Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Spatio-Temporal Crime Modeling in São Paulo: A Comparison of SAR and LSTM Approaches

Wellington Yuanhe Zhao

wellingtonyuanhe@gmail.com

Universidade de São Paulo

Cibele M. Russo

Universidade de São Paulo

Luis Gustavo Nonato

Universidade de São Paulo

Research Article

Keywords: crime modeling, spatial data, temporal data

Posted Date: May 11th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-8752288/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Spatio-Temporal Crime Modeling in São Paulo: A Comparison of SAR and LSTM Approaches

Abstract

This study investigates the prediction of urban crime in the São Paulo metropolitan region using an integrated spatio-temporal modeling framework. We compare Spatial Autoregressive (SAR) models with Artificial Neural Networks (ANN), including Long Short-Term Memory (LSTM) architectures, leveraging a comprehensive dataset of crime incidents, demographic indicators, and educational and economic variables at the municipal level. The SAR model captures spatial dependencies and allows for interpretable regional analysis, while LSTM networks model complex nonlinear patterns and temporal dynamics. Empirical results show that although the SAR model explains crime variation effectively, LSTM achieves substantially better predictive performance (RMSE: 22.64 vs 77.18 for SAR). This suggests the potential of deep learning techniques in improving urban crime forecasting. Our findings highlight the importance of combining spatial econometrics and machine learning to support evidence-based public safety policies and targeted social interventions.

Keywords: crime modeling; spatial data; temporal data

1 Introduction

Crime remains a significant challenge for societies around the world, and its impact is particularly severe in Brazil. According to the United Nations Office on Drugs and Crime, Brazil reported a homicide rate of 27.4 per 100,000 inhabitants in 2021, placing it among the countries with the highest rates of intentional homicides globally (United Nations Office on Drugs and Crime, 2023). Complementary national reports also highlight persistent patterns of violence and regional disparities in crime distribution (Cerqueira et al., 2021).

These alarming statistics have led to increased interest in understanding and modeling the spatial and temporal dynamics of crime. The traditional methodology in criminology often adopts a micro-spatial perspective, which focuses on localized geographic units such as neighborhoods or blocks. These approaches explore the correlation between social indicators and crime occurrence. For instance, Sherman et al. (Sherman et al., 1989) developed the hot spot theory, demonstrating that a small proportion of locations accounts for a large proportion of criminal incidents. Building upon this idea, Caplan et al. (2011) proposed the Risk Terrain Modeling (RTM) approach by incorporating environmental risk factors through GIS data. More recently, Wheeler and Steenbeek (2021) applied machine learning, notably random forests, to improve crime prediction accuracy over traditional RTM models.

In Brazil, spatial crime studies have also gained traction. Melo et al. (2015) analyzed crime distribution in the city of Campinas, revealing spatial clustering and substantial variation across different types of offenses. These findings support the hypothesis that crime patterns are not randomly distributed and are shaped by local socioeconomic and demographic structures.

While micro-spatial models are valuable, an alternative approach considers broader spatial units, such as municipalities. The state of São Paulo, Brazil's most populous and economically developed region, offers a diverse setting to study these patterns. Municipal-level analysis allows researchers to account for heterogeneity in social, economic, and urban characteristics across municipalities, making it possible to examine regional determinants of crime rates.

To model the spatial dependence observed in crime data, spatial econometrics provides robust tools. In particular, the Spatial Autoregressive (SAR) model accounts for spatial autocorrelation by incorporating neighboring effects in the regression framework. Anselin's foundational work in spatial econometrics (Anselin, 1988) and subsequent developments by R. S. Bivand et al. (2013) have made SAR models a cornerstone in spatial data analysis. For example, Brown (1982) applied SAR models to identify the role of poverty and racial composition in explaining crime variations in suburban areas.

In parallel with spatial models, machine learning has emerged as a powerful tool in predictive criminology. Artificial Neural Networks (ANNs) and their extensions have shown promising results in modeling complex, nonlinear relationships in crime data. Wang et al. (2013) proposed learning temporal crime patterns using event sequences, while Chun et al. (Chun et al., 2019) developed deep learning models to assess individual risk of recidivism. Reier Forradellas et al. (2020) modeled region-wise crime counts in Buenos Aires using contextual social indicators, and Yi et al. (2019) introduced a hybrid LSTM + SDAE model for fine-grained crime prediction across US municipalities. These approaches underscore the flexibility of neural networks in handling both temporal dependencies and spatial heterogeneity.

Research Questions and Contributions

This study addresses two key research questions:

1. *Can spatial econometric models such as SAR adequately capture urban crime patterns in São Paulo, considering spatial autocorrelation and socioeconomic predictors?*

2. *Do deep learning models, such as LSTM networks, offer superior predictive accuracy compared to traditional spatial models, particularly for temporal crime forecasting?*

Our contributions are threefold:

- We construct a rich, multi-source spatio-temporal dataset covering crime, education, economic, and demographic indicators across the São Paulo metropolitan region from 2002 to 2018.

- We compare the performance of a Spatial Autoregressive (SAR) model with Artificial Neural Networks (ANN) and Long Short-Term Memory (LSTM) models for crime per capita prediction.

- We conduct extensive simulations to validate SAR model estimates and visualize prediction outputs and residuals using choropleth and diagnostic maps.

2 Methodology

Modeling and predicting crime patterns based on historical data and geospatial information is an important task for public management, particularly in complex urban regions such as the metropolitan area of São Paulo. Geospatial models are especially valuable for exploring the relationships between crime and socioeconomic, demographic, and educational predictors across municipalities, each defined by its latitude and longitude coordinates.

In this study, we employ a Spatial Autoregressive (SAR) model to describe and predict the number of crimes in municipalities within the São Paulo metropolitan region. The SAR model

accounts for spatial dependencies by incorporating a spatial neighborhood matrix, and it also captures the monthly seasonality patterns observed in the data. The modeling process involves data cleaning, transformation, and spatial weight matrix construction using R. Prior to fitting the SAR model, the Lasso regression technique is applied for variable selection, identifying the most relevant predictors such as inhabitants per household, public elementary school failure rate, public school early years enrollment, and elementary school dropout rates.

In addition to the SAR model, we also implement Artificial Neural Network (ANN) models to investigate nonlinear relationships and compare predictive performance. The objective of this work is to develop spatio-temporal models capable of identifying the key variables that influence crime rates in the São Paulo metropolitan area and to evaluate the relative strengths of traditional spatial econometric and machine learning approaches.

To support the modeling, the dataset was collect from various sources. The dataset covers the period from 2002 to 2018. Observations from 2002 to 2017 were used for model training, while data from 2018 were reserved for testing. We chose 2018 as the cutoff year because data collected after this point are likely to be significantly affected by the COVID-19 pandemic. Lockdowns, changes in police patrols, and mobility restrictions introduced in 2020–2022 have distorted typical crime dynamics and reporting behavior. Therefore, excluding post-2018 data helps ensure that model evaluation reflects stable pre-pandemic crime patterns.

2.1 Variables and Sources

The final dataset includes 102 variables, grouped into the following categories:

- **Crime indicators:** Aggregated and disaggregated data for offenses such as homicide, robbery, assault, and vehicle theft, available at monthly resolution. A composite variable, *total crime*, was created by summing reported crimes across categories.
- **Demographic indicators:** Population size, age group distributions, household size, and urbanization levels.
- **Education indicators:** School enrollment figures, dropout rates, approval rates, and distinctions by public/private and education stage (e.g., elementary, high school).
- **Economic indicators:** Municipal GDP, industrial activity, and transformed value of industrial output (VTI) across sectors.

2.2 Data Preprocessing

Data cleaning and integration were conducted using Python and R. Variables with extensive missing values, such as *total crime with 12-month lag*, were excluded from modeling. Additionally, highly collinear variables were dropped to improve model interpretability. A log transformation was applied to reduce skewness in both predictors and response variables.

The spatial structure was incorporated using a municipality-level contiguity-based spatial weight matrix, constructed using GeoJSON polygon boundaries. Crime values were normalized per capita and scaled for visualization purposes.

The final dataset contains approximately 123,000 training observations and 7,728 test observations, reflecting monthly resolution across municipalities. data were collected from multiple official sources. Crime data were obtained from Secretaria de Segurança Pública do Estado de São Paulo (2021), and a new response variable - total crime - was constructed by aggregating

all crime categories. Socioeconomic, educational, and industrial data were sourced from Fundação Sistema Estadual de Análise de Dados (SEADE) (2025), while geolocation data were provided by Instituto Brasileiro de Geografia e Estatística (IBGE) (2025). Among the socioeconomic indicators, the transformed industrial value (VTI) represents the combined value of goods and services sold by the industrial sector in each municipality.

Table 1: Description of crime-related variables, in Portuguese and English.

Variable name (PT)	Variable name (EN)
homicídio doloso	intentional homicide
homicídio doloso por acidente de trânsito	traffic-related intentional homicide
número de vítimas em homicídio doloso por acidente de trânsito	number of victims in traffic-related intentional homicide
homicídio culposo por acidente de trânsito	traffic-related manslaughter
homicídio culposo outros	manslaughter (other causes)
tentativa de homicídio	attempted homicide
lesão corporal seguida de morte	bodily injury resulting in death
lesão corporal dolosa	intentional bodily injury
lesão corporal culposa por acidente de trânsito	traffic-related unintentional bodily injury
lesão corporal culposa outras	unintentional bodily injury (other causes)
latrocínio	robbery resulting in death
número de vítimas em latrocínio	number of victims in robberies resulting in death
total de estupro	total rapes
total de roubo outros	total robberies (other types)
roubo de veículo	vehicle robbery
roubo a banco	bank robbery
roubo de carga	cargo theft
furto de veículo	vehicle theft

Source: SEADE (Sistema Estadual de Análise de Dados).

