Mesopotamian journal of Cybersecurity
Vol.5,No.2, **pp**. 577–589
DOI: https://doi.org/10.58496/MJCS/2025/035; ISSN: 2958-6542
https://mesopotamian.press/journals/index.php/cybersecurity

Research Article

Enhancing User Authentication through the Implementation of the ForestPA Algorithm for Smart Healthcare Systems

Nurul Syafiqah Zaidi [1,*], Al-Fahim Mubarak Ali [1,2], Ahmad Firdaus [1,2], Adamu Abubakar Ibrahim [3], Ghassan Saleh AlDharhani [4], Mohd Faizal Ab Razak[1]

[1] *Faculty of Computing, Universiti Malaysia Pahang Al-Sultan Abdullah (UMPSA), Pekan, Pahang, Malaysia.*

[2] *Centre for Artificial Intelligence & Data Science (CAIDAS), Universiti Malaysia Pahang Al-Sultan Abdullah (UMPSA), Lebuh Persiaran Tun Khalil Yaakob, 26300 Gambang, Kuantan, Pahang, Malaysia.*

[3] *Department of Computer Science, International Islamic University Malaysia (IIUM), Jalan Gombak, Kuala Lumpur, Selangor, Malaysia.*

[4] *Institute of Computer Science & Digital Innovation, UCSI University, Kuala Lumpur, Malaysia.*

ARTICLE INFO

Article history

Received 22 Aug 2024
Revised 06 Apr 2025
Accepted 06 May 2025
Published 26 Jun 2025

Keywords

ForestPA

Key Forest-based

Classifiers

Medical

Internet of Things

User profile authentication

ABSTRACT

The machine learning-based authentication model for smart healthcare systems represents a crucial step in addressing the needs of an ever-evolving healthcare industry. The need to protect sensitive patient data, ensure regulatory compliance, and reduce medical errors, especially in the context of telemedicine and remote monitoring, underscores the importance of such systems. Traditional authentication methods frequently lack sufficient security, resulting in potential breaches. Relying solely on usernames and passwords, without supplementary authentication measures, exposes systems to advanced security attacks. As it involves patients' health and human lives, it is important to provide additional authentication, fast machine learning-based authentication models and high accuracy at the same time. This study involves five participants with devices and performs various finger-based interactions (raising, lowering, moving the finger, applying pressure, adjusting orientation, and utilizing multiple hikes) while completing reading and image comparison tasks across multiple sessions. Each experiment lasted between 25 and 50 minutes for one participant, with reading tasks typically taking 10--15 minutes and image comparison tasks requiring 3--4 minutes, all measured in milliseconds. All these activities are recorded as a dataset for model training. A model was trained via the forest penalizing attributes (ForestPA) algorithm, which can classify profiles into real or fake profiles on the basis of their behavioral patterns. The results revealed a 99.99% accuracy rate in identifying fake profiles and avoiding them by accessing medical data even though they were able to bypass the username and password.

1. INTRODUCTION

Various factors drive the evolution of smart healthcare systems. Various elements, such as population aging, a rising prevalence of chronic disease, and increasing demand for personalized and accessible health care services, have contributed to the digital transformation of the health care industry [1], [2], [3]. There is a need to address security problems, as the healthcare sector is heading towards the digitization of services [3], [5]. Large amounts of sensitive patient data made securely electronic must be processed to be vulnerable to serious security threats. These risks can range from unauthorized data access to data breaches, which poses the risk of compromising patient privacy in healthcare establishment [6]. The demand for investment in the cybersecurity and healthcare sector, which amounts to more than 16 billion euros in 2023 [7], has more than doubled since 2019. The rapid digitization of healthcare, including the rise of electronic health records and telemedicine, has created many cybersecurity vulnerabilities. In addition to the subsequent explosion in highly publicized breaches, the protection of sensitive patient data requires additional investment in data protection. According to market analyses, global spending in this field will continue to increase and exceed 25 billion euros by 2026. This trend further

*Corresponding author. Email: fahim@umpsa.edu.my

solidifies the consideration of cybersecurity within the wide infrastructure of health care. With cybercriminals taking increasingly keen interest in healthcare data, effective cybersecurity needs to be deployed earlier. Given this, the need to integrate and implement advanced cybersecurity capabilities in the healthcare domain is increasing significantly [8], [9]. A report from 2023 revealed that a detailed examination of the first six months of 2022 revealed 337 remarkable breaches, affecting 19,992,810 consumers [10]. The statistics underscore the continuing and growing issues facing the health care sector in securing sensitive data from unauthorized access and the increasingly complex threat landscape. These trends highlight the increasing demand for strong cybersecurity frameworks and interventions to reduce the growing prevalence and magnitude of data breaches, ensuring the protection and privacy of individuals' health information. Moreover, statistics from 2023 indicate an alarming trend, whereas more than 25% of Americans face exposure to health data as a result of security breaches affecting an astonishing 87 million patients [11]. However, this shocking discovery reflects a widespread weakness within the healthcare information ecosystem. This startling discovery sheds light on a deep-seated weakness in the healthcare data ecosystem, emphasizing the urgency of implementing effective access control measures. In health care systems, user authentication is critical for ensuring that only authorized personnel have access to sensitive patient information and are able to perform certain actions. Conventional approaches, including passwords, are typically inadequate, as they can be compromised or attacked. More advanced techniques include biometric verification and multifactor authentication, which provide greater security for sensitive data. By incorporating these robust methods, organizations can minimize any potential unauthorized access while still ensuring the integrity and confidentiality of health information. These two layers of security protect patients and reduce risk in an increasingly connected, digital healthcare experience.

Currently, health devices (smart inhalers, mobile ECG devices, smart thermometers) that are worn by patients relay data to mobile devices, and these data are analysed by doctors and nurses or managers via smartphones or tablets. For various authentication smartphone users, which is a common security level, traditional factors, such as passwords and PIN, are unsecured and often tedious [12]. Usernames and passwords without secondary authentication are insufficient to secure a system. Hence, it is important to add one more layer of authentication to improve the overall security of the system [13], [14]. This prompts the development of machine learning-based authentication models that are based on the user profile to increase the security of smartphone devices that are used by doctors to monitor their patients.

However, as it involves patient health and human life, fast machine learning-based authentication models and high accuracy are needed at the same time [15], [16], [17], [18]. For this purpose, forest-based machine learning classifiers are promising options. It is a type of classifier that averages the results of individual trees, thus providing more accurate predictions and providing insights into feature importance directly as part of the model output, which is useful for understanding the driving factors behind the predictions [19], [20], [34]. Furthermore, the trees in forest-based classifiers run in parallel during training, which can significantly speed up the training process, especially on multicore processors, requiring less preprocessing of data (which do not require feature scaling, normalization or standardization), which can be a time-saving advantage in the preprocessing stage and can handle missing values in the input data to some extent, reducing the need for complex imputation processes [21], [22].

