IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2017.DOI

Spatio-temporal prediction of Baltimore crime events using CLSTM neural networks

NICOLÁS ESQUIVEL[1], ORIETTA NICOLIS[2], BILLY PERALTA[3], AND JORGE MATEU [4]
[1]Universidad Andres Bello, Chile (e-mail: n.esquivelrivera@uandresbello.edu)
[2]Universidad Andres Bello, Chile (e-mail: nicolisorietta@gmail.com)
[3]Universidad Andres Bello, Chile (e-mail: billymark@gmail.com)
[4]Universitat Jaume I of Castellon, Spain (e-mail: mateu@mat.uji.es)

Corresponding author: O. Nicolis (e-mail: orietta.nicolis@unab.cl)

This work was supported in part by the Spanish Grant PID2019-107392RB-I00 and Chilean Fondecyt Grant ID1201478.

ABSTRACT Crime activity in many cities worldwide causes significant damages to the lives of victims and their surrounding communities. It is a public disorder problem, and big cities experience large amounts of crime events. Spatio-temporal prediction of crimes activity can help the cities to have a better allocation of police resources and surveillance. Deep learning techniques are considered efficient tools to predict future events analyzing the behavior of past ones; however, they are not usually applied to crime event prediction using a spatio-temporal approach. In this paper, a Convolutional Neural Network (CNN) together with a Long-Short Term Memory (LSTM) network (thus CLSTM-NN) are proposed to predict the presence of crime events over the city of Baltimore (USA). In particular, matrices of past crime events are used as input to a CLSTM-NN to predict the presence of at least one event in future days. The model is implemented on two types of events: "street robbery" and "larceny". The proposed procedure is able to take into account spatial and temporal correlations present in the past data to improve future prediction. The prediction performance of the proposed neural network is assessed under a number of controlled plausible scenarios, using some standard metrics (Accuracy, AUC-ROC, and AUC-PR).

INDEX TERMS CNN and LSTM neural networks, crime prediction, deep learning, spatial and temporal structure

I. INTRODUCTION

CRIME related problems are at the base of important and crucial issues for many societies living in large cities worldwide. [1] showed that crime and neighborhood disorder may negatively impact the health of urban residents by increasing the resident risk of experiencing violence and impacting their mental health as a form of depression for being in constant contact with assaults, blows and shots. Thus, crime rate reduction is at the core of many local policies driven by active plans supported by police action and local authorities. In this line, the use of mathematical, statistical and/or computational models able to predict crime events beforehand would help the police to generate preventive plans for areas at high risk, and to speed up the process of solving crimes, with the consequent reduction of crime rate. Indeed, a number of studies have been developed for the analysis and prediction of crime events in many parts of the world (see, [2] and the references within), and several cities have made their data available for investigations (these are the cases of Chicago, Seattle, Detroit or Baltimore, to name just a few). Among the proposed solutions to predict crime events, the use of machine learning techniques have gained great importance due to its recent success in solving real world problems. [3] applied linear regression, additive regression and decision trees for the prediction of three types of crimes in the state of Mississippi, obtaining a correlation coefficient of 99%. [4] used five different machine learning algorithms to predict which category of crime is most likely to take place at particular times and space (places) in Chicago. A decision tree model provided the best performance, achieving an accuracy of 99.88%. [5] analyzed crime events in YD county from 2012 to 2015, and applied different prediction models such as a Bayesian networks, Random Trees and Neural networks. The work showed that the best result was obtained with Random Trees, with an accuracy of 97.4%. Some machine learning methods for crime prediction in the town of San Francisco were proposed by [6], where the main aim was to classify a criminal incident by type, depending on its occurrence at a given time and location. The results highlighted AdaBoost as the best classifier together with RandomForest using the SMOTE method, with accuracies of 81.93% and 71.43%, respectively. Due to recent approaches of machine learning in the sub-field of deep learning, new neural network models have been proposed considering spatial and temporal correlations. Typical deep learning models include the Long Short-Term Memory (LSTM) neural networks and the Convolutional Neural Network (CNN). In the context of crime prediction, [7] applied a LSTM for classifying crime incidents related to public safety. By using only five features of the data set published by Chicago city police, they were able to obtain 87.84% of accuracy. [8] applied a Feed Forward Neural Network (FFNN), a Recurrent Neural Network (RNN), a Convolutional Neural Network (CNN) and a mixture of Recurring Networks with Convolutions (RNN + CNN). Using data from Chicago city, the results showed RNN + CNN as the best neural network in terms of prediction, obtaining 75.6% and 65.3% accuracy for Chicago and Portland, respectively. Here, the authors argue that the weather and the amount of traffic influence the prediction by decreasing the accuracy of the models. [9] provided a survey on some data mining techniques based on neural networks used for detection and prediction of crimes. They propose methods consisting of collecting data, classifying and finding patterns, predicting, visualizing and taking actions with the predictions given by neural networks. A similar work using deep learning comes with the application of a recurrent LSTM proposed by [10], where the LSTM layers have been chosen for the ability to store temporal patterns. Different optimization algorithms (such as Adagrad, RMSProp, SGDNesterov optimizers, AdaDelta and Adam) were compared, and Adam resulted with the smaller loss function. In [11] the city of Chicago is divided into 77 communities, where for each community they have social information such as the number of police stations in a sector, number of schools, bookstores, type of crime and calls to 311. This information was then used for running three regression models: a polynomial regression, a self-regressive model, and the support vector regression (SVR). The SVR obtained the best results (i.e. the lowest root mean square error (RMSE)) for predicting the amount of crimes in an area.

With the growth of technology, the local authorities and police departments had to deal with large amounts of data in order to understand criminal patterns. In this context, [12] presented a spatio-temporal crime prediction approach based on spatial auto-regressive models to automatically detect high-risk crime regions in urban areas, and to reliably forecast crime trends in each region. They used crime data in Chicago over the period 2014-2016, and obtained a maximum mean absolute prediction error (MAPE) between 8.7% and 11.9%. A spatio-temporal approach was also proposed by [13] building a spatio-temporal Bayesian model to analyze spatio-temporal patterns of urban crime and determine developing trends. The model was then applied to analyze data regarding burglaries occurred in Wuhan (China) during the first eight months of 2013. Using different socio-economic features, the results showed a strong correlation between the burglary crime rate with the average resident population per community and number of local internet bars.

[14] presented a novel approach using spatio-temporal analysis and a Generalized Linear Model (GLM) for Crime Site Selection (CSS). The authors were able to find the most likely crime location and predict crime trends. The model was applied to data of India regarding vehicle thefts from 2010 to 2015, and out of nine studied districts, the method identified the three districts with higher probability of occurrence of crimes. Recently, a CNN has been used jointly with integro-difference equations (IDE) for a spatio-temporal probabilistic prediction ([15]). In this case, the CNN is used to learn about the parameters governing the dynamics from the most recent behavior of the (partially observed) process.

In this paper, a new approach based on a Convolutional Long-Short Term Memory neural network (CLSTM-NN) is proposed to predict crime events in the town of Baltimore in the period from January 2016 to December 2018. The proposed neural network model contains convolutional and long short term memory layers, using as input a stack of matrices representing the number of crimes spatially distributed in the city of Baltimore on different past days in order to predict the number of crimes of the next day. Each element of the matrices represents the number of crimes within an area defined by its latitude and longitude. In order to test the performance of the model, matrices of size 8×8 were considered jointly with two types of crimes (larceny and street robberies). Different scenarios (considering different number of past days as input data) are evaluated for choosing the best model. Typical metrics (such as accuracy, area under roc and precision-recall curves) are considered for comparing results.

The CLSTM model was proposed for the first time some few years ago by [16] to predict rainfall intensity. It then was applied in different scientific contexts such as travel demand prediction, ([17]), video segmentation ([18]), depth estimation ([19]), pollution prediction ([20]), marketing intention detection ([21]), forecasting photovoltaic system output power ([22]), to name a few cases. However, the CLSTM-NN has not yest been applied to the problem of crime prediction. Differently to other types of data such as pollution levels or rainfall where the studied variables are measurements at fixed points in space and regular in time, crime events have the particularity that the spatial distribution is not known and event occurrences are not regular in time. Also, in the case of Baltimore, the events rarely occur in space and time (on average 10 events for day in an area of approximately 100 Km^2).

