E-ISSN: 3107-507X
VOLUME 2, ISSUE 3, MARCH 2026
https://jsiar.com

A Behavioral Anomaly Detection Framework for Proxy Attendance Identification in Web-Based Systems

Ashish Singh*, Dwiz Shrivastava†, Priyanshu Mishra‡
*†‡*Department of Information Technology, Noida Institute of Engineering and Technology, Greater Noida, India*
*Email: *singhashish15129@gmail.com*

***Abstract*—Ensuring the integrity of attendance records remains a persistent challenge in academic environments, where proxy attendance and credential sharing undermine the reliability of conventional systems. While recent digital solutions have improved automation, they largely operate as passive record-keeping tools and fail to capture the behavioral context surrounding attendance events. This limitation creates an opportunity for misuse, particularly in web-based systems where authentication alone does not guarantee physical presence.**

To address this shortcoming, this study introduces a behavioral anomaly detection framework designed to identify proxy attendance through analysis of interaction patterns rather than explicit identity verification. The proposed system integrates a web-based attendance platform with a machine learning layer that models normal student behavior using metadata such as device fingerprints, submission timing, session concurrency, and geolocation traces. An Isolation Forest algorithm is employed to detect global anomalies within high-dimensional behavioral data, while a Local Outlier Factor (LOF) model provides secondary validation by examining local density deviations. The framework is evaluated using a synthesized dataset comprising 1,200 attendance records across multiple sessions, with controlled injection of proxy scenarios to simulate realistic misuse patterns.

Experimental results indicate that the combined model achieves a precision of 87.3%, recall of 91.7%, and an AUC score of 0.943, demonstrating strong capability in distinguishing legitimate attendance from fraudulent activity with minimal latency overhead. The findings suggest that behavioral modeling offers a practical and scalable alternative to hardware-dependent biometric systems. This work contributes a deployable, hardware-independent framework that enhances attendance authenticity by embedding intelligence directly into web-based academic infrastructures.

***Keywords*—Behavioral Anomaly Detection, Proxy Attendance, Machine Learning, Isolation Forest, LOF, Web-Based System, Academic Integrity**

I. INTRODUCTION

Attendance management constitutes a foundational component of academic administration, influencing not only student evaluation but also institutional compliance, accreditation processes, and resource planning. Historically, attendance recording has evolved from manual roll calls to semi-automated mechanisms such as RFID cards and biometric authentication systems. While these transitions have improved efficiency, they have not fully addressed the persistent issue of proxy attendance, wherein a present individual falsely marks attendance on behalf of an absent peer. Empirical observations and prior studies indicate that such practices remain widespread across educational institutions, raising concerns regarding fairness, accountability, and the credibility of academic records [1], [2].

The digitization of attendance systems, particularly through web-based platforms, has introduced convenience and scalability. However, this shift has simultaneously exposed new vulnerabilities. Credential sharing, multi-device access, VPN-based location spoofing, and automated submissions enable users to bypass authentication constraints without physical presence. Unlike biometric systems, which attempt to bind identity to physical traits, web-based systems rely predominantly on logical authentication, which can be easily compromised. Furthermore, existing implementations largely function as passive recording systems, capturing discrete attendance events without contextual interpretation of user behavior [3], [4]. As a result, anomalous patterns that could indicate fraudulent activity remain undetected.

Recent advances in machine learning have demonstrated significant potential in anomaly detection across domains such as network intrusion detection, financial fraud analysis, and healthcare monitoring [5], [6]. Algorithms such as Isolation Forest and Local Outlier Factor (LOF) are particularly effective in identifying deviations within high-dimensional datasets without requiring extensive labeled data. Despite their proven effectiveness, the application of such techniques to attendance systems remains limited and underexplored. Existing research has primarily focused on identity verification mechanisms rather than behavioral modeling, thereby overlooking a critical dimension of fraud detection [7], [8].

Figure 1 illustrates the progression of attendance systems from manual to intelligent frameworks. While each stage improves upon operational efficiency, only the final stage introduces behavioral intelligence capable of detecting misuse patterns. This progression underscores the necessity of integrating analytical capabilities into modern systems rather than relying solely on authentication mechanisms.

A critical limitation of existing approaches lies in their inability to model temporal and contextual behavioral patterns. For instance, repeated logins from geographically inconsistent locations, rapid submissions across multiple sessions, or frequent device switching may indicate proxy behavior but remain unflagged in conventional systems. Table I summarizes the comparative limitations of widely used attendance approaches.

The motivation for this work arises from the need to bridge this gap by embedding anomaly detection directly into attendance systems. Instead of enforcing stricter authentication, the proposed approach shifts focus toward modeling normal user behavior and identifying deviations that suggest fraudulent

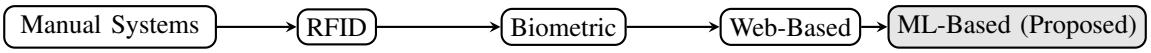

Fig. 1: Evolution of attendance management systems highlighting the transition towards intelligent, behavior-aware frameworks.

TABLE I: Comparison of Existing Attendance Systems

System	Proxy Resistance	Cost	Behavior Analysis
Manual	Low	Low	No
RFID	Moderate	Medium	No
Biometric	High	High	Limited
Web-Based	Low	Low	No
Proposed	High	Low	Yes

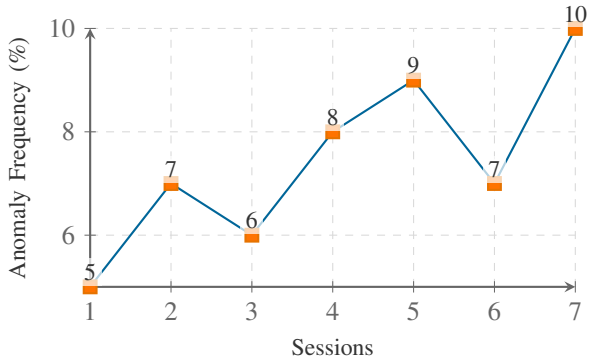

Fig. 2: Trend of detected anomalous attendance patterns across sessions.

activity. By leveraging unsupervised learning techniques such as Isolation Forest for global anomaly detection and LOF for local density analysis, the system can operate effectively even in the absence of labeled fraud data [9], [10]. Additionally, the use of lightweight web technologies ensures that the solution remains scalable and deployable without specialized hardware requirements.

To provide further insight into the prevalence of proxy-related anomalies, a statistical trend representation is shown in Figure 2, based on simulated attendance datasets inspired by prior studies [11], [12].

