Person Authentication from Behavioral Biometrics Using Smartphone and Smartwatch Sensors

Hamza Alobeidli 22010353
Ahmed Alwheibi 22010328
Ahmed Albreiki 22010325
{hamza.alobeidli, ahmed.alwheibi, ahmed.albreiki}@mbzuai.ac.ae

Abstract

Due to the rapid advancements in technology, there are now multiple ways of authenticating individuals on various devices. For instance, modern smartphones offer authentication through traditional passcodes, fingerprint scanning, and facial recognition. Alternatively, authentication can be achieved by analyzing an individual's physical actions using data collected from advanced sensors present in these devices. This project aims to investigate the effectiveness of accelerometer and gyroscope sensors, found in both smartphones and smartwatches, in identifying individuals based on their daily activities through various machine learning and deep learning models. Two experiments were conducted on the WISDM dataset, which contains gyroscope and accelerometer data for 41 users performing 18 different activities. The first experiment focused on identifying the 41 users, while the second experiment aimed to authenticate users among the 41. Three classification models were utilized for these experiments: long-short term memory (LSTM), a combination of convolutional neural network and LSTM, and finally, the random forest model. The results indicate that the random forest model performed comparably or better than the other two deep learning models, with an f1-score of 83% in the identification experiment and 95% in the authentication experiment.

1 Introduction

In our increasingly digital world, the need for secure and convenient methods of authentication has become paramount. With the widespread use of personal devices such as smartphones and smartwatches, traditional authentication methods like passcodes and PINs have proven to be susceptible to breaches and inconveniences. This has led to a quest for alternative authentication techniques that strike a balance between security and user-friendliness. One such promising approach involves leveraging data from advanced sensors, such as accelerometers and gyroscopes, to authenticate individuals based on their physical activities.

The objective of this project is to explore the effectiveness of using accelerometer and gyroscope sensors found in smartphones and smartwatches to authenticate users based on their daily activities. By collecting and analyzing data from these sensors, it becomes possible to develop machine learning and deep learning models capable of accurately identifying individuals. This innovative approach offers several advantages, including enhanced security, improved convenience, and reduced reliance on traditional authentication methods.

Beyond the realm of security, there are additional practical applications for this project. For instance, in the field of sports technology, multiple users can benefit from a shared sport smartwatch that tailors workout programs to their unique needs. By utilizing identification modeling based on individual physical actions, the smartwatch can identify each user and assign personalized workouts accordingly. This opens up new possibilities for personalized fitness experiences, allowing multiple users to utilize the same device while receiving tailored workout plans based on their individual identities.

36th Conference on Neural Information Processing Systems (NeurIPS 2022).

To investigate the viability of this approach, two experiments were conducted using the WISDM dataset, which contains gyroscope and accelerometer information from 41 users performing various activities. The first experiment focused on identifying the users among the 41 participants, while the second experiment aimed to authenticate users within this group. Three classification models, namely long-short term memory (LSTM), a combination of convolutional neural network and LSTM, and the random forest model, were employed to evaluate the performance of these identification and authentication tasks.

Overall, this project endeavors to address the pressing need for effective and user-friendly authentication methods in the digital age. By leveraging data from powerful sensors and employing advanced machine learning techniques, it aims to pave the way for more secure and convenient authentication approaches while exploring practical applications in different fields.

2 Literature Review

There have been several studies conducted in the field of authentication based on physical activities. Some of these studies focus on computer vision techniques for recognizing individuals, as seen in studies [1, 2, 3]. Other studies, similar to our own, delve into the analysis of gyroscope and accelerometer sensor data from wearable devices [4, 5, 6].

For instance, in study [4], researchers explored person identification through gait analysis using wearable devices equipped with gyroscope and accelerometer sensors. However, gait analysis only provides insights into the walking cycle, excluding information about other physical activities. In contrast, studies [5] and [6] considered a wider range of activities. In study [5], authentication and identification were performed using machine learning models such as k-nearest neighbors (kNN), support vector machines (SVMs), Bayesian Network (BN), and Decision trees (DT). The study collected data from the accelerometers and gyroscopes of 10 participants using a smartphone, encompassing activities such as walking, sitting, standing, running, climbing upstairs, and climbing downstairs.