Table 2: Description of variables related to demography, in Portuguese and in English.

Variable's Name	Full Name of Variable
Área	area
Densidade Demográfica	demographic density
Grau	degree of urbanization
Habdom	inhabitant per household
Total H	total of man
Total M	total of woman
idade de zero a catorze	age from zero to fourteen
idade de quinze a vinte e nove	age from fifteen to twenty nine
idade de trinta a cinquenta e nove	age from thirty to fifty nine
idade de sessenta e mais	age above sixty

Source: SEADE (Sistema Estadual de Análise de Dados)

The São Paulo metropolitan region includes the following municipalities: Caieiras, Cajamar, Francisco Morato, Franco da Rocha, Mairiporã, Arujá, Biritiba-Mirim, Ferraz de Vasconcelos, Guararema, Itaquaquecetuba, Guarulhos, Mogi das Cruzes, Poá, Salesópolis, Santa Isabel, Suzano, Diadema, Mauá, Santo André, São Bernardo do Campo, São Caetano do Sul, Ribeirão Pires, Rio Grande da Serra, Cotia, Embu, Embu-Guaçu, Itapecerica da Serra, Juquitiba, São Lourenço da Serra, Taboão da Serra, Vargem Grande Paulista, Barueri, Carapicuíba, Itapevi, Jandira, Osasco, Pirapora do Bom Jesus, and Santana de Parnaíba. Due to incomplete data, some municipalities were excluded from the regression analysis.

Table 3: Description of variables related to education, in Portuguese and in English. The education data mainly focus on the enrollment number of different degrees of school.

Term	Term meaning
tx aprovacao	approval rate
tx reprovacao	disapproval rate
tx abandono	abandon rate
fai publica	public elementary school early years
fai privado	private elementary school early years
faf publica	public elementary school final years
faf privado	private elementary school final years
medio publica	public high school
medio privada	private high school
matrícula creche	daycare enrollment
matriculas pre escola	pre-school enrollment
matriculas ESAI	elementary school early years enrollment
matriculas ESAF	elementary school final years enrollment
matriculas EM	high school enrollment
municipal	municipal
estadual	state
particular	private

Source: SEADE (Sistema Estadual de Análise de Dados)

2.3 Linear Regression and LASSO

Linear regression is a fundamental statistical method widely used across various fields such as economics, finance, medicine, and engineering. It models the relationship between a dependent variable (also referred to as the response or target) and one or more independent variables (also called predictors or features).
The multiple linear regression model is expressed as:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_k x_{ki} + \epsilon_i, \quad i = 1, 2, \ldots, n, \tag{1}$$

where y_i is the dependent variable, $\beta_0, \beta_1, \ldots, \beta_k$ are the regression coefficients, and ϵ_i is the random error term. It is typically assumed that the errors satisfy $\mathbb{E}[\epsilon_i] = 0$, $\text{Var}(\epsilon_i) = \sigma^2$, and are uncorrelated and normally distributed with constant variance (Yan & Su, 2009).
In matrix notation, the model can be compactly written as:

$$y = X\beta + \epsilon, \tag{2}$$

The most common method for estimating the coefficients β is the ordinary least squares (OLS) estimator, which minimizes the residual sum of squares (RSS). The OLS estimator is obtained by solving the optimization problem:

$$\hat{\beta} = \arg\min_{\beta} (y - X\beta)^T (y - X\beta). \tag{3}$$

To find the solution, we compute the gradient of the objective function with respect to β and set it equal to zero:

$$\frac{\partial}{\partial \beta} \left[(y - X\beta)^T (y - X\beta) \right] = 0. \tag{4}$$

Table 4: Description of variables related to industry, in Portuguese and in English. The data mainly describe the total value of a specific sector in a city.

Variable's Name	Full Name of Variable
Valor PIB	GDP value
Valor servico ex	external service value
Valor Agropecuaria	value of agriculture and livestock
Valor Industria	value of industry
Valor Impostos líquido	value of liquid tax
Valor Serviços adm	value of administrative service
Valor Serviços	value of service
Valor Adi	additional value
Valor PIB per capita	value of GDP per capita
VTI alimento	transformed value of food industry
VTI bebidas	transformed value of beverage industry
VTI texteis	transformed value of textile industry
VTI vestuario	transformed value of wearing industry
VTI couro	transformed value of leather industry
VTI madeira	transformed value of wood industry
VTI celulose	transformed value of cellulose industry
VTI impressao	transformed value of print industry
VTI petroleo	transformed value of petroleum industry
VTI biocombustíveis	transformed value of biofuel industry
VTI quimica	transformed value of chemical industry
VTI farmaceutica	transformed value of pharmaceutical industry
VTI borracha	transformed value of rubber industry
VTI min n met	transformed value of mineral and metal industry
VTI metal	transformed value of metal industry
VTI metalurgia	transformed value of metallurgy industry
VTI eletroinformatico	transformed value of electroinformatic industry
VTI maquina	transformed value of machinery industry
VTI outrosequi	transformed value of others equipment industry
VTI moveis	transformed value of furniture industry
VTI prod div	transformed value of diverse products industry
VTI manutencao	transformed value of maintenance industry

Source: SEADE (Sistema Estadual de Análise de Dados)

Expanding and simplifying the above expression yields the normal equations:

$$X^T X\hat{\beta} = X^T y. \tag{5}$$

Assuming that $X^T X$ is invertible (i.e., X has full rank), the solution is given by:

$$\hat{\beta} = (X^T X)^{-1} X^T y. \tag{6}$$

This solution provides the best linear unbiased estimator (BLUE) of β under the Gauss-Markov assumptions.
On the other hand, in statistics, the term *shrinkage* refers to the process of reducing the influence of sampling variability by imposing constraints on model parameters. One of the most well-known shrinkage methods is the Least Absolute Shrinkage and Selection Operator (Lasso), introduced by Tibshirani (1996) and described in detail by Hastie et al. (2009). Lasso performs both variable selection and regularization, improving the prediction accuracy and interpretability of the resulting model.
The Lasso estimator is defined as the solution to the following constrained optimization problem:

$$\hat{\beta}^{\text{lasso}} = \arg\min_{\beta} \sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{k} x_{ij}\beta_j \right)^2 \quad \text{subject to} \quad \sum_{j=1}^{k} |\beta_j| \leq t, \tag{7}$$

where $t \geq 0$ is a fixed threshold that determines the amount of shrinkage applied to the regression coefficients.
An equivalent formulation of the problem is the Lagrangian form:

$$\hat{\beta}^{\text{lasso}} = \arg\min_{\beta} \left\{ \frac{1}{2} \sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{k} x_{ij}\beta_j \right)^2 + \lambda \sum_{j=1}^{k} |\beta_j| \right\}, \tag{8}$$

where $\lambda \geq 0$ is a regularization parameter that controls the strength of the penalty. As λ increases, more coefficients are driven towards zero, effectively performing variable selection. The key feature of Lasso is its ability to produce sparse models by setting some coefficient estimates exactly to zero, which is particularly useful when dealing with high-dimensional data. In this study, Lasso was employed to select the most relevant predictors for crime modeling before applying further spatial analysis using the SAR model.

2.4 SAR model

The SAR (Spatial Autoregressive) model (Kazar & Celik, 2012), also known as spatial lag model or mixed regressive model, is an extension of the linear regression model and is represented by Equation (9)

$$y_i = \rho \sum_{j=1}^{n} W_{ij} y_j + \sum_{q=0}^{k} x_{iq}\beta_q + \epsilon_i \tag{9}$$

where ρ is the spatial autoregression (autocorrelation) parameter, W is n-by-n neighborhood matrix that accounts for the spatial relationships (dependencies) among the spatial data. ϵ_i is the iid error term, y_i is the dependent variable, x_{iq} is the independent variable and β_q is the coefficient of x_{iq}. Particularly, when $q = 0$, it is defined that $x_{i0} = 1$ and the is the β_0 is

the intercept of the model. The main difference between linear regression and SAR model is the inclusion of the spatial autocorrelation term, represented as $\rho \sum_{j=1}^{n} W_{ij} y_j$. This term is introduced to account for the spatial dependencies among the elements of the dependent variable y_i. It is important to note that when ρ is zero, the SAR model reduces to a standard linear regression model.

Incorporating spatial dependencies into models has been demonstrated to enhance overall classification or prediction accuracy in comparison to using linear regression (Kazar & Celik, 2012). Spatial dependency is defined as the connections between adjacent pixels within a spatial framework, typically represented by a regular grid (in this work, a grid could be considered as the municipality). When considering a four-neighborhood, the neighboring pixels of the (i, j)th pixel on the regular grid are depicted in equation (10).

$$neighbors(i,j) = \begin{cases} (i-1,j), 2 \geqslant i \geqslant \phi, 1 \geqslant j \geqslant q, NORTH, \\ (i,j+1), 1 \geqslant i \geqslant \phi, 1 \geqslant j \geqslant q-1, EAST, \\ (i+1,j), 2 \geqslant i \geqslant \phi-1, 1 \geqslant j \geqslant q, SOUTH, \\ (i,j-1), 1 \geqslant i \geqslant \phi, 2 \geqslant j \geqslant q, WEST, \end{cases} \tag{10}$$

where ϕ is the row dimension of the spatial framework (image) and q is the column dimension of spatial framework (image). This (i, j)th pixel of the surface will be the $(\phi \times (i-1) + j)$th row of the matrix C, which is the non-row-standardized (non-normalized) neighborhood matrix of dimension $\phi \times q$ by $\phi \times q$ or n by n.

