Article

Not peer-reviewed version

A Robust Behavioral Biometrics Framework for Smartphone Authentication via Hybrid Machine Learning and TOPSIS

Moceheb Lazam Shuwandy * , Qutaiba Alasad , Maytham M Hammood , A. A. Yass , Salwa khalid Abdulateef , Rawan A. Alsharida , Sahar Lazim Qaddoori , Saadi Hamad Thalij , Maath Frman , Abdulsalam Hamid Kutaibani , Noor S. Abd

Posted Date: 4 March 2025

doi: 10.20944/preprints202503.0084.v1

Keywords: smartphone authentication systems; cybersecurity; behavioral biometrics; machine learning; TOPSIS

 doi:10.20944/preprints202503.0084.v1

Article

A Robust Behavioral Biometrics Framework for Smartphone Authentication via Hybrid Machine Learning and TOPSIS

Moceheb Lazam Shuwandy [1,*], Qutaiba Alasad [1], Maytham M Hammood [1], A. A. Yass [1], Salwa khalid Abdulateef [2], Rawan A. Alsharida [1], Sahar Lazim Qaddoori [3], Saadi Hamad Thalij [1], Maath Frman [2], Abdulsalam Hamid Kutaibani [2] and Noor S. Abd [1]

[1] Cybersecurity Department, College of Computer Science and Mathematics, Tikrit University, Iraq
[2] Computer Science Department, College of Computer Science and Mathematics, Tikrit University, Iraq
[3] Electronic Engineering Department, Electronics Engineering College, Ninevah University, Iraq
* Correspondence: moceheb@tu.edu.iq

Abstract: Significant vulnerabilities in traditional authentication systems have been demonstrated due to the highly dependency on smartphone hardware devices to execute many different and complicated tasks. PINs, Passwords, and static biometric techniques have been shown to be subjected to various serious attacks, such as environmental limitations, spoofing, and brute force attacks, and this in turn mitigates the security level of the entire system. In this study, a robust framework for smartphone authentication is presented. Touch dynamic pattern recognitions, including trajectory curvature, touch pressure, acceleration, 2 dimensional spatial coordinates, and velocity, have been extracted and assessed as behavioral biometric features. TOPSIS, Technique for Order of Preference by Similarity to Ideal Solution, methodology has also been incorporated to get the most affected and valuable features, in which they are then fed as input to three different Machine Learning (ML) algorithms: Random Forest (RF), Gradient Boosting Machines (GBM), and K-Nearest Neighbors (KNN). Our analysis, supported by experimental results, ensure that the RF model outperforms the two other ML algorithms by getting F1-score, accuracy, recall, and precision of 95.1%, 95.2%, 95.5%, and 94.8%, respectively. In order to further increase the resiliency of the proposed technique, data perturbation approach, including temporal scaling and noise insertion, has been augmented. Also, the proposal has been shown to be resilient against both environmental variation-based attacks by achieving accuracy above 93% and spoofing attacks by obtaining a detection rate of 96%. This emphasizes that the proposed technique provides a promising solution to many authentication issues and offers user-friendly and scalable method to improve the security of the smartphone against cybersecurity attacks.

Keywords: smartphone authentication systems; cybersecurity; behavioral biometrics; machine learning; TOPSIS

1. Introduction

The need for resilient and secure authentication systems has been highly requested due to the continuous incremental increase in the dependency on smartphones in daily activities for professional, personal, and financial operations. It has been pointed out that even though traditional authentication approaches, such as PINs, passwords, and static biometrics, still provide a certain security level in a system, they have been shown to be subjected to many threats, including data sensitivity leakage, phishing, theft, and brute force attacks [1]. Other static biometric methods, e.g., facial recognition and fingerprint, have enhanced and increased the security of the system; however, it has been proven that they are subject to static data constraints, environmental conditions, and

 doi:10.20944/preprints202503.0084.v1

spoofing-based attacks [2]. To address the aforementioned shortcomings and improve usability and security, dynamic promising solutions should be adopted [3].

Behavioral biometrics have been considered as one of the best prominent solutions for system authentication issues since they analyze in-depth the personal specific interaction patterns, including swipe gestures, touch pressure, and (X, Y) coordinate of spatial movements. These approaches can repeatedly provide authentication and also be adapted to dynamic user behavior compared to traditional approaches, such as static techniques. Motion data and touch dynamics have been investigated by many prior works in order to precisely recognize unique user features; however, they are not strong enough to resist real-world variations and conditions, e.g., counterfeit attempts and environmental and climate modifications [4,5]. Such problems ensure the real need for more resilient techniques.

Interestingly, in order to further refine authentication systems, the three-dimensional (3D) touch sensors have been widely leveraged in many different applications. For instance, to get unique interaction patterns for each person, contemporary techniques have been implemented using 3D touch position and pressure sensitivity, and this significantly elevates the security level and the robustness of the system [6]. Other methodologies, e.g., audio and sound sensors, have been augmented with the 3D touch sensors to be employed for mobile healthcare in order to maintain the data of the patient private and secure [7]. To significantly refine the security level of a design, the biometric data, e.g., electroencephalogram (EEG) signal processing, can be combined with sensor data [8]. While these advances contribute to improved security, there is still a need to integrate touch pressure and location data into a unified framework for more robust authentication.

Previous works have not focused on integrating 3D touch pressure data with on-screen finger location data to create a unified framework using decision-making techniques, such as TOPSIS. In this work, we focus on addressing these limitations by presenting a novel smartphone authentication framework. The TOPSIS-based decision-making approach has been augmented with ML techniques to identify and rank critical and valuable features in user behavior. By leveraging only valuable features, selected using the TOPSIS approach, and enhancing model resilience using ML algorithms, the proposed technique provides a secure, scalable, and user-friendly authentication system that can thwart many attacks, including cybersecurity threats. In order to refine the resiliency and robustness of the proposed technique, temporal scaling, noise insertion, and spatial perturbation approaches have been incorporated to mimic real-world user responses. This work proposes a secure and lightweight technique against strong threats and addresses current authentication dilemmas by integrating TOPSIS methodology with ML algorithms and dynamic pattern recognition.

The main contributions of this paper are:

1. A smartphone authentication application was developed on the Android platform to collect data from 30 participants, each performing 10 attempts.
2. A machine learning-based approach was developed for system authentication, where various machine learning algorithms were evaluated to identify the most effective model.
3. The TOPSIS method was employed to select key behavioral features, improving the authentication system's performance by focusing on the most impactful data.
4. The system's resilience was enhanced by applying data perturbation techniques, including noise injection and temporal scaling, to simulate real-world variations.
5. The system was tested against four types of cybersecurity attacks—spoofing, lighting variations, orientation changes, and noise injection—to assess its robustness and security.

