Computing and Informatics, Vol. 42, 2023, 568–590, doi: 10.31577/cai_2023_3_568

MODEL FOR SPATIOTEMPORAL CRIME PREDICTION WITH IMPROVED DEEP LEARNING

Ature Angbera

Universiti Sains Malaysia
School of Computer Sciences
Pulau Pinang 11800, Malaysia
&
Department of Computer Science
Joseph Sarwuan Tarkaa University Makurdi
Benue State, Nigeria
e-mail: angberaature@student.usm.my

Huah Yong Chan

Universiti Sains Malaysia
School of Computer Sciences
Pulau Pinang 11800, Malaysia
e-mail: hychan@usm.my

Abstract. Crime is hard to anticipate since it occurs at random and can occur anywhere at any moment, making it a difficult issue for any society to address. By analyzing and comparing eight known prediction models: Naive Bayes, Stacking, Random Forest, Lazy:IBK, Bagging, Support Vector Machine, Convolutional Neural Network, and Locally Weighted Learning – this study proposed an improved deep learning crime prediction model using convolutional neural networks and the xgboost algorithm to predict crime. The major goal of this research is to provide an improved crime prediction model based on previous criminal records. Using the Boston crime dataset, where our larceny crime dataset was extracted, exploratory data analysis (EDA) is used to uncover patterns and explain trends in crimes. The performance of the proposed model on the basis of accuracy, recall, and f-measure was 100 % outperforming the other models used in this study. The analysis of the proposed model and prediction can aid security services in making better use of their resources, anticipating crime at a certain time, and serving the society better.

Keywords: Crime prediction, deep learning, spatiotemporal, data mining, ensemble learning

1 INTRODUCTION

Human behavior disorder is the leading cause of crimes that wreak havoc on society in many of ways. A crime is a societal illness that affects every sector of the society in a region where it happens. The crime rate is very high in the developing countries. Governments around the globe expend a lot of resources trying to deal with crime, but since crime is very complex in its nature [1], it is always very difficult to tackle it manually in traditional ways. The information communication technology (ICT) can efficiently help dealing with this problem.

Like a disease, crime is a society issue that tends to proliferate in geographic clusters. Since crime is a geographic phenomenom its hotspots, spatial clusters, spatial correlations of various indicators and forecasts provide the common topics for the crime research [2]. Spatiotemporal crime prediction with the latest artificial intelligent techniques is very important. And for public safety and smart city operations spatiotemporal crime prediction is critical [3]. Because crime episodes are sparsely dispersed spatially and temporally, the traditional deep learning approaches backed only by a coarse location-scale can forecast crime density to a limited extent. Law enforcers require precise data regarding illegal activity in order to foresee, respond and solve spatiotemporal illegal conduct.

Anticipating when and where a crime will occur, often referred to as "predictive policing", permits a society to dispatch law enforcers to highly crime potencial regions or circumstances prior to a crime occurring. Criminal activity can be predicted spatially and temporally which is helpful for a targetable allocation of police resources and surveillance. Advanced deep learning techniques are effective tools for predicting future events based on the behavior of previous ones. However, the exponential growth of spatiotemporal data is only rarely used for anticipating crime events [4] using a repository of spatiotemporal crime data sets. The availability of spatiotemporal crime data has already facilitated the development of data-driven strategies for predicting the occurrence of crimes in recent years [5, 6].

The feature representation efficacy of neural network design distinguishes deep learning-based methods from other spatiotemporal prediction methods. Many recently proposed forecasting frameworks, such as attentional neural methods [7], convolution-based learning approach [8], and spatial relation encoder with graph neural networks [9], Spatiotemporal Sequential Hypergraph Network [10], has been focused on modeling time-evolving regularities over the temporal dimension and the underlying regional geographical dependencies over the spatial dimension. Despite their success, we believe that conventional spatiotemporal prediction models fall short meeting the particular problems that multi-dimensional crime data [11] presents. There are explicit and implicit relationships between different kinds of crimes because of the heterogeneity of crime data. Current approaches are inherently incapable of capturing cross-type crime influences in a fully dynamic scenario involving both spatial and temporal patterns due to their inherent architecture.

In this study, we therefore proposed an improved deep learning technique for spatiotemporal crime prediction, using deep convolutional neural networks as the feature extractor and a strong ensemble metal classifier known as XGBoost algorithm for final prediction. Another important angle of this study is to show that crime has always been studied from literature in quantitative terms which combines many crimes to be worked on, thus, making the designed systems less productive. Hence in our study, we have studied crime giving a room for more understanding of the crime and better police allocation. Crime is a legally punished conduct, it is detrimental to society, therefore it is necessary to comprehend crime in order to prevent criminal action [12]. The major goal of this paper is to provide an improved crime prediction model based on previous criminal records.

