Received 29 June 2024; revised 1 August 2024; accepted 17 August 2024. Date of publication 26 August 2024; date of current version 16 September 2024.

Digital Object Identifier 10.1109/OJCOMS.2024.3449691

Novel Distribution-Aware and Learning-Based Dynamic Scheme for Efficient User Incentivization in Edge Sensing Systems

OMAR NASERALLAH [1] (Student Member, IEEE), SHERIF B. AZMY [1] (Graduate Student Member, IEEE), NIZAR ZORBA [2] (Senior Member, IEEE), AND HOSSAM S. HASSANEIN [3] (Fellow, IEEE)

[1] Department of Electrical and Computer Engineering, Queen's University, Kingston, ON K7L 3N6, Canada

[2] College of Engineering, Qatar University, Doha, Qatar

[3] School of Computing, Queen's University, Kingston, ON K7L 3N6, Canada

CORRESPONDING AUTHOR: N. ZORBA (e-mail: nizarz@qu.edu.qa)

This work was supported by grants from the Natural Sciences and Engineering Research Council of Canada (NSERC) under grant number ALLRP 549919-20, Qatar University grant IRCC-2024-494 and a grant from Distributive, ltd. The Open Access funding is provided by Qatar National Library.

An earlier version of this paper was presented in part at IEEE Global Communications Conference, Rio de Janeiro, Brazil, 2022 [22] [DOI: 10.1109/GLOBECOM48099.2022.10000854].

ABSTRACT Edge sensing (ES) systems employ users' owned smart devices with built-in sensors to gather data from users' surrounding environments and use their processors to carry out edge computing tasks. Therefore, ES is emerging as a potential solution for remote sensing challenges. Additionally, ES systems are recognized for their favorable characteristics, including efficient time and cost management, scalability, and the ability to gather real-time data. To improve the performance of ES systems, enormous efforts have been made to enhance the quality of data (QoD) and the systems' spatiotemporal coverage. Moreover, the research community has focused on developing better incentive schemes, as user incentivization is essential for enhancing system performance. In this study, we assess the impact of users' mobility and availability on the spatiotemporal coverage and QoD of ES systems, taking into account the heterogeneity of users. We propose a distribution-aware and learning-based dynamic incentive scheme. Specifically, we consider the randomness of users' mobility and velocity using a 2-dimensional random waypoint (RWP) model and support the learning-based incentive scheme with a long short-term memory (LSTM) model. The LSTM model utilizes the users' historical data to predict their availability to perform the sensing tasks. The learning-based incentive scheme is further used to enhance system performance and effectively manage the trade-off between quality and cost, by recruiting users based on the required quality and cost constraints, to meet the minimum quality requirement within a constrained incentivization budget.

INDEX TERMS Edge sensing, quality of data, mobility, incentive, prediction.

I. INTRODUCTION

THE PROLIFERATION of smartphones has been the driving force behind the surge of mobile crowd sensing (MCS) [1]. However, MCS is confined to data collecting and sharing functionalities. With the continuous advancements in smartphones' computational resources, there is a growing demand for solutions that can fully exploit the progress made in smartphone development. Edge sensing (ES) represents a paradigm shift that goes beyond data collecting and sharing, encompassing processing and computations directly on edge devices, enabling a diverse range of applications, including smart cities and environmental and traffic monitoring [2]. Concerns regarding task assignment, cost reduction, and quality of data (QoD) arise due to the user-centric architecture of ES systems. Achieving higher data quality and broader coverage often requires recruiting more users, which introduces a cost-quality trade-off. Implementing sparse edge sensing by dividing the area of interest (AoI) into spatiotemporal cells [3] helps manage this trade-off by optimizing task allocation and balancing data quality, spatiotemporal coverage, and budget limitations.

To ensure the effectiveness of ES systems, it is essential to tackle the challenges associated with expanding spatiotemporal coverage and ensuring the QoD. Users' varying competence levels and experiences pose a significant concern in ensuring control and maintaining data quality within the ES system [4]. Users' heterogeneity encompasses both users' skills and the quality of their smartphones. Therefore, addressing the challenges associated with users' heterogeneity is essential for effective control and achieving high-quality sensed data in the ES system. Moreover, due to the users-centric architecture, the data shared by users can be prone to errors and faults, posing significant implications for QoD [5]. In this regard, ES systems attempt to complete tasks with an appropriate QoD and coverage to ensure that no blind spots exist within the system, thereby enhancing its overall sensing effectiveness and reliability.

ES leverages on the mobility of users to achieve extensive coverage [6], distinguishing it from other remote sensing approaches like wireless sensor networks (WSN). Instead of relying on stationary sensors, ES utilizes the users' owned smart devices to perform sensing tasks, resulting in a more accurate representation of the AoI. The inherent mobility of users also makes ES highly adaptable for various applications. However, users' mobility introduces uncertainty that affects the QoD, allocation of tasks, and system functionality [7]. Moreover, users' movement incurs costs related to resource consumption, including energy and data allowances, which must be accounted to ensure optimal system performance.

Users or participants carry out the sensing tasks serve as the centerpiece of any ES system; hence, incentivizing them to complete the sensing tasks is crucial. Since ES systems must operate within a limited budget, incentive schemes must be designed for each system to ensure optimal performance [8]. Monetary and non-financial incentives, including services offered to users, are the two basic types of incentives. These types can be alternatively divided into fixed and dynamic incentives. Fixed incentives provide users with a consistent reward or benefit regardless of changing circumstances, while dynamic incentives are more flexible and more efficient as they adjust to QoD, participation time, or system requirements, among others.

As aforementioned, users' mobility, incentives, and QoD can be considered as the three main challenges ES systems face. We believe mobility prediction can be beneficial in overcoming these challenges. However, knowing users' locations is not sufficient in realistic scenarios, as the users might be present in the vicinity/cell but not available to do the tasks. The advancement in machine learning techniques enables the ES system to predict the users' willingness and availability to improve the system performance and optimize the incentive schemes [9]. Time series forecasting is a suitable option due to the temporal component. It has many real-world uses, such as weather forecasting, economic forecasting, and social studies forecasting, among others [10]. Long short-term memory (LSTM) networks have been created in response to the growing need to perform context-based learning and extract temporal correlations from sequential data [11]. As it can handle non-linear issues and has a time memory, LSTM-based time series analysis methods can produce better prediction outcomes [12].

To mitigate the potential negative impact of users' mobility and availability on QoD, this paper

- utilizes random waypoint (RWP) model to investigates the influence of users' mobility on the QoD in ES systems,
- introduces a novel distribution-aware and learning-based dynamic incentive scheme to minimize the impact of users' mobility, and availability on QoD,
- employs the LSTM model prediction to enhance task allocation, aiming to resolve the trade-off between cost and QoD.