Study [6] closely aligns with our own research. It utilizes the WISDM dataset, which contains gyroscope and accelerometer measurements from 51 individuals engaged in 18 different activities. The UCI-HAR dataset, in addition to WISDM, are both used to evaluate and compare deep learning models, including ConvLSTM, LSTM, and MSENet.

3 Methods and Materials

This project revolves around applying machine learning to tackle user identification and user authentication issues. A time-based dataset, capturing patterns of human activity, forms the backbone of this study. The data under investigation stems from everyday personal devices such as smartphones and smartwatches. These devices generate a wealth of valuable data that can be analyzed for the objectives of this project. To ensure robust and comprehensive analysis, several different machine learning models have been utilized. Each of these models offers unique strengths and provides different perspectives on the data. The models employed in this study include:

3.1 Recurrent Neural Network (RNN)

Recurrent Neural Networks (RNNs) have a unique feedback loop that influences each output using all previous outputs. Unlike typical neural networks, this makes them ideal for autocorrelated time series data. But they struggle with long-term dependencies due to gradient issues.

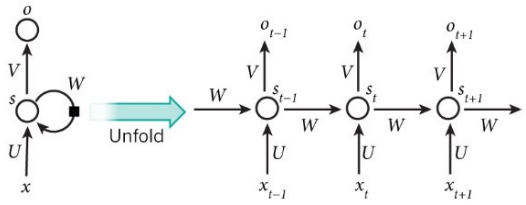


Figure 1: An expanded view of an RNN node

Each RNN cell's output is calculated by combining the previous output, current input, and a bias through an activation function, often tanh. The equation $o_t = g(Ux_t + Wo_{t-1} + b)$ represents this process, where 'g' is the activation function, 'x' the input, 'o' the output, 'U', 'W', and 'b' are the weights and bias, and 't' is the time step.

3.2 Long Short-Term Memory (LSTM)

Addressing RNNs' shortcomings, [7] proposed Long Short-Term Memory (LSTM), a variant of gated RNN. Through a series of gates and an additional line for cell state monitoring, LSTMs handle vanishing and exploding gradients effectively. This structure discerns crucial information for retention, an ability indispensable for our project, as it maintains and employs key data over long durations, suiting time sequence datasets.

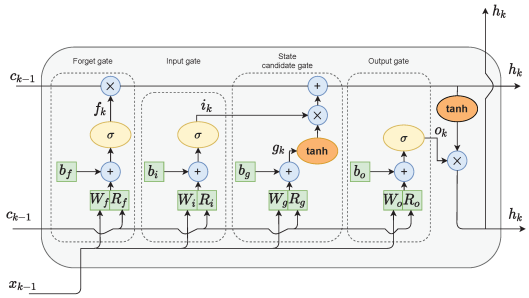
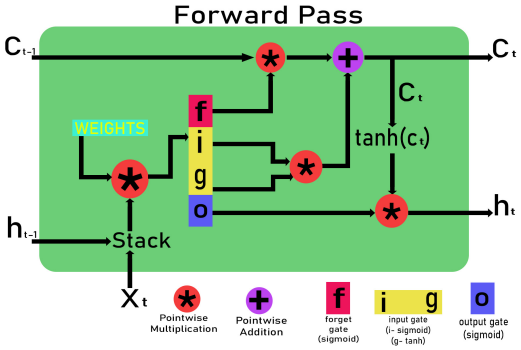


Figure 2: LSTM architechture

In the feedforward pass, data traverses the LSTM architecture, yielding an output prediction. Subsequently, the cross-entropy loss is computed. Upon determining the maximum likelihood estimation (MLE), weight adjustments occur through backpropagation using the chain rule. Gradients conform to equations in Figure 3a.

