IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received November 14, 2020, accepted November 26, 2020, date of publication December 14, 2020,
date of current version December 30, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3044495

Green CrowdSensing With Comprehensive Reputation Awareness and Predictive Device-Application Matching Using a New Real-Life Dataset

MAGGIE EZZAT GABER GENDY[1], (Student Member, IEEE),
AHMAD AL-KABBANY[1,2,3], (Member, IEEE),
AND EHAB FAROUK BADRAN[1], (Senior Member, IEEE)
[1]Department of Electronics and Communications Engineering, Arab Academy for Science, Technology & Maritime Transport, Alexandria 21937, Egypt
[2]Intelligent Systems Laboratory, Arab Academy for Science, Technology & Maritime Transport, Alexandria 21937, Egypt
[3]Department of Research and Development, VRapeutic, Cairo 11613, Egypt

Corresponding author: Ahmad Al-Kabbany (alkabbany@ieee.org; alkabbany@aast.edu)

ABSTRACT Mobile CrowdSensing (MCS) has emerged as a valuable framework for large scale mapping of phenomena of interest, thanks to the ever-growing advances and pervasiveness of sensor-rich mobile devices. Meanwhile, towards building a robust MCS system, several challenges have yet to be overcome. This includes security challenges, privacy concerns, data integrity, in addition to others. In this research, we are concerned with the power consumption of sensing campaigns, from the perspectives of service demanders and participants. We propose an opportunistic and predictive crowdsensing management framework that realizes Green Mobile CrowdSensing (GMCS) campaigns through energy-aware participant-task matching. This is achieved using two new techniques. First, we propose Green Auctioning, an auction management technique which adopts a new objective function that guarantees the selection of the devices which consume the least energy. The proposed objective function features a ***hard reputation term*** which depends on the device's energy consumption, in addition to the soft reputation term previously proposed in the literature. Second, we propose Predictive Auctioning, an auction management technique that adopts machine learning models to empower the platform to predict users' ability to complete the task at hand, using the user's battery level and internet connection status, and the task's duration. Towards this goal, we have constructed a new dataset–the ***MAGGIE*** dataset–by monitoring the energy consumption of over 100 users, over a duration of four months, using a mobile application specially developed for this purpose. To the best of our knowledge, this is the first research that addresses energy-aware auctioning by constructing a crowd-sensed dataset specially built for this purpose. We present promising results for the attained energy awareness without compromising other performance aspects including data trustworthiness and the clearance rate of sensing auctions.

INDEX TERMS Auctions, green, prediction, budget, constraints, incentive mechanisms, Mobile CrowdSensing (MCS), opportunistic crowdsensing, penalization, redundancy.

I. INTRODUCTION

Mobile CrowdSensing (MCS) is an emerging framework in which online, battery-powered mobile devices are used to collect different types of data using embedded sensors and on-board electronics. With the ever-growing number of smart phone users, MCS has attracted an increasing interest in a multitude of real-life applications, and hence from the scientific research community. MCS takes the form of sensing campaigns that are run on heterogeneous mobile devices and are usually orchestrated by a platform that receives sensing requests from service demanders and forwards those requests to participants. Since the framework is *people-powered*, MCS can be categorized as either participatory sensing or opportunistic sensing. While the former involves users that are required to actively contribute data, data collection in the

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

latter is done without the user taking action [1], [2]. Even though participatory sensing increases user involvement, and does not suffer from security [3] and privacy challenges, the quality of gathered data depends on the reliability of participants in the sensing campaign. Several factors are required to maintain the reliability of participants, one of which are effective incentive mechanisms.

In sensing campaigns, auctions has been presented in the literature as a theoretical tool for task-participant matching and for modeling incentive mechanisms. The role of incentives is to compensate the users for employing their resources in sensing, and they can be classified into monetary and non-monetary techniques [4], [5]. Incentive mechanisms are key to the success of sensing campaigns since they affect the participation level and the quality of the acquired.

In MCS systems, participants are often referred to as prosumers, since they act as both consumers and producers of crowdsensing data [6]. Many techniques have been proposed to overcome diverse challenges in the process of selecting winners in auctions by evaluating their capacities to accomplish sensing tasks. MSensing [7] dealt effectively with truthful bidding, where the users who aim at increasing their incomes participate in the auction with higher bids. Nevertheless, MSensing was shown to be vulnerable to users who send malicious information. This was handled in [8] by introducing a reputation-aware incentive mechanism. The research proposed in [9] presented different forms for the reputation factor such as voting (decentralized), statistical (centralized), anchor-assisted decentralized, and collaborative scores, reputation-based auctions. The authors of [10] proposed a redundancy factor to penalize redundant task assignment, where a task is assigned to multiple participants, which can consume the budget unnecessarily and hence decrease the clearance rate (CR) i.e., the percentage of the accomplished sensing tasks in an auction-based campaign [10]–[12].

Among other open problems, a principal challenge in the field of mobile crowdsensing is energy consumption, which is a concern for service demanders and providers. Previous research on energy-aware MCS featured techniques that aim at avoiding redundant measurements [16], reducing the sesning time [17], or minimizing the number of participants [13]. This research addresses energy-aware sensing, or *Green Mobile Crowdsensing* (GMCS) by presenting two new auction management techniques, *Green Auctioning* and *Predictive Auctioning*.

In Green Auctioning, we embed energy awareness in the task-participant matching using a newly formulated *hard reputation* term in the adopted objective function. This extends the work in [8], [9] which proposed a threshold-based hard reputation term, in addition to a soft reputation term to evaluate the trustworthiness of the acquired data. Without compromising the soft reputation, and instead of dealing with the hard reputation term as a threshold-based Boolean flag for a device's quality, our formulation of the hard reputation term computes a score for each participant/device based on the energy consumption of the specific embedded sensors that are required in each sensing task. With the proposed formulation, we aim at minimizing the power consumption of sensing campaigns by choosing the participants/devices with the least energy consumption.

The second proposed approach, Predictive Auctioning, addresses the challenge of having tasks that do not get completed after they got assigned to participants, which represents a waste of energy on the platform side and on the participants side as well. This is also closely related to the clearance rate of sensing campaigns. There are several factors that may result in this situation including the battery level, the battery temperature, the status of internet connection, and the tasks' energy demand and duration. This challenge has not been discussed before in the literature. We believe that this situation can be handled by accurate forecasting of the aforementioned factors in which we can capitalize on the advances in machine learning approaches [57].

The proposed approaches for GMCS require a real-life dataset, first, to evaluate the effectiveness of the hard reputation term using real figures for sensors' energy consumption, and second, to train machine learning models on energy and connection status forecasting using realistic data. Otherwise, the learned models would not generalize well in real-life situations. In general, there is a lack of publicly available datasets that are appropriate for our purpose. After exploring the literature, we decided to construct our own dataset–the MAGGIE dataset. The flowchart of the proposed system is shown in Fig. 1 and we elaborate on the construction and the features of MAGGIE in the following sections. Although previous research had used resource profiles [60], the proposed dataset enables us to profile users based on diverse parameters including their energy consumption, internet stability periods and charging periods, which are all key to achieve GMCS.

The contributions of this research are summarized as follows:

- We have built a real-life dataset that contains a wide variety of energy-related and internet connection-related performance data, in addition to hardware features of diverse smartphone models. The dataset was constructed by monitoring over 100 users, after having their consent, over a duration of 4 months using a mobile application specially developed for this purpose. We argue that the richness of parameters and features included in the dataset makes it a suitable test bed for future research in the field, and helps studying the patterns related to device usage. We provide this dataset publicly for the sake of transparency and for contributing to the research advances in MCS.
- To the best of our knowledge, we are the first to formulate the hard reputation of participants' devices in terms of their energy consumption. We investigate the impact of the new objective function on the energy consumption and the clearance rates of three recently proposed auction management techniques under varying number of auctions, tasks, and participants. Our extensive simulations show significant energy savings using the

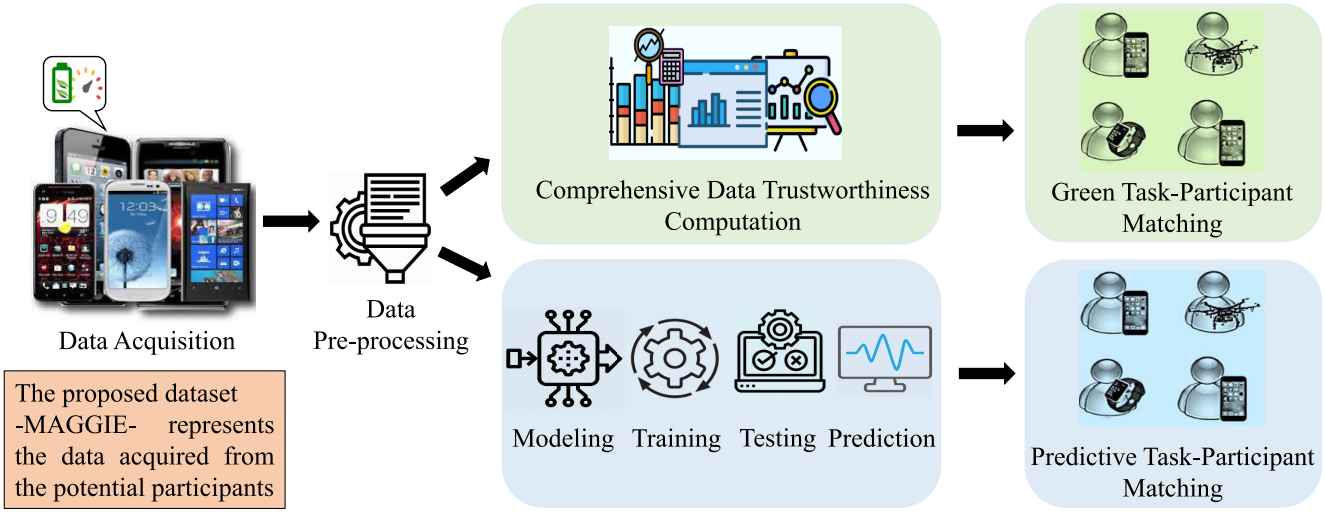

FIGURE 1. The architecture of the proposed framework. The first stage represents the acquisition of the *MAGGIE* dataset, which also represents the data gathered from potential participants in a real-life scenario. After pre-processing, the data are used to compute a hard reputation for the participating devices (towards computing comprehensive data trustworthiness) based on which we accomplish Green Auctioning. The second path represents the second proposed method–Predictive Auctioning. It involves learning energy and internet connection status patterns using LSTM neural networks. Please see text for more details.

proposed objective function, and highlight a trade-off between saving energy and maximizing the clearance rate (CR). We refer to this part of the contribution as *Green Auctioning* through the rest of this document.

- We demonstrate, using models learned from the MAGGIE dataset by means of LSTM neural networks, how we can forecast the parameters related to energy performance of the devices and the internet connection patterns of the participants. We also demonstrate how to incorporate this forecasting in developing a *Predictive Auctioning* framework. This framework aims at green crowdsening that avoids wasting energy by assigning tasks only to the participants who are capable of accomplishing them with regards to energy and internet demands. To evaluate Predictive Auctioning, we compare the performance of the original versions of three recent auction management techniques with the performance of their predictive versions. Since Predictive Auctioning is capable of choosing only highly qualified participants who can complete their tasks, we report significantly higher clearance rates achieved by the predictive versions of the techniques under consideration compared to the original versions of those techniques.

The remainder of this paper is organized as follows. We highlight related work from the literature in Section II. Then, we discuss the problem definition and explain necessary background information in Section III. Section IV is dedicated to describe the acquisition process, structure and the content of the proposed dataset. We present the proposed methods–Green Auctioning and Predictive Auctioning–in Section V and Section VI, respectively. Section VII demonstrates the effectiveness of the proposed techniques by analyzing the results of our extensive simulations, and finally section VIII concludes this research.

II. RELATED WORK

This section is meant to cover the relevant work in the recent literature. Particularly, we focus on previous research that addressed energy-aware sensing and/or that constructed datasets for this purpose.

One principal factor that affects accuracy and power consumption is the large variety of sensing devices that exist nowadays. There were many studies that focused on reducing the mobile power drawn during sensing, computing, and data transmission phases of an MCS application. For example, the authors of [13] adopted a coverage-based technique to minimize the number of participants while keeping area coverage intact. Another application-specific approach (for human body motion detection) was proposed in [14] where ultra-low-power passive static electric field sensing was used instead of traditional active accelerometer-based sensing. In [15], the authors replaced a power-hungry sensor with an energy-efficient sensor, while the sensing functionality is kept the same.

Other energy-aware techniques include [13] where a parallel transfer and delay tolerant mechanism is used to complete sensing tasks while users are placing their phone calls, in order to save energy in the data transmission phase. Usense's middleware is able to save energy using a mechanism that avoids taking measurements in those areas where there is already enough data or when the phenomenon is mostly invariant [16]. Also, the research in [17] considers allocating tasks that overlap in time to the same users, thus decreasing the overall sensing time. In this context, it is worth mentioning that a part of the literature assumes that multiple

jobs cannot be performed on a machine concurrently [18], while others [17] do not.

Previously, some researchers used mobile crowdsensing for gathering data such as the MIT Reality Mining dataset in which 100 Nokia mobile phones were given to undergraduates [45], in addition to the study in [43]. Also, the authors of [49] collected information over the course of one year from one model of Nokia smartphones. The work proposed in [46] studied installation and removal of Android applications on many devices. Due to the number and the nature of the gathered performance parameters, the number of devices' models, and the procedure of acquiring the datasets in these previous studies, we have figured out a necessity in constructing our own dataset, that lends itself well to the purpose of validating Green Auctioning and Predictive Auctioning.

