IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received September 9, 2020, accepted September 28, 2020, date of publication October 2, 2020, date of current version October 14, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3028420

# Comparison of Machine Learning Algorithms for Predicting Crime Hotspots

XU ZHANG[1,2], LIN LIU[2,3], LUZI XIAO[2], AND JIAKAI JI[4]

[1]School of Computer Sciences and Cyber Engineering, Guangzhou University, Guangzhou 510006, China
[2]Center of Geoinformatics for Public Security, School of Geographic Sciences, Guangzhou University, Guangzhou 510006, China
[3]Department of Geography, University of Cincinnati, Cincinnati, OH 45221-0131, USA
[4]School of Geography and Planning, Sun Yat-sen University, Guangzhou 510275, China

Corresponding author: Lin Liu (lin.liu@uc.edu)

This work was supported in part by the Research Team Program of Natural Science Foundation of Guangdong Province, China, under Grant 2014A030312010, in part by the National Key Research and Development Program of China under Grant 2018YFB0505500 and Grant 2018YFB0505503, in part by the Key Program of National Natural Science Foundation of China under Grant 41531178, and in part by the Key Project of Science and in part by Technology Program of Guangzhou City, China, under Grant 201804020016.

**ABSTRACT** Crime prediction is of great significance to the formulation of policing strategies and the implementation of crime prevention and control. Machine learning is the current mainstream prediction method. However, few studies have systematically compared different machine learning methods for crime prediction. This paper takes the historical data of public property crime from 2015 to 2018 from a section of a large coastal city in the southeast of China as research data to assess the predictive power between several machine learning algorithms. Results based on the historical crime data alone suggest that the LSTM model outperformed KNN, random forest, support vector machine, naive Bayes, and convolutional neural networks. In addition, the built environment data of points of interests (POIs) and urban road network density are input into LSTM model as covariates. It is found that the model with built environment covariates has better prediction effect compared with the original model that is based on historical crime data alone. Therefore, future crime prediction should take advantage of both historical crime data and covariates associated with criminological theories. Not all machine learning algorithms are equally effective in crime prediction.

**INDEX TERMS** Prediction of crime hotspots, machine learning, LSTM, built environment.

## I. INTRODUCTION

Spatiotemporal data related to the public security have been growing at an exponential rate during the recent years. However, not all data have been effectively used to tackle real-world problems. In order to facilitate crime prevention, several scholars have developed models to predict crime [1]. Most used historical crime data alone to calibrate the predictive models.

The research on crime prediction currently focuses on two major aspects: crime risk area prediction [2], [3] and crime hotspot prediction [4], [5]. The crime risk area prediction, based on the relevant influencing factors of criminal activities, refers to the correlation between criminal activities and physical environment, which both derived from the ''routine activity theory'' [6]. Traditional crime risk estimation methods usually detect crime hotspots from the historical distribution of crime cases, and assume that the pattern will persist in the following time periods [7]. For example, considering the proximity of crime places and the aggregation of crime elements, the terrain risk model tends to use crime-related environmental factors and crime history data, and is relatively effective for long-term, stable crime hotspot prediction [2]. Many studies have carried out empirical research on crime prediction in different time periods, combining demographic and economic statistics data, land use data, mobile phone data and crime history data. Crime hotspot prediction aims to predict the likely location of future crime events and hotspots where the future events would concentrate [8]. A commonly used method is kernel density estimation [9]–[12]. A model that considers temporal or spatial autocorrelations of past events performs better than those that fail to account for the autocorrelation [13]. Recently machine learning algorithms have gained popularity. The most popular methods include K-Nearest Neighbor(KNN), random forest algorithm, support vector machine (SVM), neural network

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

and Bayesian model etc. [6]. Some compared the linear methods of crime trend prediction [14], some compared Bayesian model and BP neural network [15], [16], and others compared the spatiotemporal kernel density method with the random forest method in different periods of crime prediction [12].

Among these algorithms, KNN is an efficient supervised learning method algorithm [17], [18]. SVM is a popular machine learning model because it can not only implement classification and regression tasks, but also detect outliers [4], [19]. Random forest algorithm has been proven to have strong non-linear relational data processing ability and high prediction accuracy in multiple fields [20]–[23]. Naive Bayes (NB) is a classical classification algorithm, which has only a few parameters and it is not sensitive to missing data [15], [24]. Convolutional neural networks (CNN) has strong expansibility, and can enhance its expression ability with a very deep layer to deal with more complex classification problems [25], [26]. Long Short-Term Memory (LSTM) neural network extracts time-series features from features, and has a significant effect on processing data with strong time series trends [27]–[29]. This paper will focus on the comparison of the above six machine learning algorithms, and recommend the best performing one to demonstrate the predictive power with and without the use of covariates.

## II. RELATED WORK

### A. PRINCIPLES OF THEORETICAL CRIMINOLOGY IN PREDICTION OF CRIME HOTSPOTS

The focus of crime hotspot prediction is to forecast future concentration of criminal events in a geographical space. Theoretical criminology provides the necessary theoretical basis. Specifically, several related criminological theories not only provide guidance for us to understand the important influence of location factors in the formation and aggregation of criminal events, but also provide a basic mechanism for the police to use information of crime hot spots for crime prevention or control. It mainly includes routine activity theory, rational choice theory, and crime patterns theory. These three theories are generally considered as the theoretical basis of situational crime prevention.

Routine activity theory [30] was jointly proposed by Cohen and Felson in 1979, and has now been further developed through integration with other theories. This theory believes that the occurrence of most crimes, especially predatory crimes, needs the convergence of the three elements including motivated offenders, suitable targets, and lack of ability to defend in time and space.

Rational choice theory [31] was proposed by Cornish and Clarke. The theory holds that the offender's choices in terms of location, goals, methods be explained by the rational balance of effort, risk and reward.

Crime pattern theory [32] integrates the routine activities theory and the rational choice theory, which more closely explains the spatial distribution of criminal events. People form "cognitive map" and "activity space" through daily activities. At the same time, potential offenders also need to use their cognitive maps and choose specific locations for crimes in a relatively familiar space. When committing a crime, the offender tends to avoid those places they don't know but to choose the places where the "criminal opportunity overlaps with cognitive space" based on their rational ability. The reason why these places become crime hotspots is that they have the obvious characteristics of "producing" or "attracting" crime. Therefore, the environmental factors of the places need to be considered besides historical crime data for the prediction of crime hotspots.

