Authentication of Remote IoT Users Based on Deeper Gait Analysis of Sensor Data

SAMERICA BATOOL1, ALI HASSAN1, NAZAR ABBAS SAQIB2, (Member, IEEE), AND MUAZZAM A. KHAN KHATTAK3, (Senior Member, IEEE)

1Department of Computer and Software Engineering, College of Electrical and Mechanical Engineering (E&ME), National University of Sciences and Technology, Islamabad 44000, Pakistan
2Department of Computer Science, College of Computer Science and Information Technology, Imam Abdulrahman Bin Faisal University, Dammam 31451, Saudi Arabia
3Department of Computer Sciences, Quaid-i-Azam University, Islamabad 45320, Pakistan
Corresponding author: Samera Batool (samera.batool@ce.ceme.edu.pk)

ABSTRACT In IoT based systems, authentication of users and devices is a major challenge where the traditional authentication mechanisms such as login-password are no longer supportable. It requires continuous authentication methods to ensure authenticity of the users and devices specifically for long sessions. The authentication through the IoT-Sensors based data has already gained the attention of a considerable number of researchers as it does not require direct involvement of the users and provide an extra layer of security and user privacy. Due to the instability of IoT-Sensor data, the authentication techniques need to extract a large number of features to produce high-accurate results. Moreover, the limited capabilities of computation, communication, storage, and the small battery power of IoT devices further makes its implementation hard. In this paper, we have introduced a user authentication framework for remote IoT users based on unique walking patterns to extract gait-related features with minimum data samples and cycle length. For an in-depth analysis of the distinctive features of the sensor-based gait profiles, we have applied an ECG signal processing technique. The proposed approach applies to a diverse range of IoT devices such as cellphones, smartwatches, and wearable sensors. The main contribution of the proposed approach is the deeper gait analysis with the least number of features and data for authentication purposes. We have introduced an authentication algorithm for feature comparison. The machine learning models have been applied to the gait based profiles for validation of the proposed approach. The detailed experimental analysis of different data sets has achieved an accuracy of 94% and an equal error rate (EER) of 6% higher than the existing approaches.

INDEX TERMS Data analytics, gait analysis, human activity recognition, Internet of Things, sensor data.

I. INTRODUCTION

The Internet of Things (IoT) is a means of communication between the objects of the physical world. IoT devices embedded with the latest sensors, actuators, and internet connections significantly improve the effectiveness of communication [1], [2].

Due to the advancement in technology and cheap IoT devices, many research efforts in various domains are in progress. IoT based applications and related data are sensitive, which require secure transmission to enable smart, mobile, and remote IoT services [3]. Particularly, ensuring the security of IoT users and devices is a major challenge.

The associate editor coordinating the review of this manuscript and approving it for publication was Zhipeng Cai.

Although conventional authentication methods such as passwords and biometric authentication techniques work fine, they still are not considered the ultimate solution due to security breaches such as password theft and patterns guessing, etc. The screen taps location of devices is determined by accelerometer and gyroscope sensor to obtain even complex passwords [4], [5].

Furthermore, for continuous monitoring of the users’ security, the current passwords, and security questions based authentication techniques might be irritating for users to provide the credentials again and again. Therefore, investigations for activity-based passive authentication techniques are in progress for IoT based systems. Where, in case of any malicious activity or abnormal usage patterns, the authentication can take place without requiring the direct involvement of
the user. Maintaining the security of IoT devices and data is the most important concern. Sensor data-based periodic authentication techniques are valuable as an extra layer of security at a low computational cost [6]. Sensors are the embodiment component of IoT systems, and each function in IoT systems is the result of the sensed data by IoT devices from different activities. Sensed data is of good quality and contains unique patterns of human body movement that have great potential to be used for user authentication [7].

The current models of the smartphones and wearable smart IoT devices have several built-in sensors (i.e., accelerometer, gyroscope, and fingerprint). There are many software apps such as Samsung Sheath [8], accelerometer log [9], available for collecting, storing, and communicating the sensor data. These provide a platform for the collection, storage, transfer, and further analysis of such data for various useful applications [10]. Research efforts are underway to utilize the IoT devices data for various decision making and user authentication processes. In addition to other parameters, profiles based on users’ sensors data are also an important aspect of a subject providing specific information about the user. Standards are required for customization of user gait based profiles to improve user’s privacy and security for user authentication [11].

Human gait is unique for every individual involving the synchronization between the skeletal, neurological, and muscular systems of the human body. Human gait refers to the locomotion achieved by the human limb movement. Healthcare applications of human gait analysis are under consideration for a long time [12]. Unique patterns in human gait exist for user authentication, which is proved by biometrics and healthcare research. Human gait based user authentication being biometric technique is free of threats like replication or guessing passwords [1].

Medical studies have estimated that there are twenty-four different components of the human gait, which are unique for each individual. One cycle of the human gait represents the repetition of the same step in a walk sequence [13].

The extraction of the gait information is much easier due to the progress made in sensor technology and gait identification methods. The main focus is to extract distinctive patterns in human gait [14].

The contributions of the proposed research are listed below.
1. Introduction of a continuous authentication framework for remote IoT users based on gait analysis.
2. Deeper gait analysis with ECG signal processing technique.
3. Providing high accuracy with less amount of data samples and feature keeping in view the energy and computation limitations of IoT devices.
4. Experimental evaluation of the proposed approach using different data sets.
5. Introduction of a threat model specific to user authentication techniques.