The remainder of the paper is arranged as follows. Section II includes a brief literature review for the reader to realize the current state of the art. The system model and the generated data set are addressed in Section III. Then Section IV discusses the impact of the mobility model on QoD; followed by Section V which proposes the incentive scheme. Section VI examines simulation and performance results, while Section VII concludes the paper.

II. RELATED WORKS

In relation to the QoD, a cross-validation method is introduced in [13], where participants' sensor data undergo verification by a validating crowd to ensure an accurate representation of the user's surrounding environment. Besides validation, accuracy is important to manage the quality of ES data. Reference [14] addresses the challenge of calculating similarity for unequal long-term sequences by leveraging dynamic time warping (DTW) technology. Through clustering and comparing the data in the database, authors in [14] successfully identify abnormal instances, leading to improvements in the accuracy database. To measure the accuracy, a metric is introduced in [8] which, in cases with small data, utilizes the disparity in centrality estimation to determine the source quality.

Users' mobility and coverage are very important topics discussed in the literature, where a vehicular edge sensing system is proposed in [15], as the scheme aims to align the sensing distribution of the collected data with the planned target distribution. To accommodate different desired target distributions, the incentive problem is formulated as a unique variant of the non-linear multiple-choice knapsack problem., utilizing the objective function based on the difference between the distribution of sensed data and the targeted distribution. To optimize budget utilization, a customized incentive approach is introduced, combining monetary rewards with an expected task (ride) required at the destination. Mobility characterization has a major impact on the system, and the research presented in [16] aims to explore various mobility characteristics of the real-world ES data set ParticipAct, and examines how these characteristics can be leveraged to plan an effective ES data collection campaign. The authors analyze the mobility traces gathered from ParticipAct and discuss the advantages of utilizing this information in an ES task.

Based on the mobility characteristics, a novel coverage metric for ES systems is presented in [17] to evaluate the AoI from both global and local perspectives, enabling administrators to identify regions with deficient data quality and suggest compensation strategies. The metric's performance is validated through simulations, demonstrating its effectiveness in assessing coverage quality for MCS systems.

Developing an appropriate incentive mechanism is investigated in [18] to incorporate data quality considerations while taking into account the influence of social factors. Other incentive mechanisms have been discussed in the literature, where a budget-constrained incentive mechanism is introduced in [19], which utilizes the user's prior data to determine their preferences. An approximation algorithm is then developed to select the contributor of data without violating the budget constraints.

Different artificial intelligence (AI) techniques have been proposed in literature to deal with ES systems, where the work presented in [20] utilizes deep learning to address the challenge of managing data quality in ES systems, specifically focusing on detecting malicious data introduced by adversaries or faulty components. An upgraded multi-task allocation method is presented in [1] that incorporates mobility prediction. This approach differs from other methods primarily in how it leverages workers' past movements, utilizing a fuzzy logic system for improved accuracy in predicting mobility. Additionally, it adopts a global heuristic search algorithm aimed at boosting the task completion rate. This enhancement considers the spatial and temporal traits of both workers and tasks, based on the anticipated mobility. Alternative AI schemes tackle the problem of quality-aware user recruitment and formulate it as an optimization problem [21]. The authors employ federated learning to analyze the relationship between data and context information, enabling them to forecast the quality of sensed data from different users, where an upgraded lightweight neural network model is utilized on mobile terminals.

To the best of our knowledge, no existing research specifically investigates the influence of users' mobility and availability on the QoD in heterogeneous ES systems. Existing literature discusses concerns related to QoD, user mobility, and user availability separately. However, it is important to recognize that user mobility and availability can directly impact the number of readings within a given cell, subsequently affecting both the QoD and the overall quality metric satisfaction (QMS). In this article, we aim to address this research gap by examining the impact of user mobility and availability on the quality metric previously introduced in [22]. To mitigate these challenges, we propose a distribution-aware and learning-based dynamic incentive scheme.

III. SYSTEM MODEL

The proposed system takes into account the heterogeneity of available users by categorizing them into classes according to the QoD they provide and the associated incentive costs for each class. This classification allows the system to optimize the incentive scheme and user recruitment based on the cost-quality trade-off, aiming to meet the minimum required quality $Q^{\mathrm{t}}_{\mathrm{min}}$ of each task t at the lowest possible cost. The ES system being considered consists of a total of N_{total} users distributed across the AoI. However, the effectiveness of users in completing the sensing task can be influenced by their mobility [23]. To address this, we propose a corrected user count N_m that considers their movement speeds. N_m is considered to be always equal to or less than N_{total}, as it excludes users with higher velocities. This is crucial because users with high velocities may face difficulties in ensuring the successful completion of the sensing task. Moreover, the existence of users within the system does not necessarily imply their availability or willingness to perform the sensing task, we denote the set of available users as $N^{\mathrm{k}}_{\mathrm{n}} = \{[u^1_1, u^1_2, \ldots, u^1_{\mathrm{n}}], [u^2_1, u^2_2, \ldots, u^2_{\mathrm{n}}], \ldots, [u^{\mathrm{k}}_1, u^{\mathrm{k}}_2, \ldots, u^{\mathrm{k}}_{\mathrm{n}}]\}$, which shows the count of users n in different classes k who are available to do sensing tasks, and the summation over all n and k is always less or equal to N_m. The set of sensing tasks is defined as $T = \{t_1, t_2, t_3, \ldots, t_t\}$, where each task $t \in T$ has a different sensing duration based on its nature. Available users are expected to complete the entire sensing duration of the sensing before sharing the collected data. Furthermore, the minimum required quality $Q^{\mathrm{t}}_{\mathrm{min}}$ varies for each task, reflecting the varying levels of importance among different sensing tasks.

The AoI is $X_{\mathrm{AoI}} \times Y_{\mathrm{AoI}}$ dimensions, and it is divided into L subcells to study the distribution of users over the whole AoI. Users are considered to follow the widely known 2-dimensional RWP model to assess the mobility effects [23].

To design a realistic model, we consider the heterogeneity of users participating in the sensing, as they have varying experiences and use different devices with diverse capabilities. As mentioned earlier, users are categorized into k classes based on the QoD they provide, leading to distinct incentivization costs. Furthermore, we recognize that users' presence within the AoI does not guarantee their willingness to participate. To tackle this challenge, we have integrated a learning model to predict the availability patterns of users in different classes. Since the availability pattern of users can be complex, LSTM was selected over other prediction models such as gated recurrent unit (GRU) and transformers due to its ability to capture long-term dependencies in sequential data. LSTM's architecture learns temporal correlations, which is crucial for accurately predicting user availability. Although other models can be employed, LSTM is picked to showcase that the incorporation of a learning model can significantly enhance the incentivization schemes.