### B. BUILT ENVIRONMENT DATA

At present, a large number of studies show that the urban built environment has a significant impact on urban criminal behavior, through the impact of crime opportunities to reduce and prevent crime. In the 2007 Global Habitat Report, it was pointed out that the elements of the built environment have an important impact on the occurrence of criminal acts [33]. Point of interests (POIs) data and road network density data are considered as covariates in the crime prediction model.

#### 1) POI DATA

The urban infrastructure data POI includes the location information and attribute information of various urban facilities [34], [35]. Catering facilities, shopping malls and stores are usually located in places with convenient transportation and large flow of people, gathering a large number of different groups of people to generate the targets for the criminals, while entertainment places attract criminals [36]. These POIs are selected as covariates of the prediction model.

#### 2) ROAD NETWORK DENSITY

The conventional definition of road network density refers to total length of roads divided by the size of an areal unit. The area with a denser road network attracts greater flow of people, including potential victims and criminals. Previous studies have shown that the density of road network has an impact on crime rate, especially in public space [37].

### C. CRIME PREDICTION WITH MACHINE LEARNING ALGORITHMS

The traditional methods usually detect the crime hotspot area from the historical distribution of crime cases, and assume that the past pattern is to be repeated in the future [7], [2]. This assumption tends to be reasonable for predicting long-term stable crime hotspots. The commonly used KDE method can effectively identify such stable hotspot areas [10], [11]. The KDE method based on temporal autocorrelation tends to outperform the general KDE method [38] Liu *et al.* Compared the random forest and spatiotemporal KDE method, found that the random forest algorithm is more efficient than the traditional spatiotemporal KDE method in the smaller time scale and grid space unit [12] Gabriel *et al.* used the Gated Localized Diffusion Network for crime prediction at the street segment level [39]. Compared with the traditional

Network-time KDE method, the diffusion network approach significantly increased the prediction accuracy. The ability of machine learning algorithm in processing non-linear relational data has been confirmed in many fields, including crime prediction. It has a faster training speed, can handle very high-dimensional data, and can also extract the characteristics of the data.

## III. PREDICTION MODEL

In this paper, random forest algorithm, KNN algorithm, SVM algorithm and LSTM algorithm are used for crime prediction. First, historical crime data alone are used as input to calibrate the models. Comparison would identify the most effective model. Second, built environment data such as road network density and poi are added to the predictive model as covariates, to see if prediction accuracy can be further improved.

### A. KNN

KNN, also known as k-nearest neighbor, takes the feature vector of the instance as the input, calculates the distance between the training set and the new data feature value, and then selects the nearest K classification. If k = 1, the nearest neighbor class is the data to be tested. KNN's classification decision rule is majority voting or weighted voting based on distance. The majority of k neighboring training instances of the input instance determines the category of the input instance.

### B. RANDOM FOREST

The random forest is a set of tree classifiers {h(x, $\beta$k),k = 1...}, in which the meta classifier h(x, $\beta$k) is an uncut regression tree constructed by CART algorithm; x is the input vector; $\beta$k is an independent random vector with the same distribution, and the output of the forest is obtained by voting. The randomness of random forest is reflected in two aspects: one is to randomly select the training sample set by using bagging algorithm; the other is to randomly select the split attribute set. Assuming that the training sample has M attributes in total, we specify an attribute number F $\leq$ M, in each internal node, randomly select F attributes from M attributes as the split attribute set, and take the best split mode of the f attributes Split the nodes. The multi decision tree is made up of random forest, and the final classification result is determined by the vote of tree classifier.

### C. SVM

SVM, based on statistical learning theory, is a data mining method that can deal with many problems such as regression (time series analysis) and pattern recognition (classification problem, discriminant analysis) very successfully. The mechanism of SVM is to find a superior classification hyperplane that meets the classification requirements, so that the hyperplane can ensure the classification accuracy and can maximize the blank area on both sides of the hyperplane. In theory, SVM can realize the optimal classification of linear separable data.

### D. NB

In the field of probability and statistics, Bayesian theory predicts the occurrence probability of an event based on the knowledge of the evidence of an event. In the field of machine learning, the naïve Bayes (NB) classifier is a classification method based on Bayesian theory and assuming that each feature is independent of each other. In abstract, NB classifier is based on conditional probability, to solve the probability that a given entity belongs to a certain class.

### E. CNN

CNN uses one-dimensional convolution for sequence prediction, which is the convolution sum of discrete sequences. To convolve the sequence, CNN first finds a sequence with a window size of kernel_size, and perform convolution with the original sequence to obtain a new sequence expression. The convolutional network also includes a pooling operation, which is to filter the features extracted by the convolution to get the most useful characteristics.

### F. LSTM

LSTM is a kind of deep neural network based on RNN. The core of LSTM is to add a special unit (memory module) to learn the current information and to extract the related information and rules between the data, so as to transfer the information. LSTM is more suitable for deep neural network calculation because of memory module to slow down information loss. Each memory module has three gates, including input gate ($i_t$), forget gate ($f_t$), and output gate ($o_t$). They are used to selectively memorize the correction parameters of the feedback error function as the gradient decreases. The specific structure is shown in the figure.

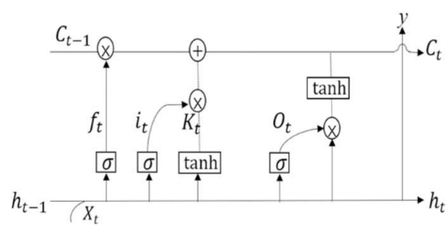



**FIGURE 1. The structure chart of LSTM algorithm.**

In the figure above, LSTM has two state chains h (hidden layer state) and C (cell state) that are passed over time, only cell state C of RNN is transmitted over time. ht-1 is the value of the current time transmitted from the hidden layer at the previous time, Xt is the input value at the current time, Ct-1 is the state value of the LSTM memory cell at the previous time, and Ct is the state value of the memory cell at the current time.