2 RELATED STUDIES

Several approaches in regards to crime prediction have been presented in recent times to provide police officers with efficient and persuasive knowledge for effective resource allocation in order to avoid future crimes [13, 14]. In [15], the article presented a crime prediction model that utilizes hotspot analysis to enhance its accuracy. The model comprises three phases: Crime Hotspot Identification, Dataset Preparation, and Crime Prediction Approach. In the initial phase, hotspot analysis is employed to pinpoint areas with high crime incidence. In the second stage, the location coordinates are replaced by the cluster number to which they belong, and the modified dataset is used to train the crime prediction model. In the third and final phase, the trained model is utilized to categorize each instance into one of 37 crime categories using advanced techniques like Naive Bayes, Decision Tree, and to ensemble learning approaches. The outcomes of the study demonstrate that incorporating hotspot analysis into the model leads to a significant improvement in crime prediction accuracy. The results indicate that Voting with Naive Bayes and REPTree produce the most reliable classification results, although deep learning could have been utilized for better results. However, the study only uses crime data from one year, which may not be adequate to capture long-term trends or changes in crime patterns. Unlike the study, our research employs a dataset that spans more than one year. In [16], the article introduces a technique for examining the strength and spatiotemporal progression of hotspots identified by the EFCM algorithm [17] for spatiotemporal hotspot detection. The proposed method in the article introduces a novel approach for analyzing the spatiotemporal evolution of hot spots in a specific area by calculating the hot spot strength index, which measures the percentage of time a selected area is affected by hot spots. Furthermore, the method can assess the reliability of the evaluation by calculating a reliability index based on a hot spot reliability measure proposed in the previous study. The application of this method in crime analysis of the City of London using a dataset of criminal events since 2011 shows a decrease in the frequency of all types of criminal events across the study area in the recent years. While the study lacks a detailed comparison of the proposed method with other existing machine learning methods, our study presents a comparison with state-of-the-art machine learning models to fully evaluate the effectiveness of our proposed method. A genetic-fuzzy system was created in [18] to produce an intelligible fuzzy knowledge base that includes patterns for forecasting future spatiotemporal crimes. The system consists of three steps: fuzzy problem space partitioning, meaningful feature selection, and fuzzy knowledge base construction. A generated dataset and a real-world dataset from Tehran, Iran were used to test the suggested system. The results suggest that the proposed approach is a good tool for detecting patterns and forecasting future crimes in contexts where crimes are concentrated in the location and timeaspect. The authors of the article reported a high computational complexity of the method, but they had no specified the extent of it. However, in our study, we proposed an improved deep learning model that can reduce the computational cost. In [19], the article introduces a novel deep learning technique called Geographic-Semantic Ensemble Neural Network (GSEN), which stacks a geographic prediction neural network and a semantic prediction neural network to improve prediction accuracy. The GSEN model combines various structures, including Predictive Recurrent Neural Network (PredRNN), Graph Convolutional Predictive Recurrent Neural Network (GC-PredRNN), and Ensemble Layer, to capture spatiotemporal dynamics from different perspectives. The RMSE of the suggested system was 0.6425 ± 0.0057. However, an improved deep learning model is presented in our study which has lower values for RMSE. In [20], an XGBoost classifier was developed for determining if a seven-day sliding time frame within a given county contains or does not contain a human trafficking-related incident. A case study was conducted with a new combined human trafficking criminal dataset that had a Matthews correlation value of 0.86. However, better advanced deep learning models would have been used for a better result. In [21], a deep learning-based model for spatiotemporal crime prediction using convolutional neural networks is proposed. The proposed approach uses a hierarchical structure to understand the timing of criminal events, with branches that focus on different time periods. Additionally, it utilizes a channel projection to better understand how past events may impact future crime risk. The effectiveness of this model is assessed using publicly available crime data sets from Chicago and Los Angeles, and compared to traditional methods. The proposed model (CNN-PT) outperforms the traditional models in terms of both AP score and RMSE score. The temporal hierarchical structure of the proposed model improves the AP performance of traditional CNN models by 1.4 % in the Chicago dataset and 1.7 % in the Los Angeles dataset. Additionally, the channel projection further improves the AP performance by 0.6 % in the Chicago dataset and 0.7 % in the Los Angeles dataset. The authors of [22], suggested a system called Crime Situation Awareness Network (CSAN) to predict future crime situations by utilizing multi-correlations and sequential context information. To achieve this, they developed a new neural structure consisting of a Conv-VAE information compression component and a Context-based Sequence Generative Model temporal component. Data preparation included creating detailed Crime Situation Awareness Graphs (CSAGs) and conducting statistical analysis. The performance of the CSAN was measured using metrics like RMSE, MSPE, and JS. In order to predict a person's likelihood of committing a crime and the type of crime they are most likely to commit based on their criminal charge history data, Chun et al. used deep neural networks (DNNs) as a machine-learning technique [23]. "Deep inception-residual networks (DIRNet)" were proposed by Ye et al. (2021) to forecast fine-grained theft-related crimes using a non-emergency service request data (311 events). The method involves identifying low-level spatiotemporal correlations from crime events and complaint records in the 311 dataset using inception units made up of asymmetrical convolution layers. Data from New York City's 311 system and theft-related offences from 2010 to 2015 are used to assess DIRNet's performance. The findings indicate that DIRNet achieves an average F1 score of 71 %, which outperforms other prediction models [3]. However, improved deep learning will produce better results for efficient policing.

The review discusses several approaches to crime prediction, including a crime prediction model that uses hotspot analysis to improve accuracy. The model has three phases: Crime Hotspot Identification, Dataset Preparation, and Crime Prediction Approach. The results show that the proposed model significantly improves the accuracy of crime prediction. Other approaches discussed in the review include an Extended Fuzzy C-means (EFCM) spatiotemporal hot spot detection algorithm, a genetic-fuzzy system, a Geographic-Semantic Ensemble Neural Network (GSEN), and an XGBoost classifier. Each method has its own advantages and disadvantages. The effectiveness of these models is assessed using different evaluation metrics, and compared to traditional methods which other studies failed to do. Hence, our study proposes an improved deep learning model which further improve the accury of the prediction model compared with the traditional methods. Finally, the researchers came to the conclusion that by incorporating dynamic variables over a wide range of criminal occurrences and with the high growth of spatiotemporal crime datasets, crime prediction performance might be greatly enhanced for better policing.

3 PROPOSED MODEL

We have covered the suggested spatiotemporal crime prediction approach in this section. The "Convolutional Neural Network (CNN)" and the "Extreme Gradient Boosting (XGBoost)" classifier, which are the components needed to make predictions in the proposed model is presented. The proposed model is depicted in Figure 1. In this study, Deep Convolutional Extreme Gradient Boosting (DeCXG-

Boost) model, which combines these two models, is proposed and will be used to predict crime spatiotemporally.

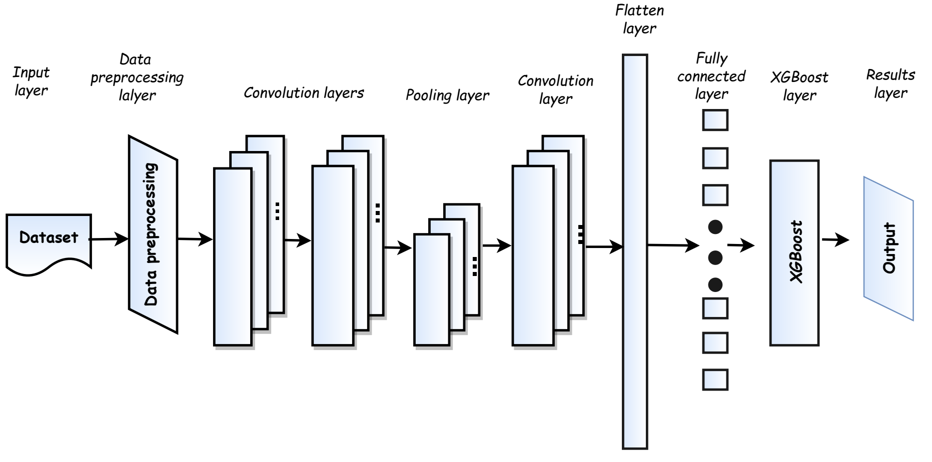

Figure 1. The proposed frame work

3.1 Convolutional Neural Networks (CNNs)

A deep, feed-forward neural network is known as a CNN [24] that is frequently used to analyse visual imagery [25]. The traditional form of CNNs is the classic multi-layer perceptron (MLP). Despite the fact that CNNs were not designed expressly for non-image data, they have been widely used in spatiotemporal data-mining applications including trajectory and spatiotemporal raster data [26]. Figure 2 depicts the architecture of a CNNs.

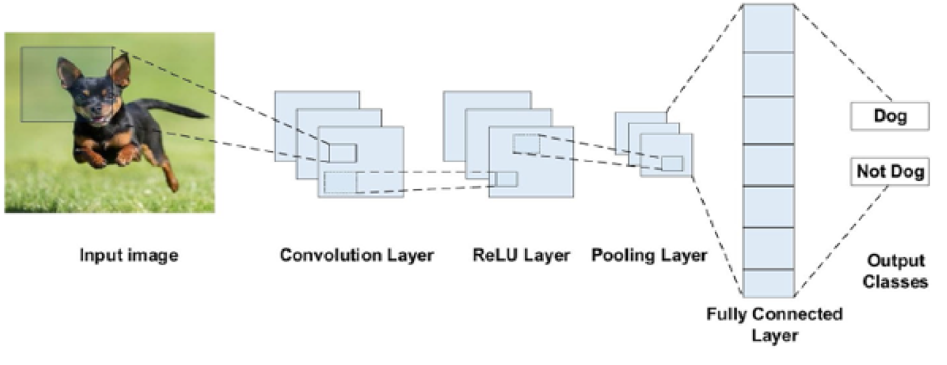

Figure 2. Architecture of CNN [27]

The feature outcome map is created by convolving a one-dimensional entry $x = (x_t)_{N-1}^{t=0}$ of size N in the first layer with a set of M_1 3-dimensional filters, w_h^1 for $h = 1; \ldots; M_1$, for which the filters are applied to all input channels [28]: see Equation (1).

$$a^1(i,h) = \left(w_h^1 \times x\right)(i) = \sum_{j=\infty}^{\infty} w_h^1(j) \times (i-j), \tag{1}$$

where $w_h^1 \in R^{1 \times k \times 1}$ and $a^1 \in R^{1 \times N-1+1 \times M_1}$.