The paper is organized as follows. In Section II, the problem and the crime data dataset of Baltimore city are presented. A description of the LSTM and CNN neural networks is provided in Section III. Section IV describes the methodology based on the CLSTM-NN model. The results are shown in Section V. A discussion is reported in Section VI. The paper ends with some final conclusions and a discussion (Section VII).

II. BALTIMORE CRIME DATA

Baltimore is an important seaport in the Maryland state of United States of America. The city is significantly known for high crime rate which ranks higher than the national average. A series of news reports and few crime studies ([23], [24]) depict the surge of violent crimes since 2011, in particular of homicides. The lowest homicide toll of 196 was recorded in the year 2011, and since then there has been a steady upward trend. According to data compiled by the Baltimore Sun (https://homicides.news.baltimoresun.com/), the number of homicides is steady above 300 with the highest value recorded in the year 2015. In the same year, Baltimore's level of violent crimes was much higher (55.4 per 100000) than the national average (5.1 per 100000). Baltimore Police Department is taking initiatives by seeking assistance from the Federal Bureau of Investigations and other federal agencies, to control antisocial activities and to provide a safer and more secure environment in the city. The city government along with Office of the Mayor provide public access to crime data in the BaltimoreOpenData portal (https://data.baltimorecity.gov/). This data is updated every week with an additional lag time due to processing time. The open data portal maintains organized primary and secondary data published by the city council, local authorities, police department and public bodies. The data is available under Creative Commons Attribution 3.0 Unported License (https://creativecommons.org/licenses/by/3.0/). The data set used in the current study contains detailed information of crimes from January 2016 to December 2018 that have occurred in the city of Baltimore, USA. This crime data set has been downloaded from the Public Safety domain of the Open Baltimore portal (https://data.baltimorecity.gov/Public-Safety/BPD-Part-1-Victim-Based-Crime-Data/wsfq-mvij). A total of 148303 records of crime are reported for these three years. Each crime record comes with both spatial (latitude and longitude) and temporal (date and time of occurrence) information along with the specific type of crime. This includes eleven different categories of crimes such as homicide, robbery, larceny or rape.

The database includes the following variables: date and time of crime occurrence, crime code, address where it happened, description of the crime (if the crime was committed within a home or outside), district, latitude, longitude, and zone (street, parking lot, or hotel).

As shown in Fig. 1, there are some types of crimes with a very low number of events, such as 'arson','rape', ' or homicide', while the 'larceny' and 'common assault' crime events are the most frequent. The present study focuses on street robbery and larceny crimes that comprise 29.4% (7.1% for street robbery and 22.3% for larceny) of the sampled data set. The decision to select these two crimes is fundamentally due to the need to experiment the proposed neural network model to data sets with high frequency of events and with a different behavior in space and time.

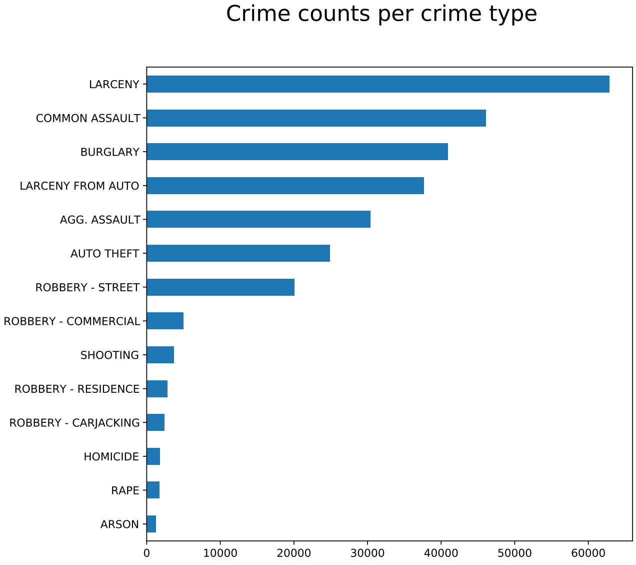


FIGURE 1. Amount of events for type of crime.

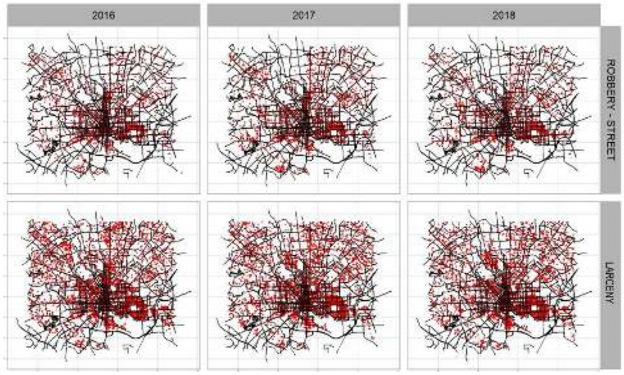


FIGURE 2. Crime events (red points) in the town of Baltimore for the years 2016, 2017, and 2018. Black line represents the road network

Both types of crime events in the town of Baltimore for the years 2016-2018 are depicted over the street network in Fig. 2 using red points. For illustrative purposes, the street network of the city has been accessed from Open Street Map (OSM) repository using R-package osmdata [25]. The package facilitates downloading OSM data using overpass API. OSM data is free and licensed under the Open Data Commons Open Database License (ODbL) by the OpenStreetMap Foundation (OSMF) [26]. Boxplots of the daily events for each year (in Fig. 3) show the different distribution of the two types of crimes over the period 2016-2018.

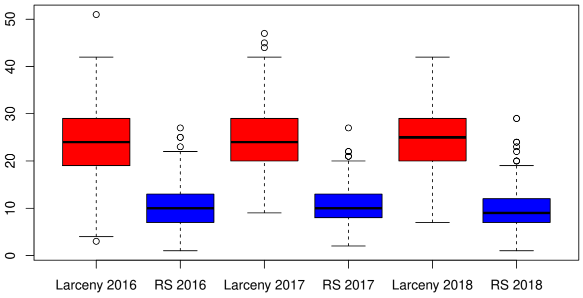


FIGURE 3. Boxplot of daily events per year for the larceny type (red colour) and for the the street robberies (blue colour) crimes.

III. LSTM AND CNN NEURAL NETWORKS

In this section two neural network architectures are described: the Long-Short Term Memory Recurrent Neural Network (LSTM-RNN) and the Convolutional Neural Network (CNN-NN), typically used for the predictions of temporal and spatial data, respectively. The compositions of these two network, called Convolutional Long-Short Term Memory Neural Network (CLSTM-NN), will be considered in this work for the spatio-temporal prediction of Baltimore events.

A. LONG-SHORT TERM MEMORY RECURRENT NEURAL NETWORKS (LSTM-RNN)

Recurrent neural networks (RNN) are a kind of neural networks with a memory given by recurrent layers, where each one takes two inputs: the output of the preceding layer, and the output of the same recurrent layer from the last point it was processed (see [27], [28]). In such way, the RNN naturally handles sequential data, being a suitable choice for dealing with time series data. Nonetheless, these networks are not able to remember the context behind longer sequences. On contrast, Long Short Term Memory (LSTM) networks, proposed by [29], are a variant of RNN designed to model the long term dependency of recurrent networks. LSTM recurrent networks have cells as processing units that apply information loops inside each one of the cells and between the cells themselves (see [30]). Each cell $s_i^{(t)}$ (for each time step t and cell i) corresponds to the basic unit of a LSTM-RNN. In addition, each cell has gating units defined as $f_i^{(t)}$ that regulate the information stream inside and between cells, then a sigmoid unit is applied in order to scale the outputs to range $[0-1]$. The formula of gating units is given by

$$f_i^{(t)} = \sigma\left(b_i^f + \sum_j U_{i,j}^f x_j^{(t)} + \sum_j W_{i,j}^f h_j^{(t-1)}\right) \tag{1}$$

where $x^{(t)}$ is the input vector, $h^{(t)}$ is the hidden layer activation vector, containing the outputs of all the cells, and b^f, U^f, W^f are biases, input weights, and recurrent weights for the forget gates, respectively. The sigmoid unit is given by σ, and the internal state of each cell is then updated with a conditional self-loop weight $f_i^{(t)}$ as follows

$$s_i^{(t)} = f_i^{(t)} s_i^{(t-1)} + g_i^{(t)} \sigma\left(b_i + \sum_j U_{i,j} x_j^{(t)} + \sum_j W_{i,j} h_j^{(t-1)}\right) \tag{2}$$

where b, U, W denote the biases, input weights, and recurrent weights into each cell, respectively. The external input gate unit $g_i^{(t)}$ is similarly computed to the forget gate as

$$g_i^{(t)} = \sigma\left(b_i^g + \sum_j U_{i,j}^g x_j^{(t)} + \sum_j W_{i,j}^g h_j^{(t-1)}\right) \tag{3}$$

Finally, the LSTM cell output is defined as $h_i^{(t)}$ which comes in terms of the output gate $q_i^{(t)}$ that uses the sigmoid unit for gating

$$h_i^{(t)} = tanh(s_i^t) q_i^{(t)} \tag{4}$$

with

$$q_i^{(t)} = \sigma\left(b_i^o + \sum_j U_{i,j}^o x_j^{(t)} + \sum_j W_{i,j}^o h_j^{(t-1)}\right) \tag{5}$$

and where b^o, U^o, W^o are biases, input weights, and recurrent weights for the output gate, respectively.