III. BACKGROUND AND PROBLEM DEFINITION

In a typical auction, buyers offer increasingly higher prices in a competition to obtain goods. In MCS campaigns, service providers' competition is manifested as under-bidding for one another. This is called a *reverse auction*. In the literature of MCS, reverse auctions are used as a principal tool for the mathematical modeling of incentive mechanisms and task assignment. The terms *auction* and *reverse auction* are going to be used interchangeably in the rest of this article.

For crowdsesning campaigns that adopt auctions, the platform is responsible for matching tasks to participants given the requirements of the service demanders and the bids received from the participants. These bids represent a remuneration that covers their *sensing costs*. Winner participants, who are chosen by the platform to accomplish the sensing tasks, are determined using an objective function. This function is designed to optimize the demand (for example, maximizing the number of accomplished tasks or minimizing the energy consumption) given the resources (for example, the platform budget or sensors' energy consumption). Each sensing campaign involves the following [36]:

- N smartphones indexed with i and represent the participants in the auction, i.e., $i \in 1, \cdots, N$.
- M campaign tasks indexed with j and their details are sent by the platform to the participants, i.e., $j \in 1, \cdots, M$.
- A bidding process in which all participants should take part. Each bidder should bid on (or be interested in) at least one task.
- Winner selection and payment determination algorithms.

Towards effective management of auctions, several types of awareness are required in the coordinating platform, among which are energy awareness, hard reputation awareness, and connection pattern awareness. In the rest of this section, we explain each of the aforementioned awareness types and we highlight its impact on the quality of the gathered data, the clearance rate, as well as the energy performance/demand of crowdsensing campaigns. Also, a summary of the symbols and the notations that are used throughout this work is given in Table 1.

TABLE 1. Frequently used notations and symbols.

Symbol	Meaning
A_i	A binary flag indicating a participant's interest in a task based on its geographic location in Green Auctioning
B	Set of bids from participants in a sensing camping
$\mathcal{B}$	A sensing campaign's budget
b_i	The battery level of user i in Predictive Auctioning
$b_i(T^{St})$	The battery level of user i in the beginning of a sensing campaign in Predictive Auctioning
b^{Th}	The battery threshold level
c	A task category in Green Auctioning
$\mathcal{C}$	Set of tasks categories in Green Auctioning
CVD	The sensing campaign validity duration in Predictive Auctioning
$\{DT\}$	Set of data trustworthiness values for participants of an auction of category c in Green Auctioning
f_{ij}	The participant-task matching flag in Predictive Auctioning
H_d	The *one-hour time during CVD* that witnesses the maximum number of participants with stable internet connections and battery levels above b^{Th} in Predictive Auctioning
I_i^{Th}	The time interval that will elapse until the battery of user i reaches the threshold level in Predictive Auctioning
l_{max}	The maximum length of a sensing interval in Predictive Auctioning
$N_{h,d}^{Stable}$	The number of participants that have a stable internet connection at hour h of day d during CVD in Predictive Auctioning
$N_{h,d}^{b^{Th}}$	The number of participants that have a battery level of more than or equal b^{Th} at hour h of day d during CVD in Predictive Auctioning
$\mathcal{P}$	Set of participants in a sensing campaign
$\mathcal{P}_c$	Set of participants that their mobile devices can perform task of category c in Green Auctioning
$\{P_c\}$	Set of payments for winners of an auction of category c in Green Auctioning
$\{P\}$	Set of payments for winners of all auctions categories in Green Auctioning
R_i^h	Hard reputation of user i
R_i^s	Soft reputation of user i
R^s	Set of soft reputations for participants in a sensing campaign in Green Auctioning
S_c	Set of winners in an auction of category c in green auctioning
S	Set of winners in a sensing campaign
$\mathcal{T}_c$	Set of tasks that falls under category c in Green Auctioning
$\mathcal{T}$	Set of tasks in a sensing campaign
T^{St}	The starting time of a sensing campaign in Predictive Auctioning
$T_i(b^{Th})$	The time at which the battery of user i reaches the threshold level in Predictive Auctioning
t_j	The duration of task j in a sensing campaign in Predictive Auctioning
$\{TI\}$	The set of tasks intervals in a sensing campaign in Predictive Auctioning
Y_i	A ranking factor for user i according to their sensor's resolution in Green Auctioning
Z_i	A ranking factor for user i according to their sensor's power consumption in Green Auctioning

A. ENERGY AWARENESS

For every participant in a sensing campaign, the energy consumption can be written as:

$$E_i = \sum_{t \in T_i} (E_{t_i}^s + E_{t_i}^b), \tag{1}$$

where E_i is the energy consumption by participant i, T_i is the set of tasks done by participant i, $E_{t_i}^s$ is the sum of the energy consumed by all the sensors involved in task t_i by participant

i, and $E^b_{t_i}$ is the energy consumed by the battery in a task t_i by participant i excluding the sensors energy, e.g., processor, display, etc. The total energy consumed by all the winners in a sensing campaign [41] is given by:

$$E_{total} = \sum_{w \in S} E_w, \tag{2}$$

where S is the set of winners in an auction, and a subset of the set of participants in that auction. So, the cardinality of the set S should be less than the number of participants, N, in a campaign. Also, E_w is the energy consumption by winner w in a campaign. Aggregate energy savings in an MCS application can be achieved by decreasing the number of recruited nodes [38] or by decreasing the energy consumed by each node. Hence, energy consumption profiling of the participating smartphones and their embedded sensors is key for controlling the energy performance of sensing campaigns through energy-aware device-task matching.

Along the same lines, assigning an energy-demanding and/or computation-hungry task that requires a long sensing time to a participant with a low battery level or a high temperature battery would result in an uncompleted campaign. This is because the task would be left unfinished which costs the platform an unpredictable loss in the clearance rate and possibly extra computations to rerun the auction. Also, it would have a negative influence on the participants as well, draining their devices' power in vain, misusing their devices, and discouraging them from future participation. Decreasing the size of the participants pool may also impact negatively the quality of gathered data.

B. HARD REPUTATION AWARENESS

Sensitive tasks should not be matched to old mobile phone models which have less accurate and low resolution sensors. Even though users can have a high soft reputation [8] (high ratio of the positive readings to the total readings), their hard reputation, i.e., the accuracy of the sensors, should also be taken into consideration.

One principal aspect for maintaining high Quality of Service (QoS) in MCS systems is to maximize data trustworthiness. For user i, this is formulated as a function of a participant's soft reputation R^s_i and hard reputation R^h_i, and is given as [9]:

$$DT_i = f(R^h_i, R^s_i), \tag{3}$$

where s in R^s_i stands for *soft*, while s in $E^s_{t_i}$ in Eqn. 1 stands for *sensors*.

While previous studies [9] took R^h_i into consideration by merely requiring the participating devices to satisfy a minimum threshold, and assumed that the quality of the acquired data can only be affected by the users' soft reputation, we considered the devices' hard reputation in more depth. Particularly, we quantify a devices' hard reputation according to the specifications (for example, accuracy and resolution) of its sensors, and match devices to applications based on the latter's requirements.

C. CONNECTION PATTERN AWARENESS

We observe that there is a pattern in the participants usage of mobile devices, e.g., their power consumption and connection to internet. This pattern can be learned by means of machine learning models, after which the prediction of different sensing device aspects and/or participants' behaviors become achievable. In opportunistic and participatory sensing, being able to recognize the pattern of internet connection is key for managing sensing campaigns effectively. Whether it is due to unstable internet connection or deliberate disconnection from the internet, the prediction of such interruptions helps the platform maintain an acceptable quality of service through informed device-application matching. Moreover, being able to predict the battery level and the temperature of the devices, the platform can determine a device's readiness to accomplish specific tasks and predict potential participants ahead of holding an auction.

This research focuses on multi-step short-term univariate energy forecasting. We assume functional dependency between historical $(\{x_1, \ldots, x_{n-1}, x_n\})$ and future $(\{x^0_{n+1}, x^0_{n+1}, \ldots, x^0_{n+k}\})$ time series data points. We aim at realizing ***Predictive Mobile CrowdSensing*** using machine learning-based forecasting. Our goal is to enhance the task allocation strategy by building the platform's capacity to choose the participants who have sufficient resources, e.g., battery level and battery health, internet connection, in order to complete the tasks they have bidden on, in a manner that results in reducing energy consumption by avoiding energy wasting.

IV. THE CONSTRUCTION OF THE *Maggie* DATASET

The principal objective of constructing a dataset is to provide a trusted testbed for mobile crowdsensing research concerning energy issues. The dataset should be realisitc in terms of energy consumption patterns, comprehensive in terms of smartphone models and their monitorable features, labelled to enable machine learning of mathematical models, and publicly available to promote transperancy and to facilitate collaborative innovation.

To the best of our knowledge, this research provides the first real-life and comprehensive power profiling dataset for energy-related studies in the area of crowdsensing. It features a rich variety of sensing device models, types of embedded sensors, and hardware parameters. Statistics on the data provided by the proposed dataset is given in Appendix. The proposed dataset is also compatible with the following set of recommendations for researchers in ubiquitous computing: consent,[1] transparency, purpose, security, and proactive privacy preservation [21]. It is worth mentioning that several smartphone datasets have been collected for research purposes in a variety of domains, e.g., social evolution and healthcare [54], [55], activity recognition, reality mining [56], security [51], user interaction and energy consumption [50], [53], and device analyzer projects [52].

[1] The dataset has been gathered after getting the consent of the participants.

However, *MAGGIE*–our proposed dataset–differs in several aspects.

Different from the datasets proposed in [50], [51], [53], [54], the *MAGGIE* dataset includes diverse sensing device models. We argue that this is a more realistic scenrio to study energy consumption patterns, especially in mobile crowdsenisng. Also, opposite to SETI@Home [49] and FoldIt [50], data collection in the proposed framework is decentralized, and data processing is centralized. Most importantly, as will be highlighted in the dataset description below, MAGGIE is unique in terms of the energy-related features acquired from the users. Similar to SETI@Home, we avoid the collection of personal information to avoid privacy-related challenges. Moreover, similar to SETI@Home and Device Analyzer, MAGGIE is collected using a mobile application that runs on the devices of volunteering participants, i.e. we did not provide a special type devices to participants for the sake of the study. We argue that using personal devices helps reveal more realistic energy consumption patterns. Nevertheless, compared to the previously proposed datasets, MAGGIE has a lower sampling rate, particularly to avoid adding a burden of extra battery consumption on the participating smart phones. The rest of this section discusses the different stages of constructing *MAGGIE*.

A. STAGES OF DATASET CONSTRUCTION

The dataset construction is comprised of three stages, namely, data acquisition, local and server-side data storage, and server-side offline data processing. These three stages are depicted in Fig. 1, and we elaborate on each of them below.

1) DATA ACQUISITION

Towards building the dataset, we developed an android mobile application for smart phones using Android Studio Version 2.0. The application is designed to have a low resource footprint, with a negligable typical power drain and an avarage amount of disk space of 2.5 MB. We have adopted two time resolution profiles by sampling the dataset gathered parameters either every one hour or every 20 minutes. We show visual examples for the gathered parameters such as the battery level and the device temperature, presented as time series 1-D signals, in the following sections. Before implementing the application, the following characteristics had to be taken into consideration:

- Battery and disk space demands have to be minimized, not only to guarantee participants' convenience, but also to avoid biasing a mobile's power profile.
- The application should be compatible with all mobile models and API levels, in order to make a successful mobile crowdsensing experiment. Our application is compatible with APIs up to 28 (Pie, Orea, Nougat, ...).
- The application should operate in the background all the time, should be on the auto-launch list when the phone is switched on, and should be excluded from battery optimization and restricted background data features on the devices.
- Data should be uploaded to a secured server using a secured protocol, e.g., SFTP over SSH on TCP port 22, to protect the users' data.
- The interaction between a participant and the application should be minimal, and any special arrangements from the side of the participants (with regards to their usage pattern) should be avoided. Hence, the devices were not required to be rooted. Also, the volunteers were not required to open the application or even to check if it is running in the background.

2) LOCAL AND SERVER-SIDE DATA STORAGE

The data is stored temporarily on the device in a *.txt* file then a batch-uploading process is carried out to a server when an internet connection (WiFi or mobile data) is available. After data uploading, the local file is deleted to save the device resources and to prevent sending multiple copies of the same data.

3) SERVER-SIDE OFFLINE DATA PROCESSING

Following the step of data uploading to the server, an offline data pre-processing (or data cleaning) stage is carried out. It includes missing value imputation and outlier detection. This is usually time-demanding because raw data is rarely 100% complete. Missing values can arise from information loss and/or failure to load the information. The presence of missing values might result in biased insights, mainly due to the small number of data points, which can eventually compromises the reliability of the dataset. Also, different approaches for handling missing values and outlier detection can drastically change the results of data analysis. Therefore, adopting an effective approach to deal with these challenges is crucial. For the missing data in our dataset, we handled it using linear interpolation. For outlier data points, which may result from recording errors, data corruption, or equipment malfunction, they were identified by defining ranges and thresholds on the collected data. Aftewards, they were deleted or replaced to avoid introducing a bias in the statistical estimates such as the mean values of different data variables.

