IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received April 3, 2022, accepted May 3, 2022, date of publication May 27, 2022, date of current version June 8, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3178635

IoTAuth: IoT Sensor Data Analytics for User Authentication Using Discriminative Feature Analysis

SAMERA BATOOL[1], ALI HASSAN[1], MUAZZAM A. KHAN KHATTAK[2], (Senior Member, IEEE), AHSAN SHAHZAD[1], AND MUHAMMAD UMAR FAROOQ[1]

[1]Department of Computer & Software Engineering, College of Electrical and Mechanical Engineering, National University of Sciences and Technology, Islamabad 44000, Pakistan

[2]Department of Computer Sciences, Quaid-i-Azam University, Islamabad 45320, Pakistan

Corresponding author: Samera Batool (samera.batool@ce.ceme.edu.pk)

ABSTRACT The revolution of IoT highly impacts on different applications such as remote sensing, smart cities, and remote digital healthcare. People use IoT devices for performing business transactions, daily tasks, and healthcare monitoring. IoT devices generate huge amounts of data assets that have potential applications. Biometrics is a potential application of sensors data. The traditional biometric methods such as PINs, passwords are exposed to numerous attacks such as replication, repeated passwords, etc. Sensors' data-based continuous authentication methods are suitable for maintaining users' privacy and security in mobile IoT systems. Most of the existing authentication methods have applied motion-based sensors for building users' identification profiles. The proposed method uses motion sensors and biomedical sensors for reliable and multi-factor user authentication. In this article, we have introduced an IoT sensors data analytics framework to construct user authentication models. We apply the fiducial points-based feature extraction method data for extracting discriminative features. These features act as unique user profiles for authentication purposes. We have performed a detailed analysis of the proposed approach using the publically available datasets. The experiments elaborate on the effectiveness of IoTauth for improved authentication results.

INDEX TERMS Authentication, biometrics, data analytics, Internet of Things, sensors.

I. INTRODUCTION

In the last two decades, the information technology industry has progressed remarkably. Technologies such as the Internet of Things (IoT), sensor networks, and wearable devices have exhibited tremendous progress. Internet connectivity is widespread in most remote areas of the world, with declined costs [1]. IoT vision defined more than five years ago has become a reality. The number of connected devices has reached up to 20 billion by the end of 2020. Moreover, in the last few months, the ongoing pandemic situation due to COVID-19, use of smart IoT devices, and the internet have increased tremendously for various applications such as online shopping, remote sensing, grocery, educational purposes, and personal health assistance [2]. Fig. 1 describes the uses of smart IoT devices in our daily routine, especially after the COVID-19 pandemic [3]. For example, healthcare systems use body-worn sensors to collect information. Robots perform efficiently in industrial applications, drones for crowd monitoring, and many other smart applications for personal activities assistance. People use smart devices to manage most of their routine tasks. It is also used for remote health monitoring avoiding unnecessary visits to hospitals. Performing daily life activities such as counting walk steps, tracking sleep hours, tracking calories, paying bills, purchasing things, finding directions, and healthcare-related information management, etc. IoT devices generate huge personal data assets containing unique patterns that can be utilized for obtaining meaningful information such as the activity of a subject etc. Data stored on these devices need protection for the security and privacy of the users.

Security issues have risen due to the widespread usage of IoT devices [9]. Authentication is a vital means for securing access to devices. Few widely used authentication methods

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

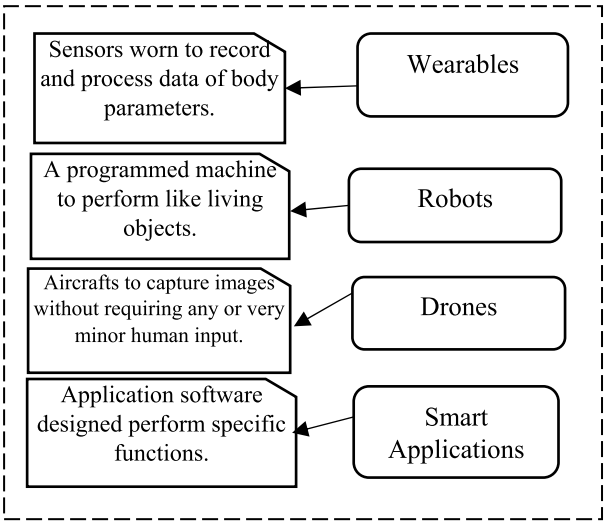

FIGURE 1. Applications of IoT sensors

include passwords, patterns lock, fingerprints, etc., which are vulnerable to security issues such as replication attacks, simple passwords, reusing the same passwords, stealing passwords, guessing patterns, etc. [4].

Effective user authentication methods are required to utilize and secure users' data passively without requiring the direct intervention of the users [5].

The behavioral models have generally been used in mobile applications for passive and continuous authentication. They improve the performance of user authentication by applying user activity patterns [6]. We introduce an IoT data analytics framework that uses sensors data for constructing users and activity recognition models by applying classification algorithms, i.e. machine learning, and deep learning, etc. These models act as the unique features of a subject, which we can verify later with real-time data for validation purposes.

These unique patterns are not easy to duplicate [6]. Smart IoT devices easily capture these signals. It enables the implementation of the proposed approach quite applicable in real-life environments. Many efforts in the literature use heart rate patterns for user identification. It is considered a unique biological human body feature that has the characteristics to change with the movement of the human body while performing different activities [7]. The application of biomedical sensors for user authentication is relatively a new field. The heterogeneous nature of IoT data and the environmental noise interruptions are a few obstacles in the practical implementation of IoT-based authentication systems.

