Crime prediction through urban metrics and statistical learning

Luiz G.A. Alves [a],*, Haroldo V. Ribeiro [b], Francisco A. Rodrigues [a]

[a] *Institute of Mathematics and Computer Science, University of São Paulo, São Carlos, SP, Brazil*
[b] *Departamento de Física, Universidade Estadual de Maringá, Maringá, PR, Brazil*

HIGHLIGHTS

- Predictive analysis is applied to crime data.
- Correlation between crime and urban metrics.
- Quantification of the importance of urban metrics in predicting crime.

ARTICLE INFO

ABSTRACT

Understanding the causes of crime is a longstanding issue in researcher's agenda. While it is a hard task to extract causality from data, several linear models have been proposed to predict crime through the existing correlations between crime and urban metrics. However, because of non-Gaussian distributions and multicollinearity in urban indicators, it is common to find controversial conclusions about the influence of some urban indicators on crime. Machine learning ensemble-based algorithms can handle well such problems. Here, we use a random forest regressor to predict crime and quantify the influence of urban indicators on homicides. Our approach can have up to 97% of accuracy on crime prediction, and the importance of urban indicators is ranked and clustered in groups of equal influence, which are robust under slightly changes in the data sample analyzed. Our results determine the rank of importance of urban indicators to predict crime, unveiling that unemployment and illiteracy are the most important variables for describing homicides in Brazilian cities. We further believe that our approach helps in producing more robust conclusions regarding the effects of urban indicators on crime, having potential applications for guiding public policies for crime control.

1. Introduction

Social phenomena increasingly attract the attention of physicists driven by successful application of methods from statistical physics for modeling and describing several social systems, including the collective phenomena emerging from the interactions of individuals [1], the spread of ideas in social networks [2], epidemic spreading [3], criminal activity [4], political corruption [5], vaccination strategies [6], and human cooperation [7]. In the particular case of crime, this interest trace back to works of Quetelet, who coined the term "social physics" in the 19th century [8]. On the one hand, traditional physics methods have proved to be useful in understanding phenomena outside conventional physics [9,10]. On the other hand, recently,

---

* Corresponding author.
 E-mail address: lgaalves@usp.br (L.G.A. Alves).

several problems from physics have been addressed through the lenses of machine learning methods, including topics related to phases of matter [11], quantum many-body problem [12], phase transitions [13], phases of strongly correlated fermions [14], among others. As physicists had added these new tools to the box [15], naturally, social physics problems could also be addressed using such ideas. In particular, in this work, we are interested in understanding the relationships between crime and urban metrics by using statistical learning.

Crime and violence are ubiquitous in society. Throughout history, organized societies have tried to prevent crime following several approaches [16]. In this context, understanding the features associated with crime is essential for achieving effective policies against these illegal activities. Studies have linked crime with several factors, including psychological traits [17,18], environmental conditions [19,20], spatial patterns [4,21,22], and social and economic indicators [23–26]. However, it is easy to find controversial explanations for the causes of crimes [27]. Methodological problems in data aggregation and selection [28,29], errors related to data reporting [30], and wrong statistical hypothesis [27] are just a few issues which can lead to misleading conclusions.

A significant fraction of the literature on statistical analysis in criminology tries to relate the number of a particular crime (*e.g.* robbery) with explicative variables such as unemployment [31] and income [32]. In general, these analyses are carried out by using ordinary-least-squares (OLS) linear regressions [33]. These standard linear models usually assume that the predictors have weak exogeneity (error-free variables), linearity, constant variance (homoscedasticity), normal residual distribution, and lack of multicollinearity. However, when trying to model crime, several of these assumptions are, often, not satisfied. When these hypotheses do not hold, conclusions about the factors affecting crime are likely to be misconceptions.

Recently, researchers have promoted an impressive progress on the analysis of cities, where one of the main findings is that the relationship between urban metrics and population size is not linear, but it is well described by a power-law function [34–40]. Crime indicators scale as a superlinear function of the population size of cities [33,35,37]. Other indicators (commonly used as predictors in linear regression models for crime forecasting) also exhibit power-law behavior with population size. These metrics are categorized into sub-linear (*e.g.* family income [37]), linear (*e.g.* sanitation [37]), and super-liner (*e.g.* GDP [33,34,37,41]), depending on the power-law exponent characterizing the allometric relationship with the population size [34]. In addition, the relationships between crime and population size as well as urban metrics and population size have some degree of heteroscedasticity [34], and most of these urban indicators also follow heavy-tailed distributions [38,42]. Thus, it is not surprising to find controversial results about the importance of variables for crime prediction when so many assumptions of linear regressions are not satisfied.