B. DATASET DESCRIPTION

The proposed dataset had been collected over a duration of 4 months from approximately 80 different device models, running Android versions from Android 5 upwards (starting from API Level 21) to guarantee the quality of the sensors and the collected data. Over 100 users have installed the software, and the dataset includes their age and gender information. MAGGIE is organized into tables such that every data table contains the following information[2]:

[2]With the consent of the volunteers, a sample of the dataset is publicly available for the research community through the following website: https://sites.google.com/view/the-maggie-dataset-webpage.

1) DEVICE AND BATTERY-RELATED DATA

- Participant identifier: This is a unique number for each device and its associated data. We used an Android ID that is given to each smartphone permanently. We could not use the IMSI (International Mobile Subscriber Identity) or the IMEI (International Mobile Equipment Identity) due to privacy issues.
- Mobile model: This is used to figure out which models consume more energy; hence, they can be avoided for the sake of green auctioning.
- Android version name
- Android version code
- Internet connection status: This is recorded with a time stamp in order to learn the connectivity pattern of the user. This allows a platform to predict the available participants ahead of receiving requests of sensing campaigns. We used the Wall-clock time, which is a real-time clock in Android with attached time zone information. It is used for displaying the local time to the user.
- Battery percentage, health status, and temperature

2) SENSORS-RELATED DATA

We gathered information about all the sensors embedded in a device. This information includes:

- Name: This makes it possible to determine which sensors are available on a device and consequently which sensing applications that device can be chosen to perform. Moreover, it shows us if the sensors are calibrated or not to help us increase the quality of the acquired data. Sensor calibration is a method of improving a sensor's performance by removing structural errors in its outputs. Structural errors are differences between a sensor's expected output and its measured output, which show up consistently every time a new measurement is taken. We can also know which sensors are wakeup.[3]
- Maximum range: This is the maximum value that can be measured by a particular sensor.
- Resolution: This is the smallest change that can be detected by the sensor.
- The power (in mA) consumed by the sensor while in use.

A sample of the text files generated by the application is shown in Fig. 2. In the rest of this document, we demonstrate two uses of the MAGGIE dataset, namely, Green and Predictable Auctioning. Green Auctioning features a device-application matching process that aims at minimizing the amount of power consumed by a sensing campaign. This is achieved using a matching objective function that penalizes energy-hungry sensors/device models. Predictive Auctioning also features a device-application matching technique, and aims that minimizing the *wasted energy*. Particularly,

[3]Wake-up sensors ensure that their data is delivered independently of the state of the system on a chip (SoC). While the SoC is awake, the wake-up sensors behave like non-wake-up-sensors. When the SoC is asleep, wakeup sensors must wake up the SoC to deliver events.

```
5ce9e81a28905cd8

HTCOne\_E8 dual sim 6.0.1 LOLLIPOP\_MR1
Oct 14, 2018 6:36:27 PM

Name:Accelerometer Sensor, maxRange:19.6133, Resolution:0.01, Power:0.17
Name:Magnetic field Sensor, maxRange:200.0, Resolution:0.01, Power:5.0
Name:CWGD  Orientation Sensor, maxRange:360.0, Resolution:0.1, Power:11.27
Name:CM36282 Proximity sensor, maxRange:9.0, Resolution:9.0, Power:0.18
Name:CM36282 Light sensor, maxRange:10240.0, Resolution:1.0, Power:0.15
Name:Magnetic Uncalibrated, maxRange:200.0, Resolution:1.0, Power:5.0
...
Battery Perc:41\%
Battery health = Good
Battery temp:38.0C
```

FIGURE 2. A sample of the generated files in the MAGGIE dataset from HTC One E8 Dual SIM.

by predicting the battery status of a candidate device, and given the task duration and the energy demand of that task, it prioritizes the devices that will be able to complete the task. Otherwise, starting the task then failing to complete it would constitute a non-rewarding loss of energy for the participants. It would also represent an additional energy consumption on the platform side since the auction needs to be re-run.

V. GREEN AUCTIONING

Towards the goal of holding green auctions, we propose a new objective function that takes into consideration the hard reputation of participants in addition to their soft reputation. This comprehensive reputation awareness enables the platform to guarantee data trustworthiness without compromising the energy consumption of a sensing campaign. The platform's energy awareness comes at the cost of more data overhead, where the hardware specifications of the sensing devices have to be sent from the participants to the platform. Although this data overhead is incurred only for new users, it is expected to be recurring (for existing users) over relatively-long periods during which the performance of the devices' components, e.g., the battery and the sensors, is expected to deteriorate. This takes the form of scheduled updates for the hardware specifications of the platform's database of devices.

To evaluate the data trustworthiness of user i as in Eqn. 3, R_i^h is computed by acquiring the resolution and the power consumption of each sensor ahead of holding auctions. For each task in a sensing campaign, the platform computes a ranking factor for every participant/device. This ranking factor depends on the sensor required to accomplish that task and on the specification identified by the service demander (for that task) as well. The ranking factor plays a key role in choosing the final set of participants in every sensing campaign. For user i, Y_i and Z_i are the ranking factors for that user according to the sensor's resolution and power consumption, respectively. Both factors, Y_i and Z_i, lie in the range [0, 1], and they are computed in a similar manner. The power consumption ranking factor, Z_i, is computed as:

$$Z_i = \frac{I_{max} - I_i}{I_{max} - 1}, \tag{4}$$

TABLE 2. A sample of the dataset showing the candidate sensing devices arranged in an ascending order according to the amount of power consumed by the on-board accelerometer.

Unique Identifier	Mobile Model	Android Version and API level	Sensor	Power	Factor (Z_i)
b0f71ded11e68c06	OPPOCPH1969	9, Pie,28	bmi160 accelerometer	0.001	1
edebda2f421a466a	samsungSM-A510F	7 M ,23	BOSCH Accelerometer Sensor	0.1	0.933
29ca73067f9d04ce	samsungSM-A750F	8, Oreo MR1,26	LSM6DSL Accelerometer	0.13	0.867
f3ab23ac989ef510	Lenovo K10a40 6	6, LOLLIPOP MR1,22	ACCELEROMETER	0.13	0.867
5ce9e81a28905cd8	HTC One E8 dual sim	6.0.1, LOLLIPOP MR1, 22	Accelerometer Sensor	0.17	0.6
980b6dc8d30a56b7	GooglePixel XL	7.1.1N,23	Accelerometer Sensor	2	0

where I_i is an index representing the order of the sensing device of user i among the other sensing devices in a campaign. This index is the output of sorting the participating devices according to a particular aspect. In Eqn. 4, this aspect is the sensors' power consumption. I_{max} is the largest index, and it is worth mentioning that I_{max} is not necessarily equal to the number of participants since we might have two or more users that get the same sorting index. Similarly, Y_i is calculated using the values of the resolution of the sensors required for a particular task. The overall R_i^h for user i is proposed to take the following formulation:

$$R^h_{i\in N} = A_i.Y_i.Z_i \;\; \forall \;\; i \in N; \;\; A_i \in \{0, 1\}, \tag{5}$$

where A_i is a binary flag indicating if a task is in a participant's region of interest or not. A participants' regions of interest is the geographic region that includes all the tasks a participant might be interested in. Beyond its borders, a participant would not bid on any task. In our simulations, we assume this region to be a circular region surrounding the participant, and we elaborate on the size of the participants' regions of interest in the Performance Evaluation section.

An example of the gathered power-related information is shown in Table 2. It represents the power consumption of the sensor under consideration–in this case, the accelerometer–in different device models, sorted in an ascending order. This structured data is a sample of the proposed dataset–MAGGIE. It is worth mentioning that MAGGIE features the two factors Z_i and Y_i for each sensor, while Table 2 is just a sample that shows Z_i only.

The previous discussion implies that the hard reputation of a participant in a sensing camping is determined by: 1) the task geographic location which is specified by the service demander, 2) the required sensor's power consumption which is specified by the nature of the task and the participating device model, 3) and the required sensor resolution which is specified by the service demander. It is worth mentioning that in case the service demander has no specific requirements for the sensors' resolution, the term Y_i can be eliminated from R_i^h. Also, the mathematical formulation of R_i^h can be extended to involve other aspects such as the maximum range of sensors which is anticipated, like the sensors' resolution, to be specified by the service demander. Furthermore, each term in the mathematical formulation of R_i^h can take the form of a hard constraint, such that if the power or the resolution do not satisfy the requirements of the service demander, the corresponding term can be set to zero, which leads to diminishing the probability of choosing that participant to join the campaign. The steps of Green Auctioning are shown in algorithm listing 1.

Algorithm 1 Green Auctioning

function Hold a Green Campaign($\mathcal{T}$, B, $\mathcal{B}$, R^s)
  Compute $\mathcal{C}$, $\mathcal{P}_c$, $\mathcal{T}_c$
  for $c \in \mathcal{C}$ **do**
    function Compute Data Trustworthiness for Participants($\mathcal{C}$, $\mathcal{P}_c$, R^s)
      for $i \in \mathcal{P}_c$ **do**
        Calculate A_i, Y_i, Z_i
        Compute R_i^h
        Compute DT_i
      end for
      return $\{DT\}$
    end function
    function Hold a Green Auction($\{DT\}$, $\mathcal{T}_c$, B, $\mathcal{B}$)
      Perform task-participant *matching*
      ⋮
      return S_c, $\{P_c\}$
    end function
  end for
  return S, $\{P\}$
end function

The algorithm starts by identifying the various categories of tasks in a sensing campaign and the participants that are capable of accomplishing each category. For each category, given the soft reputations of the participants and the characteristics of their on-board sensors, we compute the data trustworthiness of each participant–an essential step before holding the auctions. In Green Auctioning, a campaign is comprised of a set of auctions–an auction for each category. For each category, given the budget of the campaign, the data trustworthiness values, and the bids' values, we hold a green auction by performing a task-participant matching that adopts an objective function with the proposed comprehensive reputation term. Obviously, while looping over categories, the budget considered in each loop/category is the remaining budget from previous auctions. For brevity, we did

TABLE 3. A sample of a user profile in the MAGGIE dataset, showing data related to battery level/percentage, battery temperature, charging status, and internet connection status, with the corresponding time stamping information.

Date and Time	Battery Percentage	Battery Temp In C	Internet Status	Charging Status	Day Date	Day of Week	Month	Hour	Minute	Difference In Readings
8/7/2019 8:14	60	30.5	Not Connected	Not Charging	7	4	8	8	14	
8/7/2019 8:30	60	30.2	Not Connected	Not Charging	7	4	8	8	30	00:16:00
8/7/2019 8:38	59	29.9	Not Connected	Not Charging	7	4	8	8	38	00:08:00
8/7/2019 8:46	58	29.6	Not Connected	Not Charging	7	4	8	8	46	00:08:00
8/7/2019 8:58	58	29	Not Connected	Not Charging	7	4	8	8	58	00:12:00
8/7/2019 9:18	57	30.95	Not Connected	Not Charging	7	4	8	9	18	00:20:00

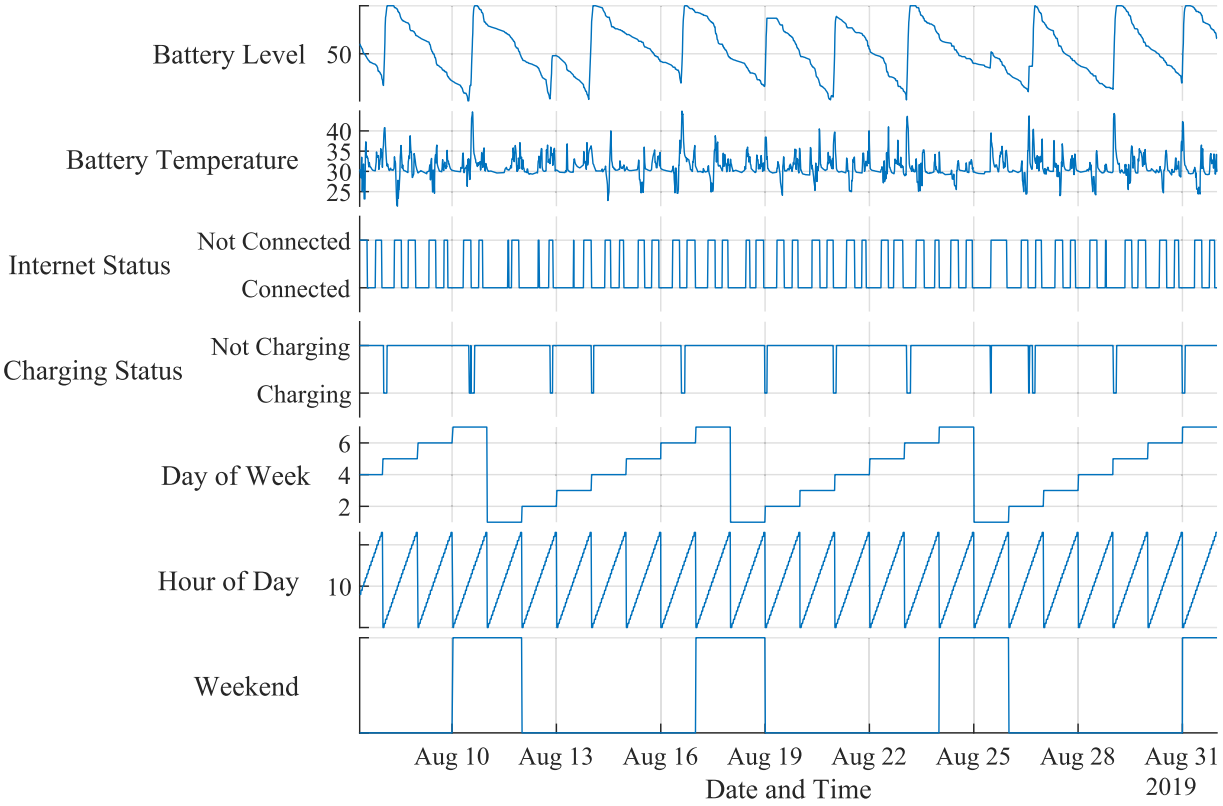

FIGURE 3. A sample of a user profile for nearly a month, showing patterns for battery level/percentage, battery temperature, charging status, and internet status.

not introduce a symbol for the remaining budget from each auction.