The rest of the paper is organized as follows: Section 2 gives a brief overview of previous works, including behavioral biometrics for authentication systems, ML techniques in authentication systems as well as decision-making methods, and current challenges in security and usability. Section 3 presents in detail our proposed technique. The experimental results, including the decision and analysis are given in Section 4. Our conclusion and future works are explained in Section 5.

2. Literature Review

 doi:10.20944/preprints202503.0084.v1

Due to the current advancements and modern sophistications in smart devices, many researchers have implemented biometric authentication to significantly improve the system security and usability [9]. Unfortunately, timeouts, and failures are two main drawbacks in conventional authentication techniques, such as pattern locks, PIN codes, basic biometric systems, and passwords. Although traditional authentications are simple, lightweight, and easy to implement, it has been demonstrated that they are subjected to brute force, password theft, and phishing assaults [3,7,10,11]. Once static credentials are compromised, an attacker can get unlimited access to break the system security [12]. To address such vulnerabilities, recent research has incorporated ML algorithms and sensor data along with the traditional authentication to offer more secure and resilient systems against many serious attacks [4,13].

2.1. Behavioral Biometrics for Authentication Systems

Biometric authentication systems, including fingerprint and facial recognition, have been introduced as alternatives to address the limitations of traditional methods. These systems offer improved security and user convenience compared to passwords [14,15]. However, facial recognition systems can be spoofed using high-quality images or masks, and fingerprint sensors are susceptible to environmental factors, such as rain, dust, and physical degradation [16,17] and vulnerable to advanced spoofing techniques [18,19]. The aforementioned challenges emphasize the need for resilient, dynamic, and adaptive authentication methods capable of addressing both security and usability issues. In [20], the authors combined fingerprint, facial recognition, and iris scanning as alternatives to traditional methods. Although the proposal utilized physiological characteristics for user verification and provides enhanced security, it is subjected to environmental factors, such as humidity, lighting conditions, and device orientation [21]. Furthermore, biometric data is susceptible to spoofing, where adversaries can mimic user credentials [19,22]. This elevates the need for behavior-based authentication approaches.

Behavioral biometrics rely on user interaction patterns, such as swipe dynamics, typing speed, and touch pressure, to enable continuous authentication. Unlike static physiological biometrics, behavioral methods are dynamic and adaptive to changes in user behavior over time, and this in turn provides resilience against spoofing attempts and adversarial attacks [6]. Many studies have shown the impact of the behavioral biometrics on the performance of the smartphone authentication. Smith et al. have employed swipe trajectories and pressure levels in order to differentiate impostors from genuine persons [23]. In [24], Wang et al have presented the touch dynamics during password input, and the results indicated that the proposed technique can highly elevate the overall performance of the authentication systems. Compared to traditional biometrics, behavioral biometrics are non-intrusive, have flexible authentication, and difficult to be duplicated. Such techniques render researchers to focus further on individuals with more features, e.g., keystroke patterns, gait, and touch dynamics [20,25]. It has been proven that touch patterns can effectively distinguish between authentic users and impostors, and this improves the system security [26–28]. Unfortunately, behavioral biometrics are not robust in real-world conditions, such as international spoofing attempts and different environmental circumstances. Such drawbacks can be mitigated by augmenting behavioral biometrics with decision-making and ML algorithms.

2.2. ML Techniques in Authentication Systems and Decision-Making Methods

ML algorithms, such as Support Vector Machines (SVM), Decision Trees, and Neural Networks have provided promising solutions in improving the system performance and identifying user-specific patterns [29]. Other ML algorithms, such as Random Forest (RF) and Gradient Boosting Machine (GBM) based multi classifiers, further improve both the generalization and the performance by decreasing overfitting [30]. ML techniques have shown exceptional performance in processing complicated behavioral data, and this allows to effectively classify user interactions. For examples, both RF and SVM algorithms have been leveraged to analyze touch dynamics data, and achieved high accuracy for user authentication [31–33]. Pryor et al. merged RF and SVM classifiers to process

 doi:10.20944/preprints202503.0084.v1

touch behavior datasets and obtained accuracy exceeding 85% [31]. However, in real-time performance for a flawless user experience, computational complexity of such technique with high protection level still pose a challenge [33].

To further reduce the computational penalty and complexity in real time performance, ML models can be incorporated with decision-making techniques, such as TOPSIS, to select only valuable feature [34,35]. The hybrid ML and decision-making frameworks can help address existing issues in authentication systems. The TOPSIS and Analytic Hierarchy Process (AHP) techniques are an example of Multi-Criteria Decision-Making (MCDM) approaches that can be employed to prioritize and rank input features in many different applications. Specifically, TOPSIS identifies the highest impact parameters on a system to provide best possible solutions [34,35], and this can help to successfully rank features in many ML models [34]. However, integrating TOPSIS and ML with behavioral biometrics for smartphone authentication is still unrevealed in the Cybersecurity field.

2.3. Security and Usability Challenges in System Authentication

It is not easy to balance between security and usability in authentication systems. Increasing the security level of a design often leads to increase the system complexity that could negatively affect the user experience [36,37]. For instance, while strong authentication systems employing multi-factor approaches provide strong resistance against both spoofing and adversarial attacks, they induce user frustration due to the increase in time complexity or repeated queries for user verification [19]. Similarly, when environmental factors, e.g., lighting or humidity, disrupt sensor detection sensitivity, highly sensitive biometric systems may not function properly and this leads to produce wrong output [38–40].

It is worth noting that repeatedly tracking biometric features of users and precisely extracting valuable features, such as touch pressure, swipe dynamics, and trajectory patterns, significantly increase the system robustness [41,42]. However, the computations of real-time processing will be highly increased, and this renders the system unsuitable for seamless use. Interestingly, decision-making techniques, e.g., TOPSIS, can be utilized to select only valuable features, and this in turn ensures that the system remains user-friendly, and secure and decreases the computational system penalty [43,44].

The goal of this work is to develop a lightweight and robust authentication framework that is dynamically adapted to the user environments and conditions in real-world applications [37]. Even though significant progress has been achieved in the fields of biometric authentication, current existing techniques are still subjected to cybersecurity attacks and are not robust to real-world conditions. To address the aforementioned shortcomings, in this paper, different ML algorithms are integrated with TOPSIS and behavioral biometrics to get secure, resilient, and lightweight prominent solution for smartphone authentication.