Numerous studies have been conducted on user authentication employing machine learning techniques. For example, [23] explored continuous authentication through touchscreen interactions, whereas [24] examined a biometric authentication system. Furthermore, [25] investigated the utilization of soft keyboard typing behavior for ongoing user recognition, and [26] leveraged the physical layer attributes of the wireless channel to identify spoofing attacks. Despite the diverse methodologies proposed and implemented within the realm of machine learning-based authentication, the rapidly evolving landscape of security threats and the proliferation of smart devices underscore the critical need for innovative approaches [27].

Therefore, methods and techniques in the healthcare area need to be continuously updated as technology progresses, as it involves human lives. This study employs a less explored and utilized updated forest-based algorithm of the ForestPA [27], which generates a set of highly accurate and diverse decision trees by exploiting the strength of all nonclass attributes available in a dataset (unlike some existing algorithms that use a subset of the nonclass attributes), it can avoid building similar trees by imposing penalties to the attributes that participated in the previous trees by randomly selecting weights from the weight ranges associated with the levels of the attributes, and it is also capable of avoiding the hyperparameter problem that affects some algorithms, such as random forest and extremely randomized trees.

Hence, the contributions of this research are as follows:

1) Investigate a novel machine learning forest penalizing attribute (ForestPA) to authenticate users within the context of smartphone devices for monitoring smart healthcare devices. It identifies profiles as either real or fake doctor profiles on the basis of their behavior and further avoids any attack that bypasses traditional user authentication.
2) Scrutinizes a dataset comprising 116,565 samples executed by five different users.
3) Finally, this research assesses and compares the results with those of previous related studies to evaluate the effectiveness of ForestPA for authenticating users.

Section 2 delineates the related work, while Section 3 presents the methodology. Section 4 illustrates the results of the experiment, and Section 5 concludes the study.

2. RELATED WORKS

In recent years, the integration of machine learning in healthcare has revolutionized the way in which studies approach medical diagnosis, treatment, and patient care. Machine learning, a subset of artificial intelligence, has shown immense potential in enhancing the efficiency, accuracy, and effectiveness of healthcare systems, and the intersection of machine learning (ML) [28] and healthcare has garnered significant attention, promising transformative changes in the delivery and optimization of healthcare services [29], [30]. With respect to previous review studies related to machine learning-based authentication models for smart healthcare systems, Table I below compares previous studies with this research.

TABLE I. COMPARISON OF PREVIOUS STUDIES

References	Authentication method	Benefits	Limitations
[23]	Touchscreen input can authenticate users based on behavioral patterns.	Provides insights on user behavior for authentication systems.	Deep neural networks (DNNs) is resource-intensive and time-consuming
[24]	Multimodal system combines fingerprint, face, age, and gender biometrics.	Enhanced security through multimodal biometric authentication system.	Multimodal systems are more expensive than unimodal systems due to the system itself, necessary computing power, and storage space for biometric data.
[25]	LSTM network improves security authentication performance for IoT network sinks.	Offers continuous authentication for improved security on smartphones.	Disadvantages include high computational cost and convergence problems.
[26]	Soft keyboard typing behavior used for continuous user recognition.	Improved authentication accuracy rate compared to non-learning methods.	LSTM method may require a lengthy input signal for high accuracy, making them less suitable for real-time systems.
This research	ForestPA is used to enhance authentication accuracy.	Provides insights on user behavior for authentication systems and combined with ForestPA classifier (fast and accurate model).	Our scope focuses on smartphone devices only. As doctors or medical assistants typically used it for easy access to monitor patients. For future work, this research may focus on other devices than smartphone (Electronic Health Record (EHR) Systems or Telemedicine Platforms).

In the comparison of the studies (Table I), each research work explores different aspects of user authentication methods, focusing on various biometric and behavioral patterns for enhancing security measures. The study [23] delves into the utilization of touchscreen input to authenticate users on the basis of their behavioral patterns. This approach provides valuable insights into user behavior for authentication systems. However, the results from the experiment yielded an accuracy of 100% for the deep neural network (DNN) for 30 features. Hence, the results are still open for improvement for fewer features. Additionally, DNNs need to be trained on very large datasets and computational resources. On the other hand, the ForestPA resides in our experiment, which is a forest-based algorithm that usually demands less computational power and can achieve decent performance with smaller datasets. Another challenge is that DNNs are often seen as “black boxes” because their decision-making processes are extremely intricate and opaque, which can make their predictions difficult to interpret and understand. In contrast, ForestPA algorithms are relatively interpretable, which makes it easier for the user to analyse and validate the decisions made by the model. Another common pitfall of DNNs occurs when the training data lack diversity, so they overfit, whereas forest-based algorithms such as the ForestPA are much more robust for this type of scenario, as they are designed as ensembles. In addition to these characteristics, ForestPA stands out as a practical and accessible method for medical applications, especially when interpretability and computational feasibility are needed. On the other hand, similar DNN types, which are gated recurrent units (GRUs) and vanilla recurrent neural networks (RNNs), both algorithms require significant computational resources and time for training, making them less efficient in resource-constrained environments. They also face interpretability challenges due to their complex architectures, which makes it difficult to follow how decisions

are made. Vanilla RNNs also include the limitations of vanishing and exploding gradients; hence, they are short-lived in learning long-term dependencies. However, the ForestPA algorithm appears to be more computationally efficient, easier to interpret, has low overfitting characteristics and makes it more suitable for several applications.

The research presented in [17], however, has a slightly different focus, as it describes its end users combining multiple biometric characteristics instead of just one and integrating them into what can be called a multimodal system consisting of the fingerprint, face, age, and gender, among others. While the idea focuses on using biological features to authorize users, this system is more costly than the former, as it involves more intricate systems and computational and storage requirements for biometric information. In this respect, since computational speed is vital for smart healthcare authentication, the system needs to deploy a fast and relevant machine learning model.

Next, the study in [25] aims to design a secure mechanism using machine learning-based biometrics to extend the Internet of Things networks against different types of spoofing attacks. This study delves into the soft keyboard typing style as one way of establishing sustained user recognition. It makes use of the physical layer aspects of wireless channels while employing neural networks in gathering channel fingerprints as a means of authentication. Qiu X et al. also developed an LSTM architecture-based detection technique to increase the number of device authentication procedures supporting intelligent algorithms, and a machine learning-based security authentication scheme focused on spoofing attacks in IoT networks. This method achieves higher degrees of accuracy in authentication and is claimed to have better results than no supervised learning methods. However, the long short–term memory (LSTM) model used in this study may require long input signals to be accurate, and this may not be suitable for real-time applications.