Kazar and Celik (2012) state the following entries of matrix C

- $\{(\phi \times (i-1) + j), (\phi \times (i-2) + j)\}$,
- $\{(\phi \times (i-1) + j), (\phi \times (i-1) + j + 1)\}$,
- $\{(\phi \times (i-1) + j), (\phi \times (i) + j)\}$
- $\{(\phi \times (i-1) + j), (\phi \times (i-1) + j - 1)\}$

will have "1" as its value, while the other elements in the row will be zero. This is illustrated in Figure 7. On the other hand, the row-standardized neighborhood matrix W is calculated in two steps: first, calculate each row sum, in which Kazar and Celik (2012) stated that W is $\phi \times q$ by $\phi \times q$ or n-by-n. Second, each element in a row is divided by its corresponding row sum. In other words, $W = D^{-1}C$ and the elements of the diagonal matrix D are defined as $d_{ii} = \sum_{i=1}^{n} c_{ij}$ and $d_{ij} = 0$ (Kazar & Celik, 2012), in which d_{ii} and d_{ij} is the element inside of the matrix D, and c_{ij} is the element inside of the matrix C; specially, d_{ii} is the diagonal element of the matrix D.

For the spatial weight matrix of São Paulo's state, the latitude and longitude polygon information was extracted from (Brugnara, 2021), containing the data in geo-json format. For a given city x in the data, if city y shares the same geographic border with x, then the element of xth line, yth column and yth line, xth column in the matrix will be 1; Otherwise, it will be zero. The spatial weight matrix was created following the approach described in R. Bivand et al. (2021), which utilized the R. Bivand (2022b) package.

A derivative-based optimization that can be used in the SAR model solution is the Newton-Raphson (root-finding) algorithm. In this algorithm, we need to compute the first derivative of the log-likelihood function. This gives the location of optimal solution for ρ parameter. The

first derivative of the log-likelihood function is given by the following equation.

$$\frac{\partial lnL(y)}{\partial \rho} = tr((I - \rho W)^{-1} \frac{\partial (I - \rho W)}{\partial \rho})) - \tag{11}$$

$$\frac{n}{2}(\frac{-y^T M^T M W y - y^T W^T M^T M y + 2\rho y^T W^T M^T M W y}{y^T (I - \rho W)^T M^T M (I - \rho W) y}) \tag{12}$$

The term M corresponds to $[I - X(X^T X)^{-1} X^T]$, where X is the design matrix that contains the predictive variables.

2.5 Artificial Neural Networks approach

The design of artificial neural networks (ANNs) is inspired by the architecture of the human brain, which is a highly complex, nonlinear, and parallel information-processing system. As noted by Haykin (1999), the brain performs tasks such as pattern recognition, perception, and motor control with remarkable speed and fault-tolerance, despite being composed of relatively slow and unreliable components. Neural networks aim to emulate this behavior in artificial systems, creating models that can learn from data through experience.
An artificial neural network is a computational model that mimics the way biological neural networks process information. It consists of interconnected nodes-referred to as neurons-organized into layers. Through the adjustment of connection weights and biases, the network learns to approximate functions that map input features to output responses. Neural networks are widely used in diverse applications including medical image diagnosis, financial forecasting, energy demand prediction, process control, and computer vision.
A typical neural network model comprises three key components: the structure of the individual neuron (processing unit), the overall network topology (how neurons are connected), and the learning algorithm (how weights are updated). These elements together determine the capacity and performance of the network.

2.5.1 Structure of an Artificial Neuron

The basic computational unit in a neural network is the artificial neuron. Its operation consists of the following steps:

- Each input signal x_j is associated with a synaptic weight w_{kj}, representing the strength of the connection between input j and neuron k.
- The neuron computes a linear combination of the inputs, producing an intermediate value v_k.
- A bias term b_k is added to shift the activation function's input, allowing more flexibility in the output.
- The result is passed through an activation function $\phi(\cdot)$, which introduces nonlinearity and regulates the output magnitude.

Mathematically, the operation of neuron k is described by the following equations:

$$v_k = \sum_{j=1}^{m} w_{kj} x_j \tag{13}$$

$$y_k = \phi(v_k + b_k), \tag{14}$$

where $x_1, x_2, \ldots, x_m$ are the input features, w_{kj} are the corresponding synaptic weights, b_k is the bias, ϕ is the activation function, and y_k is the neuron's output. The bias acts as an affine transformation of the weighted sum, improving the model's expressiveness.

2.5.2 Application of ANN to Crime Prediction

Artificial Neural Networks can be effectively applied to model security-related phenomena such as crime prediction in urban environments. By analyzing historical crime and socioeconomic data, neural networks are capable of uncovering complex, nonlinear relationships that might not be captured by traditional linear models.

In this study, we apply ANN models, including Long Short-Term Memory (LSTM) networks, to predict crime per capita in municipalities within the São Paulo metropolitan area. These models receive a set of input variables (e.g., social, demographic, and educational indicators) and process them through multiple layers of interconnected neurons. Each layer may use distinct activation functions and transformations, allowing the network to learn sophisticated mappings from input to output.

While the Spatial Autoregressive (SAR) model focuses on linear associations and explicitly accounts for spatial dependencies, it is limited in its ability to capture nonlinear interactions among predictors. The ANN models complement the SAR model by providing a more flexible learning framework capable of modeling general associations between predictors and crime levels.

Specifically, each input variable is treated as a signal to the neural network. Through training, the ANN adjusts its internal weights and biases to minimize prediction error, ultimately outputting the predicted number of crimes per capita for each municipality. This approach allows us to compare the performance of linear and nonlinear models in understanding and forecasting urban crime patterns.

2.6 Long Short-Term Memory (LSTM)

Long Short-Term Memory (LSTM) is a specialized form of recurrent neural network (RNN), designed to overcome the limitations of traditional RNNs in capturing long-term dependencies. In standard RNNs, hidden units are updated through simple recurrence relations, which often suffer from vanishing or exploding gradients when modeling long sequences. LSTM addresses this issue by introducing a memory cell structure with gates that regulate the flow of information through time (Goodfellow et al., 2016).

Instead of relying solely on standard hidden units, LSTM introduces so-called *LSTM cells*, which include mechanisms to control memory retention, input updates, and output flows. A central component of the LSTM cell is the internal memory state $s_i^{(t)}$, which evolves over time with the help of a linear self-loop, similar in spirit to a leaky integrator. A leaky unit can be expressed as:

$$\mu^{(t)} \leftarrow \alpha\mu^{(t-1)} + (1-\alpha)\nu^{(t)}, \tag{15}$$

where $\mu^{(t)}$ is the running average at time t, $\nu^{(t)}$ is the new input value, and $\alpha \in [0, 1]$ controls the memory decay. When α approaches 1, more past information is retained; when it nears 0, the past is forgotten rapidly.

Inspired by this idea, the LSTM model replaces the fixed decay α with a *forget gate* $f_i^{(t)}$, a

trainable mechanism that determines how much of the previous state should be preserved:

$$f_i^{(t)} = \sigma\left(b_i^f + \sum_j U_{i,j}^f x_j^{(t)} + \sum_j W_{i,j}^f h_j^{(t-1)}\right), \tag{16}$$

where $\sigma(\cdot)$ denotes the logistic sigmoid function, $x^{(t)}$ is the current input vector, and $h^{(t-1)}$ is the hidden state from the previous time step. The parameters b_i^f, $U_{i,j}^f$, and $W_{i,j}^f$ are the bias, input weights, and recurrent weights of the forget gate, respectively.
Compared to the artificial neuron in a standard feedforward network, the forget gate's weights U^f play a similar role to the input weights w_{kj}, and the bias b_i^f is analogous to b_k. The major difference lies in the recurrent weights W^f, which incorporate memory from the previous hidden state $h^{(t-1)}$.
The internal state $s_i^{(t)}$ of the LSTM cell is updated via:

$$s_i^{(t)} = f_i^{(t)} s_i^{(t-1)} + g_i^{(t)} \sigma\left(b_i + \sum_j U_{i,j} x_j^{(t)} + \sum_j W_{i,j} h_j^{(t-1)}\right), \tag{17}$$

where $g_i^{(t)}$ is the *input gate*, which determines how much new information to add into the state:

$$g_i^{(t)} = \sigma\left(b_i^g + \sum_j U_{i,j}^g x_j^{(t)} + \sum_j W_{i,j}^g h_j^{(t-1)}\right). \tag{18}$$

Comparing Equations (15) and (17), we see that the forget gate $f_i^{(t)}$ plays a role similar to the decay parameter α, but is learned dynamically at each time step. When $f_i^{(t)}$ is close to 1, the past state $s_i^{(t-1)}$ is retained; when close to 0, it is forgotten.
The output of the LSTM cell, $h_i^{(t)}$, is controlled by an *output gate* $q_i^{(t)}$:

$$h_i^{(t)} = \tanh(s_i^{(t)}) \cdot q_i^{(t)}, \tag{19}$$

with the output gate value computed by:

$$q_i^{(t)} = \sigma\left(b_i^o + \sum_j U_{i,j}^o x_j^{(t)} + \sum_j W_{i,j}^o h_j^{(t-1)}\right). \tag{20}$$

To summarize, an LSTM cell is composed of the internal state $s_i^{(t)}$, the input gate $g_i^{(t)}$, the forget gate $f_i^{(t)}$, and the output gate $q_i^{(t)}$. These components are neuron-like units with trainable parameters, connected recurrently and regulated by nonlinear functions. The gating mechanism allows the model to retain relevant information for long periods and discard irrelevant or outdated signals, making LSTM especially effective for time series data and sequential modeling. In this work, the LSTM model is applied to crime prediction in the study region, using historical data to learn temporal patterns in crime rates. The model receives socioeconomic variables as inputs at each time step and outputs crime-per-capita predictions. Compared to linear spatial models such as SAR, the LSTM network captures richer nonlinear dependencies over time, offering an alternative and complementary approach to urban crime forecasting.