A possible approach to overcome some of these issues is to apply a transformation to the data in order to satisfy the assumptions of linear regressions. For instance, Bettencourt et al. [35] (see also [33,37]) employed scaled-adjusted metrics to linearize the data and provide a fair comparison between cities with different population sizes. By considering these variables and applying corrections for heteroscedasticity [43], it is possible to describe 62% of the variance of the number of homicides in function of urban metrics [37]. Also, by the same approach, researchers have shown that simple linear models account for 31%–97% of the observed variance in data and correctly reproduce the average of the scale-adjusted metric [33]. However, the data still have co-linearities which can lead to misinterpretation of the coefficients in the linear models [33,37].

A better approach to crime prediction is the use of statistical learning methods (*e.g.* [44]). Regression models based on machine learning can handle all the above-mentioned issues and are more suitable for the analysis of large complex datasets [45]. For instance, decision trees are known to require little preparation of the data when performing regression [46–49]. Tree-based approaches are also considered a non-parametric method because they make no assumption about the data. Among other advantages, these learning approaches map well non-linear relationships, usually display a good accuracy when predicting data, and are easy to interpret [46–49].

Here, we consider the random forest algorithm [46,50–52] to predict and quantify the importance of urban indicators for crime prediction. We use data from urban indicators of all Brazilian cities to train the model and study necessary conditions for preventing underfitting and overfitting in the model. After training the model, we show that the algorithm predicts the number of homicides in cities with an accuracy up to 97% of the variance explained. Because of the high accuracy and easy interpretation of this ensemble tree model, we identify the important features for homicide prediction. Unlike simple linear models adjusted through OLS, we show that the importance of the features is stable under slight changes in the dataset and that these results can be used as a guide for crime modeling and policymakers.

2. Methods and results

2.1. Data

For our analysis, we choose the number of homicides at the city level as the crime indicator to be predicted. Homicide is the ultimate expression of violence against a person, and thus a reliable crime indicator because it is almost always reported. In Brazil, the report of this particular crime to the Public Health System is compulsory, and these data are aggregated at the city level and made freely available by the Department of Informatics of the Brazilian Public Health System — DATASUS [53]. As possible predictor variables of crime, we select 10 urban indicators (also at the city level) available from the Brazilian National Census that took place in 2000. They are: child labor (fraction of the population aged 10 to 15 years who is working or looking for work), elderly population (citizens aged 60 years or older), female population, gross domestic product (GDP), illiteracy (citizens aged 15 years or older who are unable to read and write at least a single ticket in the language they
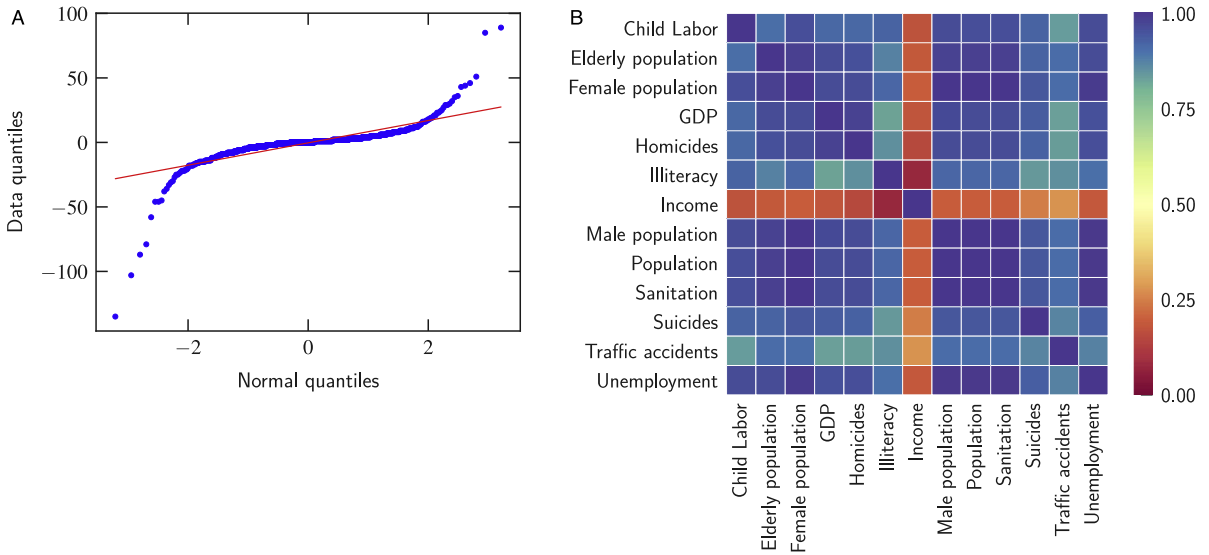
Fig. 1. Non-Gaussian residual and multicollinearity in urban metrics. (A) Quartile–quartile plot comparing the residuals of the linear with a Gaussian distribution (represented by the straight red line). We observe systematic and huge deviations from normality, which is confirmed by Kolmogorov–Smirnov normality text (*p*-value $< 10^{-16}$). (B) The matrix of the Pearson linear correlations among the urban indicators (the predictor variables). We observe that except for income, all indicators are strongly correlated to each other, confirming the presence of multicollinearity in this dataset.