The availability of smart devices with embedded sensors measures different user's specific activity and physiological patterns. In existing research, most of the works deploy accelerometer and gyroscope sensors data for user authentication purposes [8]. Most authentication techniques rely on physiological biometric like face matching, iris recognition, and fingerprints, etc. These characteristics can be duplicated and also change with time [9].

Few of these require additional hardware support and constant input from users. Moreover, the most commonly used password-based user authentication has its drawbacks, such as reusing the same passwords for multiple accounts and difficulties in memorizing many passwords for different accounts [10]. The sensor's data-based biomedical characteristics have the potential to overcome the above-stated challenges and maintain user's security and privacy [11].

In the proposed approach, we consider and address a few of the stated challenges. We have used the heart rate, pulse oximeter, and glucometer sensors along with motion sensors. These sensors' measurements provide critical information about a person's health status and unique characteristics that can be used to provide customized services. Through the proposed approach, we make the following main contributions.

- We propose an IoT data analytics framework for user authentication and activity recognition.
- We have applied the fiducial point's analysis method to extract a small set of statistical features with high accuracy and are computationally efficient.
- We performed the experimental evaluation of the proposed approach on real data set.
- We implement the proposed approach on a real application scenario for practical evaluation.

We organize the rest of the paper as follows. In Section II, we present the related work. Section III describes the proposed approach. In Section IV, we present the experiments and results. Section V is composed of the relevant discussions. Section VI establishes the conclusion and future work.

II. RELATED WORK

Sensors' data-based biometrics for user authentication has attracted the researcher's attention. The sensor-based authentication methods using physiological and behavioral biometric data have produced tremendous results [12]. Most of these methods deploy embedded motion sensors, such as accelerometers and gyroscopes [13]. The sensor-based authentication methods collect the user devices' sensor data to extract distinct patterns.

Additional biomedical sensors such as heart rate sensors, fingerprints sensor, oxygen saturation sensor embedded in IoT devices have enabled the use of such characteristics for user identification purposes. The user authentication models are of the following two categories (i) biometrics (face, fingerprints, iris, ECG, Heart rate, etc.) and (ii) behavioral biometrics (motion-based, touch gestures, activity patterns, etc.). In this section, we have discussed the sensor data-based user authentication methods under these two main categories.

A. PHYSIOLOGICAL BIOMETRIC USER AUTHENTICATION

Many biological characteristics have proved to be unique for each subject. In [14], the authors have evaluated the existing research on the use of ECG-based biometrics. They presented the state-of-the-art technique review and also highlighted the gaps and future research possibilities in this field. They also

evaluated the proposed ECG-based user authentication on different data sets and gave tremendous results. The other techniques that also used ECG signals for user identification include [15].

Latest models of smartphones and other wearable devices have embedded fingerprint sensors. These sensors' measurements are highly accurate and perform efficiently for user authentication. But they are vulnerable to various security risks such as spoofing attacks and replication of fingerprints.

In [16], the authors introduced user authentication based on multimodal physiological signals. They deployed four types of sensors signals perinasal perspiration (PER-EDA), palm EDA (P-EDA), heart rate (HR), and breathing rate (BR). The convolutional Neural Network with mono-dimensional convolution (ID-CNN) takes an input window of raw signals. They performed experiments on data set of multimodal signals from 37 subjects. It produced an accuracy of 99.51%. The limitation of the approach includes the use of a single model for experiments. The use of multimodal and uncontrolled environment data collection may test the performance.

In [17], the authors have used the fusion of accelerometer and ECG signals to construct activity models. Those activities are targeted that consume low energy or consume energy without any movement. They have used the data set from 13 subjects of the following set of activities, sitting, standing, ascending, resting, walking, and running. The accuracy of 96.35% is achieved, which is much higher than the accelerometer sensor data-based activity recognition solely. The authors also support the evidence for using human physiological data for activity recognition. ECG data have been utilized for user authentication purposes due to their unique characteristics in each person.

In [18], the authors use the chest-worn ECG sensors data of various activities. Four participants contributed to the data collection process. The experimentation of the proposed technique achieved an error rate of 6% to 13%, which can be considered well controlled. The sample size for experiments is small, and large data sets and uncontrolled data collection settings might affect the performance of the proposed approach.

In [19], the authors have used multimodal physiological sensors for user authentication purposes. The authors utilized the signals of airflow, ECG, and galvanic skin reaction. The rotation ensemble algorithm was applied to the data set of 6 subjects for experiments. It produced an accuracy of 99.6%. The small data set raises few concerns, such as the consistency in performance for a larger population in real-time environments.

B. BEHAVIORAL BIOMETRIC

Users' activities and motion-related sensor data for authentication purposes lies under behavioral biometrics. With advancements in sensor technology, more types of sensors have been explored for different applications with high accuracy [1].

In a previous research effort, we used the accelerometer sensor data for activity recognition and user identification purpose. To record the single accelerometer sensor data of the following activities standing, walking, and running. The results produced by the classification experiments using the random forest classifier are highly encouraging. But there is a certain level of overlapping in various classes of users and activities. We addressed the class overlapping by adding the heart rate sensor data along with the physical activity data set from motion sensors in [7]. The results produced with the addition of the heart rate sensor data set have reduced certain overlapping between classes compared to the single motion sensor. In IoT-based environments, the energy consumption of IoT devices is also of great importance for activity recognition.

In [20], the authors have presented the precise details of the existing classification methods of accelerometer sensor data sets. They discussed classification methods such as the decision tree, dynamic time wrapping, and support vector machines. The SVM classifier produces an accuracy of 90% in terms of activity recognition. It is higher than the other two classifiers. The gait-based user authentication techniques use the gait features for authentication purposes. In [21], the authors have introduced a gait-based authentication technique. The data is collected using the ankle-worn accelerometer sensor. The scoring system applies the threshold values calculated using the gait-related features. An EER of 20% produced. 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. THE PROPOSED APPROACH

We introduce IoTauth, a sensor-data-based authentication method using IoT sensors. The accelerometer and gyroscope are the most widely used IoT-based sensors for activity-based user authentication [20]. Biomedical sensors such as heart rate, ECG, and glucometer also contain unique patterns to support user identification [24]–[25].