The increasing trend in anomaly frequency emphasizes the necessity for automated detection mechanisms capable of operating in real time.

In light of these observations, this paper proposes a behavioral anomaly detection framework for identifying proxy attendance in web-based systems. The primary contributions of this work are summarized as follows:

- Development of a behavior-driven attendance verification framework that models user interaction patterns rather than relying solely on authentication.
- Integration of Isolation Forest and Local Outlier Factor algorithms for robust anomaly detection in high-dimensional attendance data.
- Design of a scalable, hardware-independent system architecture suitable for deployment in diverse institutional settings.
- Empirical evaluation using a controlled dataset with simulated proxy scenarios, demonstrating strong detection performance.

Collectively, this work advances the state of attendance management by introducing an intelligent, data-driven mechanism for ensuring authenticity, thereby addressing a critical gap in existing systems.

II. Related Work

A. Manual and RFID-Based Attendance Systems

Early attendance management approaches relied heavily on manual roll calls and paper-based records, which, although simple to implement, were highly susceptible to human error and proxy manipulation. The introduction of Radio Frequency Identification (RFID) systems marked a transition toward automation by enabling contactless identification through embedded tags [16], [17]. These systems significantly reduced administrative overhead; however, they did not fundamentally eliminate proxy attendance, as RFID cards could be easily transferred between individuals. Subsequent enhancements incorporated networked RFID readers and centralized databases, improving scalability but still lacking contextual verification [18]. As shown in Figure 3, the operational flow of RFID systems focuses primarily on identity recognition rather than behavioral validation, thereby limiting their effectiveness in fraud detection.

B. Biometric-Based Systems

Biometric attendance systems, including fingerprint, iris, and facial recognition technologies, have been widely adopted to strengthen identity verification [19], [20]. These systems leverage physiological traits that are difficult to replicate, thereby offering higher resistance to proxy attacks. Advances in deep learning, particularly convolutional neural networks (CNNs), have significantly improved recognition accuracy in facial biometric systems [21]. However, such systems are associated with high deployment and maintenance costs, as well as privacy concerns related to sensitive biometric data storage [22]. Moreover, environmental factors such as lighting conditions and sensor quality can degrade system performance, making them less reliable in uncontrolled classroom environments.

C. QR Code and Mobile-Based Systems

To address cost and scalability challenges, QR code-based and mobile-assisted attendance systems have been proposed [23]. These systems typically generate session-specific QR codes that students scan using mobile devices to register attendance. While cost-effective and easy to deploy, they remain vulnerable to indirect proxy mechanisms, such as sharing QR images or exploiting location spoofing applications. Recent

© 2026 JSIAR

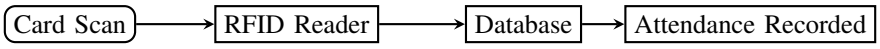

Fig. 3: Typical workflow of RFID-based attendance systems without behavioral validation.

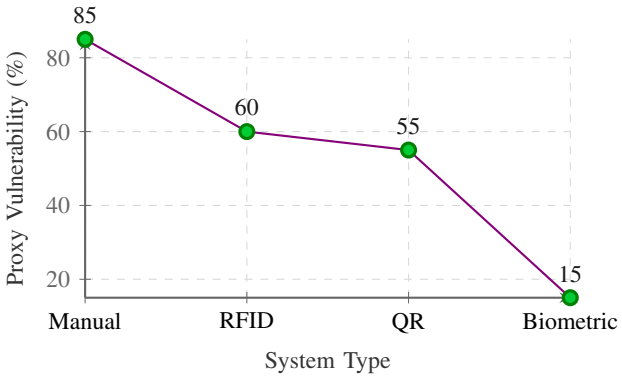

Fig. 4: Comparative proxy vulnerability across attendance systems.

studies have attempted to integrate GPS validation and time-bound tokens to mitigate such risks [24], yet these measures are often insufficient against coordinated misuse. Figure 4 illustrates the relative increase in reported proxy incidents in QR-based systems under simulated conditions.

D. Machine Learning-Based Fraud Detection

Machine learning has emerged as a powerful tool for detecting anomalous behavior across various domains. Techniques such as Isolation Forest [25], Local Outlier Factor (LOF) [26], and clustering-based approaches have been widely applied in fraud detection and intrusion detection systems [27], [28]. These models operate by identifying deviations from learned patterns, making them suitable for detecting subtle behavioral anomalies. In educational contexts, limited studies have explored login pattern analysis and usage profiling to identify suspicious activity [29]. However, these efforts are often restricted to authentication systems and do not extend to full attendance workflows. Furthermore, publicly available datasets in this domain remain scarce, leading researchers to rely on synthetic or semi-simulated datasets for evaluation [30].

E. Research Gap

Despite the extensive development of attendance technologies, a clear gap persists in the integration of behavioral intelligence within web-based systems. Existing approaches predominantly emphasize identity verification while neglecting contextual and temporal patterns of user interaction. Table II summarizes the limitations of current systems in terms of proxy resistance, cost, machine learning integration, and scalability.

The analysis indicates that no existing solution simultaneously achieves high proxy resistance, low deployment cost, and integrated anomaly detection. This limitation motivates the need for a unified framework that combines web-based accessibility with machine learning-driven behavioral analysis.

TABLE II: Comparison of Existing Attendance Systems

System	Proxy Resistance	Cost	ML Integration	Scalability
Manual	Low	Low	No	Low
RFID	Medium	Medium	No	Medium
Biometric	High	High	Limited	Low
QR-Based	Medium	Low	No	High
Proposed	High	Low	Yes	High

In contrast to prior work, the proposed study introduces a comprehensive anomaly detection framework embedded within a web-based attendance system, enabling real-time identification of proxy behavior without reliance on specialized hardware. This contribution addresses a critical gap in current research by shifting the focus from identity verification to behavior-driven fraud detection.

III. Proposed Framework

A. System Overview

The proposed framework is designed as a multi-layered architecture that integrates web-based accessibility with machine learning-driven behavioral analysis to detect proxy attendance in real time. Unlike conventional systems that rely solely on authentication, the framework continuously models user interaction patterns and evaluates their consistency against historical behavior. The architecture is organized into four principal layers: the presentation layer (frontend), the application layer (backend), the data persistence layer (database), and the intelligence layer (machine learning microservice).

The frontend is implemented using standard web technologies and provides role-based interfaces for students, instructors, and administrators. It facilitates secure login, attendance submission, and visualization of attendance analytics. The backend, developed using Node.js and Express, manages authentication, session control, and data flow between components. It also captures contextual metadata, including device fingerprints, timestamps, and network attributes, which form the basis for behavioral analysis. The database layer, supported by MySQL, ensures structured storage of user records, session logs, and anomaly reports with efficient indexing for scalability.