When ht-1 and Xt pass through the forgetting gate, the information to be discarded is calculated. The value of output to the cell state is between 0 and 1, 0 means all forgetting, and 1 means all information is reserved. Forgetting gate $f_t$ is given by the following equation:

$$f_t = \sigma(w_f \cdot [h_{t-1}, x_t] + b_f) \tag{1}$$

where $w$ and $b$ are weight matrix and bias vector in forgetting gate respectively; $\sigma$ is activation function *Sigmoid*.

There are two processes for updating new information into a cell. First, the input gate of *Sigmoid* function is used to calculate the information to be updated, and then a new value $k_t$ created by *tanh* layer is added to the cell state:

$$i_t = \sigma(w_i \cdot [h_{t-1}, x_t] + b_i) \tag{2}$$

$$k_t = \tanh(w_k \cdot [h_{t-1}, x_t] + b_k) \tag{3}$$

The results obtained from equation (2) and equation (3) are multiplied and added to the results obtained from the forgetting gate of the previous time cell state value to obtain the current time cell state value, as follows:

$$C_t = f_t * C_{t-1} + i_i * k_t \tag{4}$$

The final output depends on the cell state. First of all, *Sigmoid* classifies the output results, selects the data to be output, processes the cell state with *tanh* function, and obtains the state value $h_t$ that the hidden layer transfers to the next time. After being processed by sigmoid, $h_t$ can obtain the pre output value y at the current time, as shown in equation (5) - equation (7):

$$O_t = \sigma(w_O \cdot [h_{t-1}, x_t] + b_O) \tag{5}$$

$$h_t = O_t * tanh(C_t) \tag{6}$$

$$y = \sigma(w^{'} h^{t}) \tag{7}$$

## IV. EVALUATION INDICATOR

By comparing the prediction results of different machine learning models before and after adding covariates, the following indicators are used for evaluation. Hit Rate is one of the indexes used to evaluate the accuracy of crime prediction. The Hit Rate mainly includes Grid Hit Rate and Case Hit Rate. Grid Hit Rate $HitR_a$ refers to the ratio between the number of predicted correct hotspot grids and the total number of actual hotspot grids.

$$\mathrm{HitR}_a = \frac{a^*}{A} \tag{8}$$

where $A$ is the total number of actual hotspot grids; and $a$ is the total number of predicted correct hotspot grids.

Case Hit Rate $\mathrm{HitR_n}$ refers to the ratio between the actual number of cases in the forecast correct hot grids and the total number of cases in the study area in this period. The larger the value of $\mathrm{HitR_n}$ is, the more cases are included in the hot grids, and the higher the accuracy of prediction is.

$$\mathrm{HitR}_n = \frac{n}{N} \tag{9}$$

where n is the total number of cases in the study area, and N is the actual number of cases in the forecast hot grids.

In addition to the hit rate indexes, the Prediction Accuracy Index Hit Efficiency Index $\mathrm{HitE_n}$ can also be used to evaluate the prediction effect of the model. For the grids in a certain period, when the number of prediction grids increases, more grids can be covered. When the number of prediction grids is equal to the total number of grids, the value of $\mathrm{HitR_n}$ is 1. At this time, the value of $\mathrm{HitR_n}$ is large, but the prediction effect is not good. Therefore, $\mathrm{HitE_n}$ is needed to measure the effect of the prediction model. The higher the $\mathrm{HitE_n}$ value is, the more cases are covered with fewer prediction grids, and the higher the hit efficiency is.

$$\mathrm{HitE}_n = \frac{HitR_n}{a/A} \tag{10}$$

where a refers to the number of predicted hotspot grids and A refers to the actual number of hotspots.

## V. EXPERIMENTAL AREA AND DATA VISUALIZATION ANALYSIS

### A. EXPERIMENTAL AREA

The area XT selected in this paper is a town in a coastal megacity in Southeast China. The population density of this community is relatively large, with a total area of about 6.5 square kilometers, a total population of about 400000, and a household registration population of only 50000, suggesting that the overwhelming majority of the population domestic migrants or non-local population. The town consists of several large-scale city villages. The complex composition of built environment and population makes it a high crime area.

### B. SELECTION OF CRIME TYPES

The crime of property in public places mainly refers to the crime that takes occupying the property ownership of others as the main purpose in public places. It mainly includes theft, robbery, snatching and other types of embezzlement crimes that completely obtain property against the will of others. It is of great practical significance to choose the public property crime in this town for the prediction of crime hotspots. Accurate crime prediction can help guide the deployment of the local police resources, changing from passive policing to active prevention and control, thus improving local public security.

### C. DATA VISUALIZATION ANALYSIS

The historical crime data used in this paper comes from the police receiving data from 2015 to 2018 in the P-GIS database of the Public Security Bureau of the experimental district. The text coordinate information recorded in the database is extracted, and the case point data within the street range of the study area is extracted after it is located on the map of the study area.

In order to meet the needs of practical police work, the spatial scale of crime hot spot prediction experiments should be as small as possible. According to the calculation formula of gridding processing study area of Griffith *et al.* [40], the study area is divided into 150m $*$ 150m grids according to the investigation of actual police work and the data distribution of case points. Compared with grids with smaller spatial scales, grids divided by 150m will make case points more concentrated in certain grids and reduce the contingency of hotspot grids. Such a division will also reflect the mechanism and distribution of cases better and improve the prediction accuracy and preciseness of the crime hot spots. According to the investigation of the actual police work, 150m is the largest patrol area that a single police officer can cover in a time unit, which can better use the prediction results in crime prevention and control.

#### 1) HOTSPOT GRID PATTERN

After divide the study area XT into 369 grids, the frequency of cases in each grid is counted according to the distribution of 78 two-week historical crimes in 2015-2017. Through K-means clustering method, the optimal number of clusters is determined to be 4, so the grid is divided into four categories: stable high-risk hot grid, high-risk hot grid, occasional hot grid and non-hot grid.

