Venue-Popularity Prediction Using Social Data Participatory Sensing Systems and RNNs

KHULUD ALHARTHI[1,2], KHALIL EL HINDI[2], AND SALHA M. ALZAHRANI[1]
[1]Department of Computer Science, College of Computers and Information Technology, Taif University, Taif 21944, Saudi Arabia
[2]Department of Computer Science, College of Computers and Information Sciences, King Saud University, Riyadh 145111, Saudi Arabia

Corresponding author: Salha M. Alzahrani (s.zahrani@tu.edu.sa)

ABSTRACT Analyzing social data as a participatory sensing system (PSS) provides a deep understanding of city dynamics, such as people's mobility patterns, social patterns, and events detection. In a PSS, individuals with mobile devices sense their environment, collect, and share data. For smart cities, intelligent city dynamics analysis has many applications such as for urban planning, transportation systems, city environment, energy consumption, public safety, and city economy. This study aimed to develop an intelligent application to predict the potential number of visitors for specific venues based on the analysis of mobility patterns of individuals. The ability to accurately predict the number of visitors to a venue allows authorities to better understand the behavior of the people and allocate recourses accordingly. We formulated the venue-popularity problem as a sequence-based regression and classification problem. We employed three recurrent neural network (RNN)-based models to predict the locations of popular venues on a city map. The proposed models include basic RNN, long short-term memory (LSTM), and gated recurrent unit (GRU). We constructed several social datasets for Riyadh city using Twitter and Foursquare as the PSS. Our results revealed that modeling venue-popularity prediction as a sequence regression problem yields better results than modeling it as a sequence classification problem. For the city-popularity map prediction problem, the vector autoregression baseline model achieved better performance than the RNN-family models.

INDEX TERMS City dynamics, venue-popularity detection, participatory sensing system, recurrent neural network, social data.

I. INTRODUCTION

A participatory sensing system (PSS), or an urban sensing system, is a system in which individuals and communities sense their surrounding environment by collecting, documenting, viewing, and sharing local observations and data of the environment [1]. Social data have rapidly increased as numerous users add posts related to their places of interest or connected to their geographic locations. With the growth of social networks and the ubiquity of smartphones, social data have become one of the PSS sources to sense different aspects of city environment and to provide a deep understanding of city dynamics [2]. Understanding city dynamics involves detecting human mobility patterns [3], [4], searching for better planning decisions for city tours [5], community and social behavior detection [6], and individual behavior detection [7].

In smart cities, urban analysis has been utilized for intelligent applications such as for urban planning, transportation systems, city environment, energy consumption, public safety and security, and city economy [8]. In addition, a number of researches have explored city economic applications such as predicting real estate prices [9], searching for the optimal placement of new businesses [10], and predicting potential customers or visitors of specific venues [11].

Twitter[1] and Foursquare[2] are two important sources of the social data. Twitter is a microblogging and social networking service where users post limited-length text named tweets, which may include links to external platforms. Foursquare is a popular location-based social networking service that allow users to check into different locations and share information with their friends. By August 2015, Foursquare has a society

The associate editor coordinating the review of this manuscript and approving it for publication was Jerry Chun-Wei Lin.

[1]https://twitter.com.
[2]https://foursquare.com.

with more than 50 million active users who made more than 10 billion check-ins. The growing society of social networks that support geo-tagging services combined with the availability of their corresponding application programming interfaces (APIs) open the doors to a rich and valuable source for analyzing urban human mobility patterns [12]. Studies that use social data in the context of PSS can be categorized into five classes, namely, mobility patterns analysis, city dynamics understanding, social pattern detection, city-event detection, and human behavior analysis. Social pattern detection investigates the human relationships in a city to build various smart city applications such as community detection and recommendation systems. Event detection aims to detect natural events such as earthquakes or unnatural events such as stock market changes. Human behavior analysis aims to study human collective behaviors and opinion evaluation and to understand the reasons behind people's actions.

Prediction of the potential number of visitors to a venue in a city is considered a city economy application that can be built based on PSS human mobility data. Many researchers model the venue-popularity prediction problem as a typical feature-target machine-learning problem and do not consider the performance of these models on the dataset of venues with different distributed categories [10], [13], [14]. This emphasizes the need for investigating different modeling for the venue-popularity prediction problem with venues with different distributed categories.

Our study shares the same thought as [2], which considers social networks data as PSS sources to sense different aspects of the city. We investigated the feasibility of modeling the venue-popularity prediction problem as three sequence-based models using deep-learning algorithms with social data from different categories. The three sequence-based venue-popularity prediction models are the sequence regression model, sequence classification model, and city popularity map prediction model (based on sequences of popular locations in a city). We investigated recurrent neural network (RNN)-based models, which were useful for event sequencing and prediction. The proposed RNN-based models were models based on basic RNN, long short-term memory (LSTM), and gated recurrent unit (GRU). First, we modeled the problem as a sequence regressor for predicting the number of visitors to a venue by considering the sequence of three weeks' features of the area around a venue; second, we modeled the problem as a sequence classifier to predict the popularity class of a venue as popular, above average, below average, or unpopular also considering the sequence of three weeks' features of the area around the venue; and third, we modeled the problem of map-popularity prediction to predict the location of the top-six popular venues in the city based on the sequence of the top-six popular venues of the three previous days. To the best of our knowledge, no previous work has modeled the venue-popularity prediction as using RNN-family algorithms and social data in the context of PSS. The stability of these models across different venue categories was also examined. Accordingly, we constructed three

types of datasets for Riyadh city, based on Foursquare and Twitter data: the first dataset for a single category, the second for multiple categories, and the third for multiple categories. However, in the test set, we included a different category that was not present in the training set.

The rest of this article is organized as follows: Section II provides a detailed literature review of PSS and urban analysis methods; section III provides details of the RNN-based models, their training, and testing; section IV provides the social dataset used as a PSS from Twitter and Foursquare for a set of venues in Riyadh city, evaluation measures, and baselines; and section V presents the experimental results and discussions. Finally, section VI describes the conclusions and the future research direction.

II. RELATED WORKS

Ubiquitous computing era begins when a computing environment is available for anyone, anywhere, and at any time. One important requirement of this era is the ability to sense the environment for pervasive data. Early approaches of environment sensing depend on wireless sensor networks, but these approaches involved high costs, especially for covering large areas, in addition to the maintenance costs. Owing to these drawbacks, the focus was shifted toward potential low-cost sensing approaches that use social data generated by smart phones to sense the environment. Such approaches use humans as a source of sensing, and therefore, they such a system is called a PSS [2]. Studies on social data-based human mobility pattern detection, city dynamic for economy applications, and popularity prediction were investigated in depth for the period 2014–2020, and their results were reviewed and compared.

A. HUMAN MOBILITY PATTERN DETECTION

Human mobility patterns detection aims to analyze the mobility patterns of users, including their spatial and temporal information using different data sources such as census data, surveys, mobile phone records, GPS data, and social network data. Social network data include check-in records and geocoded tweets and photos [2], [15]. Analysis of PSS social networks contributed to the solution of different problems such as the prediction of the next location that a user is more likely to move to and the next check-in for marketing and advertising applications [16]. Similarly, the problem of recommending new places and events to attend can be solved more effectively using social locational data [17]. The problem of understanding the factors that affect user movement patterns has been examined [18] based on three aspects: human geographic movement, temporal dynamics, and social network ties. For the health side, effective vaccination strategies were planned based on the detected mobility patterns [19]. Further, mobility patterns detected from PSS social data have become an important factor in many traffic applications [20].

B. URBAN ECONOMIC COMPUTING

Analyzing city dynamics by mining different data sources can indicate the trend of the city's economy [8]. For example in [9], multiple data sources, including OpenStreetMap data, urban atlas data, and property taxes were used to predict the real estate prices using gradient-boosted trees. In another study, gradient-boosted decision tree (DT) trained with a dataset collected from the largest consumer review site in China, Dianping, was used to predict the ability of restaurants to survive and avoid closure [21]. Real estate ranking models based on estimated investment value using pairwise ranking objective and sparsity regularization in a unified probabilistic framework were proposed [22]. The study used features based on users' reviews and moving behaviors mined from a variety of mobile user-generated data such as taxi traces, smart card transactions, and location check-ins. Another real estate ranking model was proposed based on the score of the overall condition and the value of the property inferred from textual comments in real estate listing databases [23]. The model used text analysis methods based on data collected from real estate listings located across the USA originally advertised by real estate agents. The problem of finding business opportunities by answering the question of what kind of business to open and where was investigated in [24] by mining data from a location-based social network called Yelp. The solution suggested partitioning a city into different business districts, and then recommending new business categories for specific business districts.

C. VENUE-POPULARITY PREDICTION

Similar to any task in urban economic computing, for predicting venue-popularity, the dynamics of a city should be analyzed considering the economic goals. An essential ingredient of urban dynamics, which is required by venue popularity prediction, is the human mobility patterns that can be gathered from one or multiple sources of data. These patterns can be harnessed using different computational methods, in particular, machine-learning methods.

1) SOCIAL DATA-BASED VENUE-POPULARITY PREDICTION

The ranking of urban areas based on their potential needs for a new retail business was investigated in [10]. Features of the three retail chains in the city of New York were mined from Foursquare. The study examined the features related to the context of the retail business placement from two aspects: area spatial information and user mobility information. Support vector regression (SVR), M5 DTs, and linear regression (LR) were used to predict the popularity of new business in different areas. The best performance was achieved when a combination of all features was used with SVR. The use of mobility activities in each hour over a week to form temporal profiles of an area was proposed in [12]. Temporal profiles were created for all venues in an area, venues with the same general category, and venues with the same specific category. Similar temporal profiles of regions

with the desired area were computed using the k-nearest neighbor (KNN) algorithm and fed into a Gaussian model to predict the temporal pattern of the desired venue. When using a Foursquare-based dataset for London, this approach showed an improvement of 41% compared to a random selection of the areas as an input of the Gaussian model.

