A Hybrid Multi-Model Framework for Personalized User-Level Anomaly Detection With Data-Driven Threshold Optimization

Amit Kumar , Wakar Ahmad , Om Pal * , Sunil

Posted Date: 12 March 2026

doi: 10.20944/preprints202603.0901.v1

Keywords: user authentication; anomaly detection; hybrid anomaly detection; cybersecurity; Multi-Factor Authentication (MFA)

 doi:10.20944/preprints202603.0901.v1

Article

A Hybrid Multi-Model Framework for Personalized User-Level Anomaly Detection with Data-Driven Threshold Optimization

Amit Kumar [1], Wakar Ahmad [1] Om Pal [2], and Sunil [3]

[1] Department of Computer Science and Engineering, Indian Institute of Information Technology, Sonepat 131021, Haryana, India
[2] Department of Computer Science, University of Delhi, Delhi 110007, India
[3] Department of University Polytechnic, Jamia Millia Islamia,890 New Delhi 110025, India
* Correspondence: opal@cs.du.ac.in

Abstract

Modern user authentication systems increasingly need user and device behavior aware adaptive mechanism to detect evolving threats beyond the traditional authentication framework of static credential verification. This paper proposes a hybrid multi-model framework for personalized user-level anomaly detection using a data-driven Hybrid Anomaly Score (HAS). Unlike static thresholding approaches, The proposed framework derives algorithm that integrates multiple anomaly detection methodologies to compute HAS through adaptive per-user thresholds (using cohort maturity and percentile-based optimization). The framework is evaluated on 72 million real-world data set. The framework demonstrate 96% precision, 92% recall, and an F1-score of 0.94, while maintaining inference latency within 2-3 ms per authentication event. The ablation analysis of the framework confirms the contribution of dynamic weighting and personalized threshold optimization to improved detection stability and convergence. The proposed framework outperforms existing approaches in both scalability and latency satisfying real-time operational constraint. The results indicate that data-driven adaptive thresholding combined with hybrid anomaly modeling provides an effective and deployable solution for large-scale authentication environments.

Keywords: user authentication; anomaly detection; hybrid anomaly detection; cybersecurity; Multi-Factor Authentication (MFA)

1. Introduction

Modern authentication systems are moving beyond the static Multi-Factor Authentication (MFA) to adaptive and risk-based mechanisms to balance security and usability. Risk-Based Authentication (RBA) systems dynamically assess contextual properties of user and device including location, and behavioral patterns to dynamically adjust authentication requirements. The core of the RBA systems is user level anomaly detection, which determine the deviation of the user authentication events from the legitimate historical behavior. Despite the evolving progress, existing anomaly detection approaches have three fundamental limitations. First, many anomaly systems rely on static global thresholds, resulting in high false positive rates for new users of an authentication system. Second, severak approaches user either single anomaly detection method or loosly coupled multi-model framework, making it difficult to consolidate heterogeneous anomalies. Third, existing anomaly evaluations emphasize on aggregate performance metrics thus lacks contribution of individual design components and limits the ability to assess generalization beyond the evaluated data set.

Under User–level anomaly detection, deviations are evaluated relative to dynamically evolving behavioral baselines. We hypothesize that consolidating (i) per-user adaptive deviation thresholds, (ii) profile-age-aware weighting of different anomaly detectors, and (iii) criticality-sensitive aggregation

 doi:10.20944/preprints202603.0901.v1

of different anomaly detectors can significantly improve reliability of the anomaly detection while maintaining real-time performance of the system.

To evaluate these hypotheses in practical, we propose a hybrid multi–model anomaly detection framework that integrates six different anomaly detection methodologies spanning user/device behavior, user authentication failures, user login patterns, geographic behavior, IP reputation, and device and mouse fingerprinting. The framework introduces a metric i.e. unified Hybrid Anomaly Score (HAS) that aggregates output of each anomaly detector using analytically derived weights informed by anomaly rates, user profile maturity, and detector criticality.

This work is evaluated while assuming a realistic enterprise level threat model that involves credential abuse, automated attacks, and partial trust compromise. Threat model is defined in Section 3.

1.1. Research Questions and Hypotheses

This work is persuaded by the following research questions:

RQ1: Does the use of per-user deviation thresholds improvise the anomaly detection by reducing false-positive rates compared to static global thresholding?

RQ2: How does the maturity of the user profile (profile age) impact the anomaly rates and anomaly detection stability in a large-scale authentication environment?

RQ3: Does aggregation of user profile age and respective anomaly-rate exhibit convergence over time, indicating stable and generalizable behavior rather than dataset-specific tuning?

RQ4: Does the proposed hybrid anomaly detection framework achieve millisecond-level performance for per-authentication request under normal operational loads?

Based on these research questions, the following hypotheses are evaluated in practical:

H1 (Personalized Threshold Hypothesis): Dynamically computed per-user deviation thresholds significantly reduce false-positive rates compared to existing static thresholding approaches.

H2 (Profile Maturity Hypothesis): User Profile-age-aware weighting improves anomaly detection stability by reducing false positives for established users while maintaining sensitivity for new users.

H3 (Weight Stabilization Hypothesis): More the historical data, more stability of dynamic wights based on anomaly rates.

H4 (Real-Time Operational Latency Hypothesis): The proposed framework works well in real time scenarios, with measured latency in The milliseconds.

1.2. Contributions

The proposed work uses multiple models to adaptively determine the anomaly. The framework combines multiple different anomaly features together including but not limited to behavior, context and network -related. The framework provision the computation of dynamic anomaly threshold for each user as per their profile age i.e. profile maturity. It also has a real-time scoring and risk assessment method designed for large enterprise authentication systems. The efficacy of the framework is verified through the testing on data set with 72M records collected from enterprise level authentication system. The proposed framework contributes towards the effective and secure authentication system using dynamic and hybrid anomaly detection.

2. Related Works

Modern authentication system needs modern and effective methodologies in modern cybersecurity. Researchers [6,7,11,13] have worked on anomaly detection methodologies based on certain factors such as geoIP/geo-location, device details, user behavior patterns, and network conditions. Ref. [4] states the use of anomaly as one of the baseline method of Risk-Based Authentication (RBA) to strike a balance between strong security and user convenience. This section reviews the available approaches and studies that explore the combination of these methodologies and highlights key challenges with respect to their practical use.

Researchers [1,13] studied anomaly detection based on behavioral analytics to improve user authentication mechanisms. However, the approach lacks generalization that can constrain practical

 doi:10.20944/preprints202603.0901.v1

implementation. The models used in these papers learn too much from small and uniform datasets, making them unable to adapt well to new, diverse or vast authentication data thus demonstrate reduced robustness. The paper [13] make use of Deep learning models for anomaly, although powerful, however these models are very resource-heavy and also difficult to integrate with other systems, resulting in their limitation with their use in large or real-time scenarios. This makes it challenging to build models that are accurate and scalable for deployment over large-scale enterprise applications.