- $dE/dw_{xo} = dE/d_o * (d_o/dw_{xo}) = E\_delta * tanh(c_t) * sigmoid(z_o) * (1\text{-}sigmoid(z_o) * x_t$
- $dE/dw_{ho} = dE/do * (do/dw_{ho}) = E\_delta * tanh(c_t) * sigmoid(z_o) * (1\text{-}sigmoid(z_o) * h_{t-1}$
- $dE/db_o = dE/do * (do/db_o) = E\_delta * tanh(c_t) * sigmoid(z_o) * (1\text{-}sigmoid(z_o)$
- $dE/dw_{xf} = dE/df * (df/dw_{xf}) = E\_delta * o * (1\text{-}tanh^2(c_t)) * c_{t-1} * sigmoid(z_f) * (1\text{-}sigmoid(z_f) * x_t$
- $dE/dw_{hf} = dE/df * (df/dw_{hf}) = E\_delta * o * (1\text{-}tanh^2(c_t)) * c_{t-1} * sigmoid(z_f) * (1\text{-}sigmoid(z_f) * h_{t-1}$
- $dE/db_f = dE/df * (df/db_f) = E\_delta * o * (1\text{-}tanh^2(c_t)) * c_{t-1} * sigmoid(z_f) * (1\text{-}sigmoid(z_f)$
- $dE/dw_{xi} = dE/di * (di/dw_{xi}) = E\_delta * o * (1\text{-}tanh^2(c_t)) * g * sigmoid(z_i) * (1\text{-}sigmoid(z_i) * x_t$
- $dE/dw_{hi} = dE/di * (di/dw_{hi}) = E\_delta * o * (1\text{-}tanh^2(c_t)) * g * sigmoid(z_i) * (1\text{-}sigmoid(z_i) * h_{t-1}$
- $dE/db_i = dE/di * (di/db_i) = E\_delta * o * (1\text{-}tanh^2(c_t)) * g * sigmoid(z_i) * (1\text{-}sigmoid(z_i)$
- $dE/dw_{xg} = dE/dg * (dg/dw_{xg}) = E\_delta * o * (1\text{-}tanh^2(c_t)) * i * (1\text{-}tanh^2(z_g)) * x_t$
- $dE/dw_{hg} = dE/dg * (dg/dw_{hg}) = E\_delta * o * (1\text{-}tanh^2(c_t)) * i * (1\text{-}tanh^2(z_g)) * h_{t-1}$
- $dE/db_g = dE/dg * (dg/db_g) = E\_delta * o * (1\text{-}tanh^2(c_t)) * i * (1\text{-}tanh^2(z_g))$

(a) Gradients in LSTM



(b) LSTM backpropagation process

Figure 3: More information about LSTM

3.3 Convolutional Neural Network (CNN)

A Convolutional Neural Network (CNN) is a deep learning algorithm optimized for structured grid data like visual imagery and time series data due to their sequential nature.

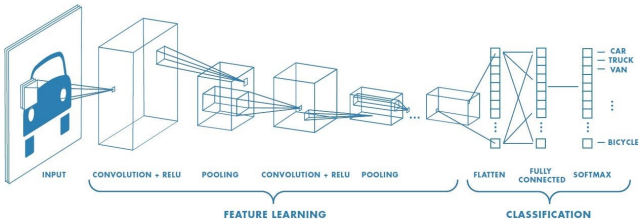


Figure 4: Typical CNN structure

Differing from regular feed-forward networks, CNNs apply convolution instead of matrix multiplication in at least one layer. Convolution layers in CNNs use learnable filters to create activation maps. CNNs exhibit translation invariance, recognizing patterns regardless of location, due to shared weights in convolutional layers, which also reduce overfitting by minimizing parameters. A CNN's structure typically includes convolution and pooling layers for feature extraction and dimensionality reduction, respectively, and fully connected layers for classification. During the forward pass, input data is processed sequentially through the CNN, computing output and calculating loss via a suitable function, such as cross-entropy loss. Weights are then updated through backpropagation using the chain rule, with gradients computed per layer type specifics.

In this project, CNNs' ability to identify local features and recognize patterns irrespective of their time sequence position is beneficial when applied to time series data.

3.4 Random Forest Classifier

The Random Forest classifier, an ensemble method, constructs multiple decision trees during training. A decision tree is a simple model that decides outcomes based on individual data features, using concepts like entropy and information gain to choose the best features for splitting. The classifier outputs the mode of individual tree classes, mitigating overfitting through ensemble averaging.