The similarity between neighboring venues and the target venue was computed in [25] based on the distance between them, main category, and subcategory. Venue popularity was predicted using the KNN algorithm, and inverse distance weight interpolation (IDW) for several venues in Texas were obtained from Foursquare and compared against the spatial-auto-regression model. The study showed an average root mean square error (RMSE) reduction of 87.50%. Another prediction model to place locations for new businesses were developed in [13], [14], [26]. The next restaurant branch was predicted using three types of features: review-based market attractiveness, review-based market competitiveness, and geographic proprieties of a candidate location restaurant [13]. Restaurant reviews were mined from Yelp of restaurants in New York City across 260 different categories, and geographic proprieties, including density, neighborhood entropy, competitiveness, and Jensen quality that measures the spatial interactions between restaurants and venues, were mined from Foursquare. Results showed that the combination of all features achieves better results than considering geographic and review-based features using ridge regression, SVR, and gradient-boosted regression trees (GBRT). A popularity predictive model of 20,877 food businesses in Singapore available from Facebook was developed in [26]. The study incorporated features of the target business and other businesses in the candidate area and the number of check-ins. The study revealed that GBRT yielded the best results. The optimal location to place a live campaign by predicting the potential number of spectators was studied in [14] based on a dataset collected for New York city from Foursquare; the features used in the study included area density, area popularity, neighbor entropy, the degree of change in area popularity, the degree of uniqueness in the total number of check-ins, the amount of available open area for the live campaign, and the most time slot the people visit this area, and the SVM had an accuracy of 72.6% [14].

Existing established stores were treated as sensors in [11] to sense the nearby human activities for retail stores. The features were collected based on geographical properties, mobility patterns, and social behaviors using a dataset built from Foursquare and Gowalla for Starbucks. An inference algorithm was developed for potentially estimating the number of customers. The effect of information diffusion on business popularity based on the social connections between visitors in the current and previous time frames, and geographical types of local and foreign visitors was studied in [27]. A combination of these two aspects of information diffusion was used to classify businesses as popular or unpopular using the Yelp dataset of different cities; for this purpose, SVM achieved the highest accuracy of 89% among all

other models. A Meetup.com dataset of three cities, namely, New York, London, and Sydney, was used in [28] to predict and improve the popularity of group-organized event using features mined according to four aspects: event location, diversity of Meetup group, temporal preferences of group members, and semantic quality of the event title and content. Classification and regression tree (CART) was used to predict the social influences of users on the decision of other users to attend the same event. Table 1 summarizes the aspects of feature mining and the model types used in different studies.

TABLE 1. Summary of features and prediction models used in previous studies for venue-popularity prediction using social data.

Model	Spatial	Temporal	Category	Text-based	Social behaviors
SVR	[10, 13]	[10]		[13]	
DT	[10]	[10]			
Linear Regression	[10]	[10]			
KNN	[25]	[12]	[25]		
Gaussian Model		[12]			
IDW	[25]		[25]		
GBRT	[13]	[26]	[26]	[13]	
Ridge Regression	[13]			[13]	
SVM	[14, 27]	[14]	[14]		[27]
Inference on Graph	[11]	[11]			[11]
CART	[28]	[28]	[28]	[28]	[28]

2) VENUE-POPULARITY PREDICTION BASED ON OTHER DATA

In many studies, a data source other than social data was used to predict the popularity of a venue and to rank a candidate location for a new business. For example, a visual analytics system called Smart Advertising Placement (SmartAdP) was developed in [29] to locate the optimal placement of billboards using GPS trajectory data, road network data, and point-of-interest (POI) data. Using a greedy heuristic method, two queries were used to extract the candidate places based on the audience coverage and costs. A matrix of the intra- and inter-spatial interaction coefficients introduced in [30] was used to define the attraction forces among commercial stores using 21 types of business in Turin, Italy. Simulated annealing was used to find a candidate place for a new store of a given type of business with the fitness function computed based on Q-index [31]. Map query data from Baidu maps and POI data were used in [32] to predict the level of demand for stores belonging to two categories, namely, ''coffee shop'' and ''express inn'' in Beijing. Wi-Fi connection data were used as an indicator of customer numbers and served as part of ground truth along with the number of queries from Baidu Maps. Several features, including density, traffic convenience, prices, popularity, distance from the city center, and specific popularity that consider the same category, were

mined. Based on the predicted number of customers, POI was ranked using different models; the random forest model was reported to provide the best performance in most cases.

Attributes of candidate areas for placing a new store were extracted in [33] based on datasets collected from Google Maps. The list of attributes includes geographical properties such as whether or not there is a nearby hospital, public transport, school, road entrance, housing area, or a parking place. Association rules were used to rank these attributes based on their correlation with a given business type. A prediction model of hotels' trading volume was proposed in [34]; the trading volume was defined as the number of optional customers times the average price of the hotel. The prediction model consists of multimodal classifiers derived from the deep forest algorithm using three types of extracted features: spatial, mobility, and business features, such as the price and quality of provided services. Besides the existing knowledge graphs such as Wikipedia and ConecptNet5, knowledge-graph convolution neural network (KG-CNN) was proposed in [35] to predict the popularity of venues using features generated from urban knowledge graphs. Table 2 summarizes the features extracted from datasets other than social data and the models used by different studies.

TABLE 2. Summary of features and prediction models from literature studies on venue popularity prediction using different data.

Model	Spatial	Temporal	Category	Business-Strategic
Search-based Model	[29, 30]	[29]		
LambdaMART	[32]	[32]	[32]	
Random Forest	[34]	[32]	[32]	
Association Rules	[33]		[33]	
Multimodal Classifiers Ensemble		[34]		[34]
KG-CNN	[35]	[35]	[35]	[35]

D. USING RNN FOR CITY DYNAMICS ANALYSIS

1) RECURRENT NEURAL NETWORK (RNN)

Deep learning is defined as a type of machine-learning technique where many layers of nonlinear information processing units called neurons are used for supervised or unsupervised feature extraction and transformation and for pattern analysis and classification [36]. RNNs are a class of neural networks that can process sequential data or time series, where time t at which at an event happens is relevant and affects the event at time $t + 1$. The target function that the RNN approximates takes the input vector X_t and the activation of the hidden layer at time $t - 1$.

Thus, the output, Y_t, is influenced not only by the current input pattern, but also by the pervious patterns $X_0, X_1, \cdots X_{t-1}$, as shown in Figure 1 [36], [37]. An RNN can be made deeper by breaking down the hidden recurrent state into groups organized hierarchically by introducing deeper computations in any of the three computation steps or by incorporating skip connections in the hidden-to-hidden
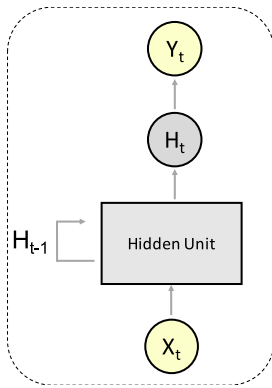
FIGURE 1. Basic Structure of a recurrent neural network (RNN).

path [36]. LSTM is an extended version of the RNN with a multiplicative gate in the hidden layer that controls an accumulative memory [38], [39] It provided encouraging results in many applications [36]. The structure of LSTM was originally proposed in [38], and it had an input gate and an output gate. A forget gate was introduced in [39]. These gates are typically trained to control the amount of memory that needs to be ignored, filtered, and forgotten. An LSTM unit processes each input vector X_t at time step t to compute a hidden state ht and a memory cell state ct that stores all the history observed until time t using the gates. The structures of LSTM and RNN are shown in Figure 2. A form of LSTM is called the GRU, which uses two gates instead of three states with no cell state [40].
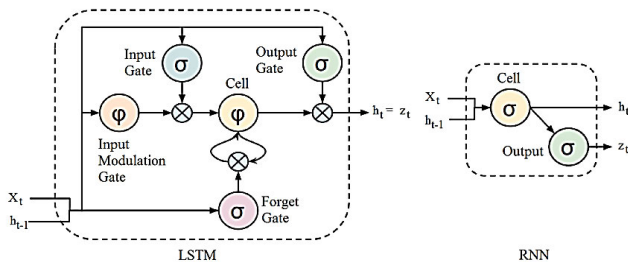

FIGURE 2. Structure of the long short-term memory (LSTM) (left) and the basic RNN (right) [41].

2) PSS-BASED CITY DYNAMICS ANALYSIS USING RNN

RNN and its variations have shown good prediction performance in different applications that require remembering sequences of events such as prediction of stock returns [41], forecasting energy market indices [42], online multi-target tracking [43], and prediction of the next location to stop-over [44]. The use of RNNs for analyzing city dynamics using social data as the PSS has been proposed in many studies. GRU was trained with backpropagation to optimize the cross-entropy loss with Adam optimizer and used to classify GPS traces [40] and to predict the next location the user is most likely to stop-over [44]. Low-dimensional and heterogeneous GPS features were mapped into distributed

vector representations to capture the high-level semantics; then, GRU was used to predict four classes: bike, car, walk, and bus [40]. Foursquare dataset and grid search were used to optimize hyperparameters to predict the next stop by visitors [44].

