Journal of Information Technology Management
ISSN-OnLine: 2980-7972
University of Tehran

HybridTouch: A Robust Framework for Continuous User Authentication by GAN-Augmented Behavioral Biometrics on Mobile Devices

Mahendra Kumar Jangir

Jawaharlal Nehru University, New Delhi, India. E-mail: Jangirnawalgarh@gmail.com

Karan Singh *

*Corresponding author, Jawaharlal Nehru University, New Delhi, India. E-mail: Karan@jnu.ac.in

Tayyab Khan

Assistant Prof., Department of Computer Science and Engineering, Indian Institute of Information Technology, Sonipat, IIIT Sonepat, India. E-mail: tayyabkhan.cse2012@gmail.com

Journal of Information Technology Management, 2025, Vol. 17, Special Issue, pp.108-129.
Published by the University of Tehran, College of Management
doi: https://doi.org/10.22059/jitm.2025.102924
Article Type: Research Paper
© Authors

Received: January 17, 2025
Received in revised form: March 03, 2025
Accepted: June 13, 2025
Published online: August 01, 2025

Abstract

With an increasing reliance on mobile devices, continuous and assured user authentication is essential to protect sensitive personal data and digital interactions from unwanted access. Based on this background, this research proposed the development of the HybridTouch framework for smartphone-based continuous and passive user authentication. The proposed HybridTouch combines Convolutional Neural Networks for spatial feature extraction and Gated Recurrent Units for temporal sequence analysis. It uses accelerometer, gyroscope, and touch data to take advantage of the unique behavioral patterns captured by it. Innovative preprocessing techniques have been incorporated into the proposed approach: Discrete Wavelet Transform is used for signal denoising, and Variable-Length Adaptive Temporal windowing is used for segmentation based on signal entropy to enhance feature representation. To eliminate the data scarcity limitation, Generative Adversarial Networks were used to synthesize realistic behavioral data that considerably augmented the dataset and enhanced model generalization capability. Extensive experiments conducted on the Hand Movement, Orientation, and Grasp (HMOG) dataset showed that the proposed HybridTouch achieved excellent results with authentication accuracy up to 98.8% with real data, growing up to 99% with GAN-augmented data. The hybrid model further has an equal error rate of 1.4% on real data and 1.25% on synthetic data, which is better than any other models currently present (Sağbas et al., 2024; Siddiqui et al., 2022; Abuhamad et al., 2020) and all implementations of standalone convolutional neural networks and gated recurrent units.

Keywords: Mobile Authentication, Touch Dynamics, Deep Learning, Smartphone Sensors, Convolutional Neural Networks.

Introduction

The rapid development of mobile devices and the growing reliance on smartphones as a means to perform almost all sorts of daily activities pose significant challenges with regard to security and data privacy. Mobile devices, such as smartphones, constitute the heart of modern existence, with 5 billion users projected for 2024 and a predicted 6 billion by 2027 (Nayak et al., 2016). With the proliferation of such devices, fears about data safety, especially data privacy, are on the rise. Since smartphones store and process substantial amounts of personal data, secure and reliable authentication methods are the need of the hour to guard against unauthorized access and maintain individual privacy (Agrawal et al., 2023). Traditionally, knowledge-based methods, which include PIN codes, passwords, and patterns, as well as static biometric methods such as fingerprints and facial recognition, have been used for authenticating mobile devices (Zhao et al., 2024). These methods are vulnerable to different types of vulnerabilities. Knowledge-based authentication is vulnerable to attacks like brute force, shoulder surfing, and smudge attacks (Nayak et al., 2016). The biometric methods that have static characteristics, though enhanced security, provide scope for spoofing and also privacy issues while collecting sensitive data.

To mitigate these disadvantages, continuous authentication techniques have been further developed by emphasizing behavioral biometrics and sensor-based authentication (Tran et al., 2020). In recent years, continuous user authentication using mobile sensors has garnered significant interest. This approach provides another layer of security (Mangal et al., 2023) through continuous verification of the user's identity, based on the behaviors of touching, gait, and patterns of device usage. The utilization of motion sensors-including accelerometers, gyroscopes, and pressure sensors-provides continuous tracking of user activity with a very high level of secure and nonintrusive authentication. However, despite the above advantages, current approaches (Anguita et al., 2013; Abuhamad et al., 2020) are still plagued by several challenges, including feature extraction from sensor data, the presence of noise in real-world environments, and the limited availability of large, labelled datasets for model training (Zou et al., 2020; Centeno et al., 2018). Since deep learning (DL) has shown its utility in so many applications, many DL approaches have been introduced to continuous authentication studies (Tran et al., 2020; Zou et al., 2020). Although our approach is generic to the development of continuous authentication, in the proposed framework called HybridTouch, touch dynamics are exploited, and CNN for feature extraction and GRU for sequential data analysis are used to leverage both spatial as well as temporal information from the sensor data for robust authentication (Hajiakhoondi et al., 2013; Alfaleh et al., 2024)

Scientific Contributions

Ensuring secure and continuous user authentication on mobile devices is a pressing challenge because of the growing reliance on smartphones for sensitive digital interactions. Traditional authentication methods (Nayak et al., 2016; Zhao et al., 2024; Lu et al., 2018; Amini et al., 2018), such as passwords or biometric verification, often fall short in providing seamless and persistent security. These methods are either prone to security vulnerabilities (Abuhamad et al., 2020) or intrusive to the user experience (Ehatisham et al., 2017). With these demands, high-level solutions that use behavioral biometrics for delivering passive, continuous, and user-friendly authentication have gained more prominence (Shoaib et al., 2013; Shoaib et al., 2014). By leveraging unique patterns of user interactions, we explore promising avenues that utilize data augmentation techniques along with deep learning (DL) to address the above-mentioned challenges. Novel key scientific contributions include the following:

a) Developing a hybrid DL model combining CNN for feature extraction and GRU for sequential data analysis to enhance continuous authentication accuracy.
b) Integrating GANs to generate synthetic user data, expand the training dataset, and improve the model's generalization capabilities.
c) Conducting extensive experiments on the HMOG dataset, validating that the proposed model outperforms standalone CNN, GRU, and prior models in terms of accuracy and robustness.
d) Performing a comprehensive performance evaluation of the model using the metrics of EER, accuracy, precision, and recall.

The paper contributes to the domain of mobile authentication using DL and data augmentation that offers a highly secure, efficient, and realistic possibility for continuous user authentication.

This research introduces HybridTouch, a novel continuous authentication framework that leverages touch dynamics data and fuses sophisticated DL with smartphone device sensor data to enhance the accuracy and robustness of authentication. The proposed framework will be able to identify and recognize characteristic user interaction behaviors on mobile devices using powerful preprocessing and modeling algorithms to give a strong identification of the users. Figure 1 shows the architecture of the proposed HybridTouch for continuous authentication. Data were collected only from the HMOG dataset (Yang et al., 2014), including static and dynamic user activities. (Table 1) shows the Summary of HMOG dataset characteristics.