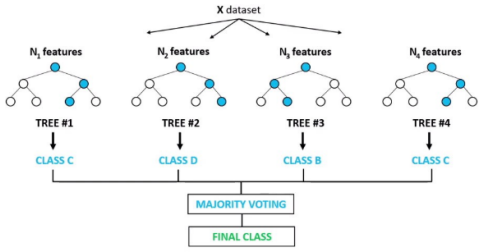


Figure 5: Simplified Random Forest Classifier illustration

Key Random Forest hyperparameters include the number of decision trees in the forest and the number of features each tree considers when splitting a node. The first boosts model robustness, while the second promotes diversity among trees, enhancing the model's resilience to noise and outliers. During training, the model learns from labeled data, associating features with labels. These learned decisions are then applied to unseen data during prediction. In this project, the Random Forest classifier's capability to manage high-dimensional data and its robustness to outliers and noise make it a fitting choice.

3.5 Hybrid CNN-LSTM Network

A hybrid CNN-LSTM network merges the power of Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks, adept at handling time series data with both temporal and spatial aspects.

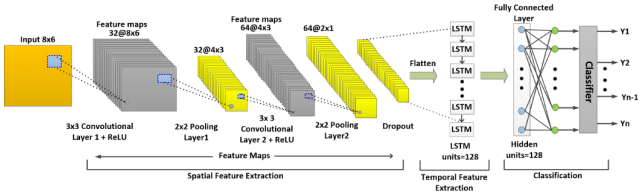


Figure 6: Hybrid CNN-LSTM Network illustration.

The model employs convolutional layers to extract spatial features from input data, identifying local patterns through shared weights and translation invariance. The LSTM layers then process these features as a sequence of feature vectors, adept at learning from long-term dependencies in the feature sequence due to their ability to manage sequence data with gating mechanisms. The CNN-LSTM fusion enables a nuanced data understanding, with CNN identifying crucial features and LSTM understanding their temporal dynamics. For this project, the hybrid model is optimal as our data, time-based activities from smartphones and smartwatches, inherently possesses spatial (features) and temporal (activity sequence) information. The model's capacity to comprehend both spatial and temporal patterns make it a robust choice for this project.

3.6 Materials and Model Parameters

Table 1: Models parameters used in the project

	Random Forest	LSTM	CNN/LSTM	1 User (neural network)	2 Users (neural network)
Layers	criterion: entropy max_depth: 30 max_features: log2 n_estimators: 200	LSTM layer: units=100. Dense layer: units=100. Dropout layer: rate=0.5.	Conv1D layer: filters=256, Dropout layer: rate=0.1. Conv1D layer: filters=512, Dropout layer: rate=0.2. MaxPooling1D: pool_size=2. Flatten layer. LSTM layer: units=256. Dropout layer: rate=0.2. Dense layer: units=100.	Add Dense layer: units= 40	Add Dense layer: units= 40
Output	majority voting	Dense layer: units=41, activation=Softmax.	Dense layer: units=41, activation=Softmax.	Dense layer: units=1, activation=Sigmoid.	Dense layer: units=3, activation=Softmax.

Techniques like GridSearch and Manual Hyperparameter tuning were used to choose such parameters.

The experiments are conducted in a set of experiments using an average Windows 11 computer specification (precisely, a Windows 11 (x64) with an i7-8665U CPU (2.11 GHz) and 16 GB of RAM) to evaluate the proposed architecture's performance. Jupyer Notebook is used with predefined libraries such as numpy, sklearn, and keras.

4 Experiments and Results

In this section, we present the experiments conducted to evaluate the performance of our proposed approach for person authentication from behavioural biometrics using smartphone and smartwatch sensors. We used the publicly available Wireless Sensor Data Mining (WISDM) dataset [8] to conduct our experiments. The first experiment aimed to recognize people based on their activities, while the second experiment focused on authenticating users based on their activity. For the latter experiment, we conducted two sub-experiments to authenticate a single user and authenticate of two users. We present our experimental setup, evaluation metrics, and the results obtained from our experiments in the following subsections: Dataset and Evaluation Metrics 4.1, Experiment 1: User Recognition 4.3, and Experiment 2: User Authentication 4.4.

