IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 27 May 2022, accepted 11 June 2022, date of publication 24 June 2022, date of current version 1 July 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3186011

A Biometrics-Based Behavioral Trust Framework for Continuous Mobile Crowd Sensing Recruitment

RUBA NASSERAN[1,2], RABEB MIZOUNI[1,2], HADI OTROK[1,2], SHAKTI SINGH[1,2], (Member, IEEE), MENATALLA ABOUOUF[1,2], AND MAHA KADADHA[1]

[1]Electrical and Computer Engineering Department, Khalifa University of Science and Technology, Abu Dhabi, United Arab Emirates
[2]Center on Cyber-Physical Systems, Khalifa University of Science and Technology, Abu Dhabi, United Arab Emirates

Corresponding author: Rabeb Mizouni (rabeb.mizouni@ku.ac.ae)

ABSTRACT The emergence of mobile crowd sensing (MCS) platforms makes it possible to collect data in a time and cost efficient manner. However, one of the challenges in MCS systems is obtaining reliable information especially in the presence of impersonators who could provide false reports without getting detected. User recruitment mechanisms adopted in MCS systems hire users such that the Quality of Information (QoI) of the submitted sensing reports is maximized. Despite that, there is still a risk of recruiting impersonators to the system. This problem is even more prominent in the case of continuous mobile sensing tasks, since multiple false sensing reports could be submitted by the impersonator during the sensing period, which can impact the quality of the sensing outcome and hence the performance of the system. Therefore, to ensure the reliability of the continuously submitted data, a more robust recruitment mechanism, which detects and eliminates impersonators during the sensing task, is needed. This work proposes a biometrics-based behavioral trust framework that can support a reliable recruitment process in continuous MCS tasks. Behavioral biometrics are unique behavioral traits that can be used to profile users based on how they naturally perform a specific activity. By leveraging machine learning techniques, these behavioral traits can be used in order to detect impersonators in the system. In this work, a unique model for each MCS worker is built based on their unique interaction patterns with the smartphone's touching screen. The proposed approach integrates the trained machine learning models with a dynamic continuous recruitment system, which continuously monitors the QoI of the submitted sensing reports and changes the recruited participants as needed. Simulation results of the proposed approach show its efficacy in detecting and eliminating impersonators in continuous sensing recruitment.

INDEX TERMS Mobile crowdsensing (MCS), quality of information (QoI), behavioral biometrics, touchscreen dynamics.

I. INTRODUCTION

The widespread use of smartphones and smart devices around the world has led to the emergence of sensing paradigms such as Mobile Crowdsensing (MCS), where user-paired devices are recruited to perform a sensing task. A typical MCS system consists of a task requester, who submits sensing requests to the management platform, a cloud-based management platform, which processes, analyzes and stores the sensing data, and the MCS workers, who perform the sensing tasks and send the sensing reports back to the management platform. In terms of user involvement, the data collection process can either take place in an opportunistic or in a participatory manner. In opportunistic sensing, users are not asked to perform a specific action, however, an application is run in the background and the data collection is performed automatically. On the other hand, in participatory sensing, a user is required to perform a task at a specific time and location and then users are rewarded based upon the quality of the data they submit [1]. The deployment of MCS platforms complements the existing Internet of Things (IoT) sensing solutions in supporting the vision of smart cities and improving the

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

quality of citizens' life. In fact, it is viewed as an important solution for building smart cities since human mobility and intelligence offer higher coverage and deeper understanding of the sensing task context.

MCS tasks can be classified into *one-time sensing tasks* and *continuous sensing tasks*. In one-time sensing tasks, only one-time readings from the devices of the recruited workers are needed. These tasks are usually triggered by event occurrence, such as detecting and reporting a car accident, monitoring bus arrival time, comparing prices of goods and healthcare applications. On the other hand, continuous sensing tasks require collecting information continuously from the Area of Interest (AoI) for a given sensing period. This type of sensing is required to accurately study the phenomena that task publisher is interested in. For example, monitoring noise, air pollution, and characterizing the coverage of WiFi intensity, all require data to be collected continuously for a specific period of time [1].

A. PROBLEM STATEMENT

In MCS systems, recruiting trustworthy users plays an important role in collecting high quality data. However, due to the human involvement in the data collection process in participatory MCS systems, malicious users could submit poisoned sensing reports to the management platform without getting detected, by impersonating other users' identities. As illustrated Figure 1, this poses a serious threat to the system, especially in continuous sensing tasks, since multiple readings will be collected from users' devices during the sensing period. Therefore, to limit the chances of successful attacks, continuous monitoring of users' behavior is needed during the sensing task so that the recruitment system can detect and interrupt impersonators as soon as possible.

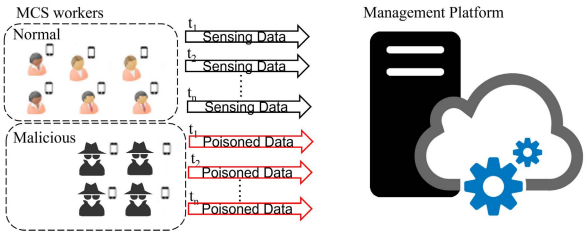

FIGURE 1. Impact of recruiting impersonators on continuous sensing tasks.

Typically, in continuous MCS tasks, users recruitment is performed such that the Quality of Information (QoI) of the submitted sensing reports meets the publisher's required quality throughout the period of sensing. Multiple works in literature proposed recruitment frameworks for continuous sensing tasks. Some of them proposed recruiting users based on the knowledge of historical data [2]–[4], while others proposed dynamically recruiting users during the task based on real-time updates of the quality of the submitted sensing reports [4]–[6]. Unfortunately, although existing recruitment systems in MCS can maintain a good QoI, there is still a risk of recruiting impersonators to the sensing task. In fact, the presence of impersonators in the system can cause a significant drop in the QoI value as will be illustrated in Section I-B. Hence, continuous recruitment systems also need to eliminate impersonators during the sensing period in order to ensure high quality sensing.

Today's smart devices are equipped with different sensors, which can be used to distinguish genuine users from impersonators in the system from how a user performs a specific activity such as walking, typing or interacting with the smartphone's touch screen. The unique behavioral traits that can be used to monitor users' action every point in time and detect impersonators is referred to as behavioral biometrics [7], [8]. Unlike one-time authentication solutions such as passwords and facial recognition, using behavioral biometrics to identify impersonators is considered a more practical solution as it offers an unobtrusive way of collecting data from users. Hence, impersonators can be detected without having to interrupt users during the sensing task. Different types of behavioral biometrics-based solutions exist, including touchscreen dynamics behavioral biometrics, which relies on the users' behavioral patterns when interacting with the smartphone's touchscreen, walking gait behavioral biometrics, where users are identified based on the motion data collected from them as they walk, and keystroke dynamics behavioral biometrics, which leverages features that describe the behavior of users when typing on the smartphone's keyboard. This work considers touchscreen dynamics behavioral biometrics since it is not scenario dependent and it only relies on the application being used and the touchscreen input data generated by the user.

All of the behavioral biometrics based solutions proposed in literature use machine learning techniques in order to make predictions about whether the user is genuine or not. However, the use of machine learning is always accompanied with uncertainties and there is no guarantee for the correctness of the predictions made by the model. Hence, in order to integrate touchscreen dynamics behavioral biometrics with continuous MCS recruitment, it is essential to take into consideration evaluating the trustworthiness of the predictions made by the machine learning models when making predictions about the users.

B. MOTIVATIONAL SCENARIO

To illustrate the impact of the presence of impersonators in the selected group of participants, an existing dynamic recruitment system for continuous sensing tasks, namely the stability-based Group-based Recruitment System (stability-based GRS) proposed in [6], was simulated in an environment where all participants are genuine (truthful environment) versus an environment where some of them are impersonated (untruthful environment). The stability-based GRS considers participants mobility during the selection process to ensure high coverage and the task requester's desired QoI [6]. In this example, the dataset of mobility traces in the city of Cologne, Germany, was used for both truthful and untruthful

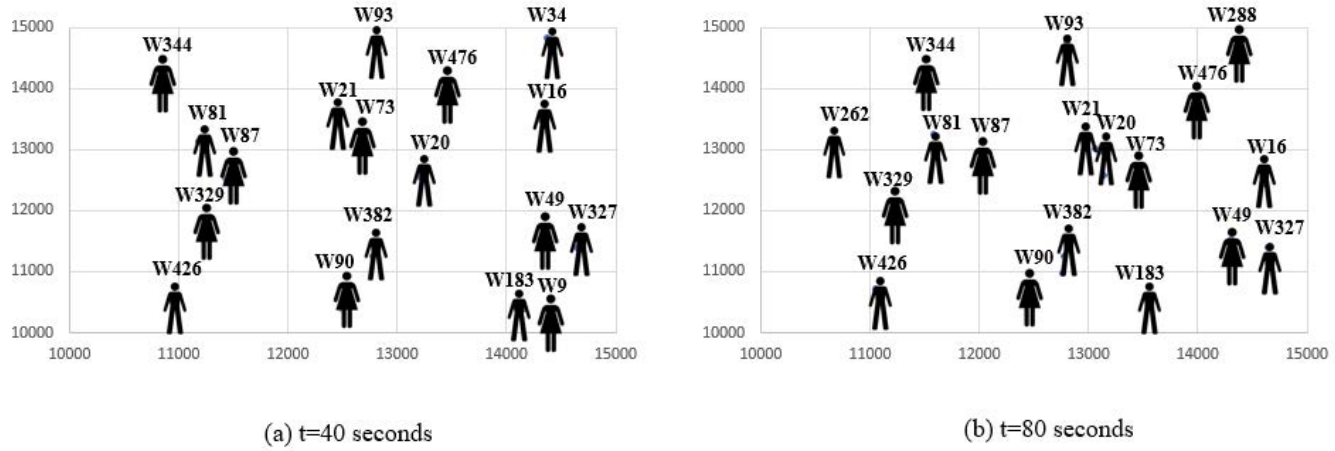

FIGURE 2. Snapshots of the selected groups in the second and fourth intervals in a truthful environment.