Here we state the organization of the rest of the paper. In Section II, we present the related work. Section III contains the security and functional requirement of the proposed approach. Section IV comprises of the threat model and security analysis of the proposed authentication framework. Section V consists of the detail of the proposed approach. In Section VI, we present the experiments and results and a comparison of our approach with the existing techniques. Section VII consists of discussions about the proposed approach. The final section is about the conclusions and future work.

II. RELATED WORK

There are several efforts in the literature that uses sensor data to provide personalized user services, especially in the field of human activity recognition and user identification. Due to the high accuracy of machine learning models and ease of data collection, the sensor data-oriented approaches are quite robust. This research focuses on the use of IoT sensor data for users’ personalized activity-based profile models for authentication purposes.

In our previous research work [16], we have introduced an IoT data analytics technique for activity and user identification. We have developed models using classification algorithms for users and activity recognition. To apply the machine learning models, we synthesize different data sets and obtained a good accuracy. With 19 subjects.

In [15], the authors have presented a user authentication method that uses the cellphone-based accelerometer data of the various basic set of activities such as walking, jogging, climbing stairs from 36 users. The authors transform the raw accelerometer sensor data into examples of 10 seconds and apply the classification algorithms on it. After segmentation, the authors collect the statistical features, such as the mean of the different axis. They have performed the experiments with machine learning classification techniques J48 and neural networks. The limited number of experiments has shown good results.

In another relevant research work [12], the authors have used the cellphone-based accelerometer sensor data for biometric gait identification. They construct feature templates for user identification, both in frequency and time domain, and identify the most pertinent cycle to all other cycles as the unique template. Their experiments in the time domain have resulted in accuracy up to 79% and 92% accuracy in the frequency domain.

In [17], the authors have introduced a user authentication method using the body mass collected from the ground forces during the walking activity. The authors have used it as a powerful feature for user identification. They have proved with the results of their study that they could collect body mass in the fraction of a second with a standard deviation of less than 1- kilogram.
In [18], the authors have presented a continuous user authentication technique. They have developed an android application for collecting physical activity data from smart bands and smartwatches. They have collected data set from 500 volunteers and applied the Support Vector Machines (SVM) algorithm on the 45 extracted features and produced 93% accuracy. The continuous collection of sensor data for authentication purpose consumes computation and processing capabilities by wearable devices with low computation and processing resources.

The authors in [19], have used ECG signals for biometric authentication of the cell phone users. The users touch two ECG electrodes to be verified to access the device. They extract fiducial points from ECG signals to create user profiles and use these features for authentication purposes. They have introduced a feature matching algorithm that calculates the scores based on the matching features and authenticate accordingly. The use of additional ECG electrodes for the creation of user-profiles and authentication purposes is a limitation of their proposed technique. The user is required to touch the electrodes every time for gaining access to the device. Moreover, certain physical scenarios can cause variations in the ECG signals, which can affect the accuracy of the proposed approach.

Another related effort is introduced in [20], which uses a gait based authentication technique. For collecting acceleration and angular velocity data, the authors have used the ankle sensor. For the authentication process, the scoring system is used based on the predefined threshold values. They have achieved an EER of 20%.

In [21], the authors have introduced a passive authentication technique. They extract time domain and frequency domain features. They have performed the experiments on different data sets by applying random forest, SVM, and decision tree classification algorithms to present a comparison of the results of each algorithm. The random forest classifier has outperformed the other two classifiers in terms of performance. A complex set of activities can cause performance deterioration.

Recently, a gait based user authentication method is introduced by [22]. The authors have developed an android application for collecting data from smartwatches and cellphones and transfer it to the authentication server. The authors extract 27 physical and statistical features related to gait for user authentication. They use the gait cycle length of 100 samples to apply machine learning algorithms for user authentication. They have achieved an EER of 8.2%. They did not consider any attack models.

The authors in [23] have introduced a gait based user authentication technique. They have developed an android application for collecting data from smart devices. They record accelerometer sensor data of 15 seconds duration after detecting the walking activity using a pedometer sensor. To detect the gait cycle, the authors use dynamic time wrapping and the local maxima. For data preprocessing, they apply the linear interpolation and Savitzky Golay filter. For performance verification of the proposed approach, the authors have hired professional actors to deploy the impersonation attacks. Even they have failed to be regular to mimic the walking style of a subject. The authors have managed to achieve the EER of 13% with the data set of 35 subjects.

In [24], the authors have performed the gait analysis by improving an existing LTN technique. For the authentication purpose, they calculate the difference in shape and sequence of the gait structure. They divide the gait cycle into five motion regions to extract the traditional and new features from the gait cycle. They have applied machine learning classification algorithms on these features. They have collected the individual prediction results and then applied the MSV fusion classifier to obtain the final results of the authentication. The new features extracted by the authors have shown good results and yielded an accuracy of 98.42%.

The authors have presented a two-step authentication technique in [25]. They have developed a fingertip device to capture the motion and physiological data. The legitimate users wear the device on the fingertip for authentication. They perform activity recognition as the first step and apply machine learning classifiers, i.e., auto encoder SVM, and KNN on the state of the identified activity. Their experiments on 40 subjects produced an accuracy of 98%.

