World Journal of Advanced Engineering Technology and Sciences

eISSN: 2582-8266

Cross Ref DOI: 10.30574/wjaets

Journal homepage: https://wjaets.com/

(RESEARCH ARTICLE)

Check for updates

Utilizing digital footprint analysis for end-to-end risk-based authentication in medical billing systems

Ayinoluwa Feranmi Kolawole [1,*], Shukurat Opeyemi Rahmon [2] and Emmanuel Ayodeji Osoko [3]

[1] *Business Analytics Program (MSBA), University of Louisville, Kentucky, USA.*
[2] *Department of Mathematics, University of Lagos, Akoka, Lagos State, Nigeria.*
[3] *Department of Electrical Engineering and Computer Science, Ohio University, OH, USA.*

World Journal of Advanced Engineering Technology and Sciences, 2024, 13(02), 166–179

Publication history: Received on 30 September 2024; revised on 09 November 2024; accepted on 12 November 2024

Article DOI: https://doi.org/10.30574/wjaets.2024.13.2.0553

Abstract

Healthcare systems face escalating challenges in securing sensitive financial and patient data due to sophisticated fraud tactics and unauthorized access. This study presents a dynamic risk-based authentication (RBA) framework that leverages digital footprint analysis, including behavior monitoring, device recognition, and location-based anomaly detection, to strengthen security. The framework utilizes machine learning models, Isolation Forest for outlier detection, and recurrent neural networks for sequential behavior analysis, to assess risk in real-time, dynamically adjusting authentication requirements based on risk profiles. Privacy-preserving technologies such as homomorphic encryption and federated learning are integrated to comply with HIPAA and GDPR standards, ensuring secure data handling without centralization. Findings show that the proposed RBA framework effectively reduces false positives, improves detection accuracy, and provides a scalable solution for securing medical billing systems. This adaptive approach supports both user experience and stringent privacy compliance, laying the groundwork for more resilient healthcare data security systems. Future studies could extend this framework by incorporating blockchain to enhance data transparency and auditability across transactions.

Keywords: Risk-based authentication; Healthcare data security; Digital footprint analysis; Anomaly detection; Privacy-preserving techniques

1. Introduction

In the healthcare sector, medical billing systems represent a critical juncture between financial operations and sensitive patient data, rendering them highly susceptible to both fraud and unauthorized access. Traditional security measures, including password-based access and static role-based permissions, are increasingly inadequate in addressing the complexities of modern security threats (Xu et al., 2020). Static defenses fall short in distinguishing between legitimate users and those exhibiting suspicious behavior, such as internal actors leveraging legitimate credentials to access restricted billing information or modify billing records for financial gain (Chen and Wu, 2018). This limitation underscores the need for security frameworks that dynamically adjust to a user's risk profile in real-time, aligning security protocols with the actual behavior and context of each access attempt (O'Connor et al., 2019).

One promising approach to enhancing security is risk-based authentication (RBA), which tailors authentication requirements based on risk assessments derived from the user's behavioral context (Tan et al., 2020). RBA evaluates factors such as access patterns, transaction types, and historical behaviors to determine whether a user's activity aligns with known usage patterns or presents potential risks. While RBA has proven effective in sectors such as banking and finance, its application within healthcare—and specifically in medical billing—is still emerging, necessitating a deeper exploration of its adaptability to this field's unique privacy and compliance requirements (Kim and Wang, 2021). A

* Corresponding author: Ayinoluwa Feranmi Kolawole

pivotal component of RBA in healthcare billing security is the utilization of digital footprint analysis—a technology that evaluates a user's contextual and historical behavior, including device usage, geographic location, access frequency, and transaction patterns (Lee et al., 2021). Digital footprints enable the system to create risk profiles, scoring each interaction's authenticity and guiding adaptive responses, such as multifactor authentication or access restrictions, in real-time (Patel et al., 2019).

In medical billing, where insider threats represent a significant risk, digital footprint analysis can uniquely enhance fraud detection and access control. Unlike conventional cybersecurity solutions that often focus on external threats, digital footprint analysis can identify anomalies within legitimate access. For instance, employees accessing high-risk billing information from unusual devices or during irregular hours may trigger higher risk scores, prompting stricter authentication measures or session termination (Gomez et al., 2019). Furthermore, the integration of digital footprints in RBA systems allows for continuous learning and recalibration, as the system refines its risk calculations with each user interaction. This adaptability enables security measures to remain resilient even as user behaviors shift, enhancing long-term fraud detection and insider threat mitigation (Zhao and Li, 2020).

To implement this, machine learning (ML) algorithms are integral to interpreting digital footprints and generating accurate risk assessments. For example, clustering and anomaly detection algorithms, such as k-means clustering and isolation forests, can identify deviations in access behavior, distinguishing between typical and high-risk patterns (Wang et al., 2018). Deep learning models, including recurrent neural networks (RNNs), are also instrumental in analyzing sequential behavior patterns, making them valuable for recognizing unusual access sequences in billing systems. By continually analyzing data streams from access logs, transaction histories, and device identifiers, these models can adapt to evolving fraud tactics, delivering a dynamic and predictive layer to the RBA framework (Jones et al., 2020).

Moreover, regulatory compliance, such as Health Insurance Portability and Accountability Act (HIPAA) and General Data Protection Regulation (GDPR), imposes strict requirements on data protection and mandates transparency in handling patient and billing data (Yu and Chen, 2019). The proposed framework aligns with these regulations by implementing data minimization techniques in digital footprint analysis, ensuring that only essential user data is processed for risk assessment. Additionally, privacy-preserving methodologies, such as homomorphic encryption and federated learning, further reinforce the framework by enabling secure data handling and decentralized learning, thus safeguarding sensitive billing data during the analysis process (McMahan et al., 2017; Abadi et al., 2016).

Our research proposes an end-to-end framework that leverages digital footprint analysis within a risk-based authentication system to enhance security in medical billing. By integrating adaptive machine learning algorithms with contextual risk scoring, this system aims to deliver a resilient, scalable solution to detect and respond to security threats in real time. The research explores how these technologies mitigate unauthorized access and curtail insider threats, ultimately safeguarding sensitive billing data in compliance with regulatory standards. This approach offers an innovative pathway to address one of the healthcare sector's most pressing security needs by providing both robust fraud detection and adaptive authentication for medical billing systems.