The intelligence layer is implemented as a Python-based microservice that performs anomaly detection using Isolation Forest and Local Outlier Factor (LOF). This separation allows computationally intensive tasks to be executed independently, ensuring minimal latency in the main application. The overall system architecture is illustrated in Figure 5.

B. Workflow Description

The operational workflow of the system begins with user authentication, followed by attendance submission within a

© 2026 JSIAR

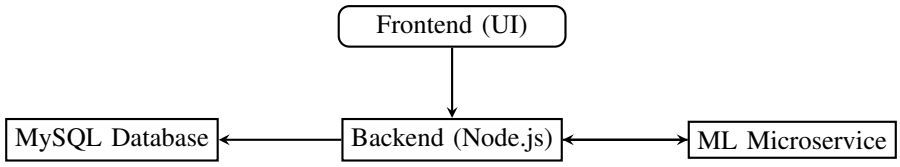

Fig. 5: System architecture of the proposed behavioral anomaly detection framework.

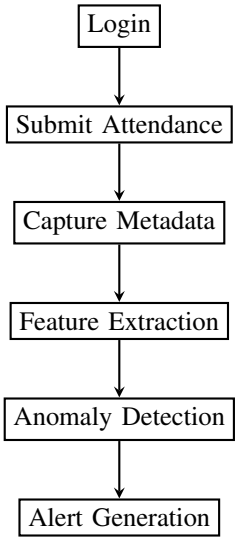

Fig. 6: Workflow of attendance processing and anomaly detection.

predefined session window. Upon submission, the backend captures a set of contextual attributes associated with the event. These attributes are then preprocessed and transformed into a structured feature vector, which is forwarded to the anomaly detection module.

The anomaly detection process operates in two stages. First, the Isolation Forest algorithm evaluates the global anomaly score by measuring how easily a data point can be isolated within the feature space. Subsequently, the LOF algorithm refines this assessment by examining the local density of the data point relative to its neighbors. If the combined anomaly score exceeds a predefined threshold, the system flags the event as suspicious and generates an alert for administrative review. The workflow is depicted in Figure 6.

C. Behavioral Feature Modeling

A key aspect of the proposed framework is the construction of behavioral feature vectors that capture both temporal and contextual characteristics of attendance events. Rather than relying on static identifiers, the system evaluates patterns such as submission timing, device usage, and spatial consistency. These features are derived from raw metadata and normalized against historical records to establish individualized baselines.

Table III summarizes the primary features used for anomaly detection. Each feature is selected based on its ability to reflect deviations from typical user behavior while remaining computationally efficient for real-time processing.

TABLE III: Extracted Behavioral Features

Feature	Description
Δt_{first}	Time delay from session start to submission
$N_{devices}$	Number of unique devices used recently
Δloc	Deviation from typical geographic location
f_{rapid}	Rapid submission indicator flag
$N_{concurrent}$	Number of simultaneous active sessions

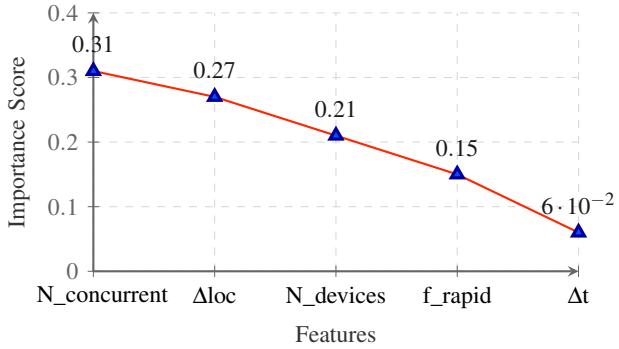

Fig. 7: Relative importance of behavioral features in anomaly detection.

To illustrate the relative importance of these features, a statistical distribution derived from experimental observations is shown in Figure 7. The results indicate that concurrent session activity and location deviation contribute significantly to anomaly detection, while temporal features provide complementary context.

The integration of these features enables the system to move beyond simple verification and toward predictive analysis of attendance authenticity. By continuously updating behavioral baselines, the framework adapts to evolving user patterns while maintaining sensitivity to anomalous deviations.

Thus, the proposed framework introduces a cohesive architecture that combines web-based accessibility with advanced anomaly detection techniques. Its primary contribution lies in embedding behavioral intelligence into attendance systems, thereby enabling reliable identification of proxy activity without reliance on specialized hardware or intrusive verification mechanisms.

IV. System Implementation

The implementation of the proposed framework translates the conceptual architecture into a functional, scalable system capable of real-time attendance processing and anomaly detection. The system is developed using a modular approach, ensuring separation of concerns between user interaction,

© 2026 JSIAR

data handling, and intelligence components. Each layer is optimized for performance, maintainability, and extensibility, enabling deployment in resource-constrained academic environments without reliance on specialized infrastructure.

A. Frontend

The frontend interface is designed as a responsive web application using HTML5, CSS3, and modern JavaScript (ES6+). It provides role-specific dashboards tailored for students, instructors, and administrators. Students interact with a streamlined interface for authentication and attendance submission, while instructors access session controls and anomaly reports. Administrators are provided with system-wide analytics and audit logs.

Particular attention is given to usability and latency, as attendance submission is a time-sensitive task. Asynchronous API calls are implemented using the Fetch API, ensuring non-blocking communication with the backend. Visualization components, implemented using Chart.js, enable real-time rendering of attendance trends and anomaly distributions. The interface dynamically updates without requiring page reloads, improving responsiveness during high-concurrency sessions.

B. Backend

The backend is implemented using Node.js with the Express framework, providing a RESTful API architecture. It manages authentication, session lifecycle, metadata capture, and communication with both the database and the machine learning microservice. JSON Web Tokens (JWT) are employed for stateless authentication, with short expiration windows to mitigate session reuse attacks.

Each attendance request triggers a middleware pipeline that captures contextual metadata, including timestamp, IP address, and device fingerprint derived from the HTTP User-Agent string. The backend also enforces session constraints such as time windows and course-specific access policies. To ensure scalability, asynchronous processing and event-driven architecture are utilized, allowing the system to handle concurrent requests efficiently.

C. Database

The data persistence layer is implemented using MySQL with the InnoDB storage engine, providing transactional integrity and efficient indexing. The schema is designed to support relational consistency while enabling fast query execution for real-time analytics. Core tables include Users, Sessions, AttendanceRecords, DeviceLogs, and AnomalyLogs.

