Using Behavioural Biometrics and Machine Learning in Smart Gadgets for Continuous User Authentication

[1]Deepthi S, [2]Mamatha Balachandra, [3]Prema K V, [4]Kok Lim Alvin Yau and [5]Abhishek A K
[1,2,5]Department of Computer Science and Engineering, Manipal Institute of Technology, Manipal Academy of Higher Education (MAHE), Manipal, Karnataka, India.
[3]Department of Computer Science and Engineering, Manipal Institute of Technology Bengaluru, Manipal Academy of Higher Education, Manipal, Karnataka India.
[4]Lee Kong Chian Faculty of Engineering and Science (LKCFES), Universiti Tunku Abdul Rahman, Sungai Long, Selangor, Malaysia.
[1]deepthi.s@manipal.edu, [2]mamatha.bc@manipal.edu, [3]prema.kv@manipal.edu, [4]yaukl@utar.edu.my, [5]abhishekak0709@gmail.com

Correspondence should be addressed to Mamatha Balachandra : mamatha.bc@manipal.edu

Article Info
Journal of Machine and Computing (http://anapub.co.ke/journals/jmc/jmc.html)
Doi: https://doi.org/10.53759/7669/jmc202404059
Received 10 January 2024; Revised from 22 March 2024; Accepted 06 June 2024
Available online 05 July 2024.

Abstract – In the ever-evolving realm of technology, the identification of human activities using intelligent devices such as smartwatches, fitness bands, and smartphones has emerged as a crucial area of study. These devices, equipped with inertial sensors, gather a wealth of data and provide insights into users' movements and behaviors. These data not only serve practical purposes, but also hold significant implications for domains such as healthcare and fitness tracking. Traditionally, these devices have been employed to monitor various health metrics such as step counts, calorie expenditure, and real-time blood pressure monitoring. However, recent research has shifted its focus to leveraging the data collected by these sensors for user authentication purposes. This innovative approach involves the utilization of Machine Learning (ML) models to analyze the routine data captured by sensors in smart devices employing ML algorithms, which can recognize and authenticate users based on their unique movement patterns and behaviors. This introduces a paradigm shift from traditional one-time authentication methods to continuous authentication, adding an extra layer of security to protect users against potential threats. Continuous authentication offers several advantages over its conventional counterparts. First, it enhances security by constantly verifying a user's identity through their interaction with the device, thereby mitigating the risk of unauthorized access. Second, it provides a seamless and nonintrusive user experience, eliminating the need for repetitive authentication prompts. Moreover, it offers robust protection against various threats such as identity theft, unauthorized access, and device tampering. The application of continuous authentication extends beyond individual devices and encompasses interconnected systems and networks. This holistic approach ensures a comprehensive security across digital platforms and services. The experiments demonstrate that the logistic regression model achieves an accuracy of 82.32% on the test dataset, highlighting its robustness for binary classification tasks. Additionally, the random forest model outperforms with a 92.18% accuracy, emphasizing its superior capability in handling complex feature interactions. In the study, the sequential neural network achieved an accuracy of 92% on the HAR dataset, outperforming traditional machine learning models by a significant margin. The model also demonstrated robust generalization capabilities with a minimal drop in performance across various cross-validation folds.

Keywords – User Authentication, Smartphone, Machine Learning, Behavioral Biometric, Human Activity Recognition.

I. INTRODUCTION

The discipline of artificial intelligence for healthcare has focused more on human activity recognition (HAR) as a study area with the goal of enhancing human health by using computers to predict human behavior or posture [1-3]. Temporal signals corresponding to human activity can be obtained using sensors, such as magnetometers, gyroscopes, triaxial accelerometers, and gravity sensors. This method can record activity data without compromising user privacy or the surroundings. In recent years, the utilization of smart gadgets equipped with inertial sensors, such as smartwatches, fitness bands, and smartphones, has revolutionized the realm of human activity recognition. These devices continuously collect vast amounts of data regarding user

movements, enabling applications across diverse domains such as healthcare and fitness tracking. From counting steps and estimating calorie expenditure to monitoring blood pressure in real time, insights derived from inertial sensor data have significantly enhanced our understanding of human behavior and health metrics. However, amidst these advancements, researchers are exploring novel applications of these sensor data beyond traditional health and fitness monitoring. One promising avenue is the development of user authentication systems that leverage machine-learning (ML) models[4-5]. Unlike conventional static authentication methods that typically rely on passwords or biometric data, this approach harnesses the wealth of information captured by inertial sensors to create dynamic and personalized authentication mechanisms. The concept involves training an ML model on routine data collected by smart gadgets over time. By analyzing patterns in user movements, gait, and other behavioral cues, the model learns to distinguish between genuine and unauthorized individuals. This approach offers several potential advantages over traditional authentication techniques [6-9]:

Continuous Authentication (CA)
Unlike one-time authentication methods, which only verify users at specific intervals, the ML model can continuously assess the authenticity of users based on their ongoing activity patterns. This dynamic authentication process adds an extra security layer, reducing the likelihood of unauthorized access.

Contextual Awareness
Inertial sensor data provide rich contextual information about user activities and environments. By considering factors such as location, time of day, and typical behavioral patterns, the ML model can adapt its authentication criteria to different scenarios, thereby enhancing its accuracy and robustness.

Non-intrusive User Experience
Because the authentication process relies on passive data collected from everyday activities, users are spared the inconvenience of manual inputs or biometric scans. This non-intrusive approach seamlessly integrates security measures into users' daily routines, minimizing disruptions, while still ensuring protection against unauthorized access.