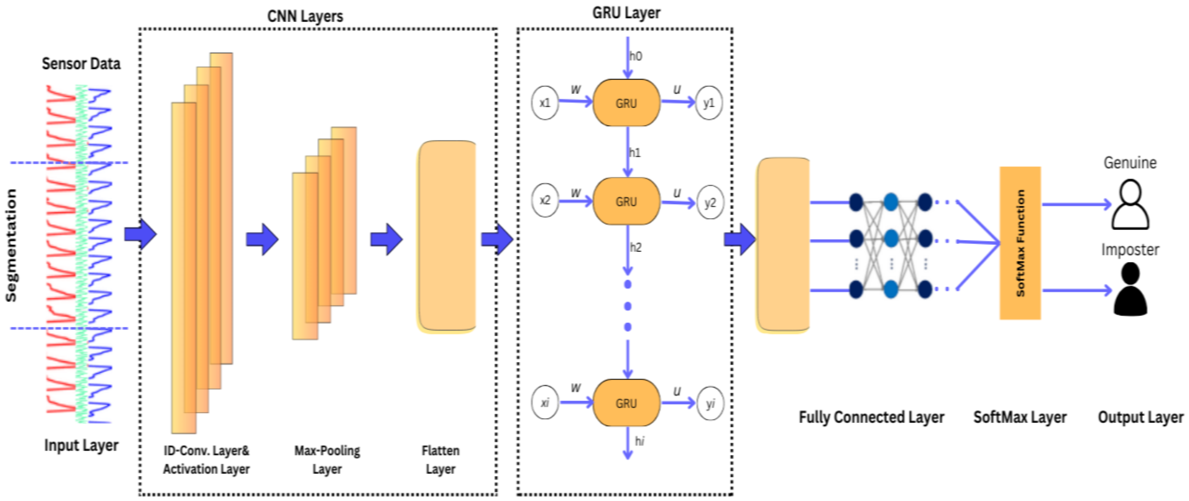

Figure 1. Proposed Hybrid Touch architecture for continuous authentication

The preprocessing stage applies DWT for noise removal and VATW for segmentation. This guarantees an optimal trade-off between feature extraction and noise reduction. To increase the user dataset from 41 to 180, a GAN is used to expand the user dataset, and the proposed model is robustly evaluated using 180 users. This augmentation improves the generalization ability of a model and explains differences in the behaviour of users. The suggested model integrates CNN for feature extraction and GRU for the temporal sequence analysis. The model utilizes raw sensor data from accelerometers, gyroscopes, and magnetometers, translating it into a high-dimensional embedding space for efficient learning. A 10-fold cross-validation is used to ensure robust performance evaluation. For smartphone-based continuous authentication, several public datasets are available for human activity recognition (HAR). However, most of these datasets are limited by their size or the controlled conditions under which they are collected, such as fixed-mounted smartphones, which restrict their utility for dynamic, real-world applications. For example, while the UCI-HAR dataset (Anguita et al., 2013) and WISDM-HARB dataset have been widely used for activity recognition and authentication research, their lack of subject-level behavioral information poses challenges for robust continuous authentication tasks. To overcome these limitations, the Hand Movement, Orientation, and Grasp (HMOG) dataset was chosen as the primary source for the proposed framework. In this publicly available dataset (Yang et al., 2014), 100 participants conducted activities over 24 sessions that included inertial sensors, such as an accelerometer, gyroscope, and magnetometer, as well as touch sensor data from performing different tasks, like reading, writing, or navigating maps. Activities were performed with the user sitting or walking, thus capturing extensive user behavior over various conditions. The design of this dataset enables continuous user authentication through analysis of fine-grained interaction patterns. Key categories for recorded data consist of accelerometer and gyroscope as well as magnetometer readings, together with touch events, for instance, taps, scrolls, or keypresses. Such extracted features support behavior interpretation in real-time through aspects such as grasp resistance-pressure during touches, and stability (i.e., how touch interactions stabilize within reasonable time intervals), and the dataset's utility does feature some infrequent irregularities or missing values in some entries.

Table 1. Summary of HMOG dataset characteristics

Category	Content
Accelerometer	Timestamp, acceleration force along x/y/z axes
Subjects	53 (males), 47 (females)
Number of Examples/Subject (Class)	in the range of 1162 to 4639

Data Preprocessing

The preprocessing phase in the proposed model is integral to the preparation of raw touch data from the HMOG dataset for use in continuous user authentication. Noise reduction, normalization, and segmentation are performed, all of which significantly contribute to the enhancement of signal quality, and the data is in a state ready for feature extraction and classification by the model.

After denoising, the data undergoes Min-Max normalization, which rescales all feature values to the range [0,1]. Mathematically, the normalization of a feature x is shown by Eq. (1) as follows.

$$x_{\mathrm{scaled}} = \frac{x - x_{\min}}{x_{\max} - x_{min}} \quad (1)$$

Here, min(x) and max(x) denote the minimum and maximum of the associated feature x, respectively. This transformation ensures that all features are considered equally in the learning process, and so features with larger magnitudes are not able to determine the learning behaviour of the model. Normalization is particularly useful when sensor data from heterogeneous sources with different ranges of values is considered.

Data segmentation is the last step of preprocessing, which is used to divide continuous raw sensor data into temporal segments to facilitate effective analysis and feature extraction. We use VATW, which is a dynamic approach to control the scaling of each segmentation window according to the amount of within-sensor variability. In this approach, one is provided with greater flexibility concerning fixed-length windows, which, if not done carefully, can fail to capture relevant data or include irrelevant data, so that temporal aspects are precisely represented. Mathematically, the signal x(t) is segmented into overlapping temporal windows w_i[t1, t2], where t1 and t2 denote the start and end points of the window. The window size$|w_i|$ is adaptively chosen according to the entropy H (w_i) of the signal contained within the window, i.e., the data variability. The entropy is shown in Eq. (2) as follows

$$H(W_i) = -\sum_{j=1}^{N} p_j \log p_j \quad (2)$$

Where N is the number of unique signal patterns in the window and p_j is the probability of occurrence of the j-th signal pattern in w_i. Window size grows when state entropy $H(w_i)$ is high, i.e., in regions of high variability (e.g., fast gestures), and shrinks when entropy is low, i.e., stable or noninformative intervals (e.g., steady hand position). This approach is inspired by the adaptive techniques used in gesture recognition studies, where the variability of human behavior is a key factor in selecting appropriate segmentation lengths (Kokal et al., 2023; Mekruksavanich et al., 2021). The adaptive property of VATW enables the system to increase the proportion of time spent analysing high-variance segments and to pick out transient events (e.g., taps, swipes, or short movements) whilst decreasing sampling rate during periods of more stable, repetitive touch patterns. Formally, the segmentation can be represented in Eq. (3) and Eq. (4) as follows:

$$W_i = [t_i, t_j + \Delta t] \tag{3}$$

$$W_{i+1} = t_i + \Delta t/2, t_j + 3\Delta t/2 \tag{4}$$

where Δt is the length of the time window, and the overlap between consecutive windows is Δt/2 in overlapping temporal windows w_i and w_{i+1}. Because of this combined structure, each time point is inside two successive windows, keeping temporal continuity and preserving inter-window dependence over the data. The data are effectively segmented by VATW in an adaptive way; the window length is dynamically determined based on the variability of the touch behavior. The segmentation stage yields data sequences that are then used as input to the proposed model for feature extraction and classification. Due to this, the system can attain improved sensitivity to small variations in touch, which is important for continuous user authentication.

Proposed Deep Learning Model

In the proposed DL model, the CNN part of the proposed scheme is developed to learn spatial information from sensor data to accomplish continuous authentication. CNNs are effective both to learn hierarchical feature representations and especially for tasks that utilize structured data (e.g., images, time-series signals). The proposed CNN architecture consists of two convolutional and max-pooling layers as shown in Figure 1. In this proposed HybridTouch model, its architecture consists of multiple layers for the extraction, transformation, and classification of both touch-based sensor data. The architecture shown in Table 2 begins with a Conv1D layer (Conv1D1) with 5 and 64 filters, followed by MaxPooling1 to filter the channels. The second Conv1D layer (Conv1D2) has a kernel size of 5 and 128 filters, and MaxPooling2 is then used to further downsample the feature maps. Both convolutional layers are characterized by a stride of 2 and padding of 1, thus preserving the input sizes and the feature extraction being efficient. A neural network followed by convolutional layers uses a GRU at the end point of sequential data. The GRU-type cells are implemented to reflect temporal dependency in the touch data; hence, both the short-term and long-term features are represented clearly. In the classification task, three Dense layers are constructed after the GRU layer. These dense layers include 128 neurons with a 0.3 dropout rate in the first layer, followed by a layer of 64 neurons with a 0.3 dropout, and the final output layer consists of two neurons for binary classification.