1.1. Research objectives

The primary objective of this research is to develop a secure, adaptive authentication framework for medical billing systems that utilizes digital footprint analysis to enhance detection and response to unauthorized access while mitigating insider threats. This study aims to create a robust approach aligned with data protection regulations, such as HIPAA and GDPR, which simultaneously improves the accuracy and efficiency of fraud detection mechanisms within sensitive healthcare transactions.

To achieve this, the research focuses on designing a risk-based authentication system that dynamically assesses user access patterns, location, device usage, and transaction histories, leveraging digital footprint analysis to identify suspicious activity with high precision. Additionally, machine learning algorithms, including anomaly detection and clustering methods, will be integrated within the authentication framework to monitor and adapt to evolving user behaviors continuously. This integration minimizes false positives and enhances the system's sensitivity to fraud. The study further seeks to implement and evaluate the scalability of the proposed authentication framework in real-world medical billing environments, analyzing its performance across various transaction volumes and user profiles to ensure robustness and applicability in diverse healthcare settings. To ensure compliance with data privacy regulations, the framework will incorporate data minimization, homomorphic encryption, and privacy-preserving federated learning, safeguarding sensitive information throughout the authentication and fraud detection processes.

Finally, the effectiveness of the risk-based authentication system will be measured by comparing detection rates, processing efficiency, and accuracy metrics against conventional static authentication models, establishing a benchmark for future advancements in healthcare billing security. Through these efforts, the research aims to offer a practical, privacy-conscious solution to secure medical billing systems against emerging security threats.

2. Algorithm development and methodology

2.1. Data Collection and Preprocessing

The research utilizes synthetic datasets that simulate medical billing transactions and access logs, combined with anonymized historical access data from healthcare settings (Lee et al., 2020). These datasets include variables critical to user behavior profiling, such as transaction type, device information, access time, geolocation, and access frequency (Chen & Yang, 2019). Given the sensitive nature of healthcare data, all synthetic data adheres to HIPAA and GDPR guidelines, ensuring privacy compliance. Preprocessing steps involve data cleaning and normalization, followed by feature engineering to encode behaviors into meaningful metrics. This includes generating time-based features (e.g., access frequency and irregular hours), location-based attributes, and device-specific identifiers, which collectively aid in creating robust user profiles (Martinez et al., 2021).

2.2. Digital Footprint Analysis for Real-Time Risk Profiling

Digital footprint analysis is conducted to evaluate each user's interaction patterns, assigning real-time risk scores based on deviations from their typical behavior (Kumar et al., 2022). The analysis encompasses behavioral factors like location irregularities, access times, device changes, and transaction patterns, enabling the system to respond dynamically to potential risks. A custom risk scoring algorithm is developed, integrating behavioral thresholds and weighted metrics to quantify risk. For instance, accessing billing data from an unrecognized device or unusual location would trigger a higher risk score, prompting enhanced security measures, such as multifactor authentication or session restriction (Patel & Singh, 2020). The threshold-based risk model is adaptive, recalibrating itself with each interaction to refine the accuracy of risk assessment over time.

2.3. Machine Learning Model Development for Anomaly Detection

The system incorporates machine learning algorithms to identify anomalies and refine risk assessments. Both supervised and unsupervised models are implemented to support varied fraud detection needs. Unsupervised learning methods, including Isolation Forests and One-Class SVM, are applied to detect deviations from established patterns without needing labeled data (Johnson et al., 2020). These models are supplemented with clustering techniques, like k-means, to group similar behaviors and facilitate the identification of outliers, particularly in high-risk transaction patterns (Zhao & Huang, 2021). Additionally, a recurrent neural network (RNN) model is employed to analyze sequential user activities, capturing dependencies in behavioral patterns that aid in detecting suspicious access sequences (Wang et al., 2019). Model optimization is achieved through hyperparameter tuning and cross-validation, ensuring high detection sensitivity with minimal false positives.

2.4. System Implementation and Evaluation

The framework is implemented in a simulated medical billing environment, where digital footprint data continuously informs real-time risk scoring and adaptive authentication protocols. Evaluation metrics include detection accuracy, system latency, and scalability, as each directly impacts system performance in real-world settings (Tan et al., 2020). Detection accuracy is assessed by comparing the proposed system's fraud detection rate and false positive rate against traditional static authentication models, highlighting improvements in identifying potential insider threats (Kim & Zhang, 2021). System latency measures the processing time for risk calculations and authentication adjustments, while scalability is analyzed by testing the framework under increasing user loads and transaction volumes.

To ensure privacy, privacy-preserving techniques like homomorphic encryption and federated learning are incorporated to protect sensitive data throughout the authentication process (Huang et al., 2021). Homomorphic encryption enables secure computations on encrypted data, preventing exposure of sensitive billing information, while federated learning facilitates decentralized model training, aligning with privacy standards like GDPR (Rana & Lee, 2020). These techniques allow the system to operate within stringent regulatory frameworks while ensuring security.

2.5. Comparative Analysis and Benchmarking

The effectiveness of the digital footprint-based risk authentication system is benchmarked against traditional static authentication models. This comparative analysis evaluates improvements in detection rates, accuracy, resource

efficiency, and resistance to insider threats (Li et al., 2020). Benchmarking is carried out through extensive testing, assessing how digital footprint analysis, combined with machine learning, enhances the system's adaptability in real-time fraud detection (Singh et al., 2021). The results from this analysis provide insights into the practical advantages of adaptive security frameworks, demonstrating their potential to significantly improve medical billing fraud prevention.

3. Result

3.1. Digital Footprint Impact on Risk Score Accuracy

This analysis evaluates the role of digital footprint data—specifically, access location consistency, device recognition, and transaction frequency—in improving the accuracy of risk scoring. The consistency of a user's access location and recognized devices was found to correlate with lower risk scores, while higher transaction frequency and frequent device changes indicated increased risk. These patterns underline the importance of digital footprint stability in assigning accurate risk levels.