As a result of the increasing use of smartphones, smartwatches, and other smart devices in the medical profession, wearable sensor-based human health-monitoring systems are emerging quickly. This is because using sensor devices embedded in smartphones to collect data on human activity is more practical, affordable, and realistic than image-based recognition. Mobile devices are replacing more things in our lives. These gadgets can be utilized for business, work, leisure, and communication, among others. They also retain sensitive information, such as images, documents, bank account information, and tickets. If these details enter an attacker, the owner may have consequences. Manufacturers have implemented various authentication alternatives beginning with patterns and PIN numbers to offer protection to users. Manufacturers of mobile devices then include the option of unlocking devices with biometric mechanisms, such as the facial scan or fingerprint of the user [1]. To prevent attacks of this kind, we can develop a CA system [2] that uses a behavioral profile [3] to identify potential attackers and lock the mobile device automatically. Currently, there is a dearth of use for this type of authentication on smartphones. This paper proposes a CA system that leverages public APIs and is built using sensors. The system can take necessary action if it determines that the user of the mobile device is not authenticated. This type of authentication, when paired with conventional static biometric authentication, can offer an additional degree of security and significantly raise the overall security of the system. The prevalence of mobile devices in our lives has led to an increasing amount of confidential data being stored on them, making security a top priority. Although PIN codes, fingerprints, and face recognition are popular authentication methods, they have shortcomings, particularly in terms of static authentication [10]. CA via behavioral biometrics can offer an additional level of security by identifying unauthorized users and automatically locking devices. CA is a way to confirm a user's identity at every moment of the session, not just when they first logged in. By continuously observing a user's hand movements while using a mobile device, this technology seeks to increase security. The system tracks and analyzes the user's hand motions using integrated sensors, such as a magnetometer, gyroscope, and accelerometer. These sensors gather data that are then processed and combined to derive attributes that can be used for authentication. Subsequently, a machine learning classifier receives these features and uses their interpretation to decide whether the current user is authorized or an intruder. A machine learning classifier can be constructed for a system in several ways [11].

The main contribution of this study is the introduction of a novel method for CA in IoT devices by leveraging individuals' behavioral profiles. This approach ensures CA across various IoT applications, offering robust liveliness detection, which poses a significant challenge to potential intruders or hackers attempting to deceive the system. Even if an individual tries to imitate a user to bypass static authentication measures, it still needs to overcome the second layer of authentication, which entails recognizing the unique attributes of the individual's behavioral profile. This dual-layer authentication strategy markedly bolsters system security and increases the difficulty for hackers or intruders to subvert it. Moreover, to align with IoT requirements, key factors, such as low algorithmic complexity, minimal power consumption, and optimal accuracy, are pivotal considerations in the design of the solution.

II. WORK IN THIS AREA

A comprehensive understanding of the current state of research and development in the CA of smartphones is essential for several reasons. Here, an overview of the literature on the CA of smartphones is presented. Dybczak et al.[1] discussed the

importance of CA using behavioral biometrics on mobile devices. It highlights the security risks associated with static, one-time authentication methods and the need for a continuous biometric system to improve security. The system utilizes built-in sensors and public APIs of smartphones to capture and analyze users' hand movements. This integration of hardware and software components enables CA of users based on their unique behavioral patterns . It compares model-based and template-based methods for registering users in the system and discusses the limitations of this type of authentication as an alternative security mechanism. Zhihao Shen et al.[2] proposed CA framework for smartphones that uses incremental learning to address touch behavior issues. It uses context-aware features to describe touch activity patterns, ensuring consistent authentication performance. IncreAuth improves security by identifying users based on hand motions throughout usage sessions, offering modern authentication accuracy and minimal system overheads. Using smartphone sensors, Liang et al. [3] presented a transformer-based deep learning architecture for human activity recognition. Conventional approaches struggle to capture the temporal and geographical correlations of sensor inputs. This problem can be solved using a transformer model that uses self-attention processes. They compared the proposed approach with LSTM and CNN networks. An average accuracy of 94% was achieved using the transformer model. However, the model's capacity for generalization was constrained because the dataset included only six activity classifications.

William et al.[12] discussed the effect of feature extraction and prediction accuracy on human activity recognition. Researchers have used the acceleration data from accelerometers and gyroscopic sensors for activity recognition. They extracted various domain features from the sensor data, such as the mean, standard deviation, correlation, and FFT coefficients. The features with the highest permutation importance were first selected and fed to different classifiers. By selecting more key features, the accuracy was increased and approached using all the features. Ganesh et al.[13] predicted HAR using a Novel CNN algorithm in comparison with a grid search algorithm and it achieved 98.65 percent accuracy compared with 89.6 percent for the grid search method. This shows that the Novel CNN performs significantly better because it incorporates the advantages of both the methods. However, the Novel CNN has some limitations, such as the difficulty in detecting complex human actions and lower segmentation accuracy. Sakorn Mekruksavanich et al.,[14] investigates the robustness of deep learning models for recognizing human actions from smartphone sensor data. Accurately identifying human activities from noisy sensor data remains a challenge. Traditional machine learning techniques rely heavily on manual feature extraction and selection, which can be error prone. Because deep learning approaches can automatically learn characteristics from raw sensor data, they have demonstrated dependable outcomes in sensor-based human activity detection in recent years. The use of Wasserstein generative adversarial networks (WGAN) for sensor data augmentation in a CA system was presented in this paper [15]. By adding new data to the training set for data augmentation, the WGAN enhances the performance of the CA. To learn and extract deep features from the sensor data, we developed a convolutional neural network and trained it using four classifiers: RF, OCSVM, DT, and KNN. The Equal Error Rate (EER) of the authentication system on the sensor data with a temporal window of 2 s was reported to be between 3.68% and 6.39% after evaluating the proposed system on the HMOG dataset.

For two-stage feature extraction, this paper [16] suggests a technique that combines manual construction with deep metric learning. In the first step, the time-series raw data of the 3 sensors were converted into 69 statistical features. To extract more features, the second stage involves fusing the statistical characteristics that were created from the three sensors into a 3-channel matrix and feeding it into a deep learning model. The proposed approach delivers a low equal error rate (EER) and good accuracy on BrainRun, hand movement, orientation, grasp, and movement datasets. Kensuke et al.[17] presented a deep learning-based model (LFP TCN) that collects local and global features to model sequential raw mobile data. This research recasts CA as an anomaly detection problem, using partial knowledge of accessible imposter profiles to enhance the inference robustness of unseen impostors. This paper proposes the generation of class-unconstrained impostor data for training using a random pattern-mixing augmentation method, making it possible to characterize different imposter patterns with little impostor data. Two public benchmark datasets, showing state-of-the-art performance in CA, were used to demonstrate the efficacy of the proposed method. Buddhacharya et al.[18] introduced CAGANet based on convolutional neural networks (CNN) that operate on smartphones. For data augmentation, the system used a “conditional Wasserstein generative adversarial network (CWGAN).” Smartphone sensors are employed to capture the motions of the phone resulting from user operation behaviors. CWGAN is utilized to generate add on sensor data that aids in the augmentation of existing data, which is then used to train the CNN that has been designed specifically for this purpose. The trained CNN was responsible for extracting deep features. Principal component analysis was further used to choose representative features for different classifiers. The enrolment phase involves the training of four one-class classifiers (EE, OC-SVM, isolation forest, and LOF). Once trained, these classifiers are used in the authentication phase to determine whether the current user is legitimate or not. On the 2-second sampling data, CAGANet with the isolation forest (IF) classifier attained the lowest EER of 3.64%. Kumar et al. [19] projected an enhanced CA method. It performs user identification using a machine learning model, and based on the current application running on the device, instantaneously selects the unimodal behavioral biometrics. Motion behavioral biometrics had an EER of 2.14% and touchscreen biometrics had an EER of 0.75% according to the results of the SVM-RBF classifier testing.