Another form of RNN called the LSTM was initially proposed in [45]; the idea was to assign one LSTM for each user to predict the human mobility pattern in a crowded context where neighboring LSTMs were connected with a pooling strategy. Each LSTM had 128 hidden units trained with backpropagation to optimize the negative log-likelihood loss, and the hyperparameters were tuned based on cross-validation on a synthetic dataset. A prediction model for traveler mobility was proposed in [46] based on input–output hidden Markov model (IO–HMM), followed by a single-layer LSTM with 128 units trained with backpropagation to optimize the negative log-likelihood with Adam optimizer. Social LSTM was trained using social data from the IgoUgo.com to generate candidate positions as a recommendation of a travel route [47].

As seen from literature studies, most research works modeled the venue-popularity prediction problem as a feature-target machine-learning model. The list in Table 3 shows that the categories covered by exiting literature were either single or multiple categories.

TABLE 3. Summary of the types of categories covered in the dataset; from literature studies on venue-popularity prediction.

Study	Categories covered in the dataset
[10]	Starbucks, Dunkin' Donuts, McDonald
[12]	Venues with Similar category
[25]	Multiple categories
[13]	Restaurants
[14]	Live Campaigns
[11]	Starbucks
[26]	Food business
[27]	Restaurants, Nightlife, Active Life, Shopping, Hotels, Beauty and Spas, Food, Arts and Entertainment, Religious Organizations and Mass Media
[28]	Event
[29]	Billboards
[30]	Hotels, Shops, Services, Gym and Restaurants.
[32, 35]	Coffee shops, Express inn
[33]	Retail Shop
[34]	Hotels

III. METHOD

A. PROBLEM FORMULATION AND GENERAL FRAMEWORK

This work involved four main stages: (i) constructing a social dataset as PSS based on data from Foursquare and Twitter for the purpose of venue-popularity prediction, (ii) constructing a vector of features that are correlated with the number of venue visitors, (iii) developing three venue-popularity prediction models namely popularity regressor, popularity classifier,
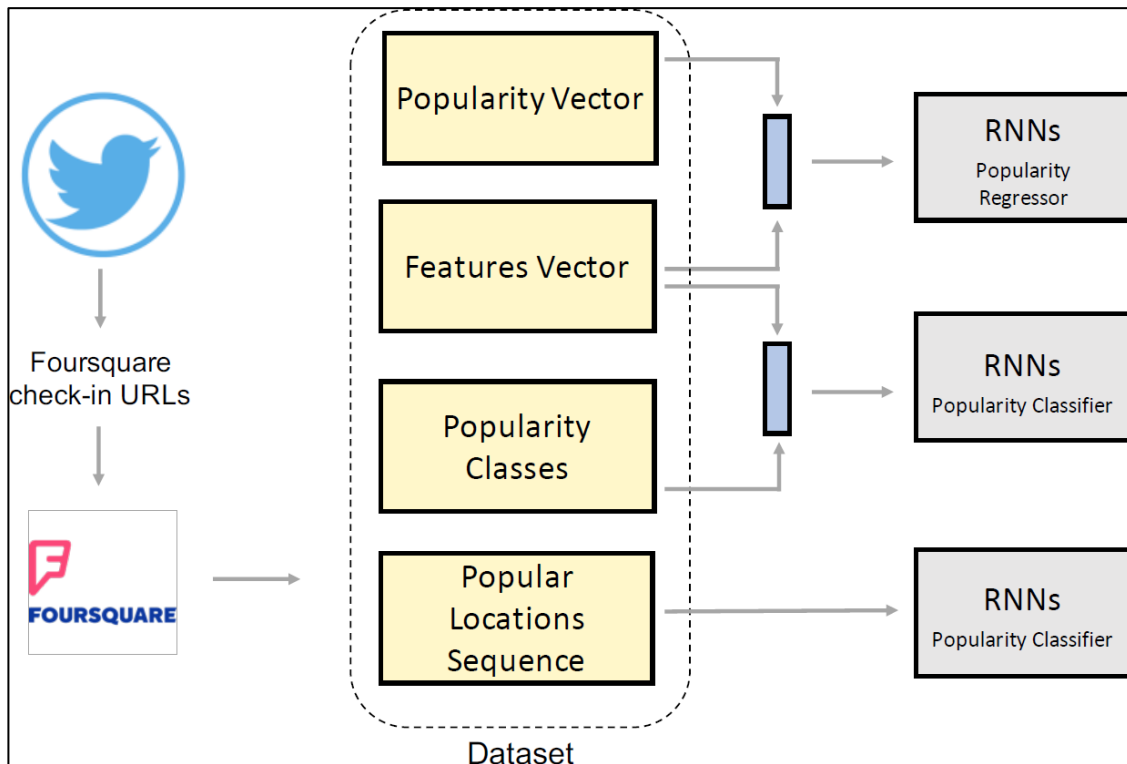
FIGURE 3. General framework of the three proposed models.

and popular-locations predictor using RNNs, and (iv) finally, evaluating the prediction models using different evaluation measures. Figure 3 shows the general framework of the three proposed models.

1) CONSTRUCTING SOCIAL DATASET

In this phase, the social data of the capital city of Saudi Arabia, Riyadh, was extracted from Twitter and Foursquare. We collected publicly shared check-in URLs from Twitter, and then, we collected information on these check-ins and their corresponding venues from Foursquare.

2) CONSTRUCTING FEATURES VECTOR

In this phase, ten urban features were selected from the constructed social dataset. The selection process was based on the correlation between these features and the number of venue visitors. The most relevant features were evaluated to construct the feature vector that describes each training instance.

3) DEVELOPING VENUE-POPULARITY PREDICTION MODEL USING RNNS

We propose three models for the venue-popularity problem:

- A sequence-regression model that is based on a sequence of venue area features in three weeks and predicts the potential number of visitors in the following week (the fourth week). Here the problem is formulated as follows: *Given the selected urban features and the number of*

visitors to different venues in three previous weeks, train an RNN as regressors to predict the potential number of visitors in each of the four timeframes (morning, afternoon, evening, and night) in the next week.

- A sequence-classification model that is also based on a sequence of venue area features in three weeks to predict the popularity-class of the venue in the following week, which could be one of four classes: popular, above average, below average, and unpopular. We formulate the problem as follows: *Given the urban features with the number of visitors to different venues in three previous weeks and their popularity class, train an RNN classifier to predict the popularity-class of the venues in the following week.*

- A city-popularity-map prediction model that predicts the locations on the map of the top-six popular venues based on a three-day sequence of popular locations and predicts the popular venue location on the following day. For this model, we formulate the problem as follows: *Given the top-six popular venues in the city for the previous three days, train an RNN regressor to predict the six most-popular venues in the city in the next day.*

4) EVALUATING THE PREDICTION MODELS

The sequence-regression model and the city-popularity-map prediction model were evaluated according to four measures: root mean squared error (RMSE), mean absolute percentage error, and the normalized versions of these two measures
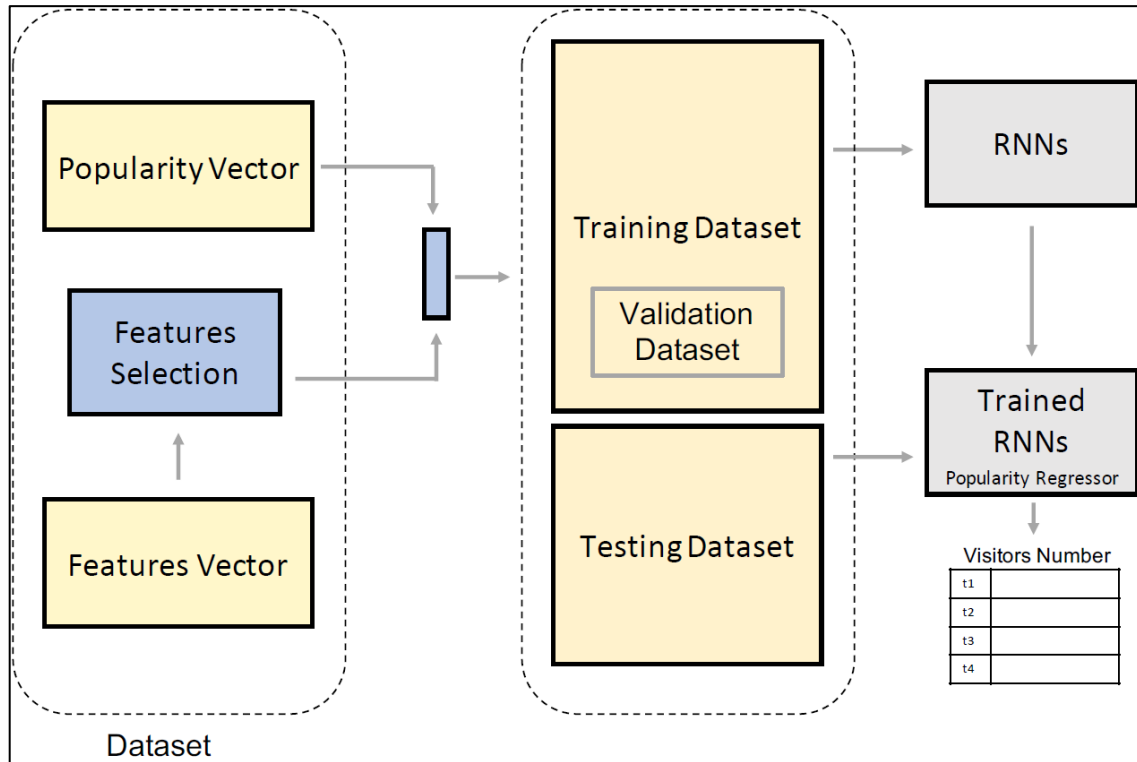
FIGURE 4. General framework of the methodology followed in the sequence regression model.