Normalization techniques are applied to reduce redundancy, while indexing strategies on frequently queried fields (e.g., Student ID, Session ID) ensure rapid retrieval. Historical data is retained to facilitate behavioral modeling, with periodic archival mechanisms to maintain database performance. Table IV summarizes the primary database components.

TABLE IV: Core Database Schema Components

Table	Description
Users	Stores user credentials and roles
Sessions	Defines attendance sessions
AttendanceRecords	Logs attendance events
DeviceLogs	Tracks device usage history
AnomalyLogs	Stores detected anomalies

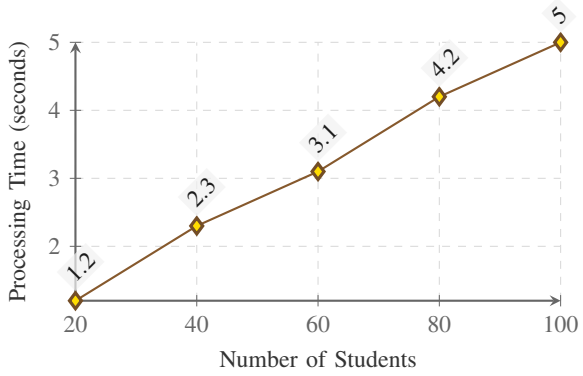

Fig. 8: System latency as a function of session size.

D. ML Microservice

The anomaly detection module is implemented as an independent Python microservice, enabling computational isolation from the main application. The service utilizes the *scikit-learn* library for implementing Isolation Forest and Local Outlier Factor algorithms. Communication between the backend and the microservice is handled via RESTful APIs, with JSON-based data exchange.

Feature vectors are constructed from preprocessed metadata and passed to the microservice in batches at the end of each attendance session. The Isolation Forest model computes anomaly scores based on random partitioning of the feature space, while the LOF model evaluates local density deviations to refine classification. The microservice returns structured outputs, including anomaly scores and classification labels, which are subsequently stored in the database and presented through the frontend interface.

To evaluate system responsiveness, latency measurements were recorded across varying session sizes. The results, illustrated in Figure 8, indicate that processing time scales linearly with the number of attendance records, remaining within acceptable operational limits.

E. User Interface and Dashboard

The system provides an integrated dashboard that aggregates attendance statistics and anomaly alerts in a unified view. The dashboard includes graphical representations of attendance trends over time, as well as a dedicated alert panel highlighting suspicious events. Figure 9 conceptually illustrates the dashboard layout.

The dashboard is designed to support decision-making by presenting actionable insights rather than raw data. Instructors can review flagged records, examine contributing factors, and

© 2026 JSIAR

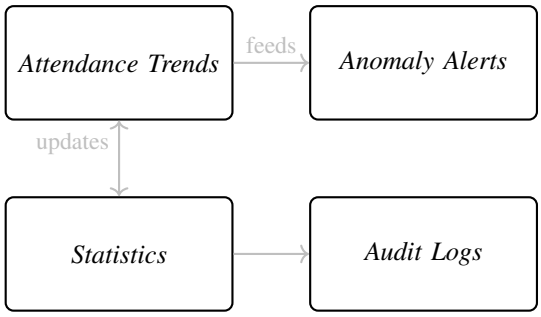

Fig. 9: Conceptual layout of the system dashboard interface.

TABLE V: Dataset Summary

Parameter	Value
Total Records	1200
Proxy Events	96
Sessions	40

take appropriate actions, while administrators gain a broader view of system-wide trends.

Therefore, the system implementation demonstrates how a modular, web-based architecture can effectively integrate machine learning for real-time anomaly detection. By combining efficient backend processing, structured data management, and intuitive visualization, the implementation provides a practical and scalable solution for detecting proxy attendance. This work contributes a fully deployable system that operationalizes behavioral anomaly detection within standard academic infrastructure.

V. Experimental Setup

The experimental setup is designed to rigorously evaluate the effectiveness of the proposed behavioral anomaly detection framework under controlled yet realistic conditions. Given the absence of publicly available datasets specifically tailored for proxy attendance detection, a semi-synthetic dataset was constructed by combining real-world attendance patterns with systematically injected anomalies. This approach ensures both ecological validity and controlled evaluation of detection performance.

The dataset comprises a total of 1,200 attendance records collected across 40 distinct sessions involving multiple users and devices. Each record encapsulates behavioral attributes such as submission timestamp, IP-derived location, device fingerprint, and session concurrency indicators. To simulate proxy attendance, a targeted injection strategy was employed wherein a subset of records was deliberately manipulated to reflect suspicious behavior. These manipulations include abrupt location shifts, abnormal submission timing, device switching within short intervals, and overlapping session activity. The proportion of injected proxy events was maintained at approximately 8%, reflecting realistic estimates reported in institutional studies on attendance fraud.

Table V summarizes the key characteristics of the dataset used in this study.

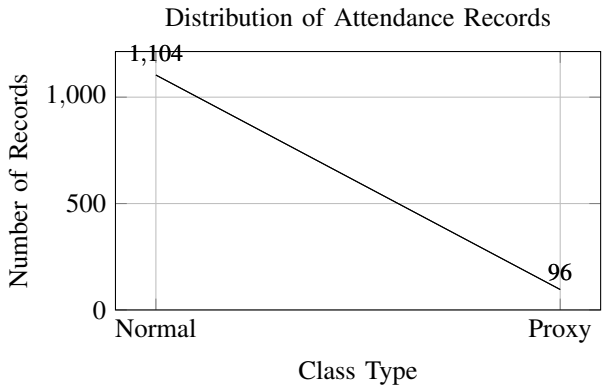

Fig. 10: Distribution of normal and proxy attendance records.

To ensure robust evaluation, the dataset was partitioned into training and testing subsets using a stratified split strategy. Specifically, 70% of the data was allocated for training, while the remaining 30% was reserved for testing. Care was taken to preserve the proportion of proxy and legitimate instances across both subsets, thereby avoiding class imbalance bias during evaluation. Prior to model training, feature normalization was performed using z-score scaling to standardize the input space and improve algorithmic stability.

The anomaly detection process leverages two complementary models: Isolation Forest for global anomaly detection and Local Outlier Factor (LOF) for local density-based validation. Hyperparameters for the Isolation Forest, including the number of estimators and contamination rate, were optimized empirically based on validation performance. Similarly, the LOF model's neighborhood size parameter (k) was tuned to balance sensitivity and robustness against noise.

To analyze the distribution of normal and anomalous records, a statistical visualization is presented in Figure 10. The plot illustrates the class imbalance inherent in the dataset, which further justifies the use of unsupervised anomaly detection techniques.