In Table 1, we present the summary of the state of the art techniques for the last three years for user authentication with their limitations.

### III. SECURITY AND FUNCTIONAL REQUIREMENTS

A user authentication system needs to ensure security and functional requirements. Here we list basic security and functional requirements for the proposed authentication technique.

1. The system should provide secure access, and no unauthorized user should be allowed to access the system resources.
2. User-specific data should not be used or shared without privileges to ensure the privacy of the users.
3. The system should provide the protection and retrieval of personal data in case of any loss.
4. The system should provide the protection and retrieval of personal data in case of any loss.
5. The system should be able to detect and prevent impersonation, spoofing, and replay attacks.
6. The system should secure data and resources in case of detecting any malicious activity to save it from further loss.

The functional requirements of the proposed approach include the following.

1. The system should provide access to legitimate users based on gait based credentials.
2. The system should not allow access to an illegitimate user based on gait based credentials.
3. The system should be user friendly to maintain passive authentication for continuous monitoring of users’ security.
IV. THREAT MODEL & SECURITY ANALYSIS

The main challenges of IoT based systems include data and device security with lots of potential threats to the IoT based user authentication. Fig. 1 presents the threat model of the proposed authentication system.

A. BATTERY FAILURE

IoT devices are limited in processing and energy capabilities. Whereas, the remote IoT users require the continuous or intermittent data collection for authentication that demands continuity of the operational devices. Battery failure or low power battery devices are a risk factor for the functioning of such systems.

B. BRUTE FORCE ATTACKS

An attacker can try to access the device source by attempting every possible combination of gait features.

Therefore, it is necessary for the authentication system to randomly generate the gait cycle features for not allowing the adversary to carry the brute force attacks.

C. DATA AND DEVICE THEFT

Commonly, the security of IoT devices is weak due to the negligence of non-technical users. Damage can be accidental or intentional theft. It can cause the failure of the user authentication system, providing an opportunity for a variety of attacks by losing private data.

D. IMPERSONATION ATTACK

Impersonation attack is known as one of the most complex & widely deliberated threats for gait based user authentication systems [26]. In this attack, the adversary tries to mimic someone’s walking style after thoroughly observing it.

E. DENIAL OF SERVICE ATTACK (DOS)

The smart IoT devices can be used as the botnets to launch denial of service attacks for an authentication server.

F. WEAK SIGNAL

The protocols such as Wi-Fi, ZigBee, etc. are used for performing the communication from an IoT device to the authentication server. These protocols have vulnerabilities, such as weak signals and coverage issues [27].

G. ACCESS CONTROL

The authentication servers on the cloud must implement strict access control mechanisms to maintain the security and privacy of the stored data for the functioning of systems.

H. DATA LOSS

The loss of data stored in the cloud to unauthorized parties is a serious concern. Third parties might use it for various malicious purposes.

I. HARDWARE FAILURE

To the authentication server residing on the cloud, another possible threat could be the accidental or technical failure of hardware resources.

The primary motivation behind this research work is to introduce an authentication scheme that is suitable for devices having limited computational and battery resources with lots of potential security threats. For performing the security analysis, we have assumed that the adversary has accessed the device or has stolen the logged-in device. The proposed authentication scheme is period based, which performs the more intense feature analysis with minimum samples of data. It solves the aforementioned threats of battery failure, limited computation, and communication capabilities.

### TABLE 1. Summary characteristics of the existing relevant techniques.

| Approach/Year | Methods | Sensors Used | Activities | No of Subjects | Features Extracted | Accuracy | Limitations |
|---------------|---------|--------------|------------|----------------|---------------------|----------|-------------|
| [21] / 2019   | Multiclass machine learning algorithms/SVM, DT and RF | Accelerometer, Gyroscope | Sitting, standing, lying, walking, upstairs, downstairs, running, jogging, jumping | 59 | 18, Time domain & Frequency domain | 95% | - Possible performance deterioration with complex activities | - Computation overhead of multiclass algorithms |
| [21] / 2019   | Machine learning algorithm i.e. RF | Accelerometer, Gyroscope | walking | 51 | 27 for one axis cycle of 100 samples | 91.8% | - Larger cycle length | - Authentication initiation events are not defined |
| [23] / 2017   | Linear interpolation, Savitzky Golay filter for preprocessing, DTW for gait point distance calculation and matching | Pedometer, Accelerometer | walking | 35 | Gait segments (start, first end), points detection, distance measurement | EER 13% | - More evaluation of the performance is required with larger datasets | - Device position variations needs to be tested for performance evaluation |
| [24] / 2019   | Multiple classifier fusion method (MSV) | Accelerometer | walking | 32 | 93, Time domain, Frequency domain, Gait points based | 98.42% | - Large no of features | - Classifier fusion causes extra overhead |
| [25] / 2018   | SVM, KNN & Decision Tree | Accelerometer, Photoplethysmography (PPG) | Motion (general) | 40 | 84, Time domain, Frequency domain, Wavelet domain | 98.5% | - Multi sensor data | - Classifier selection causes additional overhead |
We register the genuine users in the system, to curtail the probability of impersonation attack. The attacker is denied access due to the high accuracy of the feature extraction and matching process. Also, periodic authentication is applied to obtain the most recent data cycles from users after a predetermined interval to detect such attacks at an early stage, after the user has accessed the device by the mimic attack. The professional actors are not regular in intimating a person’s walk style [23].