We would like to emphasize that availability is not only due to location, but rather due to many other factors other than location (which, by the way, follows the 2D RWP

TABLE 1. Example of users' availability pattern.

Time	12:00	12:10	12:20	12:30	...
Class 1	2	4	3	3	...
Class 2	4	3	5	4	...
Class 3	10	8	8	9	...

model in [23]), such as battery power, available resources, device's computational and communicational capabilities, active user participation in the ES campaign, as well as a multitude of other factors (e.g., their satisfaction with their payment, whether they'll accept the task or not, and so on). As a consequence of this mixture of numerous factors contributing to the availability, we deemed a normal distribution as sufficient to capture the inherent randomness due to the central limit theorem. The ParticipAct dataset [16] is the closest one to our work, however it falls short due to the different nature of the task at hand (not just sensing, but also computation, which increases the dimension).

A data set that emulates users and their availability pattern is created to train the LSTM model. Since the users are assumed to be heterogeneous, the data set shows the available number of users from each class. The data set provides the number of available users in each location every 10 minutes for 8 hours daily, while more granularity in the time domain can be achieved if needed. The focus of the data set is on the central subcell to study the temporal coverage of the system, as analyzing the central subcell would account for interactions from all adjacent subcells. In our system, we used a 70/30 data split for training and testing the LSTM model, allowing for effective learning and robust evaluation.

Table 1 provides an example of the dataset used in the study. The availability patterns of different user classes ($k = 3$) are used to train the LSTM model. Utilizing the LSTM model prediction for the recruitment of users will enable the system to ensure the QMS. The synthetic data generation process was undertaken as a proof of concept due to the absence, to the best of our knowledge, of publicly available datasets that capture heterogeneous user availability patterns. The data generation process involved using normal distributions to simulate realistic user behavior. Specifically, we modeled user activity during peak hours with higher mean and standard deviation values, while using lower mean and standard deviation values for off-peak hours. This approach captures natural fluctuations in user patterns across different times of the day. This approach introduces controlled noise into the system, reflecting natural fluctuations in user activity. To evaluate the model in real-life situations where the user's availability pattern varies, a degree of randomness is introduced. This stochastic behavior is induced by employing a uniform distribution to select a subset of users for pattern alteration.

In order to evaluate the LSTM model performance under ideal conditions, we run the system without injected randomness. Fig. 1 shows the LSTM model performance. It is clear that the model is able to learn the users' availability pattern and succeeds in predicting the number of available users from each class for doing tasks at certain hours of the day. The test root mean square error (RMSE) is equal to 0.097, the mean square error (MSE) equals 0.009, and the mean absolute error (MAE) is equal to 0.069 in the zero randomness scenario. The LSTM model performs well under no randomness conditions, as the bottom line for performance in realistic systems. Its behavior by considering randomness in realistic conditions will be later analyzed in the simulations section.

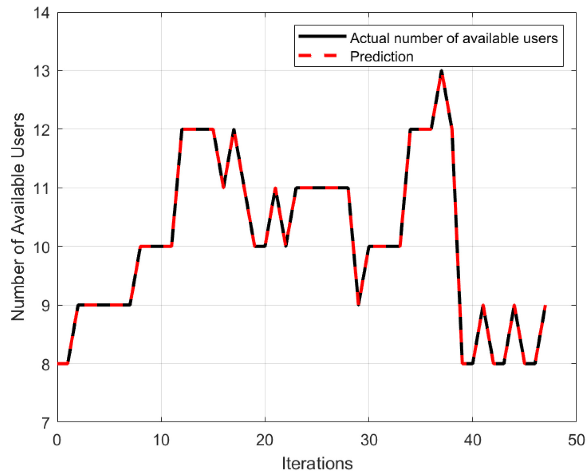

FIGURE 1. LSTM model performance.

IV. MOBILITY MODEL EFFECT ON QOD

For a user executing RWP movement along a line $[0, X_{\mathrm{AoI}}]$, their location (x) can be depicted by the probability density function (PDF) [24]

$$f_X(x) = -\frac{6}{X_{\mathrm{AoI}}^3}x^2 + \frac{6}{X_{\mathrm{AoI}}^2}x \tag{1}$$

The 1D PDF needs to be transformed into 2D to accommodate square-shaped subcells. This transformation considers two separate 1D movements: one along the x-axis and another along the y-axis, as Eqn. (2) shows.

$$p = \int_{y^\circ}^{y^\circ+\triangledown} \int_{x^\circ}^{x^\circ+\triangledown} \left(\frac{-6x^2}{{X_{\mathrm{AoI}}}^3} + \frac{6x}{{X_{\mathrm{AoI}}}^2}\right) \times \left(\frac{-6y^2}{{Y_{\mathrm{AoI}}}^3} + \frac{6y}{{Y_{\mathrm{AoI}}}^2}\right) dxdy \tag{2}$$

By solving Eqn. (2), the probability p that a user is present in the subcell is formulated, with y°, $y^\circ + \triangledown$, x° and $x^\circ + \triangledown$ representing the borders of the subcells. The probability is obtained in Eqn. (3).

$$p = \left(\frac{-2\big((y^\circ+\triangledown)^3 - (y^\circ)^3\big)}{{Y_{\mathrm{AoI}}}^3} + \frac{3\big((y^\circ+\triangledown)^2 - (y^\circ)^2\big)}{{Y_{\mathrm{AoI}}}^2}\right) \left(\frac{-2\big((x^\circ+\triangledown)^3 - (x^\circ)^3\big)}{{X_{\mathrm{AoI}}}^3} + \frac{3\big((x^\circ+\triangledown)^2 - (x^\circ)^2\big)}{X_{\mathrm{AoI}}^2}\right) \tag{3}$$

To ensure the system's reliability, the study needs to address the issue of users' mobility. Task completion becomes constrained by both space and time when users are in motion. Therefore, it is necessary to employ an appropriate approach that accounts for the impact of mobility on Eqn. (3). Since users may not always be able to complete tasks due to their mobility, it is important to quantify this impact. N_m provides a more precise estimation of users deemed suitable for allocation to the sensing tasks, as it takes into account the impact of mobility on users' spatial positioning.

Assuming that within a subcell, the distance a user travels is uniformly distributed, ranging from $D_{\min}$ to $D_{\max}$, where $D_{\max}$ represents the maximum possible travel distance within a subcell. With users moving at a fixed speed, the time spent in the subcell is inversely proportional to this speed. Based on the application of the ES system, a time threshold to guarantee task completion should be specified. For instance, compared to environmental monitoring, emergency applications have strict time constraints on sensing. The velocity threshold can be created from the time threshold.