2.7 Why Use LSTM for Crime Prediction?

While traditional time series models such as ARIMA have been widely used for forecasting applications, they rely on strong assumptions of stationarity and linearity. In contrast, crime dynamics are influenced by a complex interplay of social, economic, and spatial factors that evolve nonlinearly over time. Additionally, urban crime often displays seasonal behavior, delayed effects, and long-term memory that cannot be fully captured by simple autoregressive models.

Long Short-Term Memory (LSTM) networks are a class of recurrent neural networks (RNNs) specifically designed to learn long-range dependencies in sequential data. Unlike standard feedforward networks, LSTM models maintain internal memory states and gating mechanisms that allow them to retain relevant temporal information over extended periods. This makes LSTM particularly well-suited for modeling monthly crime data, where the influence of past events can persist for several months.

Furthermore, LSTM networks can incorporate multiple exogenous variables—such as school enrollment, unemployment rates, and population shifts—alongside the target crime variable. This flexibility allows the model to learn rich representations of time-evolving social structures and crime patterns.

3 Results

3.1 Exploratory Analysis

The dataset employed in this study encompasses not only crime-related variables but also a broad range of socioeconomic, educational, and geographic indicators for each municipality. In total, 102 variables were considered. For predictive modeling, the dataset was divided into training and test sets. The training data comprises crime per capita information from 2002 to 2017, totaling approximately 123,000 observations, while the test set consists of crime data from 2018 with 7,728 observations.

To calculate crime per capita, all crime categories in each city for the year 2017 were aggregated and divided by the city's population. Due to the small scale of the resulting values, the crime per capita figures were scaled by a factor of 1 million. The choropleth map presented in Figure 1 illustrates the spatial distribution of crime per capita across the state of São Paulo. Notably, the highest crime rates are concentrated in the metropolitan and coastal regions, suggesting that urbanized and touristic municipalities tend to exhibit higher levels of criminal activity. The horizontal axis represents latitude, while the vertical axis corresponds to longitude.

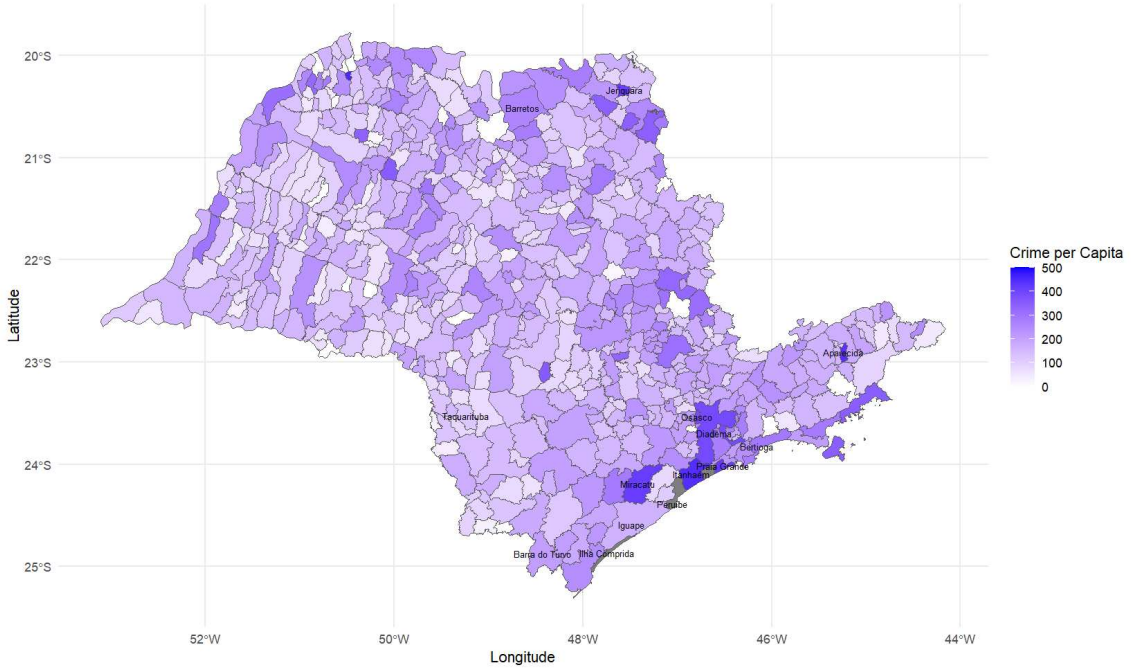


Figure 1: Choropleth map of crime per capita in São Paulo state for the year 2017.
Source: authors' elaboration.

To construct the choropleth map, geospatial polygon data (latitude and longitude boundaries) were obtained from the `geobr` Python package (Brugnara, 2021), which provides geographic data in GeoJSON format. Using the `geopandas` library in Python and `ggplot2` in R, the geographic boundaries of each municipality were plotted to visualize the spatial distribution of crime.

In addition to spatial visualization, we explored the correlation between different crime types in São Paulo state from 2002 to 2021. Each crime category was treated as an individual variable, and a correlogram was constructed to visualize the pairwise relationships. The correlogram, shown in Figure 2, reveals strong correlations among certain categories-such as between different types of rape, and between rape, assault, and homicide. The correlations were computed using Kendall's rank correlation coefficient, a non-parametric measure that assesses the strength of association between two ranked variables.

As this study also focuses on the São Paulo metropolitan region in 2017, corresponding visualizations were created for this subset. Figure 3 presents the choropleth map for crime per capita in the metropolitan region, and Figure 4 provides the correlogram for crime categories within the same area and period.

3.2 SAR model results

Based on the collected data, municipalities with populations exceeding 50,000 inhabitants were retained for the analysis. Variable selection was then performed using the LASSO method, resulting in the inclusion of the following covariates: total crime with one-month lag, total crime with two-month lag, total crime with eleven-month lag, total population, population aged over 60 years, number of students enrolled in municipal preschools, enrollments in state elementary schools (early grades), enrollments in municipal elementary schools (final grades), and enrollments in municipal high schools.

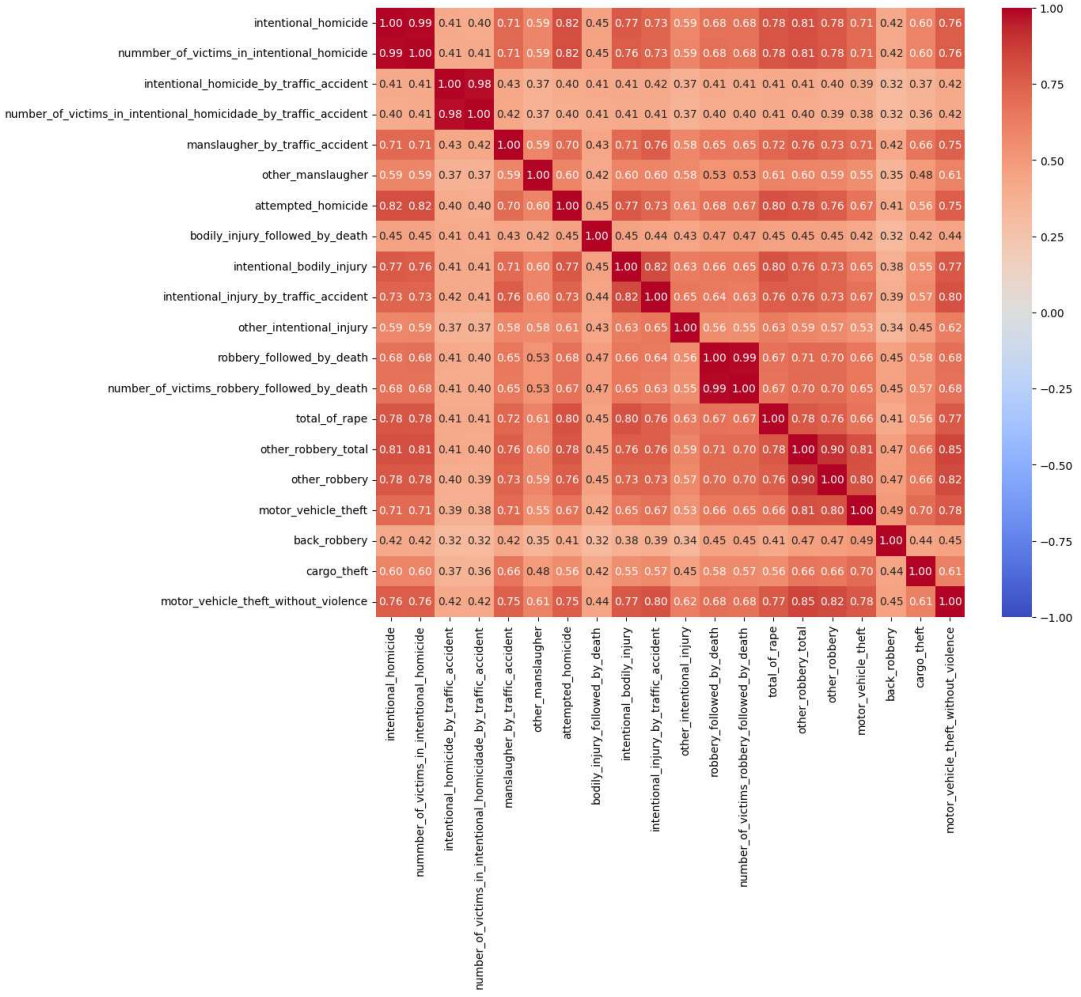


Figure 2: Correlogram's plot of different type of crimes in the São Paulo state. Source: authors' elaboration.