Another study[26] is also based on a feature selection technique called correlation and involves rapid user identification via logistic regression. The proposed method yields good classification ratios (as much as 93%) and speeds of processing as low as 0.03 ms. Continuous authentication provides a better security level than does the one-time system, which addresses problems such as password theft or forgetfulness. However, it has advantages as well as demerits such as expensive computations and difficulty in convergence.

In contrast, this study uses the ForestPA machine learning algorithm, which is known for its strong performance and is capable of addressing the complexities of datasets with high levels of accuracy and proficiency. Its construction enhances accurate predictions and analysis of the authentication of users via pattern or behavior modelling. Moreover, one of the most remarkable aspects of ForestPA is the speed at which it is able to process information. This algorithm is able to efficiently and quickly scan a greater amount of information, which makes it suitable for real-time user authentication systems where time is an essential factor. Therefore, the present study investigates the potential of this algorithm for user authentication in healthcare systems.

TABLE II. PREVIOUS STUDIES THAT INVOLVED THE FORESTPA ALGORITHM

References	Year	Domain area that utilized ForestPA algorithm
[31]	2020	Security detection
[32]	2020	Healthcare
This experiment	Ongoing	Security and healthcare

In Table II, a study [31] developed a method to predict associations between circular RNAs (circRNAs) and diseases, aiding in understanding disease pathogenesis and improving diagnosis and treatment. The graph convolutional network for circRNA–disease association (GCNCDA) combines disease semantic similarity, Gaussian interaction profile kernel similarity, and known circRNA–disease associations to predict potential associations. Wang L et al. achieved 91.2% accuracy with an AUC of 90.90% on the circR2Disease benchmark dataset, outperforming other methods. In the validation part, case studies on breast cancer, glioma, and colorectal cancer confirmed the effectiveness of GCNCDA, with most of the top predicted circRNAs validated in the literature and databases. This study proposes a computational method called GCNCDA, which is based on the deep learning fast learning with graph convolutional networks (FastGCN) algorithm, to predict potential disease-associated circRNAs. The new circRNA–disease associations are accurately predicted by the ForestPA classifier, which presents a computational method called GCNCDA for predicting circRNA–disease associations via a graph convolutional network algorithm.

Another study [32] proposed 3 meta-learner models based on the forest penalizing attributes (ForestPA) algorithm. The ForestPA uses a weight assignment and weight increment strategy to build highly efficient decision trees by exploiting the

process of all attributes (nonclass inclusive) in each dataset. The proposed meta-learners (ForestPA-PWDM, Bagged-ForestPA-PWDM, and Adab-ForestPA-PWDM) demonstrated high efficiency, with the lowest accuracy of 96.26%, a false alarm rate (FAR) of 0.004, and an ROC value of 0.994. The authors recommend the development and adoption of meta-learners based on the ForestPA for phishing website detection and other cybersecurity attacks.

Given the esteemed reputation of the ForestPA algorithm within both security and healthcare domain applications, this research endeavors to investigate its potential for user authentication within the smart healthcare domain.

This experiment significantly contributes to the ongoing research in machine learning-based authentication systems by investigating the application of an updated forest-based algorithm, ForestPA, for identifying user profiles as either real or fake on the basis of behavioral patterns. This approach enhances the security framework by adding an additional layer of authentication. Consequently, even if an attacker manages to bypass conventional authentication mechanisms, the user profile authentication system will reject access if it determines that the attacker is not an authenticated doctor authorized to use the mobile device. Implementing a dual-layer authentication process is extremely important when it consists of sensitive medical data and needs to control who actually accesses the system. This authentication could improve the effectiveness, efficiency, and user experience of mechanisms for continuous authentication and be a valuable contribution to the body of knowledge in the domain.

2.1 Forest penalizing Attribute (ForestPA) algorithm

The Forest PA algorithm [27] is a mathematical algorithm that assigns weights to nonoverlapping attributes (pieces of information) in the process of building decision trees:

a) Attributes and Weights: Within Forest PAs, each attribute has a weight that indicates its value in making predictions. A greater weight indicates the importance of the attribute, whereas a smaller weight indicates less importance.
b) Penalizing Attributed Already Featured By Trees: An attribute is punished when it is used in a decision tree; its weight decreases for coming trees. This helps return trees and encourages the algorithm to use different attributes for subsequent trees. This penalty can be understood as saying, “Let’s not depend on this attribute too much.”
c) Boosting Unused Attributes: Conversely, the weights of features that were not used in the previous tree are increased. This indicates that they assume greater weight for the coming tree.
d) Weight assignment equation: Although the precise mathematical notation might be complicated, the essential premise is that the weight of an attribute is changed in the appropriate direction if it is included in the preceding tree. For instance:
   If attribute A was used in the previous tree, its new weight could be calculated as:
   New Weight (A) = Old Weight (A) − Penalty
   If it was not used, its new weight could be as follows:
   New weight (A) = Old weight (A) + Boost
e) Random Selection: This method also randomly selects weights from a range for attributes at the same level. This randomness is critical because it ensures that the trees are different from one another and do not look too similar.

Overall Objective: The overall objective of this process of weighting is to develop a collection of models that are so different from each other that when aggregated together, they produce a stronger prediction than any of the individual trees making up the set. This is accomplished through two strategies: penalizing certain attributes and boosting them so that attributes of diverse types can be considered.

3. METHODOLOGY

The methodology used in this research to create an authentication model based on user behaviors and machine learning methods is described in Figure 1 below. This framework is segmented into three distinct phases. These steps outline the key of the framework, which represents a holistic perspective for implementing the authentication activity on the basis of sophisticated machine learning algorithms. This approach reflects the approach to authenticity as a rigorous exercise in designing machine learning solutions.