There'll be a single input pathway, and the first layer's output will be routed via the non-linear activation function $h(\cdot)$ to produce $f^1 = h(a^1)$.

A convolutional layer, a pooling layer, and a fully connected layer make up the hidden layer. Using learnable filters, the convolutional layer harvests information from various parts of the raw input or intermediate feature maps autonomously [29]. The pooling layer adds all of the items in the pooling frame together. This approach uses a max-pooling operation to reduce the dimensionality of the input tier by selecting the highest value from each subregion of the preceding layer [29]. Consequently, this level lowers the learning process's computing cost and handles any overfitting difficulties [30].

In [31], shows that, the hidden layer $l = 2; \ldots; L$, the input feature map $f^{l-1} \in R^{1 \times N_{l-1} \times M_{l-1}}$, where $1 \times N_{l-1} \times M_{l-1}$ is the size of the output filter map from the previous convolution with $N_{l-1} = N_{l-2} - k + 1$, is convolved with a set of M_1 filters $w_h^1 \in R^{1 \times k \times M_{l-1}}$, $h = 1; \ldots; M_1$, to create a feature map $a^1 \in R^{1 \times N_l \times M_l}$ as follows in Equation (2) [28].

$$a^1(i,h) = \left(w_h^l \times f^{l-1}\right)(i) = \sum_{j=\infty}^{\infty} \sum_{m=1}^{m_{l-1}} w_h^l(j,m) f^{l-1}(i-j,m). \tag{2}$$

To create the expected output, the fully connected layer flattens and incorporates the high-level obtained attributes learnt by the convolution layer. The attributes figures are then put into $f^1 = h(a^1)$ using non-linear activation functions.

After L convolutional layers, the network produces the matrix f^L, whose size is determined by the filter size and figure of filters employed in the last layer [28]. In a nutshell, the full connected layers acquire the mid and low-level characteristics and generate the high-level abstraction, that represents the final-level layers, just like in a traditional neural network. The classification scores are provided by the last layer (example SVM, etc.). Every score represents the likelihood of a particular class in a given situation [27]. In our study we chose the xgboost classifier on the last-stage layer.

3.2 XGBoost Classification Algorithm

The XGBoost discussed in [32] was created using a GBDT (Gradient Boosting Decision Tree), and it was shown to have excellent convergence and generalisation speed [33]. In [33], the XGBoost algorithm's goal function and optimization strategy were introduced. XGBoost's target function is given by Equation (3) [34].

$$Obj(\theta) = L(\theta) + \Omega(\theta), \tag{3}$$

where $L(\theta) = l(y_i', y_i)$ and $\Omega(\theta) = \gamma T + \frac{1}{2}\lambda||\omega||^2$.

The objective function is divided into 2 sections: $L(\theta)$ and $\Omega(\theta)$, which correspond to the formula's numerous parameters. The difference between the forecast y_i and the target y_i is measured by $L(\theta)$, a differentiable convex loss function. The point is to demonstrate how we can incorporate the facts into the framework [34]. Convex loss functions that are frequently employed, such as the mean square loss function in Equation (4) and the Logistic loss function shown in Equation (5), can be employed in the following equation.

$$l(y_i', y_i) = (y_i' - y_i)^2, \tag{4}$$

$$l(y_i', y_i) = y_i \ln\left(1 + e^{-y_i'}\right) + (1 + y_i)\ln\left(1 + e^{y_i'}\right). \tag{5}$$

Complex models are penalised by the regularised term $\Omega(\theta)$. T is the number of leaves in the tree, and y is the learning rate, which ranges from 0 to 1. When multiplied by T, it equals spanning tree pruning, which prevents overfitting. When compared to the classic GBDT algorithm, the XGBoost algorithm increases the term $\frac{1}{2}\lambda||\omega||^2$. The regularized parameter is λ, while w is the weight of the leaves. This item's value can be increased to control the model from fitting and to improve its generalisation capabilities. The inclusion of model penalty items with functions as parameters, on the other hand, leads in the failure of classical approaches to be optimised by the objective function in Equation (3). As a result, we must assess if we can to learn to obtain the aim y_i as seen in Equation (6) [34]:

$$L(\theta) = \sum_{i=1}^{n} l\left(y_i, y_i'^{t-1} + S_t(T_i)\right) + \Omega(\theta), \tag{6}$$

where, in the t iteration, $S_t(T_i)$ denotes the tree produced by instance i.

The optimization target in each iteration is to build a tree design that minimises the aimed function. Hence, when solving square loss function, the objective function of Equation (6) is optimal, but it becomes quite difficult when calculating other loss functions. As a result, Equation (6) translates Equation (7) using the two-order Taylor expansion, allowing further loss functions to be solved.

$$L(\theta) = \sum_{i=1}^{n} \left[l\left(y_i, y_i'^{t-1} + g_i S_t(T_i)\right) + \frac{1}{2} h_i S_t^2(T_i) \right] + \Omega(\theta), \tag{7}$$

where, $g_i = \partial_{(y')}^{t-1} l\left(y_i, y'^{t-1}\right)$ which is the 1[st] derivative of the error function and $h_i = \partial_y^{'(t-1)2} l\left(y_i, y'^{t-1}\right)$ is the 2[nd] derivative of the error function.

Because tree model needs to find the best segmentation points and then store them in a number of blocks, the algorithm ranks the eigenvalues based on the realisation of XGBoost. This structure is reused in subsequent iterations, resulting in a significant reduction in computing complexity. Furthermore, the information gain of each feature must be determined during the node splitting process, which employs the greed algorithm as shown in Algorithm 1, allowing the calculation of information gain to be parallelized [33].

Algorithm 1 Split finding greed algorithm

Require: Input I, current node's instance set
Ensure: Input d, dimension of the characteristic
  $gain \leftarrow 0$
  $G \leftarrow \sum_i \in I_{g_i}$
  for $k = 1$ to m **do**
    $G_L \leftarrow 0$
    for j in $sorted(I, by\ X_j k)$ **do**
      $G_L \leftarrow G_L + g_j$
    end for
  end for
  $G_L \leftarrow G_L + g_j$
  $G_R \leftarrow G - G_L$
  `Result: Split with max score`

In view of the above overviews of the CNN and XGBoost models, we therefore proposed an improved deep learning model for spatiotemporal crime prediction called DeCXGBoost. The DeCXGBoost model combines two machine learning algorithms, Convolutional Neural Network (CNN) and eXtreme Gradient Boosting (XGBoost), to improve the accuracy of prediction tasks. The CNN algorithm is used to extract high-level features from raw input data, such as images or time-series data. It involves several convolutional layers that perform operations on the input data to extract features and a pooling layer that reduces the dimensionality of the output. The output from the convolutional and pooling layers is then fed into a fully connected layer, which performs classification or regression. The XGBoost algorithm is a gradient boosting framework that is used for supervised learning problems. It builds a series of decision trees iteratively, with each new tree correcting the errors made by the previous one. The DeCXGBoost model combines these two algorithms to leverage their respective strengths. The CNN algorithm is used to extract high-level features from the raw input data, which are then fed into the XGBoost algorithm to make predictions. The combination of these two algorithms allows the model to extract complex features from the input data and make accurate predictions.