Table 2. HybridTouch Model Characteristics

Layer Name	Kernel Size	Kernel Number	Padding	Stride
Conv1D1	5	64	1	2
MaxPooling1	2	None	0	2
Conv1D2	5	128	1	2
MaxPooling2	2	None	0	2
GRU	-	-	-	-
Dense1	-	128	-	-
Dense2	-	64	-	-
Output	-	2	-	-

Although CNNs are good at learning features, they are not good at certain classification/learning tasks, e.g., time-series data, like smartphone sensing data. Such data dependencies are strong on the ability of the network to take current and previous inputs into account, since a historical situation influences the network's capacity to predict the next state (Shorten et al., 2019). To overcome this, RNN models, able to classify each component of a time series, are preferable. In an RNN, the term time t is dependent on the term, t^{-1}. Nevertheless, RNNs can be affected by the vanishing gradient issue when they are used with long sequences, and consequently, long-range dependencies can hardly be learned well. Description of the reset gate rt and the update gate zt that regulate the amount of the previous hidden state and the current input to be retained or discarded. The update gate zt sets the fraction of the previous hidden state that is to be kept at the current time step. In the meantime, the reset gate rt determines the degree of use of the previous hidden state in a calculation of the current state. Through these gates, the GRU can learn the long-term dependencies without suffering from the vanishing gradient effect observed in standard RNNs (Cho et al., 2014). The computational process of the GRU is as follows:

Update Gate: The update gate zt is computed as shown in Eq. (5) as follows:

$$z_t = \sigma(W_z x_t + U_z h_{t-1} + b_z) \tag{5}$$

Where Wz and Uz are weight matrices, xt is the current input, ht−1 is the previous hidden state, and bz is the bias term.

Reset Gate: The reset gate rt is calculated as shown in Eq. (6) as follows:

$$r_t = \sigma(W_r x_t + U_r h_{t-1} + b_r) \tag{6}$$

Where Wr, Ur, and br are the corresponding weight and bias terms.

Current Hidden State: The candidate hidden state ht is generated by the reset gate and the input as shown in Eq. (7) as follows:

$$\sim h_t = \tanh(W_h x_t + U_h (r_t \odot h_{t-1}) + b_h) \tag{7}$$

Where Wh, Uh, and bh are weight and bias terms, and ⊙ denotes element-wise multiplication.

Final Hidden State: The final hidden state is a weighted average of the previous hidden state and the candidate hidden state, influenced by the update gate zt as shown in Eq. (8) as follows:

$$h_t = \sigma(1 - Z_t) \odot h_{t-1} + z_t \odot h_t \tag{8}$$

The GRU's ability to control the flow of information through these gates allows it to learn long-term dependencies without the gradient-vanishing problem that affects traditional RNNs. This makes it particularly suitable for tasks involving sequential data, such as smartphone sensing for continuous user authentication (Qin et al., 2023). The model was trained using a batch size of 32 over 100 epochs, with categorical cross-entropy as the loss function and the Adam optimizer for efficient learning. Early stopping was employed to prevent overfitting by halting training when validation loss plateaued, ensuring optimal convergence without unnecessary computation (Kingma et al., 2014).

Performance Metrics

Several performance metrics were used to properly assess the proposed authentication framework for a comprehensive understanding of system dynamics. Model classification efficacy and the trade-off between sensitivity and specificity are described through such key metrics as false acceptance rate (FAR), false rejection rate (FRR), EER, and accuracy. Advanced metrics, such as the Receiver Operating Characteristic curve and Area Under the Curve (AUC), are used to further confirm the robustness of the model and its discriminatory power. FAR, FRR, accuracy, EER, precision, and recall are calculated with Eq. (9) to Eq. (12) as follows, where FA, False Acceptances refers to how many times the imposter has been mistakenly labelled as genuine and TR, True Rejections how many times an imposter was correctly rejected because it was not genuine. Again, FR is False Rejection, which indicates when a genuine user is wrongly marked as an imposter, while TA refers to True Acceptances, which refers to the right identification of a genuine user as genuine.

1. False Acceptance Rate (FAR):

$$\text{FAR} = \frac{\text{FA}}{\text{FA+TR}} \tag{9}$$

Likelihood of misclassifying an impostor as a genuine user.

2. False Rejection Rate (FRR):

$$\text{FRR} = \frac{\text{FR}}{\text{FR+TA}} \tag{10}$$

Likelihood of rejecting a genuine user as an impostor.

3. Accuracy:

$$\text{Accuracy} = \frac{\text{TA+TR}}{\text{TA+TR+FA+FR}} \tag{11}$$

Measures the overall classification performance.

4. Equal Error Rate (EER):

$$\text{EER} = \frac{\text{FAR+FRR}}{2} \tag{12}$$

Error rate at the threshold where FAR equals FRR, indicating system balance.

5. ROC Curve and AUC: ROC curves visualize the trade-off between FAR and FRR across thresholds. The AUC quantifies classification effectiveness, with higher scores reflecting superior performance.

6. Precision and Recall

Precision: $\frac{\text{TA}}{\text{TA+FA}}$ reflecting genuine predictions. (13)

Recall: $\frac{\text{TA}}{\text{TA+FR}}$ indicating sensitivity to genuine users. (14)

Data Augmentation with GAN

The GAN (Goodfellow et al., 2014) is a DL framework that contains two primary components: the generator and the discriminator. The discriminator evaluates the generated data as real or fake, making it train the generator to eventually produce data closer in nature to real data. The models develop in opposition and reach an equilibrium at their dynamic proportion during adversarial training. The generator receives some random noise-mostly uniform or normal distribution-while the output is synthetic data, and the discriminator tries to determine real data vs. generated ones (Antoniou et al., 2017). In the proposed work, WGAN is adopted for continuous user authentication data augmentation. This produces synthetic data usable for enhancing the performance of the model, especially while dealing with insufficient real-world data.

WGAN Architecture

The WGAN architecture is a simultaneous training of both the generator and the discriminator. A generator learns how to map noise to a distribution that resembles the real data. The discriminator attempts to distinguish real from fake data. The design of the architecture is such that it overcomes problems like mode collapse by measuring the Wasserstein distance between the real and generated data distributions. We define the WGAN loss function as a minimax game between the generator and the discriminator, as shown in Eq. (15), as follows.

$$\min_{G}\max_{D} L(D, G) = E_{x\sim p_r}[\log D(x)] + E_{x\sim p_g}[\log(1 - D(x))] \quad (15)$$

Where p_r is the real data distribution and p_g is the generated data distribution. To improve model performance, we replace the binary cross-entropy loss with the Wasserstein distance, as seen in Eq. (16).

$$W_{(p_r,p_g)} = \inf_{c\epsilon C(p_r,p_g)} E[||x - y||] \quad (16)$$

Where $C(p_r, p_g)$ represents the set of feasible joint distributions for the real and generated data. The loss function is adjusted using the Kantorovich-Rubinstein duality to enforce the Lipschitz continuity constraint for the discriminator shown in Eq. (17).

$$W(p_r, p_g) = \sup_{||D||_L \leq 1} E_{x\sim p_r}[D(x)] - E_{x\sim p_g}[D(x)] \quad (17)$$