NRMSE and NMAE. The classification model was evaluated with respect to the accuracy and Matthews correlation coefficient (MCC). Figure 4 shows the general framework for the sequence regression model, which was also followed in the other two models.

B. MINING URBAN FEATURES

In this section, we describe the features extracted from the constructed social dataset. Based on the findings in [10], which stated that 50% of user movements originate from venues within 200 to 300 m and 90% of movements are within 1 km, we defined the region of a venue as the area within 1-km radius of a specific location. Following [13], [25], [27], we considered four types of features: area spatial features, area popularity features, area mobility features, and area social features.

1) AREA SPATIAL FEATURES

Venue Density (N): denotes the number of established venues around the desired venue within a 1-km radius.

Venue Heterogeneity (H): measures the diversity of venues and is obtained based on the Gibbs–Martin index of diversity [54] defined as

$$H = 1 - \frac{\sum \left(\frac{CatNum}{100}\right)^2}{\left(\sum \frac{CatNum}{100}\right)^2} \tag{1}$$

where *CatNum* refers to the number of established venues of the category, cat, in the region of the venue. This feature

includes 10 manually defined categories, in addition to the medical center and school categories.

Venue Homogeneity (Hom): denotes the number of similar venues (of the same category) in the region of the venue.

2) AREA POPULARITY FEATURES

Area Popularity (AP): denotes the number of check-ins in the venues around the desired venue during the week.

Area Popularity of One Category (APC): denotes the number of check-ins in the venues around the desired venue that is of the same category during the week.

3) AREA MOBILITY FEATURES

Area Incoming Mobility (MobI): denotes the weekly number of users who check-in in a venue's area after having checked-in in venues beyond the venue's area.

Area Outgoing Mobility (MobO): denotes the weekly number of user check-ins in a venue's area and then check-in in venues beyond this area.

Local Area Mobility (MobL): denotes the weekly number of users who check-in in the other venues inside the venue's area.

4) AREA SOCIAL FEATURES

Area Male Visitors (AM): denotes the percentage of male visitors to a venue's area during the week.

Area Female Visitors (AF): denotes the percentage of female visitors to a venue's area during the week.

C. URBAN FEATURES SELECTION

The correlation between these features and the number of visitors of the venues was computed using the Pearson correlation coefficient to detect the linear dependencies between these features and the number of the visitors. The following equation defines the Pearson correlation

$$P\,(Fi, Njc) = \frac{Cov(Fi, Njc)}{\sqrt{Var\,(Fi) * (Var(Njc)}} \qquad (2)$$

where *Fi* is the feature number *i*; *Njc*, the number of visitors in venue *j* with category *c*; *Cov*, the covariance; and *Var*, the variance [48] The correlation between these features and the number of venue visitors are listed in Table 4. We calculated the correlation between each feature and the number of visitors in each timeframe and also between each feature and the number of visitors in all the timeframes (ALL_V).

**TABLE 4.** Feature correlations with the number of visitors.

| Feature | Timeframe | | | | | ALL_V |
	1	2	3	4	Avg	
AP	0.14	0.183	0.258	0.028	0.15	0.24999
APC	0.289	0.238	0.332	0.070	0.232	0.3689
MobI	0.146	0.177	0.253	0.0361	0.153	0.24856
MobO	0.122	0.142	0.230	0.0363	0.133	0.2192
MobL	0.099	0.164	0.213	0.002	0.12	0.1994
AM	-0.051	-0.067	-0.0695	0.0704	0.064	-0.0680
AF	0.055	0.069	0.0631	-0.066	0.063	0.06637
N	-0.039	0.0571	0.0427	-0.1185	0.064	0.00844
H	0.037	-0.045	-0.052	0.1067	0.06	-0.0144
Hom	0.005	0.037	0.0959	-0.043	0.045	0.0639
ALLMob	0.132	0.172	0.248	0.0268	0.145	0.23803
MobIO	0.138	0.164	0.248	0.0371	0.147	0.23968

To further analyze the correlation of the mobility features, we computed the average of all mobility features into one feature (ALLMob), and examined it in addition the incoming and outgoing mobility features into one feature (MobIO). We calculated the correlation of ALLMob and MobIO features with the number of visitors. Further, we analyzed the effectiveness of area popularity (AP), area popularity of one category (APC), incoming mobility (MobI), and the two combined features ALLMob and MobIO by comparing their predictive performance for the regression problem using LSTM. The results listed in Table 5 show that using all of the mobility features (ALLMob) with AP and APC features yield the best results in terms of low regression errors. Thus, these features

TABLE 5. Feature performance evaluation based on the regression model using LSTM.

ALL_N	RMSE	MAE	NRMSE	NMAR
AP+APC+MobI	0.06	0.0457	1.0823	0.82437
AP+APC+AllMob	0.0565	0.0382	1.0204	0.6890
AP+APC+MobIO	0.0728	0.0603	1.3132	1.0877
ALL Features	0.08	0.0708	1.4431	1.27715

were selected and used for training and testing the proposed RNN models for the sequence regression problem and the sequence classification problem. The sample structures of a sequence of the three selected features for three weeks and the popularity vector for the sequence regression and classification models are shown in Figure 5. For the city popularity map prediction model, each instance (sample) consists of a sequence of the top-six popular venue locations for the past three days as an input vector and the top-six popular venues on the fourth day as the output vector. Using different splits from the dataset, which is explained in Section IV, we evaluated this setting with a one-category dataset and with a multiple category dataset. The last setting included predicting the category of the popular venue along with its location using a multiple-category dataset, as shown in Figure 6.
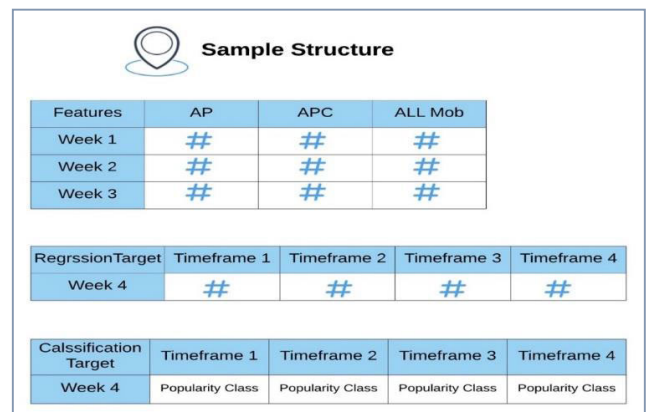
FIGURE 5. Sample structure for the sequence regression problem and sequence classification problem.
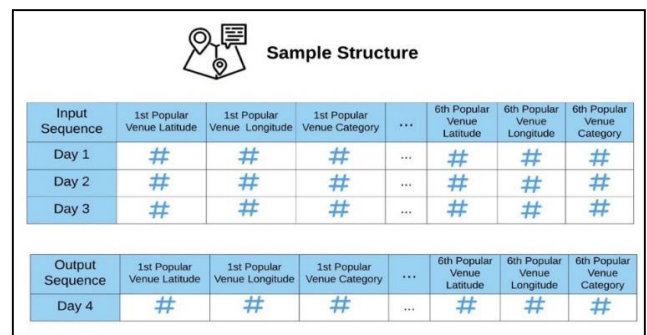
FIGURE 6. Sample structure for the city popularity map prediction model.

D. VENUE-POPULARITY PREDICTION MODEL

The proposed RNNs, including the basic RNN, LSTM, and GRU, were trained to optimize the mean squared error and mean absolute error as the loss function with Adam optimizer for regression. For classification, we used the cross-entropy as a loss function. The hyperparameters included the learning rate, number of training epochs, number of hidden layer nodes, batch size, and dropout value. These parameters were tuned using random search [49] for the basic RNN, and the

optimal parameters obtained were used in all experiments. Random search with 3-cross validation was used with a validation dataset of 20% for the training dataset. The RNNs used for the sequence classification problem have four output layers that share the same hidden layer and one output layer with the SoftMax activation function to predict the popularity class for one timeframe, as shown in Figure 7. The RNNs used for the sequence regression problem have one input layer, one hidden layer, and one output layer with a linear activation function and four nodes, where one node corresponds to each time frame, as shown in Figure 8. For the city-popularity map prediction problem, the RNN has the same structure as the regression model with an output layer with 12 units to represent the longitude and latitude of the location of the six most-popular locations and 18 units to predict the location category.
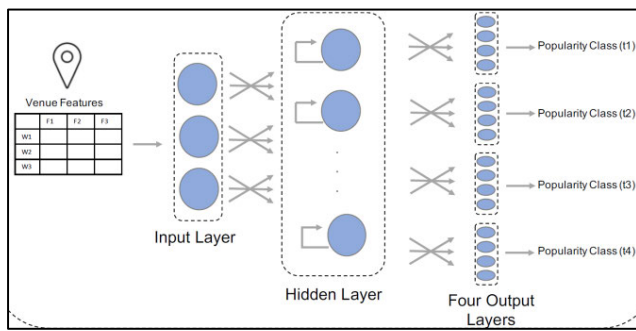

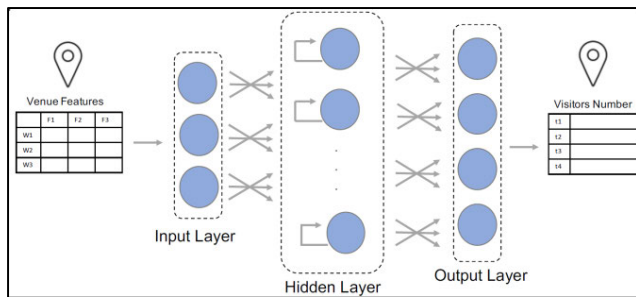**FIGURE 7.** RNN structure in the sequence classification model.