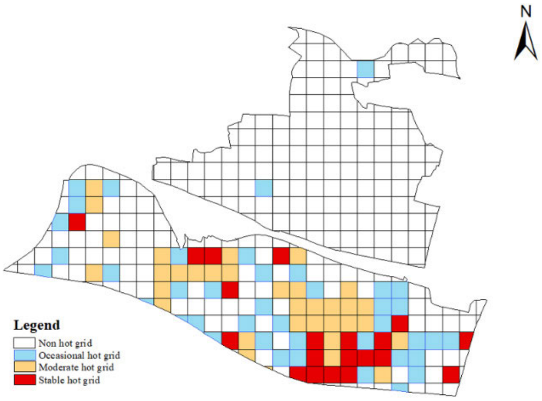



**FIGURE 2.** Hotspot grid pattern of the study area XT.

#### 2) POI DISTRIBUTION

The distribution of POI (catering, shopping malls, and entertainment facilities) in the study area XT is shown in the figure. These types of POI are spatially interpolated to obtain the POI points of the study area and are assigned to each grid as a variable.


**FIGURE 3.** Distribution of POIs of the study area XT.

#### 3) STATISTICS OF CASES BY PERIOD

In terms of the total number of cases, the number of cases in 2018 is slightly less than that in the other three years, and the number of cases in 2017 is slightly more than in 2015 and 2016. During the four-year period, the number of cases in two weeks fluctuated. The number of cases in most two-week periods ranged from 40 to 80, with an average of 58 cases every two weeks. It can be seen from Figure 9 that the case volume curve of the four years has a similar change trend. Basically, the case volume in the two weeks including holidays has a significant reduction, while the case volume in the two weeks after home holidays will pick up. The case volume in January and February of each year has a significant downward trend. The two weeks including spring holidays are the period with the least case volume in each year.

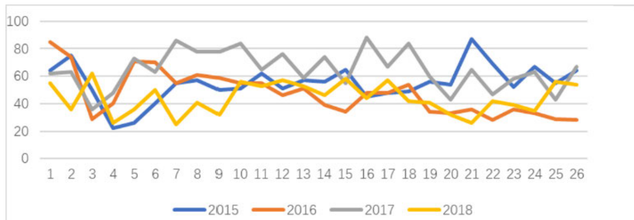



**FIGURE 4.** Statistics of biweekly cases of the study area XT.

#### 4) TIME SERIES ANALYSIS

The figure about decomposition of additive time series shows seasonality in the data, the potential trend and how crime evolved over time in four years. The top part of the figure is the original time series, the 2nd top part of the figure is the estimated trend component, the 3rd part of the figure is the estimated seasonal component and the bottom part of the figure is the estimated irregular component.

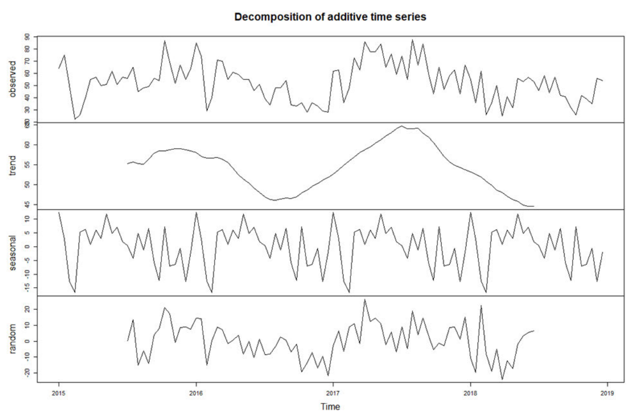



**FIGURE 5.** Decomposition of additive time series of the study area XT.

## VI. EXPERIMENTS

### A. EXPERIMENT IN XT TOWN

According to the research of Rummens [38] and Lin *et al.* [13]y, this paper takes two weeks as the time unit to predict the hot grid of property crime in public places for 13 time units from January 1 to July 1, 2018. The historical data and covariate data are used to forecast the first n hotspots with cases in the forecast period from all grids.

The variable data needed for the prediction model is mainly divided into two parts: one is the historical case data; the other is the covariate data representing the surrounding environment. First, the historical case data is located according to the address and coordinate, and the time period of the point data is divided according to the prediction time unit. We count the number of cases occurred in each grid in each period, take this part of data as the basic data of the prediction model, and select the data of the corresponding period as the training data according to the prediction target period. The second part is covariate data. In the experiment of this paper, city POI density and road network density are used to obtain the density surface of covariate in the study area through spatial interpolation of covariate spatial point data, which is used as covariate of the prediction model.

The historical data of this paper is to count the number of cases per grid in the period from 2015/2016/2017 to the same period as the target period and the four adjacent periods in front of the target period. The covariate data uses the values of two modern city data covariates, POI and road network density. The data is normalized between [0,1] using MinMaxScaler with the transformation function as follows:

$$x = \frac{(\mathrm{x} - \min)}{(\max - \min)} \tag{11}$$

Taking the two weeks from January 1 to January 14, 2018 as the prediction target, the historical data of crime hotspots prediction is divided as shown in the table below.

**TABLE 1.** Historical data division of prediction of crime hotspots from January 1 to January 14, 2018.

| | | |
|---|---|---|
| Training data (2017.12.18-2017.12.31) | 2014.12.18-2014.12.31<br>2015.12.18-2015.12.31<br>2016.12.18-2016.12.31 | 2017.10.23-2017.11.05<br>2017.11.06-2017.11.19<br>2017.11.20-2017.12.03<br>2017.12.04-2017.12.17 |
| Forecasting data (2018.01.01-2018.01.14) | 2015.01.01-2015.01.14<br>2016.01.01-2016.01.14<br>2017.01.01-2017.01.14 | 2017.11.06-2017.11.19<br>2017.11.20-2017.12.03<br>2017.12.04-2017.12.17<br>2017.12.18-2017.12.31 |

The performance of several models is shown in the figure below. Model-a is a KNN prediction model, Model-b is a random forest prediction model, Model-c is an SVM prediction model, model d is an NB prediction model, model e is a CNN prediction model, and model f is a LSTM prediction model.