4.1 Dataset and Evaluation Metrics

WISDM Dataset [8] contains activity information for 41 individuals in an activity recognition study. Each participant engaged in each of the 18 activities for three minutes, and sensor data (accelerometer and gyroscope for smartphone and smartwatch) was captured at a rate of 20 Hz, adding up to 64311 datapoints per participant per peripheral. The dataset contains both the raw data and processed data; generated by arffmagic.

Arffmagic data contains the following information: ACTIVITY (what activity the participants were doing when the datapoint was collected), Binned Distribution (grouping of 10 equal sized bins with 200 values each), Average, Time between peaks, Average Absolute Difference, Standard Deviation, Variance, Cosine distance, Correlation, and MFCC (Mel-frequency cepstral coefficients which represent a short-term power spectrum of a wave).

The different activities on the dataset are:

Table 2: Activities Represented in the dataset

Code	A	B	C	D	E	F	G	H	I
Activity	Walking	Jogging	Stairs	Sitting	Standing	Typing	Teeth	Soup	Chips
Code	J	K	L	M	O	P	Q	R	S
Activity	Pasta	Drinking	Sandwich	Kicking	Catch	Dribbling	Writing	Capping	Folding

To measure the performance of our proposed approach, we used two commonly used evaluation metrics, namely Accuracy and F1-score. Accuracy measures the proportion of correctly classified instances to the total number of instances, and is given by the following formula:

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

where TP, TN, FP, and FN represent the number of true positives, true negatives, false positives, and false negatives, respectively.

F1-score is the harmonic mean of precision and recall and is a more robust metric than accuracy when the dataset is imbalanced. It is given by the following formula:

$$F1 - Score = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

where Precision is the proportion of true positives to the total number of instances predicted as positive, and Recall is the proportion of true positives to the total number of instances that are actually positive.

4.2 Data Preprocessing

In our data preprocessing, we targeted two goals: user identification and user authentication.

For user identification, we merged all datasets into one, with each user assigned their own class for detailed identification. For user authentication, we ran experiments where only one or two users had unique classes, and the rest were grouped into a '0' class. This aimed to test the model's authentication ability. To avoid bias, we removed much of the '0' class data, balancing the dataset. We also set aside specific users (29 to 39 from class '0') solely for testing.

For both scenarios, we applied one-hot key binarization to the activity column, transforming each category into a separate binary feature. Additionally, we set the frame size to '4', preserving the time-sequential properties of the data while maintaining computational efficiency. These collective steps led to an effective and tailored preprocessing and training process for both tasks.

4.3 Experiment 1: User Recognition

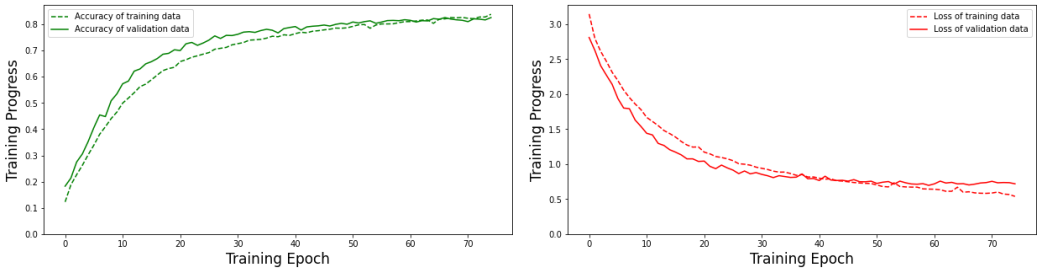
We evaluated our proposed approach for person recognition from behavioural biometrics using three different models, namely LSTM, CNN/LSTM, and Random Forest, using the WISDM dataset. The results of our experiments are summarized in Table 3. We observe that all three models achieved high accuracy and F1-score, with Random Forest achieving the highest accuracy of 84%. The LSTM and

CNN/LSTM models achieved accuracy scores of 82% and 83%, respectively. These results indicate that our proposed approach is effective in recognizing individuals based on their activities, and that different models can be used to achieve comparable results.