B. CONVOLUTIONAL NEURAL NETWORKS (CNNS)

Convolutional neural networks are a kind of neural networks which are commonly used for automatic image and data recognition tasks. They are neural networks that use convolution in place of general matrix multiplication in at least one of their layers. A CNN is composed by a set of multiple layers where each convolutional layer generally considers convolution, pooling and activation operations. This neural network searches for local patterns of features detected by the filters applied in the convolution operation. Each filter corresponds to a kernel that is represented by a matrix of numbers, which is typically consistent with the pattern that the filter is trying to detect (see [28]). A generic forward propagation of a CNN layer consists of three phases: First, the layer performs multiple convolutions in parallel to produce a set of linear matrix transformations. A convolution is given by applying the filter over the input data by all possible locations. Then, the resulting linear matrix operation given by convolution over input is processed through a nonlinear activation function. Typically, this function corresponds to sigmoidal, rectified linear or leaky rectified linear functions. Finally, CNN usually uses a pooling function to modify the output of the layer further (see [30]). The pooling layer is used to downsampling feature maps by aggregating the presence of these features. A CNN is composed by different combinations of these layers, producing an output dependent on the nonlinear function used in the last one. After the parameter training procedure, a CNN learns a series of convolutional filters sets which are optimized with respect to the cost function given by the learning task (e.g. classification or regression).

IV. METHODOLOGY

The methodology proposed in this work can be summarized as follows. First, crime events of larceny and street robberies are selected for the years 2016, 2017, 2018. Then, for each type of crime the entire area of Baltimore town is divided into a grid of size $k \times k$ cells, and the number of crime events are evaluated in each grid cell for d days ($d = 1, 2, \ldots, d_K$). Sequences of l matrices are then used as input to the CLSTM-NN. The application of the CLSTM-NN provides a matrix as output representing the crime probability in each square of the grid for the next d days. The probability in each cell is then approximated to 0 and 1 to evaluate the absence of a crime or the occurrence of at least one, respectively. Finally, some classical metrics (such as accuracy, AUC-ROC, and AUC-PR) are applied for assessing the performance of the proposed model. A scheme of the proposed methodology is represented in Fig. 4. Details of the methodology are described below.

A. BUILDING THE INPUT OF THE CLSTM-NN

As crimes are assumed to be spatially correlated, the data are represented as a tensor where each slice correspond a crime map, with each cell representing the amount of crimes at particular d days for $d = 1, 2, \ldots, d_K$. First, an empty square matrix for each d daysis assigned to represent the crime map, where the four corner cells correspond to the combinations of maximum and minimum longitudes and latitudes chosen from the georeferenced map of Baltimore. The number of rows and columns depends on the desired resolution of crime map and considers an equidistant geographical division of the city. Then, each cell (i, j) of this matrix is filled with the number of crimes committed in that area. Each cell of the matrix contains the number of crimes in d days for a particular area. Although different sizes of grids could be considered, in this work, only the 8×8 (corresponding to 806×806 squared meters) and 16×16 grids are used. Sizes smaller than 8×8 would further reduce the spatial model precision, so that they were not considered in this work. Finally, the matrices are stacked to a tensor which corresponds to the input data of the neural network. Finally, a Convolutional Long-Short Term Memory neural network (CLSTM-NN) is applied for the spatio-temporal prediction of larceny and street robbery events over the period 2016-2018.

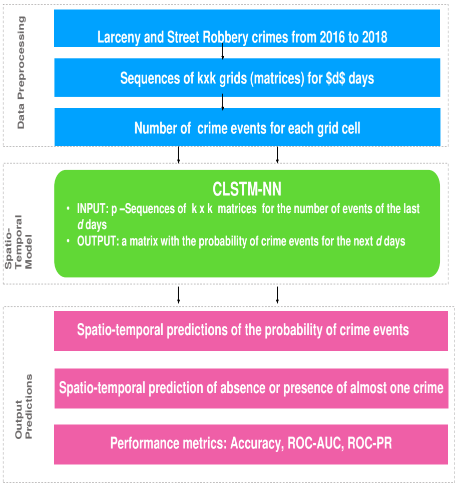

Data Preprocessing
Larceny and Street Robbery crimes from 2016 to 2018
Sequences of kxk grids (matrices) for d days
Number of crime events for each grid cell
Spatio-Temporal Model
CLSTM-NN
• INPUT: p –Sequences of k x k matrices for the number of events of the last *d* days
• OUTPUT: a matrix with the probability of crime events for the next *d* days
Output Predictions
Spatio-temporal predictions of the probability of crime events
Spatio-temporal prediction of absence or presence of almost one crime
Performance metrics: Accuracy, ROC-AUC, ROC-PR

FIGURE 4. Workflow of the methodology.

B. CONVOLUTIONAL LONG-SHORT TERM MEMORY NEURAL NETWORK

In this section a brief description of the proposed CLSTM-NN for spatio-temporal crime prediction is presented. This approach combines sequentially a convolutional neural network (CNN) and a Long-Short Term Memory (LSTM) neural network. The parameters of the LSTM are associated with each individual LSTM cell (see III-A), and the parameters of the CNN are associated with the convolution operation (see III-B). Indeed, CLSTM-NN uses the classical backpropagation algorithm in order to estimate the model parameters, i.e. to train the network.

The CLSTM-NN uses as input data a $k \times k \times l$ tensor that constitutes a sequence of maps representing the crimes for l consecutive d days and k grid divisions in Baltimore city. An example of the architecture of the network with $k = 8$ and $l = 1$ is represented in Fig. 5. Each map represents a geographical grid over the city. A grid has size $k \times k$ where each cell represents the number of crimes occurred in a day inside a specific quadrangular region of the city. The tensors represent the evolution of the patterns along the time axis. A shared convolution layer conformed by 16 filters is then applied to each temporal tensor component of the input data. The size of the convolutional kernel is 2×2, and the *TimeDistributed* wrapper of Keras library is considered. Next, a max pooling layer reducing the dimensionality to the half size of maps (from $k \times k \times l$ down to $k/2 \times k/2 \times l$) is applied. The next step is to introduce a recurrent LSTM layer with 2 output neurons considering a linear activation function. This layer allows to find the structure of the temporal patterns between time slices of the input tensor. The output layer consists of a dense layer where the number of neurons is equal to the number of cells of the output matrix, which corresponds to the $k \times k$ binary crime map of the $l + 1$ day, where a cell is assigned to 1 if it has at least one crime and 0 otherwise. As the aim is focused on the presence of crimes, this layer uses a

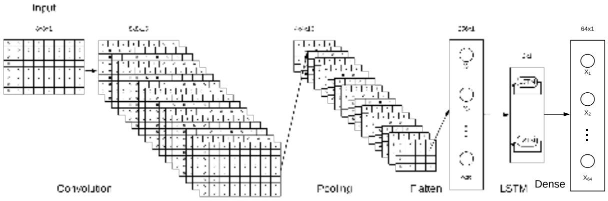


FIGURE 5. Structure of the proposed neural network. A CNN is considered where its last layer is linked to a LSTM neural network. The classification is obtained with a dense layer over the output layer of LSTM.

sigmoid activation function. This neural network is inspired by the proposal of [31], where new links are predicted in a graph, by applying convolutions and LSTM layers to the adjacency matrix. In our experiments, maps with parameter $k \in \{8, 16\}$ for $l = 5$ days stacked in chronological order are considered. A similar architecture can be used by considering an input map representing the sum of crimes at different days.