3. Methodology

In this section, we present the proposed touch-based authentication framework, as illustrated in Figure 1. The process begins by collecting data from thirty participants, each used the 3D touch screen on a Samsung Galaxy A72 smartphone with ten attempts. This data is gathered and recorded through a custom-developed application, named the Authentication Application (AA), which was implemented on the Android platform. The AA captures sensitive data, including touch pressure, X and Y coordinates, velocity, acceleration, and trajectory curvature, as entered by the participants. This data is stored in a file named "dataset collection."

Next, the TOPSIS decision-making approach is applied to identify the most impactful features, while eliminating those that have minimal effect. The dataset is then preprocessed and cleaned. Following this, three machine learning algorithms—Random Forest, Gradient Boosting, and K-Nearest Neighbors (KNN)—are implemented and trained. During the training phase, the TOPSIS approach iteratively evaluates the results from each ML model, updates the feature set, and refines

 doi:10.20944/preprints202503.0084.v1

the data processing based on the newly identified valuable features. Once the model is trained, it is ready for use in mobile authentication. The complete framework comprises five main processes:

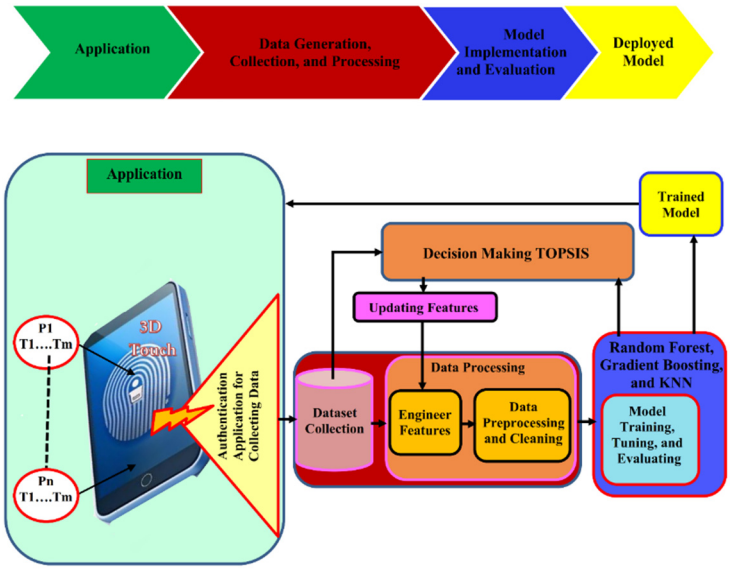

Figure 1. The life cycle of our complete framework.

3.1. Data Generation and Preprocessing

The 3D touch sensors of the Samsung Galaxy A72 device, shown in Figure 2, were experimentally utilized to collect data from touch-based behavioral biometrics. The dataset was generated with the participation of 30 individuals (15 males and 15 females) whose age range is between 18 and 50 years. Ethical considerations were a fundamental aspect of this study to ensure the responsible collection and use of data. Prior to participation, all individuals provided informed consent, ensuring their awareness and voluntary involvement in the study. To protect participant privacy, all collected data was anonymized, preventing the identification of individual contributors. Every participated person gave ten different pattern recognition samples that comprise typing phrases, unlocking the device, and performing swipe gestures. Our authentication application, implemented leveraging the Android system of Samsung Galaxy A72, received the incoming data with high resolution that are entered by the participated persons using the 3D touch screen. It is worth mentioning that the collected dataset has been encrypted to prohibit an attacker or unauthorized user get access to it. In order to verify that this work is within the conduct research role, it follows the international organization standards in research integrity as it only deals with human participants. The main five extracted features of the dynamic pattern recognition that are directly obtained from the 3D touch screen contain the following:

1. **Touch Pressure (P)**: The intensity of pressure applied during interactions.
2. **Trajectory Curvature (T):** The geometric path traced by the finger's movement.
3. **Velocity (V):** The rate of change in position during swipe gestures.
4. **Spatial Coordinates (X, Y):** The precise position of the finger on the 3D touch screen.
5. **Acceleration (A):** The change in velocity over time.

The resulting dataset spans a broad spectrum of user behaviors and scenarios, designed to enhance the robustness and adaptability of the proposed system. Noise removal, value normalization, and handling of missing data have been performed during the data preprocessing in order to ensure consistent and trustworthiness of the collected dataset. We also validated the dataset to make sure data reliability and integrity. To keep the collected data reliable and uniform, the mean for numerical features and the mode for nominal features have been leveraged. Note that, the interquartile range (IQR) approach has been used to recognize trajectory data, touch pressure, and

 doi:10.20944/preprints202503.0084.v1

velocity and then either exchanged with median values or eliminated for consistency purposes. Also, repeated records or entries have been recognized and then eliminated to avoid bias in the training process, and the entire features have been carefully normalized to a common scale in order to make sure that they are matchable with each used model.

a. The experiments b. Collected data operations

Figure 2. Dataset generation and collection using dynamic pattern recognition.

3.2. Feature Ranking and Engineering Leveraging TOPSIS Methodology

In order to rank the current entered and extracted features in terms of their effectiveness on the system performance, the TOPSIS methodology have been utilized. Similar procedure in [35,43,45] has been used to implement and run the TOPSIS methodology. First of all, the decision matrix ***D*** has been built, equation 1, in which each given column corresponds to the assessment metrics and each given row reflects the feature. The assessment metrics consist of the main measurement components, e.g., correlation, variance, and entropy. Note that, the aim of leveraging the TOPSIS methodology is to recognize only the most affected features in order to refine the accuracy of the system.

$$D = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & \dots & x_{mn} \end{bmatrix} \quad (1)$$

Next, to ensure the features' comparability, the ***D*** has been normalized leveraging the following formula, equation 2:

$$r_{ij} = \frac{x_{ij}}{\sqrt{\sum_{i=1}^{m} x_{ij}^2}} \quad \forall i, j \quad (2)$$

After ***D*** is normalized, weights w_j has been applied to normalized values, in order to drive the weighted normalized matrix, equation 3:

$$v_{ij} = w_j \cdot r_{ij} \quad (3)$$

The worst and possible values for each metric are represented by negative-ideal (***A***−) and ideal (***A***+) formulas, respectively, as follows [45], equation 4 and 5:

$$A^+ = \{\max(v_{ij}) \,|j \in J, \min(v_{ij}) | j \in J'\} \quad (4)$$

$$A^- = \{\min(v_{ij}) \,|j \in J, \max(v_{ij}) | j \in J'\} \quad (5)$$

In which ***J′*** signifies to non-beneficial criteria while ***J*** refers to beneficial criteria.

 doi:10.20944/preprints202503.0084.v1

In order to find the actual distance of each extracted feature from the negative-ideal and ideal solutions, the separation measures (***S***− and ***S***+) have been performed, as follows, equation 6:

$$S_i^+ = \sqrt{\sum_{j=1}^{n}(v_{ij} - A_j^+)^2}, S_i^- = \sqrt{\sum_{j=1}^{n}(v_{ij} - A_j^-)^2} \quad (6)$$

All given features have been ranked based on their values in C_i, in which higher values reflect high impact [46]. The relative closeness (C_i) of each feature to the ideal solution has been calculated, as follows in equation 7:

$$C_i = \frac{S_i^-}{S_i^+ + S_i^-} \quad (7)$$

It is worth to mention that the velocity, trajectory curvature, and touch pressure have been proven to represent the most valuable and affective features based on our experimental results obtained from the TOPSIS technique. Such valuable features have been demonstrated to refine the ML performance by increasing the model accuracy and mitigating the complexity of the authentication design compared to implementing the technique without incorporating the TOPSIS methodology.