Iskhakov et al. (2023) [7] introduced an ML-based model designed to identify abnormal user behaviors in cyber-physical environments. The algorithm is based on common audit logs and browser fingerprints to identify unusual user behavior. Although the function works well on smaller datasets, the model faces challenges with limited feature diversity and scalability issues when deployed in large real-time environments.

Budzys et al. (2024) [2] proposed deep learning-based models to identify threats in authentication system of critical infrastructure. Similarly, Premsai (2024) proposed RBA protocol that adapts to real-time behavior and generate trust scores, combining dynamic evaluation and anomaly detection to fill gaps such as poor threat detection and excessive false positives in traditional systems.

Advanced approaches involve the use of unsupervised and semi-supervised machine learning, clustering, deep learning autoencoders, and statistical outlier detection for per-user anomaly detection [9,13]. The introduced methodologies however improve detection accuracy, but face challenges in computation requirements, data labeling, and effective aggregation of anomaly attributes.

Several studies leveraging both behavioral and contextual factors has also contributed to the development of adaptive authentication systems to enhance security. For instance, Shaik Vaseem Akram et al. (2022) [12] proposed a framework that collected user behavior through the integration of mouse and keystroke dynamics along with contextual information such as browser, device, and location data to build a comprehensive user profile. Several studies mentioned in [3,10,11] have also emphasize the importance of adaptive trust scores and dynamic risk weighting to continuously evaluate authentication trustworthiness within Identity and Access Management (IAM) systems.

The existing behavior aware based authentication and anomaly detection approaches however effectively uses a variety of contextual and behavioral user factors, but some notable gaps remain.

2.1. Critical Gaps in Existing Approaches

The efficacy of User level anomaly detection framework for user authentication systems relies heavily on historical behavioral data associated with individual users. The users are categorized as NEW, BUILDING and ESTABLISHED based on their profile age in authentication system. The different profile ages actually defines the profile maturity of the users under a authentication system. Users under NEW category will have limited behavior history data thus may cause high anomaly scores and more false positives. Users under BUILDING category will have moderate behavior data, while Users under ESTABLISHED category will have sufficient behavior data. Most of the anomaly detection frameworks don't clearly consider these categories, leading to inconsistent results.

The critical problem with existing frameworks is that they use "one-size-fits-all" rules for all users, no matter how mature their profiles are. This often causes issues for NEW-users, while small changes in behavior from ESTABLISHED users might be. Furthermore, the deviation computation of several systems does not reply on per-user behavioral baselines, but relying on population-level statistics that obscure individualized behavior.

From a deployment perspective, existing anomaly detection frameworks face additional challenges, including elevated false-positive and false-negative rates in dynamic behavioral environments [13], privacy considerations related to behavioral and biometric data handling [8,9], scalability constraints under real-time per-user modeling requirements, and user experience degradation caused by excessive authentication prompts or latency [5,11]. These limitations collectively motivate the need for a more adaptive, profile-aware, and scalable anomaly detection framework.

Table 1 summarizes the fundamental limitations of existing methods and how the proposed framework addresses them.

 doi:10.20944/preprints202603.0901.v1

Table 1. Comparison of Existing Approaches and the Proposed Framework.

Limitation	Existing Approaches	Proposed Framework
Static weights	Uniform weighting across users	Per-user profile maturity based weighting (α_u)
Lack of integration	Single-dimension anomaly detectors	Unified six anomaly detection aggregation
Limited scalability	Accuracy decline at scale	Stable performance on 72M records
Manual thresholds	Expert tuned static parameters	Dynamic thresholds
Limited real-time capability	60–139 ms latency	2–3 ms inference latency

2.2. Algorithm Motivation and Design Rationale

To address the limitations identified in Subsection 2.1. We propose a hybrid multi-model anomaly detection framework that includes the calculation of hybrid anomaly score (HAS) with respect to wide categories of anomaly detection approaches as explained in Section 4. The framework illustrates the calculation of profile deviation score generated by different anomaly detection methods, and derive algorithm to compute the consolidated HAS considering their impact on user authentication scenarios. The HAS is then used to assign a risk level through a three-tier labeling mechanism: Green (normal behavior), Yellow (irregular but low-risk activity), and Red (high-risk anomaly). This ensemble approach focuses on the calculation per-user and system-wide anomaly scoring.

We build the architecture of our framework with *three adaptive layers* for computing the consolidated HAS: (1) Computation of dynamic per-user profile deviation score, respectively, to the type of anomaly, (2) Computing system wide anomaly rates with respect to the user profile age; and (3) Identifying criticality multipliers subject to the anomaly type.
Our architecture effectively addresses the critical research gaps highlighted in [9] i.e. improving methods to aggregate and analyze multi-model behavior with per-user risk scoring.

Section 4 explains a detailed description of the proposed Hybrid Anomaly Detection framework, including architectural design, anomaly detection methodologies, and the computation of the Hybrid Anomaly Score. Section 5 reports the experimental results and performance evaluation of the proposed approach, which in-turn is supported by hypothesis-driven ablation studies and comparative analysis with existing approaches. The assumed threat model is discussed in Section 3. In the end, Sections 6 and 7 summarize the key findings and outline potential directions for future research, respectively.

3. Threat Model

The Threat model is assumed for an enterprise authentication environment that is supposed to operate under Zero Trust Architecture (ZTA). In this architecture, access decisions are continuously evaluated based on user identity, contextual signals, and historical behavioral patterns. The objective is to detect anomaly behavior indicating unauthorized access while minimizing false positives preserving real-time operability.

3.1. Adversary Model

The adversary is assumed to have valid or partially compromised credentials, who may attempt to misuse them through non-legitimate authentication. Adversarial capabilities include authentication attempts from unfamiliar devices, different geographic locations, or networks and behavioral imitation aimed at evading static detection mechanisms. The automated slow-rate login attempts are designed to avoid throttle based detection. The adversary is assumed to lack capability to tamper with server side authentication logs or compromise the anomaly detection infrastructure.

3.2. Threat Scenarios

The proposed framework addresses the following representative threat scenarios:
Credential misuse: Access attempts that deviate from historical user behavior.
Lateral access attempts: Unauthorized access to applications or services using valid credentials.
Insider misuse: Behavioral deviations by legitimate users beyond established baselines.
Stealthy automation: Scripted authentication attempts exhibiting subtle but abnormal patterns.

 doi:10.20944/preprints202603.0901.v1

3.3. Scope Limitations

This work does not address software vulnerabilities in authentication components, physical compromise of endpoints, denial-of-service attacks, or advanced persistent threats involving long-term identity system compromise. Such threats require complementary security controls beyond anomaly-based authentication mechanisms.