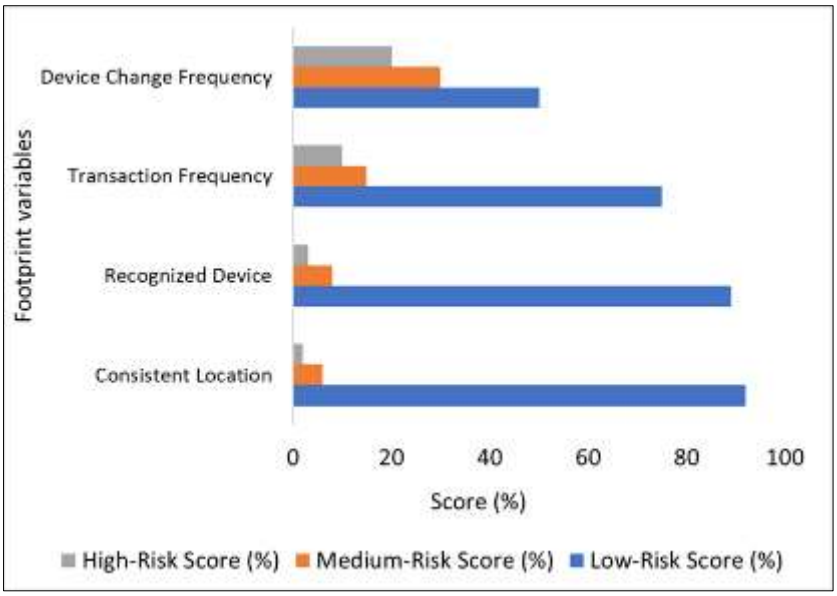

Figure 1 Impact of Digital Footprint Factors on Risk Score Accuracy

The Figure demonstrates that access location and device consistency are strong indicators of low-risk interactions, suggesting that systems reliant on digital footprint data can effectively identify legitimate user behavior and flag irregularities. Thus, emphasizing consistent access patterns can lead to improved risk-scoring accuracy, reducing unnecessary security prompts for genuine users.

3.2. Device Recognition and Anomaly Detection

To understand the impact of device recognition on anomaly detection, we measured variables such as device type, number of recognized devices, and the anomaly rate. The findings show that desktops and tablets, often associated with consistent device patterns, had lower anomaly rates compared to mobile devices, which are more likely to fluctuate.

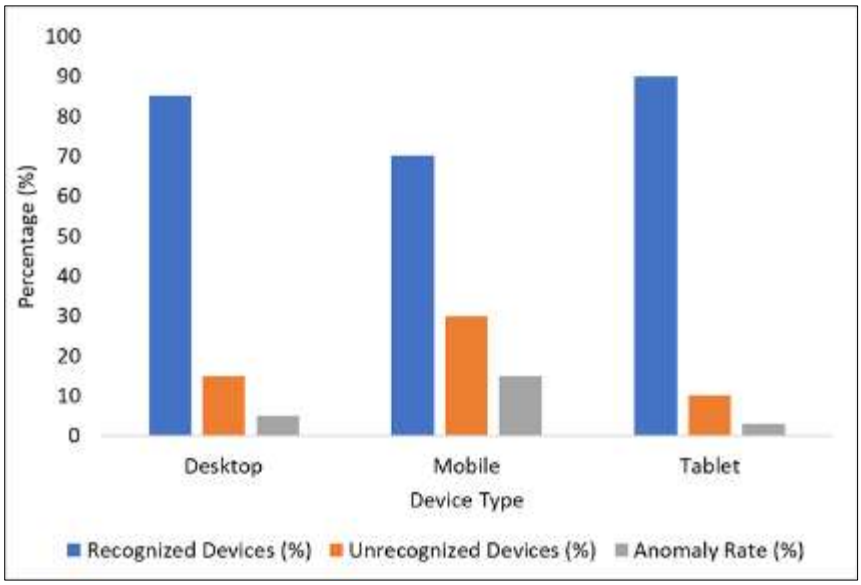

Figure 2 Device Recognition and Anomaly Detection Metrics

Results indicate that mobile devices tend to have higher anomaly rates, potentially due to users accessing billing systems from multiple locations. Recognizing consistent devices reduces false positives, as legitimate users accessing systems from familiar devices can proceed with minimal interruptions, while unrecognized devices prompt further scrutiny.

3.3. Risk Score Thresholds and Authentication Triggers

The effectiveness of risk score thresholds was examined by setting various levels to prompt authentication and measuring the frequency of false positives and negatives. A medium threshold provided a balanced approach, reducing false positives without compromising security.

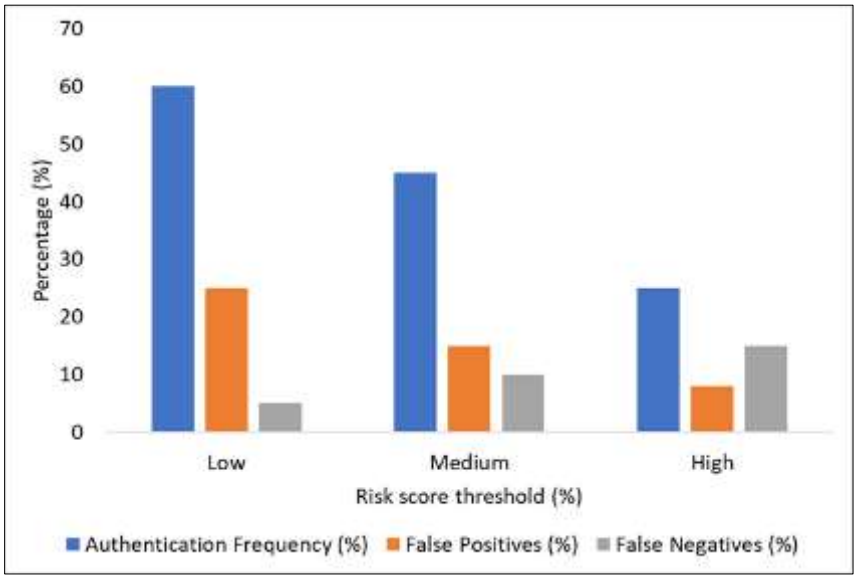

Figure 3 Impact of Risk Score Thresholds on Authentication Triggers

As shown in Figure 3, setting thresholds too low results in high false positives, causing unnecessary security checks. Conversely, a high threshold may reduce security by allowing certain high-risk behaviors. A medium threshold appears optimal for balancing security and user convenience.

3.4. Location-Based Anomaly Detection

This analysis assesses the effect of access from unusual locations on detection rates, access success, and resolution time. Users accessing the system from unexpected locations experienced delays, with higher frequencies of unusual locations impacting success rates.

