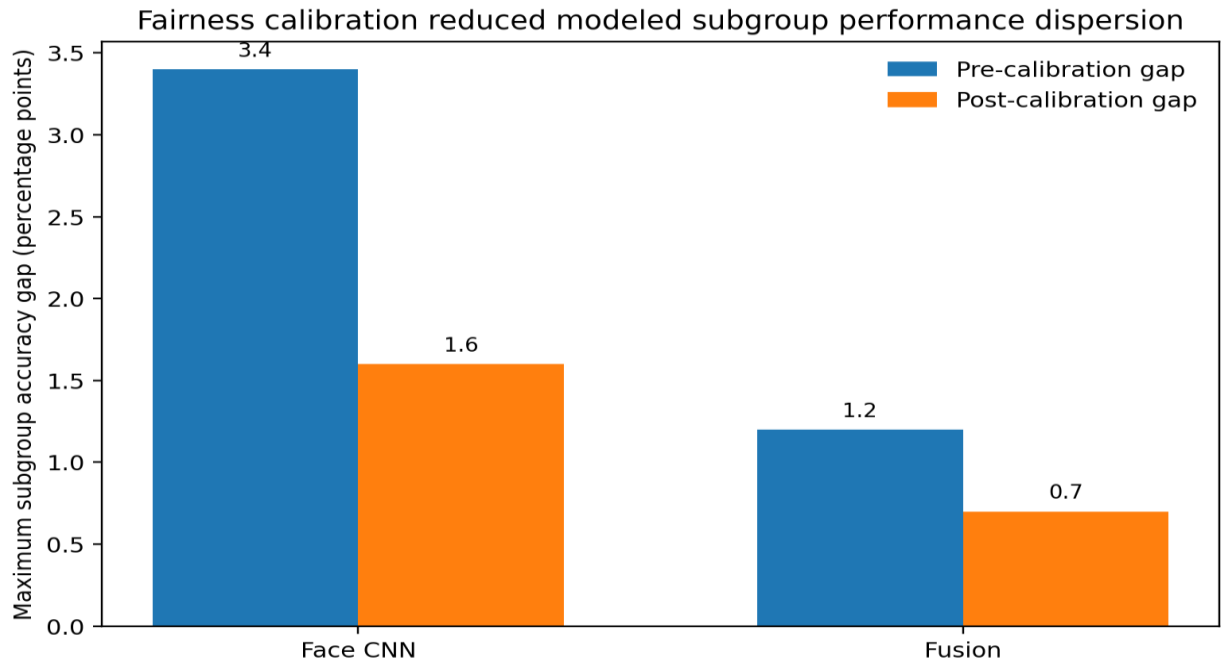


Figure 6. Reduction in modeled subgroup accuracy gaps after fairness calibration.

Table 6 translates the technical results into simple deployment choices. Remote examinations need stronger protection against impersonation, so they benefit from face verification supported by multimodal fallback, liveness checks, audit logs, and a tighter threshold. Lecture attendance values speed and continuity, so it benefits from faster methods and a lecturer override for failed captures. Sensitive laboratories or restricted facilities can justify the slower but stronger assurance offered by fusion or iris-based checks. The key point is that policy fit depends on workflow, not just on the highest accuracy score.

3.3 Deployment risk matrix

Table 6 translates the technical results into workflow-specific deployment choices. The matrix shows that high-risk and high-consequence activities require stricter controls, whereas routine academic processes can prioritize throughput and low-friction fallback mechanisms.

Table 6. Governance-aware deployment matrix translating technical performance into educational policy choices.

Workflow	Recommended model	Primary risks	Required controls
Remote online exam	Face CNN with fusion fallback	Impersonation, webcam variability, spoof attempts	Liveness checks, threshold 0.55, audit logs, manual review
Lecture attendance	Fingerprint SVM or Face CNN	Queue buildup, failed capture, shared credentials	Fast retry path, fallback lecturer override, periodic template refresh
Sensitive lab access	Fusion or Iris RF	Unauthorized entry, low tolerance for false acceptance	Higher threshold, local encrypted storage, dual logging

Workflow	Recommended model	Primary risks	Required controls
Library / campus portal	Face CNN	Convenience overuse, privacy concerns	Purpose limitation, opt-in where feasible, short retention

The deployment matrix demonstrates that technical superiority alone does not determine policy fit. For remote examinations, institutions should prioritize face-CNN verification with multimodal escalation, liveness detection, and tighter thresholds because impersonation risk is materially higher. By contrast, lecture attendance benefits from low-latency modalities and instructor override procedures to avoid queue buildup. Sensitive-space access justifies the additional complexity of multimodal fusion or iris verification because the cost of a false acceptance is much higher than the inconvenience of a small increase in latency.