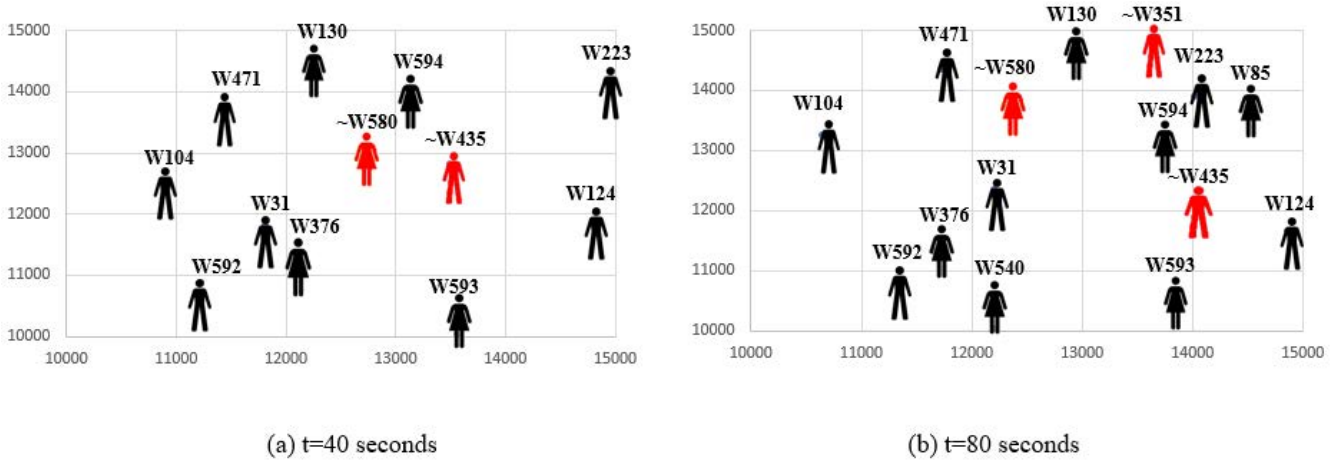

FIGURE 3. Snapshots of the selected groups in the second and fourth intervals in an untruthful environment.

environments [9]. The simulation was performed for a sensing task with AoI boundaries (10000 to 15000)x(10000 to 15000) for a duration of 80 seconds and with a required QoI of 2.7. The sensing period was divided into four equal sensing intervals of length 20 seconds each. Snapshots of the selected groups in the second and fourth intervals in both environments are shown in figures 2 and 3.

Figure 2 shows the users selected 40 and 80 seconds after the task has started in a truthful environment. At $t = 40$ seconds, 19 users were needed in order to achieve a QoI of 2.7, whereas at $t = 80$ seconds, two additional workers were needed (*W262* and *W288*) to keep the QoI of the group above 2.7. For the untruthful environment, 20% of the population was randomly chosen to be the impersonators.

The negation symbol before the identity of the worker was used to denote that this is not the true identity of the worker. As shown in Figure 3, the group members at $t = 40$ seconds included two misbehaving workers who are impersonating the identities of *W580* and *W435*. Additionally, at $t = 80$ seconds, an additional misbehaving worker who is impersonating the identity of *W351* was added to the group.

TABLE 1. Expected vs. achieved QoI in truthful and untruthful environments.

Time	Truthful Environment	Untruthful Environment	
	Achieved QoI	Expected QoI	Achieved QoI
40 sec	2.725	2.730	2.516
80 sec	2.726	2.740	2.341

The QoI of the group, at each time point, in a truthful environment is compared against the obtained values in an untruthful environment as shown in Table 1. The expected QoI is evaluated based on the selected workers without the detection of the impersonators, while the achieved QoI is computed when considering only genuine users. As can be seen from the table, the achieved QoI falls below the

required value of 2.7 in both sensing intervals. In addition, the achieved QoI is less than the expected QoI by 8.5% and 17% for both intervals respectively in the untruthful environment, which presents the impact of impersonators on the sensing outcome.

C. CONTRIBUTION

In order to ensure the reliability of the submitted sensing data in continuous sensing tasks, a more robust recruitment mechanism should also monitor users' behavior during the sensing period. This work leverages continuous authentication using behavioral biometrics in continuous participatory sensing tasks to flag potential impersonators in the system. Using supervised machine learning, a unique model per worker is trained to characterize his behavior. The generated models for all the workers are then used by the framework to help distinguish genuine users from impersonators by incorporating their predictions in the recruitment approach. Consequently, impersonators are detected, removed from the task and replaced with more reliable users. However, since the predictions made by the machine learning models are accompanied with uncertainties, the proposed system also considers evaluating a trust metric for the models before deciding to keep or remove any user. To summarize, the main contributions of this work are as follows:

- Leverage behavioral biometrics in continuous participatory MCS tasks, to eliminate impersonators from the system by building unique models for each MCS worker and integrating the predictions made by these models with the continuous recruitment process.
- Propose a trust evaluation mechanism for the workers' machine learning models to account for the uncertainties accompanied with the predictions made by the model when deciding to keep or remove a participant from the sensing task.

Simulation results of the proposed approach show its efficacy in detecting and eliminating impersonators in the stability-based GRS recruitment system using real-life datasets of mobility traces and touchscreen input data which describes users' interaction patterns with the smartphone's touchscreen. In the stability-based GRS, users are continuously added and removed from the group by considering their mobility to ensure that the required QoI value is met.

The remainder of this paper is organized as follows. Section II presents the related work. Section III presents the overall proposed approach. Section IV presents the touchscreen input dataset used and explains how the machine learning models were trained. Section V presents the proposed framework which integrates the trained models with the stability-based GRS. Section VI presents the performance evaluation of the proposed approach. Finally, Section VII concludes the paper.

II. RELATED WORK

In this section, a literature review is provided for the work done on ensuring high data quality in continuous MCS systems, in addition to the work done on impersonation attacks detection using behavioral biometrics.

A. ENSURING DATA QUALITY IN CONTINUOUS MCS SYSTEMS

Ensuring high quality sensing data is an important problem that should be addressed in continuous MCS systems. In [2], a framework to assign tasks to the best group of users who will return high quality reports within the required sensing period was proposed. The quality of information evaluated for each worker depends on the reputation, the confidence that a worker will complete the task within a given period of time and the distance between the worker and the given task. The reputation parameter computed by the MCS system depends on the historical performance of the worker. It considers that a task was performed successfully if their answers are equal to the estimated ground truth value of the system [2]. In [3], a dynamic-trust-based recruitment framework which calculates the overall trust based on real-time direct trust and indirect trust was proposed. Real-time direct trust refers to the trust value evaluated based on the satisfaction degree of the task requester in the worker in the recent past. In this trust evaluation method, the latest interaction record is given more weight than previous records. On the other hand, the indirect trust evaluation method relies on collecting feedback from the task requester after the task is over for other task requesters, as a reference in future interactions [3]. In [4], the problem of recruiting the minimum number of participants, who can achieve a certain level coverage, to multiple continuous sensing tasks was tackled. Several offline and online greedy algorithms were proposed to dynamically select a subset of participants to perform the tasks based on the probability of a user making calls at particular time and locations. In [5], an efficiency cost data collection scheme (ECDCS) where the worker is selected according to the contribution that all the data it collects have on the whole system rather than a single data samples was proposed. A matrix completion technology was adopted to recover the missing data samples with partial data while ensuring the quality of service of the task. The proposed algorithm selects workers with lower cost and better collaboration effect. In [6], a continuous group-based recruitment system was proposed which selects the best group of workers while considering their mobility patterns. The proposed system ensures that the required QoI value is met during the sensing period by continuously adding or removing members to the group [6]. Overall, different factors that affect the final sensing outcome were considered in literature in order to optimize the recruitment process in MCS systems as shown in Table 2. These factors include: the coverage, distribution, reputation and other device attributes such as the battery level and sampling frequency. As illustrated in the table, none of the proposed solutions considered the confidence that the users are genuine during the recruitment process. Therefore, this work considers the confidence parameter in addition to the other QoI parameters in order to ensure high quality sensing.

TABLE 2. Research gap.