3.3. ML Models

In order to refine the performance of the user authentication technique, three different ML algorithms—Gradient Boosting Machines (GBM), Random Forest (RF), and K-Nearest Neighbors (KNN)— have been employed. Each on these three models have been trained with and without incorporating the TOPSIS methodology.

1. **Gradient Boosting Machines (GBM):** GBM iteratively builds decision trees to minimize prediction errors, aiming to improve the model's predictive accuracy. The prediction at each iteration is represented as in equation 8:

$$F_m(x) = F_{m-1}(x) + \eta \cdot h_m(x) \quad (8)$$

where $F_m(x)$ is the updated model, $F_{m-1}(x)$ is the previous model, η is the learning rate, and $h_m(x)$ is the weak learner [47].

2. **Random Forest (RF):** RF constructs multiple decision trees during training and combines their outputs—through majority voting for classification or averaging for regression—with the objective of enhancing predictive performance. The prediction function is given as in equation 9 below:

$$f(x) = \frac{1}{N}\sum_{i=1}^{N} T_i(x) \quad (9)$$

where N is the number of trees, and $T_i(x)$ represents individual tree predictions [48].

3. **K-Nearest Neighbors (KNN):** KNN aims to classify data points by determining the majority label of their k-nearest neighbors, using a distance metric such as the Euclidean distance [49], equation 10:

$$d(x, y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \quad (10)$$

3.4. Data Perturbation Techniques

 doi:10.20944/preprints202503.0084.v1

Data perturbation approaches, noise injection, temporal scaling, and **spatial perturbations,** have been used to further refine the robustness of the proposed technique and the generalization capabilities of the ML models. The main purpose of incorporating such approaches is to imitate the real-world noise and variability in personal actions and responses through increasing the ability of the proposal to effectively deal with different conditions and circumstances.

- **Spatial Perturbations:** In order to mimic natural hand movements, small random Perturbation has been applied to the spatial coordinates (X, Y) based on the following equation 11:

$$(X', Y') = (X + \delta_x, Y + \delta_y) \tag{11}$$

where δ_y and δ_x represent the random perturbations.

- **Noise Injection:** To emulate variability in real-world interactions, random noise is augmented to the extracted features, e.g., velocity and touch pressure, equation 12:

$$X' = X + \epsilon \tag{12}$$

where $\boldsymbol{\epsilon}$ is random noise obtained from a Gaussian distribution, and $\boldsymbol{X}$ is the value of the original extracted feature.

- **Temporal Scaling:** To emulate various interaction styles, interaction durations have been further scaled to reflect variations in user speed, equation 13:

$$\boldsymbol{T'} = \boldsymbol{T} \cdot \boldsymbol{\alpha} \tag{13}$$

where $\boldsymbol{\alpha}$ is the scaling factor and $\boldsymbol{T}$ is the actual interaction duration [50] [51].

3.5. Assessing and Testing the Robustness of the Proposal

The collected dataset has been partitioned into 80% and 20% for training and validation, respectively, in order to assess the performance of the ML models. Several testing approaches and evaluation metrics have been used to evaluate the robustness and effectiveness of the proposed technique, aiming to get high reliable and accurate authentication system under different circumstances and conditions:

- **Recall (Sensitivity):** This metric quantifies the proportion of true positives among all actual positives, ensuring genuine users are accurately recognized [8], equation 14:

$$Recall = \frac{TP}{TP + FN} \tag{14}$$

- **Precision:** A critical metric for evaluating the proportion of true positives among all predicted positives, reducing false alarms, equation 15:

$$Precision = \frac{TP}{TP + FP} \tag{15}$$

High precision minimizes false-positive rates, which is crucial to preventing unauthorized access [8].

- **Accuracy:** The primary metric to measure the ratio of correctly classified instances to the total number of instances, equation 16:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{16}$$

where ***TP***, ***TN***, ***FP***, and ***FN*** denote true positives, true negatives, false positives, and false negatives, respectively [8].

- **F1-Score:** A balanced measure combining precision and recall, particularly useful in scenarios with class imbalances or where addressing the trade-off between precision and recall is critical [8], equation 17:

 doi:10.20944/preprints202503.0084.v1

$$F1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \tag{17}$$

- **Confusion Matrix (CM):** in order to evaluate the system's performance, the CM has been used, in which is can offer information about true negatives, false negatives, true positives, and false positives. This CM can provide valuable details about the model operation [53].
- **Resilience Testing [52]:** robustness of the proposed technique has been assessed under three different conditions. First, environmental variations are tested, including different lighting conditions (bright, dim, and dark) and humidity levels, to ensure stable feature extraction and classification accuracy. Second, variations in user behavior, such as changes in touch speed, pressure intensity, and swipe dynamics, are tested to assess the system's ability to adapt, with performance measured by accuracy and F1-score. Finally, the system's performance is tested under different device orientations—portrait, landscape, and tilted—to ensure consistency across various handling scenarios.
- **Spoofing Detection Rate (SDR):** The effectiveness of the proposed technique can be assessed via using the SDR measurement in order to reveal and prevent spoofing attempts-based attacks. The mathematical equation of SDR is as in the following [22], equation 18:

$$SDR = \frac{\text{Total Number of Spoofing Attempts}}{\text{Number of Detected Spoofing Attempts}} \tag{18}$$

4. Experimental Results

The performance assessment of our proposed technique based on smartphone authentication has been given in details in this section, in which the TOPSIS methodology-based feature ranking has been combined with different ML algorithms. The evaluation of our proposal is mainly concentrated on the performance criteria, a comprehensive analysis of the confusion matrix to evaluate the classification accuracy, and robustness testing under real-world conditions and circumstances. The proposed framework has been implemented on three main layers to correctly perform and authenticate each participated user. First of all, the touch-based dynamic pattern recognition data, including the gyroscope, 3D touch sensor, and accelerometer, has been used for data collection and generation layer. This layer has been leveraged to extract real-time features, such as coordinate curves for trajectories, 2D spatial coordinate, touch pressure, acceleration, and velocity. Data preparation and processing is the second layer that is used to carry out data preprocessing, including standardization, normalization, and noise mitigation or reduction. During data processing, the TOPSIS method has been implemented and applied to carefully rank the extracted valuable features. The ranked valuable features have then fed as input into three different ML models — RF, KNN, and GBM — for user classification purposes. Finally, the application layer has been used to classify the findings to the authenticated persons or users in order to give immediate feedback and input for denying or granting access.