We consider the unique sensor data patterns of each user's activities and physical biometrics produced from their smart devices. It provides an additional security cover for remote healthcare systems to improve user's privacy.

Fig. 2 describes the proposed approach and its components. Several efforts have utilized individual biomedical sensors such as ecg, heart rate, etc., for user identification. Table 1 presents the list of sensors of the mysignals-ehealth. These sensors assist in the diagnosis and treatment of many critical health conditions as well [27]–[28]. IoT sensor data analytics consists of the following phases such as data acquisition, preprocessing, and feature extraction. Here we describe each component according to the proposed method.

A. IoT SENSOR DATA ACQUISITION

The extensive usage of the internet and smart IoT devices makes it feasible collecting data about a person's lifestyle, behavior, activity patterns and biomedical features.

The MySignals-eHealth IoT KiT [26] offers a wide range of biomedical sensors for data collection and also provides a platform for building IoT applications for healthcare. Several efforts in the literature have deployed individual biomedical sensors such as ECG, heart rate, etc., for user identification. The proposed work is the first effort of its sort that comprehensively introduces a platform for IoT based biomedical sensor data analytics for user authentication and activity recognition. Raw sensory data produced from IoT devices is in continuous or non-continuous form that contains noise and unstructured data. The IoT devices produced data while normally operating in diverse environments.

B. DATA PREPROCESSING

1) DATA FILTERING/DATA SEGMENTATION

The processing of large IoT data sets that comprise noisy and unwanted data is a loss of resources in the resource-constrained IoT environment [29]. The data filtration phase discards the unwanted data. It reduces further processing, storage, and computational overhead. Table 1 presents the set of selected parameters for the data preprocessing phase. We divide the streaming data into the time sliding window. Fig 4 highlights the prominent points of each data cycle.

TABLE 1. List of sensors for data extraction.

Sensor	Physical parameter	Measurement Attribute
Herat rate monitor	Heart rate BPM	Heart rate in BPM
Pulseoximeter	Oxygen Saturation	Level of Oxygen Saturation
Glucometer	Glucose Mg	Blood Glucose level

2) NOISE REMOVAL

sensor data contains noise produced from the environmental sources and movement of other body parts. to remove the noise from the sensor data we use the smoothing moving average filter of the order 3. the filter is computationally efficient and removes the undesired noise from the sensors data

3) DATA LABELING

The Data Profiles Of Each Subject Contain The Sensor Measurement Produced By Their Devices. They Are Properly Arrange In Labeled Form.It Helps In Better Management Of The Data Set, Extract Subject-Related Features And Improve Performance [31].

C. FEATURE EXTRACTION

The fiducial points represent prominent characteristics in a data cycle, such as the peaks, valleys, minimum value of a signal, and the smooth path as an average value. these fiducial points elaborate the patterns in each data cycle. unique features based on these fiducial point forms the user authentication profiles. the sensors' readings are distinct for each subject, depending upon the inherent health conditions and biometric characteristics. the fiducial points-based features highlight sudden variations, extreme points, and smooth average values. they correspond to the relevant statistical features.

We extract the following features from each data segment and store them in corresponding feature vectors. Below we describe each feature.

1) MEAN

The mean feature represent the normal trend of a series and help in predicting future trends [33]. For each type of sensor, we extract the mean values of each data cycle, i.e., $\bar{H}, \bar{G}, \overline{O\chi}$ and accelerometer sensor's each axis. There is a corresponding vector for storing the value to each subject profile. The following equation represents the formula for the calculation of mean value.

$$\text{Mean}(\bar{S}) = \frac{1}{\text{N}} \sum_{\text{k=1}}^{N} \text{S}_\text{k} \tag{1}$$

Here $\bar{S}$ represents the signal, N represents the total no of values and k represents the current value of S.

2) STANDARD DEVIATION

To represent the consistency and patterns in a signal, we use standard deviation. How much the signals lies in the expected range. This feature represents the variation in normal patterns and changes in the heart rate of a subject to detect the abnormal scenarios. The decreased variation shows the stability of the subject heart rate patterns. We have calculated standard deviation for each sensor reading i.e. heart rate, oxygen saturation, glucose level and accelerometer sensor. It also shows the physical health status of a subject with unique traits.

$$\sigma = \sqrt{\frac{1}{\text{N}} \sum_{\text{i=1}}^{\text{N}} (\text{x}_\text{i} - \mu)^2} \tag{II}$$

Here N represents the total number of values, and i is the current value of x, whereas x is the set of values of sensor measurements. μ represents the mean of x value. The square root of x is sum up and divided by zero. The used symbols and notations are presented in the following Table 3

3) KURTOSIS

The kurtosis measures the sharpness and height of the peaks in a signal. Kurtosis of heart rate signal measures the intensity of the heart rate during different activities [34]. We calculate the kurtosis value of each sensor data cycle. The following formula in equation V is used for calculating $Kurt_H$.

$$\text{Kurt}_\text{H} = \frac{\sum_{\text{i=1}}^{\text{N}} \frac{(\text{H}_\text{i} - \overline{\text{H}})}{\text{N}}}{\text{s}^4} \tag{III}$$

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

TABLE 2. Existing biometric user authentication methods and their limitations.