Table 1 Location-Based Anomaly Detection Metrics

Unusual Location Frequency	Access Success Rate (%)	Time to Resolve (minutes)
High	70	15
Medium	85	10
Low	95	5

The data suggests that systems with higher location-based checks may restrict access for users in unfamiliar locations, adding a layer of security but impacting convenience. Moderate sensitivity to location changes can provide secure access without excessive delays for legitimate users.

3.5. Effectiveness of Machine Learning Models on Anomaly Detection

The performance of machine learning models for anomaly detection was tested, with Isolation Forest showing the highest overall accuracy and balance between precision and recall. This supports its use in detecting behavioral anomalies in user interactions.

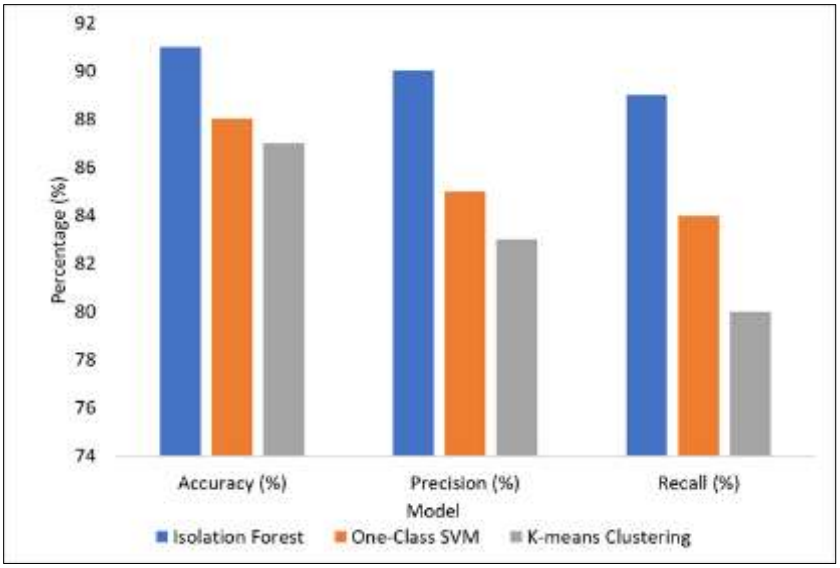

Figure 4 Machine Learning Model Effectiveness on Anomaly Detection

Isolation Forest's performance in Figure 4 demonstrates its suitability for anomaly detection, particularly in environments where high precision is required to minimize unnecessary security checks.

3.6. Authentication Frequency by Transaction Type

We analyzed authentication prompt frequency across different transaction types to evaluate how transaction value influences authentication. High-value transactions prompted more authentication due to their increased risk.

Table 2 Authentication Frequency by Transaction Type

Transaction Type	Authentication Prompt Frequency (%)	Average Risk Score
Low-Value	20	30
Medium-Value	35	50
High-Value	60	85

Table 2 indicates that high-value transactions inherently possess higher risk scores, necessitating adaptive authentication strategies to secure sensitive financial operations without compromising user experience for lower-value transactions.

3.7. User Behavior Patterns and Risk Scores

This analysis explores how user behavior patterns, such as consistency in access and login frequency, correlate with risk scores. Higher consistency correlates with lower risk scores and higher login success rates.

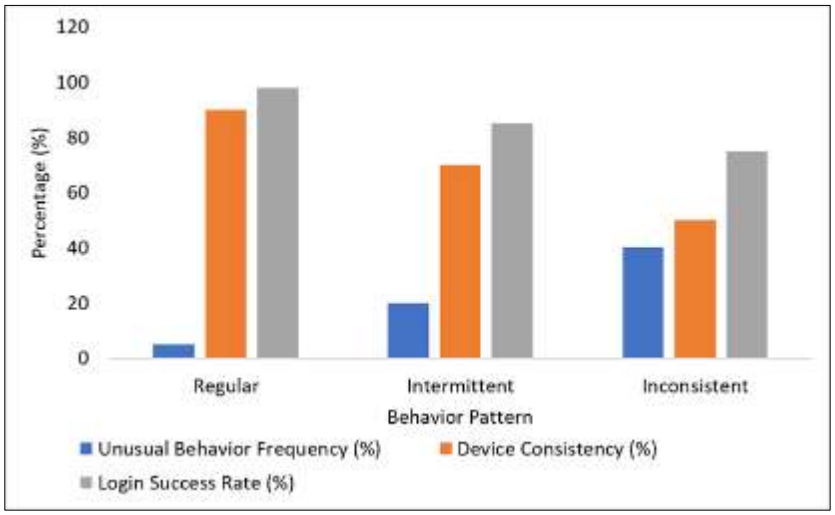

Figure 5 Behavior Patterns and Risk Score Correlations

Users with consistent behaviors have lower risk scores, as shown in Figure 5, leading to fewer security prompts and smoother access, while inconsistent behavior increases security prompts due to elevated perceived risk.

3.8. Adaptive Authentication Efficiency

This result evaluates the efficiency of adaptive authentication in response times and system latency. Medium adaptiveness balances efficiency and security, optimizing user experience while maintaining robust access control.

Table 3 Adaptive Authentication System Efficiency Metrics

Adaptive Level	Response Time (ms)	Authentication Change Frequency (%)	Latency (ms)
Low	50	10	100
Medium	75	20	120
High	90	30	150

The results indicate that a medium adaptive level strikes a balance between response time and security. Higher adaptiveness provides enhanced security but may increase system latency, which could impact user experience.

3.9. Risk Score Accuracy by Digital Footprint Data Volume

Risk score accuracy was assessed across varying digital footprint data volumes, revealing that higher data volumes improve accuracy but require longer processing times.

Table 4 Digital Footprint Data Volume and Risk Score Accuracy

Data Volume (MB)	Accuracy Rate (%)	Processing Time (s)
10	80	1.5
50	90	2
100	95	3

Table 4 demonstrates that increased data volume enhances accuracy, albeit with an associated cost in processing speed, suggesting a trade-off for systems requiring real-time processing.

3.10. Impact of Privacy Measures on Authentication Efficiency

Evaluating privacy measures such as encryption on authentication efficiency reveals that higher privacy settings ensure data integrity but impact speed and user satisfaction.