In prediction experiments in the first half of 2018, consisting of 13 time units, the overall prediction performance of the LSTM model (Model-d) is the best among the four different prediction models (Tables 2 & 3). Taking the LSTM prediction model with covariate data as an example, the average grid hit rate can reach 44.8%, and in this more than half of the predicted correct grids, it can cover an average of 45.8% of cases in the study area. The advantage of LSTM prediction model is not only to memorize the feature information extracted from time series data in short and long term, but also

**TABLE 2.** Experiment results of HitRa based on KNN, RF, SVM, NB, CNN and LSTM models.

| Prediction period | Model-a | Model-b | Model-c | Model-d | Model-e | Model-f |
|---|---|---|---|---|---|---|
| **0101-0114** | 0.152 | 0.175 | 0.221 | 0.243 | 0.295 | 0.575 |
| **0115-0128** | 0.156 | 0.281 | 0.063 | 0.173 | 0.304 | 0.500 |
| **0129-0211** | 0.027 | 0.324 | 0.054 | 0.261 | 0.036 | 0.486 |
| **0212-0225** | 0.450 | 0.368 | 0.150 | 0.285 | 0.386 | 0.450 |
| **0226-0311** | 0.182 | 0.424 | 0.242 | 0.278 | 0.415 | 0.303 |
| **0312-0325** | 0.053 | 0.237 | 0.342 | 0.154 | 0.362 | 0.368 |
| **0326-0408** | 0.368 | 0.316 | 0.421 | 0.384 | 0.453 | 0.684 |
| **0409-0422** | 0.067 | 0.421 | 0.133 | 0.278 | 0.166 | 0.200 |
| **0423-0506** | 0.179 | 0.179 | 0.107 | 0.042 | 0.324 | 0.500 |
| **0507-0520** | 0.085 | 0.319 | 0.085 | 0.153 | 0.267 | 0.170 |
| **0521-0603** | 0.102 | 0.306 | 0.245 | 0.239 | 0.454 | 0.612 |
| **0604-0617** | 0.054 | 0.324 | 0.135 | 0.134 | 0.184 | 0.541 |
| **0618-0701** | 0.051 | 0.231 | 0.077 | 0.141 | 0.278 | 0.436 |

**TABLE 3.** Experiment results of HitRn based on KNN, RF, SVM, NB, CNN and LSTM models.

| Prediction period | Model-a | Model-b | Model-c | Model-d | Model-e | Model-f |
|---|---|---|---|---|---|---|
| **0101-0114** | 0.109 | 0.212 | 0.230 | 0.174 | 0.357 | 0.600 |
| **0115-0128** | 0.196 | 0.353 | 0.078 | 0.217 | 0.382 | 0.500 |
| **0129-0211** | 0.036 | 0.438 | 0.091 | 0.348 | 0.049 | 0.436 |
| **0212-0225** | 0.378 | 0.305 | 0.126 | 0.239 | 0.320 | 0.500 |
| **0226-0311** | 0.182 | 0.583 | 0.250 | 0.278 | 0.571 | 0.250 |
| **0312-0325** | 0.045 | 0.200 | 0.283 | 0.131 | 0.305 | 0.435 |
| **0326-0408** | 0.368 | 0.260 | 0.478 | 0.384 | 0.373 | 0.696 |
| **0409-0422** | 0.067 | 0.281 | 0.384 | 0.278 | 0.111 | 0.167 |
| **0423-0506** | 0.179 | 0.179 | 0.158 | 0.042 | 0.324 | 0.553 |
| **0507-0520** | 0.071 | 0.269 | 0.071 | 0.128 | 0.225 | 0.143 |
| **0521-0603** | 0.101 | 0.339 | 0.226 | 0.237 | 0.503 | 0.629 |
| **0604-0617** | 0.100 | 0.480 | 0.200 | 0.248 | 0.273 | 0.600 |
| **0618-0701** | 0.059 | 0.178 | 0.078 | 0.163 | 0.214 | 0.451 |

**TABLE 4.** Experiment results of HitEn based on KNN, RF, SVM, NB, CNN and LSTM models.

| Prediction period | Model-a | Model-b | Model-c | Model-d | Model-e | Model-f |
|---|---|---|---|---|---|---|
| 0101-0114 | 0.139 | 0.160 | 0.202 | 0.222 | 0.270 | 0.525 |
| 0115-0128 | 0.155 | 0.279 | 0.063 | 0.172 | 0.302 | 0.497 |
| 0129-0211 | 0.022 | 0.261 | 0.043 | 0.213 | 0.029 | 0.391 |
| 0212-0225 | 0.366 | 0.299 | 0.122 | 0.232 | 0.314 | 0.366 |
| 0226-0311 | 0.125 | 0.291 | 0.166 | 0.191 | 0.285 | 0.208 |
| 0312-0325 | 0.058 | 0.259 | 0.374 | 0.169 | 0.396 | 0.402 |
| 0326-0408 | 0.375 | 0.322 | 0.430 | 0.391 | 0.462 | 0.698 |
| 0409-0422 | 0.074 | 0.464 | 0.146 | 0.307 | 0.183 | 0.220 |
| 0423-0506 | 0.135 | 0.135 | 0.081 | 0.032 | 0.244 | 0.378 |
| 0507-0520 | 0.062 | 0.234 | 0.062 | 0.112 | 0.196 | 0.125 |
| 0521-0603 | 0.104 | 0.312 | 0.250 | 0.244 | 0.463 | 0.623 |
| 0604-0617 | 0.053 | 0.317 | 0.132 | 0.132 | 0.180 | 0.530 |
| 0618-0701 | 0.034 | 0.154 | 0.051 | 0.094 | 0.185 | 0.291 |

to memorize and share the modified weights. This advantage can help LSTM model save a part of the time of weight correction in the process of crime hot spot prediction, and has a certain applicability for the prediction of hotspot grids.

**TABLE 5. Experiment results of LSTM model before and after adding covariates.**

| Prediction period | Model-f HitRa | Model- F HitRa | Model-f HitRn | Model-F HitRn | Model-f HitEn | Model-F HitEn |
|---|---|---|---|---|---|---|
| 0101-0114 | 0.575 | 0.550 | 0.600 | 0.564 | 0.525 | 0.502 |
| 0115-0128 | 0.500 | 0.531 | 0.500 | 0.500 | 0.497 | 0.528 |
| 0129-0211 | 0.486 | 0.541 | 0.436 | 0.491 | 0.391 | 0.435 |
| 0212-0225 | 0.450 | 0.600 | 0.500 | 0.625 | 0.366 | 0.487 |
| 0226-0311 | 0.303 | 0.606 | 0.250 | 0.600 | 0.208 | 0.415 |
| 0312-0325 | 0.368 | 0.474 | 0.435 | 0.478 | 0.402 | 0.518 |
| 0326-0408 | 0.684 | 0.632 | 0.696 | 0.652 | 0.698 | 0.645 |
| 0409-0422 | 0.200 | 0.500 | 0.167 | 0.571 | 0.220 | 0.551 |
| 0423-0506 | 0.500 | 0.643 | 0.553 | 0.711 | 0.378 | 0.486 |
| 0507-0520 | 0.170 | 0.596 | 0.143 | 0.625 | 0.125 | 0.437 |
| 0521-0603 | 0.612 | 0.612 | 0.629 | 0.645 | 0.623 | 0.623 |
| 0604-0617 | 0.541 | 0.541 | 0.600 | 0.600 | 0.530 | 0.530 |
| 0618-0701 | 0.436 | 0.667 | 0.451 | 0.725 | 0.291 | 0.446 |