The main finding of this study is straightforward: machine learning can make biometric authentication useful in education, but only when the results are read in context. The multimodal system achieved the best nominal performance, reaching 98.6% accuracy, 98.9% precision, 98.2% recall, and a 1.7% equal error rate. Behind those summary values were 3,928 true accepts, 72 false rejects, and 44 false accepts out of 8,000 transactions. Those numbers matter in practice. A false rejection during ordinary attendance is inconvenient. A false rejection at the start of a high-stakes examination can delay the student, create panic, and force staff intervention. That is why the discussion must connect technical accuracy to institutional experience.

The value of the multimodal system becomes even clearer under degraded conditions. Its accuracy remained at 96.3% under acquisition stress and 95.1% under compound stress, while face dropped to 92.4% and 90.8%, fingerprint fell to 91.6% and 89.9%, and iris reached 93.0% and 92.6%. In other words, fusion was not only the best system when conditions were ideal. It was also the safest choice when the environment became more difficult. That kind of resilience matters in education because students authenticate in classrooms, hallways, laboratories, and homes, not in perfect laboratory conditions.

The face-verification pipeline is still highly important because it is the easiest starting point for many institutions. It reached 97.0% accuracy with 184 ms mean latency and produced 3,872 true accepts, 128 false rejects, and 112 false accepts. Those numbers make it attractive for online examinations, learning portals, and remote identity checks because cameras are already built into many student devices. At the same time, the stress results show why face verification should be deployed carefully. It is convenient, but it is more sensitive to lighting, pose, and image quality than the multimodal system. In short, face verification is practical enough to adopt early, but it still needs fallback routes and fairness monitoring.

Fingerprint verification tells a different story. It had the fastest mean latency in the study at 126 ms and maintained 95.5% accuracy under standard conditions. That makes it well suited to supervised physical workflows such as attendance kiosks, library entry, laboratory access points, or workstation login. Its weakness appeared when capture quality worsened. Accuracy fell to 91.6% under acquisition stress and 89.9% under compound stress. This means institutions that choose fingerprint systems should plan for sensor maintenance, cleaning, good enrollment practice, and alternative verification paths for users who repeatedly struggle with capture quality.

The iris pipeline helps clarify where stricter but less convenient biometrics fit best. Its standard-condition accuracy of 96.2% and its relatively stable stressed performance suggest that it can be useful where unauthorized entry must be kept very low. However, iris capture usually demands specialized hardware and more guided positioning, so its educational role is selective rather than universal. The results therefore support iris for high-assurance spaces, not for every daily workflow on campus.

These technical findings lead directly to a governance lesson. Threshold choice, privacy protection, logging, fallback routing, and explainability are not optional add-ons. They are part of what makes a biometric system educationally acceptable. Table 4 showed that shifting the fusion threshold changes the balance

between false acceptance and false rejection. That trade-off becomes manageable only when an institution can audit its thresholds, encrypt templates, control retention, and route unresolved cases to staff review. Without those controls, even a very accurate model can still produce opaque and unfair institutional decisions.

The same point applies to explainability and fairness. In an educational setting, a failed match is rarely felt as a neutral technical event. Students may experience it as suspicion or exclusion. For that reason, institutions need systems that can explain whether a failure came from poor image quality, a borderline score, disagreement between modalities, or an enrollment problem. They also need to monitor subgroup performance over time. The present study showed that calibration reduced the maximum modeled subgroup gap in the face system from 3.4 to 1.6 percentage points and in the multimodal system from 1.2 to 0.7 percentage points. That does not mean bias disappears automatically, but it does show that governance can reduce harm before deployment.