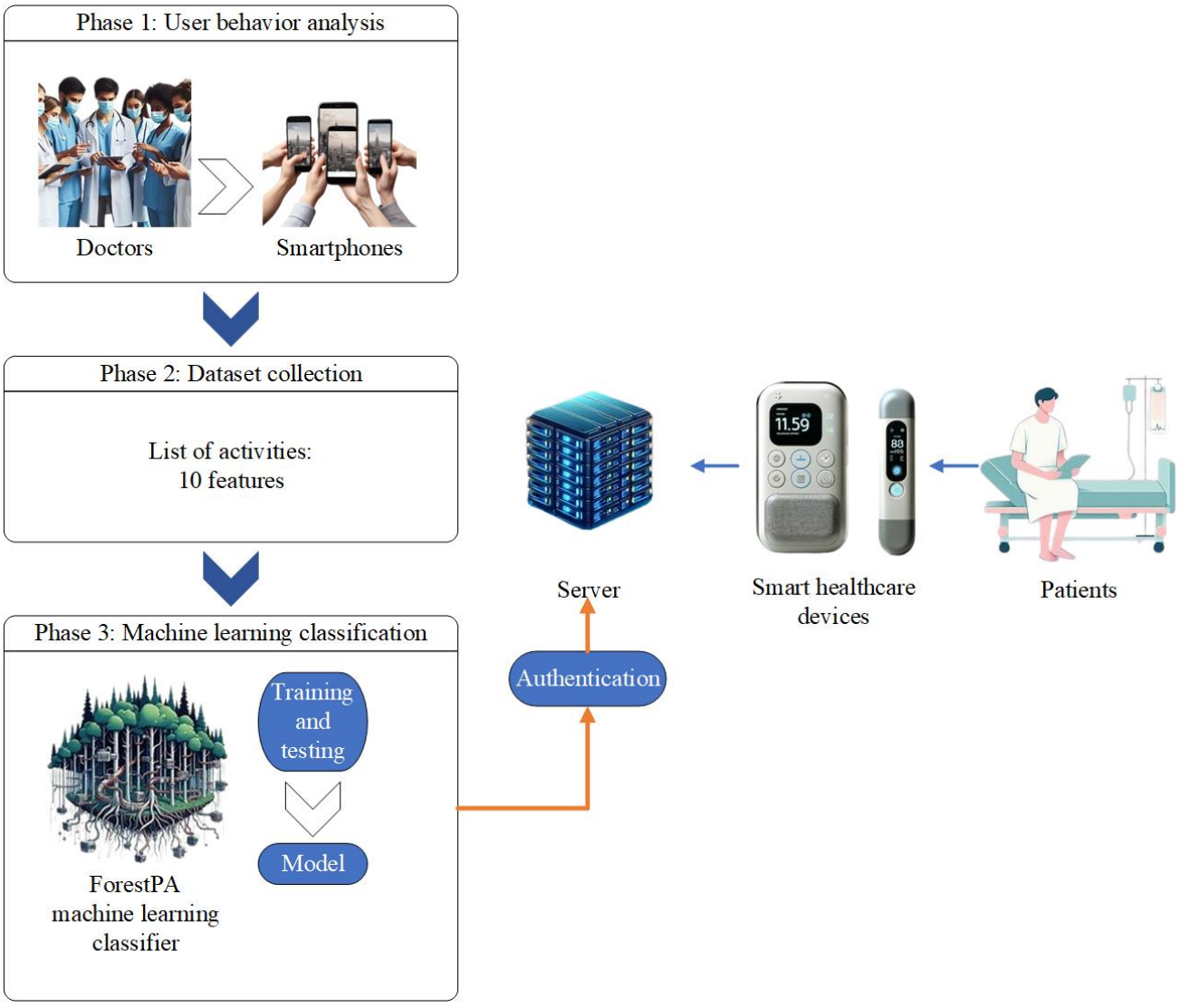

Fig. 1. Methodology

3.1 Phase 1: User behavior analysis

The methodology depicted in the figure delineates the initial phase, which encompasses the conduct of medical professionals (including doctors, nurses, or medical assistants) who are authorized to access patients' records. These healthcare personnel utilize smartphones or tablets for convenient access and mobility within the hospital premises, facilitating patient monitoring and examination. Comprehensive records of each medical personnel's activities and interventions are meticulously documented to maintain a comprehensive log of their actions.

3.2 Phase 2: Dataset collection

During this phase, the smartphones captured the users' interactions through various touch-related activities. These processes included motions such as raising the finger, lowering the finger, moving the finger, applying pressure on the screen, the area of the screen covered by the finger, the orientation of the finger, and occurrences of multitouch. Each experiment in the reading or image comparison test took a set amount of time for each one, measured in milliseconds. The time to conduct each experiment for each participant was between 25 and 50 minutes. On average, an experiment involving reading was conducted within a period ranging from 10--15 minutes, and the amount of time it took to conduct every experiment in image comparison amounted to approximately 3--4 minutes. In this period, the ForestPA machine learning classification was conducted by researchers with five different users' profiles with several features in this dataset.

3.2.1 Features in the dataset

For this purpose, the validation of diverse user profiles with machine learning techniques involved a different group of five people, and all of them used different devices, each doing a series of activities in diverse sessions. The features used in the experiment are presented in Table III. An extensive theoretical foundation of this research was a dataset of actions performed by these participants, amounting to 116,565 actions.

TABLE III. FEATURES USED IN THE EXPERIMENT

Features	More information
phoneID	Indicates the phone: Nexus 1, Nexus S, Samsung Galaxy S and Droid Incredible.
docID	Indicates the activities of the user.
time [ms]	Absolute time of recorded action.
action	Touch down, touch up, or move finger on screen.
phone_orientation	Values returned from the Android API at the current actions.
x_coordinate	
y_coordinate	
pressure	
area_covered	
finger_orientation	