FIGURE 8. RNN structure in the sequence regression model.

IV. EXPERIMENTAL SETUP

A. DATASET

Twitter and Foursquare were utilized as sources for constructing our venue-popularity datasets for Riyadh city. As Foursquare API restricts access to the number of the weekly venue check-ins, check-in URLs publicly shared through Twitter were extracted using the Twitter API. More details about these links were then extracted through the Foursquare API, and these details are listed in Table 6.

The maximum volume of 500 geocoded tweets with the location set to Riyadh city and containing a check-in URL were extracted in every 12-h for the period October 8, 2018, to March 24, 2019. The number of tweets extracted in each week were between 3500 to 5000, and consequently, the total

TABLE 6. Description of check-in details extracted from Foursquare API.

Check-ins Details	Description
Check-in Date	The specific date and time a check-in were made
User ID	The ID of the user who made the check-in
User Gender	The gender of the user who made the check-in
Venue ID	The ID of the venue that the user checks in at.
Venue Name	The name of the venue that the user checks in at.
Venue Latitude	The latitude of the venue location.
Venue Longitude	The longitude of the venue location.
Venue Category ID	The ID of the category of the venue specified by Foursquare API.
Venue Category Name	The name of the category of the venue specified by Foursquare API.

number obtained over a period of 24 weeks was around 100,000. The extracted social data have some shortcomings. First, some extracted venues were assigned to a general category such as "Hotels" and some were assigned to a specific category such as "Resorts." Second, some venues were assigned to categories that were not within the scope of the study: for example, schools, universities, and government buildings (Selected categories are listed in Table 7). Thus, cleaning and rearrangement steps were necessary. Venues were assigned to 10 manually defined categories, and the rest with categories that are not listed in Table 10 were deleted. We defined these categories to reflect the different weekly needs of visitors seeking different business locations. Table 7 shows the 10 categories and the number of instances (samples) for each category in the dataset.

TABLE 7. Description of the dataset.

Category	Number of Samples
Cat1: Shop and Services	369
Cat2: Art and Entertainments	73
Cat3: Drink and coffee shops	2113
Cat4: Outdoor events	15
Cat5: Sandwiches and Snacks	181
Cat6: Restaurants	498
Cat7: Hotels	30
Cat8: Gym and Fitness	159
Cat9: Bread and Sweets	95
Cat10: Food Trucks	31
Total	3564

Each instance (sample) consists of the features of a venue for each of the three weeks, and the popularity vector computed based on four timeframes of the fourth week and served as an output vector. The four timeframes used in this study were as follows: Timeframe1 (morning) from 5 AM to 12 PM, Timeframe2 (afternoon) from 12 PM to 6 PM, Timeframe3 (evening) from 6 PM to 12 AM, and Timeframe4 (night) from 12 AM to 5 AM.

In the regression problem, the popularity label for each timeframe was the number of visitors captured in the form of Foursquare check-in, whereas in the classification problem, the number of visitors was transformed into one of the following popularity classes:

- Popular: Assigned to instances when the number of visitors exceeding 75% of the number of visitors in the corresponding timeframe.
- Above average: Assigned to instances when the number of visitors was between 25% and 75% of the number of visitors in the corresponding timeframe.
- Below average: Assigned to instances when the number of visitors was below 25% of different number of visitors in the corresponding timeframe.
- Unpopular: Assigned to the instances when there were no visitors in the corresponding timeframe.

B. BASELINES

The sequence regression model was compared to a baseline model that combined vector autoregression (VAR) and linear regression (LR) [50], [51]. VAR was used to handle the time series nature of the input vector, while LR was used to handle the regression part. Therefore, these two can accept the same sequence of inputs as the RNNs; that is, they take an input vector representing a sequence of three weeks and produce the predicted popularity vector for the following week. VAR is a multivariate autoregressive model that predicts the next event in the time series, while, the LR model represents the linear relationship between a scalar and a response [52], [53]. Similarly, the classification model was compared to the baseline model that combines a VAR model with a DT model. The DT model employs a tree model in which each node represents a feature, each branch represents a decision, and each leaf represents the target class [54]. The structure of the baselines for the sequence regression model and the sequence classification model are shown in Figure 9. The third model that handles the problem of predicting the top-six popular venues on the following day based on the data of the previous three days is compared to the VAR model only.
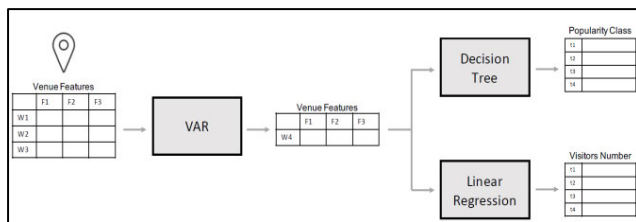

FIGURE 9. Structure of the baseline.

C. EVALUATION

The sequence regression model and city popularity map prediction model were evaluated using the RMSE and mean absolute error (MAE). We also used the normalized version of both measures (NRMSE and NMAE) obtained by dividing the measures with the mean of the values of the observed popularity vector. RMSE and MAE were computed as follows:

$$RMSE = \sqrt{\frac{1}{N} \sum (observed\ popularity\ vector - predicted\ popularity\ vector)^2} \quad (3)$$

$$MAE = \frac{1}{N} \sum |observed\ popularity\ vector - predicted\ popularity\ vector| \quad (4)$$

The classification model performance was evaluated using the accuracy and Matthews correlation coefficient (MCC). MCC is calculated based on number of true positive samples (TP), number of true negative samples(TN), number of false positive samples(FP) and number of false negative samples (FN). Accuracy and MCC were computed as follows:

$$Accuracy = \frac{Number\ of\ correctly\ classified\ samples}{Numbet\ of\ all\ \ samples} \quad (5)$$

$$MCC = \frac{TP * TN - FP * FN}{\sqrt{(TP+FP)*(TP+FN)*(TN+FP)*(TN+FN)}} \quad (6)$$

V. EXPERIMENTAL RESULTS AND DISCUSSION

Three types of datasets were constructed for the regression and classification models. The first dataset contained instances of different categories (MC); the second contained only instances from the same category (1C); and the third contained instances from different categories. The testing set contained instances from categories that were not in the training set (M1C). For the city-prediction map model, we collected data for a 24-week period with a total of 165 instances and divided them into three datasets: (i) popular venues for restaurants only, (ii) popular venues of all the categories, and (iii) an extension of the second dataset but with the input and output vectors that contain the location of the top-six popular venues as well as their categories. We used 80% of the samples for training and 20% for testing; for validation, we used 20% of the training instances. Details of the different dataset splits are listed in Table 8.

TABLE 8. Details of different dataset splits used in the study.

Dataset	Type	Categories	Samples
Validation dataset	Validation dataset	Cat10, Cat9, Cat7, Cat4, Cat5	352
MC	Training set	Validation dataset, Cat2 (Cat1,Cat3,Cat6) - (Testing set)	2805
	Testing set	First 400 samples of Cat3 First 100 samples of Cat6 First 100 samples of Cat1	600
M1C	Training set	Validation dataset, Cat3, Cat6	2480
	Testing set	Cat1, Cat2, Cat8	501
1C	Training set	80% of Cat3	1690
	Testing set	20% of Cat3	423

A. RESULTS OF THE SEQUENCE REGRESSION MODEL

Table 9 lists the results obtained from different models of RNNs: basic RNN, LSTM, and GRU, and the baseline VAR with LR trained on three datasets: MC, 1C, and M1C. The data in the table show that the basic RNN model achieved

TABLE 9. Results obtained using the RNN, LSTM, GRU and VAR + LR sequence regression models with MC, 1C, and M1C datasets.

Dataset	Model	MAE	RMSE	NRMSE	NRMAE
MC	RNN	1.5846	2.8600	1.6179	0.8963
	LSTM	1.7432	3.1917	1.8053	0.9860
	GRU	1.6378	2.8650	1.6206	0.9264
	VAR+LR	1.6220	2.9530	1.6703	0.9174
1C	RNN	1.3206	2.1855	1.5780	0.9537
	LSTM	1.3649	2.2179	1.6016	0.9857
	GRU	1.3810	2.2568	1.6298	0.9973
	VAR+LR	1.4368	2.2901	1.6538	1.0376
M1C	RNN	1.5286	3.2949	1.9114	0.8867
	LSTM	1.4697	3.0743	1.7834	0.8526
	GRU	1.4757	3.0961	1.7961	0.8560
	VAR+LR	1.8355	6.6572	3.8620	1.0648

TABLE 10. Results obtained using the RNN, LSTM, GRU, and VAR + DT sequence classification models with MC, 1C, and M1C datasets.

Data	Model	Acc1	Acc2	Acc3	Acc4	MCC1	MCC2	MCC3	MCC4	Avg. Accuracy	Avg. MCC
MC	RNN	54.67	65.6	67.17	88.83	0.0161	0	0.0558	0	69.0675	0.0172
	LSTM	54.67	65.5	67.17	88.83	0.0279	0	0.0558	0	69.0425	0.0209
	GRU	54.4	65.5	67	88.83	0.0089	0	0.0262	0	68.9325	0.0088
	VAR+DT	42.17	50.34	50.17	77.67	0.0749	0.0573	0.0084	0.0014	55.0875	0.0355
1C	RNN	54.61	74.7	64.78	85.34	0	0	0	0	69.8575	0
	LSTM	54.61	74.7	64.78	85.34	0	0	0	0	69.8575	0
	GRU	54.61	74.7	64.78	85.34	0	0	0	0	69.8575	0
	VAR+DT	44.2	53.66	53.66	80.85	0.0876	-0.0545	0.0839	0.02	58.0925	0.0343
M1C	RNN	54.58	59.9	57.9	85.19	0.0244	0	0	0	64.3925	0.0061
	LSTM	53.74	59.9	57.9	85.19	-0.0177	0	0	0	64.1825	-0.0044
	GRU	54.74	59.9	57.9	85.19	0.0065	0	0	0	64.4325	0.0016
	VAR+DT	43.93	49.08	48.25	75.37	0.0471	0.0027	0.0442	-0.0154	54.1575	0.01965

the best results for the multiple- and one category datasets (MC and 1C). For different categories in the testing set (M1C), the LSTM yielded better results than the basic RNN and GRU. Overall, RNNs outperformed the baseline models according to the NRMSE measure. Although the basic RNN yields better results for MC and 1C datasets and LSTM had less error for M1C dataset, GRU was the most stable model for all three datasets MC, 1C, and M1C.