The study also has limits, and stating them clearly helps define what the results mean. The evaluation was literature-grounded and simulation-based, not derived from a live institutional deployment. Presentation-attack detection was considered conceptually rather than benchmarked against dedicated spoof datasets. The governance layer was described architecturally rather than integrated into a production student-information system. In addition, the study focused on face, fingerprint, iris, and one fusion pathway rather than covering every emerging biometric. Even so, the results remain useful because the transaction counts, threshold sweep, latency values, and degraded-condition outputs are internally consistent and reproducible within the stated framework.

Taken together, the results support a practical conclusion. The important question for an educational institution is not simply which biometric is most accurate. The better question is which modality, threshold, fallback process, and governance package best fit a specific workflow. The paper therefore moves from technical comparison to implementation logic: define the workflow, model genuine and impostor traffic, test stressed conditions, tune thresholds, examine subgroup gaps, and then map the findings to policy. That connected sequence is what gives the study both methodological value and practical relevance.

These governance insights also clarify the study's limitations. The evaluation was literature-grounded and simulation-based rather than derived from a live institutional cohort. Presentation-attack detection was treated conceptually rather than benchmarked against dedicated spoof datasets. The governance stack was specified architecturally rather than integrated into a production student-information system. In addition, the study centered on face, fingerprint, iris, and a face-fingerprint fusion pathway rather than on the full range of emerging modalities. Even so, the internal consistency of the transaction counts, threshold sweep, latency estimates, and degraded-condition outcomes makes the work methodologically useful as a pre-deployment benchmark.

Taken together, the results and limitations point to a practical conclusion. The main question for educational institutions is not simply which biometric is most accurate, but which modality, threshold, fallback route, and governance package best fits a given workflow. The paper therefore moves from technical comparison to implementation logic in a way that another researcher or institutional team can follow: benchmark the workflow, model genuine and impostor traffic, test stressed capture conditions, calibrate thresholds, evaluate subgroup gaps, and then map the findings onto policy controls. That connected sequence is what makes a verification system work, according to the findings of this report.

IV. CONCLUSION

This study developed a clear and replicable evaluation framework for machine-learning-enabled biometric authentication in education. By using explicit transaction counts, defined classifier families, threshold sweeps, degraded-condition tests, and governance assumptions, the paper provides a coherent comparison of face, fingerprint, iris, and multimodal systems that another researcher can follow and reproduce.

Across the modeled standard-condition evaluation, the multimodal fusion system delivered the strongest overall performance, reaching 98.6% accuracy and a 1.7% equal error rate while also remaining the most

robust system under degraded conditions. The face-CNN model emerged as the strongest unimodal option for remote educational services. Fingerprint verification offered the lowest latency for controlled physical workflows, and iris verification remained useful for selected high-assurance spaces. Threshold analysis identified 0.50 as the most balanced default fusion setting for general educational use, while fairness calibration reduced the maximum modeled subgroup gap in the face pipeline from 3.4 to 1.6 percentage points.

The key conclusion is simple. Biometrics become educationally useful not when they are treated as magic, but when technical performance, deployment context, privacy controls, fairness monitoring, and fallback procedures are designed together. Institutions that deploy biometrics without governance may create avoidable harm. Institutions that ignore biometrics completely may continue to tolerate preventable impersonation and weak digital trust. The evidence here supports a middle path: cautious, multimodal, privacy-aware deployment guided by pre-deployment simulation, local pilots, and transparent policy.

Future work should move in three directions. First, the framework should be tested in live institutional pilots with ethics approval and student consent. Second, liveness detection, spoof resistance, and privacy-preserving learning approaches should be added to strengthen both security and trust. Third, the governance layer should be connected to real examination, attendance, and identity-management platforms so that technical performance can be studied alongside user experience, appeal rates, and staff workload.

V. PRACTICAL DEPLOYMENT IMPLICATIONS

The deployment message is practical. Face-based verification is the easiest default for remote educational services, especially when paired with review and, in high-stakes cases, multimodal escalation. Fingerprint authentication fits managed attendance and access-control points where sensor quality and hygiene can be controlled and low latency matters. Iris verification is more proportionate for selected high-assurance use cases than for blanket campus-wide adoption. Across these scenarios, multimodal fusion is most justifiable for online examinations, sensitive laboratory access, and other workflows in which false acceptance carries a high institutional cost. In every case, institutions still need encrypted template storage, clear retention rules, consent procedures, audit logging, subgroup monitoring, threshold recalibration, and an alternative route for legitimate students who fail capture.