PhoneID: The phoneID feature refers to specific phone models, which include Nexus 1, Nexus S, Samsung Galaxy S and Droid Incredible. Each device has its own properties, such as screen size, resolution, and touch sensitivity, which may affect the natural behavior of a user. This implementation is crucial, as it considers the different versions of application runtimes across various device types, yielding the most accurate and fine-tuned data on the basis of the analysis performed upon data collection from a variety of device models.
docID: The docID feature is an exclusive identity for every user session or activity, which allows researchers to monitor user activities over time. This enables a better understanding of user–device interactions in different contexts. The docID feature, for example, recognizes whether the person is browsing, gaming, or typing and analyses the patterns for each. This helps provide context for user interactions and detect trends or anomalies.
time [ms]: The time [ms] field records the accurate time of each user action in milliseconds, providing the temporal data needed to analyse the timing and sequence of events. Such data are crucial for extracting insights about user behavior (how long they take to respond to a stimulus or how long they engage in a particular task). Thus, temporal analysis helps in understanding the dynamics of user interactions and ultimately helps in interface design.
Action: The action feature describes the type of user interaction with the screen (touch down, touch up or move). This classification helps in understanding user navigation behavior. For example, touch down and touch up actions can represent typing or tapping, whereas move actions can signify scrolling or dragging. This examination of these activities allows researchers insights into the way in which users behave, ultimately contributing to the creation of better intuitive, user-friendly interfaces.
phone_orientation: The phone_orientation value, obtained from the Android API, indicates the orientation of the device at the time of user interaction. These data indicate whether users prefer portraits or landscape modes or change orientations often. These insights can be leveraged to make application layouts and functionalities more in line with user interests.
x_coordinate and y_coordinate: The x_coordinate and y_coordinate features correspond to the horizontal and vertical positions of a touch event, respectively, indicating their positions in pixels from the left edge and top edge of the screen. This spatial information, however, is essential for identifying touch behavior patterns, for example, which areas of screen users commonly interact. Popular touch zones may indicate commonly used features or navigation paths.
Pressure: The pressure value represents how much pressure the user's finger is putting on the screen, which is useful for differentiating between a light touch and a strong touch. These data can be helpful in monitoring interaction strength and discovering touch pressure differences. Higher pressure might signal deliberate actions, for example, while lighter touches could suggest casual browsing. Touch pressure aids in predicting user intent and consequently makes the touch interface more responsive.
area_covered: The area_covered value represents the area of the screen covered by the finger of the user during the touch. This indicates the size of the touch area, which can differ depending on the finger's contact surface. For example, larger touch areas could mean multiple fingers at once or a wider touch, whereas smaller areas might suggest more precise interactions. The data from touch are a great way to understand user interactions and accordingly improve or fine-tune the touch response, sensitivity or accuracy.
finger_orientation: This is an indication of the angle at which a user's finger touches the screen, therefore also providing insight into the manner in which a device is used. By collecting and analysing the data, which are information on how users grip or move their device, designers can create more usable products with respect to ergonomics. Certain types of finger

movements can be more common in certain tasks, such as typing or swiping, and identifying how those types of movements appear in real-world datasets could be used to assist in future processing and overall human–computer interactions.
Together, these features provide a sturdy dataset for investigating user interactions with touchscreens, laying the groundwork for this work.

3.3 Phase 3: Machine learning classification

In this study, the recently well-established advanced form of the ForestPA, a type of efficient machine learning algorithm, analyses the characteristic touch input patterns of classes for classification and good separation within the available performance. The model needs to learn very specific nuances of each person's interaction patterns and context to correctly classify them. This method guarantees clarification to be both speedy and precise, taking advantage of trivial differences within user interactions to attain superior outcomes.
The ForestPA algorithm is an approach for building a decision forest. It is an ensemble of decision trees. Since the algorithm seeks to find the best splitting attribute in the tree at each node of the tree, it first assigns the properties of the dataset, or rather features, into weights and then uses those weights to pick the best attribute for splitting. Additionally, this approach modifies the weights of the attributes upon constructing each of the trees. This approach aims to prevent excessive fitting and to stimulate variance among the trees.
The performance of the training model on a wide range of performance measures was evaluated comprehensively during the final phase of the project. The ForestPA training model was developed to differentiate among users. This comprehensive evaluation approach was implemented to assess the efficiency and robustness of the model.

3.4 Authentication part of the methodology

Once the ForestPA machine learning model has been completed, it is deployed for server authentication. Initially, authentication is performed through standard procedures, where users must provide valid credentials, such as usernames, passwords, PINs, and pass antivirus checks on their mobile devices. Successful initial authentication grants medical staff the ability to access patient data via smart healthcare devices, such as smart thermometers and wearable biosensors worn by patients.
Following initial authentication, the ForestPA operates by continuously monitoring the authenticated user's behavior through various indicators, including typing patterns, device handling, application usage, and access patterns to healthcare data. This behavioral analysis is compared against previously established legitimate user profiles constructed through an initial training phase. If an attacker successfully bypasses the mobile device’s primary security measures, the ForestPA model continues to analyse the ongoing behavior patterns.
If any observed behavior deviates from these legitimate user profiles, the ForestPA classifies the activity as fraudulent, automatically logs out the attacker, and triggers an alarm to notify security personnel. By continuously monitoring user behavior post authentication, the ForestPA model provides an enhanced, persistent layer of security that remains effective even when primary credentials, such as passwords or PINs, have been compromised. This real-time detection mechanism helps swiftly identify and respond to suspicious activities, automatically logging out intruders and triggering alerts to minimize potential damage or data loss. Furthermore, the proactive and continuous nature of this system fosters greater user trust, as it offers an advanced security measure that extends beyond traditional static authentication protocols.

4. RESULTS

A detailed experimental study of the proposed method is included in this section of the paper. It includes details of the experimental setup, performance metrics, and performance review.

4.1 Experimental setup

The performance of the ForestPA algorithm was rigorously analysed through a comprehensive set of experiments, each tailored to explore various aspects of its operation:

a) For the first time, a split of 80/20 data points was determined using only 80% of the data for training and 20% of the data for testing. This was a more realistic definition of algorithm performance.
b) A 70/30 division of data followed, where 70% was used for training and the other 30 for testing. This setup enabled an assessment of the performance of the algorithm while considering a larger portion of the dataset aside for evaluation.
c) A fivefold cross-validation process was used to estimate the generalization ability of the algorithm. In this approach, every fivefold subset is selected as a test set in which the specific subset and the other four remaining subsets are trained.
d) Finally, the cross-validation method for training was more intrusive when the dataset was split into seven folds. This approach, which increases the number of folds, aims to increase the accuracy and robustness of the algorithm.

4.2 Performance metrics

Several performance measures are used in delivering machine learning categorization, with an estimate of the accuracy and efficacy of single models. These measures provide a brief overview of how the ForestPA model performs in accurately predicting outputs. The explanations of the most important performance measures for this experiment are detailed in the section below. These main measures are accuracy, the root mean square error (RMSE), the mean absolute error (MAE), and the relative absolute error (RAE).

a) *Accuracy:* accuracy is the measure of how accurate the model is in predicting outcomes. It counts correctly predicted according to all sample instances. This is a very straightforward and direct measure of the accuracy of a model in making correct predictions.
b) *Root mean square error (RMSE):* the square root of the mean of the squared differences between the expected and observed values. In this way, the test is fair in estimating the model's performance because it does not suffer from sensitivity by means of extreme values.
c) *Mean absolute error (MAE):* The mean average of the absolute errors in every set of forecasts, irrespective of the direction, is a measure of the mean absolute error (MAE). The results are represented by a very simple judgment of the accuracy of the forecasts.
d) *Relative absolute error (RAE):* The RAE provides a measure in which the MAE stands against a baseline predictor that predicts the mean of the real values every time. It provides a normalized method of calculating the model's inaccuracy in relation to an elementary estimate.