Mathematically, the DeCXGBoost model can be represented as follows: CNN: The output of the ith convolutional layer is given by:

$$O_i = f_i(w_i * O(i-1) + b_i), \tag{8}$$

where w_i is the ith set of convolutional filters, $O(i-1)$ is the output of the previous layer, b_i is the bias term, and f_i is the activation function.

XGBoost: The output of the model is given by:

$$Y = F(X) = \sum f_t(X), \tag{9}$$

where X is the input data, f_t is the t^{th} decision tree, and $\sum$ is the sum over all decision trees.

DeCXGBoost: The DeCXGBoost model combines the two algorithms by using the output of the final fully connected layer in the CNN as input to the XGBoost algorithm:

$$Y^* = F(X) = \sum f_t(O_L), \tag{10}$$

where O_L is the output of the final fully connected layer in the CNN, and f_t is the t^{th} decision tree in the XGBoost algorithm. Overall, the DeCXGBoost model is a powerful machine learning algorithm that can be used for a wide range of prediction tasks, particularly in areas that involve complex input data such as images and time-series data.

3.3 Dataset

The Boston Police Department's (BPD) criminal event records were employed in this study that documented the incidents to which BPD officers respond. This was a collection of data from the new crime incident reports, which was designed to capture the sort of incidents as well as when and where it occurred. Table 1 shows the attributes of the crime dataset.

	Incident-Number	Offense-Code	Offense-Desc	...	Street	Lat
0	I182070945	619	Vandalism	...	Lincoln ST	42.35779134
1	I182070915	614	Auto theft	...	Hecla ST	42.30682138
2	I182070893	613	Verbal dispute	...	Dehil ST	42.32701648
...	...	...	...	...	...	...
319071	I030217815-08	1843	Larceny	...	Capen ST	42.28647012
319072	I030217815-08	301	Harassment	...	Lawn ST	42.3256949
319073	I010370257-00	3801	Trespassing	...	Hecla ST	42.31731905
	[319074 × 17]					

Table 1. Features for the crime dataset

This spatiotemporal crime dataset consists of seventeen (17) features (columns) and three hundred nineteen thousand and seventy-four (319074) samples (rows).

There are eight categories (Incident number, offense code group, offense description, district, occurred on date, day of week, UCR part, street) and nine numerical (offense code, reporting area, shooting, year, month, hour, lat, long, location) qualities.

3.4 Data Preprocessing

Data cleansing is a process that must be completed prior to data analysis. It entails tasks including filling in missing data, removing discrepancies, and finding outliers [35]. One of the most significant transformations to perform to data is feature scaling. The numeric features employed in the input should not have different scales for ML algorithms to perform properly [35]. As a result, the min/max normalisation approach was used to rescale the data set so that the values on distinct scales in the data set varied in the range of 0–1. The following formula (see Equation (11)) is used to translate a value that falls within the range of 0 to 1.

$$x_{new} = \frac{x - \min(x)}{\max(x) - \min(x)}. \tag{11}$$

3.5 Modeling

The model was trained and tested for the study, so the dataset was split in half in a 75:25 ratio for the models, 75 % of the dataset was utilized to train the model, while 25 % was used to test it. The process of modeling was carried out using the proposed methodology depicted in Figure 3.

4 EXPERIMENTAL RESULT AND ANALYSIS

In this paper, the Boston Police Department's (BPD) crime dataset is used to extract the features of the larceny crime data which we used in this study for analyzing and predicting crime as a type of crime, which is one of our objectives and motivations for this study. According to [36], there exist eight distinct categories of larceny offenses. Out of the eight categories, six are categorized as "non-occupational" offenses, which involve crimes like shoplifting, theft from a vehicle, theft of vehicle parts, pocket-picking, purse snatching, and theft from a coin-operated device. The two other categories are "theft from a building" and "all other larcenies", which are partially categorized as non-occupational and partially as indeterminate. In our study we categorized them as larceny and larceny from motor vehicle.

4.1 Exploratory Data Analysis (EDA)

A script was ran to investigates numerous distinct categories of larceny offences in the dataset, which we classified into two crime categories as previously described. The distribution of crime is depicted in Figure 4. Larceny is the most common sort of crime, followed by larceny from motor vehicles, as shown in Figure 4 below.

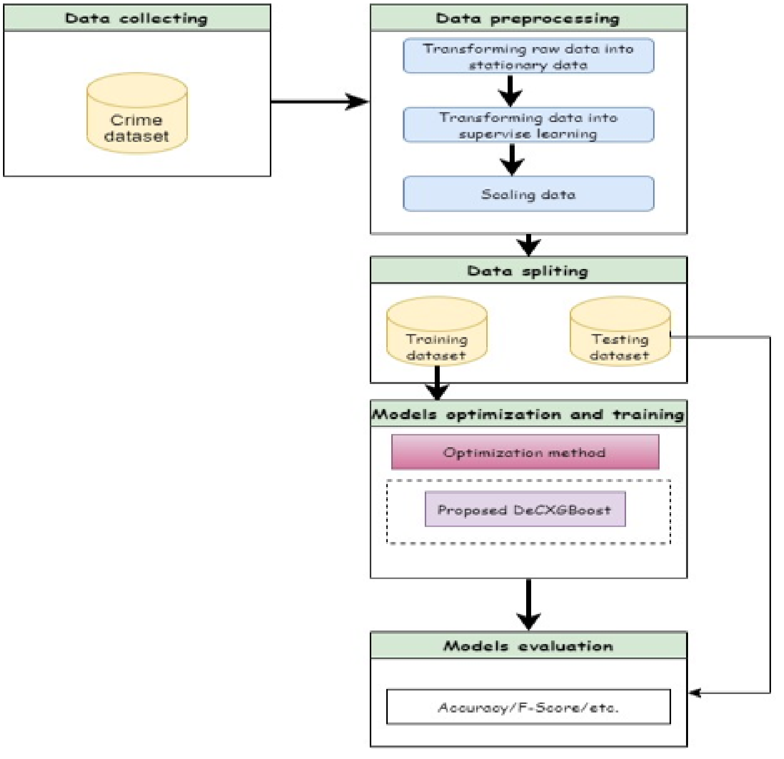

Figure 3. Flow diagram of modeling

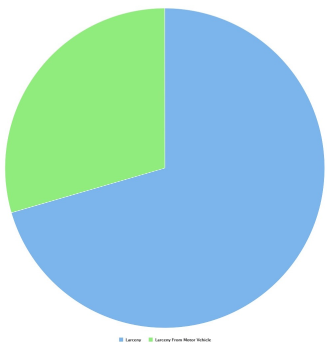

Figure 4. Larceny crime dataset distribution

The extracted crime dataset has 36 782 crime observations, 25 935 of this crimes are committed in shoplifting, pocket picking, and 10847 of this crimes were committed from motor vehicles. Figure 5 shows the hourly committing of these crimes, with larceny out of motor vehicle mostly committed.