Reference	Factor affecting the final sensing outcome					
	Reputation	Willingness	Coverage	Distribution	Device attributes	Confidence that users are genuine
[2]	✓	✓	✗	✗	✗	✗
[3]	✓	✗	✗	✗	✗	✗
[4]	✗	✓	✗	✗	✗	✗
[5]	✗	✗	✓	✗	✗	✗
[6]	✓	✓	✓	✓	✓	✗
Proposed solution	✓	✓	✓	✓	✓	✓

B. BEHAVIORAL BIOMETRICS

The unique behavioral traits that can be used to continuously profile users based upon their natural interactions and without having to constantly interrupt them during the session is referred to as behavioral biometrics. The most commonly used behavioral biometrics in literature include keystroke dynamics, walking gaits and touchscreen dynamics behavioral biometrics [7], [8]. In keystroke dynamics behavioral biometrics, features describing the typing rhythm such as the keystroke length, the pressure exerted on each key while typing and the time difference between consecutive strokes are used [8]. The performance of a range of anomaly detection algorithms employed to authenticate users based on keystroke dynamics was evaluated and compared in [10]. The top-performing detectors found were Manhattan, Nearest Neighbor and Outlier Count (z-score) [10]. Additionally, two binary classifiers: BayesNet and Random forest were used by [11] to perform authentication using keystroke features along with features describing the user's phone holding behavior obtained from the smartphone's built in sensors. The proposed scheme showed acceptable authentication rates with data that was collected in six different user positions: sitting, standing, walking, walking upstairs, walking downstairs and lying on the sofa [11]. Although continuous authentication based on keystroke dynamics provides unobtrusive data collection, the variability of typing behavior is expected to appear across different sittings which makes this type of behavioral biometrics scenario dependent. In addition, the flexibility of the input text requires the need to gather as much typing input as possible, which translates to longer waiting time before the authentication can be performed efficiently [12].

Multiple works in literature have shown that individuals can be recognized by their gait provided that proper motion measurements are taken. Walking gaits behavioral biometrics refers to the characteristic and mannerism in which an individual walks [13]. Today's smartphones and smart devices are equipped with built in motion sensors such as accelerometers and gyroscopes which contribute data that can be used to extract unique features using the prevalent signal processing techniques. In [14], a convolutional neural network-based deep learning model was proposed to identify smartphone users in crowd sensing systems based on the data produced by the accelerometer sensors in their smartphones. It was concluded that the Fourrier transform is a simple but very powerful technique in the feature extraction process which has improved the accuracy of the model [14]. In [15], a deep neural network based scheme which relies on the unique physical features of WiFi signals during the daily activities of mobile users was proposed. The system extracts 6 time domain features and 3 frequency domain features from both the amplitude and the phase channel of the channel response of WiFi signals. The extracted features are then used in a three-layer stacked autoencoder to perform activity recognition and then user authentication [15]. In order to accelerate the authentication process without having to perform feature extraction at an earlier stage [16] introduced a deep learning approach that self-learns the necessary network traffic features to authenticate the MCS users. Using a stacked autoencoder, the first layer learns first-order features which are then used in the second layer to learn the features corresponding to the patterns from the previous features [16]. Continuous authentication based on walking gaits can be affected by several internal factors including psychological conditions and illness as well as external factors such carrying a load and the type of footwear. In addition, prior studies and experiments were conducted in a controlled environment, which is not the case in the real physical world. As a result, it is expected that models trained using a dataset in a certain situation would introduce bias when applied to other situations [8].

The third type of behavioral biometrics is touchscreen dynamics behavioral biometrics. In touchscreen dynamics behavioral biometrics, features related to the on-screen sliding movements that represent the user's unique interaction patterns with the smartphone's touchscreen are used. A real-time re-authentication scheme for smartphones using touchscreen dynamics behavioral biometrics was proposed in [17]. Five machine learning algorithms were employed to authenticate users including decision tree, naive Bayesian, K-nearest neighbor, logistic regression and SVM. SVM classifier was found to be the best suited for authentication with lower equal error rate (EER) and better performance than the other machine learning methods. The data used for training and testing was obtained from users as they used their smartphones in a routine manner over a period of one month [17]. The performance of ten touch-based

authentication classification algorithms were evaluated in [18]. The best performing algorithm found with the lowest EER was the logistic regression machine learning algorithm. The data used was obtained from users as they answered multiple choice questions on their smartphones over two sessions that were at least one day apart [18]. In [19], SVM was adopted to perform touch stroke based authentication using data collected from users while normally using their smartphones during a 15 minute session for 21 days. To model a valid user in the authentication classifier, the user's data during the previous 20 days were used. On the other hand, in order to model the attacker in the authentication classifier, data obtained from users who were selected randomly from the remaining users in the dataset was used. Overall, an improvement in the average error rate was observed as the number of the randomly selected users to model the attacker increased [19]. In [20], touch stroke features were used in two different classifiers, K-nearest-neighbors (KNN) and support vector machine (SVM) to authenticate users in three different experimental settings: inter-session authentication, inter-week authentication and intra-session authentication.The data used was collected from users as they performed two different tasks: a reading task and an image comparison task. Overall, the authentication difficulty seemed to increase with the increase in the time difference between training and testing, and SVM always achieved a lower EER than KNN algorithm [20].

Authentication based on touch operations provides a natural way to collect user interaction data. Each user generates unique touch patterns, which depend on the application being used. Therefore, this type of authentication is well suited to protect against access of unauthorized individuals to important mobile applications [8].

Different machine learning techniques have been deployed in literature to learn users behavior and detect impersonators in the system. However, none of these solutions considered integrating behavioral biometrics in continuous MCS recruitment systems. Since the predictions made by the machine learning models are not guaranteed to be correct, the integration of behavioral biometrics with continuous MCS recruitment requires ensuring that the model is trustworthy before taking any decision of removing or keeping a worker. The proposed approach is discussed more in details in the next section.

III. OVERALL PROPOSED APPROACH

To address the problem of detecting impersonators in MCS systems, a novel approach is proposed which verifies that the recruited workers in continuous sensing tasks are genuine. This work leverages touchscreen behavioral biometrics to monitor the workers' behavior during the sensing period. Based on the real-time prediction made by the trained models, the proposed system eliminates from the task workers who have high likelihood of being impersonators. As mentioned previously, MCS systems typically consist of three main entities: *task requesters*, the *management platform* and *workers*. The sensing process starts by a task requester sending a sensing task to the management platform. Based on the requirement of the task, the platform will start recruiting the appropriate participants. Once the task is completed, the management platform evaluates and aggregates the submitted reports and forwards the aggregated sensing reports back to the task requester [21]. The proposed approach builds unique behavioral models that can identify genuine and non-genuine users from the way they interact with their smartphones' touchscreens. A high level architecture of the proposed system is illustrated in Figure 4. In order to incorporate worker's behavior predication in the recruitment process, the crowd sourcing platform is designed to include two main modules: *Behavioral Biometrics Model Generation*, and *Worker Selection*, which are summarized below:

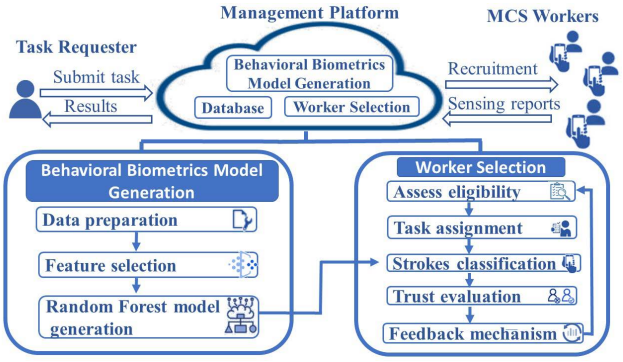

FIGURE 4. High level architecture of the proposed system.

- **Behavioral Biometrics Model Generation**: This module is responsible for the generation of the machine learning model for each worker. The features describing the user's behavior of swiping on the smartphone's touch screen are used to generate a customized training model for unique profiling of a worker. At first, the system needs to prepare the data for each worker to be able to run it through the machine learning algorithm. This includes filtering out the training instances that belong to the user, using the remaining data to model the impersonator's behavior, and combining multiple strokes together by applying a sliding window technique. After that, the most important features to train the worker's model are selected. Finally, the machine learning models are generated using Random Forest (RF) algorithm.
- **Worker Selection**: This module is responsible for the continuous selection of the workers that will perform the task. Initially, the eligibility of candidate workers is assessed and those who are available in the AoI and can satisfy the task requirements are selected. Each continuous task is divided into equal sensing intervals. After each interval, the received touchscreen data as well as the previously generated behavioral biometrics models are used in order to make predictions about whether the worker is genuine or not. The trustworthiness of the

predictions made by the classifier each sensing interval is then evaluated and used to decide whether to keep or remove the participant. Finally, a feedback mechanism is adopted to consider the current observations about the user's behavior in future sensing intervals.

IV. TOUCHSCREEN INPUT BEHAVIORAL BIOMETRICS

Using the touchscreen input data collected from the workers during the sensing period, unique behavioral models, which characterize their unique interaction patterns with the touchscreen, are built for each worker. A single touch stroke is defined as the sequence of touch data that begins with touching the screen and ends with lifting the finger. This work uses supervised machine learning to train each worker's behavioral model. The overall experimental workflow for the training phase is summarized in Figure 5 and will be explained in the subsequent sections.