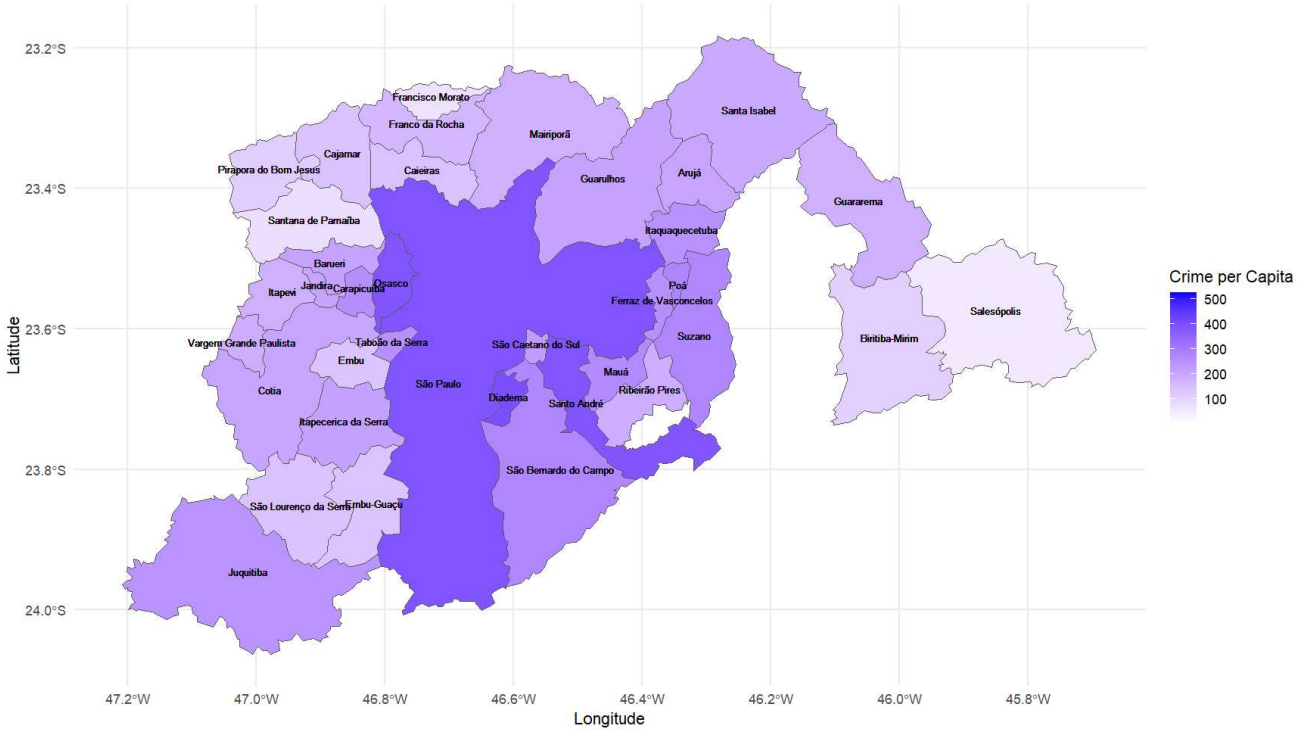


Figure 3: Choropleth plot of the metropolitan area of São Paulo's crime per capita. Figure elaborated by the authors

Subsequently, following the framework described in R. Bivand et al. (2013), implemented in the `spatialreg` package in R, the selected variables were used to fit a spatial autoregressive (SAR) model. The estimation results are reported in Table 5.

Table 5: Fitted SAR model for crimes at the region of greater São Paulo

Variables and Intercept	Estimate	Std. Error	z value	p-value
ρ	0.0022			0.6156
(Intercept)	-47.8745	29.2237	-1.6382	0.1014
total crime 1m of lag	0.0211	0.0441	0.4784	0.6324
total crime 2m of lag	0.0268	0.0432	0.6212	0.5344
total crime 11m of lag	0.5864	0.0559	10.4952	< 0.0010
Population	0.0003	0.0003	1.0246	0.3055
population of age 60+	0.0032	0.0023	1.4104	0.1584
registration at the municipal pre school	0.0107	0.0106	1.0126	0.3112
registration at the state elementary school(beginning)	0.0045	0.0040	1.1377	0.2552
registration at the municipal elementary school(final)	0.0183	0.0067	2.7204	0.0065
registration of municipal high school	-0.0501	0.0258	-1.9417	0.0522

Source: authors' elaboration.

Interestingly, while short-term lags (1–2 months) in crime do not show significant predictive power in the original SAR model, the 11-month lag is highly significant. This may suggest seasonal patterns in crime occurrence, where certain events or institutional cycles repeat annually. For example, school holidays, festivals, or fiscal year-end dynamics could contribute to systematic spikes in crime around similar months each year. This supports the case for using LSTM

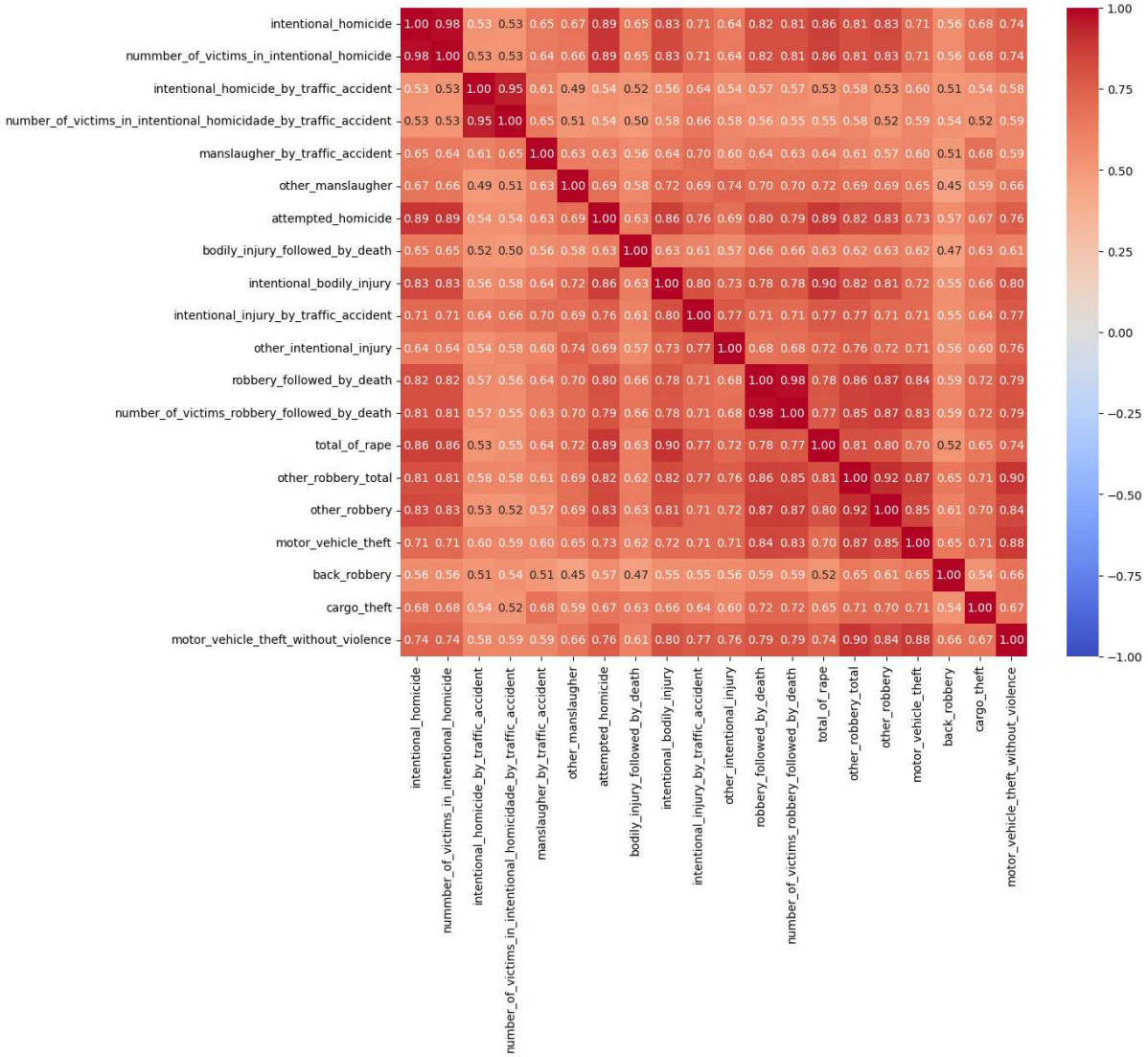


Figure 4: Correlogram's plot of different type of crimes in the metropolitan area of São Paulo. Figure elaborated by the authors

models, which are capable of capturing such complex seasonal effects over long sequences. Preliminary exploratory analysis indicated a pronounced spatial polarization, with systematically higher values concentrated in major metropolitan areas such as São Paulo and Guarulhos. To mitigate skewness and excessive concentration, a logarithmic transformation was applied to all variables. Figure 5 illustrates the resulting improvement in distributional homogeneity.