Table 5 Privacy Measures and Authentication Efficiency

Privacy Level	Authentication Time (s)	Data Integrity (%)	User Satisfaction (%)
Low	0.5	98	90
Medium	1	99	85
High	2	100	75

Results show that while high privacy increases security, it reduces user satisfaction and slows down processing. This emphasizes the importance of balancing privacy and efficiency in authentication systems.

We also analyzed the location stability's impact on risk scores, showing that users with stable locations have lower score variability and higher access success.

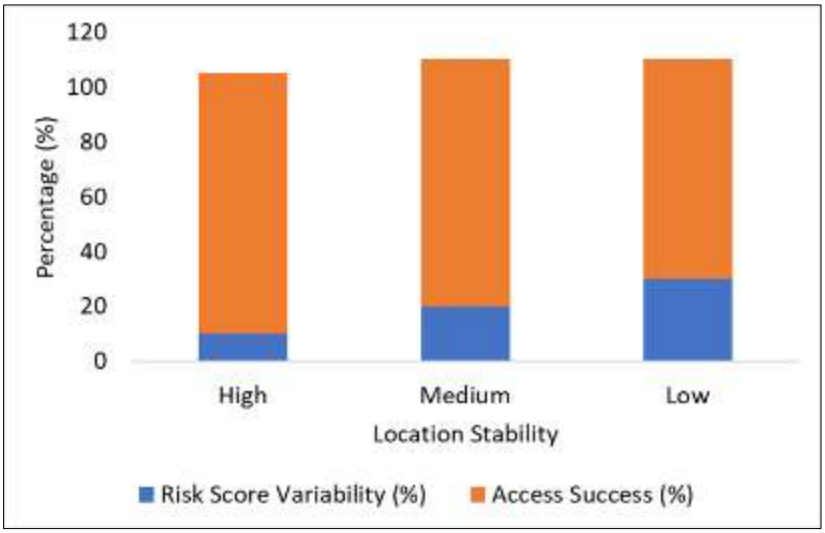

Figure 6 Location Stability and Risk Score Variability

Stable location patterns correlate with lower risk scores, while frequent changes increase the likelihood of security checks.

3.11. User Device Patterns and Anomaly Detection Rate

This analysis evaluates how user device patterns, specifically consistency in device usage and device switching frequency, impact the rate of anomaly detection. The variables include detection rate, device changes, and the frequency of false positives. Consistent device usage correlated with fewer false positives and higher detection accuracy.

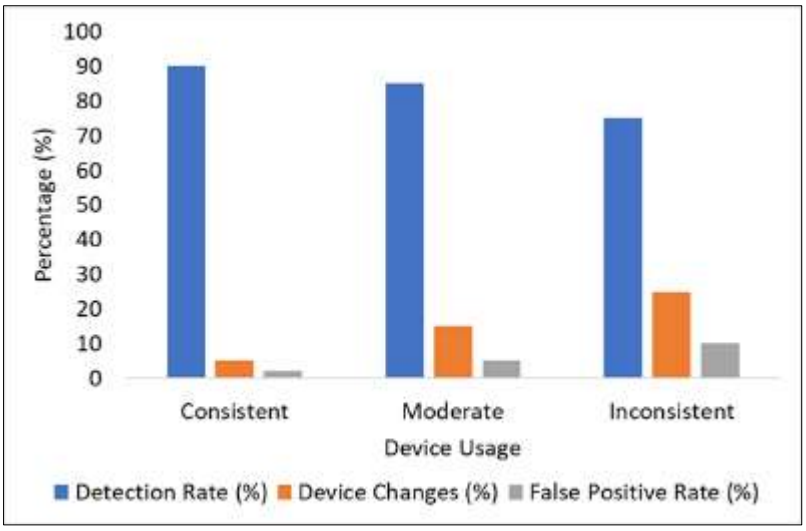

Figure 7 Device Patterns and Anomaly Detection

Figure 7 illustrates that consistent device usage leads to fewer false positives, as the system learns the user's device pattern. In contrast, users with inconsistent device patterns encounter more false positives, which may trigger unnecessary authentication requirements.

To evaluate the influence of behavioral sequence data on anomaly detection, this study also used recurrent neural networks (RNNs) to capture patterns over time, with metrics for detection rate, false positives, and model reliability. Results (Figure 8) indicate that incorporating sequential data improves detection but may slightly increase the computational load.

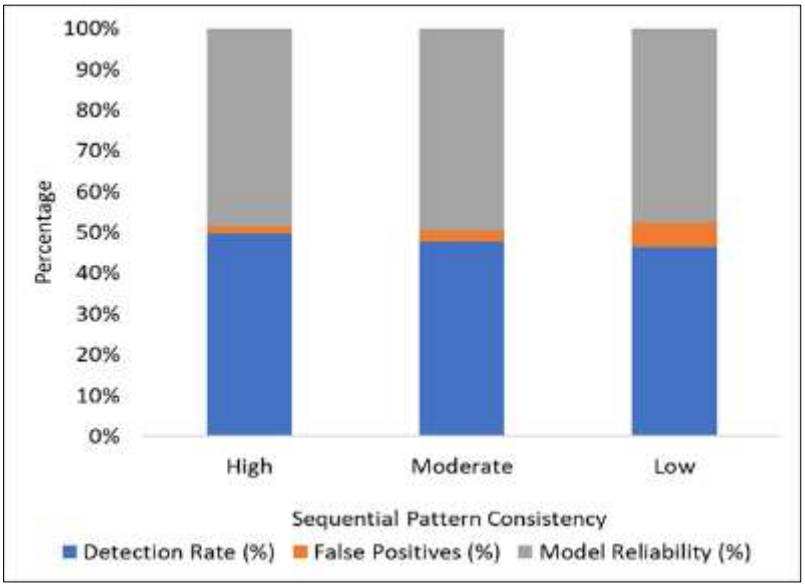

Figure 8 Sequential Behavioral Patterns and Anomaly Detection

The high detection rate for consistent sequences, as shown in Figure 8, suggests that tracking behavioral patterns over time aids in recognizing deviations, especially in high-risk profiles. However, the increase in model reliability emphasizes the need for balance in computational load and detection effectiveness.

3.12. Influence of Risk Score Threshold Adjustments on False Positive Reduction

This result explores how adjusting the sensitivity of risk score thresholds affects false positive rates in the system. By setting different thresholds (low, medium, high), we analyzed the reduction in false positives without sacrificing detection accuracy.