The proposed comprehensive formulation for data trustworthiness, which counts for both R_i^s and R_i^h, lends itself well to be used with other frameworks that aim at maximizing other aspects in MCS, such as the clearance rate (CR) [10]–[12]. In Section VI, we present the results of extensive simulations that compare the performance of various recently proposed CR-maximizing techniques that adopt comprehensive reputation (R_i^s and R_i^h), to their performance while considering R_i^s only.

VI. PREDICTIVE AUCTIONING

Given that MCS requires online, battery-powered systems, we propose a greedy algorithm that matches tasks to participants by taking into consideration that: 1) a task can be allocated to a device if and only if its sensing interval is smaller than the time expected for the device's battery to reach the *low battery* status, 2) a task can be allocated to a device if and only if it can maintain an acceptable battery health and battery temperature throughout the interval of the sensing task, and 3) a task can be allocated to a device if and only if it can maintain a sustainable internet connection throughout the interval of the sensing task. A sample of the MAGGIE dataset that is used for the simulations on Predictive Auctioning is shown in Table 3. It is worth mentioning that this is also the type of data acquired from the potential participants in a real-life implementation of Predictive Auctioning. We depict this framework in Fig. 1.

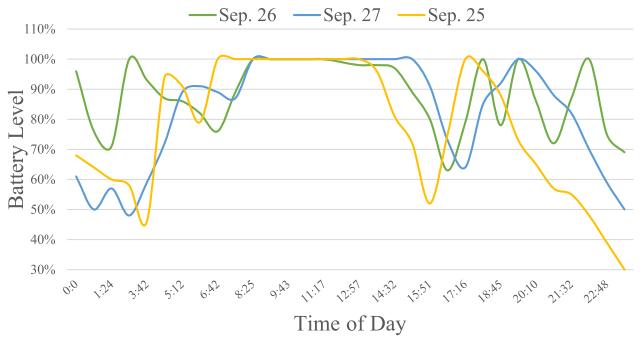

FIGURE 4. A sample of battery level variations over three successive days.

A. SYSTEM MODEL

Figure 3 shows the usage pattern for one participant which had been acquired over nearly a month. Particularly, it features an example for the data acquired by the platform from potential campaign participants for Predictive Auctioning, namely, battery level/percentage, battery temperature, internet connection status, and charging status. Similar to the system specifications that are acquired for Green Auctioning, the various time series for each user (such as those shown in Fig. 3) are expected to be updated according to a predefined schedule by the platform.

Based on our observation, there is a pattern in the aforementioned time series that can be learned, and hence predicted, for proactive management of sensing auctions. Among the user-specific patterns that we have observed in the collected data are:

- The battery reaches its maximum level at a specific time everyday. We argue that this could be used as the best time to recruit that user for battery-demanding tasks. For example, in Fig. 4, it can be seen that the device retains a maximum battery level from 6:45 am to 12:00 pm and has its least battery levels between 10:00 pm and 2:00 am.
- For some users, almost the same within-day pattern takes place for a particular day every week, yet the pattern is different from week to another. This is shown in Fig. 5 which represents a user's battery usage pattern for five consecutive Sundays.
- Some users tend to charge their mobiles when they reach a certain battery level as in Fig. 6, while other users charge their devices only during night.
- Profiling the battery temperature as in Fig. 7 showed that some smartphones become overheated during charging which can be really harmful for the battery. Assigning tasks to those devices at these times should be done after warning the participant.

For time series forecasting, this work has explored various machine learning methods. These methods are divided into traditional statistical methods and deep learning-based techniques, such as Auto-Regressive Integrated Moving Average (ARIMA) [62], and recurrent neural networks including long short-term memory (LSTM) networks [63], bi-direction

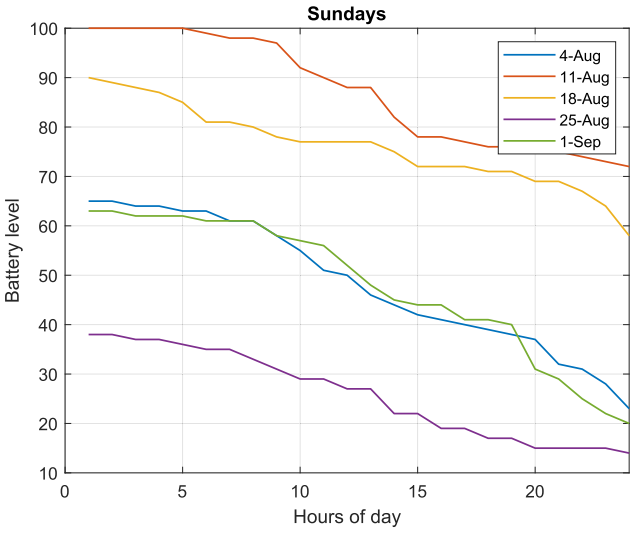

FIGURE 5. A sample of time series representing battery level variations on Sundays for 5 weeks.

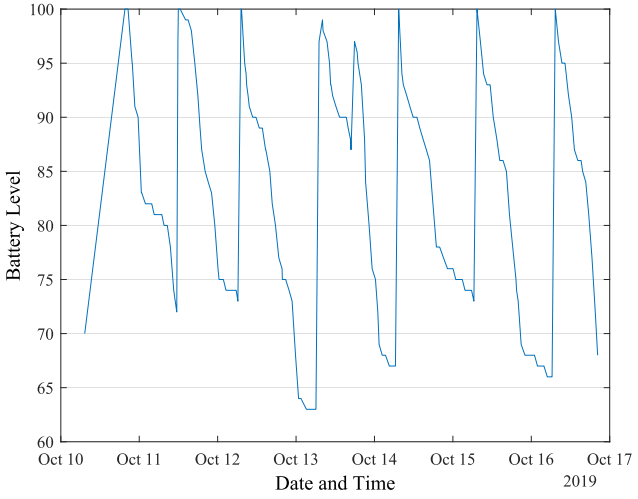

FIGURE 6. A sample of a user's energy profile showing a minimum battery level threshold after which mobile charging usually starts.

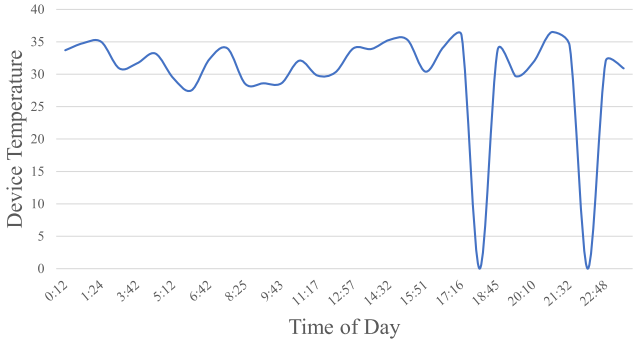

FIGURE 7. A sample of data showing variations of mobile temperature over a complete day.

recurrent neural networks, Elman networks, Jordan networks [64], and recursive neural networks. Particularly, we have explored the ability of different network configurations to capture the behavior of the data, and their

TABLE 4. Different models of time series prediction and their corresponding RMSE.

Models of Prediction	RMSE
Auto Regressive Moving Integrated Average (ARMIA)	9.9
Nonlinear Auto Regressive (NAR)	8.0
Long Short-Term Memory (LSTM)	6.79
Bidirectional LSTM (BiLSTM)	5.7
Group Method of Data Handling (GMDH)	9.68
Artificial Neural Network (ANN)	8.24

performance, measured in terms of the root mean squared error (RMSE), is shown in Table 4.

B. ALGORITHM DESIGN

In this sub-section, we present a greedy algorithm for task-participant matching that relies on predicting the battery level and internet connection status for the participants of sensing campaigns. In our model, each of the M tasks in a campaign is associated with a sensing interval indicating the period required to complete that sensing task. The set of sensing intervals, $\{TI\}$, that are associated with the sensing tasks are assumed to be known to the platform. Also, even though MCS requires online devices, not all sensing campaigns are required to be held online. For example, the service demander might have the flexibility to permit the platform to hold the auction whenever it has access to more participants, higher data trustworthiness, or any other aspect that is considered favorable, by the system designers, for the quality of the campaign. We argue that this scenario is valid for many real-life applications.

In the beginning of every campaign, the platform receives the following information from the service demander: the tasks of the campaign, the set of sensing intervals of all tasks ($\{TI\}$), and the campaign validity duration (CVD). The latter is defined as the duration throughout which the sensing is allowed. The platform then predicts the internet connection status and the battery level, throughout CVD, for each individual in the platform's users database. Since we have not started the auction yet, we call the users in the platform's database *potential participants*. From the predicted patterns, the platform determines H_d, which is defined as the *one-hour time during CVD* that witnesses the maximum number of individual profiles predicted to have stable internet connection and battery level above b^{Th}–the *battery threshold level*, which is set to 20% in this research. It is worth mentioning that a recent measurement study [12] showed that an energy-hungry sensor like the GPS can consume up to 15% of the total energy. This is the reason behind our choice for the battery threshold level. The optimization problem that we solve to determine H_d is given as:

$$H_d = \underset{h,d}{\text{argmax}} \ (N_{h,d}^{Stable} + N_{h,d}^{b^{Th}}), \tag{6}$$

where $N_{h,d}^{Stable}$ is the number of potential participants that have a stable internet connection at hour h of day d during CVD, and $N_{h,d}^{b^{Th}}$ is the number of potential participants that have battery levels above or equal b^{Th} at hour h of day d during CVD. Equation 6 is optimized by brute-force search. The starting time of the auction, T^{St}, is then set to be the onset of H_d. With regards to pattern prediction, and considering the results obtained in Table 4, we focus on LSTM and BiLSTM in this research. Figure 8 and Fig. 9 present an example for the time series forecasting obtained using LSTM and BiLSTM, respectively, on MAGGIE.

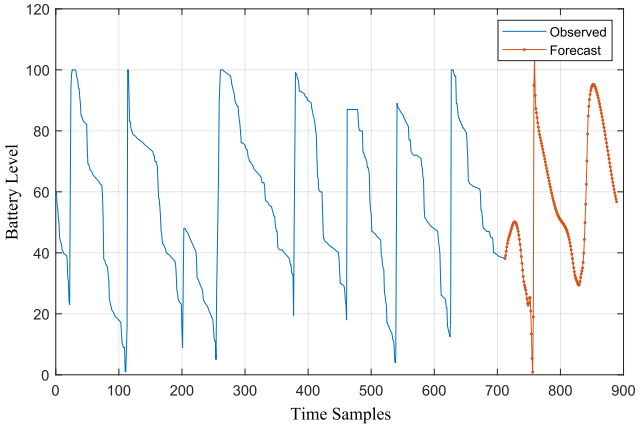

FIGURE 8. A data sample showing LSTM-based forecasting of battery level.

Based on the predicted patterns, and after identifying T^{St}, the platform determines the battery level of each potential participant at the onset of the auction (the onset of H_d). This level is denoted as $b_i(T^{St})$, such that b_i is the battery level for user i. The platform also determines the time at which the battery of user i reaches the threshold level (b^{Th}), which is denoted as $T_i(b^{Th})$. The time interval that will elapse until the battery reaches that level can be computed as:

$$I_i^{Th} = T^{St} - T_i(b^{Th}). \tag{7}$$

Even though the battery threshold level, b^{Th}, is set to 20% in this research, it can be allowed to vary according to the power requirements of the tasks and the allowed greediness level of auctions.

In our experiments, we also classified the participants into 5 different categories according to their $b_i(T^{St})$. These categories are shown in Table 5 and Fig. 10. According to the category of the users, they are allowed to bid on particular tasks. For each user, these tasks must require less time to be completed than the user's I_i^{Th}, which ensures the completion of these sensing tasks given the user's battery level in the beginning of the auction. An example is illustrated in Fig. 11 to clarify the concept. In this example, there are three sensing tasks A, B, C, each with different sensing interval. User 1 is allowed to bid on task A or B or both of them, but bidding on task C is not allowed as it needs more sensing time than I_1^{Th}. Whereas user 2 is permitted by the platform to bid on any of these tasks as they all require less than I_2^{Th}.