C. EXPERIMENTAL SETTING

The prediction performance of the proposed neural network are assessed by considering the following three scenarios for the larceny and street robbery crimes in Baltimore:

1) An l sequence of daily crime maps of size 8×8 are used as input to CLSTM-NN to predict the presence or absence of events at the next day. Up to 7 past days are used as input for the prediction on the next day (see Tables 1 and 2).
2) The total of crime events for the past $d = 1, ..., 7$ days over a map of size 8×8 are used as input to CLSTM-NN to predict the presence or absence of events on the next d days (see results in Tables 3 and 4).
3) The total of crime events for the past $d = 1, ..., 7$ days over a map of size 16×16 are used as input to CLSTM-NN to predict the presence or absence of events on the nextthe d days (see Tables 5 and 6).

For all scenarios, a hold-out validation scheme is considered with the 70% of the data assigned to the training set, and the remaining 30% to the testing set. To check the performance of the corresponding neural network, the following metrics were considered: the accuracy, the AUC-ROC curve, and the AUC-PR curve (see, [32]–[34]). For the neural network training, the Adam optimizer (see, [35]) was used for assigning the weight to the neurons. This algorithm works as a stochastic gradient algorithm, where the learning rate is dynamically adapted for minimizing the loss function at each iteration. As the problem corresponds to a classification task, the loss function was given by the binary cross-entropy function. The training process in our experiments considered 100 epochs and a batch value of 32 (i.e. the number of matrices used in each training iteration). The machine used to carry out the experiments had a GTX 1080Ti card able to process the training and tests of the network, each training being carried out in three minutes given the capacity of the card. The most time consuming process is the mapping of the city in matrices with the number of crimes, being the creation of grids of size 64×64 a process that takes approximately two hours.

V. RESULTS

The results of the three scenarios for the testing sets are shown in Tables 1 to 6. For the first scenario, maps of size 8×8 of larceny and street robbery crime events were used as input to the CLSTM-NN neural network, for predicting the presence of crime events one day ahead using d past days, with $d = 1, \ldots, 7$ (see Table 1 for larceny crime and Table 2 for street robbery). From a preliminary analysis, temporal correlations resulted very low for more than 7 days. The best accuracy resulted by using a number of past days less or equal to 5 (see Table 1 for larceny crimes). The use of past 6 and 7 days as input caused a slight overfitting in the training sets causing lower values of the performance metrics in the testing sets. Moreover, although the total accuracy reached relatively good values (more than 73%), the area under the ROC precision recall curves always resulted under 0.62 and 0.54, respectively, especially due to the low number of events per day (since only 25% of the cells has at least one event, the model tends to predict 0 events when the number of events is very low). A similar situation resulted for street robbery crimes (in Table 2) where the accuracy reached the value 0.89 and the AUC-PR is less than 0.50 with only 11% of cells with at least one event. Also, the best accuracy resulted by taking the last 5 days for predicting ahead. However, the AUC- ROC and AUC-PR are slightly better using the last 6 and 7 days.

TABLE 1. Performance metrics (Accuracy, AUC-ROC, and AUC-PR) on the testing set for one day ahead prediction of larceny crimes using different number of past days as input to the CLSTM-NN neural network.

Days of input	Accuracy	AUC-ROC	AUC-PR
1	0.7635	0.6075	0.5480
2	0.7614	0.6146	0.5461
3	0.7589	0.6118	0.5402
4	0.7571	0.6114	0.5371
5	0.7592	0.6147	0.5431
6	0.7533	0.6124	0.5325
7	0.7534	0.6102	0.5317

TABLE 2. Performance metrics (Accuracy, AUC-ROC, and AUC-PR) on the testing set for one day ahead prediction of robbery street crimes using different number of past days as input to the CLSTM-NN neural network.

Days of input	Accuracy	AUC-ROC	AUC-PR
1	0.8911	0.5924	0.3914
2	0.8918	0.5916	0.3922
3	0.8909	0.6018	0.3997
4	0.8923	0.6079	0.4120
5	0.8923	0.6065	0.4096
6	0.8919	0.6085	0.4110
7	0.8911	0.6121	0.4122

For solving the problem of unbalanced data, sequences of aggregated numbers of crime events of the last k days were used as input to the proposed CLSTM-NN for the experiments in the second scenario. Tables 3 and 4 show the results of k days ahead prediction of both types of crimes using five sequences of d aggregated events as input for the CLSTM-NN neural network. For example, for $d = 1$, 5 sequences of 8×8 matrices were used as input to the CLSTM-NN for predicting the 8×8 matrix of the next day; for $d = 2$ 5 sequences of 8×8 matrices were used for predicting the 8×8 matrix of the next two days; and so on until $d = 7$.

TABLE 3. Performance metrics (Accuracy, AUC-ROC, and AUC-PR) on the testing set using a 8×8 grid for the larceny crimes. The metrics are calculated on d days ahead prediction using five sequences of matrices of events in d days events as input to the CLSTM-NN neural network.

d	Accuracy	AUC-ROC	AUC-PR	% Events
1	0.7592	0.6147	0.5431	25.91
2	0.7267	0.7145	0.7399	41.12
3	0.7445	0.7422	0.8191	51.10
4	0.7817	0.7662	0.8671	58.09
5	0.8126	0.7753	0.8938	63.18
6	0.8378	0.7901	0.9136	67.03
7	0.8612	0.7993	0.9270	69.99

In the case of larceny crimes (Table 3) almost all the considered metrics (Accuracy, AUC-ROC, and AUC-PR) (except the accuracy for $k = 1$) increase with k running from 1 to 7, reaching an accuracy of 0.86, an AUC-ROC of 0.80 and an AUC-PR of 0.93. This means that the model improves the predictions when more days are grouped together. Note that for $k = 7$ it means that the probability that in the next 7 days there will be at least one larceny crime event is predicted. In the case of $k = 1$ the data set includes only 26% of the cells with at least one event reaching 70% for $k = 7$. This explains the low AUC-PR for $k = 1$. The accuracy, the AUC- ROC, and AUC-PR improve significantly for $k = 3$ where the percentage of at least one event is around 50%.

TABLE 4. Performance metrics (Accuracy, AUC-ROC, and AUC-PR) on the testing set using a 8×8 grid for the street robbery crimes.The metrics are calculated on d days ahead prediction using five sequences of matrices of events in d days events as input to the CLSTM-NN neural network.

d	Accuracy	AUC-ROC	AUC-PR	% Events
1	0.8922	0.6064	0.4096	11.23
2	0.8528	0.6742	0.5771	18.81
3	0.8289	0.7441	0.6666	24.45
4	0.8317	0.7946	0.7417	28.80
5	0.8353	0.8218	0.7839	32.26
6	0.8361	0.8330	0.8067	35.12
7	0.8432	0.8436	0.8277	37.53

A similar behavior is shown for street robbery predictions in Table 4 where all metrics increase as the d increases. In this case the percentage of the cells with at least one event is only 11% for $k = 1$ explaining the high level of accuracy and very low AUC-ROC (the model tends to predict quite all zeros). For $k = 7$ all the indexes are relatively high with only 37.5% of the cells with one or more events. This means that the model is able to provide good predictions with a relatively small number of events. As expected, these performance indicators improved for higher d reaching the values of 0.84 for the accuracy and AUC-ROC, and 0.82 for the AUC-PR.

The same effect can be noted by increasing the spatial resolution. Tables 5 and 6 show the results for larceny and street robbery, respectively, by considering a grid of 16×16 as described by the third scenario.

TABLE 5. Performance metrics (Accuracy, AUC-ROC, and AUC-PR) on the testing set using a 16×16 grid for the larceny crimes.The metrics are calculated on d days ahead prediction using five sequences of matrices of events in d days events as input to the CLSTM-NN neural network.

d	Accuracy	AUC-ROC	AUC-PR	% Events
1	0.9175	0.5137	0.3427	8.09
2	0.8548	0.5381	0.4251	14.72
3	0.8123	0.5961	0.5085	20.35
4	0.7810	0.6304	0.5704	25.22
5	0.7619	0.6679	0.6286	29.48
6	0.7504	0.6883	0.6716	33.25
7	0.7442	0.7064	0.7071	36.61

TABLE 6. Performance metrics (Accuracy, AUC-ROC, and AUC-PR) on the testing set using a 16×16 grid for street robbery crimes. The metrics are calculated on five days ahead prediction using d sequences of five days' aggregated events as input to the CLSTM-NN neural network.