Where the discriminator function D satisfies the Lipschitz constraint. Such a constraint may be realized through weight clipping of the discriminator during training. Generator and discriminator architecture: The generator and discriminator network architecture are as follows: The noise is used as the input in the generator. It passes through a CNN layer with 32 filters, followed by ReLU activation. The first layer of the generator is a dense layer with 32 units and sigmoid activation. This adds a dropout layer with dropout at 50%. Finally, it's attached to a batch normalization layer followed by one unit in a dense layer activated using the Tanh function. The discriminator structure starts with an input CNN layer, which is pretty much identical to that of the generator, this time using 32 filters but using ReLU activation. The first layer consists of 16 units with Tanh activation, followed by another dense layer with 8 units and sigmoid activation. Then comes the output layer, which is a final dense layer with just one unit, with sigmoid activation. This suggests a model using Wasserstein loss, along with weight clipping to combat the mode collapse that occurs during optimization within GAN models. Hence, the trained WGAN yields supplementary synthetic sensor data of increased quality and diversity of training data to complement the real HMOG dataset. Therefore, the designed model has far more significant performance than before in the continuous authentication scenario and subsequently better robustness and higher accuracy in the actual deployment.

Experimental and Results Analysis

The experiments were conducted using the HMOG dataset, which includes data on tactile dynamics for continuous user authentication. We aimed to assess the efficacy of the proposed model relative to baseline continuous authentication techniques (Abuhamad et al., 2020; Mekrusavanich et al., 2021), utilizing criteria including accuracy, EER, precision, and recall score. We present the outcome of these experiments in this section and compare this result by applying baseline standard models like CNN and GRU. The subsequent subsections discuss the architectural design of the proposed model, its training procedure, and experimental results. In particular, we compared the proposed model with both a CNN model and a GRU model. The CNN Model is a deep convolutional neural network that does not utilise a sequential architecture, focusing on the extraction of spatial characteristics while disregarding temporal relationships. The GRU Model utilises exclusively GRU layers to capture temporal correlations in the touch data, devoid of any spatial feature extraction. These models were selected because of their ability to account for spatial structures and sequential dependencies in time-series data, which are both qualities desirable for touch-based authentication. The HMOG dataset, which includes 100 users' time-series data, was exploited for training the model and its evaluation.

Experiment Setup

The experiments were performed on the HMOG dataset, which contains data from 100 users performing six different activities. To ensure a robust evaluation, the proposed model, HybridTouch was trained for 100 epochs using a 10-fold cross-validation technique to avoid overfitting and ensure homogeneous results across all users (Anguita et al., 2013). TensorFlow/Keras framework is used for implementing the model. The batch size for the training protocol was 32, and a learning rate of 0.001 was provided to the Adam optimizer to ensure convergence to the best solution. Since it was a multi-class classification, the categorical cross-entropy loss function was suitable as it would be an ideal optimization metric in this case. This set of hyperparameters was determined through iterative testing to consistently achieve high performance. The training set is divided into 90% training and 10% testing for each fold of data. In this experiment, it randomly shuffled the dataset to avoid leakage. In the performance measurement, the proposed model has been evaluated using such metrics as accuracy, precision, recall, and EER. Early stopping with the criterion validation accuracy has been applied during training so that overfitting may be avoided in this training (Kingma et al., 2014). Two different scenarios test the performance of the proposed model. In the first case, the baseline was a training of the original HMOG dataset without any augmentation. The second one would use GANs to synthesize synthetic data that could further augment the training dataset for generalization and robustness. Figures 8 and 9 present accuracy trends across epochs for the original and GAN-augmented datasets. The results were averaged over all folds of the cross-validation to ensure that they are smooth and reliable, as they are during training.

Results

The proposed model with and without GAN-generated data outperforms all the existing state-of-the-art DL models (Sağbas et al., 2024; Siddiqui et al., 2022; Frank et al., 2012) in average accuracy and EER across all the interactions, as shown in Table 3. Accuracy and EER performance of the proposed model in comparison to the existing schemes (Sağbas et al., 2024) are further depicted in Figures 2 and 3, respectively. It introduces an enhanced level of accuracy and decreased EER because of its unique architectural design of a hybrid framework. The integration of CNNs for spatial features extraction and GRUs for the modeling of time sequences effectively extracts complex patterns from user behavior. Advanced preprocessing methods employed include noise elimination using the Discrete Wavelet Transform, adaptive temporal windowing for dynamic segmentation, etc., to better represent features. Furthermore, GANs are used for data augmentation in the training set to enhance robustness and generalization. Thorough testing by 10-fold cross-validation with optimized hyperparameters and early stopping ensures that the proposed framework has better performance compared to other existing models and proves its ability to integrate spatial, temporal, and synthesized data insights. Figures 4 to 7 depict the results obtained by the HMOG dataset. The achieved results are presented with a significant gain in accuracy and EER regarding all the activities using the baseline models (CNN and GRU). More precisely, the proposed model using data generated by GAN demonstrates greater accuracy (99.0%–99.2%) and lower EER (1.25%) compared to the baseline models (Giorgi et al., 2021).