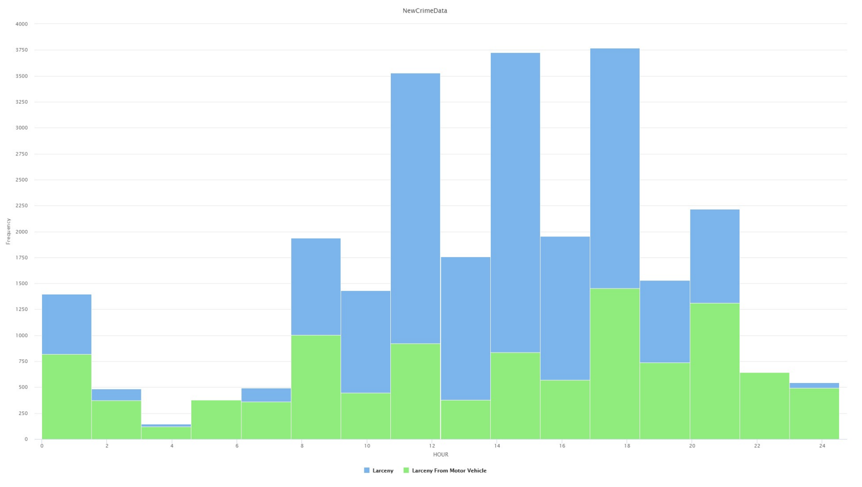

Figure 5. Hourly distribution of crimes committed

Our spatiotemporal crime dataset also displays the locations where the crimes were committed, as seen in Figure 6. It shows that more crimes were committed on Lincoln Street and Lime Street on Wednesdays, Saturdays, and other days. Additionally, Figure 7 shows the districts and the crimes committed on a weekly basis. District D4 has more crimes committed on Fridays, Saturdays, and Wednesdays. With this information and the aid of accurate crime prediction models, security personnel can be more proactive rather than reactive, resulting in a significant reduction in crime within society.

4.2 Prediction Models

The buildup and results of our proposed deep learning improved model is presented, with a comparison to other state-of-the-art models such as Naïve Bayes (NB), Stacking (STK), Random Forest (RF), Lazy:IBK (IBK), Bagging (BAG), Support Vector Machine (SVM), CNNs, and Locally weighted learning (LWL). Our suggested model, as well as the other eight models in this work, were trained and presented with a variety of setting parameters and feature choices. Both time-related and geographic variables are essential, according to the data exploration section, which explains the spatiotemporal interest. All of the models were trained and tested for the study, so the dataset was split in a 75:25 ratio for all of the models. The model was trained on 75 % of the dataset and tested on the remaining 25 %, as previously mentioned. The

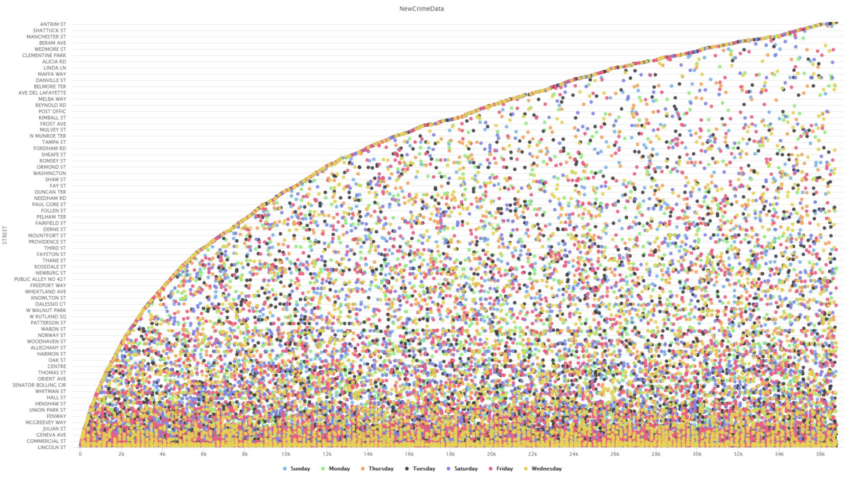

Figure 6. Weekly distribution of crime on streets

CNNs models were constructed using python-3.8, tensorflow-1.01, and keras-1.0 on an Intel core i5 desktop computer. In our proposed model, samples were provided as input. Batch normalization technology and ReLU activation functions were used in all convolution layers. The deep learning model used binary cross-entropy and adaptive moment estimation (Adam) methods as the loss function and optimizer, respectively. Additionally, the He initialization approach was used to initialize the model. WEKA tool was used for the other models. Figure 8 depicts our proposed

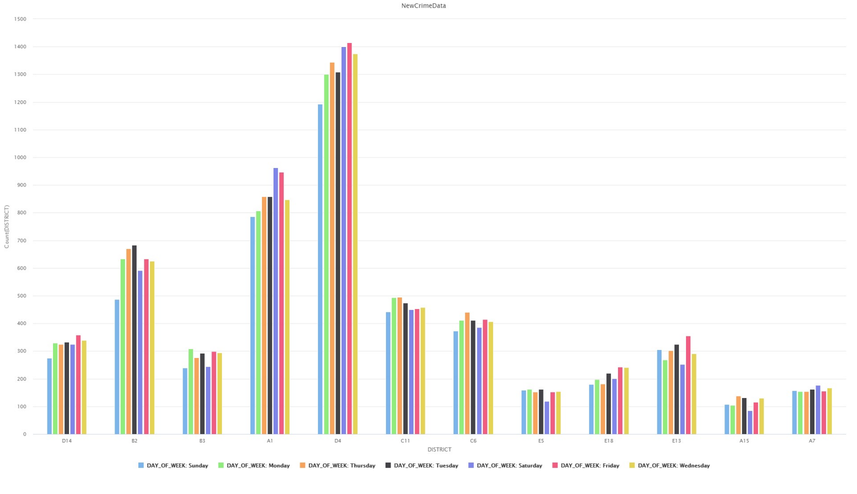

Figure 7. Weekly larceny crimes committed in districts

model's training and testing loss logs, respectively. Figure 9 depicts our suggested model's accuracy log for both training and testing.