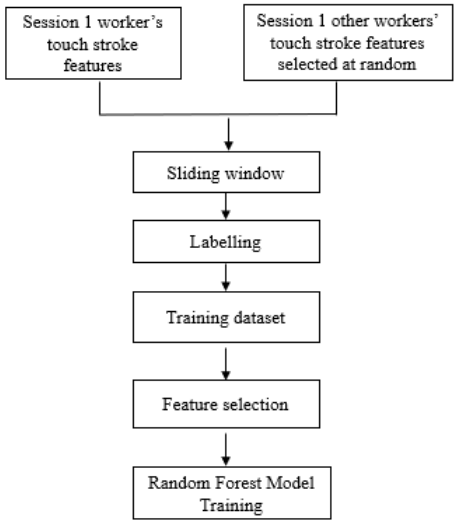

FIGURE 5. Flowchart depicting the experimental workflow.

A. DATASET

One touch stroke s is a trajectory encoded as a sequence of vectors $s_n = \{x_n, y_n, t_n, p_n, A_n, o_n^f, o_n^{ph}\}$. The parameters used to model a stroke vector are described in Table 3. The moment the touchscreen sensor detects that the finger is touching the screen, a series of touch data, which describes the location and orientation of the finger on the touch screen, how much screen area is covered by the finger, the pressure exerted by the finger and the orientation of the smartphone, is produced. For each touch stroke, different features which demonstrate different aspects of users' behavior such as the swiping time, the stroke length and the stroke direction, can be used to characterise a user. Furthermore, the way in which users interact with the smartphone's touch screen mainly depends on the interface design of the application. Consequently, the designers can anticipate what type of navigational gestures will be generated by the users more frequently. This work considers the touch strokes generated when users slide horizontally over the touchscreen. This is typically done when browsing through images or navigating to the next page of icons in the main screen. The dataset used is the Touchalytics dataset, which includes 30 touch stroke features collected from 40 different users who were asked to spot differences between pairs of similar images over two sessions [20]. Table 4 describes all the features used in this work, which could be either temporal features, spatial features or statistical features.

TABLE 3. Parameters used to describe a single touch stroke.

Symbol	Definition
x_n	x coordinate on the screen (number of pixels along the horizontal axis)
y_n	y coordinate on the screen (number of pixels along the vertical axis)
t_n	absolute time stamp of the recorded action (in ms)
p_n	pressure on the screen
A_n	screen area covered by the finger
o_n^f	orientation of the finger with respect to the screen
o_n^{ph}	orientation of the phone (1 for landscape mode and 2 for portrait mode)

B. DATA PREPARATION

In touchscreen based behavioral biometrics, touch stroke features obtained from one session can be used to build unique models for each worker. Therefore, in the training phase, every classifier is trained using the data obtained in the first session whereas in the testing phase, the data obtained from the second session is used. For each worker, the dataset is organized as $(\vec{x}, y) = (x_1, x_2, \ldots, x_m, y)$, where m is the number of stroke features, $\vec{x}$ is the features' vector and $y \in Y$ is the dependant variable which represents the label given to $\vec{x}$. To model the impersonator's touching behavior, touch strokes are randomly selected from other users in the dataset such that the number of training instances of the genuine user equals those of the impersonator. Before training each model, a sliding window technique is applied to the genuine user and the impersonator's touch stroke features. Multiple consecutive strokes were combined together using a sliding window in order to capture the temporal behavioral characteristic of the user when swiping on a touchscreen. In the sliding window method, a window of length n moves over each stroke feature, and computes the average of the data in the window. After that, the data is labeled such that $y = 1$ indicates that the stroke was generated by a genuine user and $y = 0$ indicates that the stroke was generated by an impersonator.

C. CHOICE OF CLASSIFIERS

Random Forest is an ensemble machine learning algorithm, which relies on generating many decision trees and aggregating their results in the end by taking the majority vote. Every decision tree randomly selects different subsets of features

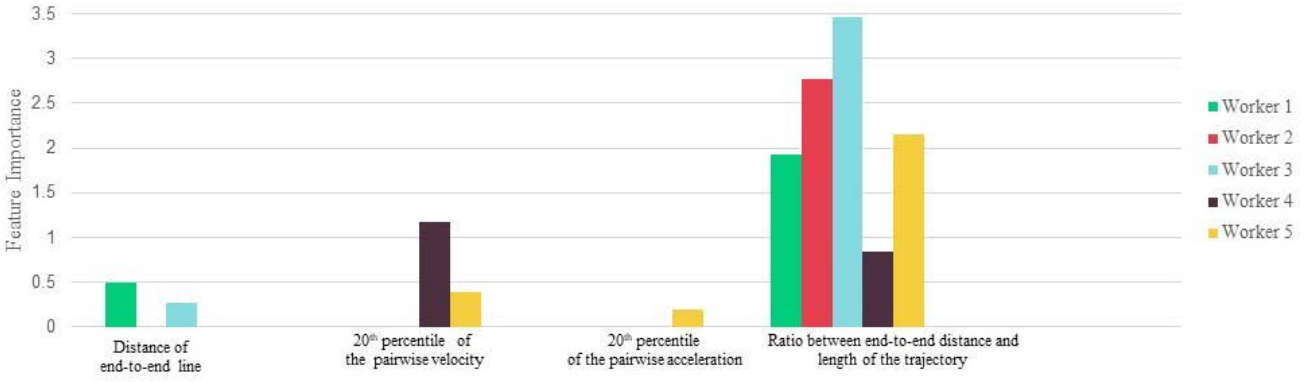

FIGURE 6. Feature importance for 5 different workers.

and different training instances from the entire training data set. In this work, a total number of 60 decision trees was used since it was found that increasing the number or trees more than that did not improve the performance of the models. This was also confirmed in [22]. The main reasons behind choosing RF algorithm are summarized below:

1) The touch strokes generated during the sensing period sometimes might not represent the true behavior of the user which leads to having outliers in the dataset. Therefore, Random Forest was chosen because it is less sensitive to outliers since it splits the data into groups based on a threshold value.
2) The touch stroke features dataset used includes continuous data types, like the average velocity of the stroke, as well as discrete data types, such as the direction of moving the screen. Hence, Random Forest was chosen as it is accurate, stable and efficient for datasets with continuous or discrete form.
3) It considers workers' possible behavioral changes over time since the final model aggregates a number of different temporal models.

D. FEATURES SELECTION

Not all touch stroke features collected can be used to uniquely distinguish a worker from other users across multiple sessions in the same way. In reality, each worker may have some specific behavioral features that are important to characterise his/her behavior. Permutation feature importance (PFI) can be used to compute the importance score of the features for each worker. First, RF is employed and each model is trained using the original training data. Then, the models are trained another time using the same data but after shuffling the training instances of the feature in the dataset. The performance of the generated models without shuffling is then compared with the performance of the models with shuffling. The features that show higher sensitivity to the shuffling operation are considered of higher importance to the model. Additionally, features with low importance are removed from the final training. Figure 6 illustrates the importance of the features used in this work for 5 different workers.

V. TOUCHSCREEN INPUT BEHAVIORAL BIOMETRICS IN CONTINUOUS MCS RECRUITMENT

Continuous sensing tasks often seek the cooperation of multiple workers to collect data that cannot be collected only by a single worker, which increases the risk of recruiting impersonators to the system. In order to select the most appropriate group to perform the sensing task, a continuous group-based recruitment system that considers the participants' collective QoI was proposed in [6]. In the proposed recruitment system, the QoI of the group depends on parameters which could be either related to the AoI, the device or the user. The system takes into consideration that some of these device-related parameters and AoI-related parameters can change during the period of sensing, due to participants' mobility or reduction in their devices battery levels. Hence, the QoI is monitored by the system periodically and participants are added and removed until the required QoI value is achieved. In this work, the touch screen input behavioral biometrics models are integrated with the dynamic recruitment system introduced in [6] to improve the recruitment by detecting and eliminating potential impersonators.

A. PARAMETER FORMULATION AND QoI EVALUATION

In MCS systems, every worker can be defined as $w_i =< l_i, SA_i, SF_i, RE_i, P_i, S_i, ML_i >$. The different attributes used to define a worker are summarized in table 5. Based on the task publisher's requirements, the management platform finds the set of eligible participants by choosing those who are available in the AoI and have the required sensors to perform the sensing task. Some publishers might also have some constraints on the reputation and confidence of each worker. The reputation of a worker represents how committed the worker is to the sensing task, since some workers might leave the AoI before the end of the sensing period. It can be

TABLE 4. Touch stroke features used in this work.