In addition to the results presented in Table 3, we also analyzed the accuracy and loss history of the LSTM and CNN/LSTM models during training using plot diagrams. The accuracy and loss histories for the LSTM model are shown in Figure 7, while those for the CNN/LSTM model are shown in Figure 8. As expected, the accuracy of both models increased over time, while the loss decreased which implies the effective of the methods and no overfitting is occurring. The CNN/LSTM model achieved slightly higher accuracy and lower loss than the LSTM model.

Table 3: User Recognition Scores

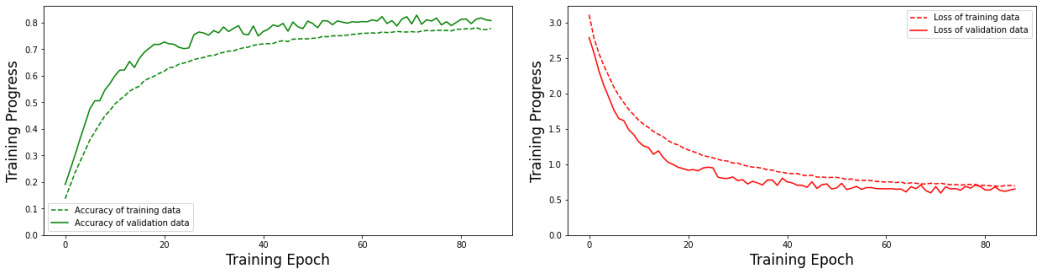
Model	Precision	Recall	F1-Score	Accuracy
LSTM	82%	82%	82%	82%
CNN/LSTM	83%	83%	83%	83%
Random Forest	84%	84%	83%	84%



(a) LSTM accuracy history

(b) LSTM loss history

Figure 7: Accuracy and Loss for LSTM model



(a) CNN/LSTM accuracy history

(b) CNN/LSTM loss history

Figure 8: Accuracy and Loss for CNN/LSTM model

4.4 User Authentication

In the user authentication experiment, we conducted two separate tests: one-person authentication and two-person authentication. In the one-person authentication test, we used the activity behavior data of a single user to determine whether they were the correct person or not. Meanwhile, in the two-person authentication test, we used the activity behavior data of two users to determine whether they were both the correct people or not. This allowed us to evaluate the performance of the models in scenarios where multiple users were involved in the authentication process. Next, a detailed results of each experiment using graphs and table containing the evaluation metrics.

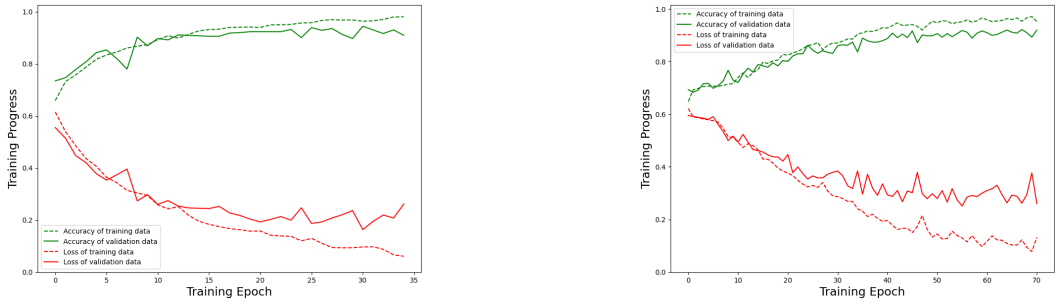
4.4.1 Single User Authentication Results

Based on the results of the 1 user authentication experiment, it appears that all three models achieved high levels of accuracy, with the CNN/LSTM model performing the best with 94% accuracy. The LSTM and Random Forest models also performed well with 91% and 93% accuracy, respectively. Looking at the precision, recall, and F1-score metrics, all three models achieved high values ranging from 94% to 98%.

The loss and accuracy history plots that in Figure 9 indicate that the models were able to learn and improve over time, with the CNN/LSTM model showing a better performance comparing to LSTM model. An early stopping mechanism is used to prevent model from overfitting. The confusion matrices in Figure 10 show that the models were able to correctly classify most of the activities, with only a few misclassifications. Overall, the results suggest that these models are capable of accurately authenticating a single user based on their activity behavior.