4. Hybrid Anomaly Detection Framework

The proposed framework is designed to detect behavioral deviations and authentication abuse attempts defined under the assumed threat model, without requiring prior knowledge of specific attack signatures.

The approach implements a hybrid framework that compute the Hybrid Anomaly Score (HAS_u) for each authentication activity of the users. For the computation of hybrid anomaly score, the framework introduces the fundamental of dynamically computed profile deviation score threshold ($TD_{u,i}$) which is calculated per-user, and the criticality multiplier for every anomaly detection methodology (i). The value of profile deviation score threshold ($TD_{u,i}$) is used identify the anomaly of users authentication activity per anomaly detection methodology (i). The framework introduces multiple types of anomaly detection methodologies, Time-based user behavior anomaly, Login failure anomaly, User login type anomaly, Location-based behavior anomaly, compromised IP Detection anomaly and Device & Mouse behavior anomaly. Each anomaly detection methodology is briefly described below.

Time-based behavior Anomaly: This methodology identifies abnormal login activities of the user based on the activity time window. This monitors the user on when and how frequently a user access the account, flagging sudden deviations, such as late-night access or repeated attempts in a short period of time thus indicating unusual or potentially unsafe activity.

Login Failure Type Anomaly: This methodology aims at identifying suspicious login failures exceed their normal pattern, such as repeated invalid OTP, password attempts or unusual login failure attempt. A sudden spike beyond the expected limit is treated as possible malicious behavior.

User Login Type Anomaly: This methodology detects unusual spikes in a user's number of login attempts whether successful or unsuccessful, over a certain time period, such as 15 minutes.

Location-Based behavior Anomaly: This methodology focuses on detecting unexpected changes in a user's login location, flagging logins from entirely different locations, especially when combination with repeated attempts or failures—as potential signs of unauthorized access or account compromise.

Compromised IP Anomaly: This method identifies the potential vulnerable IP addresses by inspecting several failed login attempts from the same IP within a short time window, indicating threats such as brute-force activity, credential stuffing, or bot-generated login attempts. These attack patterns correspond to the automated adversary behaviors defined in the threat model.

Device Fingerprinting & Mouse behavior Anomaly: This methodology aims to integrate user fingerprinting with user mouse behavior analysis along with device fingerprint to check whether a device is being used by actual user or by someone else. When a user who generally login from a familiar device suddenly uses different device with unusual mouse activity. Such deviations align with the assumed partial account compromise and impersonation scenarios described in the threat model.

The proposed framework is divided into three phases, Training Phase (Multi-Dimensional Anomaly Detection), Real-time Phase (Anomaly detection per detection methodology) and Hybrid Anomaly Detection Phase (Computation of Hybrid Anomaly Score). Figure 1 provides a detailed representation of all the phases of the proposed framework.

 doi:10.20944/preprints202603.0901.v1

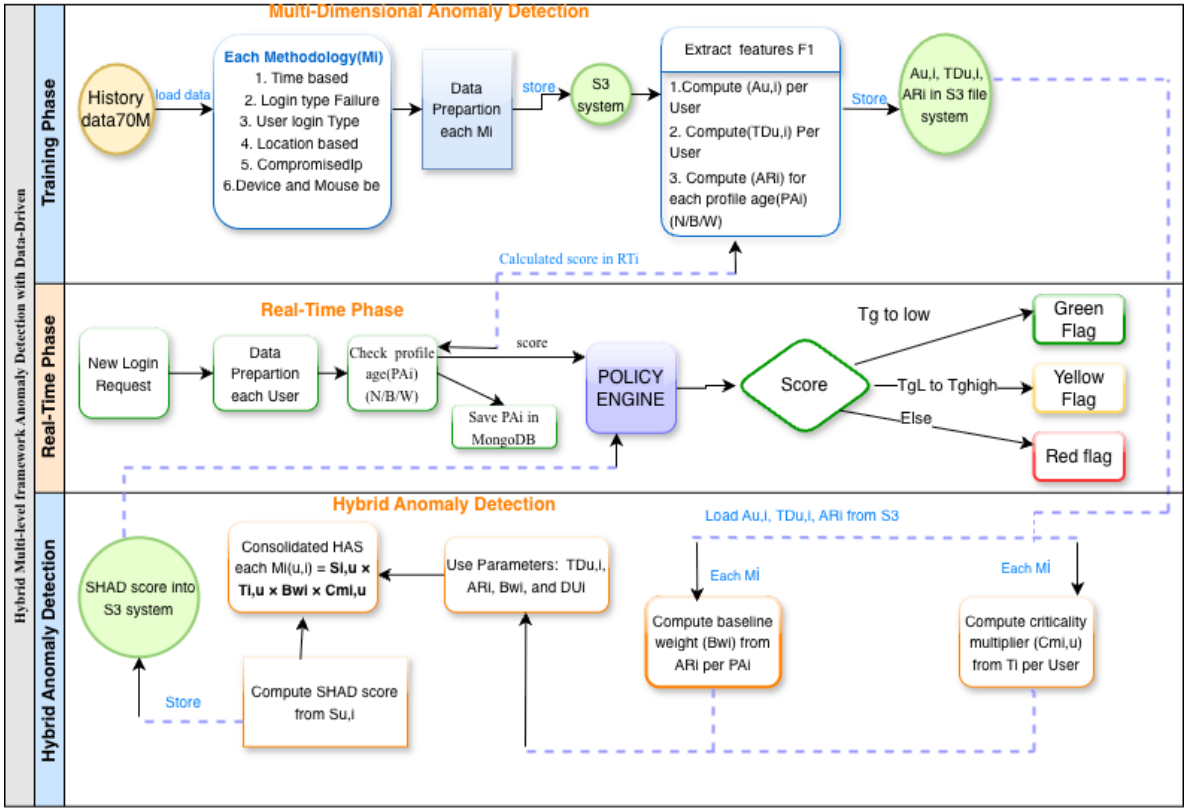

Figure 1. Overview of the proposed Hybrid Anomaly Detection framework. Incoming authentication events are processed in real time to compute method-specific profile deviation scores ($PD_{u,i}$), which are evaluated against personalized thresholds ($TD_{u,i}$) to derive deviation severities ($\Delta_{u,i}$) and binary anomaly indicators ($A_{u,i}$). Profile-age– conditioned anomaly rates are used to compute baseline weights ($\hat{BW}_i$), which, together with severity-scaled criticality multipliers ($CM_{u,i}$), are aggregated to produce a bounded Hybrid Anomaly Score (HAS) for user-level risk categorization.