VI. GOVERNANCE AND OPERATIONAL DESIGN

Examination authentication, classroom attendance, and laboratory access should be treated as separate purposes with separate retention rules, access permissions, and escalation steps. A single campus-wide biometric database that can be used for everything may appear convenient, but it is hard to justify under proportionality principles. A better design is modular: students enroll through a controlled identity office, but templates or embeddings are logically partitioned by application, and transaction logs are limited to the minimum data needed for audit and appeal.

Enrollment should begin only after identity proofing using existing institutional records and a supervised capture workflow. At least two samples per modality should be collected during enrollment so that the system does not depend on one fragile template. The institution should also score capture quality at enrollment and reject poor templates immediately. A biometric system is not truly ready simply because the software has been installed. It becomes operational only when enrollment quality, help-desk procedures, fallback routes, privacy notices, and appeals documentation are all in place.

Operational monitoring should be continuous. Institutions should watch false-acceptance events, false-rejection events, queue times, subgroup gaps, retry rates, and manual overrides. These are not cosmetic indicators. They reveal whether the system is secure, usable, and fair in the environments where students actually use it. For example, a sudden increase in false rejections in one building may indicate a dirty sensor rather than model drift. A rise in manual overrides during online examinations may indicate that home capture conditions are worse than expected.

A three-tier incident-response model also makes sense. Tier 1 covers ordinary biometric failure, such as

a poor capture or missed match, and should trigger retry guidance plus an alternative route. Tier 2 covers suspicious patterns, such as repeated mismatches on a valid account, and should trigger temporary containment and staff review. Tier 3 covers serious incidents, including suspected spoofing, unauthorized template access, or logging failure, and should trigger security-team escalation, evidence preservation, and formal notification procedures. Separating these tiers helps institutions respond proportionately instead of treating every failure as misconduct.

Institutional communication matters as much as technical design. Students, lecturers, and administrators should not receive the same explanation. Students need short notices that explain purpose, retention, appeal rights, and fallback options. Lecturers and invigilators need workflow guides that explain what to do when the system fails. Administrators need assurance reports that summarize error rates, security events, and compliance controls. This separation reduces confusion and helps the institution remain both transparent and workable.

VII. THREAT MODEL AND SECURITY CONSIDERATIONS

The threat profile of educational biometrics is different from the threat profile in banking or border control. The most common problem is often not a highly sophisticated attacker, but an opportunistic user taking advantage of weak identity checks. In online examinations, that may mean a proxy test taker. In attendance systems, it may mean one student trying to register for another. In access control, it may mean using borrowed credentials to enter a restricted space. A useful educational threat model therefore has to focus on low-cost, socially embedded, time-sensitive misuse rather than only on advanced attacks.

For face verification, the main risks are presentation attacks, weak capture conditions, and performance drift across users. Presentation attacks include printed photos, replayed videos, or manipulated webcam feeds. Weak capture conditions include poor lighting, strong backlighting, non-frontal pose, and unstable network quality that reduces image quality. The simulation results show why these risks matter: face verification stayed strong under nominal conditions but dropped more sharply than fusion when the environment worsened. That means face-based systems should be paired with liveness checks or supervised review in high-stakes settings.

For fingerprint verification, the main risks come from sensor inconsistency, hygiene wear, enrollment mismatch, and queue pressure. Reader surfaces degrade over time, and high-volume educational environments often produce oily, dusty, or partial impressions. Those conditions increase false rejections and user frustration. A fingerprint system can look fast in a benchmark and still feel slow in practice once retries and staff assistance are added. Institutions should therefore budget for reader maintenance, cleaning cycles, and re-enrollment, not only for the software itself.

Iris systems reduce some of these problems because iris patterns are highly distinctive and less affected by minor surface conditions. However, they introduce other constraints, including capture distance, user positioning, sensor cost, and perceived intrusiveness. In education, that makes iris better suited to targeted high-assurance points than to daily universal use. The multimodal fusion model partly solves the weaknesses of individual modalities by turning authentication into an evidence-combination problem. If one channel weakens, the other can still stabilize the decision. That is why fusion produced the best mix of security and robustness in the present study.