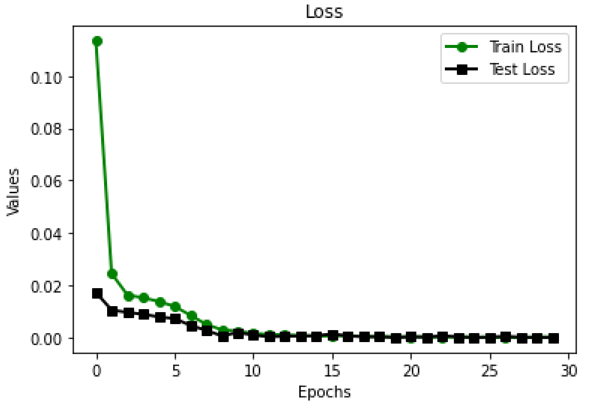

Figure 8. Proposed model loss log

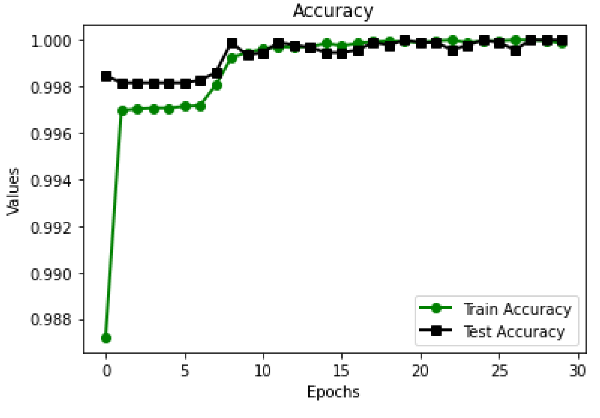

Figure 9. Proposed model accuracy log

The study evaluates the performance of multiple models using various metrics, such as MAE, RMSE, Recall, Accuracy, and F-Measure. The outcomes of these models are shown in Table 2. MAE measures the average absolute deviation between the predicted and actual values, while RMSE is the square root of the average squared deviation between the predicted and actual values. Recall is used to determine the percentage of actual positive cases that were correctly identified by the model. Accuracy is the proportion of correct predictions made by the model, and F-Measure is the harmonic mean of Precision and Recall. The results show that several models achieved high scores in various performance measures. The NB, RF, BAG, and our proposed model achieved perfect scores in Recall and F-Measure, indicating that they have correctly identified all the positive cases. The CNN model has a relatively low F-Measure score, indicating that it may not perform well in identifying positive cases. The LWL model has the highest RMSE, indicating that it has the largest errors in its predictions. Overall, our proposed model outperformed all the other models, achieving perfect scores in both Recall and F-Measure and the lowest MAE and RMSE scores. The results of this study suggest that the proposed model is highly effective in predicting outcomes in the studied domain.

Models	MAE	RMSE	Recall	Accuracy	F-Measure
NB	0.0005	0.0134	1.000	99.782	1.000
STK	0.4178	0.4592	0.698	69.8097	0.822
RF	0.0629	0.084	1.000	99.891	1.000
IBK	0.0481	0.2192	0.952	95.193	0.952
BAG	0.0001	0.0076	1.000	99.891	1.000
SVM	0.0005	0.0233	0.999	99.7456	0.999
CNN	0.0004	0.0209	1.000	99.565	0.846
OUR's	0.0000	0.0001	1.000	100.000	1.000
LWL	0.913	0.2161	0.946	94.6166	0.945

Table 2. Results of models used in the study

On our well-preprocessed crime dataset with hyperparameter settings and feature selections, Figure 10 illustrates a comparison of the MAE and RMSE of our proposed model and various other models employed in this study. The proposed model is seen to edging out the other models significantly. This is because the lower or small the figure of MAE and RMSE the greater the model.

Figure 11 shows the RECALL and F-Measure of our suggested model vs. the RECALL and F-Measure of other models used in this study. The proposed model appears to greatly outperform the other models. Also Figure 12 compares the prediction accuracy of the proposed model and the other eight models used in this study. The accuracy of our proposed model was higher.

4.3 Discussion

Our proposed (DeCXGBoost) model achieved the best performance across all metrics, with perfect scores in Recall, Accuracy, and F-Measure (see Table 2). The SVM, NB, BAG, and RF models also performed well, achieving high scores in Recall, Accuracy, and F-Measure (see Figures 11 and 12). In contrast, the STK and LWL models had relatively poor performances, with higher MAE and RMSE scores (see Figure 10), and lower Recall and F-Measure scores. The CNN model achieved a high Recall score but had a lower F-Measure score compared to the other models. The use of machine learning algorithms for spatiotemporal crime prediction has been an active area of research in recent years. Various machine

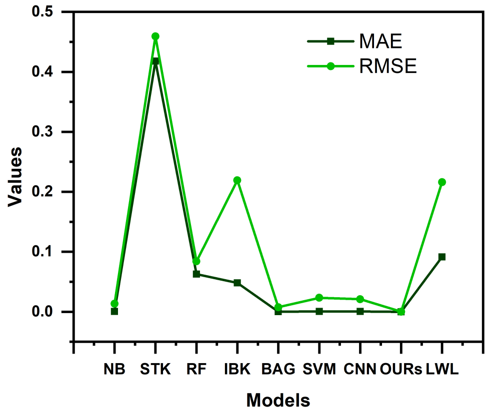

Figure 10. Comparison of our proposed model's performance to the other models using MAE and RMSE

learning algorithms, including deep learning and ensemble methods, have been employed to improve the accuracy of spatiotemporal crime prediction models. One recent study [4] proposed the use of a spatiotemporal convolutional neural network (ST-CNN) to predict crime incidents based on spatiotemporal data. The model used a combination of convolutional neural network and long short-term memory networks to capture both spatial and temporal patterns in crime incidents. The study achieved an accuracy of 86 % in predicting crime incidents. An-

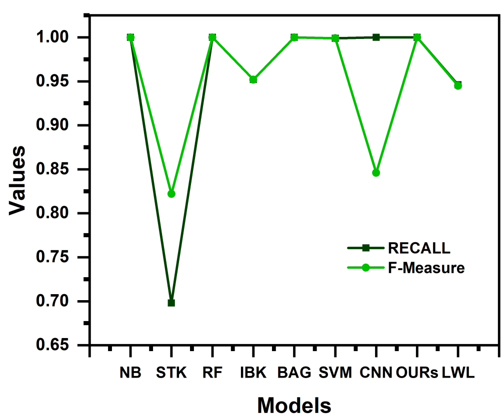

Figure 11. Comparison of our proposed model's performance to the other models using Recall and F-Measure

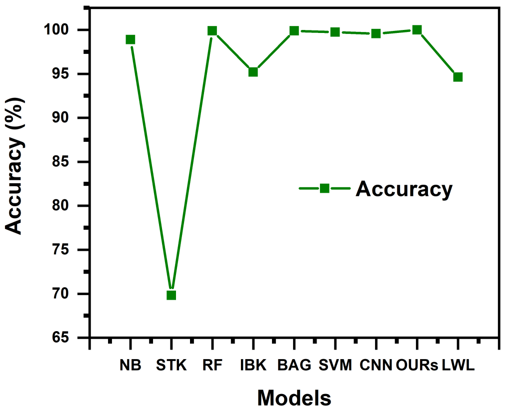

Figure 12. Comparison of our proposed model's performance to the other models using Accuracy

other study [37] used a deep learning approach, specifically the Gated Recurrent Unit (GRU), to predict crime incidents in Los Angeles and Chicago. The model incorporated weather data and social media data in addition to spatiotemporal data to improve its predictions. The study achieved an accuracy of 83.9 % and 86.3 % in predicting crime incidents. Ensemble models, which combine multiple machine learning algorithms, have also been used in spatiotemporal crime prediction. One study [12] proposed an ensemble random forest algorithm to predict crime incidents. The model achieved an accuracy of 99.16 % in predicting crime incidents.

The results of our proposed (DeCXGBoost) model in this study demonstrate the effectiveness of machine learning models in spatiotemporal crime prediction (see Table 2 and Figure 12). The DeCXGBoost model also has the lowest Mean Absolute Error (MAE), and Root Mean Square Error (RMSE) of 0.0000 and 0.0001 respectively making the model the best and more robust when compared to other baseline models used in this study. The study can inform the development of better models and algorithms in the future how to improve the accuracy and efficiency of spatiotemporal crime prediction. Also, the findings can inform the development of better models for prediction tasks in related fields, potentially leading to improvements in various applications, such as healthcare, finance, and cybersecurity.