The proposed framework processes the incoming dynamic touch information in real-time, in which optimized algorithms have been leveraged for low-latency performance in order to ensure better smooth operation. The primary sequence of the given events is as follows: the personal touch data is first entered and preprocessed; then, the ranked affected features are fed into the trained ML models; and next, the classification results of the participated persons are pulled up, and the system either refuses or allows the user to enter.

4.1. Feature Importance Analysis

The most affected and significant features for personal authentication are listed in Table 1, in which the ranking of the dynamic pattern recognition (behavioral biometrics) based on the TOPSIS methodology is presented. As illustrated in the Table, the three top ranked values of the affected and

 doi:10.20944/preprints202503.0084.v1

valuable features are: the touch force (pressure), abscissa (X coordinate), and duration (velocity). Note that, these features have been chosen based on their highest essentiality. Then, they are fed as primary input into each of the implemented ML models, instead of considering all of the entered features by the users. This helps to further refine the performance of the system and optimize the entire framework.

Table 1. TOPSIS-Based Feature Ranking.

Feature	TOPSIS Score	Rank
Touch Pressure	0.5053	1
X Coordinate	0.4602	2
Velocity	0.4562	3
Y Coordinate	0.4353	4

4.2. Performance Evaluation

The performance assessment of the three implemented ML models—KNN, RF, and GBM—with and without incorporating TOPSIS methodology-based feature ranking is elucidated in Table 2. Based on the experimental results, it has been pointed out that the KNN model performed good on small datasets; however, encountered a problem with scalability when the data size is significantly elevated. GBM model accomplished high accuracy due to its iterative boosting process, yet it needed longer training time. RF model offered a good solution balance between computational efficiency and accuracy, and this renders the system more appropriate for real-time applications and processing systems. Note that, the RF model produced the highest accuracy compared to the two other models, KNN and GBM. Also, it has been pointed out that classification accuracy has been refined when TOPSIS-methodology based feature chosen ranking is augmented, and this further emphasizes that feature selection approach is valuable. A good balance between computational efficiency and performance could be obtained based on this comparative assessment leveraging the most suitable model for authentication system purposes.

Table 2. Performance Metrics of the Proposed Models.

Metric	RF		GBM		KNN	
	Without TOPSIS	With TOPSIS	Without TOPSIS	With TOPSIS	Without TOPSIS	With TOPSIS
Accuracy	90.4%	94.8%	90.13%	94.7%	83.42%	93.8%
Precision	88.9%	94.8%	91.09%	94.3%	81.24%	93.5%
Recall	90.89%	95.5%	89.78%	94.9%	83.31%	94.1%
F1-Score	90.15%	95.1%	89.35%	94.6%	81.37%	93.8%

4.3. Confusion Matrix Results

In order to verify both areas and strengths of our proposal, the confusion matrix (CM) has been leveraged to further show whether our design is able to successfully classify personal interactions based on the selected features. Figure 3 illustrates CM outcomes for the classification performance of our proposed technique based on RF model for smartphone authentication system.

Firstly, the matrix illustrates diagonal clarity, where the majority of predictions are correctly classified. For example, most of the users have been successfully classified based on the prediction accuracy, and this shows the great performance of our proposal. The given simulation results indicate that our proposed technique is able to correctly recognize between user touch dynamic patterns based on valuable selected features, including spatial coordinates (X, Y), interaction duration or velocity, and touch pressure intensity.

However, with some minor emerged errors, the matrix also reveals misclassification cases, represented by off-diagonal values, and this could be due to overlapping behavioral patterns among

 doi:10.20944/preprints202503.0084.v1

participated users, leading to the possibility of overlapping among classes. For example, one sample from each of the users 6, 7, and 8 has been misclassified as users 8, 19, and 15, respectively, due to some overlapping in dynamic pattern recognition, e.g., similar swipe trajectories and / or similar touch pressure levels among some participated users. Note that, some minor deviations in performance may lead to missing data in some experiments when taking into consideration the differences in some samples among the participated users.

Even though some numbers have not been classified correctly, the correct classifications shown on the main diagonal of the CM elucidates the significantly effectiveness of the TOPSIS methodology-based feature extraction. Also, the high correct predictions in the matrix diagonal implies that the selected features, e.g., spatial X coordinates, trajectory curvature, and touch pressure intensity, are the most promising ones. By incorporating only valuable features and optimizing the hyperparameters of the model, the proposed technique can offer an excellent balance among different classes, and this in turn could lead to mitigate the wrong classification and obtain better performance.

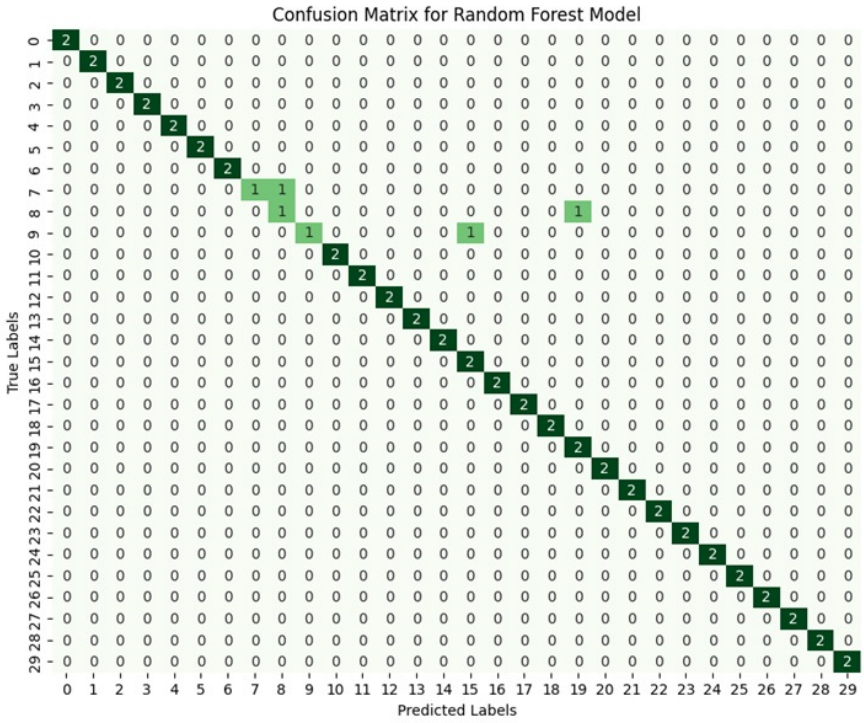

Figure 3. The Confusion Matrix (CM) outcomes of our proposal.