V. THE PROPOSED APPROACH

In security-critical systems such as IoT, user authentication is the main requirement. Periodic authentication can be useful in such systems. The current user authentication techniques like usernames & passwords require frequent interaction from users such as answering the secret questions etc. for monitoring the security. We have introduced an authentication framework that uses activity-based sensor data for the periodic authentication without requiring direct interaction of the user. Fig. 2 displays the main elements and workflow of the proposed approach.

A. IOT DATA ACQUISITION

The smartphones, wearables, smartwatches with embedded sensors are examples of smart IoT devices. For data collection, these devices are deployed in an unrestrained environment [28]. These devices are held by the subjects while performing a different set of activities. The advancement in technology and embedded software apps enable the collection and transmission of data produced by sensors such as gyroscope, accelerometer, heart rate, etc. Wi-Fi, ZigBee, and Bluetooth are the most widely used IoT data transmission protocols [29]. We use built-in or newly developed apps to collect data from IoT devices. Emerging IoT devices such as the Internet of Drones are also making their impact on data collection [30]. Table 2 contains the set of attributes used for feature extraction.

The class attribute represents the activity, subject id is for the participant identifier. The other three attributes represent the acceleration of each axis. Here we present the data sets deployed for the experiments and results collection.

1) DATA SET [31]

In this data set [30], 51 subjects participated in the data collection process. The devices used for data collection include the Google Nexus 5/5x, Samsung s5 Cellphones, and LG G5.
smartwatch. The sensor data at a rate of 20 Hz is collected. Each user has performed different activities for 3 minutes each. To perform the experiments, we use data on the walking activity.

2) DATA SET [32]
The data set in [32] is collected from 30 subjects during walk activity while carrying the Samsung S-II device on the waist. Accelerometer and gyroscope sensor data at the rate of 50 Hz, are collected. We use only walking activity data for experimentation purposes.

3) PROPOSED APPROACH DATA SET
We also collected data on walk activity from 30 subjects. We have used the Samsung S7 Edge device and android app accelerometer log [7] for the data collection process. The participants of the age range from 15 to 34 have participated in the data collection process. The walking activity is performed at the same place by different subjects.

B. DATA PREPROCESSING
The data preprocessing phase prepares data for the feature extraction phase. We present below the steps for the data preprocessing:

1) DATA FILTERING
In this step, to remove the irrelevant attributes from the data set and keeping the required attributes data filtration step is performed as displayed in Table 2.

2) NOISE REMOVAL
The data obtained from smart IoT devices contain noise that is caused by the movement of other body parts and the environment. To remove the noise, we apply the simple moving average (SMA) method on a set of n values representing data of each axis. The SMA method applies the sliding window strategy to take the average of the set of n numbers i.e. $x_1, x_2, \ldots, x_n$ over a period. It represents the equally weighted mean of the previous n data points, where M is the size of the window. We use $M = 3$ for noise removal.

$$a_{SM} = \frac{x_n + x_{n-1} + \ldots + x_{M-(n-1)}}{M} \quad (1)$$

$$a_{SM} = \frac{1}{M} \sum_{i=0}^{n-1} x_{M-i} \quad (2)$$

For calculating the subsequent rolling average values, we add the new value to the sum and drop the previous value [33].

$$a_{SM} = a_{SM-\text{prev}} + \frac{1}{n} (x_M - x_{M-n}) \quad (3)$$

Fig. 3 presents the x-axis signal before and after the noise removal process.

3) DATA CYCLE SEGMENTATION
The sensory data of different activities contain unique patterns. For segmenting the data of each axis, we apply a sliding window of length 30, which is appropriate for the cycle length. Fig. 3 clearly shows similar data patterns repeat after approximately 30 samples.

ECG based signal processing technique relies on the extraction of fiducial points as these are distinguishable characteristics from person to person. Clear high magnitudes can be observed on each axis during walk activity as the sole touches the ground [34]. Fig. 4 shows one gait cycle of walk activity along with the highlighted fiducial points.

C. FEATURE EXTRACTION
The number and type of features depend on the feature extraction technique. The highlighted fiducial points in Fig. 4 contain unique values for each subject depending upon the walking style and walking speed.
TABLE 3. Notations and symbols.

| Symbol/Notation | Description         |
|-----------------|---------------------|
| $a_{SM}$        | Simple Moving Average|
| $\bar{x}$       | Mean value of $x$   |
| $\tilde{x}$     | Median value of $x$ |
| $\hat{x}$       | Mode value of $x$   |
| $SK_x$          | Skewness of $x$     |
| $Kurt_x$        | Kurtosis of $x$     |

TABLE 4. Vectors to store the fiducial points indexes.

|   | A   | B   | C   | D   | E   |
|---|-----|-----|-----|-----|-----|
| a1 | b1  | c1  | d1  | e1  |
| a2 | b2  | c2  | d2  | e2  |
|   |     |     |     |     |
| aN | bN  | cN  | dN  | eN  |

For the extraction of features, these fiducial points are used to apply an ECG signal processing technique [19]. The fiducial point values are stored in their corresponding vectors separately along with their indexes, as presented in the following Table 4. The total no of indexes is represented by $S$ where $1 \leq s < S$ where $s$ starts from 1 to $S$ for each feature.