The study [20] covered a DeepAuthen architecture that uses deep learning classifiers to identify smartphone users based on their patterns of physical activity. The system combines deep metric learning techniques with manual construction in a two-stage feature-extraction process. In the first step, time-series raw data from multiple sensors are converted into statistical features. In feature extraction, the statistical characteristics constructed from the sensors are combined into a

channel matrix and input into a deep-learning model. The elliptic envelope technique was employed to classify users as either authentic or fraudulent. The proposed approach achieved low equal error rates and high accuracy when tested on two public datasets. This study emphasizes the significance of CA in overcoming the drawbacks of conventional one-time authentication techniques. The proposed approach improves user identification while safeguarding confidential data stored on devices. By presenting a novel method that combines feature fusion, deep metric learning, and manual construction, this research advances the subject of CA. This research endeavors to develop a behavioral profile for users by analyzing both their static and dynamic activities. Previous studies utilized the UCI-HAR dataset [21] to recognize human activities and apply continuous user authentication [22]. However, [22] employed a deep learning model that is computationally demanding for IoT devices. Consequently, this study introduces a new approach to authenticating users continuously based on their activities.

III. PROPOSED METHODOLOGY

The objective of this study is to develop Machine Learning Models and train them to provide Owner Authentication by Classifier prediction. The ML models were trained using a UCI-HAR dataset [21], which contained data collected using Inertial Sensors. The steps involved in developing the model:

Dataset Description

The UCI Human Activity Recognition (HAR) dataset is a popular set of sensor data that records people's movements while performing several types of physical activities. The dataset, which was gathered using gyroscopes and accelerometers on smartphones, included six distinct activities: sitting, standing, walking, walking downstairs, and walking upstairs. Thirty subjects participated in these activities, ranging in age from 19 to 48 years, while sporting a smartphone around their waists, providing the data. The triaxial angular velocity and acceleration signals were included in the dataset, which provided comprehensive movement data for the subjects. A detailed description of the HAR dataset is presented in **Table 1**.

Table 1. Dataset Description

Dataset Name	Year	# Rows	#columns	# Annotations	Source	Variations
UCI HAR Dataset	2021	7352	563	404K	Kaggle	Inertial Sensors Data

Data Processing

Data processing and cleaning are essential phases in the creation of a machine learning (ML) model. To ensure that the raw dataset was of high quality, consistent, and appropriate for training the model, it was thoroughly inspected and prepared during this phase. Addressing outliers, handling missing values, and finding and fixing any errors or inconsistencies in the data are all parts of data cleaning. This procedure frequently involves encoding categorical variables, scaling numerical features, and inputting missing values using appropriate methods. In the HAR dataset, the last column is modified by adding Boolean values (0 and 1) to support the classifier models, duplicate columns are removed, null values are handled, and the overall raw dataset is reshaped accordingly to support the goal. The quality of the input data has a significant impact on the effectiveness of a machine learning model; therefore, approximately 80 percent of the dataset is dedicated to training our classification model, and the remaining 20 percent is used for testing purposes.

Model Training

Logistic Regression Classifier Model (LR)

Logistic regression is a widely used statistical method in machine learning for binary classification tasks. A linear combination of input features was mapped to a probability score between 0 and 1 using a logistic function. The final binary prediction was then made based on the threshold of this probability. Since logistic regression is straightforward, effective, and easily interpreted, it is especially well suited for situations in which there is an assumed linear relationship between the features and the target variable. Mathematically, the logistic regression model can be written as follows:

- Given a set of input features $A=(x^1,x^2,...,x^n)$ and corresponding weights $\beta=(\beta_0,\beta_1,...,\beta_n)$, the logistic regression model calculates the log-odds of the probability of the binary outcome variable Y being in a particular class.
- The log-odds are transformed using the logistic function to obtain the probability of Y belonging to a certain class.
- The logistic function is defined as

$$P(B=1|A)=\frac{1}{1+e^{-(1+e\hat{}(-(\beta^0+\beta^1x^1+...+\beta^nx^n)))}} \quad (1)$$

where $P(B=1|A)$ is the probability of B being in class 1 given input features A.

Logistic regression is favored for its simplicity, interpretability, and effectiveness in many practical applications, despite its linear decision boundary.

Random Forest Classifier Model (RF)

The Random Forest classifier is a flexible and potent machine-learning model. It is a member of the ensemble learning family that combines several decision tree predictions to produce predictions that are more reliable and accurate. Each Random Forest decision tree receives training and develops its own predictions with a distinct subset of the training set. For classification tasks, the model uses a voting mechanism to aggregate these predictions; for regression tasks, averaging is used. The capacity of Random Forest to manage many features, recognize crucial features for prediction, and reduce overfitting is one of its main advantages. It also automatically yields a feature importance measure that aids in feature selection. This process introduces diversity among the trees and helps in reducing overfitting. The final prediction of a random forest model for a classification task can be expressed as:

$$\hat{y} = mode\{m_1(x), m_2(x), \ldots, m_n(x)\} \tag{2}$$

where $\hat{y}$ is the predicted class, x is the input feature vector, $m_i(x)$ represents the prediction of the i-th decision tree, and n is the total number of trees in the forest.