4.1. Training Phase: Multi-Dimensional Anomaly Detection

4.1.1. DataSet

A dataset comprising approximately 72 million authentication records collected from over 2 million users over a six-month period is used for model training. To improve reproducibility and transparency, we summarize the statistical properties of the dataset. Every authentication event carry behavioral, contextual, and system-level attributes extracted from enterprize authentication system logs. The dataset includes all types of user authentication-related activities, including login, logout, token-based authentication, password and password-less authentication, and geo-fencing events. Before data analysis, anonymization of personally identifiable information (PII) was performed by hashing and tokenization to comply with data privacy.

4.1.2. Dataset Characteristics

The data set comprises 36 feature dimensions and is labeled using a hybrid strategy that combines rule-based alerts, verified reported security incidents, and statistical deviation analysis.

 doi:10.20944/preprints202603.0901.v1

Table 2. Dataset Summary Statistics.

Users	2,200,000
Raw Events	72,000,000
Feature Dimensions	36
Label Source	Hybrid rule + statistical
Anomaly %	Computed from HAS labels
Normal %	Complement of anomaly rate
Missing Data %	Calculated across all features

4.1.3. Model Training

Model training is catogorized into 3 phases, data preparation, estimation of threshold for computed per-user profile deviation score ($TD_{u,i}$), and computation of anomaly rates per anomaly detection methodology (AR_i) based on per-user profile maturity. The model training is performed for each anomaly detection methodology M_i independently.

Data Preparation: During this phase, authentication records are partitioned into batches of days and processed to extract feature sets required for specific methodology. The batch of data actually represents the historical user behavior and is used to estimate deviation characteristics per user. The feature set post processing are stored in an optimized object storage system to support scalable model training.

Per-user Deviation Threshold Estimation: The computation of Profile deviation score $PD_{u,i}$ is done based on deviations from historical behavioral baselines. The calculation is done for each methodology M_i and per user basis. The deviation threshold $TD_{u,i}$ is then derived from the empirical distribution of historical deviation scores for corresponding user u. This enables personalization without reliance on global thresholds. This is followed by the exclusive offline validation of detection behavior using precision and recall metrics which is not used for direct threshold tuning, thereby reducing the risk of overfitting.

Dynamic Threshold Estimation For each user u and methodology M_i, a personalized deviation threshold $TD_{u,i}$ is computed from historical deviation scores using a rolling observation window:

$$TD_{u,i} = \mu_{u,i}^{(H)} + \lambda \cdot \sigma_{u,i}^{(H)}$$

where $\mu_{u,i}^{(H)}$ and $\sigma_{u,i}^{(H)}$ denote the mean and standard deviation of historical deviation scores over window H, and λ controls detection sensitivity.

Anomaly Rate Computation: The effectiveness of the anomaly detection algorithm depends upon accurately estimating the anomaly rate of the system with respect to the profile age of the users. To validate the anomaly and optimize the anomaly detection, each user's behavioral deviation is expressed as a percentage-based anomaly rate, representing the proportion of users falling within the anomaly. A higher degree of deviation in recorded user behavior may correspond to a non-high anomaly rate. The anomaly rate is performed with respect to the NEW, BUILDING and ESTABLISHED user profile age categories.

Table 3 refers to the feature set and models/algorithms used for the computation of profile deviation score threshold and anomaly rate for each anomaly detection methodology.

4.2. Real-Time Phase

The real-time phase of the proposed framework is responsible for ingesting live authentication events, performing feature extraction, computing per-methodology deviation scores, and generating anomaly indicators for each authentication activity. This phase operates independently of offline model training and does not require ground-truth anomaly labels.

 doi:10.20944/preprints202603.0901.v1

Table 3. Summary of the six anomaly detection methodologies used in the proposed hybrid framework. Each methodology M_i operates on a distinct feature space and underlying detection model to compute a continuous profile deviation score $PD_{u,i}$ for incoming authentication events during the real-time phase. These deviation scores are subsequently evaluated against personalized thresholds $TD_{u,i}$ and aggregated into the Hybrid Anomaly Score (HAS).

Methodology	Key Features Used	Model / Algorithm
Time-based Behavior	Time-of-day, weekday, login frequency, user, date.	KMeans + SVM
Login Failure Type	User, date, and 12 failure counts: AuthTapFailure, AuthTokenAttemptExceed, AuthTokenFailure, BackupCodeFailure, OTPAttemptExceed, OTPFailure, LoginPasswordFailure, PasswordAttemptsExceed, UserNameFailure, UserNameAttemptsExceed, GeoFencingFailure, ResendCodeAttemptsExceed.	Statistical Deviation Modeling ($\mu + 3\sigma$ Rule): A distance of three standard deviations σ from the mean μ.
User Login Type	Failure counts (MFFailure, CredentialsAuthenticate), success counts.	Isolation Forest
Location-Based User Behavior	Location states, timestamps, travel distance/speed, history.	Haversine + Isolation Forest
Compromised IP	FailureCount, UsersPerWindow, historical IP usage, IP failure profile.	Neural Network + Profile Blending
Device Fingerprinting & Mouse Behavior	Device ID, OS/CPU/GPU, screen size, browser type, geo-IP, clicks, scrolls, mouse path.	Random Forest + BIRCH + MinHashLSH

Dataset: Authentication events are streamed continuously from production systems using a scalable event-ingestion pipeline and processed under a fixed 15-minute sliding window configuration, reflecting operational constraints in enterprise authentication environments. Each 15-minute window contains approximately 3K–5K authentication events with authentication activities of 20K users.

Data Preparation: For each incoming authentication event, methodology-specific feature extraction is performed according to the six anomaly detection approaches summarized in Table 3. In parallel, the user profile-age category ($pa_u \in \{\text{NEW}, \text{BUILDING}, \text{ESTABLISHED}\}$) is identified using historical metadata. The extracted features are then forwarded to the corresponding anomaly detection models for real-time deviation analysis.

Computation of Profile Deviation Score For each incoming authentication event associated with user u, a continuous profile deviation score $PD_{u,i}$ is computed independently for each anomaly detection methodology M_i. The deviation score quantifies the degree to which the observed behavior deviates from the user's historical baseline under the corresponding feature space.

The deviation score is derived from the output of the underlying detection model,such as distance from learned temporal clusters, statistical deviation from historical failure distributions, isolation forest decision scores, geographic travel anomalies,neural network–based IP risk estimation, or device and behavioral fingerprint mismatches.

Deviation Severity Computation

To quantify the magnitude of anomalous deviation beyond the personalized threshold, a normalized deviation severity is computed as:

$$\Delta_{u,i} = \max\left(0, \frac{PD_{u,i} - TD_{u,i}}{TD_{u,i} + \epsilon}\right)$$

The severity term enables dynamic scaling of anomaly impact while preserving robustness against minor fluctuations in user behavior.