The notations and symbols used for feature extraction and noise removal are presented in Table 3.

The following features are extracted from one gait cycle of each axis, as presented in Table 5. F1 represents the first deepest valley in a gait cycle, whereas C and E represent the other two deepest valleys. B and D represent the peaks or highest acceleration points in a gait cycle. F6 to F9 represents the statistical and non-statistical features.

1) MEDIAN ($\bar{x}$)
We calculate the median value for each gait cycle. The median value is the middle value from a set of $N$ numbers in the ascending order.

2) MEAN ($\tilde{x}$)
We calculate the mean value for each gait. The mean value represents the normal range of values at each axis.

3) SKEW ($SK_x$)
The skew value represents the acceleration of data in the early or later phases of a gait cycle [26]. The early phase acceleration represents the negative skewness, and the later phase acceleration of the gait cycle represents positive skewness. Here $n$ represents a set of values of a cycle. We calculate the skew of a gait cycle, as presented below.

$$Sk_x = \frac{3(\bar{x} - \tilde{x})}{s}$$

Here $\bar{x}$ represent the mean value
$\tilde{x}$ represents the mode value
$s$ represents the standard deviation.

4) KURTOSIS ($Kurt_x$)
The kurtosis is the feature that represents the sharpness of the peaks. In the gait cycle, the kurtosis represents the intensity and force of the walking style and movement of a subject. The formula for calculating the kurtosis for each cycle is below.

$$Kurt_x = \frac{\sum_{i=1}^{N} (x_i - \bar{x})^4}{N}$$

Here $\bar{x}$ represents the mean value of a gait cycle.
$s$ represents the standard deviation.
$N$ represents the sample size of a set of values $n$.

In addition to the above-stated features, more fiducial points-based features can be extracted, such as starting from F10 to F18.

The vector $Data[x, y, z]$ contains the acceleration values of each axis of length $N$. For the feature extraction process, we apply the indices and values stored in the corresponding vectors. F10 corresponding to AB represents the difference between points A and B. We apply the same process for other time-based features i.e. BC, CD, DE, and BD. We collect the amplitude-based features F15 to F18 using the actual values of the fiducial points contained in the $Data$ vector. The distinct features show a variant pattern for each user, as shown in Fig. 6. To reduce the number of features and select the
appropriate features, we have chosen only those features that have a high impact on accuracy. The reduced set of features have produced high accuracy as elaborated in the experiments and results section.

D. USER AUTHENTICATION
To evaluate the features for user authentication, we use two approaches, the scoring system, where we apply the feature matching algorithm and machine learning. The other method is by applying the machine learning algorithm on the extracted features to train the models and test with new data for user authentication purposes. We give a brief introduction of both the methods.

1) SCORING SYSTEM AND MATCHING ALGORITHM BASED USER AUTHENTICATION
For constructing the user authentication templates, we extract the above-stated features randomly from 10 cycles of each subject. To construct the authentication and input templates, we apply the ratio of 70% to 30%, respectively. It calculates the average of each feature and stores an authentication template in the training phase of the user authentication process. \( F^A_1 \) represents feature 1 of the authentication template and the same for other features.

\[
F^A_1, F^A_2, F^A_3, F^A_4, F^A_5, F^A_6, F^A_7, F^A_8, F^A_9
\]

For the construction of the input template, we use three cycles. Then the average of each feature is calculated. \( F^I_1 \) represents the F1 of the input template and the same for other features.

\[
F^I_1, F^I_2, F^I_3, F^I_4, F^I_5, F^I_6, F^I_7, F^I_8, F^I_9
\]

To compare the input templates to the authentication templates, we introduce the feature matching algorithm. In Fig. 5 the flow chart of the template matching algorithm is elaborated. We use the threshold value range for the comparison. We select the lowest values and the highest values randomly to use their averages for setting the upper and lower limits as the threshold range. We follow the same process for matching each feature. We only compare the features from F1 to F9.

To compare the features of the input templates with the user authentication templates, we use the threshold values range. We select the lowest values and the highest values randomly to use their averages for setting the upper and lower limits as the threshold range. We follow the same process for matching each feature. We only compare the features from F1 to F9.

2) MACHINE LEARNING TRAINING & TESTING BASED USER AUTHENTICATION
Machine learning algorithm random forest uses the features as input. In the training process, we create models using 10 fold cross-validation and save them in the system as the unique user profiles. For authentication purposes, we apply the user-supplied data as test input.

VI. EXPERIMENTS AND RESULTS
A. PROPOSED FEATURE MATCHING ALGORITHM BASED EXPERIMENTS
For better visualization and presentation constraints, we present the experiments of 10 subjects to make the readers
FIGURE 6. Feature variation among user profiles (a) F1. (b) F2 (c) F3. (d) F4. (e) F5. (f) Kurt. (g) Skew. (h) Median.
understand the working of the proposed approach. We have randomly selected 10 gait cycles of each participant for the construction of profile templates. We store the user authentication templates as the mean value of the features to verify the subjects’ input templates.