To realize this strategy for participant-task matching, the platform calculates a matching flag for each

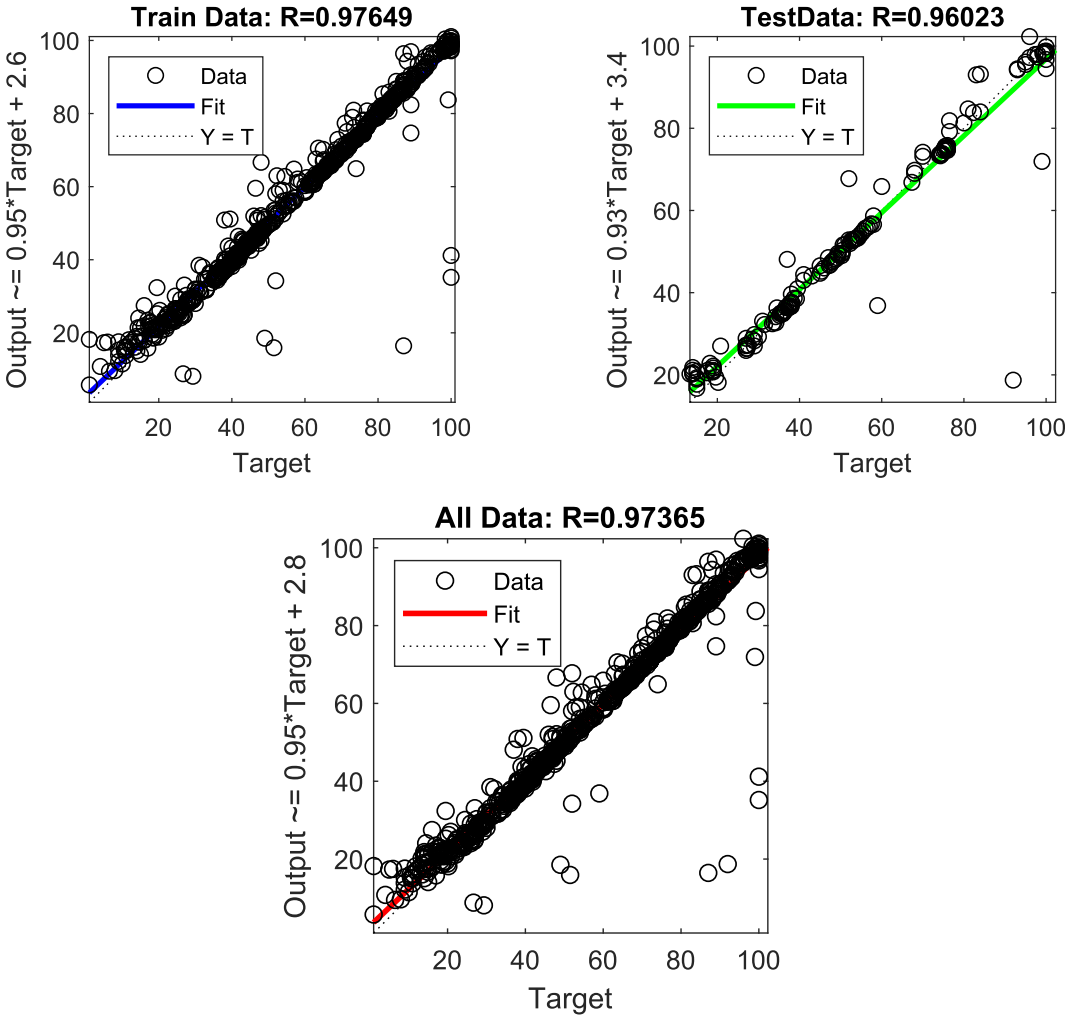

FIGURE 9. Using BiLSTM to forecast time-series on MAGGIE. The figure plots linear regression between the targets and outputs on the training and test samples. The best linear fit (the neural network simulation and predictions) is denoted by a solid line, whereas the perfect fit (outputs exactly equal to targets) is represented by the dotted line. The correlation coefficient (R) - mentioned above each plot - represents the proximity between the best linear fit and the perfect fit of the model. In this figure, with R nearly equals 1, the best linear regression–whose equation is written on the y-axis–nearly overlaps with the perfect linear fit which indicates a very strong correlation between the targets and outputs of the prediction model.

TABLE 5. Different categories of mobile users according to their battery level at the of starting time a sensing campaign.

Category	Battery Level
1	$80 \leq b_i(T^{St})$
2	$60 \leq b_i(T^{St}) < 80$
3	$40 \leq b_i(T^{St}) < 60$
4	$20 \leq b_i(T^{St}) < 40$
5	$b_i(T^{St}) < 20$

task-participant pair such that:

$$f_{i,j} = \begin{cases} 1, & \text{if } t_j \leq I_i^{Th} \\ 0, & \text{otherwise,} \end{cases}$$

where t_j is the duration of task j in a sensing campaign. After computing the matching flag, the auction is held by sending each potential participant only the tasks that they can complete, i.e., the tasks that correspond to a flag of value 1. This empowers the users of all categories, especially 3 and 4 (please see Table 5), to bid/focus on more achievable tasks, rather than tasks they will not be able to complete, and would burden the batteries of their devices to no avail. Category 5 users are not permitted to participate as they have all-zero flags. The steps of Predictive Auctioning are given by algorithm listing 2.

VII. PERFORMANCE EVALUATION

In this section, we present an evaluation for the performance of Green Auctioning and Predictive Auctioning using the

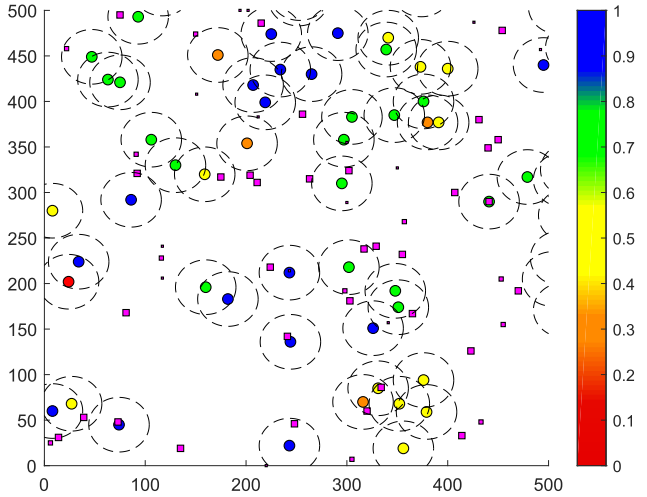

FIGURE 10. Participants are shown as dots with different colors that refer to their different categories as shown in Table 5. The color scale is shown to take a range 0-1 which corresponds to 0% and 100% battery level, respectively. The participants' areas of interest are depicted as dashed circles and the tasks are shown as violet squares with different sizes according to their sensing interval.

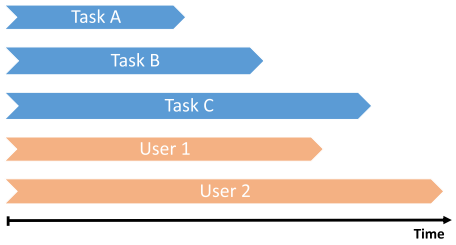

FIGURE 11. An illustration of our task allocation strategy that integrates the required sensing interval of tasks with the battery level of participants. In this example, User 1 is allowed to bid on Task A and Task B, but not on Task C since its interval is longer than what the battery of User 1 can afford. User 2, however, is allowed to bid on the three available tasks.

Algorithm 2 Predictive Auctioning

function Hold a Predictive Auction($\mathcal{T}$, B, $\mathcal{B}$, $\{TI\}$, CVD)
  Predict *connection status during CVD*
  Predict *battery levels during CVD*
  Compute $N_{h,d}^{Stable}$, $N_{h,d}^{b^{Th}}$
  Compute H_d
  Determine T^{St}, $b_i(T^{St})$, $T_i(b^{Th})$
  Compute I_i^{Th}
  Compute f_{ij}
  Perform task-participant *matching*
  return S, $\{P\}$
end function

MAGGIE dataset. For both techniques, and to facilitate the comparison with recently proposed methods, we have followed the literature in setting the values for the simulation parameters, and we elaborate on this point below. The area of the simulation was set to 1000 m × 1000 m where the tasks and the participants are distributed uniformly. For the

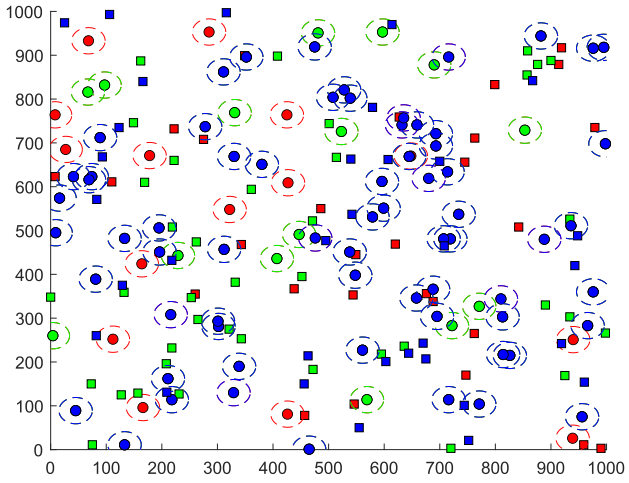

FIGURE 12. An illustration showing the area of the campaign, with participants depicted as dots and tasks as squares. The tasks and participants are shown in 3 different colours (blue, green and red), which represent accelerometer, gyroscope and magnetometer sensor-specific tasks and participants. Each participant's area of interest is marked with a dashed circle.

auctions, the collective bids of the participants were varied uniformly in the range [1, 10], while the values of the tasks were varied in the range [1, 5]. The per-task bids were varied uniformly in the range $[v_j - \alpha, v_j + \alpha]$, and we set $\alpha = 2$ in our simulations. The soft reputations of the participants were varied uniformly from 0.6 to 0.9, where the area of interest around each participant was set to a radius of 30 m as depicted in Fig. 12.

A. PERFORMANCE EVALUATION FOR GREEN AUCTIONING

For evaluating the effectiveness of the Green Auctioning algorithm, we developed a *green version* of three algorithms from the recent literature, namely, *TSCM* [8], *HMCDB* [10] and *RPB* [10], and we compare the performance of the green versions with the original versions. These algorithms are all reputation- aware (RA), but the main difference in the green versions is that they adopt the newly proposed objective function for task-participant matching, while the original versions adopt an objective function that includes R^s (soft reputation) only. It is worth mentioning that we mapped the R^h (hard reputation) factor to the range [0.5, 1] in order to be close to the range of the participants' soft reputation so that they have nearly equal influence, and Z_i from the range [0, 1] to [0.1, 1].

With the simulation results shown below, we aim at highlighting the impact of involving the R^h as a factor in managing auctions. Three aspects are considered (allowed to vary) in our simulations which are: the number of campaigns, the number of tasks, and the number of participants. Table 6 summarizes the simulated scenarios and their corresponding parameter values. It is worth mentioning that in Green Auctioning, a campaign is comprised of multiple auctions, each for every task category.

In the first row of Fig. 13, we show a comparison with regards to the attained clearance rate between the green

TABLE 6. A summary of the different simulated scenarios for Green Auctioning and their corresponding parameter values.

	Impact of #Campaigns		Impact of #Tasks		Impact of #Participants	
Parameters	*Values*	*Case*	*Values*	*Case*	*Values*	*Case*
#Tasks	100	Constant	100-500	Increasing by 100	100	Constant
#Participants	100	Constant	1000	Constant	100-1000	Increasing by 100
#Auctions	100-1000	Increasing by 100	5 for each tasks case	Constant	5 for each participants case	Constant

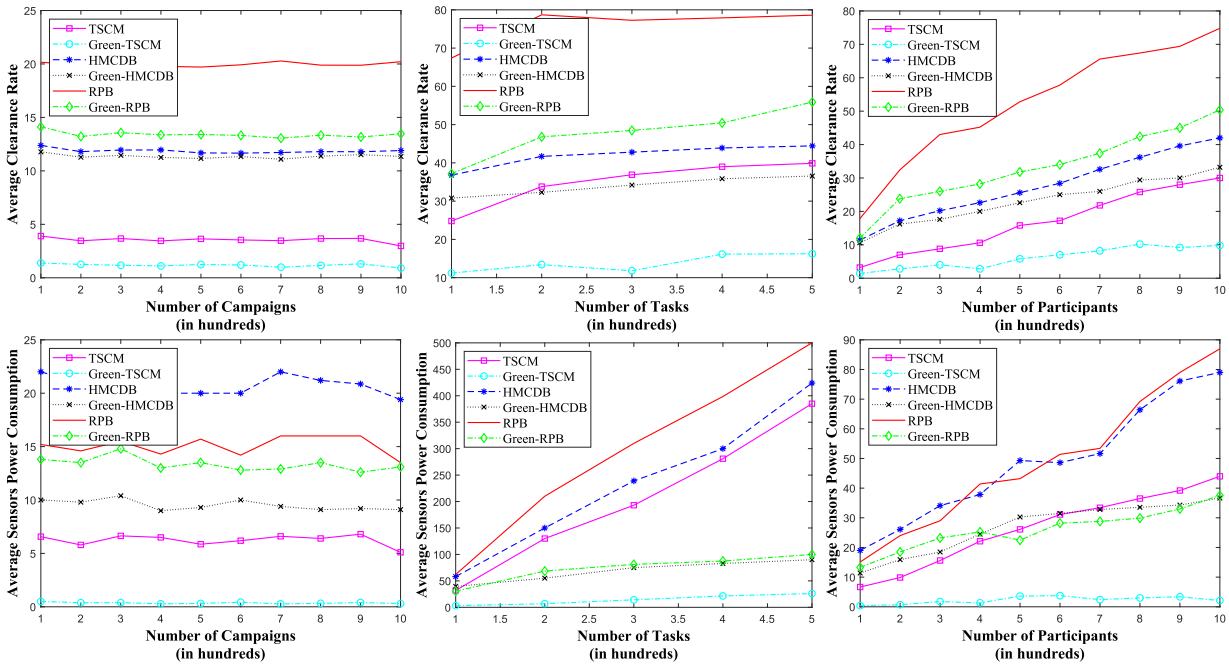

FIGURE 13. The impact of varying the number of campaigns, tasks, and participants on the average power consumption and the attained clearance rate (CR). The comparison takes into consideration three recent auction management techniques, namely, TSCM, HMCDB, and RPB. The figures compare the performance of the original versions of these three techniques with the performance of their green versions, i.e., the versions that adopt the proposed objective function for task-participant matching. Please see text for more details. As for the units of the vertical axes: For the upper row, the clearance rate is the percentage of completed tasks from the total number of tasks in a sensing campaign. For the lower row: The Android operating system responds with Ampere readings to queries regarding power consumption, so the power consumption is measured in mA.

versions of TSCM, HMCDB, and RPB and their original versions. This is done under varying number of campaigns (first column), varying number of tasks (second column), and varying number of participants (third column). The second row shows a comparison between the same candidates with regards to the consumed power. Generally speaking, there is a trade-off between clearance rate and power consumption. This is expected because even with the ability to identify the devices that consume the least power, requiring more tasks to be done will require more power consumption from the sensors. Meanwhile, it is of significant importance to compare how much power has been saved for how much loss in clearance rate, and vice versa.