Also, the variable total crime with 12 months of lag was deleted due to the large proportion of missing values. Then, SAR model was applied again on the data, which generated the result below in table 6

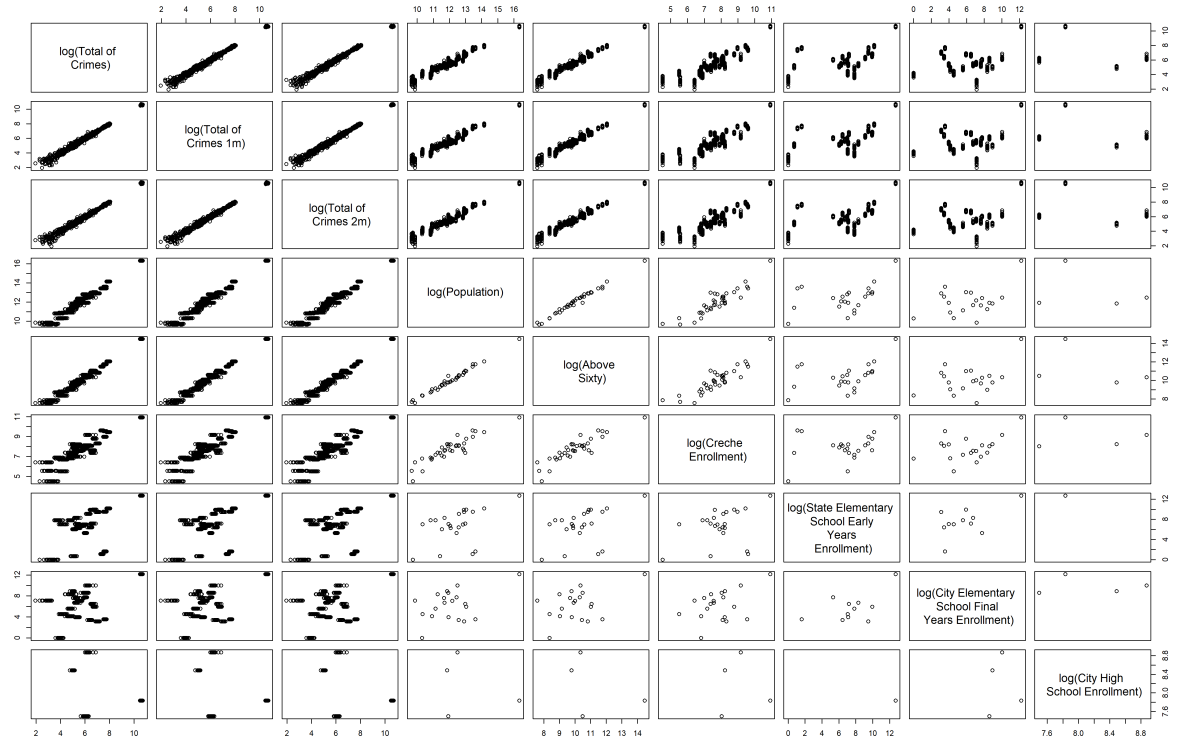


Figure 5: Pair-plot between the predictor variables after applying logarithmic function figure elaborated by the authors

Table 6: Fitted SAR model for crimes at the region of greater São Paulo after applied logarithmic function

Variables and Intercept	Estimate	Std. Error	z value	p-value
ρ	0.07823			< 0.0001
(Intercept)	-5.5047	0.1006	-54.7019	< 0.0001
total crime 1m of lag	0.0055	0.0052	1.0572	0.2904
total crime 2m of lag	0.0110	0.0038	2.9261	0.0034
Population	0.0166	0.0176	0.9446	0.3449
population of age 60+	1.0218	0.0207	49.3454	< 0.0001
registration at the municipal pre school	0.0184	0.0153	1.2076	< 0.0001
registration at the state elementary school(beginning)	0.0022	0.0019	1.1757	< 0.0001
registration at the municipal elementary school(final)	0.0046	0.0016	2.8258	0.0047
registration of municipal high school	-0.0109	0.0026	-4.1751	< 0.0001

Source: authors' elaboration.

Subsequently, the explanatory variable total crime was replaced by crime per capita. The SAR model was then fitted to this transformed dataset, in which all variables had already been log-transformed.

3.3 SAR model simulation

To assess the estimation accuracy of the SAR model, a simulation study was conducted. The simulation procedure consists of the following steps:

1. The matrix of independent variables $\mathbf{X}$ and the spatial weight matrix $\mathbf{W}$ are fixed, together with predefined values of the regression coefficients $\boldsymbol{\beta}$ and the spatial autoregressive parameter ρ, with $\rho \in (-1, 1)$.

2. Random error terms $\boldsymbol{\varepsilon}$ are generated from a normal distribution with mean zero and constant variance.

3. The variable total crime (12-month lag) was excluded due to a high proportion of missing values. Additionally, to address the strong spatial polarization observed in the data, a logarithmic transformation was applied to the covariates in $\mathbf{X}$ in order to mitigate extreme values.

4. The dependent variable $\mathbf{Y}$ is computed according to the SAR data-generating process.

5. Using $\mathbf{Y}$ together with the fixed $\mathbf{X}$ and $\mathbf{W}$, the parameters $\boldsymbol{\beta}$ and ρ are estimated. This procedure is repeated 1,000 times, and the mean of the estimated parameters is computed and compared with the true values.

With the theoretical values fixed and the error term defined as the variance $\sigma^2 = 0.01$ and $\sigma^2 = 0.1$ for every element in the vector, the obtained results are described in Tables 7, 8. Comparing with Table 6, we can see that the results between the model and the simulations match well.

Table 7: SAR model results when $\sigma^2 = 0.01$

Parameters	Theoretical	Estimated	Bias	RMSE
ρ	0.0780	0.0780	<0.0001	0.0004
Intercept	−5.5050	−5.5050	<0.0001	0.0042
Crime lag 1m	0.0050	0.0050	<0.0001	0.0002
Crime lag 2m	0.0110	0.0100	<0.0001	0.0002
Population	0.0160	0.0170	<0.0001	0.0007
Pop. age 60+	1.0220	1.0220	<-0.0001	0.0008
Municipal preschool	0.0180	0.0180	<-0.0001	0.0007
State elem. school (beg.)	0.0020	0.0020	<0.0001	0.0001
Municipal elem. school (fin.)	0.0050	0.0050	<-0.0001	0.0001
Municipal high school	-0.0110	-0.0110	<0.0001	<0.0001

Source: authors' elaboration.

Table 8: SAR model results when $\sigma^2 = 0.1$

Parameters	Theoretical	Estimated	Bias	RMSE
ρ	0.0780	0.0780	<0.0001	<0.0001
Intercept	−5.5050	−5.5050	<0.0001	0.0060
Crime lag 1m	0.0060	0.0060	<0.0001	<0.0001
Crime lag 2m	0.0110	0.0110	<0.0001	<0.0001
Population	0.0170	0.0170	<0.0001	<0.0001
Pop. age 60+	1.0210	1.0220	0.0000	<0.0001
Municipal preschool	0.0190	0.0190	<0.0001	<0.0001
State elem. school (beg.)	0.0020	0.0020	<0.0001	<0.0001
Municipal elem. school (fin.)	0.0050	0.0050	<0.0001	<0.0001
Municipal high school	−0.0110	−0.0110	<0.0001	<0.0001

Source: authors' elaboration.

From Tables 7 and 8, it is possible to recover the parameters values, since the bias and RMSE have small values in every parameter.

Additionally, simulations are also performed with ρ is 0.3 or 0.7, since the two values are neither big nor small, and subsequently repeating the procedure with σ^2 = 0.1. The obtained results are described in Tables 9,10,11,12,13. The simulation results show that the model is not sensitive to different parameter setting, with the exception of ρ.

Table 9: Mean values of betas in different ρ and σ^2 values in SAR model.

parameters	ρ= 0.3, σ^2=0.01	ρ=0.7, σ^2=0.01
ρ	0.3	0.7
Intercept	-5.5050	-5.5060
total crime 1m of lag	0.0060	0.0060
total crime 2m of lag	0.0110	0.0110
Population	0.0170	0.0170
population of age 60+	1.0220	1.0220
registration at the municipal pre school	0.0190	0.0190
registration at the state elementary school(beginning)	0.0020	0.0020
registration at the municipal elementary school(final)	0.0050	0.0050
registration of municipal high school	-0.0110	-0.0110

Source: authors' elaboration.