Model-f is LSTM prediction model based on historical data, and Model-F is LSTM prediction model based on historical crime data and built environment covariates. According to the experimental results, we found that the prediction accuracy of the prediction accuracy of the LSTM model was also improved after adding built environment covariates, and the average prediction index-HitRa of 13 experimental periods increased by percentage points increased by 12.8 percentage points, the average prediction index-HitRn of 13 experimental periods increased by 14 percentage points, and the average prediction index-HitEn of 13 experimental periods increased by 10.4 percentage points.

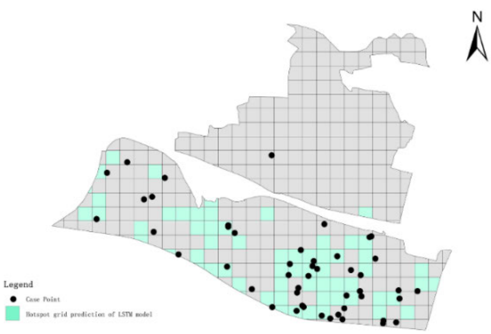



**FIGURE 6. Prediction results of LSTM model using only history data.**

Taking the biweekly period from June 18 to July 01, 2018 as an example, the comparison of the overall prediction results of the study area is shown in the figure. It can be seen that the LSTM model and the LSTM model with covariates have higher prediction accuracy based on their own high self-learning and advantages of processing time series data. The LSTM model with built environment covariates is better than the LSTM in the prediction of crime hot spots. Therefore, the LSTM model with built environment covariates has better application value in the prediction of crime.

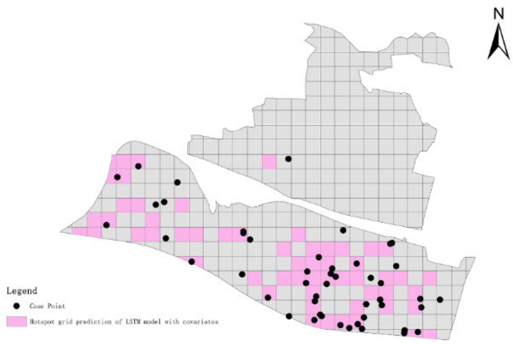



**FIGURE 7. Prediction results of LSTM model using history data and Covariate data.**

**TABLE 6. Experiment results of HitRa of JZ based on KNN, RF, SVM, NB, CNN and LSTM models.**

| Prediction period | Model-a | Model-b | Model-c | Model-d | Model-e | Model-f |
|---|---|---|---|---|---|---|
| 0101-0114 | 0.243 | 0.395 | 0.256 | 0.357 | 0.485 | 0.519 |
| 0115-0128 | 0.249 | 0.472 | 0.327 | 0.381 | 0.496 | 0.636 |
| 0129-0211 | 0.06 | 0.102 | 0.075 | 0.173 | 0.122 | 0.26 |
| 0212-0225 | 0.154 | 0.202 | 0.086 | 0.043 | 0.065 | 0.169 |
| 0226-0311 | 0.202 | 0.424 | 0.287 | 0.388 | 0.431 | 0.627 |
| 0312-0325 | 0.151 | 0.256 | 0.139 | 0.146 | 0.164 | 0.314 |
| 0326-0408 | 0.118 | 0.258 | 0.128 | 0.122 | 0.143 | 0.487 |
| 0409-0422 | 0.235 | 0.373 | 0.254 | 0.32 | 0.331 | 0.493 |
| 0423-0506 | 0.163 | 0.325 | 0.225 | 0.261 | 0.349 | 0.486 |
| 0507-0520 | 0.14 | 0.421 | 0.216 | 0.252 | 0.306 | 0.417 |

### B. VALIDATION IN JZ TOWN

The models are validated in JZ, another town in the same city as the study area XT. JZ is located is at the junction of urban and rural areas, with an area of 1.34 times that of XT town and a population of 29.7% of XT. From 2015 to 2018, the total number of crimes of crime types studied in this paper is 33.2% of that in XT. Through the modeling research of each machine learning algorithm, it is found that the performance of each algorithm is basically consistent with that of XT (Tables 5), with the LSTM model still performing the best.

## VII. CONCLUSION

In this paper, six machine learning algorithms are applied to predict the occurrence of crime hotspots in a town in the southeast coastal city of China. The following conclusions are drawn:1) The prediction accuracies of LSTM model are better than those of the other models. It can better extract the pattern and regularity from historical crime data. 2) The addition of urban built environment covariates further improves the prediction accuracies of the LSTM model. The prediction results are better than those of the original model using historical crime data alone.

Our models have improved prediction accuracies, compared with other models. In empirical research on the prediction of crime hotspots, Rummens *et al.* used historical crime data at a grid unit scale of 200 m×200 m, using three models of logistic regression, neural network, and the combination of logistic regression and neural network [41]. In the biweekly forecast, the highest case hit rate for the two-robbery type is 31.97%, and the highest grid hit rate is 32.95%; Liu *et al.* Used the random forest model to predict the hot spots in multiple experiments in two weeks under the research scale of 150 m × 150 m [23]. The average case hit rate of the model was 52.3%, and the average grid hit rate was 46.6%. The case hit rate of the LSTM model used in this paper was 59.9%, and the average grid hit rate was 57.6%, which was improved compared with the previous research results,

For the future research, there are still some aspects to be improved. The first is the temporal resolution of the prediction. Felson *et al.* revealed that the crime level changes with time [43] Some studies have shown that it is useful to check the variation of risks during the day [44]. We chose two weeks as the prediction window. It does not capture the impact of crime changes within a week, let alone the change within a day. The sparsity of data makes the prediction of crime event difficult if the prediction window is narrowed down to day of a week or hour within a day. There is no viable solution to this challenging problem at this time. The second is the spatial resolution of the grid. In this paper, the grid size is 150m ∗ 150m. Future research will assess the impact of changing grid sizes on prediction accuracy. Third, the robustness and generality of the findings of this paper needs to be tested in other study areas. Nonetheless, the findings of this research have proven to be useful in a recent hotspot crime prevention experiment by the local police department at the study size.