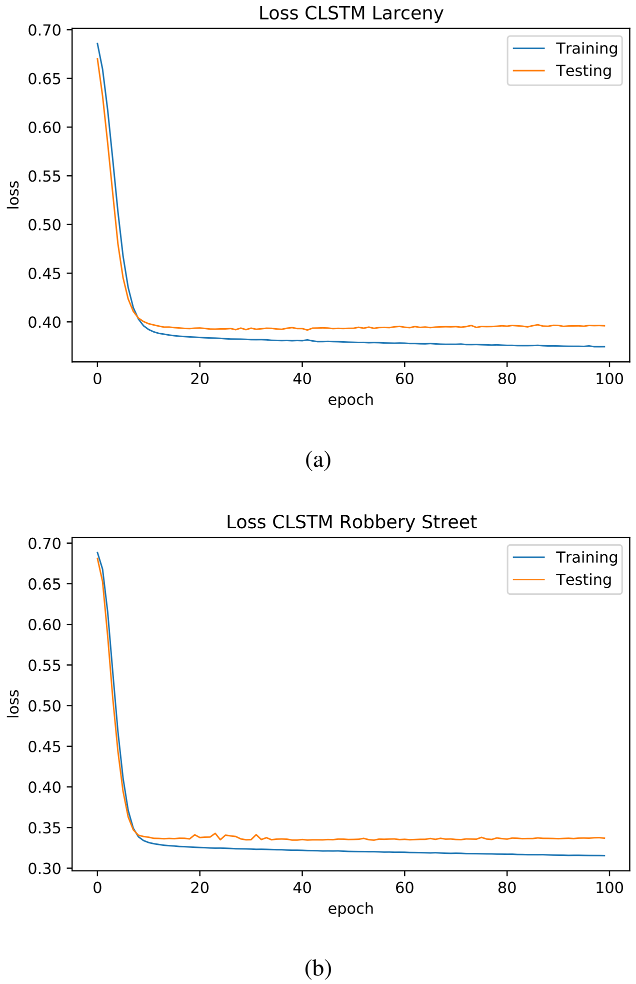
d	Accuracy	AUC-ROC	AUC-PR	% Events
1	0.9684	0.5000	0.5157	3.33
2	0.9406	0.5333	0.2903	6.19
3	0.9194	0.5771	0.3870	8.72
4	0.9020	0.6205	0.4528	10.97
5	0.8899	0.6558	0.5115	13.00
6	0.8803	0.6818	0.5544	14.84
7	0.8719	0.6985	0.5838	16.51

As expected, in both cases the accuracy decreases and the AUC-ROC and AUC-PR increase with higher values of d.

Given the poor representativeness of the predictions on the 16×16 grid size (due to the low number of cells with at least one event) in the following some particular results of the second scenario will be described.

As an example, the losses (for the training and testing sets) for the larceny and street robbery crimes with $d = 5$ and for a grid size 8×8 are shown in Fig. 6a) and Fig. 6b), respectively. As can be noted in Fig. 6, the loss of the model converges rapidly in the first 20 epochs, with the training curve providing a loss lower than the testing one. The difference between the two curves slightly increases with the iterations. To visualize the learning skills of these models, the AUC-ROC and AUC-PR curves are shown in Fig. 7a) and Fig. 7b) for larceny crimes, and in Fig. 8a) and Fig. 8b) for street robberies.

Fig. 7a) represents the ROC curve for $d = 5$ (in Table 5) when predicting larceny events. The area under this curve approximately equals to 0.77 (higher than 0.5 given by the dashed line indicating that the CLSTM-NN model is able to correctly classify the presence of larceny cases in more than 77% of the cases. The Precision-Recall metric represented by the area under the curve shown by Fig. 7b) indicates that the proposed model procedure is able to correctly classify positive cases (presence of at least one crime) in the 89% of



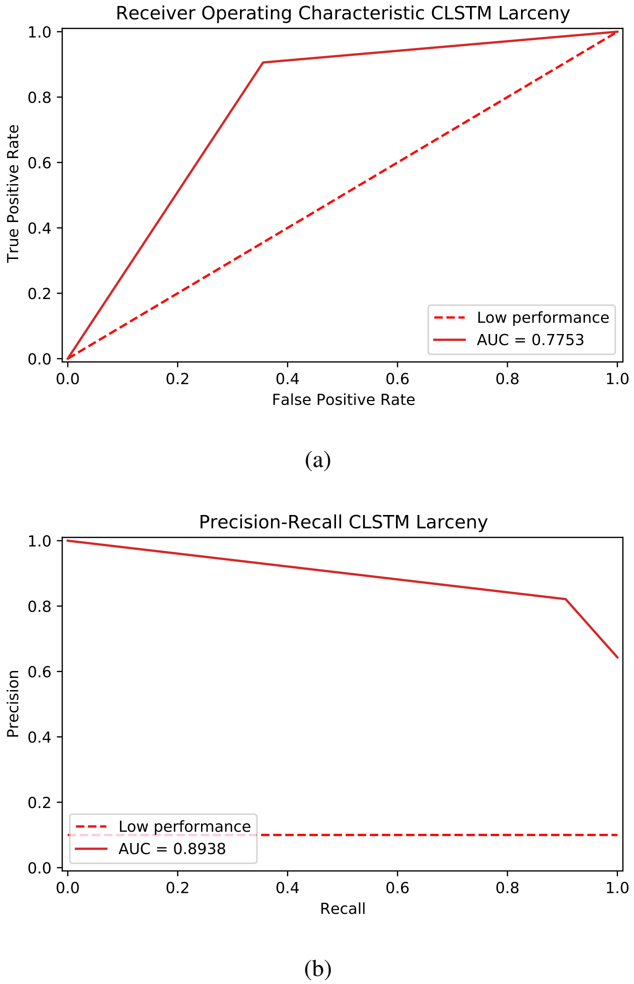
(a)

(b)

FIGURE 6. Loss functions for the CLSTM model with $d = 5$ and grid size 8×8 for larceny (a) and robbery street (b) crimes. The blue line represents the loss obtained by the model in the training set at each iteration, while the red line indicates the loss for the testing set.

the cases. A similar behavior can be noted in Fig. 8a) and Fig. 8b) for street robbery cases, where the CLSTM model is able to correctly classify the presence of street robbery cases in more than 82% of the cases (Fig. 8a), and the 78% positive cases (Fig. 8b).

Fig. 9 and Fig. 10 represent an example of the predictions and the observed crime events for the larceny and street robbery crimes, respectively, for a period of $d = 5$ days ($d = 5$ is considered a reasonable period for the decision maker in taking decisions for preventing crime events and therefore it was worth visualizing the corresponding results). In both cases the output of the CLSTM-NN model was a matrix consisting of the cumulative events in a 8×8 size map while 5 sequences of 5 days were taken as input. For both crime types, the total events on the days February, 5, 6, 7, 8 and 9 of the year 2018 were considered as input to the CLSTM-NN model for predicting the presence of at least one crime in the next 5 days (February 10, 11, 12, 13 and 14) of the same year.

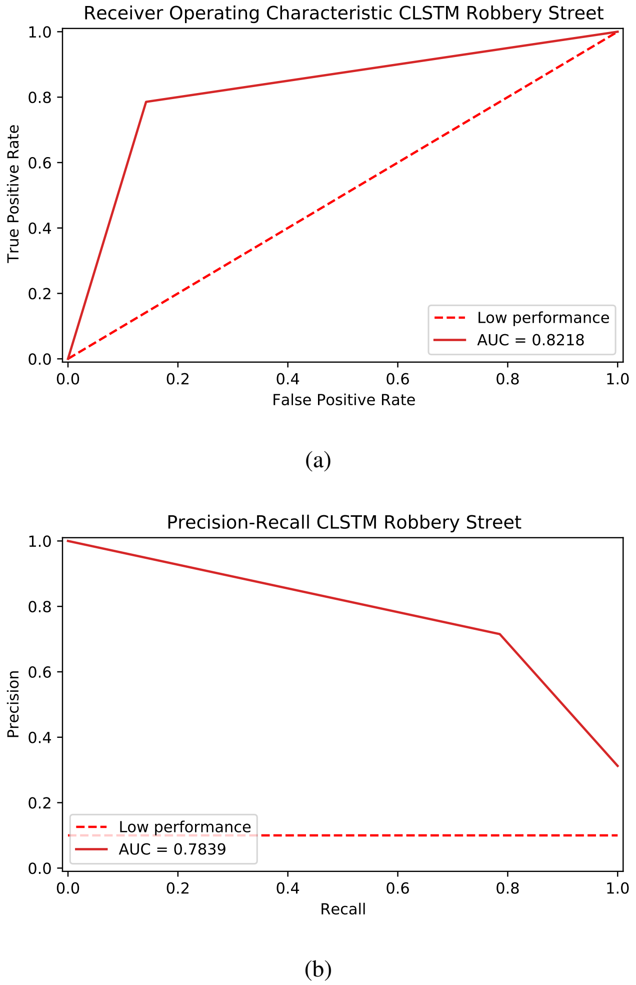


(a)

(b)

FIGURE 7. Prediction performance for the CLSTM model with $k = 5$ and grid size 8×8 for larceny crimes. (a) ROC curve: the area under the red line represents the percentage of cases correctly classified while the dashed red line indicates a randomly classification. (b) Precision-recall curve: the area under the red line represents the percentage of true cases which are correctly classified, while the red dashed line indicates a model which is not able to classify an event as a true case.