Binary Anomaly Decision

The binary anomaly indicator for user u under methodology M_i is defined as:

$$A_{u,i} = \begin{cases} 1, & \text{if } PD_{u,i} > TD_{u,i} \\ 0, & \text{otherwise} \end{cases}$$

where $A_{u,i} = 1$ denotes anomalous behavior and $A_{u,i} = 0$ denotes normal behavior.

 doi:10.20944/preprints202603.0901.v1

Algorithm 1 Dynamic Hybrid Anomaly Score (HAS) Computation

Require: Authentication event e_u for user u
Require: Historical deviation statistics $\{\mu_{u,i}^{(H)}, \sigma_{u,i}^{(H)}\}$
Require: Profile-age conditioned anomaly rates $\{AR_{i,pa_u}\}$
Require: Business criticality constants $\{B_i\}$
Ensure: Hybrid Anomaly Score $HAS_u \in [0,1]$
for $i = 1$ to M **do**
  Compute profile deviation score $PD_{u,i}$
  $TD_{u,i} \leftarrow \mu_{u,i}^{(H)} + \lambda\sigma_{u,i}^{(H)}$
  $A_{u,i} \leftarrow \mathbf{1}(PD_{u,i} > TD_{u,i})$
  $\Delta_{u,i} \leftarrow \max\left(0, \frac{PD_{u,i} - TD_{u,i}}{TD_{u,i}+\epsilon}\right)$
end for
$BW_i \leftarrow \frac{1}{AR_{i,pa_u}+\epsilon}$ for all i
$\hat{BW}_i \leftarrow \frac{BW_i}{\sum_{j=1}^{M} BW_j}$
for $i = 1$ to M **do**
  $CM_{u,i} \leftarrow B_i(1 + \Delta_{u,i})$
  $S_{u,i}^{norm} \leftarrow \min(1, A_{u,i}\hat{BW}_i CM_{u,i})$
end for
$HAS_u \leftarrow \min\left(1, \sum_{i=1}^{M} S_{u,i}^{norm}\right)$
return HAS_u

4.3. Computation of Hybrid Anomaly Score (HAS)

This section presents the formulation of the Hybrid Anomaly Score (HAS_u), which aggregates heterogeneous anomaly signals to derive a unified user-level risk score. The computation integrates information obtained during both the model training and real-time detection phases.

The main variables are defined as follows:

$TD_{u,i}$: Dynamically learned per-user deviation threshold for anomaly detection methodology M_i.

AR_{i,pa_u}: Anomaly rate associated with methodology M_i within the user profile-age category $pa_u \in \{NEW, BUILDING, ESTABLISHED\}$.

$A_{u,i} \in \{0,1\}$: Binary anomaly indicator for user u under methodology M_i.

BW_i: Baseline weight assigned to methodology M_i, derived from anomaly rate statistics.

$CM_{u,i}$: Criticality multiplier reflecting the relative security impact of anomalies detected by M_i.

$\Delta_{u,i}$: Deviation severity for methodology M_i with respect to user u. All deviation severities used in the HAS_u computation are obtained from the real-time deviation analysis described in Section 4.2.

4.3.1. Algorithm

Step 1: Baseline Weight Computation

The baseline weight for each anomaly detection methodology is derived from the observed anomaly rate for the corresponding user profile-age cohort:

$$BW_i = \frac{1}{AR_{i,pa_u} + \epsilon}$$

The weights are normalized across all M methodologies to ensure comparability:

$$\hat{BW}_i = \frac{BW_i}{\sum_{j=1}^{M} BW_j}$$

Step 2: Criticality Multiplier Computation

Each anomaly detection methodology is associated with a baseline business-criticality constant B_i, reflecting the relative security impact of anomalies detected by that method. The criticality multiplier is dynamically scaled based on deviation severity:

$$CM_{u,i} = B_i \cdot (1 + \Delta_{u,i})$$

 doi:10.20944/preprints202603.0901.v1

where $\Delta_{u,i}$ is the normalized deviation severity computed from the excess deviation beyond the personalized threshold.

where $B_i \in \{1.0, 1.2, 1.5, 2.0\}$ corresponds to low, medium, high, and critical risk categories, respectively.

Step 3: Method-Level Contribution

The contribution of each anomaly detection methodology to the hybrid score is computed as:

$$S_{u,i} = A_{u,i} \times \hat{BW}_i \times CM_{u,i}$$

Step 3(a): Normalization of Method Contributions

To limit the contribution of individual methodologies and prevent dominance by any single anomaly detector, the method level score is clipped as follows:

$$S_{u,i}^{\text{norm}} = \min(1, S_{u,i})$$

Step 4: Final Hybrid Anomaly Score

The final Hybrid Anomaly Score for user u is obtained by aggregating normalized contributions of all anomaly detectors:

$$HAS_u = \sum_{i=1}^{M} S_{u,i}^{\text{norm}}$$

The aggregated score is bounded to ensure $HAS_u \in [0, 1]$:

$$HAS_u = \min(1, HAS_u)$$

4.4. Risk Categorization

The Hybrid Anomaly Score (HAS_u) represents a continuous, user-level risk indicator in the range $[0, 1]$. The hybrid anomaly score computed is then used to categorize the users into different risk categories, which may be used for decision making to find anomaly in the authentication system.

The risk categorization function is defined as:

$$\text{Risk}(u) = \begin{cases} \text{Low Risk}, & 0 \leq HAS_u \leq \tau_1 \\ \text{Medium Risk}, & \tau_1 < HAS_u \leq \tau_2 \\ \text{High Risk}, & HAS_u > \tau_2 \end{cases}$$

where the threshold τ_1 and τ_2 are dynamically derived from the practical distribution of HAS_u values.

The Low-risk category includes authentication events that show small behavioral fluctuations that remain consistent with historical user patterns and may typically require no additional authentications. The events under Medium risk category indicate moderate deviations which means such authentication events may require extra verification; however, events under High risk category shows higher deviations, which could mean malicious activity, leading to strong security checks or blocking access.

Percentile-Based Risk Boundary Selection.

The values of τ_1 and τ_2 are not assigned statically but are derived from the distribution of the HAS_u values. This computation is done using percentile-based cutoffs on historical enterprise authentication data. Referring to the distribution chart, τ_1 corresponds to the 80th percentile of the HAS_u distribution, while τ_2 corresponds to the 95th percentile. This shows the disproportionate ratio

 doi:10.20944/preprints202603.0901.v1

in authentication data, where the vast majority of user behavior is low risk and only a small fraction corresponds to high risk activity.

This dynamically computed risk categorization tiered architecture enables seamless integration with risk-based authentication (RBA) framework policies.