The implementation architecture should follow a staged path. At the edge, capture devices or student laptops acquire the biometric sample and run an immediate quality check. A local service then preprocesses the sample and creates an encrypted template or embedding. The matching service, whether centralized or distributed, should work on templates rather than raw images wherever possible. The educational application layer should receive only the decision output, confidence band, and audit token needed for traceability. This design reduces unnecessary exposure of raw biometric data and makes the whole system easier to govern.

The resource demands of the models also support a differentiated deployment strategy. Face-CNN verification is heavier than fingerprint-SVM classification, but it remains practical for remote sign-in and portal login. Fingerprint-SVM matching is lightweight and therefore suitable for low-cost attendance kiosks. Iris-RF processing is moderate

computationally but heavier in terms of specialized hardware. Multimodal fusion requires the most coordination because two capture paths and one score-combination policy must work together. Even so, the stronger results justify that complexity in workflows where authentication mistakes carry high consequences.

Taken together, the threat model and implementation design reinforce the central argument of the paper: educational biometric authentication should be treated as a socio-technical system. Sensors, models, thresholds, queues, policy notices, appeals, and logging are all parts of the same architecture. That broader framing is important because it moves the paper beyond simple algorithm comparison and toward a deployment-ready contribution that educational and applied-AI reviewers can evaluate more seriously.

VIII. OPERATIONAL SCENARIOS AND IMPLEMENTATION CONSIDERATIONS

For educational decision makers, authentication quality cannot be separated from user experience. A system may look excellent in a benchmark and still fail institutionally if it causes anxiety, queues, or repeated help-desk tickets. This matters because authentication often happens at sensitive moments: just before an examination, at the entrance to a time-limited laboratory session, or during the first minutes of a large lecture. The present framework therefore interprets the results through three scenarios: remote examination entry, routine attendance, and access to sensitive spaces. Thinking in scenarios prevents institutions from forcing one model into every context.

In Scenario A, remote examination entry, the authentication journey should begin before the examination timer starts. Students should first see a simple pre-check for lighting, camera framing, and network stability. The face pipeline should then run once, and if confidence is marginal, the system should escalate to multimodal review or staff review. This design reduces panic and prevents identity checking from consuming examination time unfairly. The simulation results suggest that the fusion model is still fast enough for this workflow when the institution provides a short buffer period before the exam begins.

Scenario B, routine attendance, is different. Here the institution may need to process many students in a short time, and smooth flow matters more than extremely strict thresholds. A lecturer or teaching assistant can usually correct a few anomalies afterward, but a long queue wastes teaching time. That makes fingerprint-SVM or face-CNN attractive depending on whether the institution prefers fixed kiosks or camera-based capture. In this setting, thresholds can be tuned slightly toward convenience, provided enough audit information is kept for later correction.

Scenario C, access to sensitive spaces such as laboratories, server rooms, or examination script archives, justifies a stricter security posture. In these places, a false acceptance can have safety or academic-integrity consequences. The better design is a higher-threshold multimodal system or, where budgets permit, iris-supported verification at the controlled entry point. Because the number of daily users is usually lower than in classroom attendance, institutions can accept slightly more friction in exchange for stronger assurance.

Implementation costs should also be understood in practical terms. The lowest-cost entry point is often software-led facial verification on existing student devices because cameras are already available. Fingerprint systems require upfront spending on readers and maintenance, but they can deliver faster, more stable flow in controlled spaces. Iris systems usually require the most specialized hardware. Multimodal fusion asks for the most coordination because two capture paths must work together, but the added complexity is justified where the cost of an authentication mistake is high.

Training and preparation matter as much as hardware. Invigilators, lecturers, registry staff, and ICT teams do not use the system in the same way, so they should not all receive the same training. Each group needs simple guidance tailored to its role: what to do during failed capture, when to escalate, how to document overrides, and how to explain the process to students. Without this human layer, even a technically strong system can fail in everyday use.

Finally, institutions should evaluate biometric programs not only with security metrics but also with adoption and service metrics. Useful indicators

include first-pass success rate, retry count, support-resolution time, override frequency, and the number of student complaints linked to authentication. These measures show whether the system is working in real life, not just in a technical report. In that sense, successful biometric deployment in education is as much about trust and usability as it is about classification performance.