4.4. Evaluating the Strength of Our Proposal

In order to exam the resiliency of the proposal-based RF model under various real-world circumstances, four different cybersecurity attacks have been incorporated, as elucidated in Table 3. The reason behind selecting the RF model for the testing is that the FR model provides better accuracy compared to the two other models as shown previously in Table 2. The proposed technique showed excellent resiliency against spoofing attempts-based attacks in which the spoofing detection rate reached 96%, and this emphasized that the proposed techniques can successfully prohibit adversarial access-based attacks. Moreover, ever when random noise has been inserted into the primary input dataset, the framework illustrated to be strong against such noise insertion data-based attacks via accomplishing 90.8% accuracy. The proposal can also prevent environmental variations or conditions, including device orientation adjustments and lighting changes, via achieving 93.2% and 92.3% accuracies, respectively. Given the aforementioned experimental results, the proposal is shown to be resilient against different conditions and circumstances, and this makes it appropriate for real-world processing systems and applications.

 doi:10.20944/preprints202503.0084.v1

Table 3. Testing the resiliency of the proposed technique-based RF model against four serious cybersecurity threats.

Test Scenario	Metric	Result
Spoofing Detection	Detection Rate	96%
Lighting Variations	Accuracy	92.3%
Device Orientation Changes	Accuracy	93.2%
Noise Injection in Data	Accuracy	90.8%

4.5. Discussion Summary

It has been proven that the TOPSIS methodology can effectively be leveraged to select the most affected features, and this in turn increase the security level of the authentication system. The three extracted features: velocity, X coordinate, and touch pressure, have been recognized to be the most affected and valuable features, and can significantly elevate the classification accuracy. as previously explained via the experimental findings, the hybrid ML models and these affected features have refined the performance of the system significantly. It has been pointed out that when the RF combined with the TOPSIS methodology-based feature chosen, it can accomplish the highest accuracy than the two other algorithms. Under different environmental circumstances and conditions, e.g., device orientation and modifications in lighting density, our proposed technique has been carefully evaluated in terms of resiliency and robustness with minimal performance penalty and with high detection rates in spoofing attempts-based attacks. Moreover, the proposal keeps achieving a high level of accuracy, underscoring the combined strength of the chosen features and robustness's model in refining both accuracy and reliability in real-world applications and real system processing even when noise was inserted into the original dataset.

Even though the proposed framework has shown many advancements, it also has some limitations. For examples, environmental conditions and highly dependency on variations in device orientation and lighting density can slightly impact on collected and generated dataset. Also, with small sample size of the provided dataset, it could not possible to capture all user behavioral biometrics, and this encourages us to gather and enlarge the data samples in our future works. Finally, significant additional resources might be requested in some of the designed and implemented ML algorithms, e.g., GBM, and this in turn potentially impact on both the computational penalty and affected the real-time performance.

5. Conclusion and Future Work

A smartphone framework-based authentication technique has been designed and implemented leveraging TOPSIS technique-based extracted feature ranking, dynamic pattern recognition (behavioral biometrics), and different ML models. It has been proven that when RF is incorporated, the proposed technique performs better than the other two ML algorithms, with a reported accuracy of 95.2%. The system is also resilient against serious cybersecurity attacks, achieving over 93% accuracy against environmental changes and 96% detection rates against spoofing attacks. By considering only critical features, such as touch pressure, velocity, and trajectory curvature, the framework balances security and usability, enabling real-time adaptability for end users. The proposed technique provides a lightweight, robust, secure, and scalable authentication solution when the TOPSIS decision-making technique and ML algorithms are integrated into the design. Despite promising results, this work has limitations, including the use of a small dataset and the lack of testing against extreme environmental factors. In future work, we plan to expand our datasets further by considering multi-modal biometrics, such as facial recognition and fingerprints, and leveraging deep learning algorithms, such as CNN, DNN, and RNNs, to enhance performance in real-world applications and improve energy efficiency.

 doi:10.20944/preprints202503.0084.v1

Author Contributions: Conceptualization, Moceheb Lazam Shuwandy; methodology, Moceheb Lazam Shuwandy, Qutaiba Alasad, and Maytham M. Hammood; software, Maytham M. Hammood and Noor S. Abd; validation, Moceheb Lazam Shuwandy, Ayad A. Yass, and Rawan A. Alsharida; formal analysis, Saadi Hamad Thalij and Salwa Khalid Abdulateef; investigation, Moceheb Lazam Shuwandy and Qutaiba Alasad; resources, Maath Frman and Abdulsalam Hamid Kutaibani; data curation, Sahar Lazim Qaddoori and Noor S. Abd; writing—original draft preparation, Moceheb Lazam Shuwandy; writing—review and editing, Qutaiba Alasad, Maytham M. Hammood, and Noor S. Abd; visualization, Rawan A. Alsharida and Saadi Hamad Thalij; supervision, Moceheb Lazam Shuwandy; project administration, Moceheb Lazam Shuwandy.

Funding: This research received no external funding.

Data Availability Statement: The data is unavailable due to privacy or ethical restrictions.