Table 6 presents the values of the y-axis templates. The label subject id represents the participants, and F1 to F9 are the features calculated for each subject. Fig. 7 displays the variation in authentication feature templates of each subject.

**B. MACHINE LEARNING BASED EXPERIMENTS**

We also performed the experiments on the machine learning-based tools for validation of the proposed approach and comparison of the results. The random forest classifier performs better than other algorithms for activity and user Identification [21], [22]. We select 10 gait cycle features of each subject for training and testing purposes of the ratio of 70 to 30. To build the models, we apply the 10 fold cross-validation. The supplied test data consist of the input template features.

The confusion matrix in Table 7 presents the results of the random classifier. For better readability and understanding, we only present experiments of only 10 subjects. For calculating the accuracy of each user, we apply the method used by [22]. Using the confusion matrix, we can calculate True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN).

$$\text{Accuracy}_i = \frac{TP_i + TN_i}{TP_i + TN_i + FP_i + FN_i}$$

Here ‘i’ represents the subject id. To calculate the accuracy for each subject, it uses the values of the confusion matrix. TP$_i$ represents the value underlined where rows and columns of subject 7 intersect, i.e. TP$_7$ = 8. FP$_7$ = 1. FN$_7$ = 2. The sum of the TP of all subjects except subject 7 in this case is 27.
represents \( TN_i \), i.e. \( TN_i = 9 + 10 + 10 + 10 + 10 + 9 + 9 = 86 \). The accuracy of subject 7 is 96.90.

The comparison of the proposed approach is presented in Table 8 with the existing methods for gait based user authentication.

### VII. DISCUSSIONS

Smart IoT devices produce a diverse range of smart services to improve the quality of life in various domains. These devices also keep personal and sensitive data about users and lifestyle. The valuable data assets produced by IoT devices demands that the security of these devices must be ensured [35]. Authentication of users and these devices is essential. Specifically, security issues arise when operating in mobile environments [36]. These devices share data to the cloud that enhances the possible threats and authentication issues [37]. Therefore, continuous authentication can be very effective in such systems [38].

But the limited battery, computation, and communication capabilities of these smart devices must be taken into consideration. In this research work, we introduced a passive, continuous authentication technique using gait analysis. We have also utilized the least amount of data with a minimum number of features for producing highly accurate and reliable results. The results of the experiments are highly encouraging as compared with the existing approaches. Here we list a few limitations of the proposed approach in its current form.

1. We have not performed the experiments on the proposed approach in real-world scenarios, with a large number of users.
2. We have not yet evaluated the other threats on the proposed approach.
3. We have only evaluated our proposed approach to the normal walking style.
4. The technical issues are the major barrier that requires a proper solution for the practical implementation of the proposed authentication technique.

### VIII. CONCLUSION AND FUTURE WORK

In this research paper, we have presented a novel user authentication approach based on gait analysis using sensor data. For user authentication like other biometric methods, sensor data-based gait profiles are a valuable and potential solution. The proposed approach utilizes the ECG signal processing technique for the creation of sensor data-based gait templates with the least amount of data samples. We have taken into consideration the energy constraints of the smart IoT devices. We have introduced a feature matching algorithm. To perform the cross-validation of the proposed approach features, we have performed the experiments using the features on the machine learning algorithm. We have introduced a threat model-specific to the proposed approach.

We have resolved major threats such as impersonation attacks effectively. In future work, we will address the limitations of the proposed approach and implement it on a large number of subjects. We will customize the gait based profiles to be more distinct and unique for each subject. The extracted features are also helpful for deeper analysis of the human gait for various domains such as healthcare etc.

### REFERENCES

[1] A. K. Das, M. Wazid, N. Kumar, A. V. Vasilakos, and J. J. P. C. Rodrigues, “Biometrics-based privacy-preserving user authentication scheme for cloud-based industrial Internet of Things deployment,” *IEEE Internet Things J.*, vol. 5, no. 6, pp. 4900–4913, Dec. 2018.
[2] B. Guo, D. Zhang, Z. Wang, Z. Yu, and X. Zhou, “Opportunistic IoT: Exploring the harmonious interaction between human and the Internet of Things,” J. Netw. Comput. Appl., vol. 36, no. 6, pp. 1531–1539, Nov. 2013.

[3] K. Mansoor, A. Ghanzi, S. Chaudhry, S. Shamshirband, S. Ghayyur, and A. Mosavi, “Securing IoT-based RFID systems: A robust authentication protocol using symmetric cryptography,” Sensors, vol. 19, no. 21, p. 4752, Nov. 2019.

[4] G. D. Clark and J. Lindqvist, “Engineering gesture-based authentication systems,” IEEE Pervas. Comput., vol. 14, no. 1, pp. 18–25, Jan. 2015.

[5] N. Shone, C. Dobbins, W. Hurst, and Q. Shi, “Digital memories based mobile user authentication for IoT,” in Proc. IEEE Int. Conf. Comput. Inf. Technol., Ubiquitous Comput. Commun., Dependable, Autonomic Secure Comput., Pervasive Intell. Comput., Oct. 2015, pp. 1796–1802.

[6] J. Lin, W. Yu, N. Zhang, X. Yang, H. Zhang, and W. Zhao, “A survey on Internet of Things: Architecture, enabling technologies, security and privacy, and applications,” IEEE Internet Things J., vol. 4, no. 5, pp. 1125–1142, Oct. 2017.