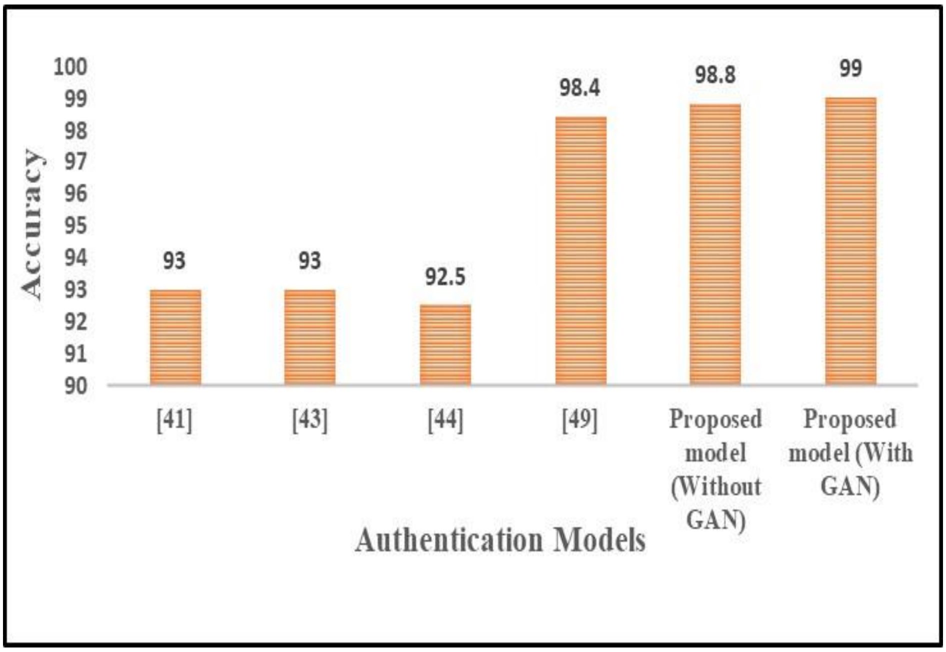

Figure 2. Comparison of the accuracy of the proposed model with existing schemes

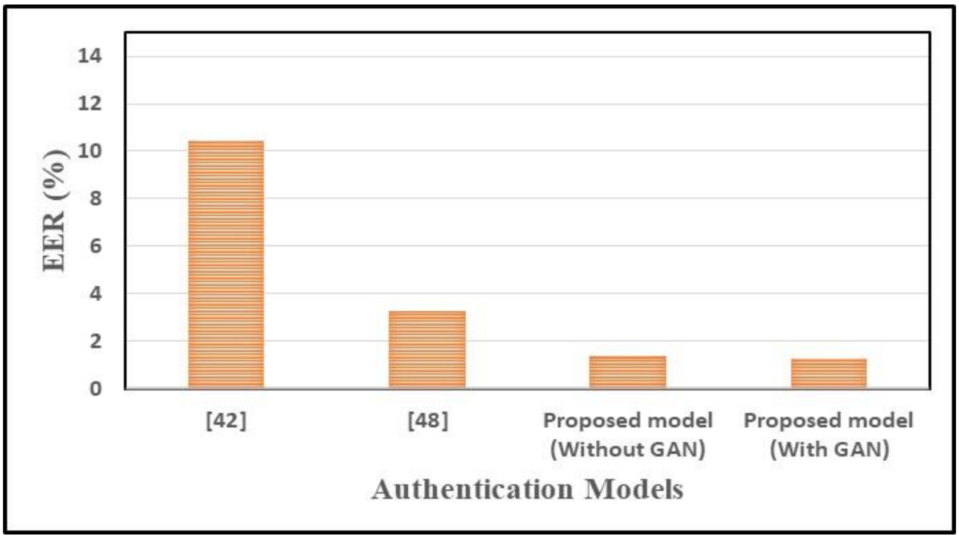

Figure 3. Comparison of EER of the proposed model with existing schemes

Table 3. Average Metrics on Classifier Evaluation of DL models using HMOG Dataset

Activity	HybridTouch Model	CNN Model	GRU Model
Read Sit	Accuracy: 98.9%(±0.5%) EER: 1.5% (±0.6%)	Accuracy: 99.0%(±0.4%) EER: 1.4% (±0.5%)	Accuracy: 98.6%(±0.6%) EER: 1.6% (±0.7%)
Read Walk	Accuracy: 98.8%(±0.6%) EER: 1.6% (±0.7%)	Accuracy: 98.8%(±0.5%) EER: 1.6% (±0.7%)	Accuracy: 98.4%(±0.8%) EER: 1.7% (±1.0%)
Write Sit	Accuracy: 99.0%(±0.4%) EER: 1.3% (±0.5%)	Accuracy: 98.8%(±0.6%) EER: 1.4% (±0.7%)	Accuracy: 98.5%(±0.9%) EER: 1.5% (±0.8%)
Write Walk	Accuracy: 98.7%(±0.7%) EER: 1.7% (±0.9%)	Accuracy: 98.6%(±0.9%) EER: 1.6% (±1.0%)	Accuracy: 98.3%(±1.0%) EER: 1.8% (±1.2%)
Map Sit	Accuracy: 98.9%(±0.5%) EER: 1.4% (±0.6%)	Accuracy: 99.0%(±0.3%) EER: 1.3% (±0.4%)	Accuracy: 98.6%(±0.7%) EER: 1.5% (±0.8%)
Map Walk	Accuracy: 98.8%(±0.6%) EER: 1.5% (±0.7%)	Accuracy: 98.6%(±0.8%) EER: 1.4% (±0.7%)	Accuracy: 98.5%(±0.8%) EER: 1.6% (±0.9%)

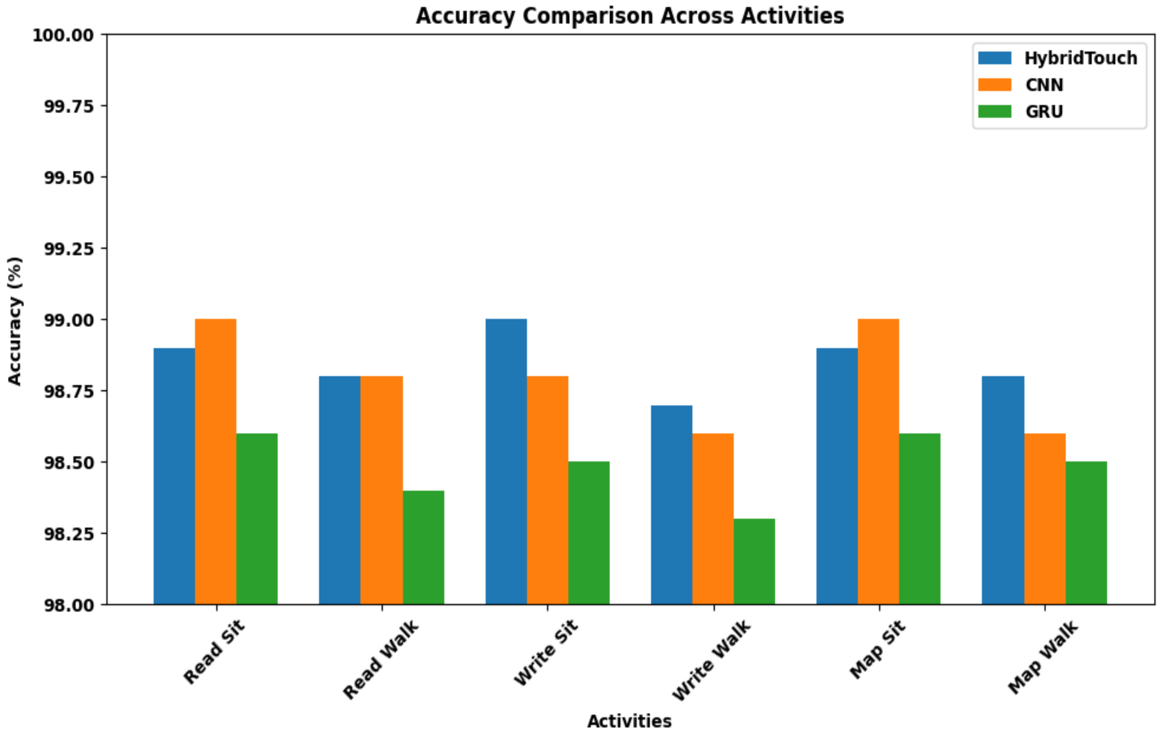

Figure 4. Comparison of Accuracy Across Activities

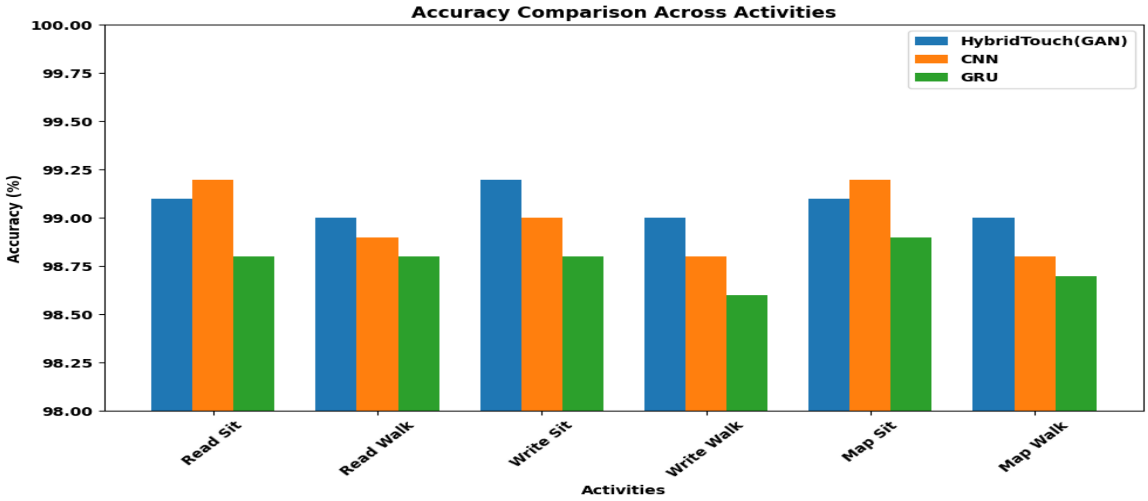

Figure 5. Comparison of Accuracy Across Activities (GAN-based data)