4.3 Performance review

The evaluation of performance metrics is the most critical part of the development and checking of machine learning models. This helps in determining whether the model has learned the underlying patterns and can generalize well to new data, which is particularly important in smart healthcare applications. Table IV below tabulates the results of the ForestPA algorithm.

TABLE IV. PERFORMANCE EVALUATION

Performance metrics	Training and testing		Cross validation	
	80% train, 20% test	70% train, 30% test	5 folds	7 folds
Accuracy	99.9957%	99.9990%	99.9991	99.9983
Mean absolute error	0.0005	0.0002	0.0005	0.0005
Root mean squared error	0.0075	0.0049	0.0076	0.0069
Relative absolute error	0.1592	0.0629	0.1714	0.1656

Accuracy is the most intuitive performance measure, and it provides a general idea of how often the model makes the correct prediction. It is the ratio of the number of correct predictions to the total number of input samples. The accuracy values provided (ranging from 99.9957% to 99.9991%) are exceptionally high, which generally indicates that the ForestPA algorithm performs excellently on the dataset used.

The MAE measures the average magnitude of the errors in a set of predictions without considering their direction. It is the average over the test sample of the absolute differences between the prediction and actual observations where all individual differences have equal weights. The MAEs reported are very low (0.0002--0.0005), suggesting that the predictions made by the model are, on average, very close to the actual values. This demonstrates a positive sign, especially in healthcare applications within the IoMT, where precise predictions can be critical.

The RMSE is a quadratic scoring rule that also measures the average magnitude of the error. It is the square root of the average of the squared differences between the prediction and actual observations. The RMSE values are slightly higher than the MAE values are, which is expected since the RMSE gives more weight to larger errors. This means that the RMSE is sensitive to outliers. A low RMSE (0.0049 to 0.0076) indicates good predictive performance, with the model's predictions deviating little from the true values.

The RAE, also known as the mean absolute percentage error (MAPE), is a measure of the prediction accuracy of a forecasting method in statistics. It compares the absolute error between the predicted value and the actual value over the sum of the absolute differences between the actual values and the mean of all the actual values. The lower values of the RAE (0.0629--0.1714) indicate that the model has a lower prediction error relative to the simple baseline model for predicting the mean, hence demonstrating the effectiveness of the ForestPA algorithm.

The ForestPA algorithm appears to perform exceptionally well across all the metrics. The consistency of the high accuracy and low error rates across different data splits and cross-validation folds suggests that the model is robust and generalizable well to unseen data. The results are indicative of a well-fitting model that could be highly suitable for IoMT environments where predictive accuracy is crucial for decision-making, such as in patient monitoring systems or diagnostic tools.

In the context of smart healthcare systems, where models may be used for critical health-related predictions, the ability to accurately predict new samples is crucial. In regard to patient monitoring, personalized healthcare, and other medical applications, this reliability is absolutely necessary because inaccurate forecasts could have major implications. The efficiency of the ForestPA model was evaluated through the use of fresh samples, which were assessed as part of this research.

TABLE V. TESTING NEW SAMPLES

New test samples (excluded from training set)	Number of samples	Results
user1	Ten	Correctly predicted
user2	Ten	
user3	Ten	
user4	Ten	
user5	Ten	

From Table V, the model was given ten new samples (for each user) for prediction. It is clear that the model has learned the underlying patterns in the data rather than simply memorizing the training set, as evidenced by the fact that all of the predictions were accurate for each individual user. A well-generalized model is distinguished by its ability to apply previously learned principles to new contexts.

Further analysis of these results indicates that the model has a strong ability to generalize user-specific behavioral traits effectively. The perfect prediction accuracy demonstrates that ForestPA not only accurately captures individual user patterns but also robustly identifies deviations, validating its potential efficacy in real-world deployment scenarios. Additionally, the consistency of correct predictions across diverse user samples underscores the model's reliability in identifying authentic user behaviors and proactively detecting unauthorized activities.

Moreover, these findings suggest that the ForestPA model can be effectively scaled to larger user populations without significant loss in accuracy. The robustness observed in the test predictions implies that the model can reliably adapt to varying user behaviors and patterns, even under conditions involving dynamic or evolving usage scenarios. Consequently, this adaptability ensures sustained effectiveness in practical deployments, reinforcing ForestPA's ability as a valuable security enhancement tool within smart healthcare systems.

TABLE VI. THE RESULTS COMPARISON WITH SIMILAR DATASETS

References	Number of features	Accuracy (%)
[33]	34	100.00 (Deep Neural Net)
This experiment	10	99.999 (ForestPA)

Moreover, for results comparison, to conduct a meaningful performance comparison, it is essential to compare the results with those of other studies that use the same dataset because consistency in the data conditions is ensured, thus enabling a fair and reliable assessment of different methods or models and facilitating a clearer evaluation of their relative strengths and weaknesses.

When different datasets are used, the results may not be directly comparable because variations in factors such as data distribution, labelling standards, or complexity can lead to inconsistent performance assessments, thereby undermining the reliability and fairness of the comparison. Specifically, data distribution differences might reflect varying user demographics, behavioral diversity, or environmental contexts, causing models to perform differently even with similar methodologies. Labelling standard discrepancies can result from inconsistent criteria for identifying or classifying behaviors, leading to varying interpretations of what constitutes legitimate or fraudulent activities. Additionally, dataset complexity, which includes variations in the volume of data, the level of behavioral granularity, and noise levels, can significantly impact model performance and its ability to generalize effectively. Therefore, careful consideration and standardization of these factors are crucial when interpreting results across different datasets.

Since no other studies have used this specific dataset, our results can only be compared with those of that study [33], which we used for datasets in our experiment as well. They implemented a deep neural network to achieve 92% accuracy. They used 34 features and applied a deep neural network to achieve 100% accuracy. Compared with the research presented in [33], where 34 features were used to achieve 100% success by a deep neural network, the current experiment provides a highly competitive success of 99.999% while using only 10 features via the ForestPA model. The new method drastically decreases the number of features from a total of 34 to only 10, reducing the associated computational overhead and making it more feasible for constrained mobile devices where processing power and battery life are at a premium. This efficient but near-perfect identification of fakes is crucial in the real-time detection of fake user profiles since it decreases energy consumption and minimizes the pressure of hardware on devices. Furthermore, there is often a faster inference time with fewer features,