## REFERENCES

[1] U. Thongsatapornwatana, "A survey of data mining techniques for analyzing crime patterns," in *Proc. 2nd Asian Conf. Defence Technol. (ACDT)*, Jan. 2016, pp. 123–128.

[2] J. M. Caplan, L. W. Kennedy, and J. Miller, "Risk terrain modeling: Brokering criminological theory and GIS methods for crime forecasting," *Justice Quart.*, vol. 28, no. 2, pp. 360–381, Apr. 2011.

[3] M. Cahill and G. Mulligan, "Using geographically weighted regression to explore local crime patterns," *Social Sci. Comput. Rev.*, vol. 25, no. 2, pp. 174–193, May 2007.

[4] A. Almehmadi, Z. Joudaki, and R. Jalali, "Language usage on Twitter predicts crime rates," in *Proc. 10th Int. Conf. Secur. Inf. Netw. (SIN)*, 2017, pp. 307–310.

[5] H. Berestycki and J.-P. Nadal, "Self-organised critical hot spots of criminal activity," *Eur. J. Appl. Math.*, vol. 21, nos. 4–5, pp. 371–399, Oct. 2010.

[6] K. C. Baumgartner, S. Ferrari, and C. G. Salfati, "Bayesian network modeling of offender behavior for criminal profiling," in *Proc. 44th IEEE Conf. Decis. Control, Eur. Control Conf. (CDC-ECC)*, Dec. 2005, pp. 2702–2709.

[7] W. Gorr and R. Harries, "Introduction to crime forecasting," *Int. J. Forecasting*, vol. 19, no. 4, pp. 551–555, Oct. 2003.

[8] W. H. Li, L. Wen, and Y. B. Chen, "Application of improved GA-BP neural network model in property crime prediction," *Geomatics Inf. Sci. Wuhan Univ.*, vol. 42, no. 8, pp. 1110–1116, 2017.

[9] R. Haining, "Mapping and analysing crime data: Lessons from research and practice," *Int. J. Geogr. Inf. Sci.*, vol. 16, no. 5, pp. 203–507, 2002.

[10] S. Chainey, L. Tompson, and S. Uhlig, "The utility of hotspot mapping for predicting spatial patterns of crime," *Secur. J.*, vol. 21, nos. 1–2, pp. 4–28, Feb. 2008.

[11] S. Chainey and J. Ratcliffe, "GIS and crime mapping," *Soc. Sci. Comput. Rev.*, vol. 25, no. 2, pp. 279–282, 2005.

[12] L. Lin, W. J. Liu, and W. W. Liao, "Comparison of random forest algorithm and space-time kernel density mapping for crime hotspot prediction," *Prog. Geogr.*, vol. 37, no. 6, pp. 761–771, 2018.

[13] C. L. X. Liu, S. H. Zhou, and C. Jiang, "Spatial heterogeneity of micro-spatial factors' effects on street robberies: A case study of DP Peninsula," *Geograph. Res.*, vol. 36, no. 12, pp. 2492–2504, 2017.

[14] M. I. Jordan and T. M. Mitchell, "Machine learning: Trends, perspectives, and prospects," *Science*, vol. 349, no. 6245, pp. 255–260, Jul. 2015.

[15] X. Zhao and J. Tang, "Modeling temporal-spatial correlations for crime prediction," in *Proc. Int. Conf. Inf. Knowl. Manag. Proc.*, vol. F1318, 2017, pp. 497–506.

[16] A. Babakura, M. N. Sulaiman, and M. A. Yusuf, "Improved method of classification algorithms for crime prediction," in *Proc. Int. Symp. Biometrics Secur. Technol. (ISBAST)*, 2015, pp. 250–255.

[17] Q. Zhang, P. Yuan, Q. Zhou, and Z. Yang, "Mixed spatial-temporal characteristics based crime hot spots prediction," in *Proc. IEEE 20th Int. Conf. Comput. Supported Cooperat. Work Design (CSCWD)*, May 2016, pp. 97–101.

[18] A. R. Dandekar and M. S. Nimbarte, "Verification of family relation from parents and child facial images," in *Proc. Int. Conf. Power, Autom. Commun. (INPAC)*, 2014, pp. 157–162.

[19] G. R. Nitta, B. Y. Rao, T. Sravani, N. Ramakrishiah, and M. BalaAnand, "LASSO-based feature selection and Naïve Bayes classifier for crime prediction and its type," *Serv. Oriented Comput. Appl.*, vol. 13, no. 3, pp. 187–197, 2019.

[20] H. Tyralis and G. Papacharalampous, "Variable selection in time series forecasting using random forests," *Algorithms*, vol. 10, no. 4, p. 114, Oct. 2017.

[21] K. K. Kandaswamy, K.-C. Chou, T. Martinetz, S. Möller, P. N. Suganthan, S. Sridharan, and G. Pugalenthi, "AFP-pred: A random forest approach for predicting antifreeze proteins from sequence-derived properties," *J. Theor. Biol.*, vol. 270, no. 1, pp. 56–62, Feb. 2011.

[22] V. F. Rodriguez-Galiano, B. Ghimire, J. Rogan, M. Chica-Olmo, and J. P. Rigol-Sanchez, "An assessment of the effectiveness of a random forest classifier for land-cover classification," *ISPRS J. Photogramm. Remote Sens.*, vol. 67, pp. 93–104, Jan. 2012.

[23] L. Lin, J. Jiakai, S. Guangwen, L. Weiwei, Y. Hongjie1, and L. Wenjuan, "Hotspot prediction of public property crime based on spatial differentiation of crime and built environment," *J. Geo-Inf. Sci.*, vol. 21, no. 11, pp. 1655–1668, 2019.