In summary, this study evaluated the performance of various machine learning models for spatiotemporal crime prediction. The proposed DeCXGBoost model achieved the best performance across all metrics, with perfect scores in Recall, Accuracy, and F-Measure. Other models like SVM, NB, BAG, and RF also performed well. In contrast, the STK and LWL models had relatively poor performances, with higher MAE and RMSE scores and lower Recall and F-Measure scores. The CNN model achieved a high Recall score but had a lower F-Measure score compared to the other models.

The study also discussed other studies that employed machine learning algorithms for spatiotemporal crime prediction, such as the use of spatiotemporal convolutional neural network (ST-CNN) and the Gated Recurrent Unit (GRU) models. The proposed DeCXGBoost model outperformed these models, achieving a perfect score across all metrics.

The results of this study demonstrate the effectiveness of machine learning models in spatiotemporal crime prediction, and the proposed DeCXGBoost model is highly robust and accurate when compared to other baseline models. The study provides valuable insights into the development of better models and algorithms in the future to improve the accuracy and efficiency of spatiotemporal crime prediction.

5 CONCLUSIONS

This article analyses the outcomes of extracted larceny crime data from the Boston crime dataset and presents exploratory data analysis with a novel proposed spatiotemporal crime prediction model based on classification approaches. Python was used to implement the proposed model. The experimental findings reveal that our suggested DeCXGBoost model outperformed other crime categorization models for all eight methods. For both accuracy and recall, our proposed model received a perfect score. Our proposed methodology can help law enforcement agencies fight crime more effectively, channel resources more efficiently, foresee crime to some extent and serve society. The presented proposed crime prediction model can be used to make predictions and manage resources on any dataset or criminal data. For future improvement, real time crime prediction is an open direction for this work with more advanced technologies.

REFERENCES

[1] ToppiReddy, H. K. R.—Saini, B.—Mahajan, G.: Crime Prediction and Monitoring Framework Based on Spatial Analysis. Procedia Computer Science, Vol. 132, 2018, pp. 696–705, doi: 10.1016/j.procs.2018.05.075.

[2] Matijosaitiene, I.—Zhao, P.—Jaume, S.—Gilkey Jr, J. W.: Prediction of Hourly Effect of Land Use on Crime. ISPRS International Journal of Geo-Information, Vol. 8, 2019, No. 1, Art. No. 16, doi: 10.3390/ijgi8010016.

[3] Ye, X.—Duan, L.—Peng, Q.: Spatiotemporal Prediction of Theft Risk with Deep Inception-Residual Networks. Smart Cities, Vol. 4, 2021, No. 1, pp. 204–216, doi: 10.3390/smartcities4010013.

[4] Esquivel, N.—Nicolis, O.—Peralta, B.—Mateu, J.: Spatio-Temporal Prediction of Baltimore Crime Events Using CLSTM Neural Networks. IEEE Access, Vol. 8, 2020, pp. 209101–209112, doi: 10.1109/ACCESS.2020.3036715.

[5] Wang, H.—Kifer, D.—Graif, C.—Li, Z.: Crime Rate Inference with Big Data. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16), 2016, pp. 635–644, doi: 10.1145/2939672.2939736.

[6] Huang, C.—Zhang, C.—Zhao, J.—Wu, X.—Yin, D.—Chawla, N.: MiST: A Multiview and Multimodal Spatial-Temporal Learning Framework for Citywide Abnormal Event Forecasting. The World Wide Web Conference (WWW '19), 2019, pp. 717–728, doi: 10.1145/3308558.3313730.

[7] Yao, H.—Tang, X.—Wei, H.—Zheng, G.—Li, Z.: Revisiting Spatial-Temporal Similarity: A Deep Learning Framework for Traffic Prediction. Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 33, 2019, No. 1, pp. 5668–5675, doi: 10.1609/aaai.v33i01.33015668.

[8] Zhang, J.—Zheng, Y.—Qi, D.: Deep Spatio-Temporal Residual Networks for Citywide Crowd Flows Prediction. Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 31, 2017, No. 1, doi: 10.1609/aaai.v31i1.10735.

[9] Zhang, X.—Huang, C.—Xu, Y.—Xia, L.—Dai, P.—Bo, L.—Zhang, J.—Zheng, Y.: Traffic Flow Forecasting with Spatial-Temporal Graph Diffusion Network. Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 35, 2021, No. 17, pp. 15008–15015, doi: 10.1609/aaai.v35i17.17761.

[10] Xia, L.—Huang, C.—Xu, Y.—Dai, P.—Bo, L.—Zhang, X.—Chen, T.: Spatial-Temporal Sequential Hypergraph Network for Crime Prediction with Dynamic Multiplex Relation Learning. Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence (IJCAI-21), 2021, pp. 1631–1637, doi: 10.24963/ijcai.2021/225.

[11] Yang, B.—Liu, L.—Lan, M.—Wang, Z.—Zhou, H.—Yu, H.: A Spatio-Temporal Method for Crime Prediction Using Historical Crime Data and Transitional Zones Identified from Nightlight Imagery. International Journal of Geographical Information Science, Vol. 34, 2020, No. 9, pp. 1740–1764, doi: 10.1080/13658816.2020.1737701.

[12] Hossain, S.—Abtahee, A.—Kashem, I.—Hoque, M. M.—Sarker, I. H.: Crime Prediction Using Spatio-Temporal Data. In: Chaubey, N., Parikh, S., Amin, K. (Eds.): Computing Science, Communication and Security (COMS2 2020). Springer, Singapore, Communications in Computer and Information Science, Vol. 1235, 2020, pp. 277–289, doi: 10.1007/978-981-15-6648-6_22.

[13] Mushtaq, H.—Siddique, I.—Malik, B. H.—Ahmed, M.—Butt, U. M.—Ghafoor, R. M. T.—Zubair, H.—Farooq, U.: Educational Data Classification Framework for Community Pedagogical Content Management Using Data Mining. International Journal of Advanced Computer Science and Applications, Vol. 10, 2019, No. 1, pp. 329–338, doi: 10.14569/IJACSA.2019.0100144.

[14] Brayne, S.—Christin, A.: Technologies of Crime Prediction: The Reception of Algorithms in Policing and Criminal Courts. Social Problems, Vol. 68, 2021, No. 3, pp. 608–624, doi: 10.1093/socpro/spaa004.

[15] Hajela, G.—Chawla, M.—Rasool, A.: A Clustering Based Hotspot Identification Approach for Crime Prediction. Procedia Computer Science, Vol. 167, 2020, pp. 1462–1470, doi: 10.1016/j.procs.2020.03.357.

[16] Cardone, B.—Di Martino, F.: Fuzzy-Based Spatiotemporal Hot Spot Intensity and Propagation — An Application in Crime Analysis. Electronics, Vol. 11, 2022, No. 3, Art. No. 370, doi: 10.3390/electronics11030370.

[17] Di Martino, F.—Sessa, S.—Barillari, U. E. S.—Barillari, M. R.: Spatio-Temporal Hotspots and Application on a Disease Analysis Case via GIS. Soft Computing, Vol. 18, 2014, No. 12, pp. 2377–2384, doi: 10.1007/s00500-013-1211-7.