To improve the performance of the system by giving tasks to users who will take the appropriate time to finish them, it is, therefore, necessary to take into account a velocity threshold for user velocity. The quality metric's performance will be enhanced by the consideration of the corrected number of users, as it will provide a more accurate user number after removing users with higher velocities and lower task completion rates. The corrected number of users N_m is then formulated as

$$N_{\mathrm{m}} = N_{\mathrm{total}}\left(1 - \frac{V_{\max} - V_{\mathrm{th}}}{V_{\max} - V_{\min}}\right) \tag{4}$$

where $V_{\max}$ and $V_{\min}$ are the maximum and minimum velocities of users, respectively, and V_{th} is the velocity threshold. Notice that driving the system to extreme values of $V_{\mathrm{th}} = V_{\max}$ or $V_{\mathrm{th}} = V_{\min}$ will result in logical results.

To further enhance the estimation of the number of users, the LSTM model is utilized to predict the available number of users $N_{\mathrm{n}}^{\mathrm{k}}$ based on their historical data and their availability pattern. The prediction of $N_{\mathrm{n}}^{\mathrm{k}}$ will enable the system to have an estimation of the achievable quality and to maintain the quality requirement satisfied.

Among the typical time-series forecasting techniques, LSTM was chosen due its capability to understand long-term sequences of observations. Unlike traditional recurrent neural networks (RNNs), LSTM is an advanced RNN architecture designed to more accurately model temporal sequences and their long-range dependencies [25]. The LSTM model will help cut expenses and save unneeded incentives by predicting $N_{\mathrm{n}}^{\mathrm{k}}$ to complete the sensing tasks at a certain time. The LSTM model uses the users' historical data to predict their availability. Then the system uses the prediction data to check the satisfaction of the quality metric; this will enable the system to optimize the incentivization process as it will only incentivize the users to do the sensing in specific spatiotemporal subcells based on the need.

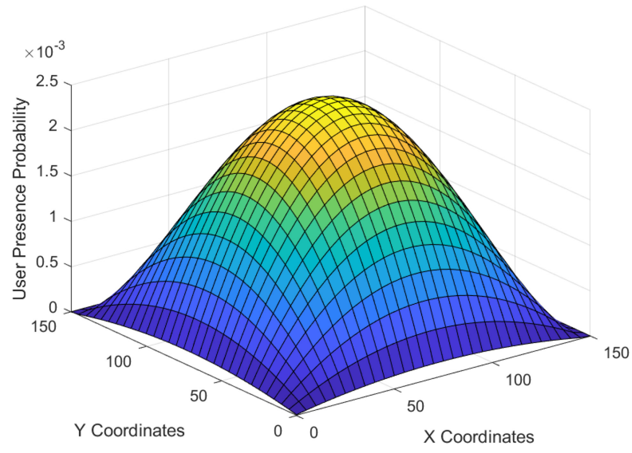

FIGURE 2. Users' distribution over the AoI.

By substituting the coordinates (x, y) of the subcell into Eqn. (3), the probability p that a user is present in the subcell is determined. This step is crucial for computing the probability $P(N)$, which represents the likelihood of N users being in a specific subcell, as

$$P(N) = \binom{N_{\mathrm{n}}^{\mathrm{k}}}{N} p^N q^{N_{\mathrm{n}}^{\mathrm{k}} - N} = \frac{N_{\mathrm{n}}^{\mathrm{k}}!}{(N_{\mathrm{n}}^{\mathrm{k}} - N)!N!} p^N q^{N_{\mathrm{n}}^{\mathrm{k}} - N} \tag{5}$$

where q is the complementary probability $1 - p$.

The probability that the quality metric fails to be met for any specified minimum number of users $N_{\min}$ can thus be articulated as

$$P(N < N_{\min}) = \sum_{N=0}^{N_{\min}} \binom{N_{\mathrm{n}}^{\mathrm{k}}}{N} p^N q^{N_{\mathrm{n}}^{\mathrm{k}} - N} \tag{6}$$

The distribution of users within the system's subcells, as shown in Fig. 2, demonstrates a non-uniform spread due to the RWP mobility model, illustrating that users are dispersed unevenly across the subcells.

V. QOD INCENTIVE SCHEME

To address the impact of users' mobility in the previous section, we now propose a distribution-aware and learning-based dynamic-incentive scheme.

A. NECESSITY FACTOR

To guarantee the completion of tasks with an acceptable QoD, we now propose a necessity factor (Ω), where the demand for incentive to complete the task is indicated by the necessity factor. Normalized from various parameters, Ω indicates the degree of incentivization required to ensure task completion. For instance, $\Omega = 0$ if there are sufficient users and the QoD is met. The optimal necessity factor may be determined by considering a variety of scenario parameters. By designing attraction regions inside the targeted subcells, this factor is used to determine a dynamic incentive for each scenario, shifting the distribution of users across subcells. Our formulated Ω is proposed as

$$\Omega = w_1 x_1 + w_2 x_2 + w_3 x_3 \tag{7}$$

where x_1, x_2, and x_3 reflect the impact of three scenario parameters, which are the probability of satisfying the minimum number of sensing users, QMS, and deadline on the necessity factor, respectively. Based on how important each parameter is to achieve the task goal, the administrator can set each parameter weight (w_1, w_2, and w_3), while the summation of the weights is unity. Each of these weights ensures guaranteed unitless summation within Ω.

Each subcell has a different number of available users. This distinction should be considered in the incentive system to ensure that all tasks, regardless of location, adhere to the minimum required number of sensing users.

$$x_1 = \ln(2 - P(N \geq N_{\min})) \tag{8}$$

where $P(N_{\min})$ is the probability of satisfying the minimum required number of sensing users $N_{\min}$, obtained from Eqn. (5). This parameter is essential in establishing the necessity factor because, as a consequence of the mobility model, central subcells are more likely to satisfy the minimum required number of sensing users than subcells on the border. It is important to note that $P(N \geq N_{\min})$ is inversely related to the need for incentives, where a decrease in $P(N \geq N_{\min})$ will drive a larger x_1 value.

The required QMS should be met in order to reach a satisfactory QoD. As a result, the need for incentives is influenced by the number of measurements

$$x_2 = \ln\left(2 - \frac{Q^{\mathrm{t}}_{\mathrm{ach}}}{Q^{\mathrm{t}}_{\min}}\right) \tag{9}$$

where $Q^{\mathrm{t}}_{\mathrm{ach}}$ represents the estimated achievable quality based on the LSTM prediction and $Q^{\mathrm{t}}_{\min}$ is the minimum required quality for each task. With higher achievable quality, it is clear that incentives are less necessary.