	Features
Temporal	Time difference between two consecutive strokes
	Stroke duration
Spatial	x start position of the stroke
	y start position of the stroke
	x end position of the stroke
	y end position of the stroke
	End to end distance of the stroke
	Direction of the end to end distance of the stroke
	Mean direction for each (x_n,y_n),(x_{n+1},y_{n+1}) pair
	Average velocity of the stroke
	Encoded direction of moving the screen: 1 up, 2 down, 3 left, 4 right
	Pressure exerted by the finger in the middle of the stroke
	Ratio between end to end distance and length of the trajectory
	Change of the finger orientation during the stroke
	Phone orientation Screen area covered by the finger in the middle of the stroke
	Sum of pairwise distance through the stroke Mean of pairwise distances through the stroke
Statistical	20^{th} percentile of the pairwise velocity
	50^{th} percentile of the pairwise velocity
	80^{th} percentile of the pairwise velocity
	20^{th} percentile of the pairwise acceleration
	50^{th} percentile of the pairwise acceleration
	80^{th} percentile of the pairwise acceleration
	Median of the last three points of the velocity
	Largest deviation from end to end line: the maximum distance between the direct end-to-end line and the line of the trajectory
	20^{th} percentile of the deviation from end-to-end line
	50^{th} percentile of the deviation from end-to-end line
	80^{th} percentile of the deviation from end-to-end line
	Median acceleration over the first five points

evaluated as shown in (1).

$$P_i = \frac{\text{number of tasks the worker was committed to}}{\text{total number of assigned tasks}} \quad (1)$$

The confidence of the system in an individual worker can be obtained as given in (2). Every sensing interval, the system makes predictions about whether the user is genuine or not. Based on that, as more intervals detect that a certain worker is impersonated during the sensing tasks, smaller value is

TABLE 5. Different attributes used to define a worker.

Symbol	Definition
l_i	location of worker w_i
SA_i	set of the sensors available in the device of worker w_i
SF_i	sampling frequency of the the sensor in the device of worker w_i
RE_i	residual energy of worker w_i
P_i	reputation of worker w_i
$Conf_i$	confidence that worker w_i is genuine based on the predictions made every interval
ML_i	behavioral biometrics trained machine learning model of worker w_i

obtained for the confidence of the system in that worker.

$$Conf_i = 1 - \frac{\text{duration the worker was impersonated}}{\text{duration the worker was committed}} \quad (2)$$

Using these attributes, different parameters that characterize a group of workers can be evaluated as shown in table 6. These parameters can be classified into three main categories: AoI-related parameters, user-related parameters and device-related parameters. From the GPS location of the workers, two AoI related parameters can be evaluated: the coverage and the distribution. To evaluate the coverage of the group, first the AoI is divided into smaller sub-regions and then it can be evaluated by dividing the number of sub-regions which include at least one group member over the total number of sub-regions in the AoI. Full coverage of is achieved when the group submits sensing data from every sub-region in the AoI. On the other hand, the distribution parameter $D(g)$ measures how uniformly the participants are distributed in the AoI. It uses the Chi Square test in order to determine whether the observed values of the true number of users in each sub-region meet the theoretical assumption that participants are evenly distributed among all sub-regions. In addition to the AoI-related parameters, two device related parameters that reflect the capability of the participants' devices to sense the requested data are used: the sampling frequency of the sensors in the group $SF(g)$ and the residual energy of the devices in the group $RE(g)$. Finally, to reflect the users' properties that can affect the sensing outcome, two user-related parameters are considered: the reputation of the group $P(g)$ and the confidence of the group $S(g)$. The reputation parameter represents how committed the workers are to completing the sensing task. On the other hand, the confidence metric reflects the certainty of the system that the group of users are genuine, based on the touchscreen input data obtained from them from previous sensing intervals. The previously introduced parameters can be used to find the QoI of the group as in (3).

$$QoI(g) = w_1 \times C(g) + w_2 \times D(g) + w_3 \times SF(g) + w_4 \times RE(g) + w_5 \times P(g) + w_6 \times Conf(g) \quad (3)$$

In this equation, $w_1 - w_6$ are the weights assigned to each of the parameters as specified by the task publisher [6].

TABLE 6. Parameters used in the QoI evaluation.

	Parameter	equation
AoI-related	Coverage	$C(g) = \frac{\text{no. of covered sub-regions}}{\text{total no. of sub-regions}}$
	Distribution	$D(g) = 1 - \frac{s^2}{\theta_U}$; for $\theta_U \neq 0$
Device-related	Sampling frequency for sensor n	$SF(g,s) = \frac{1}{n}\sum_{i \in g} SF_i(s) \times e^{-SD(SF_i(n))}$
	Residual Energy	$RE(g) = \frac{1}{n}\sum_{i \in g} RE_i \times e^{-SD(RE_i)}$
User-related	Reputation s	$P(g) = min_i\{P\}$
	Confidence	$Conf(g) = min_i\{Conf\}$

1) STABILITY

In continuous sensing tasks, some participants might leave the AoI or lose connectivity with the management platform during the sensing period. Since the sensing is continuous, recruiting such participants is not desirable. Nevertheless, it is possible to predict participants' future mobility patterns using their historical mobility traces. Based on the predictions of the participants' locations over the required sensing period in the AoI, the coverage that is expected to be achieved at the beginning of each sensing interval (C_i) can be evaluated. The stability parameter reflects the availability of the participants in the AoI by the summing the expected coverage at every sensing interval during the sensing period, as given in (4). This parameter can be used to assess the likelihood of the workers to stay in the AoI during the entire sensing period. Therefore, it is necessary to take into consideration that the stability is maximized when selecting participants before the beginning of the sensing task [6].

$$Stability = \sum_{i=1}^{\text{number of sensing intervals}} C_i \tag{4}$$

2) TRUST EVALUATION

After every sensing interval, the system uses participants' touch screen input and their behavioral biometrics models in order to predict whether or not each participant is genuine. However, the use of machine learning models is always accompanied by uncertainties regarding their outcomes. Since no guarantee is provided for the correctness of the predictions made by the models, the system needs to deal with inherent uncertainty in its outcome in order to make the models more dependable. Trusting a machine learning model can be interpreted as a special type of belief, where the model believes that users behave in a certain way. In order to quantitatively evaluate the trustworthiness of a machine learning model, the uncertainty in the belief needs to be taken into consideration. In the context of this work, three possible scenarios can take place: The model believes that the user is a genuine user without any uncertainties, the model believes that the user is not genuine without any uncertainties, or the model is uncertain about whether or not the user is genuine.
The probability that user i is genuine can be evaluated as given in (5), by considering the classifier's probabilistic score $p_j(ML_i, s_j)$ obtained for each genuine stroke prediction and the fraction of the positively labeled strokes by the model (m/M), where m is the number of positively labeled strokes and M is the total number of predictions made by the model.

$$p_i = \frac{m}{M}\sqrt{\prod_{j=1}^{m} p_j(ML_i, s_j)} \tag{5}$$

Using the evaluated probability, the entropy function can be evaluated as given in equation 6.

$$H(p_i) = -p_i \log_2(p_i) - (1 - p_i)\log_2(1 - p_i) \tag{6}$$

The trust metric of a machine learning model of user i can be evaluated using entropy as given in (7), where p_i is the probability that the user is genuine based on the machine learning model's predictions and $H(p_i)$ is the entropy function which represents the model's uncertainty. The trust metric gives a positive value if the model believes that user is genuine without any uncertainties, a negative value if the model believes that the user is not genuine without any uncertainties, and 0 if the model is uncertain about whether or not the user is genuine. Therefore, in the proposed approach, whenever the trust value is negative, the user is eliminated since the model is confident that the user is an impersonator. On the other hand, if the trust value is positive or zero, the user doesn't get eliminated and remains in the group.

$$T_{ML_i} = \begin{cases} 1 - H(p_i), & 0.5 \leq p_i < 1 \\ H(p_i) - 1, & 0 < p_i < 0.5 \\ 1, & p_i = 1 \\ 0, & p_i = 0 \end{cases} \tag{7}$$

B. INTEGRATING BEHAVIORAL BIOMETRICS IN CONTINUOUS MCS RECRUITMENT

Once a sensing task is publicized, the goal of the recruitment system is to find the group with the highest stability and confidence whose members are able to achieve the required QoI during the period of sensing. A flowchart of the recruitment system is illustrated in Figure 7 and summarized below:

1) The system starts initially by searching for a group of participants whose stability and confidence are maximized using the genetic algorithm. The fitness function used in the genetic algorithm is shown in (8).

$$F = Conf(g) + stability \tag{8}$$

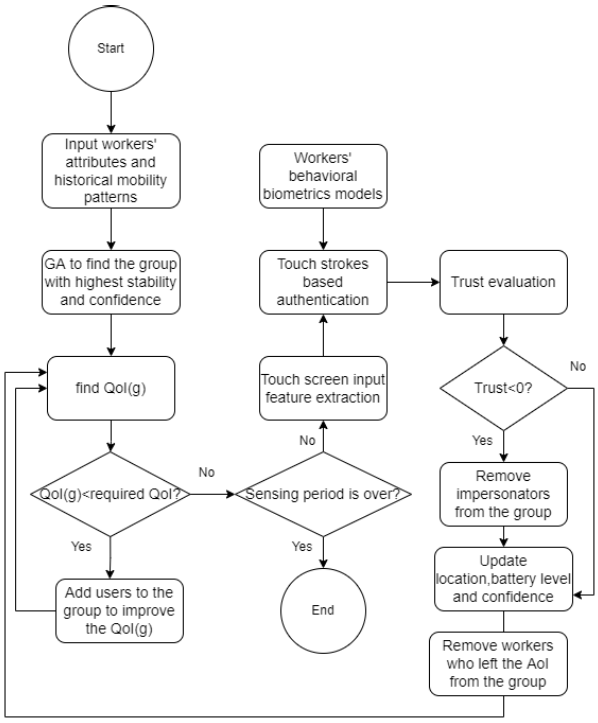

FIGURE 7. Flowchart of the proposed approach.