Fig. 9a) and Fig. 10a) show the probability of occurrence of at least one event with the observed events overlapped (red points); Fig. 9b) and Fig.10b) show the probability of occurrence of at least one event with the observed events overlapped (red points), and Fig. 9c) and Fig. 10c) show the distribution of the areas where the events occurred (white squares). Observing these results, it is noted the highest crime rates are concentrated at Baltimore downtown. Comparing the predictions with the observed points, there are few predictions errors (given by the red points on the gray squares) showing the ability of the proposed neural network to learn about the spatial and temporal structures and predict crime presence on the future. Also, by comparing the probability output with the location of observed events, only few events (red points) correspond to the lowest probabilities (darker squares).



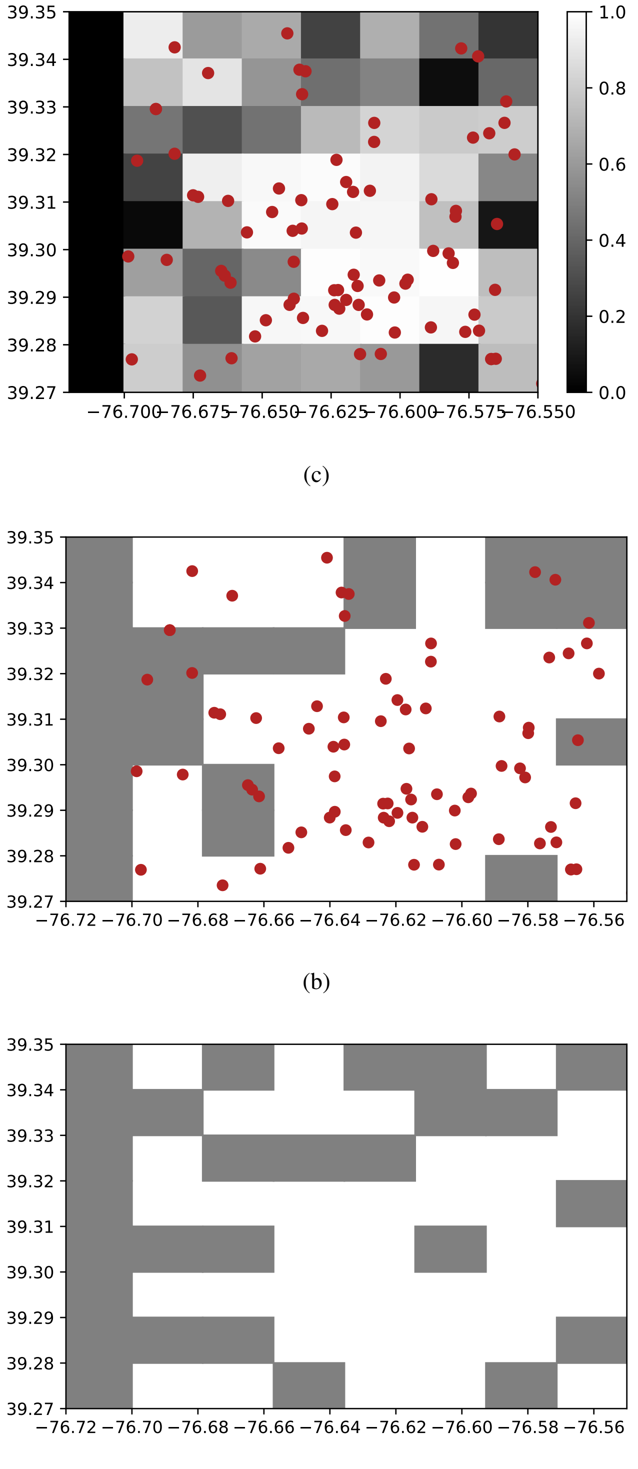
(a)

(b)

FIGURE 8. Prediction performance for the CLSTM model with $d = 5$ and grid size 8×8 for robbery street crimes. (a) ROC curve: the area under the red line represents the percentage of cases correctly classified, while the dashed red line indicates a randomly classification. (b) Precision-recall curve: the area under the red line represents the percentage of true cases which are correctly classified, while the red dashed line indicates a model which is not able to classify an event as a true case.

VI. DISCUSSION

This work proposes a deep learning model for predicting crimes using LSTM and Convolutional layers to generate predictions at crime sub-zones. Past number of events per day in each sub-zone have been considered as input to the model, and a map of 0 and 1 (0=zero events, 1=at least one event) as output. Since the generated matrices of events by day included a large number of 0 values (more than the 50%) providing an unbalanced dataset, aggregation of events over different days are considered in order to have input matrices with higher percentages of events. Two types of crimes, larceny and street robberies, were chosen for this study in the city of Baltimore. In order to assess the performance of the model, the following metrics have been applied: the accuracy, the area under the ROC and precision-recall curve. The best results were obtained for larceny crime by considering a 8×8 grid map, and 5 sequences of 7 days as input. In this case the

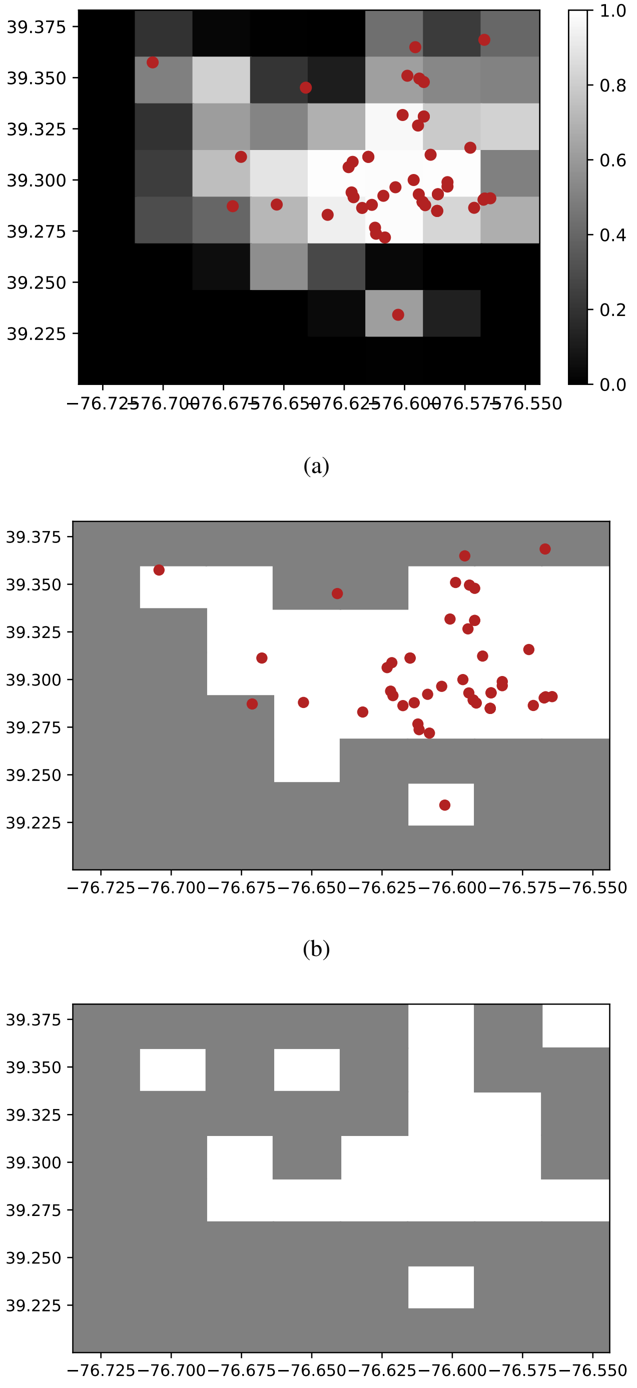


(c)

(b)

(a)

FIGURE 9. Prediction of larceny crimes in a 8×8 grid from 2/10/2018 to 2/14/2018 using 5 sequences of 5 days as input: (a) prediction of probability of one or more events with observed events overlapped (darker cell: less predicted probability of crime; true crime events: red points); (b) Prediction of absence/presence of events (gray color: absence; white color: presence of at least one event) with observed events overlapped (red points); (c) Matrix of absence/presence (gray color: absence; white color: presence of at least one event).



