Comparison of Machine Learning Algorithms for Predicting Crime Hotspots

[1]Dr.M Jaganathan(Associate professor),
[2]Nikitha, [3]Harika, [4]Murari, [5]Ajay,
Department of CSE,
MALLA REDDY INSTITUTE OF TECHNOLOGY AND SCIENCE, Telangana, Hyderabad.

ABSTRACT

The development of police plans and the execution of crime prevention and control measures are greatly aided by accurate crime prediction. At the moment, the most popular way to make predictions is using machine learning. Unfortunately, there has been a dearth of research that compares various machine learning approaches to crime prediction in a systematic way. This study compares the prediction potential of several machine learning algorithms using public property crime data from a portion of a big coastal metropolis in southeastern China from 2015 to 2018. Evidence from past criminal cases alone suggests that LSTM fared better than other models, including convolutional neural networks, naive Bayes, support vector machines, random forests, and KNN. Furthermore, the LSTM model is fed with built environment data in the form of urban road network density and sites of interests (POIs). Results show that when comparing the original model—which relied just on historical crime data—with the one that incorporates built environment variables, the latter produces superior prediction results. Historical crime data and criminological theory-related factors should, therefore, be used for future crime prediction. Predicting criminal behavior is a challenging task for many machine learning systems.

I. INTRODUCTION

There has been a meteoric rise in the amount of spatiotemporal data pertaining to public safety in the last few years. On the other hand, not all data sets have been put to good use in solving actual issues. Several researchers have created crime prediction models to help with crime prevention efforts [1]. When calibrating their prediction algorithms, the majority relied only on past crime data. Two main areas of crime prediction research are the prediction of crime risk areas [2, 3] and crime hotspots [4, 5]. The "routine activity theory" is the foundation for both the physical environment and the important in_uencing variables of criminal activities, which together form the crime risk area forecast. [6]. Once upon a time, there was a system for estimating the likelihood of criminal activity that relied on looking at crime statistics and making the assumption that patterns would continue into the future [7]. To illustrate, the terrain risk model is useful for reliable, long-term hotspot prediction since it takes into account the closeness of crime locations and the aggregation of crime components,

as well as environmental characteristics connected to crime and crime history [2]. There has been a lot of empirical study on crime prediction throughout many time periods, using data from demographic and economic statistics, land use, cell phone records, and criminal history. By analyzing historical data, crime hotspot prediction attempts to pinpoint exactly where criminal activity is most likely to cluster in the future [8]. Kernel density estimation is one of the most used methods [9, 12]. Better results are achieved by models that take into consideration the geographical or temporal autocorrelations of previous occurrences [13]. In recent times, machine learning algorithms have become more popular. Support vector machines (SVMs), neural networks, Bayesian models, K-Nearest Neighbor (KNN), and random forest algorithms are among the most used approaches [6]. Some have contrasted linear approaches for predicting crime trends [14], others have compared Bayesian models with BP neural networks [15], and yet others have contrasted the spatiotemporal kernel density method with the random forest method across time ranges of crime prediction [12].

One efficient supervised learning approach algorithm is KNN [17], [18]. SVM's versatility makes it a go-to model for machine learning applications; in addition to classi_cation and regression, it can also spot outliers [4], [19]. Numerous fields have shown the great prediction accuracy and robust non-linear relational data processing capabilities of the random forest method [20] [23]. One traditional classification technique is Naive Bayes (NB), which handles missing data with ease and has minimal parameters [15], [24]. With the addition of a very deep layer, convolution neural networks (CNNs) may improve their expression ability and handle more complicated classification tasks [25], [26]. Processing data with strong time series trends is significantly impacted by the Long Short-Term Memory (LSTM) neural network's ability to extract time-series features from features. (27, 29). In order to show the predictive capacity with and without covariates, this study will compare and contrast the six machine learning methods mentioned above and then suggest the best performing one.