2) The QoI of the selected group is evaluated and new members are added in case the obtained QoI value is less than the required value by the task publisher.
3) After every sensing interval, the group members' behavioral biometric models are used to predict whether the touch strokes were generated by a genuine worker or an impersonator.
4) A trust value is evaluated to find the confidence in the predictions made by the machine learning model. If the obtained trust value is negative, then the system removes impersonators from the group. Otherwise, if the trust value is positive or zero, then the worker gets to stay in the group. The trust metric is further explained in the next section.
5) The system updates workers' locations, confidence parameters, and devices battery levels.
6) Workers who leave the AoI are removed from the group.
7) Finally, the QoI is checked again and new members are added to the group in case its value was found to be less than the required value by the task publisher.

The continuous recruitment system algorithm used is shown in Algorithm 1. Initially, the genetic algorithm is implemented by trying different combinations to find the best group of workers. The maximum group size in the genetic algorithm was set to 20. Every sensing interval, the marginal QoI of all workers in the population is evaluated. Based on these values, the algorithm greedily adds workers to the group whenever the QoI was found to drop below the required value.

Once a decision is made by the system to keep or remove each participant in the group every sensing interval, all confidence values are updated, as given in (9), by taking the weighted sum of the previous and new confidence values.

$$Conf_i = (1-\alpha)Conf_i + \alpha(1 - \frac{duration\ the\ worker\ was\ impersonated}{duration\ the\ worker\ was\ committed}) \tag{9}$$

Furthermore, after every interval, the members of the group get paid in return for their submitted sensing reports. As a result, once impersonators get eliminated, they do not get paid for the subsequent intervals. The total payment made to the group can be evaluated using (10), where the maximum QoI represents the maximum QoI value that a group can achieve if all parameters were maximized, and the maximum interval budget is the maximum budget assigned by the publisher for each interval.

$$TP = \frac{QoI(g)}{max\ QoI} \times\ max\ interval\ budget \tag{10}$$

Every participant in the group is paid based on their contribution to the QoI(g). The contribution of a worker can be evaluated using (11), by finding the marginal QoI. In equation (11), g' is equivalent to the group g but without worker w_i. Based on that, the payment is evaluated for each participant every sensing interval as given in (12), where $payment^a_{w_i}$ is the payment for worker w_i at interval a. At the end of the task, all participants are paid for all the intervals they participated in using (13).

$$contribution\ of\ worker w_i = QoI(g) - QoI(g') \tag{11}$$

$$payment^a_{w_i} = \frac{contribution\ of\ worker w_i \times TP}{QoI(g)} \tag{12}$$

$$TP_{w_i} = \sum_{a=1}^{\text{num of sensing intervals}} payment^a_{w_i} \tag{13}$$

C. DATASET

The dataset used in order to obtain users' locations over a certain period of time is the vehicular mobility traces of the city of Cologne dataset [9]. In all experiments, the AoI boundaries were set to (10000 to 15000) x (10000 to 15000), which narrowed down the total number of users in the AoI to 600 users. The remaining user attributes including the reputation, residual energy, sampling frequency and confidence were randomly generated following a uniform distribution. In addition, the previously trained behavioral biometrics models and touchstroke features were assigned for each user in the dataset at random.

VI. EVALUATION RESULTS

In this section, simulation results of the proposed approach are provided to validate its efficacy by running the continuous recruitment system with and without integration of the proposed mechanism for the detection and elimination of

Algorithm 1 Continuous Recruitment System Algorithm

```
Input: participants' dataset, mobility patterns, trained models, required QoI, numOfIntervals, AoI_X_Begin,
AoI_Y_Begin, AoI_X_End, AoI_Y_End, numOfStrokes, numOfYBlocks, numOfXBlocks
Output: best group of participants that can achieve the required QoI
max_group_size= 20;
most_stable_group= GA(participants' dataset, mobility patterns, AoI_X_Begin, AoI_Y_Begin, AoI_X_End,
  AoI_Y_End, max_group_size, numOfYBlocks, numOfXBlocks);
qoi= find_qoi(most_stable_group, participants' dataset);
population_size= length(participants' dataset);
for a=1 to numOfIntervals do
    while qoi<required_qoi do
        for i=1 to population_size do
            if i≠ most_stable_group then
                marginal_qoi_change(i, 1)= i;
                temp_group= [most_stable_group; i];
                temp_qoi= find_qoi(temp_group, participants' dataset);
                if temp_qoi≥qoi then
                    marginal_qoi_change(i,2)= temp_qoi- qoi;
                end
            end
        end
        index= maxIndex(marginal_qoi_change(:, 2));
        most_stable_group=[most_stable_group; marginal_qoi_change(index,1)];
        group_size= length(most_stable_group);
        qoi= find_qoi(most_stable_group, participants' dataset); alpha= 0.9;
        for g=1 to group_size do
            model= get_model(most_stable_group(g), trained models);
            TestData= get_test(most_stable_group(g), participants' dataset, numOfStrokes);
            [labels_predicted, scores]= predict(model, TestData);
            trust= findTrust(label_predicted, scores);
            if trust<0 then
                update_confidence(participants' dataset, alpha, most_stable_group(g), Impersonator=True);
                most_stable_group(g)=[];
                else
                    update_confidence(participants' dataset, alpha, most_stable_group(g), Impersonator=False);
                    most_stable_group(g)=[];
                end
            end
        end
        most_stable_group= check_X_Y(most_stable_group, participants' dataset, a+1);
        qoi= find_qoi(most_stable_group, participants' dataset);
    end
end
return most_stable_group;
```

impersonators. First, workers' behavioral models are trained and evaluated using different machine learning algorithms including: Random Forest, decision tree, and Support vector machines (SVM) with an rbf kernel. Then, the trained models with the best performance are used with the stability-based continuous recruitment system.

A. EVALUATION OF THE MACHINE LEARNING MODELS

All workers' models that were trained using the data obtained from the first session are tested using data obtained from the second session. Three metrics are used for performance evaluation: precision, recall and f1 score, where the positive class was chosen to be the genuine user class and the negative class was chosen to be the impersonator class.

To illustrate the effect of combining multiple strokes on the classification performance using the sliding window technique, Figure 8 shows the f1 score average value of all 40 behavioral models over 10 runs for a varying number of strokes. The performance of three machine learning algorithms including RF, SVM and decision tree classifier

was compared. In order to perform the hyperparameter tuning of the SVM models, cross validation with five folds was used. As shown in the figure, the models trained with RF outperformed those trained with SVM and decision tree classifier. Therefore, these models will be used at the recruitment stage in order to detect and eliminate the impersonators during the sensing task. Furthermore, as depicted in the figure, the performance of the models trained with RF started to converge towards 95% at n= 7 strokes. Therefore, this number was used to combine the strokes during training and testing. Table 7 shows the average and the standard deviation of the precision, recall and f1 scores of the behavioral models when tested using the data obtained from the second session. Despite the fact that the models trained using RF showed the best performance, the system still needs to consider that these predictions are accompanied with some uncertainties, and hence, it must ensure that these predictions are trustworthy before making any decisions about the user during the sensing task.

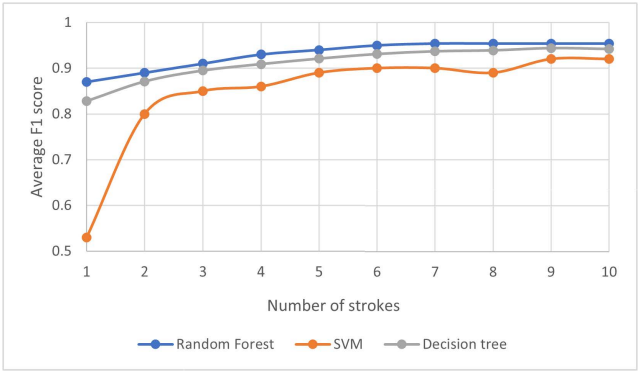

FIGURE 8. F1 score value averaged over 10 runs for a varying number of strokes.

TABLE 7. Evaluation of the machine learning models trained using RF, SVM and decision tree classifier.

	Precision		Recall		F1 score	
	mean	σ	mean	σ	mean	σ
RF	0.956	0.035	0.951	0.039	0.954	0.037
Decision tree	0.947	0.064	0.93	0.071	0.937	0.067
SVM	0.901	0.16	0.912	0.13	0.90	0.14

B. EVALUATION OF CONTINUOUS MCS TASKS WITH BEHAVIORAL BIOMETRICS

In this section, the proposed approach is simulated in an untruthful environment to prove its robustness. No works exist in literature that integrate behavioral biometrics with a continuous MCS recruitment system. Therefore, the performance of the proposed approach is compared with the stability-based group based recruitment system (stability-based GRS) proposed in [4] which gives a higher emphasis to participants' mobility during the sensing period in order to ensure a certain QoI value. This selection mechanism is simulated using the same equations proposed in section V-A.