To ensure that the tasks are completed within the given time frame, the deadline is a crucial factor that should be taken into account. The need for incentives to ensure the accomplishment of the sensing task increases as the deadline approaches.

$$x_3 = \ln\left(1 + \frac{1}{j - i - 1}\right) \tag{10}$$

where j is the deadline of the task, we can see that as sensing duration i gets larger, the value of x_3 increases then the need for an incentive is higher.

B. ATTRACTION AREA

Based on the information from the preceding subsection, with $0 \leq x_1 \leq \ln(2)$, $0 \leq x_2 \leq \ln(2)$, and $0 \leq x_3 \leq \ln(2)$, along with the condition that $w_1 + w_2 + w_3 = 1$, it follows that $0 \leq \Omega \leq \ln(2)$. Consequently, the calculation of the normalized necessity factor proceeds as follow

$$\overline{\Omega} = \frac{\Omega}{\ln(2)} \tag{11}$$

The calculation of the necessity factor aims to facilitate the creation of an attraction zone within the specified subcell, which can be implemented as follows

$$g(x, y) = \left(1 - \overline{\Omega}\right)A_t + \left(\frac{\overline{\Omega}}{p}\right)A_c \tag{12}$$

with p obtained from Eqn. (3) as mentioned earlier to represent the users' presence probability within the targeted subcell, and where A_t represents all the points within the AoI as

$$\begin{aligned} A_t = (u(x_d + X_{\mathrm{AoI}}) - u(x_d - X_{\mathrm{AoI}})) \\ \cdot (u(y_d + Y_{\mathrm{AoI}}) - u(y_d - Y_{\mathrm{AoI}})) \end{aligned} \tag{13}$$

while A_c corresponds to all the points within the targeted subcell as

$$\begin{aligned} A_c = \left(u\left(x_d - x^\circ + \triangledown\right) - u\left(x_d - x^\circ\right)\right) \\ \cdot \left(u\left(y_d - y^\circ + \triangledown\right) - u\left(y_d - y^\circ\right)\right) \end{aligned} \tag{14}$$

The coordinates of each point are evaluated using Heaviside unit step functions [24]. Using the Heaviside unit step function, it is possible to distinguish between points inside and outside the targeted subcell.

Eqn. (12) is used to scale the distribution of users' locations by creating an attraction area in the subcell $[x^\circ, x^\circ + \triangledown][y^\circ, y^\circ + \triangledown]$, while $\overline{\Omega}$ is the normalized necessity factor between 0 and 1. A higher $\overline{\Omega}$ will result in more users attracted to the targeted subcell. Eqn. (12) and Eqn. (3) can be multiplied to obtain the distribution of the new users, which is thought to be a response to the demand for further users within the targeted subcell in order to complete the sensing task successfully.

C. RECRUITMENT OPTIMIZATION

The total cost for each task $C(t)$ is the incentivization amount of the recruited users to perform the task $C(t) = \sum_{k=1}^{K}\sum_{n=1}^{N} C_n^k$. To overcome the task assignment challenges associated with users' heterogeneity, the following binary integer programming (BIP) optimization is formulated

$$\min \; C(t) \tag{15a}$$

$$\text{s.t.} \; \sum_{k=1}^{K}\sum_{n=1}^{N} x_n^k C_n^k \leq B_t \tag{15b}$$

$$\sum_{k=1}^{K}\sum_{n=1}^{N} x_n^k Q_n^k \geq Q^{\mathrm{t}}_{\min} \tag{15c}$$

$$x_n^k \in \{0, 1\}, \quad \forall n \in \left[N_n^k\right], \quad \forall k \in [K] \tag{15d}$$

The system optimization aims to minimize incentivization costs while ensuring QMS. The variables C_n^k, and Q_n^k represent the cost, and quality of users from different classes. The decision parameter x_n^k determines which users are selected for the sensing task. The system operates within the task budget constraints B_t and a minimum quality threshold

Algorithm 1 User Recruitment and Task Allocation in ES Systems

Require: N_{total}, T, User dataset, LSTM model, $Q^{\text{t}}_{\text{min}}$, B_t
1: **for** each task t in T **do**
2: Utilize RWP model to estimate the spatial distribution of available users
3: Calculate the corrected number of users N_m
4: Obtain the more corrected number N_n^k using LSTM prediction
5: Compute quality metric satisfaction (QMS) for task t
6: Calculate necessity factor (Ω) for task t and start the assignment starting by the most necessary task
7: **for** each user in N_n^k **do**
8: Initialize cost matrix **W**, and necessary variables
9: Apply Hungarian method to cost matrix **W**
10: **while** QMS not met **do**
11: Perform row and column reductions on **W**
12: Cover the matrix **W** with the minimum number of lines
13: **if** QMS met **then**
14: **return** binary assignment matrix **A** for t
15: **else**
16: Update Ω
17: Find the smallest uncovered element in **C**
18: Modify the matrix to improve the assignment for task t
19: **end if**
20: **end while**
21: **end for**
22: **end for**
23: **Output:** Task assignments for all tasks in T with optimized quality and cost

$Q^{\text{t}}_{\text{min}}$ to guide the recruitment and incentivization process of users.

The assignment of tasks is conducted through the optimization process over a cost-weighted graph, which is represented by the matrix $\mathbf{W} \in \mathbb{R}^{N_n^k \times T}$. This process begins with the formulation of the cost matrix **W**, incorporating all costs associated with participant recruitment. For the optimization of task assignments, the Hungarian algorithm is utilized, as indicated in Algorithm 1, specifically from lines 9 to 19. Participant allocation is denoted by the binary matrix $\mathbf{A} \in 0, 1^{N_n^k \times T}$. Recognized for its efficiency in computation, the Hungarian method plays a pivotal role in facilitating rapid task assignments, a critical aspect for real-time systems. Additionally, it reliably delivers optimal solutions for one-to-one matching scenarios, which aligns with the objectives of the ES system. The simplicity and versatility of this method in adapting to various cost and objective functions. Most importantly, it is capable of addressing real-life limitations such as participant availability and task deadlines, thus ensuring that the assignments adhere to the system's operational requirements.