Random forests are known for their high accuracy, robustness to overfitting, and capability to handle large datasets with higher dimensionality. Additionally, they provide insights into feature importance, which can be valuable for understanding the underlying data.

Sequential Neural Network for Binary Classification (SNN)

An activation function is added to the output layer (more precisely, a sigmoid activation function). With a sigmoid function, the output is reduced to a probability value between 0.0 and 1.0. By switching to binary cross-entropy, a loss function is designed specifically for binary classifiers. This model is characterized by a series of layers where each layer consists of neurons that perform linear transformations followed by nonlinear activation functions. As a result, we set the metrics to accuracy such that the history object that fits returns includes the accuracies calculated by the loss function. The structure of a SNN can be mathematically described as follows:

$$a^{(l)} = \sigma\left(W^{(l)}a^{(l-1)} + b^{(l)}\right) \tag{3}$$

where the bias vector for layer l is $b(l)$, the activation vector of layer l is represented by $a(l)$, the activation function (such as ReLU, sigmoid, or tanh) is denoted by σ, and the weight matrix connecting layer l to layer l-1 is represented by $W(l)$. The ultimate prediction is provided by the network's output, $a(L)$, which comes after L layers and is given into the network as a(0). This architecture allows the network to learn hierarchical representations of the input data, capturing complex patterns and relationships. Sequential neural networks are highly versatile and can be adapted to various problem domains by adjusting the number of layers, the number of neurons per layer, and the type of activation functions used.

Authentication Model

In this model, the user logs into using static authentication. The user activities are identified, and the authentication module authenticates the user throughout the session with the help of a trained model. If the user is a signed user, they will remain logged into the system. Otherwise, the system alerts the user and locks on the smartphone. This feedback was provided to the ML model to improve future authentication decisions.

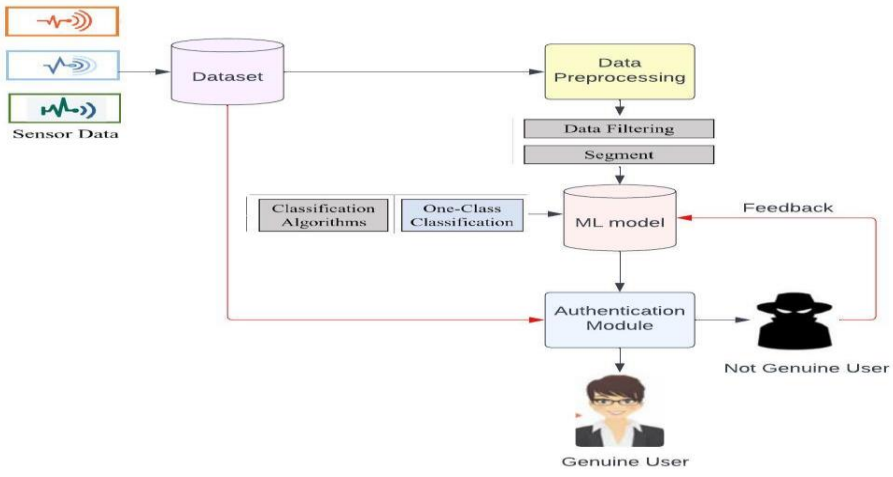

Fig 1. Proposed Methodology.

Fig 1 shows the overall methodology of the proposed system. After the user registration phase, if a registered user wants to access the system, the user is prompted for static authentication. After successful static authentication, CA will be invoked and run until the current session is completed. If any change in the behavior of the registered user is identified, the system logs the user out.

Algorithm 1 User Registration and Session Management

Input : Static user authentication credentials and behavioral profile
Output : Authentication status (Authenticated/Not Authenticated)

After user registration phase:
If registered user wants to access the system:
Prompt for static authentication (e.g., password or biometric)
If static authentication is successful:
While session is active:
Invoke Continuous Authentication (CA)
If change in behavior is detected:
Log out the user
Terminate session
Feedback is given to ML models for future predictions

User Registration and Session Management algorithm depicted as algorithm 1, is critical for initiating and maintaining the continuous authentication system. After the initial user registration phase, which involves collecting and storing the user's biometric and behavioral data, any registered user attempting to access the system is first prompted for static authentication, such as a password or biometric verification. Upon successful completion of this static authentication, the continuous authentication (CA) system is activated and will operate for the duration of the user's session. During this active session, the system continuously monitors the user's behavior using the previously described machine learning models to ensure consistent authentication. If the system detects any significant deviation from the expected user behavior, it assumes a potential security risk and immediately logs out the user, thereby terminating the session. This dual-layer approach enhances security by combining static initial authentication with ongoing behavioral verification, ensuring that only legitimate users maintain access to the system throughout their session. Algorithm 2 provides a detailed framework for the proposed continuous authentication system.

Algorithm 2 Continuous Authentication System Using Machine Learning Models

Input: Behavioral profile of registered user
Output: Authentication status (Authenticated/Not Authenticated)
1. Initialize:
- Define time window size (T)
- Set authentication threshold (τ)
- Load pre-trained ML models (logistic regression, random forest, sequential neural network)

2. Preprocessing

Function preprocess_data(buffer):
Normalize data
Handle missing values and outliers
Modify the last column of the HAR dataset by adding Boolean values (0 and 1) to support classifier models
Segment the processed dataset into training and testing sets with an 80-20 splits.
Return segmented data

3. Model Selection and Prediction:
Function predict_authentication(feature_list):
Initialize predictions
For each model in [logistic_regression, random_forest, sequential_neural_network]:
Predict probability of authentication for each feature set in feature_list
Append model predictions to predictions
Return average of predictions

4. Decision Making:

Function authenticate_user(predictions, threshold):

For each prediction in predictions:
If prediction < threshold:
Trigger authentication challenge
Return "Not Authenticated"
Return "Authenticated"

5. Continuous Authentication:

While system is running:
If buffer contains data from dataset for at least one time window:
segmented_data = preprocess_data(buffer)
feature_list = extract_features(segmented_data)
predictions = predict_authentication(feature_list)
status = authenticate_user(predictions, τ)
Output status
Clear processed data from buffer

6. Adaptation and Learning:

Periodically throughout the session:
Collect new labelled data
Retrain models with updated data
Validate and update models to improve accuracy

IV. EXPERIMENTAL ANALYSIS & RESULTS

After processing and feature extraction, the dataset was split for testing and training purposes; approximately 80%of the data were fed to the model as training data and 20% for testing purposes. After training, the Accuracy Score for both models was calculated and compared to the prediction efficiency; the owner was greater in the Random Forest Classifier than the Logistic Regression. After training and testing, The Accuracy Score for both models was calculated, and it was observed that the accuracy of the SNN was better than that of the other two models.