Approach/Year	Physiological /Behavioral	Sensors Used	No of Subjects	Types of Features Extracted	Accuracy	Limitations
[21]/2011	Behavioral	Accelerometer	10	Gait Cycle Analysis	20% EER	• Small data set for analysis with very few aspects no of gait cycle are tested
[19] /2015	Physiological	Airflow, ECG, galvanic skin response sensor	6	Time domain and Frequency domain	99.6%	• Testing in real time uncontrolled environments is required • Larger data sets should be used for validation of the proposed approach
[17]/2016	Behavioral	Accelerometer	(6 activities * 100 samples each)	Fast Fourier Transform coefficient	92.17%	• The focus of the experiments is only in terms of energy efficiency
[8] /2017	Behavioral	Accelerometer	19	Time and Frequency domain	93%	• Large data sets with more number of activities needs to be tested
[22]/2017	Behavioral	Accelerometer	35	Dynamic Time Wrapping/Local Maxima		• Only the Impersonation attack based analysis is performed for continuous authentication
[26] /2017	Behavioral	Accelerometer	35	Gait segments/ Fiducial points	EER 13%	• Performance evaluation in different real time environments is required
[27]/2018	Behavioral	Accelerometer, Photoplethysmography (PPG)	40	Time domain and Frequency domain	98.5%	• Additional overhead of multiple classification models selection
[16] /2019	Physiological	ECG	29	Fiducial points based distance features	97%	• The scalability of data set and features on the performance needs further testing • Longer periods feature stability requires to be tested
[18] /2019	Physiological	heart rate (HR), breathing rate (BR), palm electro dermal activity (P-EDA), and perinasal perspiration (PER-EDA)	37	Time and frequency domain features	99%	• The data set collected in uncontrolled environment needs evaluation • More activity scenario needs to be tested
[44] /2020	Behavioral	Accelerometer	50	Cross correlation	11. 2	• Need to check the performance of the proposed approach in a multi-sensor environment for the same user
[45]/2021	Behavioral	Accelerometer	30	LSTM based Features	7.9	• The result proves as individual training produces better results with respect to the aggregation of the walking types.

4) HEART RATE VALLEYS H_{VAL}

The set of minimum values that in a data cycle represents valleys. The deepest valley represents the minimum value in a cycle. It is stored as a unique attribute in the profile of each subject. We calculate the valleys of each sensors data cycle.

If H is a series of heart rate values. It is denoted as H_{val}.

5) HEART RATE PEAKS H_{PEAK}

The maximum value represents the highest peak in a data cycle. These values are also distinct for each subject based on physical characteristics and activity patterns. We store the maximum value of the heart rate as a unique feature. We also calculate Peak values of data cycle for sensors. It is denoted as

If H is a series of heart rate values. It is denoted as H_{peak}.

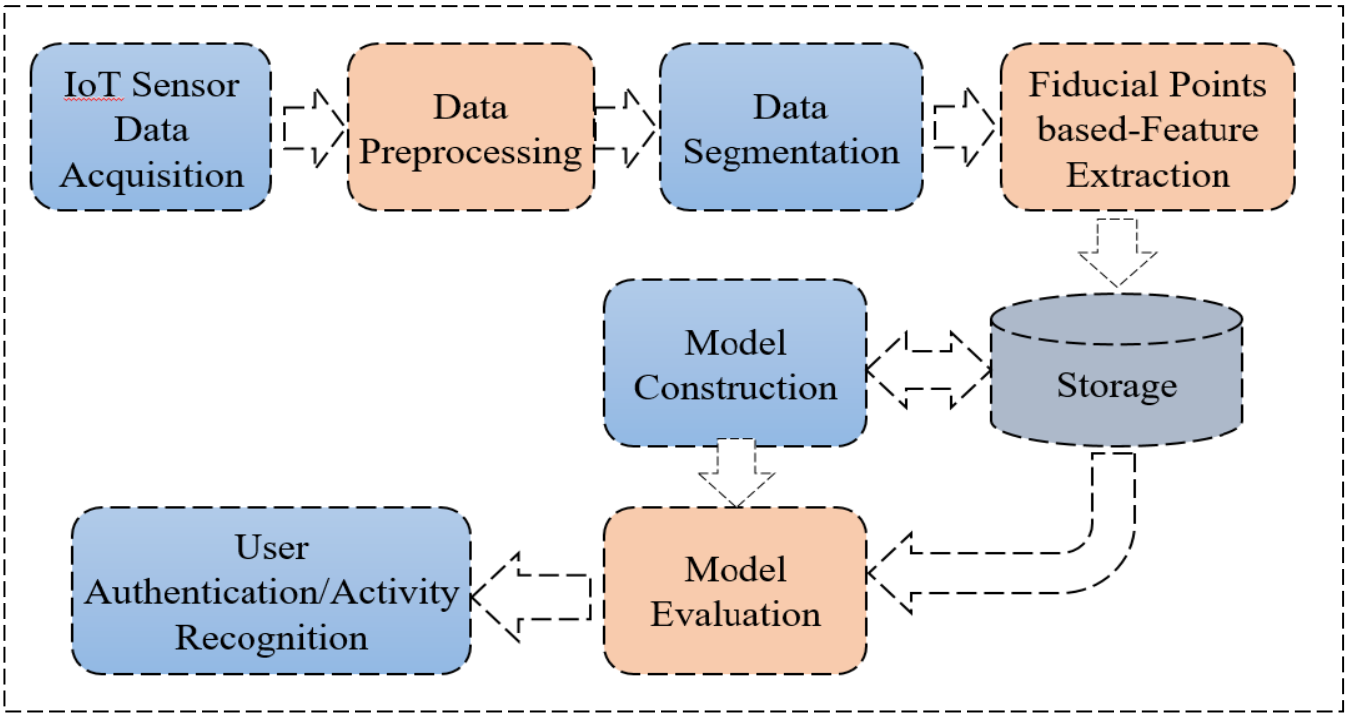

FIGURE 2. The proposed framework for IOT based user authentication.