II. RELATED WORK

A. THEORETICAL CRIMINOLOGY PRINCIPLES FOR THE PREDUCTION OF HIGH-INTELLIGENCE CASES

Predicting where criminal activities will cluster in the future is the main goal of crime hotspot prediction. The essential theoretical groundwork is supplied by theoretical criminology. In particular, a number of interconnected criminological theories not only help us make sense of the role that physical locations play in the genesis and consolidation of criminal incidents, but they also lay out the groundwork for how law enforcement can use data on problem areas to bring them under control. Theories like rational decision, routine behavior, and criminal patterns make up the bulk of it. Theoretically, situational crime prevention is based on these three beliefs. In 1979, Cohen and Felson jointly introduced routine activity theory [30], which has since been expanded upon by incorporating additional ideas. According to this view, the three conditions—motivated criminals, appropriate victims, and an inability to protect oneself in terms of both time and space—must come together for the commission of any crime, but notably predatory crimes. It was Cornish and Clarke that first put forward rational choice theory [31]. Theoretically, the criminal's decisions on setting, objectives, and means may be accounted for by a logical equilibrium between danger, effort, and reward.

Part B: Information on Man-Made Environments

Many recent studies have shown that crime prevention and reduction chances are significantly influenced by the physical environment of cities, and hence by urban criminal behavior. According to the 2007 Global Habitat Report, built environment factors significantly affect the frequency of criminal activities [33].

The crime prediction model takes into account data on road network density and point of interests (POIs) as variables.

(1) Point-of-Interest Information

The information about city infrastructure Location and attribute data for a wide range of urban amenities are part of POI [34], [35]. Places with plenty of people and easy access to transportation tend to have catering facilities, shopping centers, and shops, which means that these establishments attract a wide variety of individuals and might serve as easy targets for criminals [36]. The POIs are chosen to be included in the prediction model as covariates.

2) The Density of Road Networks

The traditional way of looking at road network density is by dividing the total length of roads by the size of an area unit. A higher concentration of roads means more traffic, which means more opportunities for both victims and offenders. The density of the road network affects the crime rate, particularly in public spaces, according to previous research [37].

C. Suggesting Criminal Acts Using Machine Learning Techniques

Using the distribution of crime cases across time, conventional approaches often identify areas with a high crime rate and make the assumption that this pattern will continue in the future [7], [2]. When trying to forecast where crime would remain steady over the long run, this assumption is usually realistic. For such persistent hotspots, the widely-used KDE technique works well [10], [11]. When compared to the general KDE technique, the one based on temporal autocorrelation tends to perform better [38]. Citation: Liu et al. When looking at short time scales and grid space units, the random forest approach was determined to be more efficient than the standard spatiotemporal KDE technique [12]. For their street-level crime prediction study, Gabriel et al. [39] used the Gated Localized Diffusion Network. The prediction accuracy was significantly improved using the diffusion network methodology as compared to the classic Network-time KDE method. Machine learning algorithms have shown themselves capable of handling non-linear relational data in several fields, including crime prediction. It can handle very high-dimensional data, train quickly, and extract data properties.

III. PREDICTION MODEL

This study uses many algorithms for crime prediction, including random forest, KNN, SVM, and LSTM. To begin, the models are calibrated using just crime data from the past. The best model might be found by comparing them. Additionally, in order to assess whether there is room for improvement in the forecast accuracy, built environment statistics like road network density and poi are included as variables in the predictive model.

Okay. KNN

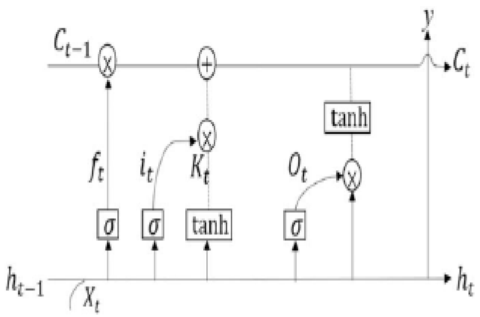
K-nearest neighbor (KNN) is a method for selecting the K-nearest classifications; it takes an instance's feature vector as input, determines the distance between the training set and the new data's feature value, and finally makes its selection. The data to be examined is the closest neighbor class if k is less than 1. The classification determination rule of KNN is either a weighted vote depending on distance or a majority vote. The input instance's categorization is determined by the majority of its k nearby training examples.