In Fig. 14, we show the relative percentage change in the consumed power and the clearance rate between the green and original versions of TSCM (row 1), HMCDB (row 2), and RPB (row 3). This was simulated under varying number of campaigns (column 1), tasks (column 2), and participants (column 3). The relative percentage change in the power consumed is defined as:

$$\epsilon\% = \frac{E_{original} - E_{green}}{E_{original}} \times 100, \tag{8}$$

where $E_{original}$ and E_{green} are the energy consumed by the original and green versions of the technique under consideration, respectively. Similarly, the relative percentage change in the clearance rate is defined as:

$$\eta\% = \frac{{r_c}^o - r'_c}{{r_c}^o} \times 100, \tag{9}$$

where r_c^o and r'_c are the clearance rate attained by the original and green versions of the technique under consideration, respectively. The blue and red bars in Fig. 14 represent $\eta\%$ and $\epsilon\%$, respectively. For TSCM and HMCDB, even though

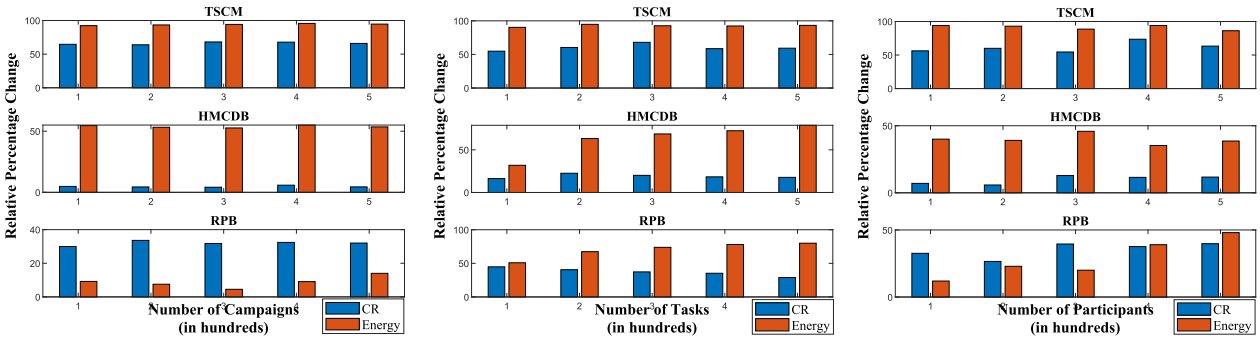

FIGURE 14. The relative percentage change in energy consumption and clearance rate due to Green Auctioning. This is computed for TSCM, HMCDB, and RPB, under varying number of auctions, tasks, and participants. There is a trade-off between energy consumption minimization and CR maximization. Nevertheless, the figure shows that in the majority of the comparison cases, the attained energy savings are more than the resulting CR decline. Please see text for more details.

TABLE 7. A summary of the different simulated scenarios for Predictive Auctioning and their corresponding parameter values.

	Impact of #Auctions		Impact of #Tasks		Impact of #Participants		Impact of Max. Intervals Length	
Parameters	*Values*	*Case*	*Values*	*Case*	*Values*	*Case*	*Values*	*Case*
#Tasks	100	Constant	100-500	Increasing by 100	100	Constant	100	Constant
#Participants	100	Constant	1000	Constant	100-1000	Increasing by 100	100	Constant
#Auctions	100-1000	Increasing by 100	5 for each tasks case	Constant	5 for each participants case	Constant	5 for each Interval case	Constant
Maximum Sensing Interval	40	Constant	40	Constant	40	Constant	20-120	Increasing by 20

the green versions of both techniques have resulted in CR decline, the energy savings are more than the CR loss, since the red bars ($\epsilon\%$) are bigger than the blue bars ($\eta\%$). This is consistently the case with varying number of campaigns, tasks, and participants.

With RPB, though, energy savings do not consistently outweigh CR losses. Particularly, with varying number of auctions and in some cases of varying the number of participants, the green version results in energy savings but at a high cost in terms of CR decrease. In part, this is expected since RPB is an auctioning technique that focuses on maximizing CR by penalizing redundant task assignment and by giving high priority to tasks with less bidders to increase the odds of their accomplishment. So, it is more inclined towards accomplishing more tasks, and hence more energy consumption. This explains why its green version finds it challenging to achieve energy savings that are higher than the *CR loss*.

B. PERFORMANCE EVALUATION FOR PREDICTIVE AUCTIONING

For evaluating predictive auctioning, we used the same simulation area and the same settings for the value of each task, the participants' bids, and the participants' area of interest as the case with Green Auctioning. Also, similar to Green Auctioning, the performance was investigated under varying number of auctions/campaigns, tasks, and participants. We also investigated the impact of varying the maximum allowable sensing interval as will be shown later in this section. It is worth mentioning that in Predictive Auctioning, a campaign is comprised of one auction, different from Green Auctioning where auction is comprised of multiple auctions, each for every task category. In our simulations, we set the interval of the tasks to be directly related to the values of the tasks, i.e. a task with a high value implies that it has a long sensing interval. The maximum length of a sensing interval is denoted by l_{max}. The length of the sensing interval of a task is chosen randomly in [4, $lmax$] time slots. The time slot is one minute. Table 7 summarizes the simulated scenarios and their corresponding parameter values.

As mentioned in subsection VI-B, and according to the RMSE values obtained and shown in Table 4, our simulations for predictive auctioning adopted the BiLSTM model. The initial learning rate was set to 0.005, the learning rate drop factor was set to 0.2, the learning rate drop period was set to 125. The number of hidden layers in both units was set to 50, the maximum number of training epochs was set to 250 and the minimum batch size was set to 120. A train-test split

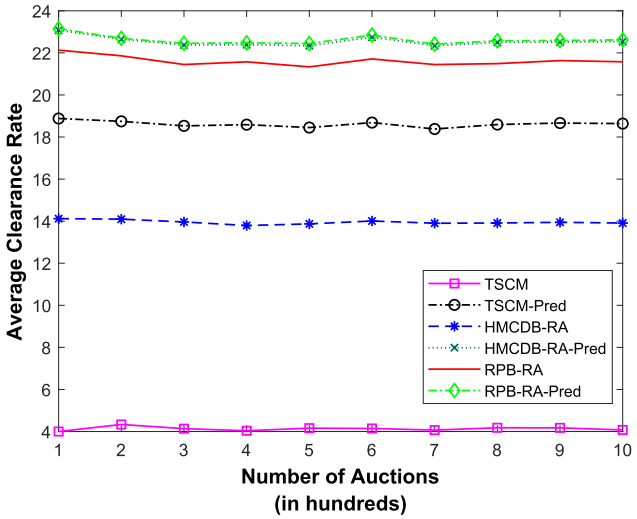

FIGURE 15. The impact of varying the number of auctions on the performance of the reputation-aware (RA) original and predictive versions of TSCM, HMCDB, and RPB with regards to the attained clearance rate.

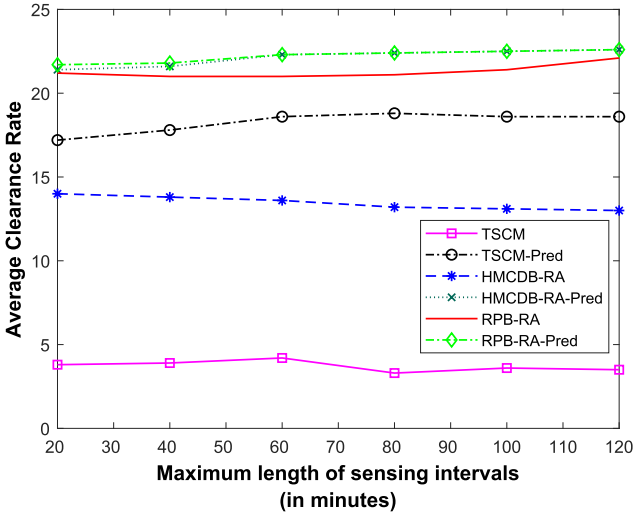

FIGURE 17. The impact of varying the tasks' maximum sensing interval on the performance of the reputation-aware (RA) original and predictive versions of TSCM, HMCDB, and RPB with regards to the attained clearance rate.

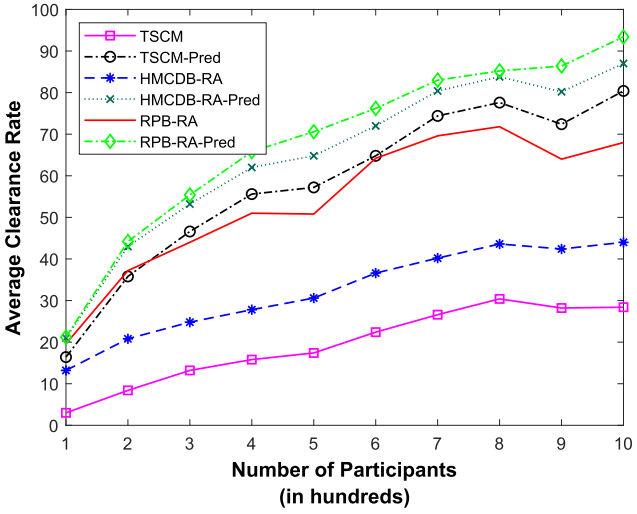

FIGURE 16. The impact of varying the number of participants on the performance of the reputation-aware (RA) original and predictive versions of TSCM, HMCDB, and RPB with regards to the attained clearance rate.

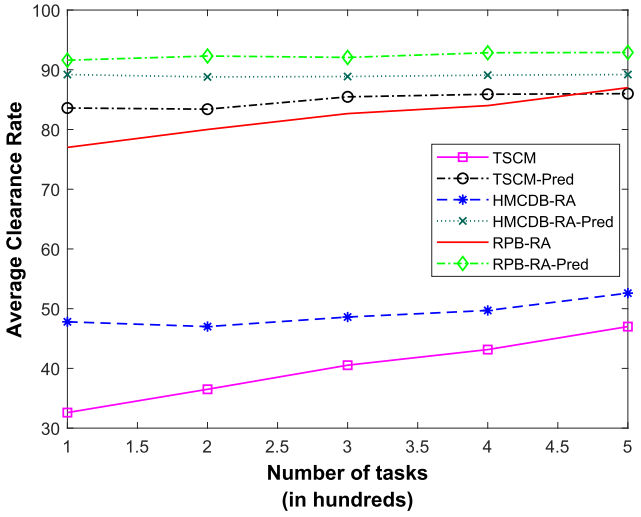

FIGURE 18. The impact of varying the number of tasks on the performance of the reputation-aware (RA) original and predictive versions of TSCM, HMCDB, and RPB with regards to the attained clearance rate.

of 80-20 was adopted and the model was built using the Deep Learning Toolbox of Matlab.

As a performance baseline, we chose the same auctioning techniques that were used to evaluate Green Auctioning, namely, TSCM [8], HMCDB [10], and RPB [10], which are all representatives of reputation-aware (RA) techniques. In the rest of this sub-section, we compare the performance of the original versions of the aforementioned techniques with the performance of their predictive versions. Particularly, we focus on the clearance rate, and we show that predictive auctioning consistently yields an increase in the number of accomplished tasks in a sensing campaign.

With regards to the number of auctions, Fig. 15 shows that the predictive versions of the algorithms under considerations attain higher clearance rate compared to the original versions that are reputation-aware only. In addition to being reputation aware, the predictive versions are capable of choosing tasks to participants in a manner that maximizes the number of accomplished tasks. The predictive TSCM attained a CR that is more than 4 times that of its original version. The predictive HMCDB and the predictive RPB achieved an equal CR of 23%, while their original versions attained 14% and 22% respectively.