6) VARIANCE

The following equation is used to calculate the variance of the signal.

$$\text{var}(S) = \frac{1}{N}\sum_{k=1}^{N}\left(S_k - \bar{S}\right)^2 \tag{IV}$$

7) MEDIAN

The median value is calculated for all the sensor data cycles. It is used to seperate the lower and higher half data cycle.

8) MEDIAN ABSOLUTE DEVIATION

The median absoloute deviation value is used to calculate the average distance between each data point and the mean. We calculate this feature for motion sensor each axis data.

$$\text{mad}(S) = \text{median}\left(\left|\left(S_k - \bar{S}\right)\right|\right) \tag{V}$$

Here S represents the signal and k.

9) INTERQUARTILE RANGE

The inquartile range measures the fifty percent of the middle values in a data set. We calculate it for each axis of accerlerotmer data cycle. Following formula is used to calculate the value.

$$\text{iqr}(S) = Q3(S) - Q1(S) \tag{VI}$$

TABLE 3. Result Matrics for the experiment on MySignals eHealth Dataset.

Accuracy	Precision	EER	Specificity
88%	0.86	12%	86%

The interquartile range is found using the quartiles. It subtracts the first quartile from the third quartile.

10) AUTOREGRESSION (AR) COEFFICIENTS

This is a model of time series data that learns from the previous time steps as input to a regression equation and predicts the value for next time. We calculate this value for each axis of accelerometer sensor data.

The equation for auto regression calculation is presented below.

$$a = \text{arburg}(S, 4), \quad a \in R \tag{VII}$$

Here we use arburge function of Matlab for calculation of autorgression.

11) ENERGY

The energy of a signal is measured as the signal strength. It is calculate using the formula below.

$$E_f(S) = \Sigma|S|f\|^2 \tag{VIII}$$

TABLE 4. Results Matric for the user authentication experiment on MySignals Dataset.

Subject Id	Subject 1	Subject 2	Subject 3	Subject 4	Subject 5	Subject 6	Subject 7	Subject 8	Subject 9	Subject 10
Subject 1	**50**	0	0	0	0	0	0	0	0	0
Subject2	0	**48**	0	0	0	1	0	10	0	0
Subject3	1	2	**47**	0	0	0	0	0	0	0
Subject4	0	0	0	**50**	0	0	0	0	0	0
Subject5	0	1	0	0	**48**	0	0	1	0	0
Subject6	0	0	0	0	0	**50**	0	0	0	0
Subject7	0	1		1	0	0	**46**	0	2	0
Subject8	2	0	1	0	0	0	0	**47**	0	0
Subject9	3	0	5	2	2	3	1	1	**30**	3
Subject10	6	5	2	7	1	3	1	5	1	**22**

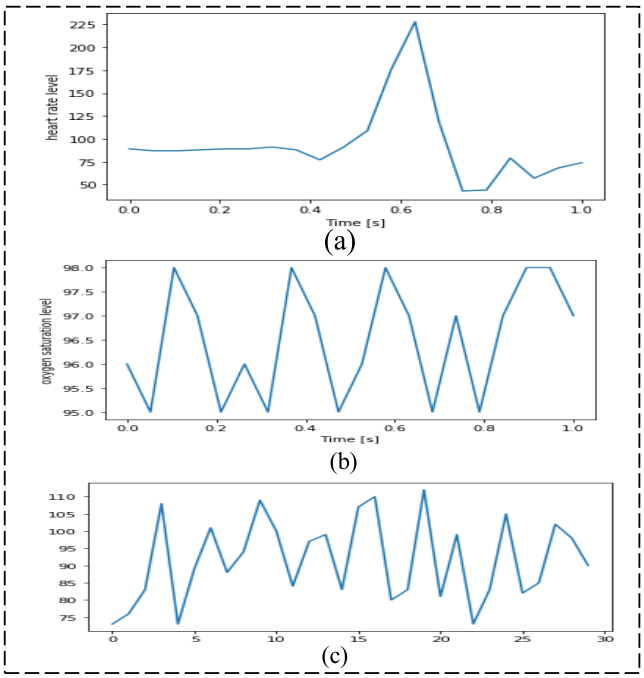

FIGURE 3. Fiducial points (a) Heart rate (b) Oxygen Saturation (c) Glucose mg.

12) SUM OF PEAKS

In this feature we calculate the sum of all the peak values of a data cycle. it is calculated for each axis of accelerometer sensor data.

We performed an empirical analysis and extracted the following set of statistical features from motion sensors data. We present the list of selected features in the following table 5. We have extracted each feature along three axis of motion sensors i.e (x, y, z) in this section we describe only those features that are not discussed for biomedical sensors features.

a: MODEL CONSTRUCTION

In this step, we construct the models for user authentication from the extracted features set. This phase applies machine learning tools. We give input of features. We have applied weka for constructing unique models for each user for authentication purposes.

TABLE 5. WISDM dataset description.

Devices	No of people	Age group	Activities
Smart Phone	36	30-40	6 Basic ADL

TABLE 6. Results metric of the experiment on WISDM dataset.

Accuracy	Precision	EER	Specificity
99%	0.94	1%	99%

b: USER AUTHENTICATION

In the user authentication phase, We compare the saved features profile models with real-time data of the users

IV. EXPERIMENTS & RESULTS

In this section, we elaborate the validation of the proposed approach through experiments to collect results.

Data set Description.