(a)

(b)

(c)

FIGURE 10. Prediction of street robbery crimes in a 8×8 grid from 2/10/2018, to 2/14/2018 using 5 sequences of 5 days as input: (a) prediction of probability of one or more events with observed events overlapped (darker cell: less predicted probability of crime; true crime events: red points); (b) Prediction of absence/presence of events (gray color: absence; white color: presence of at least one event) with observed events overlapped (red points); (c) Matrix of absence/presence (gray color: absence; white color: presence of at least one event).

proposed model provided an accuracy of 86%, and an AUC-ROC of 77%, and an AUC-PR of 93%. For street robbery crime, the best results, obtained with the same structure ($d = 7$), were 83%, 81% y 77% for the three indicators.

There are not other published works on the application of machine learning on crimes at Baltimore town according to our revision; however, by comparing our results with those obtained by the application of deep learning techniques to the prediction of crime at other cities in the world, our method results competitive, by outperforming existing proposals in many occasions. For example, [8] implemented a recurrent neural network (RNN) and a convolutional one (CNN) for predicting the next day crime counts in the cities of Chicago and Portland. By considering additional variables related to weather, census data and transportation, they reached an accuracy of 75.6% and 65.3% for Chicago and Portland, respectively. An LSTM model was used by [7] for predicting future class labels of crime incidents. They validated their method over Chicago by using some input features obtaining an accuracy rate of 87.4% and showing its ability to work with big data sets. Differently to our work, no spatial variability was taken into account in this work.

A spatio-temporal approach has been considered by [11] and [12]. [11] used a regression model (polynomial regression, support vector regression, and auto-regressive regression) for predicting crime activity in the city of Chicago using social information sources from network analytic techniques. By comparing the models, the support vector regression provided better performances in terms of RMSE. [12] presented the design and implementation of an approach based on spatial analysis and auto-regressive models to automatically detect high-risk crime regions in urban areas, and to reliably forecast crime trends in each region. The results (in terms of MAE) showed that crimes decrease for smaller areas. A different approach using autoencoder architecture with convolutions for predicting the number of crimes in the town of Chicago has been proposed by [36]. The results provided a very good performance (around 97% of R^2) when using a small dataset (less or equal to a year). The use of a spatio-temporal approach in this work is mainly justified by the presence of spatial and temporal correlations in crime data. The assumption of spatial auto-correlation can also be confirmed by the results of Moran´s I test ([37]), presented in Table 7, that always rejected the null hypothesis of spatial randomness when taking temporal aggregation of crime events.

TABLE 7. Moran's I test for the larceny and street robbery crime events in the years 2016, 2017, and 2018: Moran's I statistic, Standard Deviate (SD), and P-value.

Year	Crime	Moran's I stat.	SD	P-value
2016	Rob.Street	0.5124	8.4434	2.20e-16
	Larceny	0.3526	6.6328	1.64e-11
2017	Rob.Street	0.5006	8.0897	2.99e-16
	Larceny	0.3230	6.1368	4.21e-10
2018	Rob.Street	0.4931	7.9654	8.23e-16
	Larceny	0.3808	6.8521	3.64e-12

VII. CONCLUSIONS AND FURTHER DEVELOPMENTS

In the context of machine learning, the use of neural networks plays an important role in data analysis. In particular, neural networks are shown to be useful to learn from the past patterns to predict the future. In this paper, the Convolutional Long Short Memory Neural Network (CLSTM-NN) is proposed to predict the presence of crime events over the city of Baltimore (USA). The model constitutes a novelty for this kind of data since the CLSTM-NN is more often used in the transport field and other applications (see, [38]). Also, there are no works that predict the crimes of Baltimore town using deep learning.

Three scenarios were considered by taking a different number of events for predicting the crime events in the next d days. The CLSTM-NN provided better results in the second scenario by reaching and accuracy of 0.86 and a AUC-PR of 0.93 for the larceny crime using sequences of matrices of events occurred in $d = 7$ days. The obtained results show that the network model is able to learn past spatial patterns to predict future presence of crimes. The main results could be summarized as follows: (i) the temporal and spatial resolution is relevant in the performance of the model; (ii) the type of crime could influence the performance of the neural network; (iii) an opportune pre-processing of data and the use of an optimal architecture of neural network are important ingredients for crime prediction. A good prediction of when and where will happen the future crime events would help the police to make better use of limited resources. However, the main limitation of CLSTM-NN is due to the poor performance for low percentage of crime events (given by sparse spatio-temporal matrices) which affects the spatio-temporal resolution of predictions. The problem of sparsity has been partially solved by [39] by representing the hourly crime data of Chicago and Los Angeles towns with a spatio-temporal weighted graph (STWG). In order to improve the predictions on higher resolution scales, the graph approach of [39] together with other deep learning methods (see, the reviews [38] and [40] and the references within) will be explored in future works. Also, attention-based neural networks will be applied to improve crime prediction considering exogenous variables related to crime events (such as weather and socio-demographic data). Finally, other types of crime (such as common assault, burglary, etc.) will be considered for the spatio-temporal prediction.

REFERENCES

[1] A. Curry, C. Latkin, and M. Davey-Rothwell, "Pathways to depression: The impact of neighborhood violent crime on inner-city residents in baltimore, maryland, usa," *Social science & medicine (1982)*, vol. 67, pp. 23–30, 08 2008.

[2] B. David. H and A. Suruliandi, "Survey on crime analysis and prediction using data mining techniques," *ICTACT Journal on Soft Computing*, vol. 7, pp. 1459–1466, 04 2017.

[3] L. McClendon and N. Meghanathan, "Using machine learning algorithms to analyze crime data," *Machine Learning and Applications: An International Journal*, vol. 2, pp. 1–12, 03 2015.

[4] J. Q. Yuki, M. M. Q. Sakib, Z. Zamal, K. M. Habibullah, and A. K. Das, "Predicting crime using time and location data," in *Proceedings of the 2019 7th International Conference on Computer and Communications Management*, 2019, pp. 124–128.

[5] S. Wu, C. Wang, H. Cao, and X. Jia, "Crime prediction using data mining and machine learning," in *International Conference on Computer Engineering and Networks*. Springer, 2018, pp. 360–375.

[6] N. Shama, "A machine learning approach to predict crime using time and location data," Ph.D. dissertation, BRAC University, 2017.

[7] J. A. M.-T. Ulises M. Ramirez-Alcocer, Edgar Tello-Leal, "Predicting incidents of crime through lstm neural networks in smart city domain," in *SMART 2019 : The Eighth International Conference on Smart Cities, Systems, Devices and Technologies*. IARIA, 07 2019, pp. 32–37.

[8] A. Stec and D. Klabjan, "Forecasting crime with deep learning," *arXiv preprint arXiv:1806.01486*, 2018.

[9] A. S. Bhardwaj, M. Divakar.K., A. AshiniK, S. DevishreeD., and S. M. Younis, "Deep learning architectures for crime occurrence detection and prediction," *International Journal of Advance Research, Ideas and Innovations in Technology*, vol. 5, pp. 822–824, 2019.

[10] A. Krishnan, A. Sarguru, and A. S. Sheela, "Predicitve analysis of crime data using deep learning," *International Journal of Pure and Applied Mathematics*, vol. 118, no. 20, pp. 4023–4031, 2018.

[11] S. K. Dash, I. Safro, and R. S. Srinivasamurthy, "Spatio-temporal prediction of crimes using network analytic approach," in *2018 IEEE International Conference on Big Data (Big Data)*. IEEE, 2018, pp. 1912–1917.