Table 10: Mean values of betas in different ρ and σ^2 values in SAR model.

parameters	$\rho = 0.3, \sigma^2 = 0.1$	$\rho = 0.7, \sigma^2 = 0.1$
ρ	0.3	0.7
Intercept	-5.5060	-5.5050
total crime 1m of lag	0.0060	0.0060
total crime 2m of lag	0.0110	0.0110
Population	0.0170	0.0160
population of age 60+	1.0220	1.0220
registration at the municipal pre school	0.0180	0.0190
registration at the state elementary school(beginning)	0.0020	0.0020
registration at the municipal elementary school(final)	0.0050	0.0050
registration of municipal high school	-0.0110	-0.0110

Source: authors' elaboration.

Table 11: Parameter's bias and RMSE when $\sigma^2 = 0.1$ and $\rho = 0.3$ in SAR model.

parameters	bias	RMSE
ρ	0	0
Intercept	<0.0001	0.006
total crime 1m of lag	<-0.0001	<0.0001
total crime 2m of lag	<0.0001	<0.0001
Population	<0.0001	<0.0001
population of age 60+	<0.0001	<0.0001
registration at the municipal pre school	<-0.0001	<0.0001
registration at the state elementary school(beginning)	<-0.0001	<0.0001
registration at the municipal elementary school(final)	<-0.0001	<0.0001
registration of municipal high school	<-0.0001	<0.0001

Source: authors' elaboration.

Table 12: Parameter's bias and RMSE when $\sigma^2 = 0.01$ and $\rho = 0.3$ in SAR model.

parameters	bias	RMSE
ρ	0	0
Intercept	<-0.0001	0.001
total crime 1m of lag	<0.0001	<0.0001
total crime 2m of lag	<0.0001	<0.0001
Population	<0.0001	<0.0001
population of age 60+	<0.0001	<0.0001
registration at the municipal pre school	<-0.0001	<0.0001
registration at the state elementary school(beginning)	<0.0001	<0.0001
registration at the municipal elementary school(final)	<-0.0001	<0.0001
registration of municipal high school	<0.0001	<0.0001

Source: authors' elaboration.

Table 13: Parameter's bias and RMSE when $\sigma^2 = 0.1$ and $\rho = 0.7$ in SAR model.

parameters	bias	RMSE
ρ	0	0
Intercept	<0.0001	0.009
total crime 1m of lag	<0.0001	<0.0001
total crime 2m of lag	<0.0001	<0.0001
Population	<0.0001	<0.0001
population of age 60+	<0.0001	<0.0001
registration at the municipal pre school	<0.0001	<0.0001
registration at the state elementary school(beginning)	<0.0001	<0.0001
registration at the municipal elementary school(final)	<0.0001	<0.0001
registration of municipal high school	<0.0001	<0.0001

Source: authors' elaboration.

For the bias and RMSE suppose Y_i is the value of ith element of a vector, $\hat{Y}$ is the fixed value of the parameter and X_i is the value of ith element of another vector. Then, the bias formula can be shown as

$$bias = \frac{\sum_{i=1}^{N}(Y_i - \hat{Y})}{N}$$

the formula of RMSE is

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(Y_i - \hat{Y})^2}{N}}$$

The purpose of these simulations is to assess the estimation accuracy of the SAR model as implemented in the R package described in (R. Bivand, 2022a). Based on the simulation results, it can be concluded that the SAR model implementation provides reliable parameter estimates. Across simulations using both the original parameter values and user-defined parameter settings, the model exhibits good performance in terms of bias and root mean squared error (RMSE).

3.4 Fitted SAR Model and Simulation at the Region of Greater São Paulo from 2002-2017

The dataset includes all explanatory variables spanning the period 2002 to 2017. Consequently, We replicate the same procedure as previously outlined. Notably, the variable "total of crimes" is substituted with "crime per capita". The findings from this analysis are presented in Table 14.

Table 14: Fitted SAR model for crimes at the region of greater São Paulo after applied logarithmic function.

Variables and Intercept	Estimate	Std. Error	z value	p-value
ρ	0.5058			<0.0001
(Intercept)	3.1220	0.1530	20.3960	<0.0001
total crime 1m of lag	0.0040	0.0020	2.6700	0.0080
total crime 2m of lag	0.0060	0.0010	5.1800	<0.0001
Population	−0.7400	0.0280	−26.0820	<0.0001
population of age 60+	0.7860	0.0272	28.9180	<0.0001
registration at the municipal pre school	0.0440	0.0080	5.7846	<0.0001
registration at the state elementary school(beginning)	−0.0030	0.0010	−3.3880	0.0010
registration at the municipal elementary school(final)	0.0080	0.0007	10.1030	<0.0001
registration of municipal high school	−0.0140	0.0010	−11.4370	<0.0001

Source: authors' elaboration.

Evidently, there is a notable enhancement in the model's performance, particularly in the context of the ρ value. This indicates that the geographical factor has provided a constructive explanation for the crime dynamics within the São Paulo metropolitan region.
The fitted SAR model (Table 14) shows that crime per capita is significantly influenced by both demographic and educational variables. Notably, municipalities with a larger proportion of elderly residents (`age 60+`) exhibit higher crime rates, possibly due to their increased vulnerability. In contrast, higher enrollment in municipal high schools is associated with a statistically significant reduction in crime, highlighting the protective role of secondary education. Short-term temporal lags (1 and 2 months) also show significant positive effects, confirming temporal autocorrelation in urban crime dynamics.

3.5 Artificial Neural Networks and LSTM

Artificial neural networks (ANN) and LSTM models were also fitted. The main differences between the SAR model and the ANN/LSTM models lie in the set of covariates and in the treatment of spatial effects. While the SAR model incorporates lagged crime variables (e.g., total crime at 1- and 2-month lags) and explicitly models spatial dependence, the ANN and LSTM models exclude these lagged spatial terms. Instead, population density is included as a proxy for spatial heterogeneity, although spatial dependence is not explicitly modeled as in the SAR framework.

Based on the test data, the results in Table 15 demonstrate the superior predicting performance of the ANN, LSTM models when compared to SAR model for predicting crimes in the metropolitan area of São Paulo state.

Finally, based on the artificial neural networks model, a choropleth plot illustrates the predicted crime per capita in the metropolitan area of São Paulo (Figure 6).

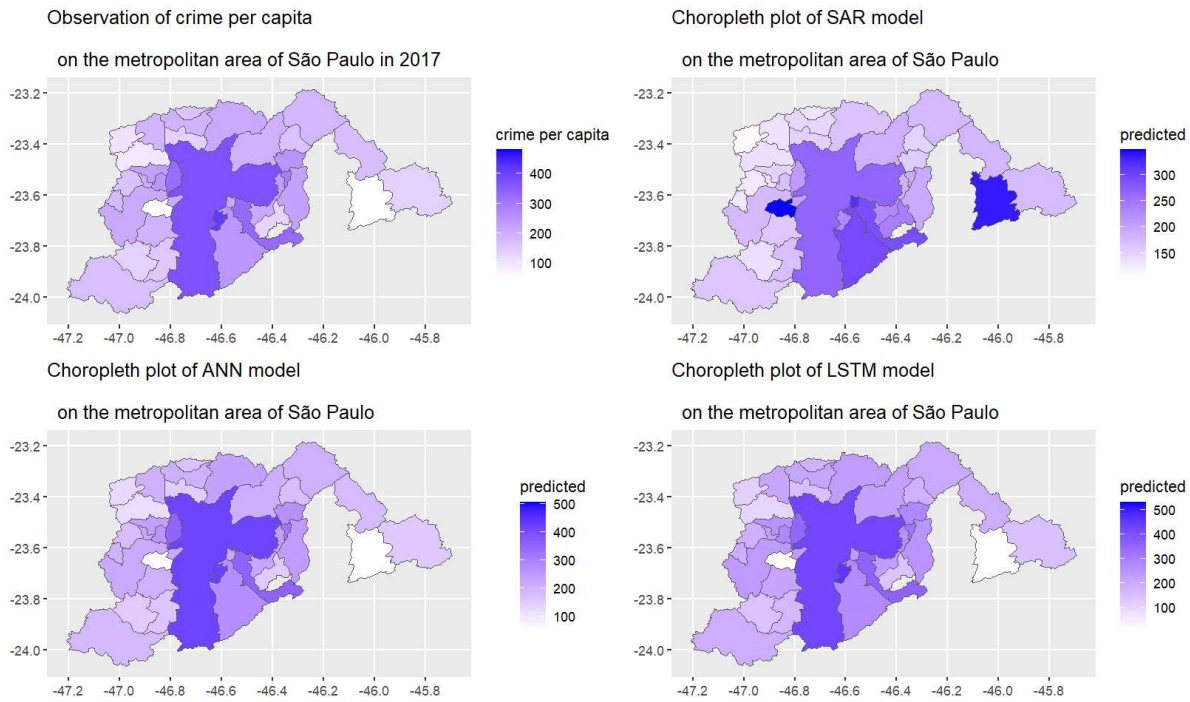


Figure 6: Choropleth plot of crime per capita in the metropolitan area of São Paulo, based on the predicted data.