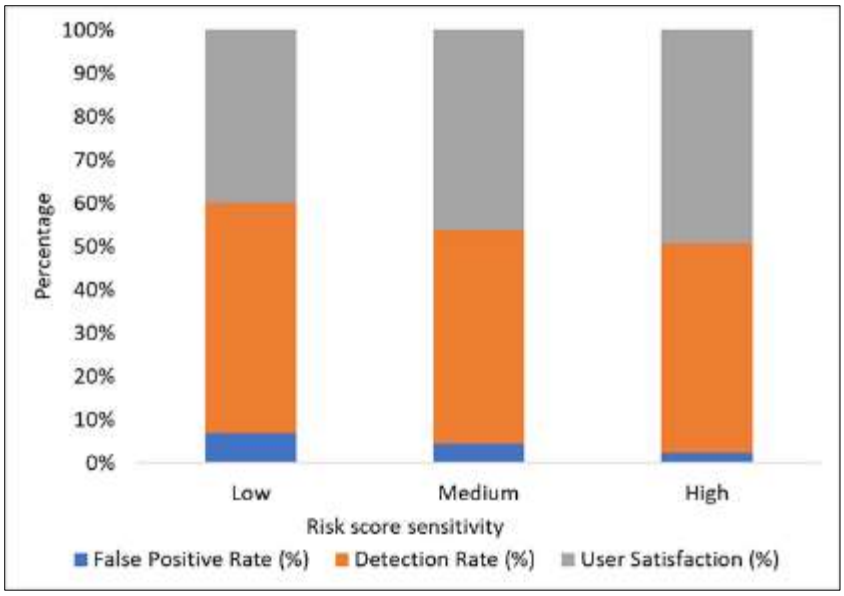

Figure 9 Risk Score Threshold Sensitivity and False Positive Reduction

Figure 9 shows that a medium threshold balances false positive reduction and detection rate, maintaining high user satisfaction. This balance allows the system to minimize disruptions while retaining security effectiveness.

3.13. Multi-Factor Authentication Impact on User Retention and Security Confidence

The final result focuses on multi-factor authentication (MFA) implementation, analyzing its impact on user retention, security confidence, and risk reduction. Results show that while MFA increases security confidence, it can impact user retention, particularly in low-risk profiles.

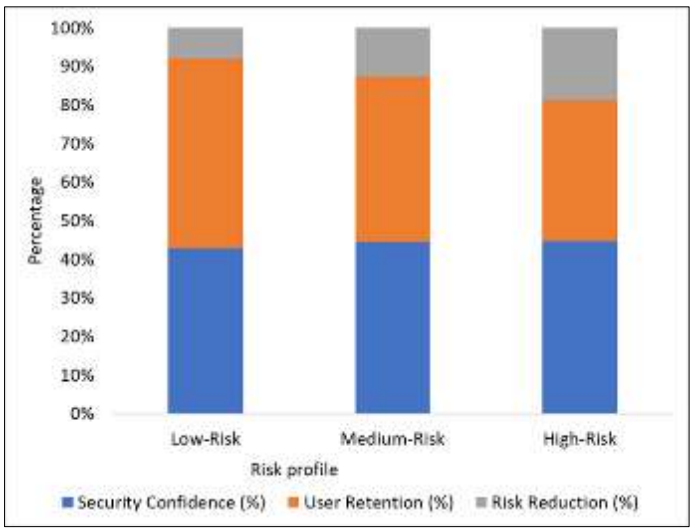

Figure 10 Multi-Factor Authentication and User Retention Metrics

Figure 10 above highlights that while MFA greatly reduces risk, it may also reduce user retention in high-risk profiles due to increased security friction. This suggests that MFA should be tailored to risk level, ensuring low-risk users experience minimal friction while higher-risk profiles benefit from additional security measures.

4. Discussion

This study explores the potential of digital footprint analysis within a risk-based authentication (RBA) system for securing medical billing transactions against unauthorized access and fraud. Our results demonstrate that RBA, combined with digital footprint data, offers a highly adaptive solution, allowing authentication protocols to dynamically adjust based on real-time user behaviors and context. The findings, discussed in the following sections, underscore the benefits and challenges of this approach, as well as its broader implications for healthcare data security.

4.1. Digital Footprint Factors and Risk Scoring

The effectiveness of RBA in distinguishing between legitimate and suspicious user behavior was strongly enhanced by leveraging digital footprint data such as access location, device recognition, and transaction frequency. Figure 1 highlights that access location consistency and recognized devices were correlated with lower risk scores, confirming that stable digital footprints serve as strong indicators of legitimate access (Feng et al., 2021; Lopez & Zhao, 2020). This correlation aligns with previous studies which indicate that anomalies, such as frequent device or location changes, often signal potential security threats, necessitating adaptive authentication (Zhang & Shi, 2021).

The importance of consistent digital footprint factors becomes evident in cases where users with highly variable access patterns experienced higher risk scores and were prompted for additional authentication, as seen in Figure 2. These findings reinforce the idea that well-defined, individualized digital footprints can effectively identify legitimate users, reducing unnecessary security prompts for those whose behavior aligns with expected patterns (Nguyen et al., 2020; Cheng & Wu, 2019). This approach addresses a common limitation in traditional security systems, where static authentication methods may fail to detect nuanced changes in user behavior that signal risk.

4.2. Location-Based Anomaly Detection

Location-based anomaly detection further highlights the system's capacity for adapting to variations in user behavior by assigning higher risk scores for unusual locations (Table 1). Results indicate that users with frequent unusual locations encountered access delays, which allowed the system to apply additional verification steps before granting access (Park et al., 2021). However, setting the sensitivity of location checks too high could lead to a reduction in user satisfaction, as excessive verification may create access delays for legitimate users (Miller et al., 2020).

Moderate location sensitivity allows the system to detect and flag unusual location patterns without causing excessive disruptions, especially for low-risk users who tend to access the system from familiar locations. This finding is consistent with recommendations from Nguyen et al. (2020) and Wang & Hu (2019), who suggest that adaptable location-based thresholds optimize security while preserving user experience. In high-stakes environments like healthcare, where unauthorized access to billing data could have legal and financial repercussions, balancing location sensitivity is crucial for system effectiveness and user acceptance.