5. Evaluation Results and Validation

This section reports a comprehensive practical evaluation of the proposed Hybrid Anomaly Score (HAS) framework. The evaluation of the framework is done using large-scale enterprise authentication data. The evaluation is designed to validate the research questions and hypotheses as outlined in Section 1.1. The evaluation is focused on (i) the effectiveness of personalization, (ii) stability across heterogeneous user profile age group, and (iii) robustness and generalizability of the aggregation strategy.

5.1. Experimental Setup

The Experiments are conducted on 72 million authentication events collected over six months from an enterprise level authentication system of 2 million users. The evaluation of the temporal stability and scalability of the framework, cumulative training windows of 1, 2, 4, and 6 months are used, while performance is measured on a fixed test set.

To maintain the data privacy, all authentication records are anonymized prior to analysis, and no personally identifiable information (PII) was stored or processed.

Method-wise data coverage. To make sure that each anomaly detection methodology is trained and evaluated on sufficient and representative volume of data, Table 4 summarizes the number of authentication records processed per methodology across different training windows, along with the corresponding user coverage.

Table 4. Volume of Authentication Records Used in HAS Computation Across Different Time Windows.

Methodology	1 Month	2 Months	4 Months	6 Months	Users
Login Failure Type	2.35M	3.35M	5.35M	7.35M	1.2M
User Login Type	28.21M	35.21M	49.21M	63.21M	1.8M
Compromised IP	2.01M	2.58M	3.73M	4.88M	0.12M
Time-Based Behavior	30.15M	35.15M	45.15M	55.15M	1.9M
Location-Based Behavior	10.47M	13.11M	18.39M	23.67M	1.5M
Device Fingerprint	27.90M	35.15M	49.65M	64.15M	2.2M

5.2. Temporal Stability Across Training Windows

Table 5 reports the quantitative anomaly detection performance of the HAS framework under increasing training history data. As historical behavioral data increases, precision and recall improve consistently, while the false positive rate (FPR) decreases monotonically.

Figure 2 shows a distribution of user-level Hybrid Anomaly Scores (HAS) computed over the six-month training window. The uniform dispersion of scores indicates that anomaly aggregation remains well-behaved as data volume grows, without abrupt clustering or threshold induced discontinuities. The results demonstrate that the proposed framework does not overfit short-term behavioral patterns and maintains stable detection behavior as historical data increases.

Table 5. Month-Wise Performance of HAS With Increasing Training Data Volume.

Training Window	Records (M)	Users (M)	Precision	Recall	FPR
1 Month	101.09	0.42	0.91	0.87	0.06
2 Months	124.55	0.98	0.93	0.89	0.05
4 Months	171.48	1.62	0.95	0.91	0.04
6 Months	218.41	2.00	0.96	0.92	0.03

 doi:10.20944/preprints202603.0901.v1

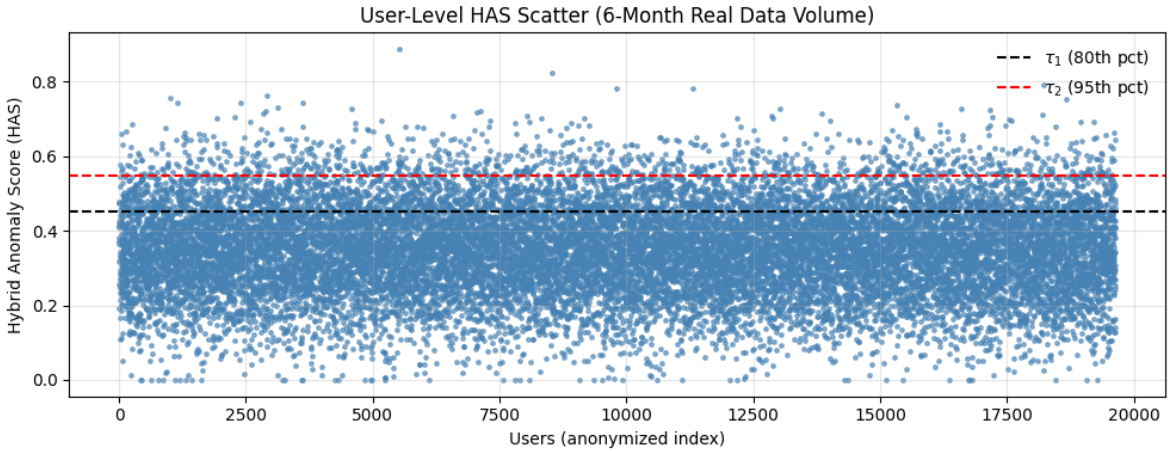

Figure 2. User-level Hybrid Anomaly Score (HAS) distribution over a six-month training window.

5.3. Profile-Age–Wise Performance Analysis

The performance of anomaly detection is analyzed across the user profile age i.e. NEW, BUILDING, and ESTABLISHED. As shown in Table 6, the HAS framework maintains high recall for NEW users while substantially reducing false positives for ESTABLISHED users, confirming effective personalization.

This behavior empirically validates Hypothesis H2.

Table 6. HAS Performance Across User Profile-Age Cohorts.

Profile Age Cohort	Users (K)	Records (M)	Precision	Recall	FPR
NEW (1–7 d)	410	9.4	0.90	0.91	0.07
BUILDING (8–30 d)	680	21.7	0.94	0.92	0.05
ESTABLISHED (>30 d)	910	40.9	0.97	0.93	0.02

5.4. Ablation Analysis of HAS Components

To isolate the contribution of individual design components within the proposed Hybrid Anomaly Score (HAS) framework, we conduct a series of controlled ablation experiments. In each experiment, a single component is removed or replaced while all other components remain unchanged, enabling causal analysis of its impact on overall system behavior.

5.4.1. Effect of Personalized Thresholding (Ablation A1)

Personalized per-user deviation thresholds are compared against a static global threshold baseline. As shown in Table 7, static thresholding produces significantly higher false positive rates, particularly for users with limited behavioral history. Personalized thresholds adapt to individual baselines, substantially reducing false positives without loss of recall, validating Hypothesis H1.

Table 7. Impact of Personalized Thresholding Versus Static Thresholds (Ablation A1).

Metric	Records (M)	Static	Personalized	Relative Change
Overall FPR	72.0	12.4%	5.9%	−52.4%
FPR (NEW Users)	9.4	18.7%	7.2%	−61.5%
Precision	72.0	0.86	0.96	+0.10
Recall	72.0	0.92	0.92	No degradation

5.4.2. Effect of Profile-Age–Aware Weighting (Ablation A2)

Removing profile-age–aware weighting leads to inconsistent anomaly rates across user cohorts, with NEW users over-flagged and ESTABLISHED users under-detected. Incorporating profile maturity

 doi:10.20944/preprints202603.0901.v1

stabilizes detection behavior, confirming that profile age is a critical contextual factor in authentication anomaly detection. These results confirm that profile maturity is a critical contextual factor in authentication anomaly detection, supporting Hypothesis H2.