Section B. Unpredictable Forest

The random forest is a collection of tree classifiers {h(x, _k),k D1... }, where the meta classifier h(x, _k) is a linear regression tree built using the CART algorithm, x is the input vector, _k is an independently distributed random vector, and the forest's output is obtained through voting. There are two parts to random forest that contribute to its randomness: first, the bagging method is used to randomly choose the training sample set, and second, the split attribute set is also randomly selected. Assuming there are a total of M attributes in the training sample, we assign an integer value to each internal node representing an attribute number, choose F attributes at random from the set of M attributes to serve as the split attribute, and then determine the optimal split mode for the set of f attributes. Cut the nodes in half. The multi decision tree uses a random forest, and the tree classifier votes on the final classification result.

C. SVM

The statistical learning theory-based support vector machine (SVM) is a data mining tool that excels at a wide variety of tasks, including pattern recognition (classification issue, discriminate analysis) and regression (time series analysis). Finding a better classification hyper plane that satisfies the classification criteria is the mechanism of support vector machines (SVM). This hyper plane should guarantee the classification accuracy and maximize the blank area on both sides. Optimized classification of linearly separable data is theoretically possible using SVM.



The LSTM algorithm's structural chart (FIGURE 1).

In the figure above, we can see that whereas RNN simply transmits cell state C, LSTM contains two state chains—h for hidden layer and C for cell—that are conveyed across time. In the above equation, ht-1 represents the value of the current time that was sent from the hidden layer at a prior time, Xt stands for the input value at the current time, Ct-1 represents the state value of the LSTM memory cell at a previous time, and Ct describes the current state value of the memory cell. It is possible to determine what data should be lost when ht-1 and Xt go through the forgetting gate. There is a range of values for the output to the cell state; 0 indicates complete forgetfulness and 1 indicates complete reservation of knowledge. The following equation gives the forgetting gate ft:

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

where Sigmoid is the activation function and w and b are the weight and bias matrices of the forgetting gate, respectively. Adding additional data to a cell may be done in two ways. A new value kt is added to the cell state V when the Sigmoid function's input gate calculates the information that needs updating.

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

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

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

How the cell is in its final state determines the output. Sigmoid first determines which output data to output, uses the tanh function to process the cell state, and then gets the state value ht, which the hidden layer transmits to the next iteration. Equations (5)–(7) indicate that after passing through the sigmoid, ht may get the pre-output value y at the present moment:

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

$$h_t = O_t * tanh(C_t) \quad (6)$$

$$y = \sigma(w' h^t) \quad (7)$$

IV. EVALUATION INDICATOR

These metrics are used to evaluate machine learning models by comparing their prediction performance before and after the addition of variables. One metric used to measure the precision of crime prediction is the hit rate. Grid and Case Hit Rates make up the Hit Rate primarily. Grid Success Rate The ratio of projected accurate hotspot grids to total actual hotspot grids is called HitRa.

$$\mathrm{HitR}_a = \frac{a^*}{A} \quad (8)$$

$$\mathrm{HitR}_n = \frac{n}{N} \quad (9)$$

$$\mathrm{HitE}_n = \frac{HitR_n}{a/A} \quad (10)$$

where "a" denotes the expected number of hotspot grids and "A" is the actual list of hotspots.

V. EXPERIMENTAL AREA AND DATA VISUALIZATION ANALYSIS

Section A: The Test Field

A town in a coastal megacity in Southeast China is the location XT chosen for this article. With an area of approximately 6.5 square kilometers and a total population of around 400000, this community has a relatively large population density. However, only 50,000 of these residents are listed as registered households, which suggests that most of the residents are either domestic migrants or not from the area. Multiple large-scale city villages make up the town. This neighborhood has a high crime rate due to the complicated mix of the built environment and inhabitants.