B. RESULTS OF THE SEQUENCE CLASSIFICATION MODEL

The RNN models were designed to predict the popularity class for each timeframe (morning, afternoon, evening, and night). The experimental results showed different performance results for each timeframe (see Table 10). For all datasets (MC, 1C, and M1C), our results reveal that the models achieved comparable poor results. This is because low MCC values indicated that the achieved decent accuracy values are almost based on random prediction. However, the baseline models achieved better results with respect to MCC. Here, note that the classification accuracy of all models degraded when using M1C dataset rather than the MC and 1C; however, based on the MCC, the MC dataset showed less indication of classification randomness. Thus, the MC dataset is a better choice for this sequence classification problem.

C. RESULTS OF THE CITY POPULARITY MAP PREDICTION MODEL

When using the dataset constructed for *Restaurant* category only (1C), the error was small in all of the proposed models; however, the baseline model VAR achieved the best results followed by GRU (see Table 11). With regard to the MC dataset, the error was also relatively small in all the proposed models; here too, the baseline, VAR, achieved the best results followed by GRU. With a dataset constructed from multiple categories that predict the location of the popular venues along with their categories (M1C), the results were the same as those for the previous two settings. When predicting the popular venue locations in the city, the baseline outperformed most of the RNN models with respect to all measures except for the RMSE and NRMSE measures of the 1C dataset.

D. DISCUSSION

The results show that the performance of the venue-popularity prediction models was better when the problem was considered as a regression problem instead of a classification problem. This difference can be attributed to the weak correlation between the features and number of visitors, which has a greater effect on the learning process in the classification problem than in the regression problem. The results

TABLE 11. City popularity map prediction results obtained using the RNN, LSTM, GRU, and VAR models with 1C, MC dataset, and MC dataset with category prediction.

Dataset	Model	MAE	RMSE	NRMSE	NRMAE
MC	RNN	3.0811	3.2123	0.08996	0.0863
	LSTM	2.4232	2.5984	0.0728	0.0679
	GRU	1.3676	1.4589	0.0409	0.0383
	VAR	0.0422	0.0539	0.0015	0.0012
1C	RNN	1.5325	1.8813	0.0744	0.0606
	LSTM	2.4656	2.8029	0.1108	0.0975
	GRU	1.2113	1.7271	0.0683	0.0479
	VAR	0.8637	1.8901	0.0747	0.0341
M1C	RNN	1.8105	1.9308	0.0541	0.0507
	LSTM	2.9571	3.2072	0.0898	0.0828
	GRU	1.9173	2.0465	0.0573	0.0537
	VAR	0.0350	0.0469	0.0013	0.0010
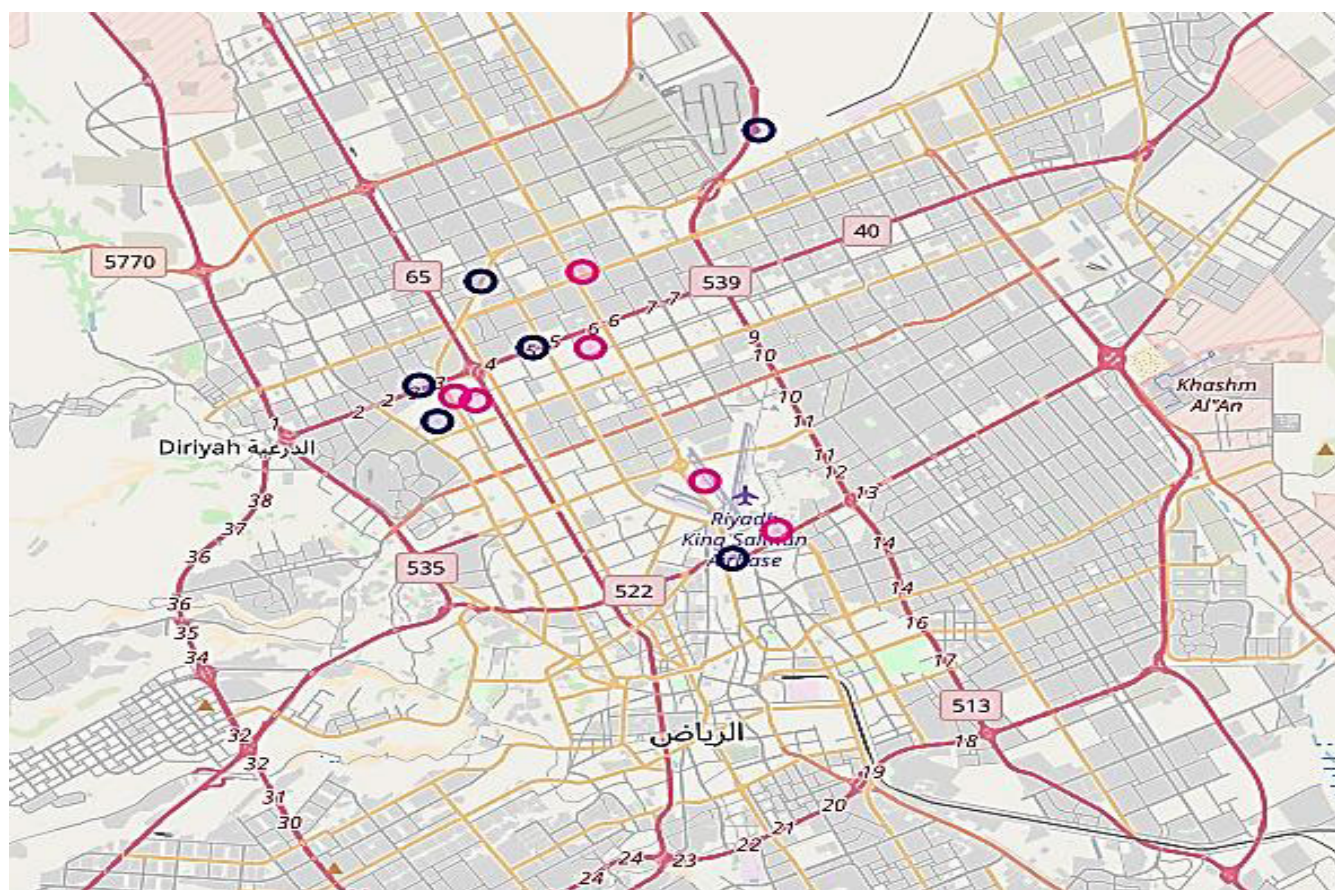
FIGURE 10. Popular locations: Observed and predicted using VAR.

can also be attributed to the sparsity of labels, which require more data for sufficient learning by RNN models; therefore, modeling the problem as a classification problem introduced a higher level of difficulty. In the regression experiments, RNN achieved the best results, indicating that the basic RNN can handle the problem of popularity prediction when modeled as a sequence regression problem.

Overall, our results show that modeling the popularity prediction as sequence regression is perhaps the best modeling approach for the venue-popularity prediction problem. Most of the models performed better with the dataset containing instances from the same category, but when used with a dataset with multiple categories in the training and testing sets, the performance was degraded. The most challenging

FIGURE 11. Popular locations: Observed and predicted using GRU.

dataset to learn from was the M1C dataset, where different categories were used in the training and testing sets.

From the viewpoint of complexity, the three datasets can be ranked as follows: 1C being the least complex and M1C, the most complex. The GRU model had the highest stability in performance in terms of the least degradation in performance with the variant-complexity datasets; however, this model did not always achieve the optimum results.

For city popularity map prediction, compared to the RNN-family models, VAR achieved better results with the least errors. The dataset for this problem may have been insufficient for the RNNs to learn as the data were collected over a period of 6 months and yielded only 165 samples. Constructing a larger dataset means extending the window of time. However, because of the dynamics of the city, extending the time period can increase the complexity of the patterns in the dataset. Thus, we can argue that for the present problem, a simpler model such as VAR is a better choice over deep-learning methods (i.e., RNN in this study).

The effect of the error magnitude in such a real application is noticeable. As the model predicts the latitude and longitude of a popular location, a small error in the prediction of these two values leads to a significantly distant location. For instance, an error of 0.0010 can predict a location that is around 100 m away from the correct one, and this example

becomes more evident when we look at the real predicted maps shown in Figures 10 and 11. In Figure 10, the observed popular locations are indicated in blue and the popular locations predicted using VAR with 1C dataset are in pink. In Figure 11, the observed popular locations and those predicted using GRU with 1C are in navy and pink, respectively. The distance between the observed and predicted locations seen in Figure 11 originated from the small error in the GRU model, indicating that even minor errors cannot be tolerated in such a location prediction problem.