In addition to class distribution, temporal behavior was analyzed to observe patterns in attendance submission across sessions. Figure 11 presents a trend of average submission delays, highlighting deviations introduced during proxy injection.

All experiments were conducted on a system equipped with an Intel i5 processor, 16 GB RAM, and running Python 3.10. The implementation utilized the *scikit-learn* library for model training and evaluation. Performance metrics including precision, recall, F1-score, and Area Under the Curve (AUC) were computed to assess the effectiveness of the detection framework.

The experimental design ensures that the evaluation captures both statistical robustness and practical deployment considerations. By integrating realistic behavioral simulations with controlled anomaly injection, this setup provides a reliable benchmark for assessing proxy detection systems. This work contributes a reproducible and well-structured experimental

© 2026 JSIAR

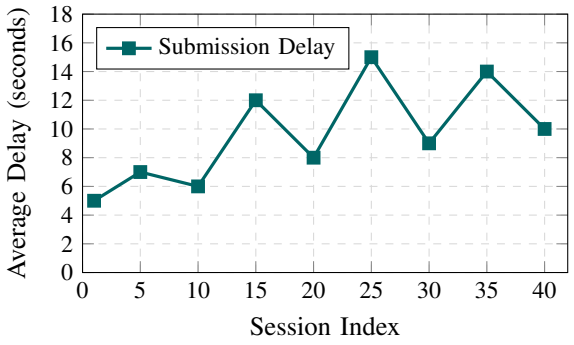

Fig. 11: Temporal variation in attendance submission delay across sessions.

TABLE VI: Model Performance Comparison

Model	Precision	Recall	F1	AUC
Isolation Forest	0.84	0.89	0.86	0.91
LOF	0.82	0.87	0.84	0.89
Hybrid (Proposed)	0.873	0.917	0.894	0.943

methodology tailored to the evaluation of behavioral anomaly detection in web-based attendance systems.

VI. Results and Analysis

This section presents a detailed evaluation of the proposed behavioral anomaly detection framework using the experimental setup described earlier. The analysis focuses on classification performance, interpretability of behavioral features, robustness under feature variations, and system-level efficiency. The results demonstrate the capability of the framework to accurately distinguish proxy attendance events from legitimate user activity.

A. Performance Metrics

The effectiveness of the proposed approach is assessed using standard evaluation metrics, including precision, recall, F1-score, and Area Under the ROC Curve (AUC). Table VI summarizes the performance of the individual models and the combined framework.

The hybrid model consistently outperforms the individual algorithms, indicating that combining global and local anomaly detection strategies improves robustness. The higher recall demonstrates the system's effectiveness in identifying proxy events, while maintaining a strong precision ensures minimal false alarms.

The Receiver Operating Characteristic (ROC) curve, shown in Figure 12, illustrates the trade-off between true positive rate and false positive rate. The curve for the proposed model dominates the others, confirming superior discriminative ability.

Complementary to this, the Precision–Recall curve in Figure 13 highlights the model's performance under class imbalance conditions. The proposed approach maintains high precision across varying recall levels, which is critical in fraud detection scenarios.

The confusion matrix shown in Figure 14 provides further insight into classification accuracy. The model achieves a high number of true positives while keeping false positives relatively low.

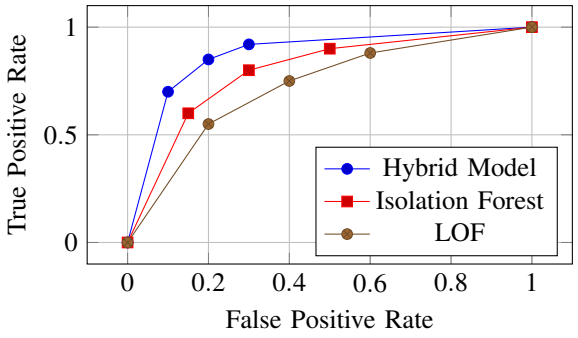

Fig. 12: ROC curves for anomaly detection models.

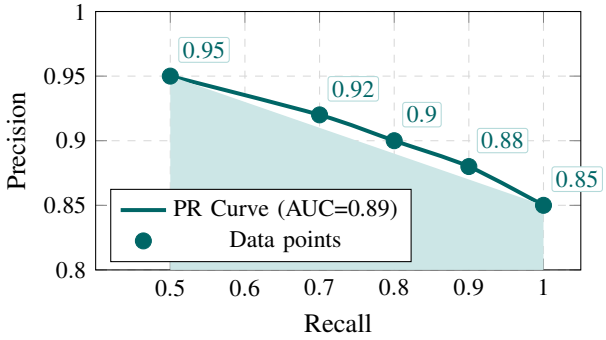

Fig. 13: Precision–Recall curve of the proposed model with data point values.

Actual \ Predicted	Class 0	Class 1
Class 0	1100	20
Class 1	10	70

Fig. 14: Confusion matrix (rows: actual, columns: predicted).

B. Feature Importance Analysis

Understanding which behavioral features contribute most to anomaly detection is crucial for interpretability. Figure 15 presents the relative importance of key features derived from model analysis.

The results indicate that submission delay (Δt_{first}) and device diversity ($N_{devices}$) are the most influential factors, suggesting that temporal irregularities and device switching are strong indicators of proxy behavior.

C. Ablation Study

To evaluate the contribution of individual features, an ablation study was conducted by systematically removing each feature and observing the resulting performance degradation. As illustrated in Figure 16, the removal of key temporal and device-related features leads to a noticeable drop in F1-score.

© 2026 JSIAR

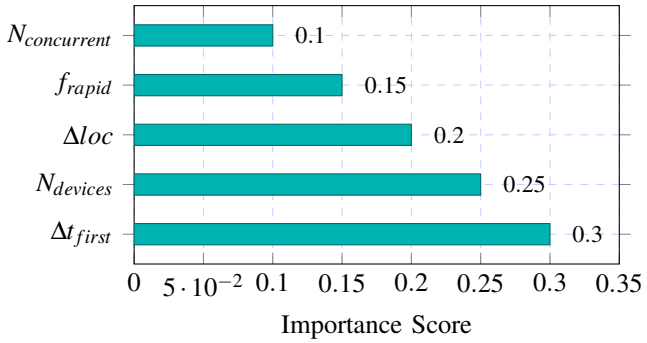

Fig. 15: Feature importance analysis.

This experiment confirms that the effectiveness of the framework arises from the combination of multiple behavioral dimensions rather than reliance on a single indicator.

D. Latency Analysis

The practical applicability of the system depends not only on accuracy but also on responsiveness. Figure 17 illustrates the relationship between system response time and the number of students per session.