A. MYSIGNALS-EHEALTH DATASET

For the collection of biomedical sensors data we have used the MySignals-eHealth IoT Kit. It contains several biomedical sensors for the measurement of different body parameters. We have used the biomedical sensors listed in the Table 1. The participants measured the sensor readings using the sensor device and transfer to the Mysginals-eHealth cloud storage.

We saved the sensor readings of each subject against their ids in excel. Each data cycle consists of a fixed-length window of 5 seconds from which we extract data samples. It also reduces the computational overhead of processing huge features set. We extracted the features listed in Table 3 for each

TABLE 7. Confusion matrix for the experiment on WISDM dataset.

Subject Id	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
A	64	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B	0	64	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C	0	0	59	2	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
D	0	0	0	52	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
E	0	0	0	0	65	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
F	0	0	0	0	0	69	0	0	0	0	0	0	0	0	0	0	0	0	0	0
G	1	0	0	0	0	0	56	0	0	0	0	0	0	0	0	0	0	0	5	0
H	0	0	0	0	0	0	0	62	0	0	0	0	0	0	0	0	0	0	0	0
I	0	0	0	0	0	0	0	0	87	0	0	0	0	0	2	0	0	0	0	0
J	0	0	1	0	0	0	0	0	0	57	0	0	0	0	0	0	0	0	1	0
K	0	0	0	0	0	0	0	0	0	0	65	0	0	0	0	0	0	0	0	0
L	0	0	0	0	0	0	0	1	1	0	0	62	0	0	0	0	0	0	0	0
M	0	0	0	0	0	0	0	0	0	0	0	0	35	0	0	0	0	0	0	0
N	0	0	0	0	0	0	0	0	0	0	0	0	0	32	0	0	0	0	0	0
O	0	0	0	0	0	0	0	0	0	0	0	0	0	0	35	0	0	0	0	0
P	0	1	0	0	3	0	0	0	0	0	0	0	1	0	0	31	0	0	0	0
Q	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	34	0	0	0
R	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	66	0	0
S	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	62	0
T	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	2	69

data cycle of a subject and saved in.csv format excel data based. For this experiment, we applied the random forest classifier with 10 fold cross-validation. Table 3 describes the results using different matrics. Table 4 presents the confusion matrix of the experiment.

Experiments for User Authentication on WISDM data set.

B. WISDM DATASET

The WISDM dataset contains data of 36 volunteers. The participants carried the phone in their front leg pocket and performed the different activities, i.e., walking, jogging, ascend stairs, descend stairs, sit, and stand for a specific time. We collected the data every 50ms, and there were 20 samples per second. The following table 5 describes the details of WISDM dataset.

We have extracted the following set of statistical features from the WISDM data set. These features correspond to the fiducial points in data cycles. And for each axis these features are selected that comprise of $(11^*3) = 33$.

We constructed the models using the Random Forest and SVM classifier with 10 fold cross validation. The results of the experiments are presented in Table 6. We present the confusion matrix of only first 20 subjects due to space constraints and better visibility in Table 7.

The results of the activity wise user authentication experiments are presented in the following Table 8. Here we have described the detailed results of classifier performing the best. Compared with other classifiers the random forest classifier provided the best results.

V. DISCUSSIONS

IoT devices such as sensor-enabled smartphones, wearables, etc., produce a large number of personal data assets. These data sets provide different contextual information about a person's activity, biometric, biological and behavioral, and social patterns. In IoT applications, there is a frequent disconnection in mobile environments. Continuous authentication and monitoring of users are essential [38]. The existing user authentication methods which require interaction and input from a user are not suitable and troublesome for the users. It also causes the overhead and delay in real-time applications where time and efficiency is the basic requirement [39].

Therefore, the user device-generated sensor data contains unique patterns for user authentication. There exist efforts that use the sensor data from different contexts and perspectives for authentication credentials such as Touch patterns, device holding position [40], keystroke input [41], user gait [42], etc. The most widely adopted method for sensor data-based user recognition includes the user behavior related data such as the actions, activities patterns, etc.

There exists a variety of sensors in the IoT devices that produce user actions, motions, and biomedical-related information. Motion sensors are explored in the existing literature for this purpose quite effectively compared to other sensors, such as fingerprints, blood sugar level, oxygen saturation level, which got less attention. We can use these sensors individually or with fusion with existing techniques to improve the performance of user authentication. It is encouraged in existing research [43].

TABLE 8. Results matric of activity wise user authentication.

Activity	No of User	Classifier	Accuracy
Jogging	32	RF	79.6%
Downstairs	32	RF	50.0%
Upstairs	32	RF	56.5 %
Sitting	23	RF	37.5 %
Standing	24	RF	51.9%

In this study, we have introduced the IoT data-based user authentication method that evaluated the wide range of sensors on a diverse range of data sets. Moreover, we have examined the different tools for sensor-based data analytics for user recognition. The MySignals-eHealth platform provides a large number of sensors with quite a realistic accuracy level of measurements for research purposes. It has potential applications in various healthcare-related systems such as monitoring. We also used the publically available data sets such as WISDM for the experimentation. The results are evidence for future research directions. The openSMILE [37] tool provides a built-in feature set with different configuration files. These are particularly for various uses.

We present a comparative analysis to highlight the difference of features set an impact on accuracy. The measurements provided by the biomedical sensors through smart devices are a good addition for the customization of user profiles. Even though the proposed approach performs better in many aspects, but here we highlight a few limitations of the proposed approach.

MySignals-eHealth platform is still at its earlier stage of implementation and has limited data sets available till now. For testing purposes, we have used a real-time data set. The IoT devices have limited batteries, so when we are collecting real-time data continuously and storing it in the cloud. There are a few steps required for the better management of the data collection and storage process for long periods.