[7] S. A. Chaudhry, T. Shon, F. Al-Turjman, and M. H. Alsharif, “Correcting design flaws: An improved and cloud assisted key agreement scheme in cyber physical systems,” Comput. Commun., vol. 153, pp. 527–537, Mar. 2020.

[8] S. Y. Jung, J.-W. Kim, H. Hwang, K. Lee, R.-M. Baek, H.-Y. Lee, S. Yoo, W. Song, and J. S. Han, “Development of comprehensive personal health records integrating patient-generated health data directly from Samsung S-Health and apple health apps: Retrospective cross-sectional observational study,” JMIR mHealth uHealth, vol. 7, no. 5, May 2019, Art. no. e12691.

[9] Samsung Health. Accessed: May 29, 2020. [Online]. Available: https://play.google.com/store/apps/details?id=com.sec.android.app.shealth.

[10] M. M. Rodgers, V. M. Pai, and R. S. Conway, “Recent advances in wearable sensors for health monitoring,” IEEE Sensors J., vol. 15, no. 6, pp. 3119–3126, Jun. 2015.

[11] D. Gil, A. Fernández, H. Mora-Mora, and J. Peral, “Internet of Things: A review of surveys based on context aware intelligent services,” Sensors, vol. 16, no. 7, p. 1069, Jul. 2016.

[12] H. M. Thang, V. Q. Viet, N. D. Thuc, and D. Choi, “Gait identification using accelerometer on mobile phone,” in Proc. Int. Conf. Control, Autom. Inf. Sci. (ICCAIS), Nov. 2012, pp. 344–348.

[13] M. W. Whittle, Gait Analysis: An Introduction. London, U.K.: Butterworth, 2014.

[14] H. Lu, J. Huang, T. Saha, and L. Nachman, “Unobtrusive gait verification for mobile phones,” in Proc. ACM Int. Symp. Wearable Comput. (ISWC), 2014, pp. 91–98.

[15] I. Kwapisz, G. M. Weiss, and S. A. Moore, “Cell phone-based biometric identification,” in Proc. 4th IEEE Int. Conf. Biometrics, Appl. Syst. (BTAS), Sep. 2010, pp. 1–7.

[16] S. Batool, N. A. Saqib, and M. A. Khan, “Internet of Things data analytics for user authentication and activity recognition,” in Proc. 2nd Int. Conf. Fog Mobile Edge Comput. (FMEC), May 2017, pp. 183–187.

[17] J. Jenkins and C. Ellis, “Using ground reaction forces from gait analysis: Body mass as a weak biometric,” in Proc. Int. Conf. Pervas. Comput. Berlin, Germany: Springer, 2007, pp. 251–267.

[18] S. Vahdati and C. Poellabauer, “Wearable device user authentication using physiological and behavioral metrics,” in Proc. IEEE 28th Annu. Int. Symp. Pers., Indoor, Mobile Radio Commun. (PIMRC), Oct. 2017, pp. 1–6.

[19] J. S. Arteaga-Falconi, H. Al Osman, and A. El Saddik, “ECG authentication for mobile devices,” IEEE Trans. Instrum. Meas., vol. 65, no. 3, pp. 591–600, Mar. 2016.

[20] S. Terada, Y. Enomoto, D. Hanawa, and K. Oguchi, “Performance of gait authentication using an acceleration sensor,” in Proc. 34th Int. Conf. Telecommun. Signal Process. (TSP), Aug. 2011, pp. 34–36.

[21] M. N. Malik, M. A. Azam, M. Ehtasham-ul-Haq, W. Ejaz, and A. Khalid, “ADLAuth: Passive authentication based on activity of daily living using heterogeneous sensing in smart cities,” Sensors, vol. 19, no. 11, p. 2466, May 2019.

[22] P. Musale, D. Baek, N. Werrellagama, S. S. Woo, and B. J. Choi, “You walk, we authenticate: Lightweight seamless authentication based on gait in wearable IoT systems,” IEEE Access, vol. 7, pp. 37883–37895, 2019.

[23] M. Muazz and R. Mayrhofer, “Smartphone-based gait recognition: From authentication to imitation,” IEEE Trans. Mobile Comput., vol. 16, no. 11, pp. 3209–3211, Nov. 2017.

[24] Z. Huan, X. Chen, S. Lv, and H. Geng, “Gait recognition of acceleration sensor for smart phone based on multiple classifier fusion,” Math. Problems Eng., vol. 2019, pp. 1–17, Jun. 2019.

[25] G. Wu, J. Wang, Y. Zhang, and S. Jiang, “A continuous identity authentication scheme based on physiological and behavioral characteristics,” Sensors, vol. 18, no. 1, p. 179, Jan. 2018.

[26] J. Blasco and P. Peris-Lopez, “On the feasibility of low-cost wearable sensors for multi-modal biometric verification,” Sensors, vol. 18, no. 9, p. 2782, Aug. 2018.

[27] N. Kumar, N. Chilamkurti, and S. Misra, “Bayesian coalition game for the Internet of Things: An ambient intelligence-based evaluation,” IEEE Commun. Mag., vol. 53, no. 1, pp. 48–55, Jan. 2015.