leading to more rapid detection and response to potential security threats. Therefore, this experiment demonstrates a possible approach to provide strong security in mobile devices without degrading performance or exhausting system resources.
Alternatively, the experiment using the ForestPA algorithm attained an extraordinary accuracy rate of 99.999% with only ten features, indicating its remarkable performance in the experimental framework. Through the utilization of this technique, which most likely places an emphasis on decision tree-based ensemble learning, this experiment demonstrates the potential for reaching near-perfect accuracy rates in classification tasks. This is accomplished by concentrating on a very limited number of features. In addition, the reduction in the number of features of the ForestPA model makes it easier to read and makes it less likely to be overfit.
Additionally, the exceptionally high detection rate shows that ForestPA can retain its strong defense against fraudulent user activities with a significantly smaller number of input parameters. In practical terms, this illustrates a firmer security position, whereas devices can monitor for potential intrusions continuously without sacrificing performance or draining system resources. These results demonstrate an optimal trade-off between security strength, power efficiency, and operational speed, which is essential for protecting mobile devices from unauthorized access.
Furthermore, the model has potential for use in real time within smart healthcare devices. In that way, it can process data streams and analyse them in real time to provide timely insights or alarms since the model can predict all the new test samples accurately. The performance capacity of the model to exactly predict all new test samples in the context of the smart healthcare framework defines an earnest logic for the assured functioning of its operational deployment.

5. CONCLUSIONS AND FUTURE WORK

The current work was able to demonstrate the fine capability of the forest penalizing attributes (ForestPA) algorithm for classifying novel, previously unseen data. In applying such models, especially those aimed at reliably recognizing the genuine user from the authentication samples, excellent performance has been consistently observed, even under very rigorous testing using different users. The quick processing capability and high accuracy of the algorithm turn out to be a reliable and efficient choice, considering smart healthcare systems, when health information is to be handled very sensitively, and resources need to be authenticated both quickly and precisely.
In addition, the high level of accuracy and speed that the algorithm offers makes it a very good candidate for implementation in intelligent healthcare devices, where processing power is usually very constrained. The results indicate that the ForestPA algorithm can be implemented in intelligent healthcare systems for the purpose of quick and secure user authentication. This is crucial in the process of maintaining confidential patient information and privacy.
The proposed approach was evaluated in a simulation environment in this study to ensure that its performance is in line with expectations in a controlled environment. Simulation-based experimentation serves as a proof of concept and support for optimizations. Therefore, the next step in the research will be to build and test a physical hardware prototype, validating the robustness, reliability, and scalability of the system in actual operation.
Furthermore, the current study examined only five profiles, which limits the generalizability of the findings. Nonetheless, it is essential to mention that in future work, there is an opportunity to increase the number of profiles to better encompass more diverse user personalities and hence increase the strength of our experiments. More generally, this strategy can increase the generalizability of our findings and support practical implementation.

Conflicts of interest

The authors declare that they have no conflicts of interest.

Funding

This research is made possible and supported by the UMP-IIUM Sustainable Research Collaboration 2022 Research Grant (IUMP-SRCG22-014-0014) and Universiti Malaysia Pahang Al-Sultan Abdullah (UMPSA) under Internal Research grant RDU220362.

Acknowledgement

This research is made possible and supported by the UMP-IIUM Sustainable Research Collaboration 2022 Research Grant (IUMP-SRCG22-014-0014) and Universiti Malaysia Pahang Al-Sultan Abdullah (UMPSA) under the Internal Research Grant RDU220362.

References

[1] S. Namasudra, “Taxonomy of DNA-based security models,” in *Advances of DNA Computing in Cryptography*, 2018, pp. 37–52.
[2] M. Zubair *et al.*, “Secure bluetooth communication in smart healthcare systems: a novel community dataset and intrusion detection system,” *Sensors*, vol. 22, no. 21, Nov. 2022, doi: 10.3390/s22218280.

[3] D. Xu, W. H. Chan, H. Haron, H. W. Nies, and K. Moorthy, "From COVID-19 to monkeypox: a novel predictive model for emerging infectious diseases," *BioData Min*, vol. 17, no. 1, Dec. 2024, doi: 10.1186/s13040-024-00396-8.

[4] S. Das and S. Namasudra, "MACPABE: Multi authority-based CP-ABE with efficient attribute revocation for IoT-enabled healthcare infrastructure," *International Journal of Network Management*, vol. 33, no. 3, 2023.

[5] M. Rahman, A. Murmu, P. Kumar, N. R. Moparthi, and S. Namasudra, "A novel compression-based 2D-chaotic sine map for enhancing privacy and security of biometric identification systems," *Journal of Information Security and Applications*, vol. 80, Feb. 2024, doi: 10.1016/j.jisa.2023.103677.

[6] P. Kumar, M. Rahman, S. Namasudra, and N. R. Moparthi, "Enhancing security of medical images using deep learning, chaotic map, and hash table," *Mobile Networks and Applications*, 2023, doi: 10.1007/s11036-023-02158-y.

[7] P. O. A. de la Rosa, "The alarming reality of cyberattacks on our healthcare information." Accessed: Dec. 03, 2023. [Online]. Available: https://laopinion.com/2023/11/27/the-alarming-reality-of-cyberattacks-on-our-healthcare-information/

[8] J. Mohamad Arif, M. F. Ab Razak, S. Awang, S. R. Tuan Mat, N. S. N. Ismail, and A. Firdaus, "A static analysis approach for Android permission-based malware detection systems," *PLoS One*, vol. 16, no. 9, pp. 1–23, 2021, doi: 10.1371/journal.pone.0257968.

[9] A. Karim, V. Chang, and A. Firdaus, "Android botnets: A proof-of-concept using hybrid analysis approach," *Journal of Organizational and End User Computing*, vol. 32, no. 3, pp. 50–67, 2020, doi: 10.4018/JOEUC.2020070105.

[10] N. J. Palatty, "80+ Healthcare Data Breach Statistics 2023." Accessed: Dec. 03, 2023. [Online]. Available: https://www.getastra.com/blog/security-audit/healthcare-data-breach-statistics/

[11] J. Yang, "Healthcare cybersecurity spending globally 2023 Statista." Accessed: Dec. 03, 2023. [Online]. Available: https://www.statista.com/statistics/1359081/cybersecurity-spending-in-healthcare-sector-worldwide/

[12] S. Kamil, H. S. A. Siti Norul, A. Firdaus, and O. L. Usman, "The rise of ransomware: a review of attacks, detection techniques, and future challenges," in *2022 International Conference on Business Analytics for Technology and Security, ICBATS 2022*, Institute of Electrical and Electronics Engineers Inc., 2022, pp. 1–7. doi: 10.1109/ICBATS54253.2022.9759000.

[13] M. L. Shuwandy, R. A. F. Alsharida, and M. M. Hammood, "Smartphone Authentication Based on 3D Touch Sensor and Finger Locations on Touchscreens via Decision-Making Techniques," *Mesopotamian Journal of CyberSecurity*, vol. 5, no. 1, pp. 165–177, Jan. 2025, doi: 10.58496/MJCS/2025/011.