By plotting recall on the y-axis and precision on the x-axis, the recall–precision curve is usually utilized for different threshold values used in classification choices [23-24]. Every point on the curve was associated with a certain threshold. Better performance is shown by a curve that is closer to the top-right corner when recall and precision are both high. The Recall Precision curves for logistic regression and random forest are shown in **Fig 2**. It is evident that the proposed methodology places the curve close to the upper right corner.

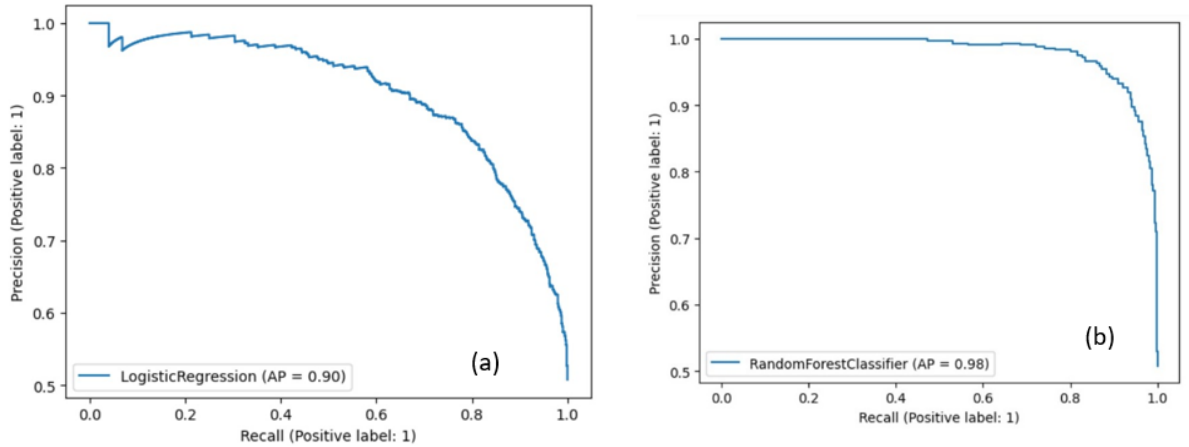

Fig. 2. Precision-Recall Trade-off Analysis: A Comprehensive Evaluation, (a) – Logistic Regression, (b) – Random Forest.

A model validation method called cross-validation [25] is used to evaluate the generalizability of the statistical findings to other datasets. Practically, it aids in estimating the accuracy of a prediction model. K-fold cross-validation is a popular method that divides data into K equal-sized portions [26]. Bias in the data may be removed by performing a k-fold=5 cross validation. K was spent repeating the testing and training procedures. The cross-validation values for random forest and logistic regression were 91% and 80%, respectively. The concept of scalability and methodology for attaining it are generic and adaptable to various CA systems. A good cross-validation score shows that the model is independent of a dataset and can be applied to other types of behavioral datasets.

Additional performance metrics, such as the area under the receiver operating characteristic (ROC) curve [27], are computed to provide a comprehensive evaluation of the model performance. Logistic regression, SNN, and RF accounted for 82%, 93%, and 93%, respectively.

Fig 3 shows a comparative analysis of multiple ML models. The performance metrics used to evaluate the models were accuracy, precision, recall, and F1 score [28-30]. **Table 2** lists the overall accuracy of the proposed system. The SNN showed better results among the three models used. Sequential Neural Networks can be trained in an end-to-end fashion, allowing them to directly optimize the entire model for the task at hand. In contrast, Random Forest and Logistic Regression may require additional preprocessing or feature selection steps, which could introduce information loss or suboptimal performance.

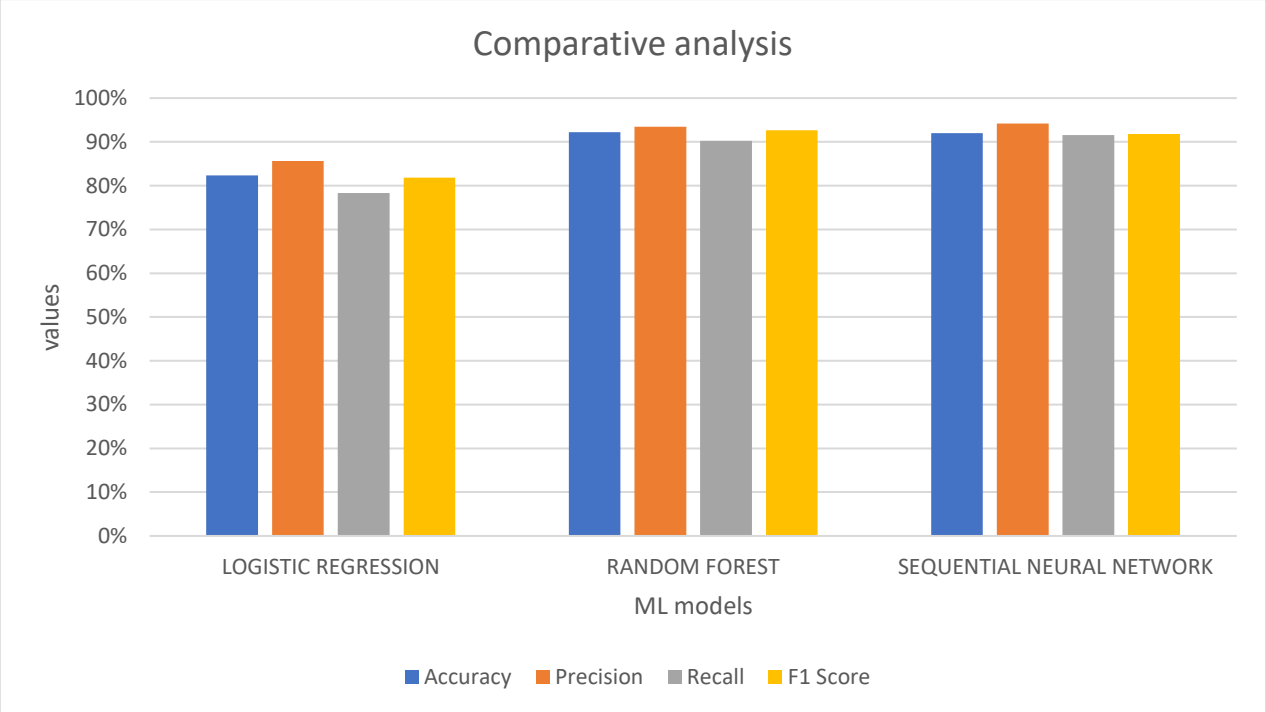

Fig. 3. Comparative analysis algorithm with different model.