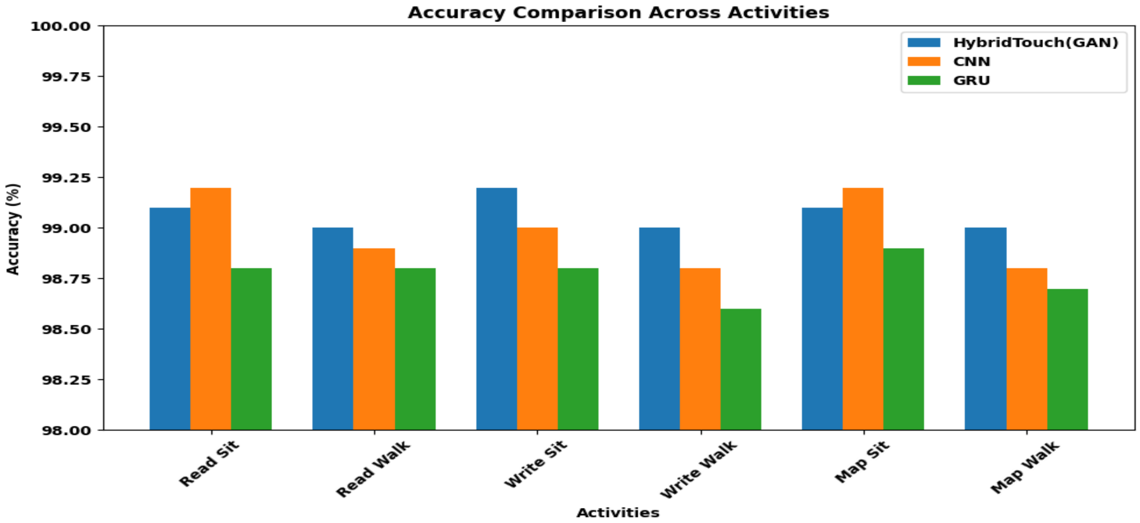

Figure 6. ERR vs. Activities (Real Data)

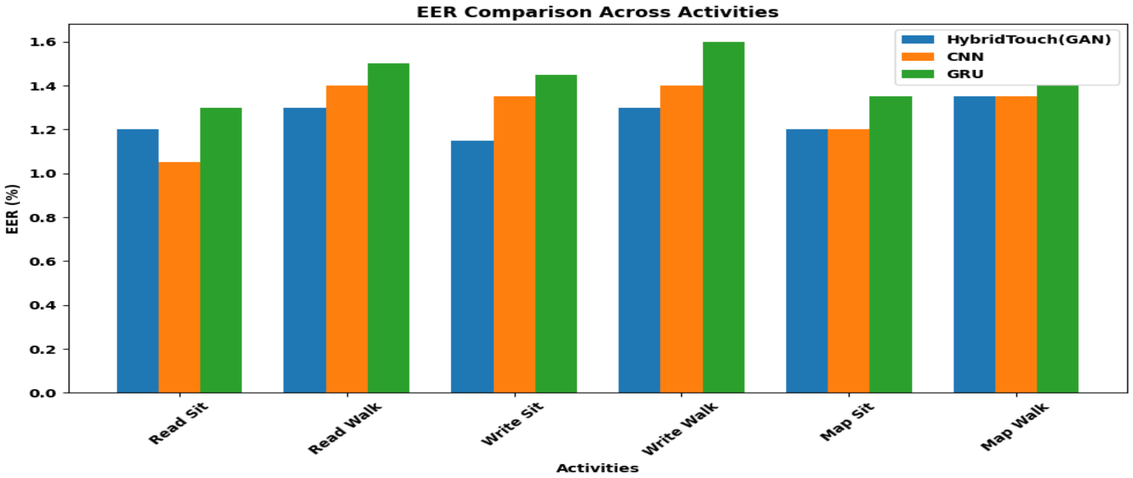

Figure 7. ERR vs. Activities (GAN-based Data)

Table 4. Average Metrics on Classifier Evaluation of GAN-based DL Models Using HMOG Dataset

Activity	Models	Accuracy (%)	Accuracy±(%)	EER (%)	EER±(%)
Read Sit	HybridTouch(GAN)	99.1	±0.4	1.20	±0.5
	CNN	99.2	±0.6	1.05	±0.7
	GRU	98.8	±0.5	1.30	±0.8
Read Walk	HybridTouch(GAN)	99.0	±0.5	1.30	±0.6
	CNN	98.9	±0.8	1.40	±0.9
	GRU	98.8	±0.7	1.50	±0.9
Write Sit	HybridTouch(GAN)	99.2	±0.4	1.15	±0.5
	CNN	99.0	±0.5	1.35	±0.7
	GRU	98.8	±0.6	1.45	±0.8
Write Walk	HybridTouch(GAN)	99.0	±0.4	1.30	±0.5
	CNN	98.8	±0.8	1.40	±0.9
	GRU	98.6	±0.7	1.60	±0.8
Map Sit	HybridTouch(GAN)	99.1	±0.3	1.20	±0.4
	CNN	99.2	±0.6	1.20	±0.7
	GRU	98.9	±0.5	1.35	±0.8
Map Walk	HybridTouch(GAN)	99.0	±0.5	1.35	±0.6
	CNN	98.8	±0.7	1.35	±0.8
	GRU	98.7	±0.6	1.40	±0.9

The HybridTouch model always shows better performance than the baseline CNN and GRU models concerning all tasks in terms of accuracy (98.8%–99.0%) and EER (1.4%), respectively, for the sitting and walking tasks. Sitting-based tasks, e.g., "Write Sit," "Map Sit," and "ReadSit," achieve the best accuracy, while walking-based tasks have slightly lower accuracy and higher EER, especially for the GRU model. This indicates that the proposed model is better suited to stable conditions, but could be improved for dynamic task conditions such as walking. When trained using the GAN-synthesized data, the proposed model features a significant performance improvement, including a variety of accuracy from 99.0% to 99.2% and EER of 1.25% decrease, as shown in Table 4. The improved performance, especially on seat-based tasks, illustrates the model's capacity to perform well on a variety of activity types, while at the same time being better than baseline models on walking tasks. These observations highlight the power of GAN data augmentation in improving the overall proposed model performance for continuous user authentication. Figure 8 shows the increase in model accuracy across epochs that can be attributed to the iterative optimization process during training. As the model is exposed to more data, its weights are progressively updated through backpropagation as it minimizes the error on predictions. Adam optimizer with learning rate 0.001 efficiently and stably converges towards the minimum while categorical cross-entropy loss is used to teach the model separations between classes. Advanced techniques in preprocessing as well as the extraction of features facilitate the model's ability to find meaningful patterns while improving predictions step by step. Moreover, early stopping prevents overfitting, and the accuracy is guaranteed to increase as the model converges towards optimal performance. The gain in model accuracy resulting from GAN-based data augmentation is attributable to the diversity and richness provided by the synthetic data, as illustrated in Figure 9. GANs generate additional training samples that replicate authentic user behaviour, hence enhancing the model's ability to generalise to unfamiliar data.