Table 4: 1 User Authentication Scores

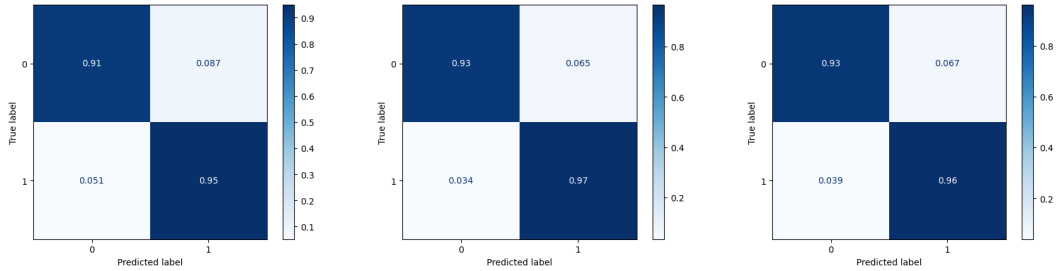
Model	Precision	Recall	F1-Score	Accuracy
LSTM	98%	91%	94%	91%
CNN/LSTM	98%	94%	95%	94%
Random Forest	98%	93%	95%	93%



(a) Accuracy and Loss history for LSTM Model

(b) Accuracy and Loss history for CNN/LSTM Model

Figure 9: Accuracy and Loss history for 1 User Authentication



(a) Confusion Matrix for LSTM Model

(b) Confusion Matrix for CNN/LSTM Model

(c) Confusion Matrix for Random Forest Model

Figure 10: Confusion Matrix for 1 User Authentication

4.4.2 Two Users Authentication Results

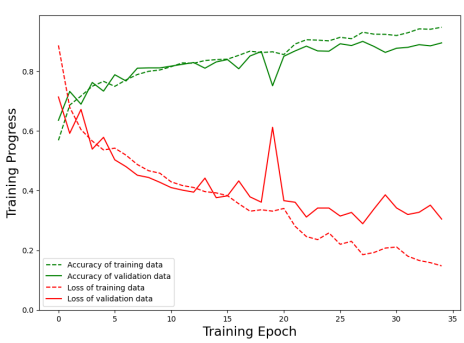
The experiment aimed to authenticate two users using their keystroke dynamics data. Three models, LSTM, CNN/LSTM, and Random Forest, were trained on the dataset and evaluated based on their performance measures. The results show that all models achieved high precision, recall, F1-score,

and accuracy. Random Forest achieved the highest performance with a precision of 97%, recall of 94%, F1-score of 95%, and accuracy of 94%. LSTM and CNN/LSTM models also performed well, achieving a precision of 97%, recall of 89%, F1-score of 91-92%, and accuracy of 89%.

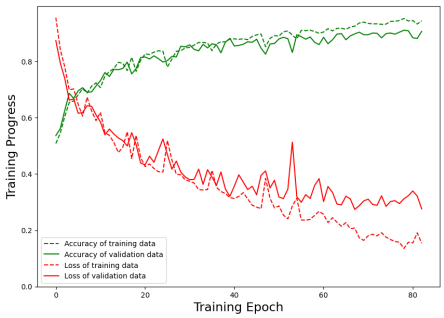
The accuracy and loss history for the LSTM and CNN/LSTM models in Figure 11 show that they both converged quickly and achieved a stable performance. The confusion matrices for all models in Figure 12 show that they correctly classified most of the samples, with only a few false positives and false negatives.

Table 5: 2 Users Authentication Scores

Model	Precision	Recall	F1-Score	Accuracy
LSTM	97%	89%	91%	89%
CNN/LSTM	97%	89%	92%	89%
Random Forest	97%	94%	95%	94%

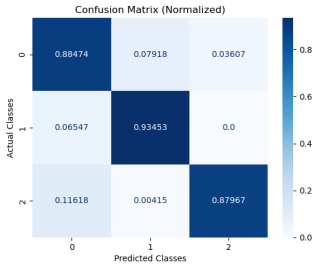


(a) Accuracy and Loss history for LSTM Model

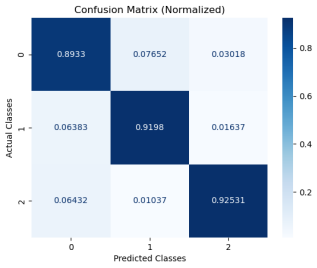


(b) Accuracy and Loss history for CNN/LSTM Model

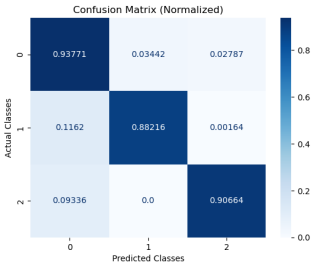
Figure 11: Accuracy and Loss history for 2 Users Authentication