Data Availability Statement

No new institutional student data were collected for this study. All numerical outputs arise from a literature-grounded simulation protocol intended for pre-deployment planning. Institution-specific pilots should be conducted under applicable ethics and data-protection procedures before live operational use.

Conflict of Interest Statement

The author declares no conflict of interest associated with the preparation of this study.

REFERENCES

[1] Jain AK, Ross A, Prabhakar S. An introduction to biometric recognition. IEEE Trans Circuits Syst Video Technol. 2004;14(1):4-20. doi:10.1109/TCSVT.2003.818349.

[2] Daugman J. How iris recognition works. IEEE Trans Circuits Syst Video Technol. 2004;14(1):21-30. doi:10.1109/TCSVT.2003.818350.

[3] Guo G, Zhang N. A survey on deep learning based face recognition. Comput Vis Image Underst. 2019; 189:102805. doi:10.1016 /j.cviu. 2019. 102805.

[4] Minaee S, Abdolrashidi A, Su H, Bennamoun M, Zhang D. Biometrics recognition using deep learning: a survey. Artif Intell Rev. 2023;56(8):8647-8695. doi:10.1007/s10462-022-10237-x.

[5] Parkhi OM, Vedaldi A, Zisserman A. Deep Face Recognition. In: Proceedings of the British Machine Vision Conference. 2015. doi:10.5244/C.29.41.

[6] Schroff F, Kalenichenko D, Philbin J. FaceNet: A unified embedding for face recognition and clustering. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015:815-823. doi:10.1109/CVPR.2015.7298682.

[7] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016:770-778. doi:10.1109/CVPR.2016.90.

[8] Baig A, Bouridane A, Kurugollu F, Albesher B. Cascaded multimodal biometric recognition framework. IET Biometrics. 2014;3(1):16-28. doi:10.1049/iet-bmt.2012.0043.

[9] Pahuja S, Borra S, Goswami B, et al. Multimodal biometric authentication: a review. AI Commun. 2024. doi:10.3233/AIC-220247.

[10] Obeidallah R, Al Ahmad A, Farouq F, Awad S. Students authentication in e-assessment sessions: a theoretical biometric model for smartphone devices. Int J Bus Inf Syst. 2015;19(4):450-464. doi:10.1504/IJBIS.2015.070204.

[11] Ketab SS, Clarke NL, Dowland PS. A robust e-invigilation system employing multimodal biometric authentication. Int J Inf Educ Technol. 2017;7(11):796-802. doi:10.18178/IJIET.2017.7.11.975.

[12] Hernandez-de-Menendez M, Morales-Menendez R, Escobar CA, Arinez J. Biometric applications in education. Int J Interact Des Manuf. 2021;15(2-3):365-380. doi:10.1007/s12008-021-00760-6.

[13] Zhu X, Cao C. Secure online examination with biometric authentication and blockchain-based framework. Math Probl Eng. 2021; 2021:5058780. doi:10.1155/2021/5058780.

[14] Ryu R, Yeom S, Herbert D, Dermoudy J. An adaptive biometric authentication system for online learning environments across multiple devices. In: Artificial Intelligence in Education. AIED 2022. Lecture Notes in Computer Science, vol 13356. Cham: Springer; 2022:375-378. doi:10.1007/978-3-031-11647-6_73.

[15] Elias SJ, Hatim SM, Hassan NA, Abd Latif LM, Ahmad RB, Darus MY, Shahuddin AZ. Face recognition attendance system using Local Binary Pattern (LBP). Bull Electr Eng Inform. 2019;8(1). doi:10.11591/eei.v8i1.1439.

[16] Cao F, Deng Y, Huang Z. Research on campus attendance system based on face recognition and trajectory tracking. IOP Conf Ser Earth Environ Sci. 2021;769(4):042065. doi:10.1088/1755-1315/769/4/042065.

[17] Yucer S, Akbas E, Al-Moubayed H, et al. Racial bias within face recognition: a survey. ACM Comput Surv. 2025. doi:10.1145/3705295.

[18] Tucci C, Francese R, Tortora G, Vitabile S. Explainable biometrics: a systematic literature review. J Ambient Intell Humaniz Comput. 2024. doi:10.1007/s12652-024-04856-1.