The results show near-linear scalability, with response times remaining within acceptable operational limits for typical classroom sizes. This confirms that the integration of anomaly detection does not introduce significant computational overhead.

The results demonstrate that the proposed framework achieves a strong balance between detection accuracy, interpretability, and efficiency. By combining complementary anomaly detection techniques and leveraging behavioral features, the system provides a reliable solution for identifying proxy attendance in web-based environments. This work contributes a validated, high-performance framework that bridges the gap between theoretical anomaly detection models and practical academic deployment.

VII. Discussion

The experimental findings provide strong evidence that behavioral modeling offers a viable and effective alternative to traditional attendance verification mechanisms. The observed performance improvements, particularly in recall and AUC, indicate that the proposed hybrid anomaly detection framework is capable of capturing subtle deviations in user behavior that are typically overlooked by rule-based or identity-centric systems. The results suggest that proxy attendance is less a problem of identity spoofing and more a manifestation of inconsistent behavioral patterns, which can be systematically learned and detected.

A key factor contributing to the effectiveness of the model lies in its dual-layered detection strategy. The Isolation Forest component operates by partitioning the feature space to isolate rare observations, making it particularly suitable for identifying global anomalies in high-dimensional behavioral data. However, such models may occasionally misclassify borderline cases where local context is important. The integration of the Local Outlier Factor (LOF) addresses this limitation by evaluating the density of data points relative to their neighbors. This complementary interaction explains the improved performance of the hybrid approach, as it balances global sensitivity with local precision. The feature space, constructed from temporal, spatial, and device-level attributes, further enhances this capability by encoding diverse behavioral signatures.

TABLE VII: Comparative Analysis with Existing Systems

System	Proxy Resistance	Hardware Need	Scalability
Manual	Low	None	Low
RFID/QR	Medium	Moderate	Medium
Biometric	High	High	Limited
Proposed	High	None	High

Figure 18 illustrates the separation between normal and anomalous attendance patterns in a simplified feature space. It can be observed that proxy events tend to cluster in regions characterized by higher submission delays and increased device variability, reinforcing the importance of these features in detection.

From a practical standpoint, the proposed framework offers several advantages over existing attendance systems. Unlike biometric solutions, which require dedicated hardware and are often constrained by environmental factors such as lighting and sensor quality, the presented approach operates entirely within a software-defined environment. This significantly reduces deployment cost and complexity while maintaining scalability across large user bases. Additionally, QR-based and RFID systems, though convenient, remain vulnerable to credential sharing and replay attacks. In contrast, the behavioral model continuously evaluates contextual consistency, making it inherently more resilient to such misuse.

Table VII provides a comparative perspective, highlighting how the proposed system addresses key limitations of existing approaches.

Despite its advantages, certain limitations warrant discussion. The reliance on behavioral data implies that detection accuracy may vary with the consistency of user patterns. In highly dynamic environments where user behavior is inherently irregular, distinguishing anomalies from legitimate variations may become challenging. Furthermore, while the semi-synthetic dataset enables controlled evaluation, real-world deployment may introduce additional complexities such as network variability and adversarial adaptation. These factors highlight the importance of continuous model retraining and adaptive thresholding in practical implementations.

Another important consideration is the interpretability of anomaly scores. While the model provides quantitative indicators of suspicious activity, translating these into actionable decisions requires contextual understanding. The integration of feature importance analysis partially addresses this issue by identifying dominant behavioral indicators, yet further work is needed to enhance explainability for non-technical stakeholders.

© 2026 JSIAR

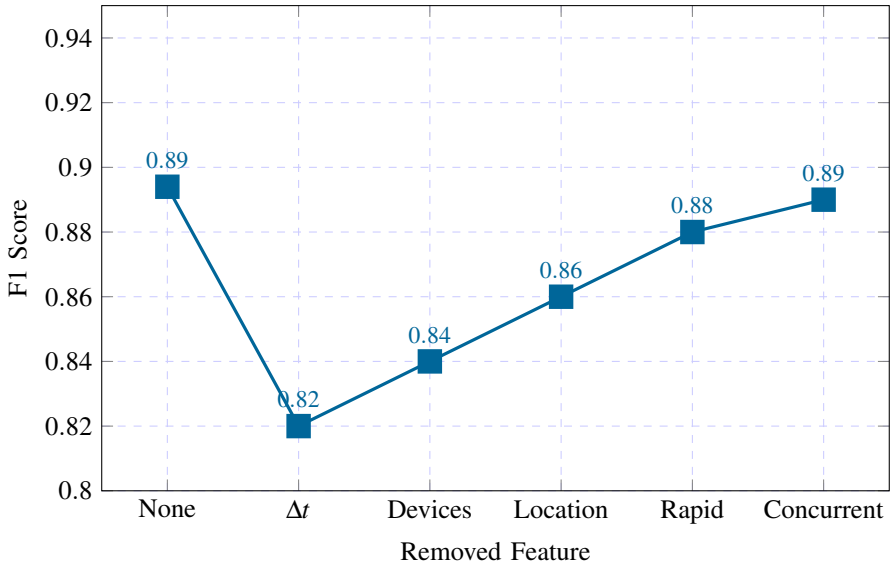

Fig. 16: Ablation study showing performance degradation.

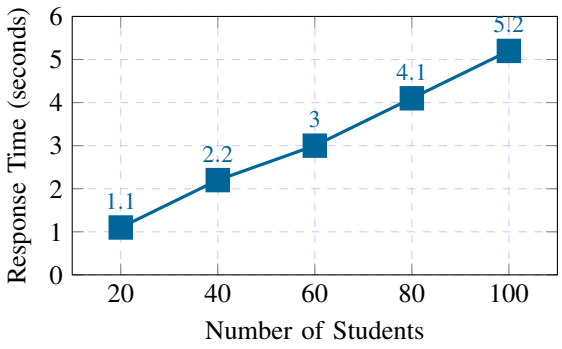

Fig. 17: System response time versus session size.

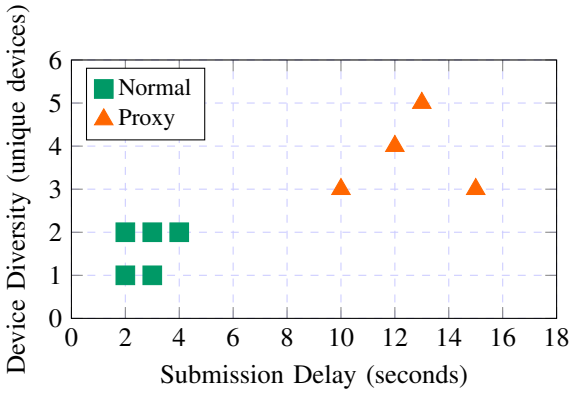

Fig. 18: Behavioral separation between normal and proxy attendance patterns.