[14] S. A. Hussein *et al.*, "Integrating Law, Cybersecurity, and AI: Deep Learning for Securing Iris-Based Biometric Systems," *Mesopotamian Journal of CyberSecurity*, vol. 5, no. 2, pp. 319–336, May 2025, doi: 10.58496/MJCS/2025/020.

[15] C. Chen, H. Guo, Y. Wu, B. Shen, M. Ding, and J. Liu, "A lightweight authentication and key agreement protocol for IoT-enabled smart grid system," *Sensors*, vol. 23, no. 8, Apr. 2023, doi: 10.3390/s23083991.

[16] S. Namasudra and P. Roy, "A new table based protocol for data accessing in cloud computing," *Journal of Information Science and Engineering*, vol. 33, no. 3, pp. 585–609, May 2017, doi: 10.6688/JISE.2017.33.3.1.

[17] N. N. C. Razali, N. A. Ghani, S. I. Hisham, S. Kasim, N. S. Widodo, and T. Sutikno, "Rainfall-runoff modelling using adaptive neuro-fuzzy inference system," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 17, no. 2, pp. 1117–1126, Feb. 2020, doi: 10.11591/IJEECS.V17.I2.PP1117-1126.

[18] Soni, M. A. Remli, K. M. Daud, and J. Al Amien, "Ensemble learning approach to enhancing binary classification in Intrusion Detection System for Internet of Things," *International Journal of Electronics and Telecommunications*, vol. 70, no. 2, pp. 465–472, Jun. 2024, doi: 10.24425/ijet.2024.149567.

[19] L. Breiman, "Random forests," *Mach Learn*, pp. 5–32, 2001, doi: 10.1023/A:1010933404324.

[20] A. Nur Iman and T. Ahmad, "Improving intrusion detection system by estimating parameters of Random Forest in Boruta," in *International Conference on Smart Technology and Applications (ICoSTA)*, 2020, pp. 1–6. doi: 10.1109/ICoSTA48221.2020.1570609975.

[21] Y. M. Putra Abdullah, S. Abu Bakar, W. N. J. Wan Yussof, R. Hamzah, R. A Hamid, and D. Satria, "A comparative study of image retrieval algorithm in medical imaging," *International Journal on Informatics Visualization*, no. 8, pp. 1567–1572, Nov. 2024, [Online]. Available: www.joiv.org/index.php/joiv

[22] M. Christine, H. H. Nuha, M. Irsan, A. G. Putrada, and S. Izhar Hisham, "Object tracking in surveillance system using extended Kalman Filter and ACF detection," in *International Conference on Decision Aid Sciences and Applications (DASA)*, Institute of Electrical and Electronics Engineers Inc., 2024. doi: 10.1109/DASA63652.2024.10836597.

[23] M. Frank, R. Biedert, E. Ma, I. Martinovic, and D. Song, "Touchalytics: On the applicability of touchscreen input as a behavioral biometric for continuous authentication," *IEEE Transactions on Information Forensics and Security*, vol. 8, no. 1, pp. 136–148, 2013, doi: 10.1109/TIFS.2012.2225048.

[24] R. Brown, G. Bendiab, S. Shiaeles, and B. Ghita, "A novel multimodal biometric authentication system using machine learning and blockchain," in *12th International Network Conference*, 2020, pp. 31–46. doi: 10.1007/978-3-030-64758-2_3.

[25] E. A. Sağbaş and S. Ballı, "Machine learning-based novel continuous authentication system using soft keyboard typing behavior and motion sensor data," *Neural Comput Appl*, 2024, doi: 10.1007/s00521-023-09360-9.

[26] X. Qiu, X. Sun, and X. Si, "Machine learning-based security authentication for IoT networks," in *Lecture Notes of the Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering, LNICST*, Springer Science and Business Media Deutschland GmbH, 2021, pp. 106–115. doi: 10.1007/978-3-030-69514-9_10.

[27] M. N. Adnan and M. Z. Islam, "Forest PA: Constructing a decision forest by penalizing attributes used in previous trees," *Expert Syst Appl*, vol. 89, pp. 389–403, Dec. 2017, doi: 10.1016/j.eswa.2017.08.002.

[28] E. H. Tusher, M. A. Ismail, M. A. Rahman, A. H. Alenezi, and M. Uddin, "Email spam: A comprehensive review of optimize detection methods, challenges, and open research problems," *IEEE Access*, vol. 12, Oct. 2024, doi: 10.1109/ACCESS.2024.3467996.

[29] E. H. Tusher, M. A. Ismail, and A. F. Mat Raffei, "Email spam classification based on deep learning methods: a review," *Iraqi Journal for Computer Science and Mathematics*, vol. 6, no. 1, Feb. 2025, doi: 10.52866/2788-7421.1236.

[30] M. Al-Zubaidie and W. Jebbar, "Transaction security and management of blockchain-based smart contracts in e-banking-employing microsegmentation and yellow saddle goatfish," *Mesopotamian Journal of CyberSecurity*, vol. 4, no. 2, pp. 71–89, May 2024, doi: 10.58496/mjcs/2024/005.

[31] L. Wang, Z. H. You, Y. M. Li, K. Zheng, and Y. A. Huang, "GCNCDA: A new method for predicting circRNA-disease associations based on Graph Convolutional Network Algorithm," *PLoS Comput Biol*, vol. 16, no. 5, May 2020, doi: 10.1371/journal.pcbi.1007568.

[32] Y. A. Alsariera, A. V. Elijah, and A. O. Balogun, "Phishing website detection: forest by penalizing attributes algorithm and its enhanced variations," *Arab J Sci Eng*, vol. 45, pp. 10459–10470, 2020.

[33] M. Montgomery, P. Chatterjee, J. Jenkins, and K. Roy, "Touch analysis: an empirical evaluation of machine learning classification algorithms on touch data," in *Information Security and Cryptology 8th International Conference, Inscrypt, Beijing, China*, Springer Verlag, 2012, pp. 147–156. doi: 10.1007/978-3-030-24907-6_12.

[34] L. A. Al-Haddad, H. A. H. Kahachi, H. Z. Ur Rehman, A. A. Al-Zubaidi, M. I. Al-Karkhi, and B. Al-Oubaidi, "Advancing sustainability in buildings using an integrated aerodynamic façade: Potential of artificial intelligence," *Terra Joule Journal*, vol. 1, no. 1, Art. no. 1, 2024. https://doi.org/10.64071/3080-5724.1006