[12] C. Catlett, E. Cesario, D. Talia, and A. Vinci, "A data-driven approach for spatio-temporal crime predictions in smart cities," in *2018 IEEE International Conference on Smart Computing (SMARTCOMP)*. IEEE, 2018, pp. 17–24.

[13] T. Hu, X. Zhu, L. Duan, and W. Guo, "Urban crime prediction based on spatio-temporal bayesian model," *PloS one*, vol. 13, no. 10, 2018.

[14] R. Yadav and S. K. Sheoran, "Computational model for spatio temporal crime event prediction," *International Journal of Advanced Research in Computer Science*, vol. 8, no. 1, 2017.

[15] A. Z. Mangion and C. K. Wikle, "Deep integro-difference equation models for spatio-temporal forecasting," *CoRR*, vol. abs/1910.13524, 2019. [Online]. Available: http://arxiv.org/abs/1910.13524

[16] S. Xingjian, Z. Chen, H. Wang, D.-Y. Yeung, W.-K. Wong, and W.-c. Woo, "Convolutional lstm network: A machine learning approach for precipitation nowcasting," in *Advances in neural information processing systems*, 2015, pp. 802–810.

[17] D. Wang, Y. Yang, and S. Ning, "Deepstcl: A deep spatio-temporal convlstm for travel demand prediction," in *2018 International Joint Conference on Neural Networks (IJCNN)*, 2018, pp. 1–8.

[18] M. Rebol and P. Knöbelreiter, "Frame-to-frame consistent semantic segmentation," *arXiv preprint arXiv:2008.00948*, 2020.

[19] V. Patil, W. V. Gansbeke, D. Dai, and L. V. Gool, "Don't forget the past: Recurrent depth estimation from monocular video," 2020.

[20] V.-D. Le, T.-C. Bui, and S. K. Cha, "Spatiotemporal deep learning model for citywide air pollution interpolation and prediction," 2019.

[21] Y. Wang, K. Ma, L. Garcia-Hernandez, J. Chen, Z. Hou, K. Ji, Z. Chen, and A. Abraham, "A clstm-tmn for marketing intention detection," *Engineering Applications of Artificial Intelligence*, vol. 91, p. 103595, 2020.

[22] C.-J. Huang and P.-H. Kuo, "Multiple-input deep convolutional neural network model for short-term photovoltaic power forecasting," *IEEE Access*, vol. 7, pp. 74 822–74 834, 2019.

[23] R. S. Levine, R. P. Schneid, R. J. Zoorob, and C. H. Hennekens, "A tale of two cities: Persistently high homicide rates in baltimore city compared with significant declines in new york city," *The American journal of medicine*, vol. 132, no. 1, pp. 3–5, 2019.

[24] D. Weisburd and C. White, "Hot spots of crime are not just hot spots of crime: Examining health outcomes at street segments," *Journal of contemporary criminal justice*, vol. 35, no. 2, pp. 142–160, 2019.

[25] M. Padgham, B. Rudis, R. Lovelace, and M. Salmon, "osmdata," *The Journal of Open Source Software*, vol. 2, no. 14, jun 2017. [Online]. Available: https://doi.org/10.21105/joss.00305

[26] OpenStreetMap contributors, "Planet dump retrieved from https://planet.osm.org ," https://www.openstreetmap.org, 2017.

[27] L. Medsker and L. Jain, *Recurrent neural networks: design and applications*. CRC Press, 1999.

[28] F. Cady, *The Data Science Handbook*. John Wiley & Sons, Incorporated, 2017.

[29] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[30] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016, http://www.deeplearningbook.org.

[31] J. Chen, X. Xu, Y. Wu, and H. Zheng, "Gc-lstm: Graph convolution embedded lstm for dynamic link prediction," *arXiv preprint arXiv:1812.04206*, 2018.

[32] J. Davis and M. Goadrich, "The Relationship Between Precision-Recall and ROC Curves," in *ICML '06: Proceedings of the 23rd international conference on Machine learning*. ACM, 2006, pp. 233–240.

[33] N. Japkowicz, *Assessment Metrics for Imbalanced Learning*. John Wiley & Sons, Ltd, 2013, ch. 8, pp. 187–206.

[34] P. Branco, L. Torgo, and R. Ribeiro, "A survey of predictive modelling under imbalanced distributions," 2015.

[35] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[36] N. Esquivel, B. Peralta, and O. Nicolis, "Crime level prediction using stacked maps with deep convolutional autoencoder," in *2019 IEEE CHILEAN Conference on Electrical, Electronics Engineering, Information and Communication Technologies (CHILECON)*. IEEE, 2019, pp. 1–7.

[37] P. A. P. Moran, "The interpretation of statistical maps," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 10, no. 2, pp. 243–251, 1948.

[38] S. Wang, J. Cao, and P. S. Yu, "Deep learning for spatio-temporal data mining: A survey," 2019.

[39] B. Wang, X. Luo, F. Zhang, B. Yuan, A. L. Bertozzi, and P. J. Brantingham, "Graph-based deep modeling and real time forecasting of sparse spatio-temporal data," *arXiv preprint arXiv:1804.00684*, 2018.

[40] N. Gao, H. Xue, W. Shao, S. Zhao, K. K. Qin, A. Prabowo, M. S. Rahaman, and F. D. Salim, "Generative adversarial networks for spatio-temporal data: A survey," 2020.

NICOLÁS ESQUIVEL was born in Santiago, Chile in 1995. In 2014 he started studying a B.S. in computer science at the Andrés Bello University, Chile, now. He is now a student of the Master in Computer Sciences at the same University. He is co-author of a IEEE publication on crime prediction and he's doing his thesis on the same topic. His research interests include deep-learning, artificial intelligence, and spatio-temporal prediction models.

ORIETTA NICOLIS earned a degree in Economics in 1995 at the University of Verona, Italy. She obtained her PhD degree in Statistics in 1999 at the University of Padua (Italy) and a postdoctoral fellowship in statistics at the University of Brescia (Italy) in the following two years. From 2002 to 2012, she has worked as Researcher and Aggregate Professor of Statistics at the University of Bergamo (Italy). From 2012 to 2018 she was working at University of Valparaiso where she was Director of the PhD Program in Statistics. Since August 2018, she is Full Profesor at the Engineering Faculty of the University Andres Bello en Vinã del Mar (Chile) where she is director of the Master in Computation Sciences. She is also responsable of national projects on artificial intelligence and statistical models. Dr. Nicolis is author of more than 60 international publications and reviewer of several international scientific journals. Her research interests include the study of spatio-temporal models, machine learning methods, deep learning, big data, computer science, wavelet-transforms, fractional, and multifractal processes.

BILLY PERALTA was born in Lima, Perú in 1981. He received the B.S. in system engineering from the Universidad Nacional de Trujillo, Perú, in 2001, and the M.S and Ph.D. degree in Computer Sciences from Pontificia Universidad Católica, Chile, in 2008 and 2013, respectively. From 2004 to 2018, he was an Assistant Professor with the Universidad Católica de Temuco, Chile. Since 2018, he has been an Assistant Professor with the Universidad Andres Bello, Chile. He is the author of 22 articles, and he is member of IEEE Society and Chilean Society of Computer Science. His research interests lie in the area of machine learning and computer vision. In recent years, he has focused on applications of deep learning applications. He has collaborated actively with researchers in several other scientific disciplines.

JORGE MATEU Jorge Mateu graduated in Mathematical Sciences at University of Valencia (Spain) and fulfilled the PhD at the University of Valencia with long visiting periods to University of Lancaster (UK), with Prof. Peter Diggle. He is currently a full professor of Statistics at the Department of Mathematics of University Jaume I of Castellon, where he has worked for the past 20 years. His main fields of interest are stochastic processes in their wide sense with a particular focus on spatial and spatio-temporal point processes and geostatistics. He has published more than 250 papers in peer-reviewed international journals, and he is co-author of several proceedings and research books. He has organised several international conferences with a focus on modelling space-time processes, and leads the organising committee of a series of biannual conferences (called METMA, ten by now) co-sponsorised by TIES, for which he was also Secretary. Prof. Mateu is also director of the Unit "Statistical Modelling of Crime Data", based in the Department of Mathematics, University Jaume I of Castellon, and he is co-director of the Erasmus Mundus Master in Geospatial Technologies, funded by the European Commission.