VI. CONCLUSION & FUTURE WORK

User device-generated data contain important user information. Moreover, the biomedical sensors included in the smart devices have enabled more options for the collection of subject-related data. These sensors have been used for healthcare monitoring and activity recognition. Although they contain unique behavioral patterns of each user. In this research, we have proposed a user authentication framework based on IoT data. We use biomedical sensors such as heart rate, oxygen saturation, and glucometer sensors for authentication purposes. We have performed detailed experimentation on the real data sets. The MySignal-eHealth produced an accuracy of 97%. The experiments on the WISDM data set with the same feature set produced an accuracy of 99%. The feature set of openSmile provided less accuracy with a large number of features that causes computational overhear. The construction of user sensor profiles and experiments for user authentication has proved the effectiveness of these sensor data based profiles for user authentication.

REFERENCES

[1] 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.

[2] F. Schaub, R. Deyhle, and M. Weber, "Password entry usability and shoulder surfing susceptibility on different smartphone platforms," in *Proc. 11th Int. Conf. Mobile Ubiquitous Multimedia*, 2012, pp. 1–10.

[3] M. Nasajpour, S. Pouriyeh, R. M. Parizi, M. Dorodchi, M. Valero, and H. R. Arabnia, "Internet of Things for current COVID-19 and future pandemics: An exploratory study," 2020, *arXiv:2007.11147*.

[4] F. Schaub, R. Deyhle, and M. Weber, "Password entry usability and shoulder surfing susceptibility on different smartphone platforms," in *Proc. 11th Int. Conf. Mobile Ubiquitous Multimedia (MUM)*, 2012, pp. 1–10, doi: 10.1145/2406367.2406384.

[5] M. Abuhamad, T. Abuhmed, D. Mohaisen, and D. Nyang, "AUToSen: Deep-learning-based implicit continuous authentication using smartphone sensors," *IEEE Internet Things J.*, vol. 7, no. 6, pp. 5008–5020, Jun. 2020.

[6] S. A. Israel, J. M. Irvine, A. Cheng, M. D. Wiederhold, and B. K. Wiederhold, "ECG to identify individuals," *Pattern Recognit.*, vol. 38, no. 1, pp. 133–142, Jan. 2005.

[7] M. Komeili, N. Armanfard, and D. Hatzinakos, "Liveness detection and automatic template updating using fusion of ECG and fingerprint," *IEEE Trans. Inf. Forensics Security*, vol. 13, no. 7, pp. 1810–1822, Jul. 2018.

[8] S. Batool, N. A. Saqib, M. K. Khattack, and A. Hassan, "Identification of remote IoT users using sensor data analytics," in *Proc. Future Inf. Commun. Conf.* Germany: Springer, 2019, pp. 328–337.

[9] M. N. Malik, M. A. Azam, M. Ehatisham-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.

[10] M. Tamviruzzaman, S. I. Ahamed, C. S. Hasan, and C. O'brien, "EPet: When cellular phone learns to recognize its owner," in *Proc. 2nd ACM Workshop Assurable Usable Secur. Configuration*, 2009, pp. 13–18.

[11] D. Mail, "Do YOU suffer from password rage? A third of people have thrown a tantrum after forgetting login details," Tech. Rep., 2015.

[12] T. Neal and D. Woodard, "Surveying biometric authentication for mobile device security," *J. Pattern Recognit. Res.*, vol. 11, no. 1, pp. 74–110, 2016.

[13] M. Hammad, Y. Liu, and K. Wang, "Multimodal biometric authentication systems using convolution neural network based on different level fusion of ECG and fingerprint," *IEEE Access*, vol. 7, pp. 26527–26542, 2018.

[14] J. R. Pinto, J. S. Cardoso, and A. Lourenço, "Evolution, current challenges, and future possibilities in ECG biometrics," *IEEE Access*, vol. 6, pp. 34746–34776, 2018.

[15] 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, 2018.

[16] S. Bianco and P. Napoletano, "Biometric recognition using multimodal physiological signals," *IEEE Access*, vol. 7, pp. 83581–83588, 2019.

[17] J. Lee and J. Kim, "Energy-efficient real-time human activity recognition on smart mobile devices," *Mobile Inf. Syst.*, vol. 2016, pp. 1–12, Jan. 2016.

[18] S. Sprager, R. Trobec, and M. B. Jurič, "Feasibility of biometric authentication using wearable ECG body sensor based on higher-order statistics," in *Proc. 40th Int. Conv. Inf. Commun. Technol., Electron. Microelectron. (MIPRO)*, May 2017, pp. 264–269.

[19] C. Camara, P. Peris-Lopez, J. E. Tapiador, and G. Suarez-Tangil, "Non-invasive multi-modal human identification system combining ECG, GSR, and airflow biosignals," *J. Med. Biol. Eng.*, vol. 35, no. 6, pp. 735–748, Nov. 2015.

[20] I. Bisio, A. Delfino, F. Lavagetto, and A. Sciarrone, "Enabling IoT for in-home rehabilitation: Accelerometer signals classification methods for activity and movement recognition," *IEEE Internet Things J.*, vol. 4, no. 1, pp. 135–146, Feb. 2017.