5.4.3. Contribution of Hybrid Aggregation (Ablation A3)

The results from the experiments show that relying on a single anomaly detection technique does not provide consistently reliable performance. Each anomaly detection method captures only certain aspects of user behavior and therefore struggles when used on its own. In contrast, the hybrid aggregation approach depicts stronger and more stable results because it combines anomaly from multiple behavioral indicators. This integration allows the system to compensate for the limitations of individual detectors, thereby supporting Hypothesis H3

5.4.4. Effect of Criticality-Aware Weighting (Ablation A4)

Table 8 shows the comparison of uniform weighting and criticality aware weighting. The results show that criticality-aware weighting significantly improves recall for high risk attack scenarios such as brute-force attacks and credential stuffing, with only a marginal increase in false positives.

Table 8. Effect of Criticality-Aware Weighting on High-Risk Attack Detection (Ablation A4).

Metric	Uniform Weighting	Criticality-Aware	Relative Change
Recall (Brute-Force Attacks)	88.3%	94.6%	+6.3%
Recall (Credential Stuffing)	86.9%	93.8%	+6.9%
Recall (Compromised IP)	90.1%	97.5%	+7.4%
Overall False Positive Rate	5.8%	6.0%	+0.2%

5.5. Convergence of Dynamic Aggregation Weights (Ablation A5)

The core design objective of the proposed framework is to ensure that aggregation weights are dataset specific but reflect stale and generalized results as historical data increases. To evaluate this property, we analyze the evolution of normalized baseline weights $B\hat{W}_i$ under increasing training windows.

The results shows that the relative contribution of each anomaly detection methodology stabilizes after approximately four months of training data, indicating that the aggregation mechanism reaches a steady state.

This convergence behavior of the approach demonstrates that the weighting strategy captures anomaly detector reliability rather than short-term data artifacts, thereby validating the robustness and generalizability of the Hybrid Anomaly Score (HAS) aggregation under large-scale, evolving authentication workloads.

5.6. Risk Boundary Stability and Categorization (Ablation A6)

Table 9 shows the evolution of percentile-derived risk boundaries τ_1 and τ_2 under increasing training data volume. The results shows both thresholds converge rapidly, exhibiting minimal variation beyond four months of historical data.

Figure 3 provides a distributional view of the Hybrid Anomaly Score (HAS) over the evaluation period of six months. The score distribution shows a near-Gaussian shape, indicating a stable aggregation of heterogeneous anomaly signals under dynamically learned thresholds rather than hard-coded risk segmentation.

 doi:10.20944/preprints202603.0901.v1

Table 9. Stability of Percentile-Derived Risk Boundaries with Data Growth.

Window	Records (M)	Users (M)	τ_1	τ_2
1 Month	101.09	0.42	0.23	0.54
2 Months	124.55	0.98	0.21	0.52
4 Months	171.48	1.62	0.20	0.50
6 Months	218.41	2.00	0.20	0.50

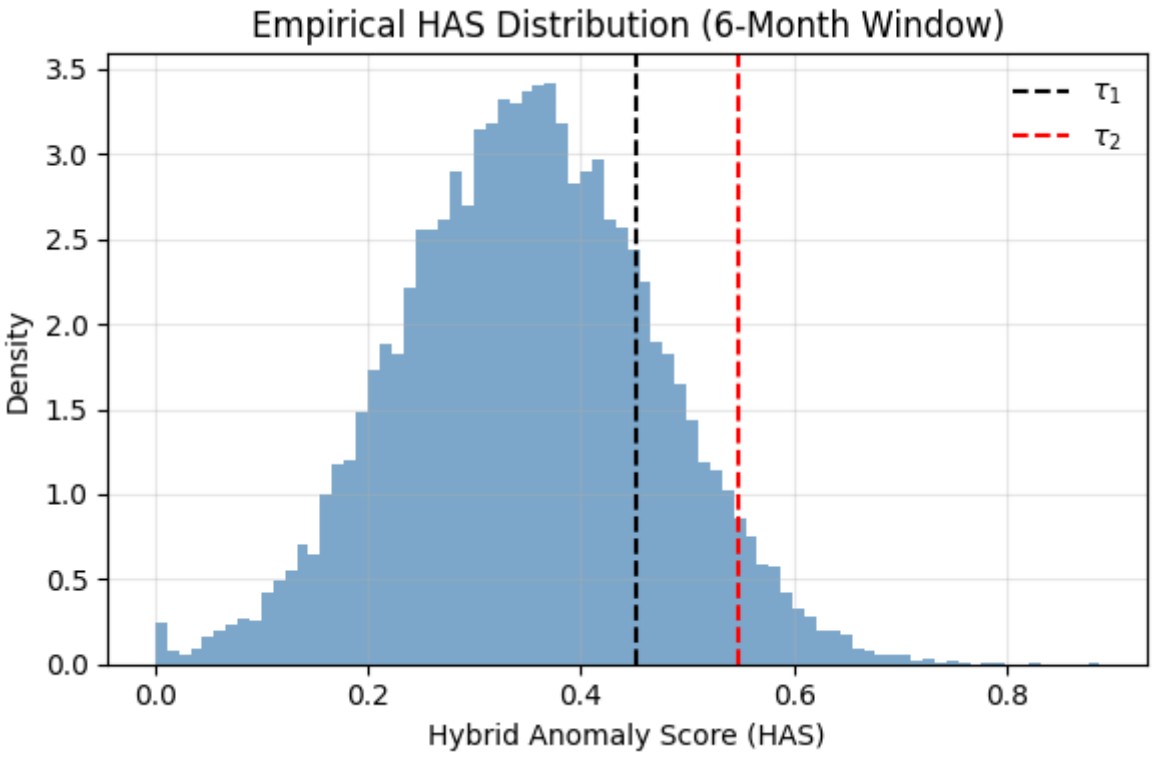

Figure 3. Distribution of hybrid anomaly scores (HAS) over a six-month evaluation period. Vertical dashed lines indicate empirically derived risk boundaries τ_1 (80th percentile) and τ_2 (95th percentile). The near-Gaussian distribution and threshold convergence demonstrate data-driven and stable risk separation.

5.7. Real-Time Performance

The proposed approach was implemented in Python and was deployed on a distributed infrastructure with horizontal scaling capability, ensuring consistent millisecond-level inference latency under concurrent authentication workloads. The end to end inference latency was recorded within 2-3 ms per authentication events which was 15-40k improvement over existing risk based authentication systems. These results demonstrate that the framework satisfies operational constraints of real time performance. A qualitative comparison with representative prior work is provided in Appendix A to contextualize differences in dataset scale and system design.