Thus, this enriched data set allows capturing a wider gamut of patterns of behavior at training time, resulting in more appropriate predictions. Since the GAN-augmented data helps cope with the effects of data poverty, the augmented model is a lot more rugged and learns using a much better distribution of different patterns of behavior across epochs. These consequently improve the efficiency of the prediction model. Figure 10 clearly shows the improvement in the AUC when using GAN-generated data, reflecting enhanced model robustness and better discrimination between classes. The proposed model with GAN achieves a higher AUC, indicating improved True Positive Rates (TPR) at lower False Positive Rates (FPR). The findings show a significant performance improvement when augmented data is used, by improving accuracy, precision, and recall across all users. Tasks, like Map Sit and Write Sit, show the best performance, demonstrating the model's ability to perform well in stable conditions. Even if the metrics in the activities that involve walking are lower, the model still outperforms the baseline methods. Introducing augmented data dramatically increases not only the overall accuracy and recall but also the robustness of authentication. These results highlight the essential contribution that data augmentation plays in enhancing performance in a range of continuous authentication tasks, which are in agreement with prior work that has focused its power on data scarcity and how to generalize the model (Goodfellow et al., 2014; Shorten et al., 2019).

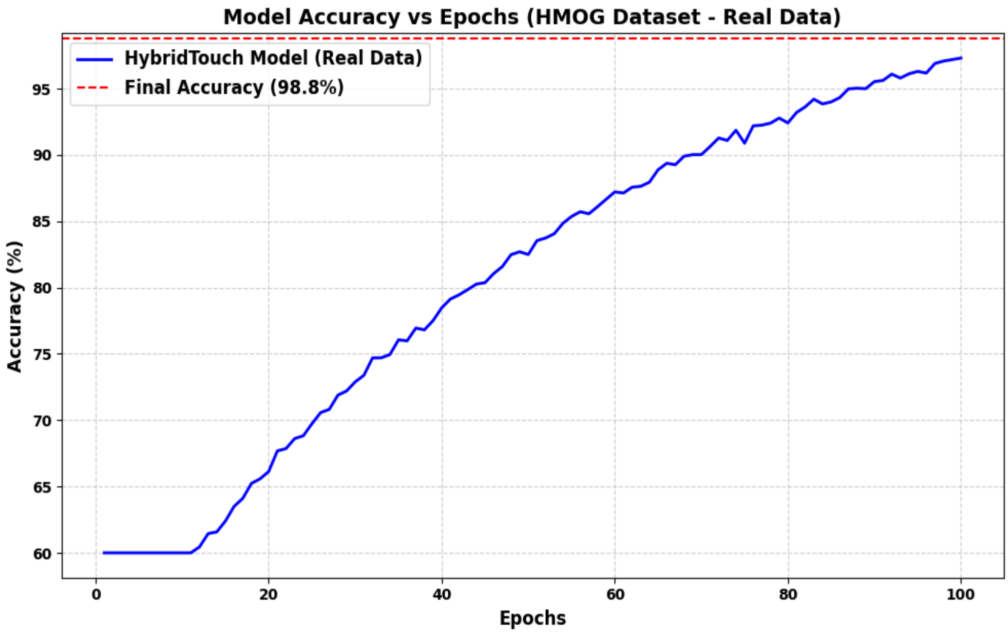

Figure 8. Model Accuracy vs. Epoch

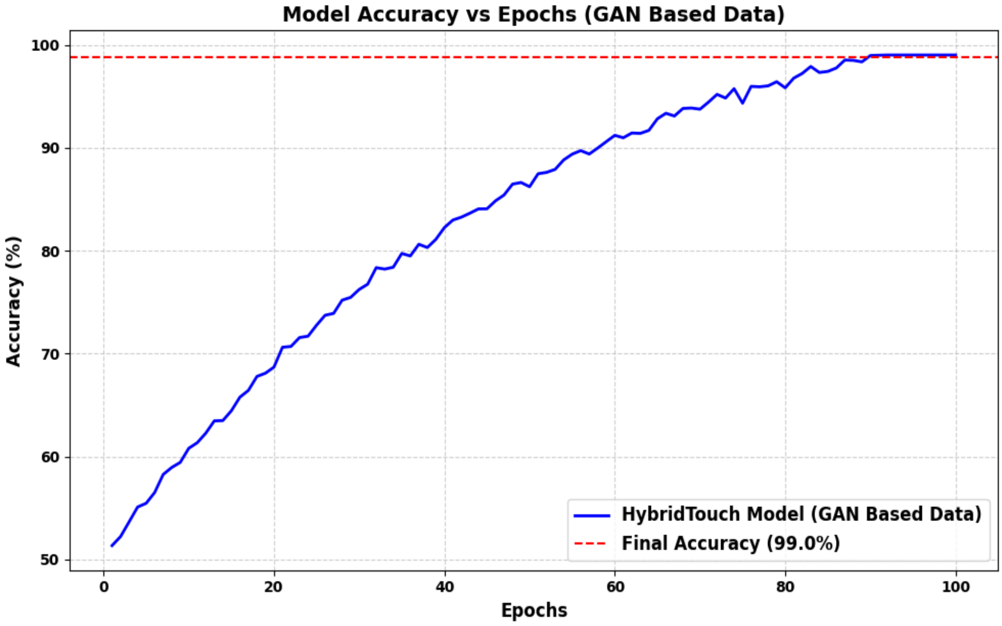

Figure 9. Model Accuracy vs. Epoch (GAN-based Data)

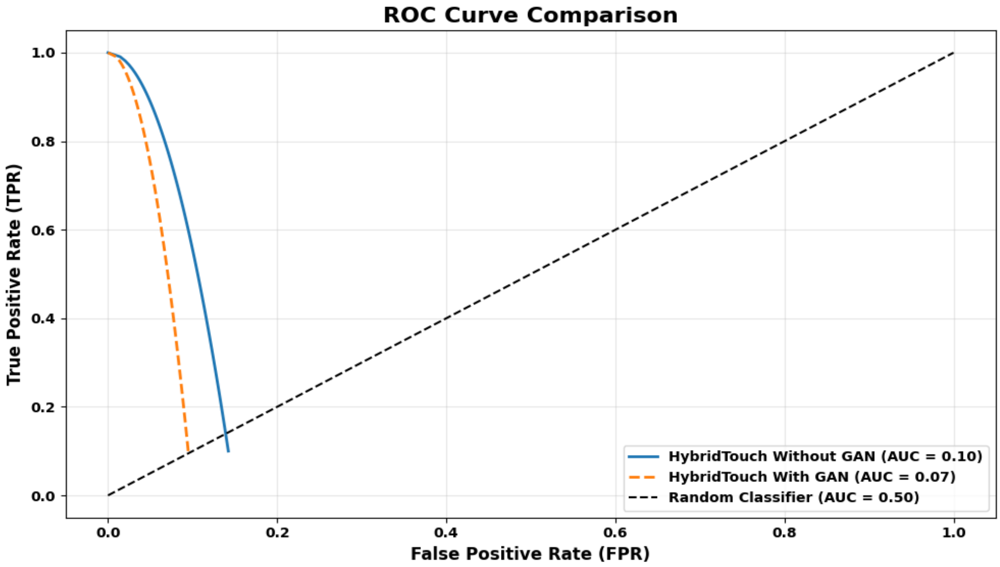

Figure 10. ROC Curve Comparison

Conclusion

Smartphones have become ubiquitous devices that are used for operations, e.g., mobile banking, communication, and storage of personally identifiable information, and therefore, secure mechanisms of protection of sensitive information are necessary. To meet the increasing need for continuous and uncompromised authentication schemes, we propose a framework for continuous authentication that uses data from smartphone sensors. The framework combines CNN for feature extraction and GRU for sequential pattern analysis, which allows for an in-depth understanding of user behaviour. In particular, GANs are used for creating synthetic data that solves the data scarcity problem and considerably improves the model's usability and generalization process. Compared with baseline models including common LSTMs, GRUs, and CNN alone, the proposed model achieves significantly high performance in terms of authentication accuracy (98.8%) and an EER (1.4%). The integration of GANs has significantly enhanced the proposed framework by augmenting the dataset with high-fidelity synthetic samples, capturing the variability in user activity patterns, and enabling robust generalization across diverse behaviors. This feature also reduces overfitting, which is a common problem in authentication models. The proposed system, with continuous, non-intrusive, and secure user authentication, is a significant step forward in mobile security. This framework uses advanced modeling and novel data augmentation techniques, which are a significant advancement for DL-based continuous authentication systems on cellphones and facilitate future advancements. Although GAN-based augmentation effectively tackles the problem of data scarcity, future work might explore other generator models, for example, Variational Autoencoders (VAEs) and Diffusion Models, to be able to present diverse and realistic synthetic data. In addition, transfer learning and self-supervised learning (SSL) have more promising opportunities with unlabelled data and better generalisation of the models. Federated Learning improves privacy-preserving training by decentralized model construction over user data, which improves security as well as usability. The further research has to focus on efficiency and scalability that guarantees the use of such models on mobile phones without sacrificing either accuracy or experience.