[21] 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.
[22] M. Muaaz and R. Mayrhofer, "Smartphone-based gait recognition: From authentication to imitation," *IEEE Trans. Mobile Comput.*, vol. 16, no. 11, pp. 3209–3221, Nov. 2017.
[23] W.-H. Lee and R. B. Lee, "Multi-sensor authentication to improve smartphone security," in *Proc. Int. Conf. Inf. Syst. Secur. Privacy (ICISSP)*, 2015, pp. 1–11.
[24] S. N. Ramli, R. Ahmad, M. F. Abdollah, and E. Dutkiewicz, "A biometric-based security for data authentication in wireless body area network (WBAN)," in *Proc. 15th Int. Conf. Adv. Commun. Technol. (ICACT)*, 2013, pp. 998–1001.
[25] A. Buriro, B. Crispo, and M. Conti, "AnswerAuth: A bimodal behavioral biometric-based user authentication scheme for smartphones," *J. Inf. Secur. Appl.*, vol. 44, pp. 89–103, Feb. 2019.
[26] *MySignals-eHealth*. Accessed: Sep. 24, 2020. [Online]. Available: http://www.my-signals.com/
[27] T. Ahrens, "The most important vital signs are not being measured," *Austral. Crit. Care*, vol. 21, no. 1, pp. 3–5, Feb. 2008.
[28] X.-F. Teng, Y.-T. Zhang, C. C. Y. Poon, and P. Bonato, "Wearable medical systems for p-health," *IEEE Rev. Biomed. Eng.*, vol. 1, pp. 62–74, 2008.
[29] R. Krishnamurthi, A. Kumar, D. Gopinathan, A. Nayyar, and B. Qureshi, "An overview of IoT sensor data processing, fusion, and analysis techniques," *Sensors*, vol. 20, no. 21, p. 6076, Oct. 2020.
[30] N. V. Chawla and G. Karakoulas, "Learning from labeled and unlabeled data: An empirical study across techniques and domains," *J. Artif. Intell. Res.*, vol. 23, pp. 331–366, Mar. 2005.
[31] *Rolling Average*. Accessed: Sep. 24, 2020. [Online]. Available: https://www.portent.com/blog/analytics/rolling-averages-mathmoron.htm
[32] *Moving Average in Pandas*. Accessed: Sep. 24, 2020. [Online]. Available: https://www.datacamp.com/community/tutorials/moving-averagesin-pandas
[33] P. Musale, D. Baek, N. Werellagama, 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.
[34] 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.
[35] S. Batool, A. Hassan, N. A. Saqib, and M. A. K. Khattak, "Authentication of remote IoT users based on deeper gait analysis of sensor data," *IEEE Access*, vol. 8, pp. 101784–101796, 2020.
[36] M. Nasajpour, S. Pouriyeh, R. M. Parizi, M. Dorodchi, M. Valero, and H. R. Arabnia, "Internet of Things for current COVID-19 and future pandemics: An exploratory study," 2020, *arXiv:2007.11147*.
[37] 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. ESANN*, vol. 3, 2013, p. 3.
[38] *OpenSMILE*. Accessed: Dec. 2, 2020. [Online]. Available: https://www.audeering.com/opensmile/
[39] S. Amini, V. Noroozi, A. Pande, S. Gupte, P. S. Yu, and C. Kanich, "DeepAuth: A framework for continuous user re-authentication in mobile apps," in *Proc. 27th ACM Int. Conf. Inf. Knowl. Manage.*, Oct. 2018, pp. 2027–2035.
[40] W.-H. Lee, X. Liu, Y. Shen, H. Jin, and R. B. Lee, "Secure pick up: Implicit authentication when you start using the smartphone," in *Proc. 22nd ACM Symp. Access Control Models Technol.*, Jun. 2017, pp. 67–78.
[41] A. Buriro, B. Crispo, and Y. Zhauniarovich, "Please hold on: Unobtrusive user authentication using smartphone's built-in sensors," in *Proc. IEEE Int. Conf. Identity, Secur. Behav. Anal. (ISBA)*, Feb. 2017, pp. 1–8.
[42] F. Bergadano, D. Gunetti, and C. Picardi, "User authentication through keystroke dynamics," *ACM Trans. Inf. Syst. Security*, vol. 5, no. 4, pp. 367–397, 2002.
[43] 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, 2018.
[44] S. Deb, Y. Ou Yang, M. C. H. Chua, and J. Tian, "Gait identification using a new time-warped similarity metric based on smartphone inertial signals," *J. Ambient Intell. Hum. Comput.*, vol. 11, no. 10, pp. 4041–4053, Oct. 2020.
[45] G. Giorgi, A. Saracino, and F. Martinelli, "Using recurrent neural networks for continuous authentication through gait analysis," *Pattern Recognit. Lett.*, vol. 147, pp. 157–163, Jul. 2021.

SAMERA BATOOL graduated in economics from Punjab University and received the M.S. degree in CS from the Institute of Information Technology, University of Arid and Agriculture Rawalpindi, in 2013, the M.C.S. degree from the National University of Modern Languages, Islamabad, and the Ph.D. degree from the College of Electrical and Mechanical Engineering, NUST, Islamabad. She has published her research work at various reputable platforms.

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 an Associate Professor with the Department of Computer and Software Engineering, College of Electrical and Mechanical Engineering, NUST. His research interests include machine learning and speech processing.

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 with 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 a Professor, an Associate Dean, and a Tenured Associate, in 2017. He has been with the School of Computer Science, University of Ulm, Germany, and the Networking and Multimedia Laboratory, 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, block chain-based information and network security, vehicular *ad-hoc* networks, and acoustic networks.

AHSAN SHAHZAD received the Ph.D. degree in electrical engineering and computer science from the Gwangju Catholic College, in 2019. He is currently working as an Assistant Professor with the Department of Computer & Software Engineering, College of Electrical & Mechanical Engineering, NUST, Islamabad. His research interests include healthcare informatics, biomedical signal processing, and machine learning.

MUHAMMAD UMAR FAROOQ received the Ph.D. degree in computer science from the University Politehnica of Bucharest. He is an Inventor of 1 awarded U.S. patent and has several years of industry and teaching experience. He is currently working as an Assistant Professor with the College of Electrical & Mechanical Engineering, NUST, Islamabad.