VI. CONCLUSION AND FUTURE WORK
Social data are considered a good source of PSS that aims to sense different aspects of city dynamics. Social data can be harnessed to analyze city dynamics in terms of its mobility patterns, social patterns, event detection, and collective human behavior and have motivated many researchers to develop intelligent applications for smart cities, such as urban planning, transportation systems, city environment, energy consumption, public safety and security, and city economy. Predicting the potential number of visitors to a specific venue by analyzing the mobility patterns can have positive implications for the city economy. In this study, we investigated different RNN-family models, including the basic RNN, LSTM, and GRU models, in addition to some baseline models,

for venue-popularity prediction. To evaluate the models, we constructed different datasets for Riyadh city using Twitter and Foursquare data. Ten urban features were mined from these social data, and the best features based on their correlation with the number of visitors to different venues were selected. Three sequence-based popularity prediction models were developed and evaluated using data sets with different categories. RNNs achieved encouraging performance for predicting sequences of events as regression models. Our experimental works show that modeling the venue-popularity prediction problem as a regression model achieves better results than modeling it as a classification problem. The basic RNN yielded the best results for sequence regression with MC and 1C datasets; however, GRU provided the most stable performance across different dataset categories.

With regard to the city-popularity-map prediction problem, the basic VAR achieved a better result than the RNNs. Because the outcomes of this research will benefit city planners, business owners, campaign planners, traffic systems, and advertisement companies to make more informed decisions and plans, our future works will include extending the datasets to other cities and to wider timeframes. Another research direction is to use the datasets and feature mining for one category of venues and experiment with different deep-learning models. Sophisticated models of LSTM are suggested for future works such as using a sequence of LSTM units each for a specific purpose such as popular venue, peak hour, days to visit, and so on.

REFERENCES

[1] M. Becker, S. Caminiti, D. Fiorella, L. Francis, P. Gravino, M. Haklay, A. Hotho, V. Loreto, J. Mueller, F. Ricchiuti, and V. D. P. Servedio, "Awareness and learning in participatory noise sensing," *PLoS ONE*, vol. 8, no. 12, Dec. 2013, Art. no. e81638, doi: 10.1371/journal.pone.0081638.

[2] T. H. Silva, P. O. S. V. de Melo, J. M. Almeida, and A. A. F. Loureiro, *Social Media as a Source of Sensing to Study City Dynamics and Urban Social Behavior: Approaches, Models, and Opportunities* (Lecture Notes in Computer Science), M. Atzmueller, A. Chin, D. Helic, and A. Hotho, Eds. Berlin, Germany: Springer, 2012, pp. 63–87.

[3] Y.-T. Zheng, Z.-J. Zha, and T.-S. Chua, "Mining travel patterns from geotagged photos," *ACM Trans. Intell. Syst. Technol.*, vol. 3, no. 3, pp. 56:1–56:18. May 2012, doi: 10.1145/2168752.2168770.

[4] Z. Cheng, J. Caverlee, K. Lee, and D. Z. Sui, "Exploring millions of footprints in location sharing services," in *Proc. 5th Int. AAAI Conf. Weblogs Social Media*, Jul. 2011, pp. 81–88.

[5] R. Ji, X. Xie, H. Yao, and W.-Y. Ma, "Mining city landmarks from blogs by graph modeling," in *Proc. 17th ACM Int. Conf. Multimedia (MM)*. Beijing, China: ACM, Oct. 2009, pp. 105–114, doi: 10.1145/1631272.1631289.

[6] J. Cranshaw, E. Toch, J. Hong, A. Kittur, and N. Sadeh, "Bridging the gap between physical location and online social networks," in *Proc. 12th ACM Int. Conf. Ubiquitous Comput.* Copenhagen, Denmark: ACM, Sep. 2010, pp. 119–128, doi: 10.1145/1864349.1864380.

[7] K. Joseph, C. H. Tan, and K. M. Carley, "Beyond 'local' 'categories' and 'friends': Clustering foursquare users with latent 'topics,'" in *Proc. ACM Conf. Ubiquitous Comput. (UbiComp)*, vol. 8. Pittsburgh, PA, USA: ACM, 2012, pp. 919–926, doi: 10.1145/2370216.2370422.

[8] Y. Zheng, L. Capra, O. Wolfson, and H. Yang, "Urban computing: Concepts, methodologies, and applications," *ACM Trans. Intell. Syst. Technol.*, vol. 5, no. 3, p. 38, Sep. 2014, doi: 10.1145/2629592.

[9] M. De Nadai and B. Lepri, "The economic value of neighborhoods: Predicting real estate prices from the urban environment," in *Proc. IEEE 5th Int. Conf. Data Sci. Adv. Analytics (DSAA)*. Turin, Italy, Italy: IEEE, Oct. 2018, pp. 323–330, doi: 10.1109/DSAA.2018.00043.

[10] D. Karamshuk, A. Noulas, S. Scellato, V. Nicosia, and C. Mascolo, "Geo-spotting: Mining online location-based services for optimal retail store placement," in *Proc. 19th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining (KDD)*. Chicago, IL, USA: ACM, 2013, pp. 793–801, doi: 10.1145/2487575.2487616.

[11] H.-P. Hsieh, C.-T. Li, and S.-D. Lin, "Estimating potential customers anywhere and anytime based on location-based social networks," in *Proc. Joint Eur. Conf. Mach. Learn. Knowl. Discovery Databases* (Lecture Notes in Computer Science), A. Appice, P. P. Rodrigues, V. S. Costa, J. Gama, A. Jorge, and C. Soares, Eds. Cham, Switzerland: Springer, Sep. 2015, pp. 576–592.

[12] K. D'Silva, A. Noulas, M. Musolesi, C. Mascolo, and M. Sklar, "Predicting the temporal activity patterns of new venues," *EPJ Data Sci.*, vol. 7, no. 1, p. 13, Dec. 2018, doi: 10.1140/epjds/s13688-018-0142-z.

[13] F. Wang, L. Chen, and W. Pan, "Where to place your next restaurant?: Optimal restaurant placement via leveraging user-generated reviews," in *Proc. 25th ACM Int. Conf. Inf. Knowl. Manage.*, Indianapolis, IN, USA: ACM, Oct. 2016, pp. 2371–2376, doi: 10.1145/2983323.2983696.

[14] M. K. Rahman and M. A. Nayeem, "Finding suitable places for live campaigns using location-based services," in *Proc. 4th Int. ACM Workshop Manag. Mining Enriched Geo-Spatial Data (GeoRich)*. Chicago, IL: ACM, 2017, pp. 1–6, doi: 10.1145/3080546.3080630.

[15] H. Barbosa, M. Barthelemy, G. Ghoshal, C. R. James, M. Lenormand, T. Louail, R. Menezes, J. J. Ramasco, F. Simini, and M. Tomasini, "Human mobility: Models and applications," *Phys. Rep.*, vol. 734, pp. 1–74, Mar. 2018, doi: 10.1016/j.physrep.2018.01.001.v.

[16] Y. Zhuang, S. Fong, M. Yuan, Y. Sung, K. Cho, and R. K. Wong, "Location-based big data analytics for guessing the next foursquare check-ins," *J. Supercomput.*, vol. 73, no. 7, pp. 3112–3127, Jul. 2017, doi: 10.1007/s11227-016-1925-2.

[17] C. Bothorel, N. Lathia, R. Picot-Clemente, and A. Noulas, "Location recommendation with social media data," in *Social Information Access* (Lecture Notes in Computer Science), P. Brusilovsky and D. He, Eds. Cham, Switzerland: Springer, 2018, pp. 624–653.

[18] E. Cho, S. A. Myers, and J. Leskovec, "Friendship and mobility: User movement in location-based social networks," in *Proc. 17th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining (KDD)*. San Diego, CA, USA: ACM, Aug. 2011, pp. 1082–1090, doi: 10.1145/2020408.2020579.

[19] W. Luo, P. Gao, and S. Cassels, "A large-scale location-based social network to understanding the impact of human geo-social interaction patterns on vaccination strategies in an urbanized area," *Comput., Environ. Urban Syst.*, vol. 72, pp. 78–87, Nov. 2018, doi: 10.1016/j.compenvurbsys.2018.06.008.

[20] X. Wang, M. Chen, Z. Han, D. O. Wu, and T. T. Kwon, "TOSS: Traffic offloading by social network service-based opportunistic sharing in mobile social networks," in *Proc. IEEE Conf. Comput. Commun. (INFO-COM)*. Toronto, ON, Canada: IEEE, Apr. 2014, pp. 2346–2354, doi: 10.1109/INFOCOM.2014.6848179.

[21] J. Lian, F. Zhang, X. Xie, and G. Sun, "Restaurant survival analysis with heterogeneous information," in *Proc. 26th Int. Conf. World Wide Web Companion (WWW Companion)*. Perth, WA, Australia: International World Wide Web, 2017, pp. 993–1002, doi: 10.1145/3041021.3055130.

[22] Y. Fu, Y. Ge, Y. Zheng, Z. Yao, Y. Liu, H. Xiong, and J. Yuan, "Sparse real estate ranking with online user reviews and offline moving behaviors," in *Proc. IEEE Int. Conf. Data Mining*. Shenzhen, China: IEEE, Dec. 2014, pp. 120–129, doi: 10.1109/ICDM.2014.18.

[23] M. Shahbazi, J. R. Barr, V. Hristidis, and N. N. Srinivasan, "Estimation of the investability of real estate properties through text analysis," in *Proc. IEEE 10th Int. Conf. Semantic Comput. (ICSC)*. Laguna Hills, CA, USA: IEEE, Feb. 2016, pp. 301–306, doi: 10.1109/ICSC.2016.85.