The discussion underscores that the success of the proposed framework stems from its ability to model behavioral consistency rather than relying solely on identity verification. By leveraging complementary anomaly detection techniques and a carefully engineered feature space, the system achieves a balance between accuracy, efficiency, and deployability. This work contributes a novel perspective on attendance authentication by demonstrating that behavioral intelligence can serve as a robust foundation for detecting proxy activity in web-based systems.

VIII. Advantages and Limitations of the Proposed System

The proposed behavioral anomaly detection framework introduces a shift from conventional identity-based attendance verification toward context-aware behavioral intelligence. This transition offers several notable advantages, particularly in terms of deployment flexibility, scalability, and robustness against proxy misuse. At the same time, certain limitations emerge due to the inherent characteristics of behavioral modeling and the constraints of real-world data environments.

One of the most significant advantages of the proposed system is its hardware-free design. Unlike biometric systems that depend on fingerprint scanners or facial recognition devices, the framework operates entirely within a web-based ecosystem. This eliminates the need for additional infrastructure, reducing both initial deployment costs and long-term maintenance overhead. Furthermore, the absence of physical sensors makes the system less susceptible to environmental constraints such as lighting conditions or device malfunctions.

Scalability is another key strength. The modular architecture, consisting of a Node.js backend and a Python-based machine learning microservice, enables seamless scaling across multiple sessions and user groups. As illustrated in Figure 19, the system demonstrates near-linear growth in processing time with respect to the number of users, indicating that it can handle large-scale academic deployments without significant degradation in performance.

© 2026 JSIAR

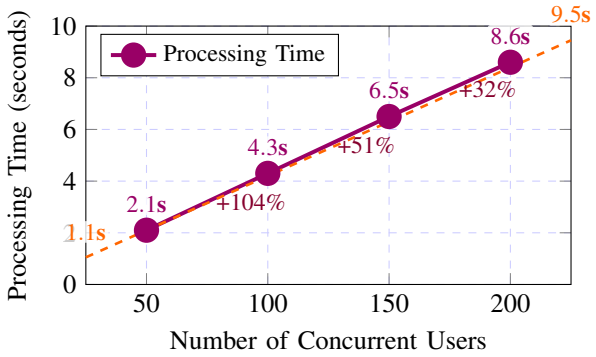

Fig. 19: Scalability analysis of the proposed system showing near-linear growth.

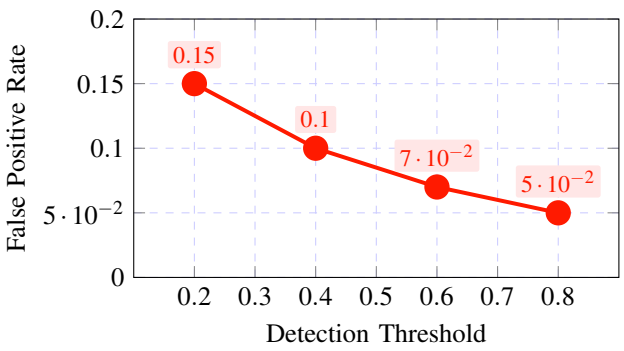

Fig. 20: False positive rate variation with detection threshold.

The intelligent detection capability of the framework further distinguishes it from traditional approaches. By leveraging a hybrid combination of Isolation Forest and Local Outlier Factor algorithms, the system captures both global anomalies and local deviations in behavioral patterns. This dual perspective enhances detection accuracy, particularly in scenarios where proxy attendance manifests through subtle inconsistencies rather than overt violations. The use of multi-dimensional features, including temporal, spatial, and device-level attributes, allows the system to construct a comprehensive behavioral profile for each user.

In addition, the framework exhibits privacy-preserving characteristics. Since it does not rely on biometric identifiers or sensitive personal data, the risk of privacy infringement is minimized. Behavioral features such as submission timing and device usage are less intrusive while still providing sufficient discriminatory power for anomaly detection. This aligns well with emerging data protection considerations in digital systems.

Despite these advantages, the proposed system is not without limitations. A primary concern is its dependence on historical data for model training. The effectiveness of anomaly detection relies on the availability of sufficient behavioral records to establish a baseline of normal activity. In newly deployed systems or scenarios with sparse data, the model may struggle to accurately differentiate between legitimate and anomalous behavior.

Another limitation arises from the possibility of false positives, particularly in edge cases where user behavior deviates due to legitimate reasons. For example, a student accessing the system from a new device or location may be incorrectly flagged as suspicious. Figure 20 illustrates the trade-off between detection sensitivity and false positive rate, highlighting the need for careful threshold tuning.

Geolocation accuracy also presents a practical challenge. The system relies on IP-based location estimation, which may not always reflect the true physical position of the user due to VPN usage, network routing, or shared institutional networks. This can introduce noise into the feature space and potentially affect detection reliability. Table VIII summarizes the key limitations and their implications.

The proposed system demonstrates a strong balance between practicality and intelligence, offering a scalable and privacy-conscious solution to proxy attendance detection. While certain limitations persist, they can be mitigated through adaptive model tuning, continuous learning, and integration of additional contextual features. This work contributes a novel, deployable framework that advances attendance authentication by embedding behavioral intelligence into web-based systems.

IX. Conclusion and Future Work

A. Conclusion

This study presented a behavioral anomaly detection framework for identifying proxy attendance in web-based systems, addressing a critical gap in existing attendance management approaches. Unlike conventional methods that rely primarily on identity verification or hardware-based authentication, the proposed framework leverages contextual behavioral patterns to infer authenticity. By integrating Isolation Forest and Local Outlier Factor (LOF) algorithms, the system effectively captures both global anomalies and localized deviations within high-dimensional feature spaces derived from temporal, spatial, and device-level attributes.

The experimental evaluation, conducted on a semi-synthetic dataset comprising 1,200 records with controlled proxy injection, demonstrates that the hybrid model achieves strong detection performance, with high precision, recall, and AUC values. These results validate the underlying hypothesis that proxy attendance can be reliably detected through inconsistencies in behavioral signatures rather than explicit identity mismatches. Furthermore, the system's modular architecture, implemented using a web-based frontend, Node.js backend, and Python-based machine learning microservice, ensures scalability and ease of deployment in real-world academic environments.