[24] Z. Jun and H. Wenbo, "Recent advances in Bayesian machine learning," *J. Comput. Res. Develop.*, vol. 52, no. 1, pp. 16–26, 2015.

[25] J. T. Huang, J. Li, and Y. Gong, "An analysis of convolutional neural networks for speech recognition," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, South Brisbane, QLD, Australia, Apr. 2015, pp. 4989–4993.

[26] Z. Feiyan, J. Linpeng, and D. Jun, "Review of convolutional neural network," *Chin. J. Comput.*, vol. 40, no. 6, pp. 1229–1251, 2017.

[27] Y. Yang, J. Dong, X. Sun, E. Lima, Q. Mu, and X. Wang, "A CFCC-LSTM model for sea surface temperature prediction," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 2, pp. 207–211, Feb. 2018.

[28] X. Hong, R. Lin, C. Yang, N. Zeng, C. Cai, J. Gou, and J. Yang, "Predicting Alzheimer's disease using LSTM," *IEEE Access*, vol. 7, pp. 80893–80901, 2019.

[29] L. Mou, P. Zhao, and Y. Chen, "Short-term traffic flow prediction: A long short-term memory model enhanced by temporal information," in *Proc. 19th COTA Int. Conf. Transp. Prof. CICTP Transp. China-Connect. World*, 2019, pp. 2411–2422.

[30] L. E. Cohen and M. Felson, "Social change and crime rate trends: A routine activity approach," *Amer. Sociol. Rev.*, vol. 44, no. 4, p. 588, Aug. 1979.

[31] G. Gudjonsson, "The reasoning criminal. Rational choice perspectives on offending," *Behav. Res. Therapy*, vol. 26, no. 3, pp. 246–287, 1988.

[32] P. Brantingham and P. Brantingham, "Criminality of place—Crime generators and crime attractors," *Eur. J. Crim. Policy Res.*, vol. 3, no. 3, pp. 5–26, 1995.

[33] *Enhancing Urban Safety and Security. Global Report on Human Settlements 2007*, UN-Habitat, Nairobi, Kenya, 2007.

[34] G. Owusu, C. Wrigley-Asante, M. Oteng-Ababio, and A. Y. Owusu, "Crime prevention through environmental design (CPTED) and built-environmental manifestations in Accra and Kumasi, Ghana," *Crime Prevention Community Saf.*, vol. 17, no. 4, pp. 249–269, Nov. 2015.

[35] Y. Wenhao and A. Tinghua, "The visualization and analysis of POI features under network space supported by kernel density estimation," *Acta Geodaetica et Cartographica Sinica*, vol. 44, no. 1, pp. 82–90, 2015.

[36] G. Song, L. Xiao, S. Zhou, D. Long, S. Zhou, and K. Liu, "Impact of residents' routine activities on the spatial-temporal pattern of theft from person," *Acta Geography Sinica*, vol. 72, no. 2, pp. 356–367, 2017.

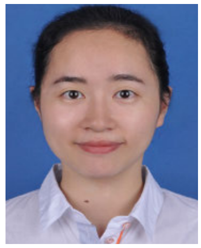
[37] L. Lin, D. Fang-Ye, X. Lu-Zi, S. Guang-Wen, and J. C. L. Kai, ''The density of various road typesand larceny rate: An empirical analysis of ZG city,'' *Hum. Geeography*, vol. 32, no. 6, pp. 32–39, 2017.
[38] C. Xu, L. Liu, and S. H. Zhou, ''The comparison of predictive accuracy of crime hotspot density maps with the consideration of the near similarity: A case study of robberies at DP Peninsula,'' *Scientia Geographica Sinica*, vol. 36, no. 1, pp. 55–62,2016.
[39] G. Rosser, T. Davies, K. J. Bowers, S. D. Johnson, and T. Cheng, ''Predictive crime mapping: Arbitrary grids or street networks,'' *J. Quantum Criminol.*, vol. 33, no. 3, pp. 569–594, 2017.
[40] D. Griffith, *Multivariate Statistical Analysis for Geographers*. Upper Saddle River, NJ, USA: Prentice-Hall, 1997.
[41] A. Rummens, W. Hardyns, and L. Pauwels, ''The use of predictive analysis in spatiotemporal crime forecasting: Building and testing a model in an urban context,'' *Appl. Geography*, vol. 86, pp. 255–261, Sep. 2017.
[42] S. Favarin, ''This must be the place (to commit a crime). Testing the law of crime concentration in Milan, Italy,'' *Eur. J. Criminol.*, vol. 15, no. 6, pp. 702–729, Nov. 2018.
[43] M. Felson and E. Poulsen, ''Simple indicators of crime by time of day,'' *Int. J. Forecasting*, vol. 19, no. 4, pp. 595–601, Oct. 2003.
[44] A. Sagovsky and S. D. Johnson, ''When does victimisation occur?'' *Austral. New Zealand J. Criminol.*, vol. 40, no. 1, pp. 2–16, 2007.

**XU ZHANG** received the B.S. degree in computer science and technology and the M.S. degree in electronic and communication engineering from Hainan University, Haikou, China, in 2011 and 2015, respectively. He is currently pursuing the Ph.D. degree in cyberspace security with Guangzhou University, Guangzhou, China.

His current research interests include big data analysis, spatiotemporal data mining, machine learning, and crime geography.

**LIN LIU** received the B.S. and M.S. degrees in geography and remote sensing from Peking University, Beijing, China, and the Ph.D. degree in GIScience from The Ohio State University, Columbus, OH, USA.

He is currently a Professor of geography with the University of Cincinnati. He is also a Professor with Guangzhou University. His current research interests include crime analysis and applications of GIScience.



**LUZI XIAO** received the B.S. and Ph.D. degrees in human geography from the School of Geography and Planning, Sun Yat-Sen University, Guangzhou, China.

She is currently a Lecturer with the School of Geographical Sciences, Guangzhou University. Her research interests include crime geography, GIS, and spatiotemporal big data analysis.

**JIAKAI JI** received the B.Sc. degree in geography information science from South China Normal University, Guangzhou, China, in 2017, and the M.S. degree in cartography and geography information system from Sun Yat-sen University, Guangzhou, in 2020. His research interests include criminal geography and machine learning.