[28] E. Curry, W. Dergeuch, S. Hasan, C. Kourruptopoulou, and U. H. Hassan, “A real-time linked dataspaces for the Internet of Things: Enabling ‘pay-as-you-go’ data management in smart environments,” Future Gener. Comput. Syst., vol. 90, pp. 405–422, Jan. 2019.

[29] A. Yassine, S. Singh, M. S. Hossain, and G. Muhammad, “IoT big data analytics for smart homes with fog and cloud computing,” Future Gener. Comput. Syst., vol. 91, pp. 563–573, Feb. 2019.

[30] C. Lin, D. He, N. Kumar, K.-K.-R. Choo, A. Vinel, and X. Huang, “Security and privacy for the Internet of Drones: Challenges and solutions,” IEEE Commun. Mag., vol. 56, no. 1, pp. 64–69, Jan. 2018.

[31] J. R. Kwapisz, G. M. Weiss, and S. A. Moore, “Activity recognition using cell phone accelerometers,” ACM SIGKDD Explorations Newsl., vol. 12, no. 2, pp. 74–82, Mar. 2011.

[32] D. Anguita, A. Ghio, L. Oneto, X. Parra, and J. L. Reyes-Ortiz, “A public domain dataset for human activity recognition using smartphones,” in Proc. Ensam, 2013, pp. 1–6.

[33] Moving Average in Pandas. Accessed: May 29, 2020. [Online]. Available: https://www.datacamp.com/community/tutorials/moving-averages-in-pandas.

[34] M. O. Derawi, C. Nickel, P. Bours, and C. Busch, “Unobtrusive user-authentication on mobile phones using biometric gait recognition,” in Proc. 6th Int. Conf. Intell. Inf. Hiding Multimedia Signal Process., Oct. 2010, pp. 306–311.

[35] A. Ghanzi, K. Mansoor, S. Mehmood, S. A. Chaudhry, A. U. Rahman, and M. N. Saqib, “Security and key management in IoT-based wireless sensor networks: An authentication protocol using symmetric key,” Int. J. Commun. Syst., vol. 32, no. 16, p. e4139, 2019.

[36] A. K. Das, M. Wazid, N. Kumar, M. K. Khan, K.-K.-R. Choo, and Y. Park, “Design of secure and lightweight authentication protocol for wearable devices environment,” IEEE J. Biomed. Health Inform., vol. 22, no. 4, pp. 1310–1322, Jul. 2018.

[37] S. Kumar, M. Karuppiah, A. K. Das, X. Li, F. Wu, and N. Kumar, “A secure authentication scheme based on elliptic curve cryptography for IoT and cloud servers,” J. Supercomput., vol. 74, no. 12, pp. 6428–6453, Dec. 2018.

[38] A. J. A. Majumder, Y. A. ElSaadany, R. Young, and D. R. Ucci, “An energy efficient wearable smart IoT system to predict cardiac arrest,” Adv. Hum.- Comput. Interact., vol. 2019, pp. 1–21, Feb. 2019.

[39] A. H. Johnston and G. M. Weiss, “Smartwatch-based biometric gait recognition,” in Proc. IEEE 7th Int. Conf. Biometrics, Appl. Syst. (BTAS), Sep. 2015, pp. 1–6.
**ALI HASSAN** received the B.E. and M.S. degrees in computer engineering from the College of Electrical and Mechanical Engineering (CEME), NUST, Islamabad, Pakistan, in 2004 and 2007, respectively, and the Ph.D. degree in electrical engineering from the University of Southampton, U.K., in 2012. He is currently the Head of Department of the Department of Computer and Software Engineering, College of Electrical and Mechanical Engineering, NUST. His areas of specialization are machine learning and speech processing.

**NAZAR ABBAS SAQIB** (Member, IEEE) received the M.Sc. and M.Phil. degrees in electronics from Quaid-i-Azam University, Islamabad, in 1993, and the Ph.D. degree in electrical engineering from CINVESTAV-IPN, Mexico. He is currently working as a Full Professor with the College of Computer Science and Information Technology (CCSIT), Imam Abdulrahman Bin Faisal University (IAU), Dammam, Saudi Arabia. He has coauthored a book titled *Cryptographic Algorithms on Reconfigurable Hardware* with Springer New York, in 2005. His research interests include computer and communication security, cryptographic hardware, and FPGA based system design.

**MUAZZAM A. KHAN KHATTAK** (Senior Member, IEEE) received the Ph.D. degree in computer sciences as a sandwich program from IIUI and UMKC, USA, in 2011. He was a Postdoctoral Researcher from the University of Missouri, Kansas City, MO, USA, in 2016. He joined the School of Electrical Engineering and Computer Science (SEECs), NUST, Islamabad, Pakistan, as an Assistant Professor, in 2013, and promoted to an Associate Dean and a Tenured Associate Professor in 2017. He has been with the School of Computer Science, University of Ulm, Germany, and with the Networking and Multimedia Lab, School of Computer and the Electrical Engineering, University of Missouri, as a Research Fellow. He is currently working as a Tenured Associate Professor with the Department of Computer Science, Quaid-i-Azam University, Islamabad. His research interests include the Internet of Things, next generation intelligent networks, blockchain based information and network security, vehicular ad-hoc networks, and acoustic networks.

* * *