CRediT authorship contribution statement

Mahendra Kumar Jangir: Writing – original draft, Methodology. Karan Singh: Supervision, Methodology, Software, Validation, Visualization. Tayyab Khan: Supervision, Conceptualization, Formal analysis, Investigation, Editing.

Data availability: The research data supporting the results of this study are available at https://www.cs.wm.edu/~qyang/hmog.html.

Conflicts of Interest: The authors declare that they have no competing interests.

Funding Declaration: We declare that this manuscript received no funding from any sources.

Conflict of interest

The authors declare no potential conflict of interest regarding the publication of this work. In addition, the ethical issues including plagiarism, informed consent, misconduct, data fabrication and, or falsification, double publication and, or submission, and redundancy have been completely witnessed by the authors.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

References

Abuhamad, M., Abuhmed, T., Mohaisen, D., & Nyang, D. (2020). AUToSen: Deep-learning-based implicit continuous authentication using smartphone sensors. *IEEE Internet of Things Journal, 7*(6), 5008–5020.

Agrawal, K., & Bhatnagar, C. (2023). F-mim: Feature-based masking iterative method to generate the adversarial images against the face recognition systems. *Journal of Information Technology Management, 15*(Special Issue: EIntelligent and Security for Communication, Computing Application (ISCCA-2022)), 80–93.

Alfaleh, K., Alabdultif, A., & Aladhadh, S. (2024). Artificial intelligence-driven cyberbullying detection: A survey of current techniques. *Journal of Information Technology Management, 16*(4), 38–63.

Anguita, D., Ghio, A., Oneto, L., Parra, X., & Reyes-Ortiz, J. L. (2013, April). A public domain dataset for human activity recognition using smartphones. In *Esann* (Vol. 3, No. 1, pp. 3–4).

Antoniou, A., Storkey, A., & Edwards, H. (2017). Data augmentation generative adversarial networks. *arXiv preprint* arXiv:1711.04340.

Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning phrase representations using RNN encoder-decoder for statistical machine translation. *arXiv preprint* arXiv:1406.1078.

Ehatisham-ul-Haq, M., Azam, M. A., Loo, J., Shuang, K., Islam, S., Naeem, U., & Amin, Y. (2017). Authentication of smartphone users based on activity recognition and mobile sensing. *Sensors, 17*(9), 2043.

Giorgi, G., Saracino, A., & Martinelli, F. (2021). Using recurrent neural networks for continuous authentication through gait analysis. *Pattern Recognition Letters, 147*, 157–163.

Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2014). Generative adversarial nets. *Advances in Neural Information Processing Systems, 27.*

Hajiakhoondi, E., Hashemzadeh Khorasgani, G., Rahmany Youshanlouei, H., & Mirkazemi Mood, M. (2013). Proposing a model to evaluate communication technologies in mobile communication industry. *Journal of Information Technology Management, 5*(4), 47–66.

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint* arXiv:1412.6980.

Kokal, S., Vanamala, M., & Dave, R. (2023). Deep learning and machine learning, better together than apart: A review on biometrics mobile authentication. *Journal of Cybersecurity and Privacy, 3*(2), 227–258.

Lu, C. X., Du, B., Zhao, P., Wen, H., Shen, Y., Markham, A., & Trigoni, N. (2018, October). Deepauth: In-situ authentication for smartwatches via deeply learned behavioural biometrics. In *Proceedings of the 2018 ACM International Symposium on Wearable Computers* (pp. 204–207).

Mangal, A., Garg, H., & Bhatnagar, C. (2023). Assessing the performance of Co-Saliency detection method using various deep neural networks. *Journal of Information Technology Management, 15*(Special Issue: EIntelligent and Security for Communication, Computing Application (ISCCA-2022)), 23–34.

Mekruksavanich, S., & Jitpattanakul, A. (2021). Deep learning approaches for continuous authentication based on activity patterns using mobile sensing. *Sensors, 21*(22), 7519.

Nayak, A., & Bansode, R. (2016). Analysis of knowledge-based authentication system using persuasive cued click points. *Procedia Computer Science, 79*, 553–560.

Qin, Z., Yang, S., & Zhong, Y. (2023). Hierarchically gated recurrent neural network for sequence modeling. *Advances in Neural Information Processing Systems, 36*, 33202–33221.

Sağbaş, E. A., & Ballı, S. (2024). Machine learning-based novel continuous authentication system using soft keyboard typing behavior and motion sensor data. *Neural Computing and Applications, 36*(10), 5433–5445.

Shoaib, M., Bosch, S., Incel, O. D., Scholten, H., & Havinga, P. J. (2014). Fusion of smartphone motion sensors for physical activity recognition. *Sensors*, *14*(6), 10146–10176.

Shoaib, M., Scholten, H., & Havinga, P. J. (2013, December). Towards physical activity recognition using smartphone sensors. In *2013 IEEE 10th International Conference on Ubiquitous Intelligence and Computing and 2013 IEEE 10th International Conference on Autonomic and Trusted Computing* (pp. 80–87). IEEE.

Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. *Journal of Big Data*, *6*(1), 1–48.

Siddiqui, N., Dave, R., Vanamala, M., & Seliya, N. (2022). Machine and deep learning applications to mouse dynamics for continuous user authentication. *Machine Learning and Knowledge Extraction, 4*(2), 502–518.

Tran, L., & Choi, D. (2020). Data augmentation for inertial sensor-based gait deep neural network. *IEEE Access*, *8*, 12364–12378.

Yang, Q., Peng, G., Nguyen, D. T., Qi, X., Zhou, G., Sitová, Z., ... & Balagani, K. S. (2014, November). A multimodal data set for evaluating continuous authentication performance in smartphones. In *Proceedings of the 12th ACM Conference on Embedded Network Sensor Systems* (pp. 358–359).

Zhao, C., Gao, F., & Shen, Z. (2024). Multi-motion sensor behavior based continuous authentication on smartphones using gated two-tower transformer fusion networks. *Computers & Security*, *139*, 103698.

Zou, Q., Wang, Y., Wang, Q., Zhao, Y., & Li, Q. (2020). Deep learning-based gait recognition using smartphones in the wild. *IEEE Transactions on Information Forensics and Security*, *15*, 3197–3212.

Bibliographic information of this paper for citing:

Jangir, Mahendra Kumar; Singh, Karan & Khan, Tayyab (2025). HybridTouch: A Robust Framework for Continuous User Authentication Using GAN-Augmented Behavioral Biometrics on Mobile Devices. *Journal of Information Technology Management*, 17 (Special Issue), 108-129. https://doi.org/10.22059/jitm.2025.102924