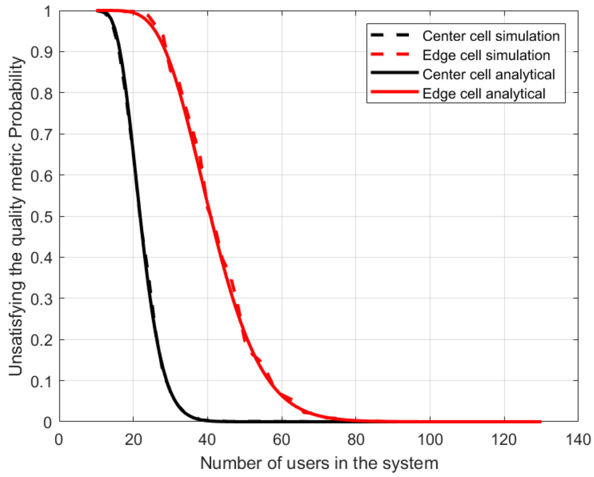

FIGURE 3. **Subcell location effect on the quality metric.**

Algorithm 1 illustrates the comprehensive operation of the proposed method, factoring in variable elements like user mobility and availability. It offers a dynamic solution to address prevailing challenges, leveraging the RWP and LSTM models. This method adeptly balances the cost-quality trade-off, optimizing the approach based on each task requirements.

VI. SIMULATION AND PERFORMANCE EVALUATION

Simulations of an ES system under various situations are performed to test the mathematical expressions derived in this paper, and the obtained results are discussed in this part. The system is composed of nine square subcells. Users move within the AoI of the system, which is considered to be $[0, X_{\text{AoI}}][0, Y_{\text{AoI}}]$. Users' velocity is uniformly distributed with $1 \leq v \leq 20$.

A one-step univariate LSTM model is developed to test the integration of learning within the proposed incentivization scheme, utilizing 70% of the data for training and 30% for testing. The model's hyperparameters include 3000 epochs, 100 neurons, and a batch size of 1. These parameters were chosen through a trial-and-error process to avoid overfitting and underfitting.

As previously shown in Fig. 2, the center subcell has a higher user density, which can significantly impact the system. It is anticipated that the quality metric's probability of not being satisfied in border subcells will be considerably higher than in center subcells. As a result, the AoI edges may turn into blind spots, preventing the complete system's coverage. Fig. 3 shows these probabilities, where it is clear that simulation results align with analytical results for both edge and center subcells.

Fig. 4 illustrates the relationship between V_{th} and the corrected users number, proving that although velocity may not impact user distribution, it does impact the number of users qualified to perform the sensing task. It is noticeable that the relationship is linear as having higher V_{th} will enable the system to include users at higher velocities, subsequently

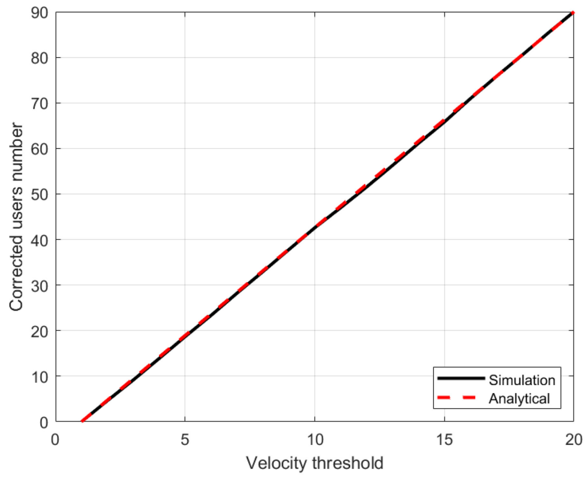

FIGURE 4. Velocity threshold effect on corrected number of users.

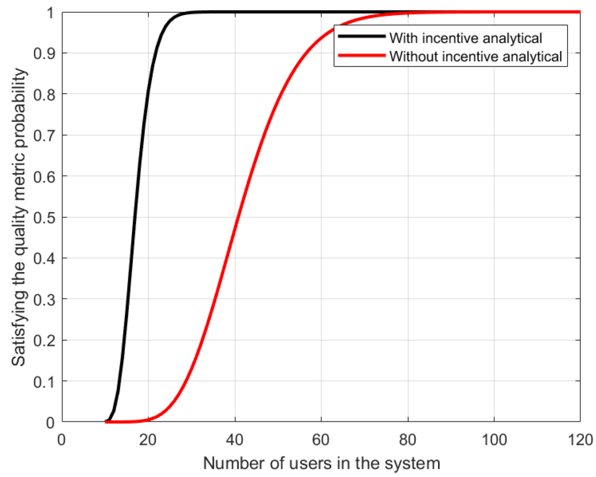

FIGURE 6. Number of users effect on the quality metric.

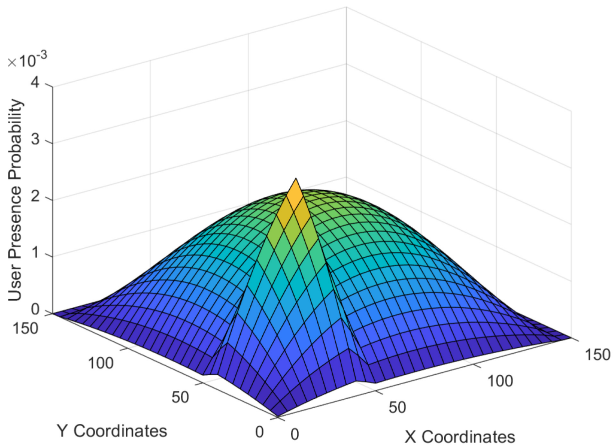

FIGURE 5. Attraction area.

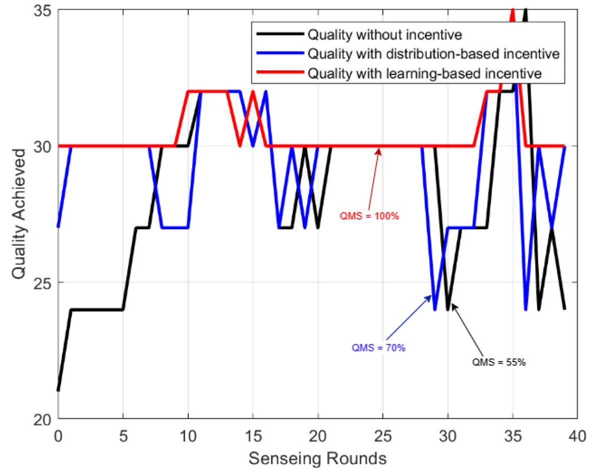

FIGURE 7. QMS Enhancement.

increasing the corrected number of users by including more users in the sensing tasks.

The ability of the incentive scheme to create an attraction area within a particular subcell is evaluated using Fig. 5 and Fig. 6. Users' attraction area was created in an edge subcell, as shown in Fig. 5, increasing the likelihood that users will be present in the targeted subcell.

The formed attraction area can resolve various problems that ES systems encounter as it improves system coverage by increasing the likelihood of users' presence in desired locations, which raises the likelihood of quality metric satisfaction, as shown in Fig. 6. Additionally, it demonstrates that the incentive scheme improved the system's coverage without the necessity to add more users, making it a suitable incentive scheme for scenarios where the budget is constrained.