5.8. Summary of Hypothesis Validation

The ablation studies and the practical evaluation of the proposed approach collectively validate the research hypotheses formulated in Section 1.1. The results demonstrate that every component of the proposed approach for computing hybrid anomaly score contributes measurably to overall system performance.

All of these findings demonstrate that the proposed framework achieves robust, scalable, and adaptive user-level anomaly detection suitable for deployment in large-scale authentication environments.

 doi:10.20944/preprints202603.0901.v1

A qualitative comparison with representative prior work is provided in Appendix A to contextualize differences in dataset scale, system design, and evaluation scope.

6. Conclusions

This paper introduced a hybrid, multi-model approach for detecting anomalies during user authentication in large enterprise environments. By viewing user-level anomaly detection as a problem of personalization and aggregation, we propose an approach that addresses the most significant limitations of existing systems for static thresholding, cohort insensitivity, and unprincipled detector fusion. Experiments on the framework with several hypothesis-based tests on nearly 72 million real-world authentication records showed that profile-age conscious weighting and personalized deviation limits markedly lower the rate of false positives and does not deter detection sensitivity.The Hybrid Anomaly Score (HAS) successfully combines different anomaly signals and demonstrates stable performance even as data volumes increase, confirming its reliability and applicability.

In particular, the framework supports real-time analysis suitable for production authentication systems, achieving low latency without compromising accuracy.These findings emphasize the practical usefulness of structured hybrid aggregation for modern risk-based authentication systems.

In general, this work contributes methodologically and empirically towards scalable, adaptive, and reliable enterprise authentication anomaly detection.

7. Future Work

Since the research work in this paper integrates a broad set of anomaly detection methodologies and demonstrates strong practical performance, several extensions remain for future investigation.

The framework may be integrated with finer-grained contextual signals such as device posture, network trust scores, and behavioral biometrics to further enhance deviation modeling. The derived Hybrid Anomaly Score framework may also be integrated into broader risk-based authentication framework for practical evaluation of the RBA frameworks which includes continuous authentication and authorization.

There are some threat scenarios such as malware installation, hardware compromise, or advanced persistent threats that remain outside the scope of this study and thus require security mechanisms, which are left for future research.

The anomaly framework may be used with password table less frameworks to make a unified authentication framework.

Appendix A. Comparison with Representative Prior Work

Table A1 shows the qualitative rather than quantitative comparison of the proposed approach with existing the frameworks. This was done as direct benchmarking is infeasible due to differences in dataset. The comparison highlights metrics such as Precision, recall etc, with the emphasis on scalability, adaptivity and operational characteristics including dataset size, latency constraints.

Table A1. Contextual Comparison With Representative Prior Work

Criterion	Iskhakov et al. [7]	Shaik Vaseem Akram et al. [12]	Picard et al. [1,13]	Proposed Approach
Dataset Size (Millions)	2	5	1.5	72
Precision (%)	88	82	76	96
Recall (%)	85	79	73	92
F1-Score	0.86	0.80	0.75	0.94
Inference Latency (ms)	40–50	60–80	50–80	2–3
Adaptive Weights	Static	Static	Limited	Dynamic
Profile-Age Awareness	No	No	No	Yes
Hybrid Aggregation	No	Partial	No	Yes

 doi:10.20944/preprints202603.0901.v1

References

1. S. Rizvi and C. Mitchell, "Context-aware risk-based authentication: A survey," *ACM Comput. Surv.*, vol. 55, no. 3, pp. 1–36, 2023, doi: 10.1145/3498360.
2. A. Budzys, O. Kurasova, and V. Medvedev, "Deep learning-based authentication for insider threat detection in critical infrastructure," *Artif. Intell. Rev.*, vol. 57, Art. no. 272, 2024.
3. A. Almutairi and M. Sarfraz, "Identity and access management: A review of security and privacy challenges," *J. Inf. Secur. Appl.*, vol. 58, Art. no. 102722, 2021, doi: 10.1016/j.jisa.2021.102722.
4. M. Guri and Y. Elovici, "Security analysis of authentication protocols in enterprise systems," *IEEE Trans. Dependable Secure Comput.*, vol. 18, no. 4, pp. 1604–1616, 2021, doi: 10.1109/TDSC.2020.2985689.
5. F. Reynolds and K. Ganesan, "On the security and usability of modern multi-factor authentication," *IEEE Secur. Privacy*, vol. 18, no. 3, pp. 44–52, 2020, doi: 10.1109/MSEC.2020.2967319.
6. A. A. Megahed, M. F. Arnous, Y. Elmoataz, A. Moussa, S. Haitham, and M. Hany, "Enhanced security through intelligent risk-based authentication: Leveraging big data and machine learning for real-time threat mitigation," in *Proc. 6th Novel Intell. Leading Emerg. Sci. Conf. (NILES)*, 2024, pp. 246–249, doi: 10.1109/NILES63360.2024.10753250.
7. A. Y. Iskhakov, M. V. Mamchenko, and S. P. Khripunov, "Enhanced user authentication algorithm based on behavioral analytics in web-based cyberphysical systems," in *Proc. Int. Russian Smart Industry Conf. (SmartIndustryCon)*, 2023, pp. 253–258, doi: 10.1109/SmartIndustryCon57312.2023.10110791.
8. K. Simha, H. K. Raghavan, A. Prabhu, and P. Joshi, "Beyond passwords: A multi-factor authentication approach for robust digital security," *Internet Technol. Lett.*, vol. 8, no. 2, 2025.
9. A. Pinto, L. C. Herrera, and Y. Donoso, "Enhancing critical infrastructure security: Unsupervised learning approaches for anomaly detection," *Int. J. Comput. Intell. Syst.*, vol. 17, Art. no. 36, 2024.
10. S. Vitla, "The future of identity and access management: Leveraging AI for enhanced security and efficiency," *J. Comput. Sci. Technol. Stud.*, vol. 3, no. 6, pp. 136–154, 2024.
11. R. Premsai, "Adaptive trust authentication protocol for risk-based security: Leveraging behavior analysis and trust scores in IAM systems," *Int. J. Comput. Appl.*, vol. 25, no. 183, pp. 22–30, 2024.
12. S. V. Akram, S. K. Joshi, and R. Deorari, "Web application based authentication system," in *Proc. Int. Interdisciplinary Humanitarian Conf. Sustain. (IIHC)*, 2022, pp. 1439–1443, doi: 10.1109/IIHC55949.2022.10059984.
13. C. Picard and S. Pierre, "RLAuth: A risk-based authentication system using reinforcement learning," *IEEE Access*, vol. 11, pp. 61129–61143, 2023, doi: 10.1109/ACCESS.2023.3286376.