1) PERFORMANCE EVALUATION OF ONE SENSING TASK

In this experiment, a sensing task with 10 sensing intervals and a required QoI of 2.5 was simulated where the percentage of impersonators in the AoI was varied from 10% to 40%. The simulation results are shown in Figures 9 and 10. As illustrated in the figures, it takes the proposed approach one interval only to meet the required QoI value and achieve a higher confidence value during the rest of the sensing task period, even when 40% of the population were impersonated. This is due to the fact that the system needs to wait to obtain touchscreen input data from the users during the first sensing interval. In addition, it is clear from the figures that the proposed approach performs better than the stability-based GRS since the stability-based GRS fails to achieve the required QoI during the sensing period due to the decreasing confidence of the group. The main reasons behind this decrease is that the impersonators who join from the beginning of the sensing period get to stay recruited to the task and more of them might also get recruited in the subsequent intervals by the recruitment system in case some group members leave the AoI.

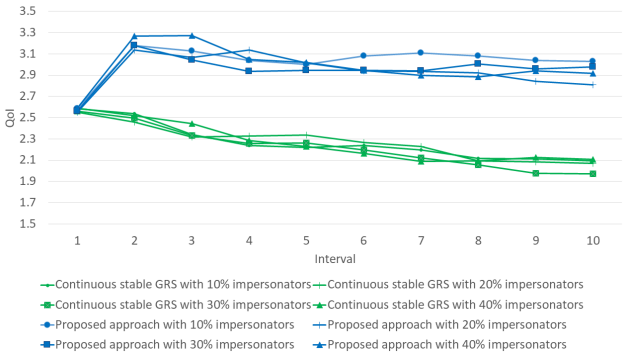

FIGURE 9. Achieved QoI by the group during the sensing task with 10% to 40% impersonators in the population.

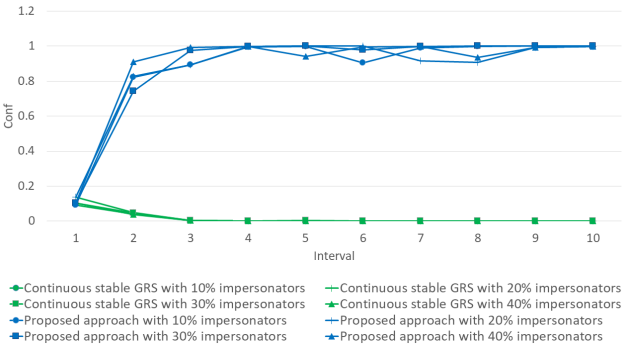

FIGURE 10. Achieved confidence by the group during the sensing task with 10% to 40% impersonators in the population.

Furthermore, the task payoff of the proposed approach was compared with the task payoff of the stability-based GRS.

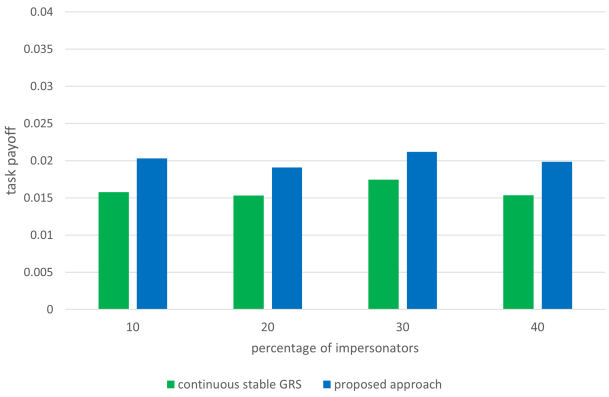

FIGURE 11. Task payoff achieved for a varying percentage of impersonators in the AoI.

The payoff of a sensing task represents the benefit gained in contrast to the costs endured by the task. In this work, the task payoff is defined as the minimum QoI achieved during the sensing task over the time needed to complete the recruitment of the users by the system, as shown in (14).

$$\text{task payoff} = \frac{\text{minimum QoI achieved during the sensing task}}{\text{running time}} \tag{14}$$

Figure 11 shows the task payoff achieved by the proposed approach and the stability-based group-based recruitment system for a varying percentage of impersonators in the area of interest. As illustrated in the Figure, the proposed approach outperforms the stability-based group-based recruitment system in each scenario which shows the efficacy of the proposed solution.

2) PERFORMANCE EVALUATION OF MULTIPLE SENSING TASKS

A total of 100 sensing tasks with different number of sensing intervals and required QoI values were simulated in an AoI, with 20% of its population being impersonated. To show how the proposed approach detects and eliminates impersonators, Figure 12 shows the total number of impersonators selections in the selected groups for every set of tasks. The simulation was repeated 10 times for every task and the average number of impersonators selection was considered. As shown in the figure, the number of selections of impersonators made by the stability-based GRS can reach six time more than the number of selections made by the proposed approach. Therefore, the proposed approach outperforms the stability-based GRS in terms of not selecting impersonators throughout the sensing period. This is due to the fact that throughout the simulations, the impersonators' selections made by the proposed approach mostly would take place at the beginning of the sensing task, on the other hand, the stability-based GRS selects impersonators every interval and does not consider reducing their confidence, which causes the number of impersonators selections

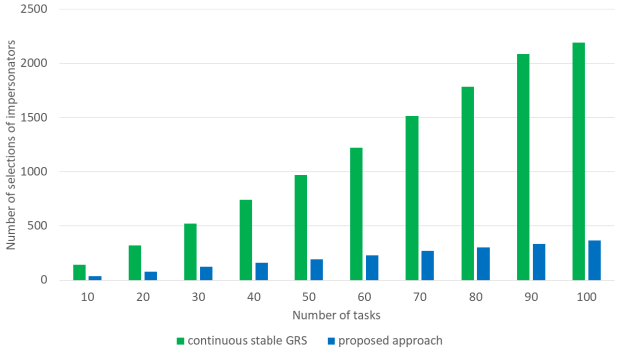

FIGURE 12. Number of selections of impersonators in the group for a different number of tasks.

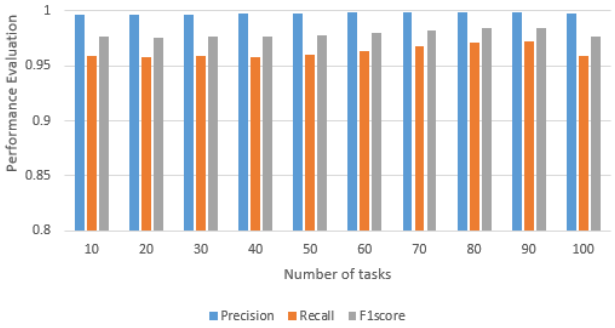

FIGURE 13. Performance evaluation of the proposed system for a different number of tasks.

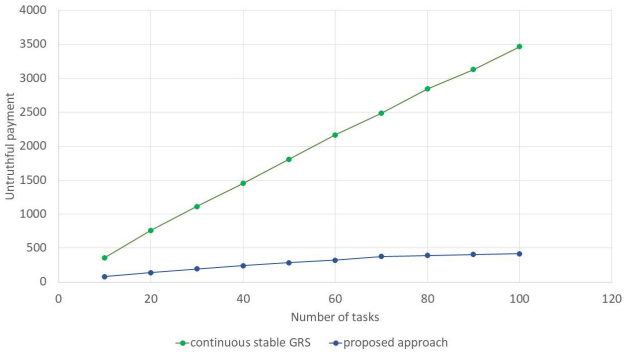

FIGURE 14. Untruthful payment for a varying number of tasks.

to be higher. The performance of the system was evaluated using recall, precision and f1score as given in Figure 13. High scores for all metrics indicate that the system was able to accurately predict whether the user is an impersonator or not, and as a result, make the right decision of keeping or eliminating the user from the task. Finally, the untruthful payment made using the proposed approach is compared to the untruthful payment made using the stability-based GRS in Figure 14. As expected, the untruthful payment made to the workers using the proposed approach is less due to the small number of impersonators' selection.

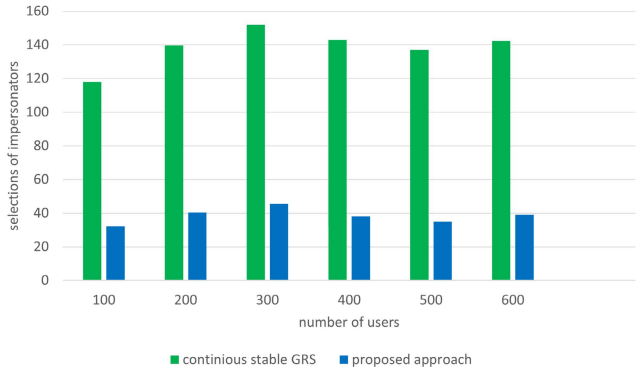

FIGURE 15. Number of selections of impersonators for a different number of users in the AoI.

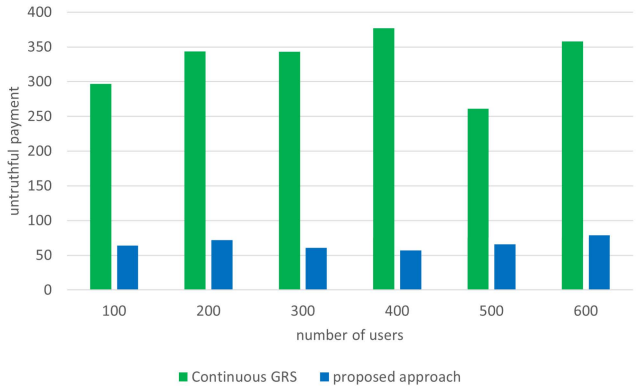

FIGURE 16. Untruthful payment made for a different number of users.