As for the number of participants, generally speaking, as the number of participants increases, the higher the attained CR, because the number of task executers becomes larger. Fig. 16, shows that the predictive RPB algorithm achieves an average of approximately 30% increase in CR compared to the original RPB over varying number of participants. Predictive RPB is closely followed by the predictive HMCDB

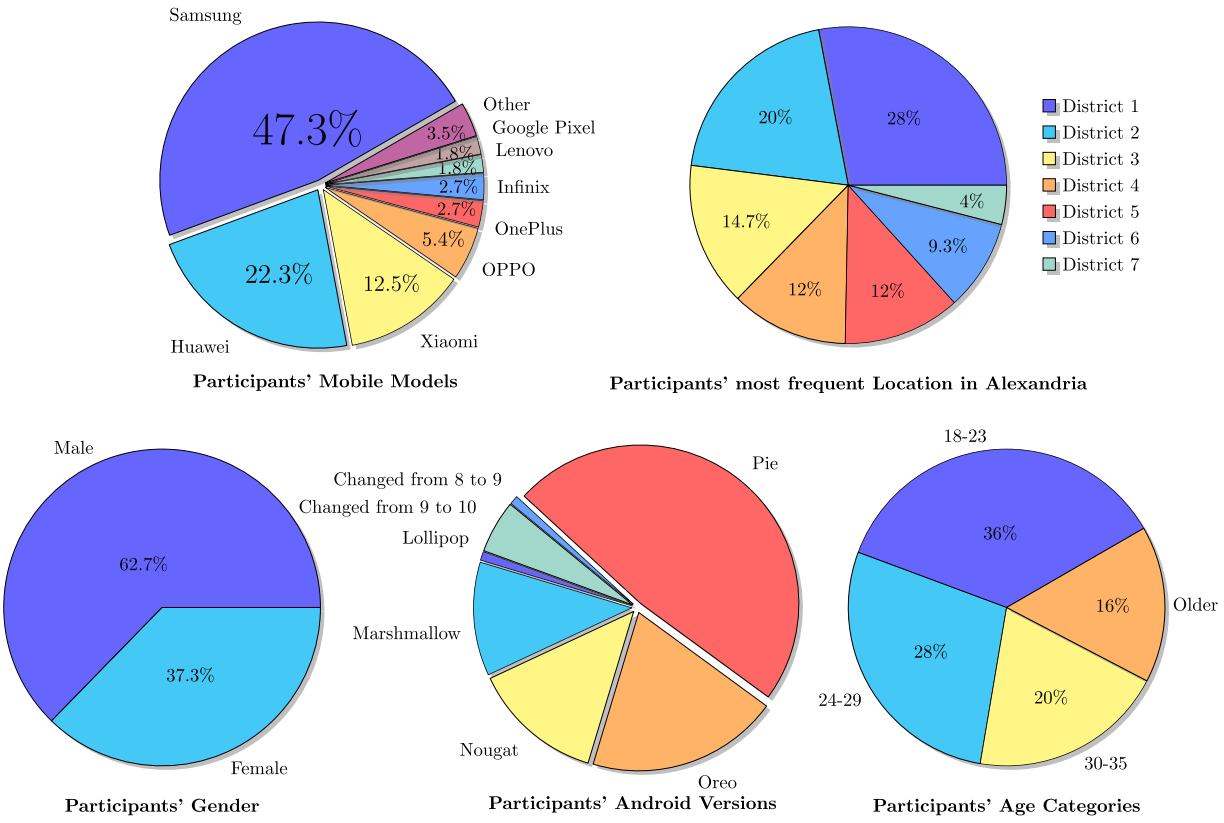

FIGURE 19. Statistics on the data provided by the MAGGIE dataset.

algorithm and the predictive TSCM. The predictive TSCM algorithm yields almost 50% larger CR compared with the original algorithm at 1000 participants.

In our simulations, the maximum length of sensing intervals is allowed to vary from 20 to 120 minutes with increments of 20 minutes. Figure 17 highlights the impact of varying the maximum sensing interval on the attained clearance rate. While the attained CR, in general, does not demonstrate significant change over the range of simulated intervals, there are some trends that are worth mentioning. For the original TSCM and HMCDB, with more tasks having longer intervals, there is a slight decrease in the CR. This is because more tasks will need larger sensing interval than the user can handle. The case with the original RPB is different, though, because it involves budget saving by penalizing redundant sensing in addition to three stages for winners selections, namely, the redundancy, the primary and the secondary winners, compared to only two stages in TSCM and HMCDB. Consequently, the three sets of winners in RPB are capable of executing more tasks. This also implies larger coverage area, and hence can handle more tasks, which saves the tasks that have less sensing intervals from being dropped, such as in the case of TSCM and HMCDB. The predictive versions of all the techniques under consideration consistently lead to increasing the clearance rate, thanks to assigning the sensing tasks to the users that are qualified, battery-wise, to accomplish them. The superiority of the predictive versions are also valid for the case of varying the number of tasks as shown in Fig. 18. The pool of tasks provided within the area of interest of each participant gets larger when the number of tasks increases. Hence, both original and predictive versions demonstrate an increase in CR with increasing the number of tasks. While the original versions show a higher rate of CR increase (larger slopes), the predictive versions have preserved consistently higher clearance rates.

It should be stated that in the above comparative analysis, the platform in all our simulations had been given an access to the same budget in each case/technique under consideration. Hence, the predictive versions were able to achieve significantly higher clearance rates using the same resources/budget.

VIII. CONCLUSION

In this research, we proposed two approaches for *Green Mobile Crowdsensing* (GMCS), aiming at reducing the power consumption in mobile crowdsensing campaigns through effective device-application (or task-participant) matching. In the first approach, Green Auctioning, we proposed a new objective function that represents a comprehensive formulation for user reputation. Previous techniques adopted an objective function with a soft reputation term that depends on data trustworthiness, in addition to a threshold-based hard reputation term which reflects the quality of the sensing device. Without compromising soft reputation, we presented

a new formulation for the hard reputation term that depends on the energy consumption of the device. By considering three recently proposed, state-of-the-art auction management techniques (TSCM, HMCDB, and RPB), we compared the performance of their original and green versions, and we showed significant energy savings in return of a decline in clearance rates—a trade-off that empowers system designers to optimize specific system characteristics. To further analyze this trade-off, under varying number of auctions, tasks, and participants, we showed that the attained energy savings were higher than the CR decline in the majority of simulated scenarios.

The second proposed approach for GMCS, Predictive Auctioning, relies on forecasting the energy and internet connection-related patterns of sensing devices using LSTM neural networks. It is a task-participant matching algorithm that assigns tasks only to users who can maintain sufficient battery levels and a stable internet connection throughout the sensing interval of the task at hand. Hence, we can avoid the waste of energy that results from the case of a user failing to accomplish a task after being assigned to it. Similar to Green Auctioning, we compared the performance of the original and predictive versions of TSCM, HMCDB, and RPB, and we showed a consistent advantage for predictive auctioning with regards to the attained clearance rates, under varying number of auctions, tasks, participants, and sensing intervals. It is worth mentioning that for the 1st and the 2nd approaches, the reported advantages were achieved without compromising other aspects such as data trustworthiness, since the proposed objective functions lend themselves well to optimizing further performance parameters.

Towards developing and evaluating the aforementioned approaches, we have built a new dataset, the MAGGIE dataset, that is comprised of battery-related, embedded sensors-related, and internet connection-related data of over 100 users, acquired over a duration of 4 months, using a mobile application that was specially developed for this purpose. To the best of our knowledge, the proposed dataset is the first to include that variety of sensing device models and their respective performance data. We provide the new dataset publicly to the research community for the sake of transparency and to contribute to the advance of MCS.

APPENDIX
STATISTICS ON THE DATA PROVIDED BY THE MAGGIE DATASET

See Figure 19.

ACKNOWLEDGMENT

The authors would like to thank this deep gratitude to Prof. A. Osman of the Construction and Building Engineering Department, Arab Academy for Science and Technology in Egypt, for generously taking the time to generate-then-provide them with the map of the city of Alexandria, Egypt, in which the dataset acquisition took place.

REFERENCES

[1] Y. Chon, N. D. Lane, F. Li, H. Cha, and F. Zhao, "Automatically characterizing places with opportunistic crowdsensing using smartphones," in *Proc. ACM Conf. Ubiquitous Comput. (UbiComp)*, 2012, pp. 481–490.
[2] P. Jayaraman, C. Perera, D. Georgakopoulos, and A. Zaslavsky, "Efficient opportunistic sensing using mobile collaborative platform MOSDEN," in *Proc. 9th IEEE Int. Conf. Collaborative Comput., Netw., Appl. Worksharing*, Oct. 2013, pp. 77–86.
[3] L. Xiao, Y. Li, G. Han, H. Dai, and H. V. Poor, "A secure mobile crowdsensing game with deep reinforcement learning," *IEEE Trans. Inf. Forensics Security*, vol. 13, no. 1, pp. 35–47, Jan. 2018.
[4] M. A. Alsheikh, Y. Jiao, D. Niyato, P. Wang, D. Leong, and Z. Han, "The accuracy-privacy trade-off of mobile crowdsensing," *IEEE Commun. Mag.*, vol. 55, no. 6, pp. 132–139, Jun. 2017.
[5] W. Nan, B. Guo, S. Huangfu, Z. Yu, H. Chen, and X. Zhou, "A cross-space, multi-interaction-based dynamic incentive mechanism for mobile crowd sensing," in *Proc. IEEE 11th Int. Conf. Ubiquitous Intell. Comput., IEEE 11th Int. Conf. Autonomic Trusted Comput., IEEE 14th Int. Conf. Scalable Comput. Commun. Associated Workshops*, Dec. 2014, pp. 179–186.
[6] G. Ritzer and N. Jurgenson, "Production, consumption, prosumption: The nature of capitalism in the age of the digital 'prosumer,'" *J. Consum. Culture*, vol. 10, no. 1, pp. 13–36, Mar. 2010.
[7] D. Yang, G. Xue, X. Fang, and J. Tang, "Crowdsourcing to smartphones: Incentive mechanism design for mobile phone sensing," in *Proc. 18th Annu. Int. Conf. Mobile Comput. Netw. (Mobicom)*, 2012, pp. 173–184.
[8] B. Kantarci and H. T. Mouftah, "Trustworthy sensing for public safety in cloud-centric Internet of Things," *IEEE Internet Things J.*, vol. 1, no. 4, pp. 360–368, Aug. 2014.
[9] M. Pouryazdan, B. Kantarci, T. Soyata, L. Foschini, and H. Song, "Quantifying user reputation scores, data trustworthiness, and user incentives in mobile crowd-sensing," *IEEE Access*, vol. 5, pp. 1382–1397, 2017.
[10] M. E. Gendy, A. Al-Kabbany, and E. F. Badran, "Maximizing clearance rate by penalizing redundant task assignment in mobile crowdsensing auctions," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, May 2020, pp. 1–7.
[11] M. E. Gendy, A. Al-Kabbany, and E. F. Badran, "Maximizing clearance rate of reputation-aware auctions in mobile crowdsensing," in *Proc. 16th IEEE Annu. Consum. Commun. Netw. Conf. (CCNC)*, Jan. 2019, pp. 1–2.
[12] M. E. Gendy, A. Al-Kabbany, and E. F. Badran, "Maximizing clearance rate of reputation-aware auctions in mobile crowdsensing," 2018, *arXiv:1810.05774*. [Online]. Available: https://arxiv.org/abs/1810.05774
[13] S. Reddy, D. Estrin, and M. Srivastava, "Recruitment framework for participatory sensing data collections," in *Proc. 8th Int. Conf. Pervas. Comput.*, May 2010, pp. 138–155.
[14] G. Cohn, S. Gupta, T.-J. Lee, D. Morris, J. R. Smith, M. S. Reynolds, D. S. Tan, and S. N. Patel, "An ultra-low-power human body motion sensor using static electric field sensing," in *Proc. ACM Conf. Ubiquitous Comput. (UbiComp)*, 2012, pp. 99–102.
[15] C. Gomez, J. Oller, and J. Paradells, "Overview and evaluation of Bluetooth low energy: An emerging low-power wireless technology," *Sensors*, vol. 12, no. 9, pp. 11734–11753, 2012.
[16] V. Agarwal, N. Banerjee, D. Chakraborty, and S. Mittal, "USense—A smartphone middleware for community sensing," in *Proc. IEEE 14th Int. Conf. Mobile Data Manage. (MDM)*, Milan, Italy, vol. 1, Jun. 2013, pp. 56–65.
[17] Q. Zhao, Y. Zhu, H. Zhu, J. Cao, G. Xue, and B. Li, "Fair energy-efficient sensing task allocation in participatory sensing with smartphones," in *Proc. IEEE Conf. Comput. Commun. (INFOCOM)*, Apr. 2014, pp. 1366–1374.
[18] J. Fakcharoenphol, B. Laekhanukit, and D. Nanongkai, "Faster algorithms for semi-matching problems," in *Automata, Languages and Programming*. Berlin, Germany: Springer, 2010, pp. 176–187.
[19] (Feb. 17, 2020). *Cisco Annual Internet Report (2018–2023) FAQ*. [Online]. Available: https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/q-and-a-c67-482177.html
[20] N. P. Owoh and M. M. Singh, "Security analysis of mobile crowd sensing applications," *Appl. Comput. Informat.*, vol. 10, Jul. 2020. [Online]. Available: https://www.emerald.com/insight/content/doi/10.1016/j.aci.2018.10.002/full/pdf?title=security-analysis-of-mobile-crowd-sensing-applications
[21] M. Langheinrich, "Privacy by design—Principles of privacy-aware ubiquitous systems," in *Ubicomp 2001: Ubiquitous Computing* (Lecture Notes in Computer Science), vol. 2201, G. D. Abowd, B. Brumitt, and S. Shafer, Eds. Berlin, Germany: Springer, 2001, doi: 10.1007/3-540-45427-6_23.