Table 2. Performance metrics were obtained using different machine-learning models.

Performance Metric	Logistic Regression	Random Forest	Sequential Neural Network
Accuracy	82.32%	92.18%	92%
Precision	85.65%	93.48%	94.21%
Recall	78.31%	90.22%	91.56%
F1 Score	81.81%	92.62%	91.64%

False Acceptance Rate (FAR)

The implication of FAR for security occurs when a user's FAR is high, which indicates that the system has mistakenly accepted them as authenticated. This opens up critical resources and data for unauthorized access, which is a serious security concern. Possible Repercussions include unauthorized access that may result in financial fraud, data breaches, or other nefarious acts that jeopardize user confidence and system integrity.

False Rejection Rate (FRR)

Consequences for Usability An authorized user may be mistakenly rejected by the system, which can cause annoyance, irritation, and perhaps lower user adoption. This was indicated by the high FRR values. Impact on User Experience: Frequently receiving erroneous rejections might worsen the user experience by encouraging more attempts at login, lowering productivity, or encouraging users to give up on the authentication process. **Table 3** shows that the given yields good results for these parameters.

The point where the False Acceptance Rate (FAR) equals the False Rejection Rate (FRR) indicates the threshold where authentication errors balance. Lower EER values indicate better performance; therefore, our logistic regression and Random Forest models had lower EER of value 0.20 and 0.08 respectively.

These measures provide important information on how well biometric authentication systems work [30-35]. These help to design and test systems to understand how security and ease of use can be balanced. **Table 3** and **Fig 4** shows the performance of the authentication system based on the behavioral biometrics. Here, logistic regression shows better values.

Table 3. Analysis of Authentication System Based on Behavioral Biometric Metrics

Accuracy Rate	Logistic Regression	Random Forest
EER	0.20	0.08
FAR	0.13	0.063
FRR	0.21	0.08

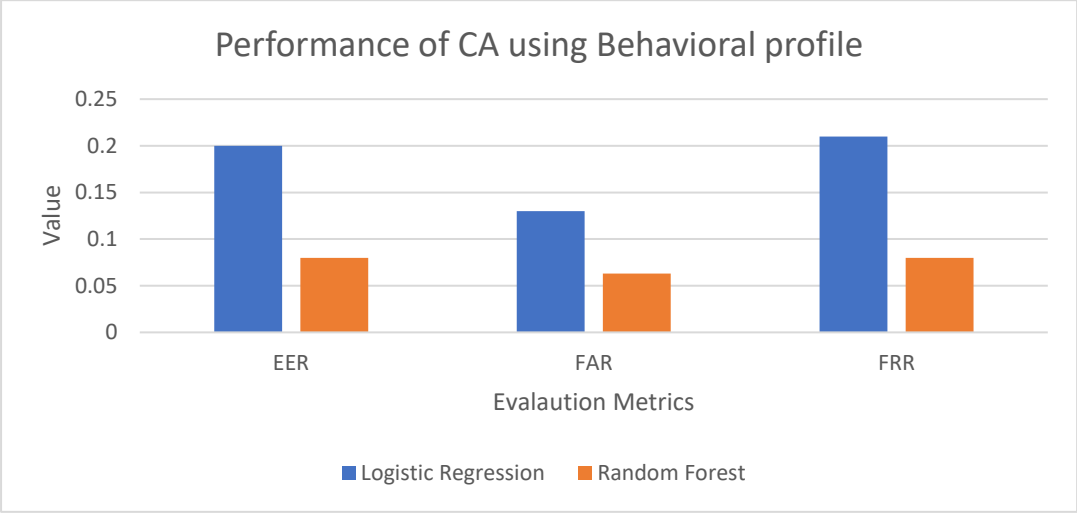

Fig 4. Performance of Continuous Authentication System.

In [21], the author used the HAR dataset to identify activity and then used it for user classification using CNN, LSTM, CNN-LSTM, and ConvLSTM deep learning (DL) models. For the USC HAD dataset, the ConvLSTM model achieved the maximum average accuracy of 87.178% for dynamic activities, whereas the CNN-LSTM model attained the highest average accuracy of 78.698% for static activities. However, when focusing solely on the top three activities—walking forward, walking left, and walking right—the CNN-LSTM model achieved the highest mean accuracy of 95.858%. Even though the accuracy of CNN LSTM is higher, this work uses deep learning models, which may not be suitable for IoT because DL is computationally intensive. **Table 4** shows a comparison using various performance metrics, and it is evident that the proposed method performs better and is suitable for CA for smart gadgets using ML models.

Table 4. Comparison with Existing Work

	Accuracy	Precision	Recall	F1-Score
Previous work [21]	91.776	91.10	89.20	90.10
Proposed work	92	94.21	91.56	91.64

V. PRACTICAL IMPLEMENTATION AND REAL-WORLD DEPLOYMENT CHALLENGES

Procedures for getting explicit user consent should be put in place before collecting and processing behavioral data in order to allay worries about privacy infringement. Enforcing robust encryption and security methods is crucial to protect user data during transmission and storage on backend servers and handsets. It could be challenging to put in place effective real-time data processing pipelines that can manage numerous users' simultaneous authentication requests and constant streams of smartphone data. Processing of this kind ought to be done on the cloud. Minimizing network bandwidth usage by compressing data transmissions and prioritizing critical authentication updates to conserve resources, especially in bandwidth-constrained environments should be done. It is essential to optimize the utilization of resources on smartphones, like as CPU, memory, and battery, in order to mitigate the effects of ongoing data gathering and model inference on device performance. Being transparent about the types of data collected from the smartphone, such as location information, device usage patterns, and biometric data, and how this data is utilized to establish the user's behavioral profile. To explanations or insights into the factors considered by the machine learning model when making authentication decisions. This could include highlighting specific behavioral patterns or anomalies that contribute to the decision. Behavioral patterns may change over time due to various factors such as cultural shifts, technological advancements, or socioeconomic changes. If the training data does not adequately represent these temporal variations, the model may be biased towards outdated patterns. These are not handled by the study. But, this will be incorporated in the future implementation of the work.