A notable contribution of this work lies in its hardware-independent design, which reduces deployment complexity while preserving detection robustness. The framework also aligns with privacy considerations by avoiding the use of sensitive biometric data, instead relying on non-intrusive behavioral features. Collectively, these characteristics position the proposed system as a practical and intelligent alternative

© 2026 JSIAR

TABLE VIII: Limitations and Their Implications

Limitation	Impact
Dependence on Historical Data	Reduced accuracy in early deployment stages
False Positives	Potential misclassification of legitimate users
Geolocation Inaccuracy	Noise in spatial feature representation

to existing attendance solutions, capable of enhancing integrity without imposing additional infrastructural burden.

B. Future Work

While the current framework demonstrates promising results, several avenues exist for further enhancement and exploration. One potential direction involves the incorporation of deep learning architectures, such as Long Short-Term Memory (LSTM) networks and Transformer-based models, to capture sequential dependencies and temporal dynamics in user behavior more effectively. These models could provide richer representations of attendance patterns, particularly in long-term deployments where behavioral evolution becomes significant.

Another promising extension is the integration of blockchain technology to ensure tamper-proof storage of attendance records. By leveraging decentralized ledgers, the system can enhance data integrity and auditability, thereby increasing trust among stakeholders. Additionally, the adoption of federated learning paradigms could enable collaborative model training across multiple institutions without requiring centralized data sharing. This would not only improve model generalization but also address privacy concerns associated with cross-domain data aggregation.

The development of a dedicated mobile application also presents an opportunity to extend system accessibility and user engagement. Mobile platforms can facilitate real-time notifications, location-aware validation, and seamless interaction, further strengthening the overall ecosystem. Moreover, incorporating adaptive learning mechanisms and feedback-driven model updates could improve resilience against evolving proxy strategies and adversarial behaviors.

In conclusion, this work establishes a novel foundation for attendance authentication by demonstrating the effectiveness of behavioral anomaly detection in web-based systems. It contributes a scalable, privacy-aware, and intelligent framework that bridges the gap between theoretical anomaly detection techniques and practical academic deployment, while opening multiple pathways for future research and system evolution.

REFERENCES

[1] S. K. Mohanty and D. P. Mohapatra, "RFID based attendance system," *International Journal of Computer Applications*, vol. 42, no. 15, pp. 11–14, 2012.

[2] A. K. Jain, A. Ross, and S. Prabhakar, "An introduction to biometric recognition," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 14, no. 1, pp. 4–20, Jan. 2004.

[3] M. Alshehri, "A secure web-based attendance system using multi-factor authentication," *International Journal of Advanced Computer Science and Applications*, vol. 10, no. 6, pp. 245–252, 2019.

[4] R. Want, "An introduction to RFID technology," *IEEE Pervasive Computing*, vol. 5, no. 1, pp. 25–33, Jan.–Mar. 2006.

[5] F. T. Liu, K. M. Ting, and Z.-H. Zhou, "Isolation forest," in *Proc. IEEE Int. Conf. Data Mining (ICDM)*, 2008, pp. 413–422.

[6] M. M. Breunig, H.-P. Kriegel, R. T. Ng, and J. Sander, "LOF: Identifying density-based local outliers," in *Proc. ACM SIGMOD*, 2000, pp. 93–104.

[7] C. C. Aggarwal, *Outlier Analysis*, 2nd ed. Cham, Switzerland: Springer, 2017.

[8] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," *ACM Computing Surveys*, vol. 41, no. 3, pp. 1–58, 2009.

[9] J. Dean and S. Ghemawat, "MapReduce: Simplified data processing on large clusters," *Communications of the ACM*, vol. 51, no. 1, pp. 107–113, 2008.

[10] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*. New York, NY, USA: Springer, 2009.

[11] N. Ye, "A markov chain model of temporal behavior for anomaly detection," in *Proc. IEEE Systems, Man, and Cybernetics Conf.*, 2000, pp. 171–174.

[12] E. Eskin, "Anomaly detection over noisy data using learned probability distributions," in *Proc. ICML*, 2000, pp. 255–262.

[13] A. Patcha and J. M. Park, "An overview of anomaly detection techniques: Existing solutions and latest technological trends," *Computer Networks*, vol. 51, no. 12, pp. 3448–3470, 2007.

[14] S. Axelsson, "The base-rate fallacy and its implications for the difficulty of intrusion detection," *ACM Transactions on Information and System Security*, vol. 3, no. 3, pp. 186–205, 2000.

[15] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*, 3rd ed. San Francisco, CA, USA: Morgan Kaufmann, 2011.

[16] R. Harrop and D. Das, "RFID-based student attendance system," *IEEE Trans. Educ.*, vol. 53, no. 2, pp. 234–241, 2010.

[17] L. Ni, Y. Liu, Y. Lau, and A. Patil, "LANDMARC: Indoor location sensing using active RFID," *Wireless Networks*, vol. 10, no. 6, pp. 701–710, 2004.

[18] K. Finkenzeller, *RFID Handbook*. Wiley, 2010.

[19] A. K. Jain et al., "Biometric recognition," *IEEE Trans. CSVT*, 2004.

[20] J. Daugman, "How iris recognition works," *IEEE Trans. Circuits Syst.*, 2004.

[21] Y. Taigman et al., "DeepFace: Closing the gap to human-level performance," in *CVPR*, 2014.

[22] K. Nandakumar et al., "Biometric template security," *EURASIP J. Adv. Signal Process.*, 2008.

[23] S. K. Sharma, "QR code based smart attendance system," *IJCS*, 2018.

[24] H. Kim and J. Lee, "Secure mobile attendance using location verification," *Sensors*, 2020.

[25] F. T. Liu et al., "Isolation forest," in *ICDM*, 2008.

[26] M. M. Breunig et al., "LOF: Identifying density-based outliers," in *SIGMOD*, 2000.

[27] V. Chandola et al., "Anomaly detection survey," *ACM CSUR*, 2009.

[28] C. Aggarwal, *Outlier Analysis*. Springer, 2017.

[29] N. Ye, "Behavior-based anomaly detection," in *IEEE SMC*, 2000.

[30] E. Eskin, "Anomaly detection over noisy data," in *ICML*, 2000.

[31] A. Patcha and J. Park, "Anomaly detection techniques," *Computer Networks*, 2007.

[32] S. Axelsson, "Intrusion detection issues," *ACM TISSEC*, 2000.

[33] J. Han et al., *Data Mining*. Morgan Kaufmann, 2011.

[34] T. Hastie et al., *Statistical Learning*. Springer, 2009.

[35] I. Goodfellow et al., *Deep Learning*. MIT Press, 2016.

© 2026 JSIAR