Table 15: Predictive performance comparison between SAR model, ANN and LSTM

Measure	SAR model	ANN	LSTM
RMSE	77.18	21.5	22.64
MAE	60.77	19.93	19.16

Then, Based on the results of different models, a choropleth plot illustrates the predicted crime per capita in the metropolitan area of São Paulo (Fig. 3).The first figure shows the choropleth plot of crime per capita on the metropolitan area of São Paulo, which is actually the Figure 7. The result of the plot shows the superior performance of the LSTM model in predicting crime per capita in the metropolitan area of São Paulo.
Finally, the standardized residual choropleth plot of the three models is made. Figure 7 presents standardized residual maps for SAR, ANN, and LSTM models. The residuals are computed as:

$$\text{Standardized Residual}_i = \frac{y_i - \hat{y}_i}{\sigma}$$

where y_i is the observed crime rate, $\hat{y}_i$ the predicted value, and σ the standard deviation of residuals. These maps highlight areas where the models over- or under-estimate crime. Notably, SAR shows systematic underestimation in high-crime areas such as São Paulo city, while LSTM provides more balanced residuals spatially.

4 Discussion and Policy Implications

This study provides empirical evidence on the spatial and temporal determinants of crime across municipalities in the São Paulo metropolitan region. By comparing the performance

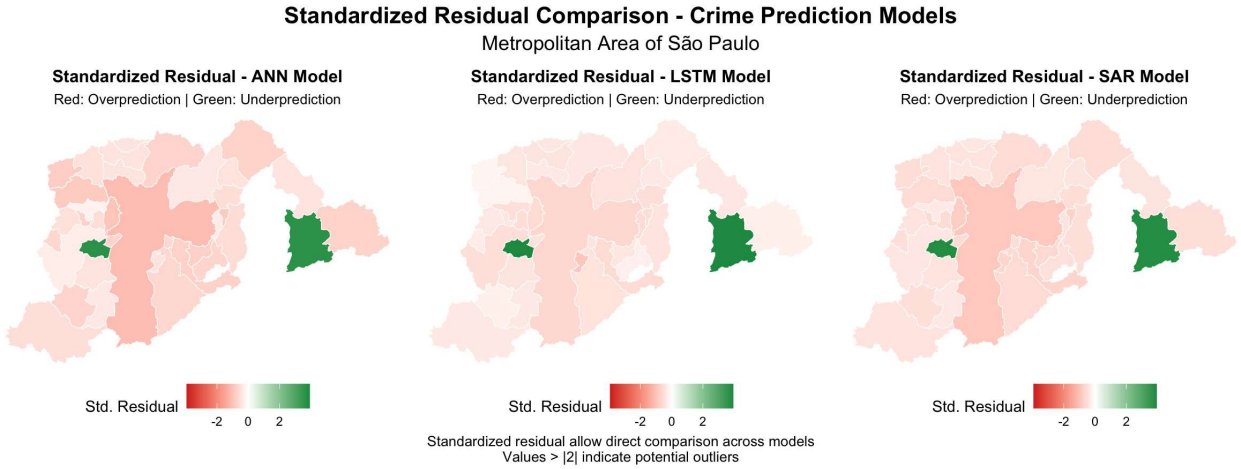


Figure 7: Standardized residual plot of crime per capita in the metropolitan area of São Paulo, based on the predicted data.

of a Spatial Autoregressive (SAR) model with that of Artificial Neural Networks (ANN) and Long Short-Term Memory (LSTM) networks, we demonstrated the complementary strengths of traditional spatial econometrics and deep learning techniques.

The SAR model offered interpretability and statistical insight into which factors are most associated with higher crime levels. Notably, the positive association between crime and the proportion of elderly residents and the negative association with high school enrollment rates underscore the importance of targeted educational and social policies. These findings support previous research that links educational attainment with lower youth vulnerability to crime, suggesting that increasing access to public secondary education may serve as a long-term crime prevention strategy.

Moreover, the strong temporal dependence observed in LSTM results indicates that crime dynamics evolve with consistent monthly and annual patterns. This temporal structure could reflect socio-institutional cycles (e.g., school calendars, holidays, economic cycles), pointing to windows of opportunity for preventive action. For example, law enforcement agencies might benefit from adjusting patrol strategies during months of historically high crime activity, especially in municipalities with identified spatial clusters.

Spatial Intelligence for Crime Prevention

The spatial visualization of standardized residuals and predicted crime rates (Figures 6, 7) offers important diagnostic tools for policy. The SAR model helps identify municipalities that consistently exhibit higher-than-expected crime rates, even after controlling for demographic and educational factors. These spatial "outliers" could be prioritized for further investigation, resource allocation, or social program targeting.

Integrating these predictive outputs into public safety planning systems could support:

- Optimized police deployment based on expected temporal and spatial crime hotspots
- More equitable distribution of educational and community support programs in high-risk areas
- Urban development initiatives tailored to social vulnerabilities (e.g., support for elderly populations, school infrastructure)

Limitations and Future Work

This study has several limitations. Due to data availability, a number of municipalities were excluded from the final model. Additionally, the use of a single aggregated crime indicator ("total crime") may mask different dynamics across crime types (e.g., theft vs. homicide). Disaggregated modeling by crime type would provide more targeted insights for intervention. Furthermore, we excluded data after 2018 to avoid pandemic-related biases; future studies could evaluate model robustness post-COVID.

Conclusion

This research highlights the value of combining interpretable spatial models with high-performance neural networks for urban crime analysis. While SAR models are essential for understanding underlying structural factors and geographic dependence, LSTM networks are better suited for operational forecasting. Together, these tools can support a more nuanced, evidence-based approach to urban safety and social policy design.

References

Anselin, L. (1988). *Spatial econometrics: Methods and models*. Springer.

Bivand, R. (2022a). *Spatialreg: Spatial regression analysis* [R package version 1.2-6]. https://cran.r-project.org/package=spatialreg

Bivand, R. (2022b). *Spdep: Spatial dependence* [R package version 1.2-7]. https://cran.r-project.org/package=spdep

Bivand, R., Hauke, J., & Kossowski, T. (2013). Computing the jacobian in gaussian spatial autoregressive models. *Geographical Analysis*, *45*(2), 150–179.

Bivand, R., Millo, G., & Piras, G. (2021). A review of software for spatial econometrics in r. *Mathematics*, *9*(11), 1276.

Bivand, R. S., Pebesma, E., & Gómez-Rubio, V. (2013). *Applied spatial data analysis with r* (2nd). Springer.

Brown, M. A. (1982). Modelling the spatial distribution of suburban crime. *Economic Geography*, *58*(3), 247–261.

Brugnara, T. (2021). *Geodata-br package* [Python package]. Retrieved September 14, 2025, from https://raw.githubusercontent.com/tbrugz/geodata-br/master/geojson/geojs-35-mun.json

Caplan, J. M., Kennedy, L. W., & Miller, J. (2011). Risk terrain modeling: Brokering criminological theory and gis methods for crime forecasting. *Justice Quarterly*, *28*(2), 360–381.

Cerqueira, D., Bueno, S., Lima, R. S., et al. (2021). Atlas da violência 2021 [Accessed: 2025-06-26]. https://www.ipea.gov.br/atlasviolencia/

Chun, S. A., Tak, S.-H., Tavakkol, S., & Choi, J. (2019). Crime prediction model using deep neural networks. *Proceedings of the 20th Annual International Conference on Digital Government Research*, 512–514.

Fundação Sistema Estadual de Análise de Dados (SEADE). (2025). *Seade foundation*. Retrieved September 14, 2025, from https://www.seade.gov.br/

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning*. Springer.

Haykin, S. (1999). *Neural networks: A comprehensive foundation*. Prentice Hall.

Instituto Brasileiro de Geografia e Estatística (IBGE). (2025). *Instituto brasileiro de geografia e estatística*. Retrieved September 14, 2025, from https://www.ibge.gov.br/

Kazar, B. M., & Celik, M. (2012). Spatial autoregression (sar) model: Parameter estimation techniques. In *Spatial multidimensional data models and techniques* (pp. 89–109). Springer. https://doi.org/10.1007/978-3-642-29613-3_5

Melo, S. N. d., Matias, L. F., & Andresen, M. A. (2015). Crime concentrations and similarities in spatial crime patterns in a brazilian context. *Applied Geography*, *62*, 314–324.

Reier Forradellas, R. F., Náñez Alonso, S. L., Jorge-Vazquez, J., & Rodriguez, M. L. (2020). Applied machine learning in social sciences: Neural networks and crime prediction. *Social Sciences*, *10*(1), 4.

Secretaria de Segurança Pública do Estado de São Paulo. (2021). *Dados estatísticos da secretaria de segurança pública do estado de são paulo (via base dos dados)*. Retrieved September 14, 2025, from https://basedosdados.org/

Sherman, L. W., Gartin, P. R., & Buerger, M. E. (1989). Hot spots of predatory crime: Routine activities and the criminology of place. *Criminology*, *27*(1), 27–56.

United Nations Office on Drugs and Crime. (2023). Global study on homicide 2023 [Accessed: 2025-06-26]. https://www.unodc.org/unodc/en/data-and-analysis/global-study-on-homicide.html

Wang, T., Rudin, C., Wagner, D., & Sevieri, R. (2013). Learning to detect patterns of crime. *Machine Learning*, *93*(1), 103–130.

Wheeler, A. P., & Steenbeek, W. (2021). Mapping the risk terrain for crime using machine learning. *Journal of Quantitative Criminology*, *37*(2), 445–480.

Yan, X., & Su, X. (2009). *Linear regression analysis: Theory and computing*. World Scientific.

Yi, F., Liu, Y., Zhang, C., Wang, J., & Li, Z. (2019). Neural network based continuous conditional random field for fine-grained crime prediction. *Proceedings of the 28th International Joint Conference on Artificial Intelligence (IJCAI)*, 4157–4163.