VI. CONCLUSION

In this study, there are two different machine learning classifier models trained on a UCI-HAR dataset containing behavioral biometric data collected through inertial sensors for classifying whether the device is used by the owner or not, and the results can be used to take appropriate actions such as the locking device and backlisting the user based on the result obtained by a prediction model. Because computational resources may be constrained in some devices, this paper proposes using Classifier Models over CNN and Deep Learning Models. Among the three machine models tested, the sequential neural network exhibited the best results. SNN are well suited for CA models based on human activities. This research can provide Real-Time Authentication to increase the security of electronic gadgets and protect them against ever-evolving security threats. CA can authenticate whether it is an authenticated user, and it can be improved to identify multiple registered users.

Data Availability

No data was used to support this study.

Conflicts of Interests

The author(s) declare(s) that they have no conflicts of interest.

Funding

No funding agency is associated with this research.

Competing Interests

There are no competing interests.

References

[1]. J. Dybczak and P. Nawrocki, “Continuous authentication on mobile devices using behavioral biometrics,” 2022 22nd IEEE International Symposium on Cluster, Cloud and Internet Computing (CCGrid), May 2022, doi: 10.1109/ccgrid54584.2022.00125.
[2]. Z. Shen, S. Li, X. Zhao, and J. Zou, “IncreAuth: Incremental-Learning-Based Behavioral Biometric Authentication on Smartphones,” IEEE Internet of Things Journal, vol. 11, no. 1, pp. 1589–1603, Jan. 2024, doi: 10.1109/jiot.2023.3289935.
[3]. Y. Liang, K. Feng, and Z. Ren, “Human Activity Recognition Based on Transformer via Smart-phone Sensors,” 2023 IEEE 3rd International Conference on Computer Communication and Artificial Intelligence (CCAI), May 2023, doi: 10.1109/ccai57533.2023.10201297.
[4]. S. Aboukadri, A. Ouaddah, and A. Mezrioui, “Machine learning in identity and access management systems: Survey and deep dive,” Computers & Security, vol. 139, p. 103729, Apr. 2024, doi: 10.1016/j.cose.2024.103729.
[5]. K. Istiaque Ahmed, M. Tahir, M. Hadi Habaebi, S. Lun Lau, and A. Ahad, “Machine Learning for Authentication and Authorization in IoT: Taxonomy, Challenges and Future Research Direction,” Sensors, vol. 21, no. 15, p. 5122, Jul. 2021, doi: 10.3390/s21155122.
[6]. J. Lin and M. E. Latoschik, “Digital body, identity and privacy in social virtual reality: A systematic review,” Frontiers in Virtual Reality, vol. 3, Nov. 2022, doi: 10.3389/frvir.2022.974652.
[7]. X. Wang, Y. Shi, K. Zheng, Y. Zhang, W. Hong, and S. Cao, “User Authentication Method Based on Keystroke Dynamics and Mouse Dynamics with Scene-Irrelated Features in Hybrid Scenes,” Sensors, vol. 22, no. 17, p. 6627, Sep. 2022, doi: 10.3390/s22176627.
[8]. A. F. Baig, S. Eskeland, and B. Yang, “Privacy-preserving continuous authentication using behavioral biometrics,” International Journal of Information Security, vol. 22, no. 6, pp. 1833–1847, Jul. 2023, doi: 10.1007/s10207-023-00721-y.
[9]. D. R. Bhuva and S. Kumar, “A novel continuous authentication method using biometrics for IOT devices,” Internet of Things, vol. 24, p. 100927, Dec. 2023, doi: 10.1016/j.iot.2023.100927.
[10]. Z. A. Zukarnain, A. Muneer, and M. K. Ab Aziz, “Authentication Securing Methods for Mobile Identity: Issues, Solutions and Challenges,” Symmetry, vol. 14, no. 4, p. 821, Apr. 2022, doi: 10.3390/sym14040821.
[11]. S. Kokal, M. Vanamala, and R. Dave, “Deep Learning and Machine Learning, Better Together Than Apart: A Review on Biometrics Mobile Authentication,” Journal of Cybersecurity and Privacy, vol. 3, no. 2, pp. 227–258, Jun. 2023, doi: 10.3390/jcp3020013.
[12]. P. William, G. R. Lanke, D. Bordoloi, A. Shrivastava, A. P. Srivastavaa, and S. V. Deshmukh, “Assessment of Human Activity Recognition based on Impact of Feature Extraction Prediction Accuracy,” 2023 4th International Conference on Intelligent Engineering and Management (ICIEM), May 2023, doi: 10.1109/iciem59379.2023.10166247.
[13]. P. Ganesh, P. Jagadeesh, and J. S. Raj. J, “Prediction of Human Activity Recognition Using Convolution Neural Network Algorithm in Comparison with Grid Search Algorithm,” 2023 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI), May 2023, doi: 10.1109/accai58221.2023.10200427.
[14]. S. Mekruksavanich, P. Jantawong, and A. Jitpattanakul, “Comparative Analysis of CNN-based Deep Learning Approaches on Complex Activity Recognition,” 2022 Joint International Conference on Digital Arts, Media and Technology with ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunications Engineering (ECTI DAMT & NCON), Jan. 2022, doi: 10.1109/ectidamtncon53731.2022.9720320.
[15]. S. Zou, H. Sun, G. Xu, C. Wang, X. Zhang, and R. Quan, “A Robust Continuous Authentication System Using Smartphone Sensors and Wasserstein Generative Adversarial Networks,” Security and Communication Networks, vol. 2023, pp. 1–11, Apr. 2023, doi: 10.1155/2023/3673113.
[16]. M. Hu, K. Zhang, R. You, and B. Tu, “Multisensor-Based Continuous Authentication of Smartphone Users With Two-Stage Feature Extraction,” IEEE Internet of Things Journal, vol. 10, no. 6, pp. 4708–4724, Mar. 2023, doi: 10.1109/jiot.2022.3219135.
[17]. K. Wagata and A. B. J. Teoh, “Few-Shot Continuous Authentication for Mobile-Based Biometrics,” Applied Sciences, vol. 12, no. 20, p. 10365, Oct. 2022, doi: 10.3390/app122010365.
[18]. S. Buddhacharya and N. Awale, “CNN-BASED CONTINUOUS AUTHENTICATION OF SMARTPHONES USING MOBILE SENSORS,” International Journal of Innovative Research in Advanced Engineering, vol. 9, no. 8, pp. 361–369, Aug. 2022, doi: 10.26562/ijirae.2022.v0908.37.
[19]. P. K. Rayani and S. Changder, “Enhanced Unimodal Continuous Authentication Architecture on Smartphones for User Identification through Behavioral Biometrics,” 2023 2nd International Conference on Vision Towards Emerging Trends in Communication and Networking Technologies (ViTECoN), May 2023, doi: 10.1109/vitecon58111.2023.10157803.