(a) Confusion Matrix for LSTM Model



(b) Confusion Matrix for CNN/LSTM Model



(c) Confusion Matrix for Random Forest Model

Figure 12: Confusion Matrix for 2 Users Authentication

The proposed approach for person recognition using behavioural biometrics was evaluated using three different models, namely LSTM, CNN/LSTM, and Random Forest, using the WISDM dataset. All three models achieved high accuracy and F1-score, with Random Forest achieving the highest accuracy of 84%. The accuracy and loss histories of the LSTM and CNN/LSTM models during training were also analyzed using plot diagrams to examine the model behaviour. To avoid overfitting, some techniques are used such as dropout, pooling and early stopping. In the user authentication experiment, all three models achieved high levels of accuracy in the one-person authentication test, with the CNN/LSTM model performing the best with 94% accuracy. In the two-person authentication test, the models achieved accuracy scores ranging from 89% to 94%. Overall, the results suggest that the proposed approach is effective in recognizing individuals based on their activities, and that different models can be used to achieve comparable results.

5 Conclusion

In this study, we presented an approach for person authentication from behavioural biometrics using smartphone and smartwatch sensors. We conducted experiments to evaluate the performance of our proposed approach using the publicly available Wireless Sensor Data Mining (WISDM) dataset. The results of our experiments indicate that our proposed approach is effective in recognizing individuals based on their activities. We also evaluated our approach for person authentication and obtained promising results, demonstrating the potential of our approach for improving security in mobile and wearable devices. We used commonly used evaluation metrics such as Accuracy and F1-score to measure the performance of our proposed approach. Our approach can be useful for applications such as secure mobile payments, access control, and identity verification. Further research can explore the use of more advanced machine learning models and larger datasets to improve the performance of our approach. An alternative method could involve utilizing models that prioritize maintaining privacy, like Federated Learning and Hierarchical Federated Learning, in order to safeguard the privacy of the model data.

References

[1] W. G. Bhargavas, K. Harshavardhan, G. C. Mohan, A. N. Sharma, and C. Prathap, "Human identification using gait recognition," in *2017 International Conference on Communication and Signal Processing (ICCSP)*, pp. 1510–1513, 2017.

[2] Niyogi and Adelson, "Analyzing and recognizing walking figures in xyt," in *1994 Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 469–474, 1994.

[3] J. P. Singh, S. Jain, S. Arora, and U. P. Singh, "Vision-based gait recognition: A survey," *IEEE Access*, vol. 6, pp. 70497–70527, 2018.

[4] N. T. Trung, Y. Makihara, H. Nagahara, Y. Mukaigawa, and Y. Yagi, "Performance evaluation of gait recognition using the largest inertial sensor-based gait database," in *2012 5th IAPR International Conference on Biometrics (ICB)*, pp. 360–366, 2012.

[5] M. Ehatisham-ul Haq, M. A. Azam, J. Loo, K. Shuang, S. Islam, U. Naeem, and Y. Amin, "Authentication of smartphone users based on activity recognition and mobile sensing," *Sensors*, vol. 17, no. 9, p. 2043, 2017.

[6] F. Luo, S. Khan, Y. Huang, and K. Wu, "Activity-based person identification using multimodal wearable sensor data," *IEEE Internet of Things Journal*, vol. 10, no. 2, pp. 1711–1723, 2023.

[7] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Computation*, vol. 9, pp. 1735–1780, 11 1997.

[8] G. M. Weiss, K. Yoneda, and T. Hayajneh, "Smartphone and smartwatch-based biometrics using activities of daily living," *IEEE Access*, vol. 7, pp. 133190–133202, 2019.