B. Crime Type Selection

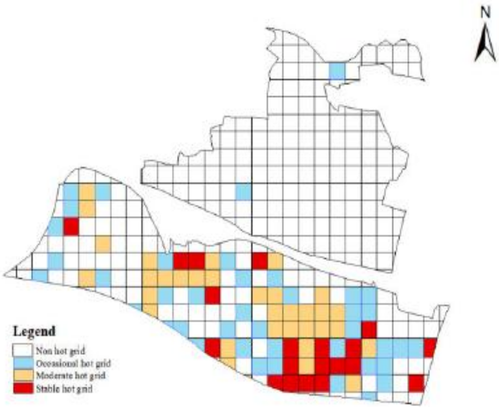
In most cases, the primary goal of the crime known as "the crime of property in public places" is to occupy the property ownership of others while in a public area. Theft, robbery, snatching, and other forms of embezzlement that include the full acquisition of someone else's property against their will mostly fall under this category. Selecting this town's public property crime as a predictor of crime hotspots is very practically significant. By shifting from reactive to proactive prevention and control, local public safety may be enhanced with the use of accurate crime prediction to direct the allocation of police resources.

C. Analysis of Data Visualization

The experimental district's Public Security Bureau's P-GIS database contains historical crime data used in this work, which spans from 2015 to 2018. After locating the case point on the study area map, the text coordinate information stored in the database is retrieved. The data within the street range of the research area is then extracted. The geographical size of experiments predicting crime hot spots should be kept as modest as feasible so that they can fulfill the demands of real police activity. The research area is partitioned into 150m x 150m grids in accordance with the examination of real police activity and the data distribution of case points, as per the calculation formula of the gridding processing study area of Grif_th et al. [40]. If the grids are split by 150 meters, the case points will be more concentrated in certain grids and the hotspot grids will be less likely to occur than in grids with lower spatial dimensions. In addition to improving the accuracy of predictions and pinpointing crime hotspots, this division will provide a clearer picture of the case distribution and process. In order to make better use of the prediction findings for crime prevention and control, the inquiry into real police work has shown that a single officer can cover a maximum of 150 meters in a single time unit while on patrol.

1) Pattern of the Hotspot Grid

Based on the distribution of 78 two-week historical crimes from 2015-2017, we split the research region XT into 369 grids and then count the frequency of cases in each grid. After finding that four clusters was the sweet spot using the K-means clustering approach, we may classify the grid as either stable high-risk, high-risk, occasional hot, or non-hot.



2. Grid structure of hotspots in the research region XT.

2) Distribution of points of interest

The figure shows the distribution of points of interest (POI) in the XT study region, which includes catering, retail malls, and entertainment facilities. To get the POI points of the research region, these POI kinds are given to each grid as a variable and then spatially interpolated.

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

3) STATISTICS OF CASES BY PERIOD

In comparison to the previous three years, 2018 had a somewhat lower total number of cases, whereas 2017 had a somewhat higher total number of cases than both 2015 and 2016. The frequency of two-week instances changed throughout the course of the four years. Nearly every two-week period had between forty and eighty cases, with an average of fifty-eight cases each two-week period. Figure 9 displays the four-year case volume curve, which exhibits a consistent pattern of change. Case volume essentially drops significantly in the two weeks leading up to vacations and then picks up again in the two weeks after the holidays. There is a noticeable decreasing tendency in the case volume during the first two months of the year. Annually, the two weeks around spring breaks saw the lowest number of cases.

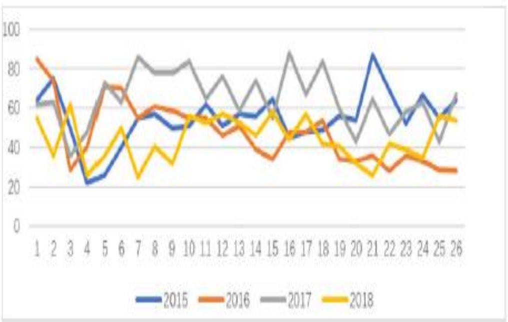


FIGURE 4. Statistics of biweekly cases of the study area XT.