[20]. S. Mekruksavanich and A. Jitpattanakul, “Deep Learning Approaches for Continuous Authentication Based on Activity Patterns Using Mobile Sensing,” Sensors, vol. 21, no. 22, p. 7519, Nov. 2021, doi: 10.3390/s21227519.
[21]. A. H and A. R, “Artificial Intelligence and Machine Learning for Enterprise Management,” 2019 International Conference on Smart Systems and Inventive Technology (ICSSIT), Nov. 2019, doi: 10.1109/icssit46314.2019.8987964.
[22]. Davide Anguita, Alessandro Ghio, Luca Oneto, Xavier Parra and Jorge L. Reyes-Ortiz. “A Public Domain Dataset for Human Activity Recognition Using Smartphones”,21th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning, ESANN 2013. Bruges, Belgium 24-26 April 2013.
[23]. L. de-Marcos, J.-J. Martínez-Herráiz, J. Junquera-Sánchez, C. Cilleruelo, and C. Pages-Arévalo, “Comparing Machine Learning Classifiers for Continuous Authentication on Mobile Devices by Keystroke Dynamics,” Electronics, vol. 10, no. 14, p. 1622, Jul. 2021, doi: 10.3390/electronics10141622.
[24]. B. Pelto, M. Vanamala, and R. Dave, “Your Identity is Your Behavior - Continuous User Authentication based on Machine Learning and Touch Dynamics,” 2023 3rd International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME), Jul. 2023, doi: 10.1109/iceccme57830.2023.10252828.
[25]. E. Klieme and C. Meinel, “All about that BASE: Modeling Biometric Authentication Systems and their Evaluations to enable a more efficient Exchange of Research Results,” 2023 IEEE International Joint Conference on Biometrics (IJCB), Sep. 2023, doi: 10.1109/ijcb57857.2023.10448589.
[26]. S. Zou, H. Sun, G. Xu, C. Wang, X. Zhang, and R. Quan, “A Robust Continuous Authentication System Using Smartphone Sensors and Wasserstein Generative Adversarial Networks,” Security and Communication Networks, vol. 2023, pp. 1–11, Apr. 2023, doi: 10.1155/2023/3673113.
[27]. K. Gorur, “Fourier Synchrosqueezing Transform-ICA-EMD Framework Based EOG-Biometric Sustainable and Continuous Authentication via Voluntary Eye Blinking Activities,” Biomimetics, vol. 8, no. 4, p. 378, Aug. 2023, doi: 10.3390/biomimetics8040378.
[28]. P. K. Rayani and S. Changder, “Continuous user authentication on smartphone via behavioral biometrics: a survey,” Multimedia Tools and Applications, vol. 82, no. 2, pp. 1633–1667, Jun. 2022, doi: 10.1007/s11042-022-13245-9.
[29]. M. Hu, K. Zhang, R. You, and B. Tu, “AuthConFormer: Sensor-based Continuous Authentication of Smartphone Users Using A Convolutional Transformer,” Computers & Security, vol. 127, p. 103122, Apr. 2023, doi: 10.1016/j.cose.2023.103122.
[30]. U. Uslu, Ö. D. İncel, and G. I. Alptekin, “Evaluation of Deep Learning Models for Continuous Authentication Using Behavioral Biometrics,” Procedia Computer Science, vol. 225, pp. 1272–1281, 2023, doi: 10.1016/j.procs.2023.10.115.
[31]. Hyung-dong Lee, KiHyo Nam, Heewoong Lee, and Mun-Kweon Jeong, “PC-based User Continuous Authentication System Using the User’s Finger Stroke Characteristics,” Research Briefs on Information and Communication Technology Evolution, vol. 9, pp. 160–177, Nov. 2023, doi: 10.56801/rebicte.v9i.173.
[32]. P. Bansal and A. Ouda, “Continuous Authentication in the Digital Age: An Analysis of Reinforcement Learning and Behavioral Biometrics,” Jun. 2023, doi: 10.20944/preprints202306.1005.v1.
[33]. S. B. Kulkarni and S. Kulkarni, “Study of the Value of π Probability Sampling by Testing Hypothesis and Experimentally,” Journal of Computers, Mechanical and Management, vol. 3, no. 1, pp. 22–29, Feb. 2024, doi: 10.57159/gadl.jcmm.3.1.240101.
[34]. N. Ranjan, “Enhancing Voting Security and Efficiency,” Journal of Computers, Mechanical and Management, vol. 2, no. 3, pp. 9–15, Aug. 2023, doi: 10.57159/gadl.jcmm.2.3.23065.
[35]. S. Kumar, U. Gupta, A. K. Singh, and A. K. Singh, “Artificial Intelligence,” Journal of Computers, Mechanical and Management, vol. 2, no. 3, pp. 31–42, Aug. 2023, doi: 10.57159/gadl.jcmm.2.3.23064.