[22] S. Reddy, D. Estrin, M. Hansen, and M. Srivastava, "Examining micro-payments for participatory sensing data collections," in *Proc. 12th ACM Int. Conf. Ubiquitous Comput.*, Sep. 2010, pp. 33–36.
[23] M. Musthag, A. Raij, D. Ganesan, S. Kumar, and S. Shiffman, "Exploring micro-incentive strategies for participant compensation in high-burden studies," in *Proc. 13th Int. Conf. Ubiquitous Comput. (UbiComp)*, 2011, pp. 435–444.
[24] J.-S. Lee and B. Hoh, "Dynamic pricing incentive for participatory sensing," *Pervas. Mobile Comput.*, vol. 6, no. 6, pp. 693–708, Dec. 2010.
[25] C. A. Le Dantec, J. E. Christensen, M. Bailey, R. G. Farrell, J. B. Ellis, C. M. Danis, W. A. Kellogg, and W. K. Edwards, "A tale of two publics: Democratizing design at the margins," in *Proc. 8th ACM Conf. Designing Interact. Syst. (DIS)*, 2010, pp. 11–20.
[26] C. A. Le Dantec, R. G. Farrell, J. E. Christensen, M. Bailey, J. B. Ellis, W. A. Kellogg, and W. K. Edwards, "Publics in practice: Ubiquitous computing at a shelter for homeless mothers," in *Proc. Annu. Conf. Hum. Factors Comput. Syst. (CHI)*, 2011, pp. 1687–1696.
[27] E. Kanjo, "NoiseSPY: A real-time mobile phone platform for urban noise monitoring and mapping," *Mobile Netw. Appl.*, vol. 15, no. 4, pp. 562–574, Aug. 2010.
[28] T. W. Malone, R. Laubacher, and C. Dellarocas, "The collective intelligence genome," *MIT Sloan Manage. Rev.*, vol. 51, no. 3, p. 21, 2010.
[29] K. Han, E. A. Graham, D. Vassallo, and D. Estrin, "Enhancing motivation in a mobile participatory sensing project through gaming," in *Proc. IEEE 3rd Int. Conf. Privacy, Secur., Risk Trust IEEE 3rd Int. Conf. Social Comput.*, Oct. 2011, pp. 1443–1448.
[30] E. Miluzzo, N. D. Lane, S. B. Eisenman, and A. T. Campbell, "Cenceme–injecting sensing presence into social networking applications," in *Proc. Eur. Conf. Smart Sens. Context*. Berlin, Germany: Springer, 2007, pp. 1–28.
[31] N. Maisonneuve, M. Stevens, M. E. Niessen, and L. Steels, "Noisetube: Measuring and mapping noise pollution with mobile phones," in *Information Technologies in Environmental Engineering*. Berlin, Germany: Springer, 2009, pp. 215–228.
[32] Y. F. Dong, S. Kanhere, C. T. Chou, and N. Bulusu, "Automatic collection of fuel prices from a network of mobile cameras," in *Proc. Int. Conf. Distrib. Comput. Sensor Syst.* Berlin, Germany: Springer, 2008, pp. 140–156.
[33] N. Bulusu, C. T. Chou, S. Kanhere, Y. Dong, S. Sehgal, D. Sullivan, and L. Blazeski, "Participatory sensing in commerce: Using mobile camera phones to track market price dispersion," in *Proc. Int. Workshop Urban, Community, Social Appl. Netw. Sens. Syst. (UrbanSense)*, 2008, pp. 6–10.
[34] B. L. Sullivan, C. L. Wood, M. J. Iliff, R. E. Bonney, D. Fink, and S. Kelling, "EBird: A citizen-based bird observation network in the biological sciences," *Biol. Conservation*, vol. 142, no. 10, pp. 2282–2292, Oct. 2009.
[35] *Floracaching*. Accessed: Sep. 15, 2020. [Online]. Available: https://research.cens.ucla.edu/education/highschoolscholars/2011/docs/Floracaching.pdf
[36] M. Yilmaz and R. V. O'Connor, "A market based approach for resolving resource constrained task allocation problems in a software development process," in *Proc. Eur. Conf. Softw. Process Improvement*. Berlin, Germany: Springer, 2012, pp. 25–36.
[37] H. Xiong, D. Zhang, L. Wang, and H. Chaouchi, "EMC3: Energy-efficient data transfer in mobile crowdsensing under full coverage constraint," *IEEE Trans. Mobile Comput.*, vol. 14, no. 7, pp. 1355–1368, Jul. 2015.
[38] A. Capponi, C. Fiandrino, D. Kliazovich, P. Bouvry, and S. Giordano, "A cost-effective distributed framework for data collection in cloud-based mobile crowd sensing architectures," *IEEE Trans. Sustain. Comput.*, vol. 2, no. 1, pp. 3–16, Jan. 2017.
[39] A. Mehrotra, V. Pejovic, and M. Musolesi, "SenSocial: A middleware for integrating online social networks and mobile sensing data streams," in *Proc. 15th Int. Middleware Conf.*, Bordeaux, France, Dec. 2014, pp. 205–216.
[40] Y. Yao, L. T. Yang, and N. N. Xiong, "Anonymity-based privacy-preserving data reporting for participatory sensing," *IEEE Internet Things J.*, vol. 2, no. 5, pp. 381–390, Oct. 2015.
[41] H. Xiong, D. Zhang, L. Wang, J. P. Gibson, and J. Zhu, "EEMC: Enabling energy-efficient mobile crowdsensing with anonymous participants," *ACM Trans. Intell. Syst. Technol.*, vol. 6, no. 3, pp. 1–26, May 2015.
[42] D. P. Anderson, J. Cobb, E. Korpela, M. Lebofsky, and D. Werthimer, "SETI@home: An experiment in public-resource computing," *Commun. ACM*, vol. 45, no. 11, pp. 56–61, Nov. 2002.
[43] M. Böhmer, B. Hecht, J. Schöning, A. Krüger, and G. Bauer, "Falling asleep with angry birds, Facebook and kindle: A large scale study on mobile application usage," in *Proc. MobileHCI*, 2011, pp. 47–56.
[44] L. Carata, A. Rice, and A. Hopper, "IPAPI: Designing an improved provenance API," in *Proc. TaPP*, 2013, pp. 1–4.
[45] N. Eagle and A. Pentland, "Reality mining: Sensing complex social systems," *Pers. Ubiquitous Comput.*, vol. 10, no. 4, pp. 255–268, May 2006.
[46] A. Girardello and F. Michahelles, "AppAware: Which mobile applications are hot?" in *Proc. MobileHCI*, 2010, pp. 431–434.
[47] F. Khatib, S. Cooper, M. D. Tyka, K. Xu, I. Makedon, Z. Popovic, D. Baker, and F. Players, "Algorithm discovery by protein folding game players," *Proc. Nat. Acad. Sci. USA*, vol. 108, no. 47, pp. 18949–18953, Nov. 2011.
[48] M. Langheinrich, "Privacy by design—Principles of privacy-aware ubiquitous systems," in *Proc. UbiComp*, 2001, pp. 273–291.
[49] J. K. Laurila, J. Blom, O. Dousse, D. Gatica-Perez, O. Bornet, and M. Miettinen, "The mobile data challenge: Big data for mobile computing research," in *Proc. PerCom*, 2012.
[50] E. Oliver, "The challenges in large-scale smartphone user studies," in *Proc. HotPlanet*, 2010, pp. 1–5.
[51] M. Yisroel, S. Asaf, R. Lior, S. Bracha, and E. I. Yuval, "SherLock vs moriarty: A smartphone dataset for cybersecurity research," in *Proc. AISec*, vol. 16. New York, NY, USA: Association for Computing Machinery, 2016, p. 112.
[52] D. T. Wagner, A. Rice, and A. R. Beresford, "Device analyzer: Large-scale mobile data collection," *ACM SIGMETRICS Perform. Eval. Rev.*, vol. 41, no. 4, pp. 53–56, Apr. 2014.
[53] C. Shepard, A. Rahmati, C. Tossell, L. Zhong, and P. Kortum, "LiveLab: Measuring wireless networks and smartphone users in the field," *ACM SIGMETRICS Perform. Eval. Rev.*, vol. 38, no. 3, pp. 15–20, Jan. 2011.
[54] N. Kiukkonen, J. Blom, O. Dousse, D. GaticaPerez, and J. Laurila, "Towards rich mobile phone datasets: Lausanne data collection campaign," in *Proc. ICPS*, Berlin, Germany, 2010.
[55] A. Madan, M. Cebrian, S. Moturu, K. Farrahi, and A. Pentland, "Sensing the 'health state' of a community," *IEEE Pervasive Comput.*, vol. 11, no. 4, pp. 36–45, Oct./Dec. 2012.
[56] N. Eagle and A. Pentland, "Reality mining: Sensing complex social systems," *Pers. Ubiquitous Comput.*, vol. 10, pp. 255–268, Nov. 2006.
[57] S. Masum, Y. Liu, and J. Chiverton, "Multi-step time series forecasting of electric load using machine learning models," in *Proc. Int. Conf. Artif. Intell. Soft Comput.* Cham, Switzerland: Springer, May 2018, pp. 148–159.
[58] I. Aizenberg, L. Sheremetov, L. Villa-Vargas, and J. Martinez-Muñoz, "Multilayer neural network with multi-valued neurons in time series forecasting of oil production," *Neurocomputing*, vol. 175, pp. 980–989, 2016. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0925231215016008, doi: 10.1016/j.neucom.2015.06.092.
[59] L. D. Marino, K. Amarasinghe, and M. Manic, "Building energy load forecasting using deep neural networks," in *Proc. 42nd Annu. Conf. IEEE Ind. Electron. Soc.*, Oct. 2016, pp. 7046–7051.
[60] A. Brunnert, K. Wischer, and H. Krcmar, "Using architecture-level performance models as resource profiles for enterprise applications," in *Proc. 10th ACM SIGSOFT Int. Conf. Qual. Softw. Archit. (QoSA)*, Lille, France, 2014, pp. 53–62.
[61] C.-J. Ho and J. W. Vaughan, "Online task assignment in crowdsourcing markets," in *Proc. AAAI*, 2012, pp. 45–51.
[62] R. Adhikari and R. K. Agrawal, "An introductory study on time series modeling and forecasting," 2013, *arXiv:1302.6613*. [Online]. Available: http://arxiv.org/abs/1302.6613
[63] A. Yadav, C. K. Jha, and A. Sharan, "Optimizing LSTM for time series prediction in Indian stock market," *Procedia Comput. Sci.*, vol. 167, pp. 2091–2100, Jan. 2020.
[64] S. A. Abdulkarim, "Time series prediction with simple recurrent neural networks," *Bayero J. Pure Appl. Sci.*, vol. 9, no. 1, pp. 19–24, 2016.

MAGGIE EZZAT GABER GENDY (Student Member, IEEE) received the B.Sc. and M.S. degrees in electronics and communications engineering from the Arab Academy for Science, Technology & Maritime Transport (AASTMT), Alexandria, Egypt, in 2016 and 2019, respectively. Since 2017, she has been an Assistant Lecturer with the Electronics and Communications Engineering, AASTMT. Her research interests include mobile crowdsensing, optimization approaches, and machine learning.

AHMAD AL-KABBANY (Member, IEEE) received the B.Sc. degree in electronics and communications engineering from the Arab Academy for Science and Technology, Alexandria, Egypt, in 2005, and the Ph.D. degree in electrical and computer engineering from the University of Ottawa, Ottawa, ON, Canada, in 2016.

From 2006 to 2010, he was a Teaching and Research Assistant with the Department of Electronics and Communications Engineering, Arab Academy for Science and Technology. In 2010, he was a Graduate Research Intern with the University of Ulster, U.K. From 2010 to 2016, he was a Graduate Research Assistant pursuing his Ph.D. degree, and then a Postdoctoral Fellow with VIVA Lab, School of Electrical Engineering and Computer Science, University of Ottawa. Since 2016, he has been an Assistant Professor of electronics and communications engineering, and a Founding Member of the Intelligent Systems Laboratory, Arab Academy for Science and Technology. During his Ph.D. studies, he got two research awards from the University of Ottawa, and one of the algorithms that he co-proposed was ranked, at the time, among the best eight techniques worldwide according to a standard benchmark in image alpha matting. He is a Reviewer of IEEE Signal Processing Letters and IEEE Transactions on Vehicular Technology. His research interests include the applications of graph optimization and machine learning in communication networks and N-D signal processing, synthesis, analysis, and communication. He is also the Founder and the CEO of VRapeutic–a research-based software startup specializing in the development of therapeutic AI-enabled virtual reality solutions, with a current focus on learning difficulties and developmental disorders. He has led VRapeutic to several achievements, including the 1st place at the Global Impact Challenge in 2019, organized by Singularity University in California, USA. VRapeutic has joined the UNICEF Innovation Fund portofolio of companies in June 2020.

EHAB FAROUK BADRAN (Senior Member, IEEE) received the B.Sc. degree (Hons.) and the M.Sc. degree in electrical engineering from Assiut University, Assiut, Egypt, in May 1995 and March 1998, respectively, and the M.Sc. and Ph.D. degrees in electrical engineering from Louisiana State University (LSU), Baton Rouge, LA, USA, in May 2001 and May 2002, respectively. From 1995 to 1998, he was an Instructor with the Department of Electrical Engineering, Assiut University, where in May 1998, he was promoted to an Assistant Lecturer. From January 2000 to May 2002, he was a Teaching and Research Assistant with the Department of Electrical and Computer Engineering, Louisiana State University, during his Ph.D. studies. From September 2002 to August 2003, he was an Assistant Professor with the Department of Electrical Engineering, Assiut University. From September 2003 to May 2007, he worked as an Assistant Professor with the Department of Electronics and Communication Engineering, Arab Academy for Science, Technology & Maritime Transport, Alexandria, Egypt. From June 2007 to May 2011, he was an Associate Professor with the Department of Electronics and Communication Engineering. In June 2011, he was promoted to be a Professor. His research and teaching interests include wireless communications, signal processing, MIMO systems, image processing, and communication networks.