[24] S. Zhao, I. King, M. R. Lyu, J. Zeng, and M. Yuan, "Mining business opportunities from location-based social networks," in *Proc. 40th Int. ACM SIGIR Conf. Res. Develop. Inf. Retr.* Shinjuku, Tokyo, Japan: ACM, Aug. 2017, pp. 1037–1040, doi: 10.1145/3077136.3080712.

[25] O. AlSonosy, S. Rady, N. Badr, and M. Hashem, "Business behavior predictions using location based social networks in smart cities," in *Information Innovation Technology in Smart Cities*, L. Ismail and L. Zhang, Eds. Singapore: Springer, 2018, pp. 105–122.

[26] J. Lin, R. Oentaryo, E.-P. Lim, C. Vu, A. Vu, and A. Kwee, "Where is the goldmine?: Finding promising business locations through facebook data analytics," in *Proc. 27th ACM Conf. Hypertext Social Media*. Halifax, NS, Canada: ACM, Jul. 2016, pp. 93–102, doi: 10.1145/2914586.2914588.

[27] A. K. Bhowmick, S. Suman, and B. Mitra, "Effect of information propagation on business popularity: A case study on yelp," in *Proc. 18th IEEE Int. Conf. Mobile Data Manage. (MDM)*. Daejeon, South Korea: IEEE, May 2017, pp. 11–20, doi: 10.1109/MDM.2017.13.

[28] S. Zhang and Q. Lv, "Event organization 101: Understanding latent factors of event popularity," in *Proc. 11th Int. AAAI Conf. Web Social Media*, May 2017, pp. 1–5.

[29] D. Liu, D. Weng, Y. Li, J. Bao, Y. Zheng, H. Qu, and Y. Wu, "SmartAdP: Visual analytics of large-scale taxi trajectories for selecting billboard locations," *IEEE Trans. Vis. Comput. Graphics*, vol. 23, no. 1, pp. 1–10, Jan. 2017, doi: 10.1109/TVCG.2016.2598432.

[30] R. M. Sánchez-Saiz, J. M. Galán, and J. I. Santos, *Localization Based on Business Interactions Through a Simulated Annealing Algorithm* (Lecture Notes in Management and Industrial Engineering), C. Hernández, A. López-Paredes, and J. M. Pérez-Ríos, Eds. Cham, Switzerland: Springer, 2014, pp. 325–331.

[31] P. Jensen, "Network-based predictions of retail store commercial categories and optimal locations," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 74, no. 3, Sep. 2006, Art. no. 035101, doi: 10.1103/PhysRevE.74.035101.

[32] M. Xu, T. Wang, Z. Wu, J. Zhou, J. Li, and H. Wu, "Demand driven store site selection via multiple spatial-temporal data," in *Proc. 24th ACM SIGSPATIAL Int. Conf. Adv. Geographic Inf. Syst.* Burlingame, CA, USA: ACM, Oct. 2016, pp. 1–10, doi: 10.1145/2996913.2996996.

[33] A. M. B. M. Rohani and F.-F. Chua, "Location analytics for optimal business retail site selection," in *Computational Science and Its Applications—ICCSA*, in Lecture Notes in Computer Science, O. Gervasi, Ed. Cham, Switzerland: Springer, 2018, pp. 392–405.

[34] J. Chen, C. Yu, and H. Jin, "Evaluation model for business sites planning based on online and offline datasets," *Future Gener. Comput. Syst.*, vol. 91, pp. 465–474, Feb. 2019, doi: 10.1016/j.future.2018.08.024.

[35] N. Zhang, S. Deng, H. Chen, X. Chen, J. Chen, X. Li, and Y. Zhang, "Structured knowledge base as prior knowledge to improve urban data analysis," *ISPRS Int. J. Geo-Inf.*, vol. 7, no. 7, p. 264, Jul. 2018, doi: 10.3390/ijgi7070264.

[36] L. Deng, "Deep learning: Methods and applications," *Found. Trends Signal Process.*, vol. 7, nos. 3–4, pp. 197–387, Jun. 2014, doi: 10.1561/2000000039.

[37] A. Graves, "Supervised sequence labelling with recurrent neural networks," Ph.D. dissertation, Dept. Comput. Sci., Tech. Univ. Munich, Munich, Germany, 2008.

[38] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997, doi: 10.1162/neco.1997.9.8.1735.

[39] F. A. Gers, J. Schmidhuber, and F. Cummins, "Learning to forget: Continual prediction with LSTM," *Neural Comput.*, vol. 12, no. 10, pp. 2451–2471, Oct. 2000, doi: 10.1162/089976600300015015.

[40] K. Cho, B. van Merrienboer, C. Gulcehre, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using RNN encoder-decoder for statistical machine translation," in *Proc. Conf. Empirical Methods Natural Lang. Process. (EMNLP)*, 2014, pp. 1724–1734.

[41] A. M. Rather, A. Agarwal, and V. N. Sastry, "Recurrent neural network and a hybrid model for prediction of stock returns," *Expert Syst. Appl.*, vol. 42, no. 6, pp. 3234–3241, Apr. 2015, doi: 10.1016/j.eswa.2014.12.003.

[42] J. Wang and J. Wang, "Forecasting energy market indices with recurrent neural networks: Case study of crude oil price fluctuations," *Energy*, vol. 102, pp. 365–374, May 2016, doi: 10.1016/j.energy.2016.02.098.

[43] A. Milan, S. H. Rezatofighi, A. Dick, I. Reid, and K. Schindler, "Online multi-target tracking using recurrent neural networks," in *Proc. 31st AAAI Conf. Artif. Intell.*, Feb. 2017, pp. 1–9.

[44] E. Palumbo, G. Rizzo, R. Troncy, and E. Baralis, "Predicting your next stop-over from location-based social network data with recurrent neural networks," in *Proc. 2nd ACM Int. Workshop Recommenders Tourism*, Como, Italy, Aug. 2017, pp. 1–8.

[45] A. Alahi, K. Goel, V. Ramanathan, A. Robicquet, L. Fei-Fei, and S. Savarese, "Social LSTM: Human trajectory prediction in crowded spaces," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Las Vegas, NV, USA, Jun. 2016, pp. 961–971.

[46] Z. Lin, M. Yin, S. Feygin, M. S. Transportation, J.-F. Paiement, and A. P. Cee, "Deep generative models of urban mobility," *IEEE Trans. Intell. Transp. Syst.*, unpublished.

[47] G. Hu, Y. Qin, and J. Shao, "Personalized travel route recommendation from multi-source social media data," *Multimedia Tools Appl.*, vol. 79, nos. 45–46, pp. 33365–33380, Dec. 2020, doi: 10.1007/s11042-018-6776-9.

[48] G. Chandrashekar and F. Sahin, "A survey on feature selection methods," *Comput. Electr. Eng.*, vol. 40, no. 1, pp. 16–28, Jan. 2014, doi: 10.1016/j.compeleceng.2013.11.024.

[49] L. A. Rastrigin, "The convergence of the random search method in the extremal control of a many parameter system," *Automaton Remote Control*, vol. 24, pp. 1337–1342, 1963.

[50] S. Weisberg, *Applied Linear Regression*. Hoboken, NJ, USA: Wiley, 2005.

[51] S. Das, *Time Series Analysis*. Princeton, NJ, USA: Princeton Univ. Press, 1994.

[52] A. Hatemi-J, "Multivariate tests for autocorrelation in the stable and unstable VAR models," *Econ. Model.*, vol. 21, no. 4, pp. 661–683, Jul. 2004, doi: 10.1016/j.econmod.2003.09.005.

[53] D. A. Freedman, *Statistical Models: Theory and Practice*. Cambridge, U.K.: Cambridge Univ. Press, 2009, p. 459.

[54] S. R. Safavian and D. Landgrebe, "A survey of decision tree classifier methodology," *IEEE Trans. Syst., Man, Cybern.*, vol. 21, no. 3, pp. 660–674, May 1991, doi: 10.1109/21.97458.

KHULUD ALHARTHI received the B.Sc. degree (Hons.) in computer science from Taif University, Saudi Arabia, in 2010, and the M.Sc. degree (Hons.) in computer science from King Saud University in 2019. She is currently pursuing the Ph.D. degree in swarm robotics with the Faculty of Engineering, University of Bristol. In 2012, she joined the College of Computers and Information Technology, Taif University, as a Faculty Member. Her research interests include in the area of machine learning, data mining, robotics, machine learning, and collective intelligence for swarm robotics in medical and urban applications.

KHALIL EL HINDI is currently a Professor with the Department of Computer Science, King Saud University. His main research interests include machine learning, classification algorithms, outlier detection, instance weighing, ensembles of classifiers, similarity distance metrics, and neural networks.

SALHA M. ALZAHRANI graduated from high School (Science). She received the B.Sc. degree (Hons.) in computer science with from Taif University in 2004 and the M.Sc. degree (Hons.) and the Ph.D. degree in computer science from the University of Technology Malaysia in 2009 and 2012, respectively. She was a Ph.D. Recognized Student at the University of Oxford, U.K., in 2011. She is currently an Associate Professor of computer science, the Vice Dean of graduate studies, and the former Vice-Dean with the College of Computers and Information Technology, Taif University. Her contributions to the field of data science are illustrated by indexed publications in high-impact journals and peer-reviewed conferences. Her research interests include the Internet of Things, big data, data mining, deep learning, plagiarism detection, computational linguistics, natural language processing, and soft information retrieval. She was awarded as the First Rank Student in Taif. She was a recipient of the University Chancellor's Award for the top ten students in the University of Technology Malaysia in 2009. She was also serving as the Web Activities Coordinator for the IEEE SMC Technical Committee on Soft Computing and coorganizing special sessions for IEEE SMC. Her total citations received are 669 till 2020 and some have been published in impact-factor journals.

● ● ●