Acknowledgments: We thank the participants who contributed their touch dynamics data to this study.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. K. Shaheed, P. Szczuko, M. Kumar, I. Qureshi, Q. Abbas, and I. Ullah, “Deep learning techniques for biometric security: A systematic review of presentation attack detection systems,” *Eng Appl Artif Intell*, vol. 129, p. 107569, 2024.
2. M. L. Shuwandy, B. B. Zaidan, and A. A. Zaidan, *Novel authentication of blowing voiceless password for android smartphones using a microphone sensor Content courtesy of Springer Nature , terms of use apply . Rights reserved . Content courtesy of Springer Nature , terms of use apply . Rights reserved .* Multimedia Tools and Applications, 2022.
3. R. Ryu, S. Yeom, D. Herbert, and J. Dermoudy, “The design and evaluation of adaptive biometric authentication systems: Current status, challenges and future direction,” *ICT Express*, vol. 9, no. 6, pp. 1183–1197, 2023.
4. M. L. Shuwandy, B. B. Zaidan, A. A. Zaidan, and A. S. Albahri, “Sensor-Based mHealth Authentication for Real-Time Remote Healthcare Monitoring System: A Multilayer Systematic Review,” *J Med Syst*, vol. 43, no. 2, 2019, doi: 10.1007/s10916-018-1149-5.
5. B. Zurita, S. Bosque, W. Fuertes, and M. Macas, “Social Engineering Shoulder Surfing Attacks (SSAs): A Literature Review. Lessons, Challenges, and Future Directions,” in *Advanced Research in Technologies, Information, Innovation and Sustainability*, T. Guarda, F. Portela, and J. M. Diaz-Nafria, Eds., Cham: Springer Nature Switzerland, 2024, pp. 220–233.
6. M. L. Shuwandy, H. A. Aljubory, N. M. Hammash, M. M. Salih, M. A. Altaha, and Z. T. Alqaisy, “BAWS3TS: Browsing Authentication Web-Based Smartphone Using 3D Touchscreen Sensor,” *2022 IEEE 18th International Colloquium on Signal Processing and Applications, CSPA 2022 - Proceeding*, no. May, pp. 425–430, 2022, doi: 10.1109/CSPA55076.2022.9781888.
7. M. L. Shuwandy et al., “mHealth authentication approach based 3D touchscreen and microphone sensors for real-time remote healthcare monitoring system: Comprehensive review, open issues and methodological aspects,” *Comput Sci Rev*, vol. 38, p. 100300, 2020, doi: 10.1016/j.cosrev.2020.100300.
8. A. Y. Younis and M. L. Shuwandy, “Biometric Authentication Utilizing EEG Based-on a Smartphone’s 3D Touchscreen Sensor,” in *2023 IEEE 14th Control and System Graduate Research Colloquium (ICSGRC)*, IEEE, 2023, pp. 169–174.
9. R. Bhagavatula, B. Ur, K. Iacovino, S. M. Kywe, L. F. Cranor, and M. Savvides, “Biometric authentication on iphone and android: Usability, perceptions, and influences on adoption,” 2015.
10. A. Constantinides et al., “Security and usability of a personalized user authentication paradigm: Insights from a longitudinal study with three healthcare organizations,” *ACM Trans Comput Healthc*, vol. 4, no. 1, pp. 1–40, 2023.

 doi:10.20944/preprints202503.0084.v1

11. Z. M. Saadi, A. T. Sadiq, O. Z. Akif, and A. K. Farhan, "A Survey: Security Vulnerabilities and Protective Strategies for Graphical Passwords," *Electronics (Basel)*, vol. 13, no. 15, p. 3042, 2024.
12. Ö. Aslan, S. S. Aktuğ, M. Ozkan-Okay, A. A. Yilmaz, and E. Akin, "A comprehensive review of cyber security vulnerabilities, threats, attacks, and solutions," *Electronics (Basel)*, vol. 12, no. 6, p. 1333, 2023.
13. M. L. Shuwandy et al., "Sensor-Based Authentication in Smartphone; a Systematic Review," *Journal of Engineering Research*, 2024.
14. K. Sathya, J. Esther, S. Kavitha, and J. Kamalakumari, "Facetpass-Intelligent Facial Recognition Authentication System Security and Usability," in *2024 2nd International Conference on Artificial Intelligence and Machine Learning Applications Theme: Healthcare and Internet of Things (AIMLA)*, IEEE, 2024, pp. 1–6.
15. D. Harikrishnan, N. Sunil Kumar, S. Joseph, and K. K. Nair, "Towards a fast and secure fingerprint authentication system based on a novel encoding scheme," *International Journal of Electrical Engineering & Education*, vol. 61, no. 1, pp. 100–112, 2024.
16. A. Wone, J. Di Manno, C. Charrier, and C. Rosenberger, "Impact of environmental conditions on fingerprint systems performance," in *2021 18th International Conference on Privacy, Security and Trust (PST)*, IEEE, 2021, pp. 1–5.
17. S. Liu, B. Yang, P. C. Yuen, and G. Zhao, "A 3D mask face anti-spoofing database with real world variations," in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 2016, pp. 100–106.
18. D. S. Ametefe et al., "Enhancing Fingerprint Authentication: A Systematic Review of Liveness Detection Methods Against Presentation Attacks," *Journal of The Institution of Engineers (India): Series B*, vol. 105, no. 5, pp. 1451–1467, 2024, doi: 10.1007/s40031-024-01066-3.
19. C.-A. Toli and B. Preneel, "Provoking security: Spoofing attacks against crypto-biometric systems," in *2015 World Congress on Internet Security (WorldCIS)*, 2015, pp. 67–72. doi: 10.1109/WorldCIS.2015.7359416.
20. M. I. Sharif, M. Mehmood, M. I. Sharif, and M. P. Uddin, "Human gait recognition using deep learning: A comprehensive review," *arXiv preprint arXiv:2309.10144*, 2023.
21. R. Alrawili, A. A. S. AlQahtani, and M. K. Khan, "Comprehensive survey: Biometric user authentication application, evaluation, and discussion," *Computers and Electrical Engineering*, vol. 119, p. 109485, 2024.
22. D. Menotti et al., "Deep representations for iris, face, and fingerprint spoofing detection," *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 4, pp. 864–879, 2015.
23. A. J. Aviv, K. Gibson, E. Mossop, M. Blaze, and J. M. Smith, "Smudge Attacks on Smartphone Touch Screens."
24. K. Wang, L. Zhou, and D. Zhang, "Biometrics-Based Mobile User Authentication for the Elderly: Accessibility, Performance, and Method Design," *Int J Hum Comput Interact*, vol. 40, no. 9, pp. 2153–2167, 2024.
25. R. V Yampolskiy and V. Govindaraju, "Behavioural biometrics: a survey and classification," *Int J Biom*, vol. 1, no. 1, pp. 81–113, 2008.
26. B. Pelto, M. Vanamala, and R. Dave, "Your identity is your behavior-continuous user authentication based on machine learning and touch dynamics," in *2023 3rd International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME)*, IEEE, 2023, pp. 1–6.
27. P. Aaby, M. V. Giuffrida, W. J. Buchanan, and Z. Tan, "An omnidirectional approach to touch-based continuous authentication," *Comput Secur*, vol. 128, p. 103146, 2023.
28. P. G. do Nascimento, P. Witiak, T. MacCallum, Z. Winterfeldt, and R. Dave, "Your device may know you better than you know yourself--continuous authentication on novel dataset using machine learning," *arXiv preprint arXiv:2403.03832*, 2024.
29. A. E. K. Khalil, J. A. Perez-Diaz, J. A. Cantoral-Ceballos, and J. M. Antelis, "Unlocking Security for Comprehensive Electroencephalogram-Based User Authentication Systems," *Sensors*, vol. 24, no. 24, p. 7919, 2024.