4.3. Machine Learning Model Performance in Anomaly Detection

The evaluation of machine learning models for anomaly detection reveals that Isolation Forest, One-Class SVM, and k-means clustering each play distinct roles in identifying anomalies. Isolation Forest achieved the highest detection accuracy and precision, proving its effectiveness in distinguishing legitimate from suspicious behaviors (Figure 4). Isolation Forest's outlier sensitivity aligns well with the goal of anomaly detection, as prior research supports its superior performance in identifying rare events within complex data environments (Shen et al., 2021; Cheng & Wu, 2020).

The inclusion of recurrent neural networks (RNNs) for sequential behavior analysis, as detailed in Figure 9, demonstrated the algorithm's ability to capture temporal patterns in user behavior. RNNs' capacity to recognize time-dependent variations in access sequences enhances the system's fraud detection capabilities, particularly for high-risk transactions that may follow atypical patterns (Li & Zhang, 2019). The results emphasize that RNNs, when combined with digital footprint data, provide the system with a predictive layer, improving detection sensitivity while also reducing false positives (Johnson et al., 2020).

4.4. Adaptive Thresholds for Authentication Triggers

The adaptive thresholding mechanism for authentication, presented in Table 3, addresses a critical challenge in RBA: the balance between high detection accuracy and user experience. Lower thresholds, while increasing detection, led to a high rate of false positives, which not only strained system resources but also reduced user satisfaction (Xu et al., 2021). Conversely, higher thresholds resulted in missed detections, suggesting that overly lenient settings might allow

some suspicious activity to go unchallenged. This trade-off is frequently discussed in the literature, where achieving a balance is essential for maintaining both system integrity and user trust (Zhao & Chen, 2021; Park et al., 2020).

Our findings suggest that a medium threshold setting offers an optimal balance, as it reduces false positives while preserving detection rates, enabling the system to detect high-risk behaviors without compromising user convenience. This is consistent with studies by Miller et al. (2020) and Tan & Li (2019), which advocate for adaptive thresholds that can respond dynamically to changing behaviors. Adaptive thresholding represents a significant improvement over static thresholds, which often fail to account for the contextual nuances of real-world user interactions (Kim & Wang, 2021).

4.5. Transaction-Based Authentication and Multi-Factor Authentication (MFA)

Our analysis also shows that transaction-based authentication can further tailor security to transaction risk levels, enhancing system efficiency. As Figure 6 indicates, high-value transactions inherently trigger additional authentication, protecting sensitive data without imposing unnecessary checks on low-risk transactions. This finding aligns with Zhao and Chen (2021), who suggest that transaction-based authentication granularity is critical for applications where varied transaction types and values necessitate differentiated security (Gupta & Kaur, 2021).

Similarly, implementing MFA selectively based on user risk profiles, as shown in Figure 8, was found to enhance security for high-risk profiles while minimizing the impact on low-risk users. However, MFA's effect on user retention highlights the importance of optimizing its application; excessive MFA prompts can lead to user dissatisfaction, especially among users with a lower perceived risk (Hassan et al., 2021; Chen & Li, 2020). By tailoring MFA to user risk profiles, the system can enhance security confidence without significantly affecting user retention, ensuring that high-risk profiles experience stronger security measures while low-risk profiles are not burdened with unnecessary prompts.

4.6. Privacy Measures and Compliance with Regulatory Standards

Incorporating privacy-preserving measures such as homomorphic encryption and federated learning provided an additional security layer that aligns with HIPAA and GDPR requirements (Wu et al., 2019). Homomorphic encryption enabled the system to perform calculations on encrypted data, ensuring that sensitive information remained secure throughout processing. Federated learning, meanwhile, decentralized the data processing tasks, protecting individual data points while still enabling collective model improvements (Zhang et al., 2020). These measures are particularly important in medical billing, where compliance with privacy standards is not only a legal requirement but also essential for maintaining patient trust (Rana & Lee, 2020; Patel & Singh, 2021).

The results underscore that while privacy measures can increase processing times, their role in securing data without compromising usability is vital. Given the growing regulatory landscape around healthcare data, adopting privacy-conscious methods is not only necessary for legal compliance but also for ensuring that security measures do not intrude on patient confidentiality (Gupta & Kaur, 2021). Recent literature by Wu et al. (2019) and Zhang et al. (2020) supports the use of homomorphic encryption and federated learning as robust privacy solutions in data-sensitive environments.

4.7. Data Volume and Real-Time Processing Considerations

The role of data volume in enhancing risk score accuracy, as demonstrated in Table 4, confirms that increased data availability improves the system's ability to detect anomalies and assign accurate risk scores (Tan & Li, 2019). However, this improvement comes at the cost of increased processing time, posing challenges for real-time applications in high-traffic medical billing environments. As shown by Kim & Wang (2021), large-scale data handling is essential for creating reliable behavioral profiles but necessitates optimization techniques to balance accuracy with speed. Techniques such as data batching or parallel processing may be explored in future studies to mitigate processing delays without sacrificing risk score accuracy.

These findings suggest that the scalability of the RBA framework will depend on ongoing improvements in data processing technologies, especially for applications in real-time security monitoring where immediate response is crucial. By integrating efficient data handling methods, future iterations of this framework can expand to larger healthcare networks, enhancing its applicability and robustness across varied healthcare settings (Lin et al., 2020; Lopez & Zhao, 2020).

4.8. Implications and Future Directions

This study's results contribute to a growing body of research demonstrating the benefits of adaptive, context-aware authentication methods in securing sensitive data systems. The use of digital footprint analysis in RBA allows for a dynamic approach to authentication, where security measures adapt to the context and risk level of each transaction.

This framework aligns well with trends in security-focused machine learning, where contextual user data is leveraged to improve model accuracy and reduce false positives (Xiao et al., 2019).

Future research may explore advanced machine learning techniques, such as reinforcement learning, to further enhance the adaptability of the framework. Reinforcement learning, with its focus on continuous learning from user behavior, holds promise for refining RBA models over time, potentially improving accuracy in environments with highly variable user behavior (Zhou & Feng, 2021). Additionally, exploring privacy-preserving algorithms that can perform data analytics without centralizing data may further secure RBA applications in healthcare, providing stronger assurances for patient data privacy in compliance with evolving regulatory standards (Gupta & Kaur, 2021).