VI. EXPERIMENTS

A. Try It Out in Next Town

This work uses findings from Rummens [38] and Lin et al. [13] to project the hot grid of public property crime for 13 time units from January 1, 2018, to July 1, 2018, using a two-week time unit as the unit of analysis. For the first n hotspots with cases in the forecast period from all grids, the history data and covariate data are used. Historical case data and covariate data reflecting the environment make up the bulk of the variable data required for the prediction model. Before dividing the time period of the point data according to the predicted time unit, the historical case data is found using the location and coordinates.For each period, we tally up the number of instances in each grid; this data serves as the foundation for the prediction model. Then, we choose the training data from the relevant period based on the prediction goal period. Next, we have the covariate data. As a covariate in the prediction model, the density surface of the research region is obtained using spatial interpolation of covariate spatial point data using city POI density and road network density. The data used in this paper's historical analysis is a count of the number of instances per grid from 2015/2016/2017 to the target period and the four periods immediately before it. Two variables used in contemporary city data, point-of-interest (POI) and road network density, provide the values for the covariate data. The following transformation function is used with MinMaxScaler to normalize the data between 0 and 1: V

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

Table 1 shows the projection of crime hotspots based on historical data from January 1, 2018, to January 14, 2018.

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

The figure below displays the results of many models. Model-a is a KNN model, Model-b is an RF model, Model-c is an SVM model, Model d is an NB model, Model e is a CNN model, and Model f is an LSTM model for prediction. Tables 2 and 3 show that out of four prediction models, the LSTM model (Model-d) had the greatest overall performance in 13 time unit prediction trials conducted in the first half of 2018. When using covariate data in an LSTM prediction model, for instance, the average grid hit rate may reach 44.8%, and in this case, more than half of the projected right grids can be covered, on average, at 45.8%.

Table 2. Results of HitRa's experiments using CNN, LSTM, KNN, RF, and SVM 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 shows the outcomes of HitRn experiments using several models, including KNN, RF, SVM, NB, CNN, and LSTM.

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

through the instances within the research region. The long short-term memory (LSTM) prediction model has the ability to remember and communicate the modified weights in addition to the short- and long-term feature information collected from time series data. This benefit may aid LSTM models in reducing the amount of time spent correcting weights during crime hot spot prediction and can be used to some extent for hotspot grid prediction. Model-F is an LSTM prediction model that takes into account both historical crime data and built environment variables, whereas Model-f is based on data collected from the past. By incorporating built-environment variables into the LSTM model, we were able to improve its prediction accuracy. Over the course of the 13 experimental periods, the average prediction index-HitRa increased by 12.8 percentage points, the average prediction index-HitRn increased by 14 percentage points, and the average prediction index-HitEn increased by 10.4 percentage points.

VII. CONCLUSION

The purpose of this study is to forecast the likelihood of crime hotspots in a town located in a city on China's southeast coast using six machine learning techniques. We may deduce the following:1) LSTM model outperforms the other models in terms of prediction accuracy. When applied to crime statistics, it improves the extraction of regularities and patterns. 2) The LSTM model's prediction accuracies are even further enhanced by include urban built environment variables. In comparison to the first model that relied only on crime statistics from the past, the current forecast results are superior. When compared to competing models, our models offer much higher prediction accuracies. Using three models of logistic regression, neural networks, and a combination of the two, Rummens et al. [41] conducted empirical study on the prediction of crime hotspots using historical crime data at a grid unit size of 200 m_200 m. Using a study size of 150m_150m, Liu et al.[23] used a random forest model to predict hotspots in several tests over the course of two weeks. The maximum case hit rate for the two-robbery type was 31.97%, while the greatest grid hit rate was 32.95%. The model has an average success rate of 52.3% for cases and 46.6% for grids. An improvement over earlier studies was seen in the case hit rate of 59.9% and the average grid hit rate of 57.6% using the LSTM model utilized in this work.
Some areas still need improvement for the study that will be conducted in the future. The first thing to consider is the prediction's temporal resolution.
According to Felson et al., the crime rate fluctuates over time [43]. It is helpful to examine the variance of hazards during the day, according to several research [44].As a forecast window, we settled on two weeks. Even a one-day variation in crime rates is too small to be captured by this measure. If the prediction window is limited to a single day or even an hour inside a single day, the data is too sparse to reliably anticipate when a crime will occur. Currently, there is no workable answer to this complex issue. Grid spatial resolution is the second. The grid size used in this article is 150m _ 150m. Examining how different grid sizes affect prediction accuracy is a topic for future study. Third, it is necessary to evaluate the generalizability and robustness of this paper's findings in different fields of research. Regardless, the local police department's most recent hotspot crime prevention effort at the study size has shown that the results of this research are beneficial.

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 pro_ling,'' 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. Fore- casting, 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, andW.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.