3) PERFORMANCE EVALUATION OF ONE SENSING TASK FOR A DIFFERENT NUMBER OF USERS

An experiment was conducted to evaluate the proposed approach with different number of users for a set of 10 tasks and 20 % impersonators in the AoI. The simulations were all repeated 10 times and the average number was considered in each case. First, the number of selections of impersonators was observed when the number of users in the AoI was varied from 100 to 600 users as shown in Figure 15. It is clear that the selections of impersonators made by the proposed system is significantly less than the selections made by the stability-based GRS in all scenarios, which proves the scalability of the proposed system.

Secondly, the untruthful payment made using the proposed approach is compared against that made by the stability-based GRS as illustrated in Figure 16. As evident from the figure, the proposed system outperforms the stability-based GRS for all scenarios with different number of users in the AoI, since impersonators get detected and eliminated during the sensing period. Finally, as shown in Figure 17, the performance of the system was evaluated using precision recall and f1 score measures. As depicted in the figure, the average f1 score achieved by the system in all scenarios is around 97% which shows the robustness of the proposed trust-based framework in detecting impersonators due to the fact that it can cope with

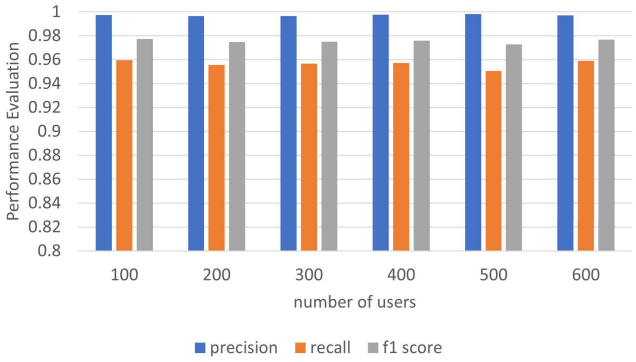

FIGURE 17. Performance evaluation of the proposed system for different number of users in AoI.

the uncertainties accompanied with the predictions made by the trained models.

VII. CONCLUSION

In this paper, touchscreen input behavioral biometrics was integrated with continuous MCS recruitment in order to detect and eliminate impersonators from the group in every sensing interval. Three machine learning algorithms including RF, SVM and decision tree, were used to classify touch strokes made by a user into genuine strokes or non genuine strokes. The performance of the trained user models were then evaluated and compared. Subsequently, the models trained using RF were adopted during MCS recruitment to detect and eliminate impersonators after every sensing interval based on the data collected from users as they interacted with the smartphone's touchscreen. A trust metric was proposed to help the system detect impersonators and genuine users with higher confidence and cope with the uncertainties in the predictions. Simulations were performed in untruthful environments for one task and multi-tasks to evaluate the performance of the proposed approach. The simulations performed in a sensing task with a specified required QoI showed that the proposed approach was able to maintain the required QoI value whenever it received touchscreen input from the users. This was achieved even when the percentage of impersonators in the group would reach 40% of the entire population. Furthermore, the task payoff, defined as the minimum QoI to running time ratio, was evaluated for each approach, and it was shown that proposed system outperforms the stability-based GRS. On the other hand, the simulations performed for multiple tasks showed that the proposed approach makes less number of selections of impersonators and consequently, decreased the untruthful payment significantly. Finally, to prove the scalability of the proposed system, its performance was evaluated for a varying number of users in the AoI.

REFERENCES

[1] A. Capponi, C. Fiandrino, B. Kantarci, L. Foschini, D. Kliazovich, and P. Bouvry, "A survey on mobile crowdsensing systems: Challenges, solutions, and opportunities," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 3, pp. 2419–2465, 3rd Quart., 2019.

[2] R. Estrada, R. Mizouni, H. Otrok, A. Ouali, and J. Bentahar, "A crowd-sensing framework for allocation of time-constrained and location-based tasks," *IEEE Trans. Services Comput.*, vol. 13, no. 5, pp. 769–785, Sep. 2020.
[3] Y. Gao, X. Li, J. Li, and Y. Gao, "DTRF: A dynamic-trust-based recruitment framework for mobile crowd sensing system," in *Proc. IFIP/IEEE Symp. Integr. Netw. Service Manage. (IM)*, May 2017, pp. 632–635.
[4] H. Li, T. Li, and Y. Wang, "Dynamic participant recruitment of mobile crowd sensing for heterogeneous sensing tasks," in *Proc. IEEE 12th Int. Conf. Mobile Ad Hoc Sensor Syst.*, Oct. 2015, pp. 136–144.
[5] Y. Ren, W. Liu, Y. Liu, N. Xiong, A. Liu, and X. Liu, "An effective crowdsourcing data reporting scheme to compose cloud-based services in mobile robotic systems," *IEEE Access*, vol. 6, pp. 54683–54700, 2018.
[6] R. Azzam, R. Mizouni, H. Otrok, S. Singh, and A. Ouali, "A stability-based group recruitment system for continuous mobile crowd sensing," *Comput. Commun. J.*, vol. 119, pp. 1–14, Apr. 2018.
[7] L. Gonzalez-Manzano, J. M. D. Fuentes, and A. Ribagorda, "Leveraging user-related Internet of Things for continuous authentication: A survey," *ACM Comput. Surveys*, vol. 52, no. 3, pp. 1–38, May 2020.
[8] Y. Liang, S. Samtani, B. Guo, and Z. Yu, "Behavioral biometrics for continuous authentication in the Internet-of-Things era: An artificial intelligence perspective," *IEEE Internet Things J.*, vol. 7, no. 9, pp. 9128–9143, Sep. 2020.
[9] S. Uppoor, O. Trullols-Cruces, M. Fiore, and J. M. Barcelo-Ordinas, "Generation and analysis of a large-scale urban vehicular mobility dataset," *IEEE Trans. Mobile Comput.*, vol. 13, no. 5, pp. 1061–1075, May 2014.
[10] K. S. Killourhy and R. A. Maxion, "Comparing anomaly-detection algorithms for keystroke dynamics," in *Proc. IEEE/IFIP Int. Conf. Dependable Syst. Netw.*, Jun. 2009, pp. 125–134.
[11] A. Buriro, B. Crispo, F. D. Frari, and K. Wrona, "Touchstroke: Smartphone user authentication based on touch-typing biometrics," in *Proc. Int. Conf. Image Anal. Process.* Cham, Switzerland: Springer, 2015, pp. 27–34.
[12] P. S. Teh, A. B. J. Teoh, and S. Yue, "A survey of keystroke dynamics biometrics," *Sci. World J.*, vol. 2013, pp. 1–24, Aug. 2013.
[13] A. Nambiar, A. Bernardino, and J. C. Nascimento, "Gait-based person re-identification: A survey," *ACM Comput. Surv.*, vol. 52, no. 2, pp. 1–34, Mar. 2020.
[14] A. I. Middya, S. Roy, S. Mandal, and R. Talukdar, "Privacy protected user identification using deep learning for smartphone-based participatory sensing applications," *Neural Comput. Appl.*, vol. 33, no. 24, pp. 17303–17313, Dec. 2021.
[15] C. Shi, J. Liu, H. Liu, and Y. Chen, "Smart user authentication through actuation of daily activities leveraging WiFi-enabled IoT," in *Proc. 18th ACM Int. Symp. Mobile Ad Hoc Netw. Comput.*, Jul. 2017, pp. 1–10.
[16] V. L. L. Thing, "IEEE 802.11 network anomaly detection and attack classification: A deep learning approach," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Mar. 2017, pp. 1–6.
[17] L. Lu and Y. Liu, "Safeguard: User reauthentication on smartphones via behavioral biometrics," *IEEE Trans. Computat. Social Syst.*, vol. 2, no. 3, pp. 53–64, Sep. 2015.
[18] A. Serwadda, V. V. Phoha, and Z. Wang, "Which verifiers work: A benchmark evaluation of touch-based authentication algorithms," in *Proc. IEEE 6th Int. Conf. Biometrics, Theory, Appl. Syst. (BTAS)*, Sep. 2013, pp. 1–8.
[19] H. Xu, Y. Zhou, and M. R. Lyu, "Towards continuous and passive authentication via touch biometrics: An experimental study on smartphones," in *Proc. 10th Symp. Usable Privacy Secur. (SOUPS)*, 2014, pp. 187–198.
[20] M. Frank, R. Biedert, E. Ma, I. Martinovic, and D. Song, "Touchalytics: On the applicability of touchscreen input as a behavioral biometric for continuous authentication," *IEEE Trans. Inf. Forensics Security*, vol. 8, no. 1, pp. 136–148, Jan. 2013.
[21] M. Abououf, S. Singh, H. Otrok, R. Mizouni, and E. Damiani, "Machine learning in mobile crowd sourcing: A behavior-based recruitment model," *ACM Trans. Internet Technol.*, vol. 22, no. 1, pp. 1–28, Feb. 2022.
[22] T. M. Oshiro, P. S. Perez, and J. A. Baranauskas, "How many trees in a random forest," in *Proc. Int. Workshop Mach. Learn. Data Mining Pattern Recognit.* Cham, Switzerland: Springer, 2012, pp. 154–168.