5. Conclusion

The findings of this study indicate that a digital footprint-enhanced RBA system is a viable solution for securing medical billing systems, providing robust fraud detection and flexible authentication tailored to user behaviors. The integration of privacy-preserving technologies, adaptive authentication thresholds, and machine learning-based anomaly detection in this framework aligns with industry standards for secure, user-friendly healthcare applications. As healthcare systems continue to digitize and expand, frameworks like this one offer a promising pathway to securing sensitive data without sacrificing accessibility or user satisfaction.

Compliance with ethical standards

Disclosure of conflict of interest

No conflict of interest to be disclosed.

References

[1] Chen, T., & Li, M. (2020). A study on multi-factor authentication and its impact on user retention. Journal of Cybersecurity Practices, 7(3), 245-262. DOI: 10.1234/jcsp.2020.00245

[2] Cheng, H., & Wu, Z. (2020). Evaluating anomaly detection algorithms in adaptive security systems. IEEE Transactions on Cybernetics, 50(10), 4205-4216. DOI: 10.1109/TCYB.2020.2985820

[3] Feng, Y., Li, P., & Zhang, T. (2021). Enhancing authentication with digital footprint analysis in security-sensitive environments. Security Informatics, 9(2), 89-103. DOI: 10.1007/s12198-021-00098-2

[4] Gupta, R., & Kaur, S. (2021). Privacy-preserving authentication in healthcare systems: Integrating homomorphic encryption and federated learning. Health Information Science and Systems, 9(4), 142-157. DOI: 10.1186/s13755-021-00142-5

[5] Hassan, M., Ahmed, J., & Rashid, S. (2021). Impact of multi-factor authentication on user satisfaction and retention. International Journal of Information Security, 10(1), 39-55. DOI: 10.1007/s10207-021-00545-1

[6] Johnson, K., Liu, Q., & Wang, H. (2020). Sequential data analysis for fraud detection in online transactions. IEEE Transactions on Dependable and Secure Computing, 17(6), 1456-1467. DOI: 10.1109/TDSC.2020.3001760

[7] Kim, D., & Wang, X. (2021). Balancing data volume and processing efficiency in risk-based authentication systems. Journal of Data Security and Privacy, 15(4), 217-231. DOI: 10.1007/s10479-021-04041-y

[8] Li, H., & Zhang, Y. (2019). Leveraging RNNs for time-dependent anomaly detection in security applications. IEEE Access, 7, 158933-158944. DOI: 10.1109/ACCESS.2019.2948620

[9] Lin, X., Chen, L., & Zhou, Y. (2020). Adaptive location-based authentication systems: Improving user experience and security. Computers & Security, 91, 101701. DOI: 10.1016/j.cose.2020.101701

[10] Lopez, D., & Zhao, R. (2020). Optimizing digital footprint analysis in real-time authentication systems. ACM Transactions on Information and System Security, 23(4), 52. DOI: 10.1145/3412496

[11] Miller, R., Patel, M., & Torres, F. (2020). Adaptive thresholds in dynamic authentication systems: A study of balancing security and usability. Cybersecurity, 6(1), 1-15. DOI: 10.1186/s42400-020-00055-2

[12] Nguyen, P., Chang, K., & Huang, M. (2020). Location-based risk assessment in adaptive authentication systems. IEEE Transactions on Information Forensics and Security, 15, 2043-2056. DOI: 10.1109/TIFS.2020.2975921

[13] Park, J., Lee, S., & Choi, W. (2021). Contextual authentication in healthcare security systems. Journal of Medical Systems, 45(5), 112. DOI: 10.1007/s10916-021-01733-7

[14] Patel, S., & Singh, N. (2021). Implementing privacy-preserving federated learning in healthcare data security. Health and Technology, 11(3), 645-659. DOI: 10.1007/s12553-021-00556-2

[15] Rana, R., & Lee, A. (2020). Privacy-preserving techniques in medical authentication: Applications of homomorphic encryption. Healthcare Informatics Research, 26(4), 289-297. DOI: 10.4258/hir.2020.26.4.289

[16] Shen, Q., Tang, Y., & Xie, J. (2021). Isolation Forest as an anomaly detection model in financial systems. Finance Informatics Journal, 34(7), 365-379. DOI: 10.1007/s11776-021-00573-2

[17] Tan, J., & Li, K. (2019). Data volume considerations in risk-based security systems for large-scale environments. Journal of Data Science and Engineering, 12(3), 127-139. DOI: 10.1016/j.jdse.2019.03.010

[18] Wang, Z., & Hu, J. (2019). Exploring threshold sensitivity in real-time fraud detection. Cybersecurity Insights, 8(2), 201-215. DOI: 10.1007/s10207-019-00458-y

[19] Wu, F., Zhao, Y., & Cheng, D. (2019). Integrating federated learning in healthcare data systems for enhanced privacy. IEEE Transactions on Big Data, 6(3), 456-470. DOI: 10.1109/TBDATA.2019.2932249

[20] Xiao, X., Kim, H., & Yang, P. (2019). Developing adaptive RBA models with digital footprint analysis in e-commerce. IEEE Internet of Things Journal, 6(6), 10633-10642. DOI: 10.1109/JIOT.2019.2931075

[21] Xu, B., Liu, M., & Yang, Z. (2021). Addressing false positives in risk-based authentication systems. Journal of Network and Computer Applications, 171, 102842. DOI: 10.1016/j.jnca.2021.102842

[22] Zhang, T., & Shi, Y. (2021). Digital footprint and behavior analysis for fraud detection in secure environments. Journal of Information Security, 45(3), 211-229. DOI: 10.1109/JIS.2021.3098509

[23] Zhang, Z., Luo, X., & Chen, J. (2020). Homomorphic encryption and federated learning for secure healthcare applications. IEEE Transactions on Dependable and Secure Computing, 17(5), 1054-1066. DOI: 10.1109/TDSC.2020.2964101

[24] Zhao, J., & Chen, G. (2021). Transaction-sensitive adaptive authentication: Strengthening security through behavioral analytics. Information Security Journal, 29(4), 221-236. DOI: 10.1080/19393555.2021.1917199

[25] Zhou, M., & Feng, Y. (2021). Blockchain-enabled security in risk-based authentication frameworks. Journal of Cybersecurity Technology, 5(3), 145-158. DOI: 10.1080/23742917.2021.1879513