Fig. 7 demonstrates how learning-based incentives exploit the LSTM model predictions to improve the system performance. We compare between the distribution-aware incentive scheme and the LSTM learning-based incentive scheme. The non-incentive scheme is also included as a baseline for comparison. It is noticeable that the learning-based incentive scheme helped the system to maintain the quality metric satisfied. Although both incentive schemes improved the QMS, the learning-based incentive scheme outperformed the distribution-aware incentive scheme under a budget-constrained scenario due to its previous knowledge of the users' availability pattern. The proposed model utilizes this knowledge to assign the tasks based on the users' availability. Moreover, this knowledge enables the system estimate the achievable QMS. Therefore, this knowledge prevented any drop in the QMS, which enabled the ES system to maintain full spatiotemporal coverage.

Concerned with the incentive schemes' normalized costs to keep the quality metric satisfied, Fig. 8 tackles the normalized cost, as the number of incentives that the system needs to satisfy the quality metric. It is shown that the prediction-supported incentive scheme satisfies the quality metric at a reduced cost due to its previous knowledge of users' availability patterns. This knowledge enabled the learning-based incentive scheme to determine the minimum

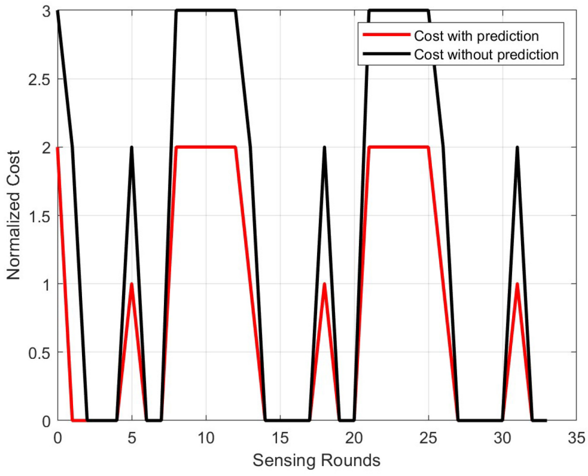

FIGURE 8. Cost comparison.

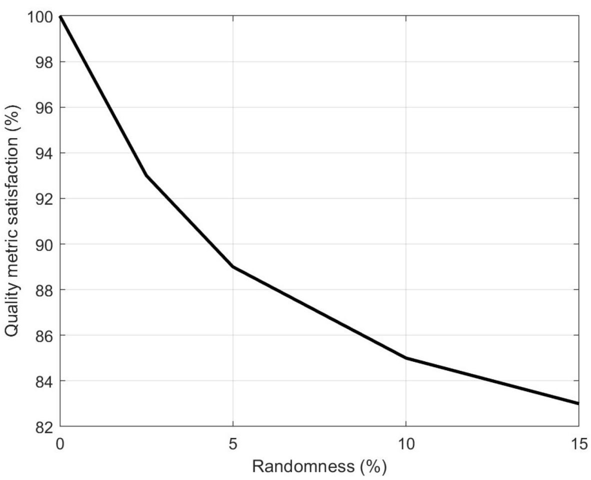

FIGURE 9. Randomness effect on QMS.

needed number of users to satisfy the quality metric and incentivize them. Thus, the learning-based incentive scheme succeeded in its objective as it was able to satisfy the QoD metric and minimize the cost, as shown in both Fig. 7 and Fig. 8.

Fig. 9 demonstrates the influence of the randomness on the performance of the proposed scheme. The results show the overall performance, taking into account all the system's parameters, such as user availability. Increasing the randomness in training and testing data results in increasing the percentage of error in the prediction, subsequently decreasing the QMS. However, as shown in Fig. 9, the randomness does not affect the QMS drastically, and the system was able to perform appropriately at a high randomness percentage.

Table 2 shows the LSTM model performance under various randomness percentages to test its ability to perform in real-life scenarios where users' availability pattern is not always fixed. As expected, as randomness increases, the percentage of error increases; however, the increment in the percentage of error does not affect the system performance drastically. The table clearly shows that the model robust under randomization situations.

TABLE 2. LSTM model performance.

Randomness	0%	2%	5%	10%
RMSE	0.097	0.19	2.24	2.79
Correct prediction	288	288	275	269
Incorrect prediction	0	0	13	19
Percentage of error	0	0	4.5%	6.6%

VII. CONCLUSION AND FUTURE WORK

This paper assesses the users' mobility and availability impact on QoD and the system's spatiotemporal coverage. RWP mobility model was adopted to emulate users' mobility and velocity. The users' distribution PDF was derived to evaluate the mobility effect on the system's spatial coverage and QoD. Moreover, we introduced the corrected number of users to reflect realistic system's performance, where users with higher velocities and lower task completion rates should not be considered.

A distribution-aware and learning-based incentive scheme was also proposed to address the challenges brought on by users' mobility. Understanding the users' distribution enabled the creation of an attraction area in the targeted subcells to satisfy the quality metric and improve the system coverage. This was achieved by calculating the necessity of the incentive based on three scenario-related parameters, which are the probability of satisfying the minimum number of sensing users, QMS, and deadline on the necessity factor. Moreover, the proposed incentive scheme enhances the incentivization process and deals with real-world challenges, where the presence of users does not mean their availability for doing the tasks. An learning model was developed to predict the availability of users for sensing. The integration of the learning model enabled the system to maintain the quality metric satisfied at the lowest possible cost and eliminated the negative impact of users' availability challenge.

This paper demonstrates the significant influence of users' mobility and availability on the quality and coverage of ES systems, highlighting the pivotal role these factors play in enhancing system performance. Through the use of LSTM model, the paper emulates critical user behavior parameters, such as availability, offering insights into how similar predictive models can be integrated into the system design. This approach underscores the importance of considering user behavior in the design of user-centric ES systems, opening avenues for leveraging behavioral models to improve the efficacy and reliability of these systems. In future work, the consideration of collecting real-world data from users to have a realistic dataset of heterogeneous users' availability patterns and other behaviors is necessary to keep advancing the research. Testing different learning models' performance for ES purposes using this data will provide valuable insights into optimizing these systems.

REFERENCES

[1] J. Zhang and X. Zhang, “Multi-task allocation in mobile crowd sensing with mobility prediction,” *IEEE Trans. Mobile Comput.*, vol. 22, no. 2, pp. 1081–1094, Feb. 2023.