[18] Farjami, Y.—Abdi, K.: A Genetic-Fuzzy Algorithm for Spatio-Temporal Crime Prediction. Journal of Ambient Intelligence and Humanized Computing, 2021, pp. 1–13, doi: 10.1007/s12652-020-02858-3.

[19] Jin, G.—Sha, H.—Feng, Y.—Cheng, Q.—Huang, J.: GSEN: An Ensemble Deep Learning Benchmark Model for Urban Hotspots Spatiotemporal Prediction. Neurocomputing, Vol. 455, 2021, pp. 353–367, doi: 10.1016/j.neucom.2021.05.008.

[20] Ahmed, S.—Gentili, M.—Sierra-Sosa, D.—Elmaghraby, A. S.: Multi-Layer Data Integration Technique for Combining Heterogeneous Crime Data. Information Processing and Management, Vol. 59, 2022, No. 3, Art. No. 102879, doi: 10.1016/j.ipm.2022.102879.

[21] Ilhan, F.—Tekin, S. F.—Aksoy, B.: Spatio-Temporal Crime Prediction with Temporally Hierarchical Convolutional Neural Networks. 2020 28th Signal Processing and Communications Applications Conference (SIU), 2020, pp. 1–4, doi: 10.1109/SIU49456.2020.9302169.

[22] Wang, Q.—Jin, G.—Zhao, X.—Feng, Y.—Huang, J.: CSAN: A Neural Network Benchmark Model for Crime Forecasting in Spatio-Temporal Scale. Knowledge-Based Systems, Vol. 189, 2020, Art. No. 105120, doi: 10.1016/j.knosys.2019.105120.

[23] Chun, S. A.—Avinash Paturu, V.—Yuan, S.—Pathak, R.—Atluri, V.—Adam, N. R.: Crime Prediction Model Using Deep Neural Networks. Proceedings of the 20th Annual International Conference on Digital Government Research (dg.o 2019), ACM, 2019, pp. 512–514, doi: 10.1145/3325112.3328221.

[24] Krizhevsky, A.—Sutskever, I.—Hinton, G. E.: ImageNet Classification with Deep Convolutional Neural Networks. Communications of the ACM, Vol. 60, 2017, No. 6, pp. 84–90, doi: 10.1145/3065386.

[25] Gao, N.—Xue, H.—Shao, W.—Zhao, S.—Qin, K. K.—Prabowo, A.—Rahaman, M. S.—Salim, F. D.: Generative Adversarial Networks for Spatio-Temporal Data: A Survey. ACM Transactions on Intelligent Systems and Technology (TIST), Vol. 13, 2022, No. 2, Art. No. 22, doi: 10.1145/3474838.

[26] Prabowo, A.—Koniusz, P.—Shao, W.—Salim, F. D.: COLTRANE: ConvolutiOnaL TRAjectory NEtwork for Deep Map Inference. Proceedings of the 6th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation (BuildSys '19), 2019, pp. 21–30, doi: 10.1145/3360322.3360853.

[27] Alzubaidi, L.—Zhang, J.—Humaidi, A. J.—Al-Dujaili, A.—Duan, Y.—Al-Shamma, O.—Santamaría, J.—Fadhel, M. A.—Al-Amidie, M.—Farhan, L.: Review of Deep Learning: Concepts, CNN Architectures, Challenges, Applications, Future Directions. Journal of Big Data, Vol. 8, 2021, Art. No. 53, doi: 10.1186/s40537-021-00444-8.

[28] Borovykh, A.—Bohte, S.—Oosterlee, C. W.: Conditional Time Se-

ries Forecasting with Convolutional Neural Networks. CoRR, 2017, doi: 10.48550/arXiv.1703.04691.

[29] Zuo, R.—Xiong, Y.—Wang, J.—Carranza, E. J. M.: Deep Learning and Its Application in Geochemical Mapping. Earth-Science Reviews, Vol. 192, 2019, pp. 1–14, doi: 10.1016/j.earscirev.2019.02.023.

[30] Hoseinzade, E.—Haratizadeh, S.: CNNpred: CNN-Based Stock Market Prediction Using a Diverse Set of Variables. Expert Systems with Applications, Vol. 129, 2019, pp. 273–285, doi: 10.1016/j.eswa.2019.03.029.

[31] Barzegar, R.—Aalami, M. T.—Adamowski, J.: Short-Term Water Quality Variable Prediction Using a Hybrid CNN-LSTM Deep Learning Model. Stochastic Environmental Research and Risk Assessment, Vol. 34, 2020, No. 2, pp. 415–433, doi: 10.1007/s00477-020-01776-2.

[32] Chen, T.—Guestrin, C.: XGBoost: A Scalable Tree Boosting System. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16), 2016, pp. 785–794, doi: 10.1145/2939672.2939785.

[33] Jiang, Y.—Tong, G.—Yin, H.—Xiong, N.: A Pedestrian Detection Method Based on Genetic Algorithm for Optimize XGBoost Training Parameters. IEEE Access, Vol. 7, 2019, pp. 118310–118321, doi: 10.1109/ACCESS.2019.2936454.

[34] Chen, Z.—Jiang, F.—Cheng, Y.—Gu, X.—Liu, W.—Peng, J.: XGBoost Classifier for DDoS Attack Detection and Analysis in SDN-Based Cloud. 2018 IEEE International Conference on Big Data and Smart Computing (BigComp), 2018, pp. 251–256, doi: 10.1109/BigComp.2018.00044.

[35] Alexandropoulos, S. A. N.—Kotsiantis, S. B.—Vrahatis, M. N.: Data Preprocessing in Predictive Data Mining. The Knowledge Engineering Review, Vol. 34, 2019, Art. No. e1, doi: 10.1017/S026988891800036X.

[36] Steffensmeier, D.—Harris, C. T.—Painter-Davis, N.: Gender and Arrests for Larceny, Fraud, Forgery, and Embezzlement: Conventional or Occupational Property Crime Offenders? Journal of Criminal Justice, Vol. 43, 2015, No. 3, pp. 205–217, doi: 10.1016/j.jcrimjus.2015.03.004.

[37] Sun, J.—Yue, M.—Lin, Z.—Yang, X.—Nocera, L.—Kahn, G.—Shahabi, C.: CrimeForecaster: Crime Prediction by Exploiting the Geographical Neighborhoods' Spatiotemporal Dependencies. In: Dong, Y., Ifrim, G., Mladenić, D., Saunders, C., Van Hoecke, S. (Eds.): Machine Learning and Knowledge Discovery in Databases. Applied Data Science and Demo Track (ECML PKDD 2020). Springer, Cham, Lecture Notes in Computer Science, Vol. 12461, 2021, pp. 52–67, doi: 10.1007/978-3-030-67670-4_4.

Ature Angbera is a current postgraduate student in the School of Computer Sciences, Universiti Sains, Malaysia. He also works with Joseph Sarwuan Tarkaa University Makurdi, Benue State in the Department of Computer Science. His main research fields are deep learning, machine learning, spatiotemporal analysis, stream processing, data mining and intelligent techniques.

Huah Yong Chan is Associate Professor in the School of Computer Sciences, Universiti Sains, Malaysia. His research fields are grid computing, cloud computing, resource allocation, and data analytics.