 doi:10.20944/preprints202503.0084.v1

30. D. Upadhyay, J. Manero, M. Zaman, and S. Sampalli, "Gradient boosting feature selection with machine learning classifiers for intrusion detection on power grids," *IEEE Transactions on Network and Service Management*, vol. 18, no. 1, pp. 1104–1116, 2020.
31. L. Pryor, J. Mallet, R. Dave, N. Seliya, M. Vanamala, and E. S. Boone, "Evaluation of a User Authentication Schema Using Behavioral Biometrics and Machine Learning," *arXiv preprint arXiv:2205.08371*, 2022.
32. C. Gupta, I. Johri, K. Srinivasan, Y.-C. Hu, S. M. Qaisar, and K.-Y. Huang, "A systematic review on machine learning and deep learning models for electronic information security in mobile networks," *Sensors*, vol. 22, no. 5, p. 2017, 2022.
33. R. A. Disha and S. Waheed, "Performance analysis of machine learning models for intrusion detection system using Gini Impurity-based Weighted Random Forest (GIWRF) feature selection technique," *Cybersecurity*, vol. 5, no. 1, p. 1, 2022, doi: 10.1186/s42400-021-00103-8.
34. Y. B. Abushark et al., "Usability evaluation through fuzzy AHP-TOPSIS approach: Security requirement perspective," *Comput. Mater. Contin*, vol. 68, pp. 1203–1218, 2021.
35. W. Alhakami, "Evaluating modern intrusion detection methods in the face of Gen V multi-vector attacks with fuzzy AHP-TOPSIS," *PLoS One*, vol. 19, no. 5, p. e0302559, 2024.
36. C. Braz and J.-M. Robert, "Security and usability: the case of the user authentication methods," in *Proceedings of the 18th Conference on l'Interaction Homme-Machine*, 2006, pp. 199–203.
37. T. O. Agboola, J. Adegede, and J. G. Jacob, "Balancing Usability and Security in Secure System Design: A Comprehensive Study on Principles, Implementation, and Impact on Usability," *International Journal of Computing Sciences Research*, vol. 8, pp. 2995–3009, 2024.
38. J. Zhang, A. R. Beresford, and I. Sheret, "SensorID: Sensor calibration fingerprinting for smartphones," in *Proceedings - IEEE Symposium on Security and Privacy*, 2019, pp. 638–655. doi: 10.1109/SP.2019.00072.
39. C. Shen, Y. Chen, and X. Guan, "Performance evaluation of implicit smartphones authentication via sensor-behavior analysis," *Inf Sci (N Y)*, vol. 430–431, pp. 538–553, 2018, doi: 10.1016/j.ins.2017.11.058.
40. O. E. Basar, G. Alptekin, H. C. Volaka, M. Isbilen, and O. D. Incel, "Resource usage analysis of a mobile banking application using sensor-and-touchscreen-based continuous authentication," *Procedia Comput Sci*, vol. 155, pp. 185–192, 2019, doi: 10.1016/j.procs.2019.08.028.
41. P. S. Teh, N. Zhang, A. B. J. Teoh, and K. Chen, "A survey on touch dynamics authentication in mobile devices," *Comput Secur*, vol. 59, pp. 210–235, 2016.
42. L. Zhang et al., "Toward Robust and Effective Behavior Based User Authentication With Off-the-shelf Wi-Fi," *IEEE Transactions on Information Forensics and Security*, 2024.
43. C. Z. Radulescu and M. Radulescu, "A Hybrid Group Multi-Criteria Approach Based on SAW, TOPSIS, VIKOR, and COPRAS Methods for Complex IoT Selection Problems," *Electronics (Basel)*, vol. 13, no. 4, p. 789, 2024.
44. A. K. M. B. Haque, B. Bhushan, and G. Dhiman, "Conceptualizing smart city applications: Requirements, architecture, security issues, and emerging trends," *Expert Syst*, vol. 39, no. 5, p. e12753, 2022.
45. R. Kumar, A. I. Khan, Y. B. Abushark, M. M. Alam, A. Agrawal, and R. A. Khan, "An integrated approach of fuzzy logic, AHP and TOPSIS for estimating usable-security of web applications," *IEEE Access*, vol. 8, pp. 50944–50957, 2020.
46. F. A. Al-Zahrani, "Evaluating the usable-security of healthcare software through unified technique of fuzzy logic, ANP and TOPSIS," *IEEE Access*, vol. 8, pp. 109905–109916, 2020.
47. H. Lu, S. P. Karimireddy, N. Ponomareva, and V. Mirrokni, "Accelerating gradient boosting machines," in *International conference on artificial intelligence and statistics*, PMLR, 2020, pp. 516–526.

 doi:10.20944/preprints202503.0084.v1

48. D. J. S. Raja, R. Sriranjani, P. Arulmozhi, and N. Hemavathi, "Unified Random Forest and Hybrid Bat Optimization based Man-in-the-Middle Attack Detection in Advanced Metering Infrastructure," *IEEE Trans Instrum Meas*, 2024.
49. R. Wang and D. Tao, "DTW-KNN Implementation for Touch-based Authentication System," *Proceedings - 5th International Conference on Big Data Computing and Communications, BIGCOM 2019*, pp. 318–322, 2019, doi: 10.1109/BIGCOM.2019.00055.
50. Y. Li, H. Hu, and G. Zhou, "Using Data Augmentation in Continuous Authentication on Smartphones," *IEEE Internet Things J*, vol. 6, no. 1, pp. 628–640, 2019, doi: 10.1109/JIOT.2018.2851185.
51. A. Mikołajczyk and M. Grochowski, "Data augmentation for improving deep learning in image classification problem," in *2018 international interdisciplinary PhD workshop (IIPhDW)*, IEEE, 2018, pp. 117–122.
52. H. Yang et al., "TapLock: Exploit finger tap events for enhancing attack resilience of smartphone passwords," *IEEE International Conference on Communications*, vol. 2015-Septe, no. 17, pp. 7139–7144, 2015, doi: 10.1109/ICC.2015.7249465.
53. M. M. Mijwil and M. Aljanabi, "A comparative analysis of machine learning algorithms for classification of diabetes utilizing confusion matrix analysis," *Baghdad Sci. J*, vol. 20, no. 10.21123, 2023.