[2] S. Liu, Z. Wan, Y. Yuan, Q. Dong, B. Yang, and F. Hao, “Efficient certificateless blind signature scheme with conditional revocation for mobile crowdsensing within smart city,” *IEEE Internet Things J.*, vol. 11, no. 9, pp. 15985–15997, May 2024.

[3] L. Wang, D. Zhang, Y. Wang, C. Chen, X. Han, and A. M’hamed, ‘Sparse mobile crowdsensing: Challenges and opportunities,” *IEEE Commun. Mag.*, vol. 54, no. 7, pp. 161–167, Jul. 2016.

[4] Y. Zhen, Y. Wang, P. He, Y. Cui, R. Wang, and D. Wu, “Popularity balanced multi-task bundling for mobile crowd sensing,” *IEEE Trans. Knowl. Data Eng.*, vol. 36, no. 7, pp. 3390–3401, Jul. 2024.

[5] X. Yang, Z. Zeng, A. Liu, N. N. Xiong, T. Wang, and S. Zhang, “A decentralized trust inference approach with intelligence to improve data collection quality for mobile crowd sensing,” *Inf. Sci.*, vol. 644, Oct. 2023, Art. no. 119286.

[6] Y. Ogawa, G. Hasegawa, and M. Murata, “Probabilistic control of dynamic crowds toward uniform spatial-temporal coverage,” *IEEE Trans. Mobile Comput.*, vol. 23, no. 2, pp. 1050–1065, Feb. 2024.

[7] X. Guo, C. Tu, Y. Hao, Z. Yu, F. Huang, and L. Wang, “Online user recruitment with adaptive budget segmentation in sparse mobile crowdsensing,” *IEEE Internet Things J.*, vol. 11, no. 5, pp. 8526–8538, Mar. 2024.

[8] S. B. Azmy, N. Zorba, and H.S. Hassanein, “Robust quality metric for scarce mobile crowd-sensing scenarios,” in *Proc. IEEE Int. Conf. Commun. (ICC)*, 2018, pp. 1–5.

[9] J. Wang et al., “Learning-assisted optimization in mobile crowd sensing: A survey,” *IEEE Trans. Ind. Informat.*, vol. 15, no. 1, pp. 15–22, Jan. 2019.

[10] G. Liu, F. Xiao, C.-T. Lin, and Z. Cao, “A fuzzy interval time-series energy and financial forecasting model using network-based multiple time-frequency spaces and the induced-ordered weighted averaging aggregation operation,” *IEEE Trans. Fuzzy Syst.*, vol. 28, no. 11, pp. 2677–2690, Nov. 2020.

[11] H. Nikam, S. Satyam, and S. Sahay, “Long short-term memory implementation exploiting passive RRAM crossbar array,” *IEEE Trans. Electron Devices*, vol. 69, no. 4, pp. 1743–1751, Apr. 2022.

[12] Y. J. Guan, “Financial time series forecasting model based on CEEMDAN-LSTM,” in *Proc. 4th Int. Conf. Adv. Comput. Technol., Inf. Sci. Commun. (CTISC)*, 2022, pp. 1–5.

[13] T. Luo, J. Huang, S. S. Kanhere, J. Zhang, and S. K. Das, “Improving IoT data quality in mobile crowd sensing: A cross validation approach,” *IEEE Internet Things J.*, vol. 6, no. 3, pp. 5651–5664, Jun. 2019.

[14] Z. Bai and R. Zhuo, “Quality management of crowd sensing data based on machine learning,” in *Proc. Int. Conf. Comput. Inf. Big Data Appl. (CIBDA)*, 2020, pp. 185–188.

[15] S. Xu, X. Chen, X. Pi, C. Joe-Wong, P. Zhang, and H. Y. Noh, “iLOCuS: Incentivizing vehicle mobility to optimize sensing distribution in crowd sensing,” *IEEE Trans. Mobile Comput.*, vol. 19, no. 8, pp. 1831–1847, Aug. 2020.

[16] S. Chessa, L. Foschini, and M. Girolami, “Understanding human mobility for CrowdSensing strategies with the ParticipAct data set,” in *Proc. IEEE Glob. Commun. Conf. (GLOBECOM)*, 2020, pp. 1–6.

[17] S. B. Azmy, N. Zorba, and H. S. Hassanein, “Quality of coverage: A novel approach to coverage for mobile crowd sensing systems,” in *Proc. Glob. Inf. Infrastruct. Netw. Symp. (GIIS)*, 2018, pp. 1–5.

[18] H. Gao, J. An, C. Zhou, and L. Li, “Quality-aware incentive mechanism for social mobile crowd sensing,” *IEEE Commun. Lett.*, vol. 27, no. 1, pp. 263–267, Jan. 2023.

[19] X. Kong, P. Li, T. Zhang, M. Tu, and Q. Liu, “A budget constraint incentive mechanism in spatial-temporal mobile crowdsensing,” in *Proc. IEEE 21st Int. Conf. High Perform. Comput. Commun. IEEE 17th Int. Conf. Smart City IEEE 5th Int. Conf. Data Sci. Syst. (HPCC/SmartCity/DSS)*, 2019, pp. 1726–1731.

[20] S. Afzal-Houshmand, S. Homayoun, and T. Giannetsos, “A perfect match: Deep learning towards enhanced data trustworthiness in crowd-sensing systems,” in *Proc. IEEE Int. Mediterr. Conf. Commun. Netw. (MeditCom)*, 2021, pp. 258–264.

[21] W. Zhang, Z. Li, and X. Chen, “Quality-aware user recruitment based on federated learning in mobile crowd sensing,” *Tsinghua Sci. Technol.*, vol. 26, no. 6, pp. 869–877, Dec. 2021.

[22] O. Naserallah, S. B. Azmy, N. Zorba, and H. S. Hassanein, “Impact of users’ mobility on the quality of edge sensing systems,” in *Proc. IEEE Glob. Commun. Conf. (GLOBECOM)*, 2022, pp. 4196–4201.

[23] M. Alzard, S. Althunibat, and N. Zorba, “On the performance of non-orthogonal multiple access considering random waypoint mobility model,” in *Proc. IEEE Int. Conf. Commun. (ICC)*, 2022, pp. 721–725.

[24] C. Bettstetter and C. Wagner, “The spatial node distribution of the random waypoint mobility model,” in *Proc. Mobile Ad-Hoc Netzwerke WMAN*, 2002, pp. 41–58.

[25] A. Joshi, P. K. Deshmukh, and J. Lohokare, “Comparative analysis of vanilla LSTM and peephole LSTM for stock market price prediction,” in *Proc. Int. Conf. Comput., Commun., Security Intell. Syst. (IC3SIS)*, 2022, pp. 1–6.

[26] H. Taha, *Operations Research: An Introduction*. London, U.K.: Pearson, 2017.