know), family income (average household incomes of family residents, in Brazilian currency), male population, population, sanitation (number of houses that have piped water and sewerage), and unemployment (citizens aged 16 years or older who are without working or looking for work). We have also considered as possible crime-predicting variables the number of traffic accidents and suicides (these data are also aggregated at the city level and from the same year of the census). We have chosen these indicators because they are common in studies correlating crime with socioeconomic indicators [27]. We have further included the number of homicides in the year 2000 as a predictor for homicides in 2010 for investigating a possible autocorrelated behavior. There are thus 13 urban indicators (including also the homicide indicator) from the year 2000 in our dataset that will be used as predictors of the number of homicides 10 years later. We choose a time interval of 10 years because the characteristic time for changes in social indicators has been estimated to be of the order of decades [33,35]. However, we also investigate how the accuracy changes as the time-lag increases from one to ten years considering homicide data from the years 2001 to 2010.

2.2. Problems with usual linear models

Many statistical models try to relate crime with possible explicative variables through OLS linear regressions [27]. Thus, it is usually assumed that crime is a linear function of the explicative variables, i.e., crime rate = f(explicative variables), where f is a linear function [27]. In the context of our dataset, a naive approach to model homicides in Brazilian cities is to consider the following linear regression

$$H(t + \Delta t) = a H(t) + \sum_{n=1}^{12} b_n X_n(t) + \xi(t), \tag{1}$$

where the dependent variable $H(t + \Delta t)$ is the number of homicides in the year 2010, $H(t)$ is the number of homicides in 2000, $X_n(t)$ is the nth ($n = 1, 2, \ldots, 12$) urban indicator in 2000, a and b_n are the linear coefficients, and $\xi(t)$ is a random noise normally distributed (with zero mean and unitary variance) that accounts for unobserved determinants of homicides. It is worth noting that this model has a lagged dependent variable as an instrumental variable, which may cause problems when the error $\xi(t)$ is autocorrelated in time [54].

The OLS linear regression of Eq. (1) requires the residuals to follow a normal distribution. However, this conditions is not satisfied as shown by the quantile–quantile plot in Fig. 1A, and confirmed by the Kolmogorov–Smirnov normality test (*p*-value $< 10^{-16}$). Usually, researchers account for this problem by applying some transformation to data. This may eventually solve the normality problem, but collinearities may exist among urban indicators, and ignoring this fact can lead to controversial results, even when linear models display high predicting accuracies.

According to the urban scaling hypothesis [33–35,37,39,40], the urban indicators in our dataset are dependent of the population size. Thus, each indicator can be written in terms of the population size N as $X_n \sim N^{\alpha_n}$, where α_n is the urban scaling exponent, a condition that introduces correlations among all indicators. This property in regression models is called

multicollinearity of the variables, and we can check for that in our data by evaluating the Pearson correlation coefficient among all urban indicators, as shown in Fig. 1B. From this correlation matrix, we verify that there are significant correlations practically between all pair of variables, violating another condition assumed by most regression models.

If we ignore the assumptions required by linear regression models and perform the OLS fitting, we could eventually find good predictions. In fact, the linear model of Eq. (1) explains 87% of the variance in our dataset. It is an appealing result since it is easy to interpret the coefficients of a linear model. However, the feature importance quantified by regression models can be very sensitive to changes in the dataset, such as by considering undersampling or bootstrapping methods [55]. For instance, by bootstrapping our dataset and applying the linear model of Eq. (1), we find that (depending on the sample used) city's income can be positively, negatively, or even uncorrelated with homicides. These facts could explain some of the inconsistencies reported in the literature about crime. For instance, Entorf et al. [56] found that higher income and urbanization are related to higher crime rates, whereas Fatjnzylber et al. [57] claimed that average income is not correlated with violent crime, but higher urbanization is associated with higher robbery rates, and not with homicides. The same disagreements happen for unemployment, punishment and deterrence, among others indicators [16,33,37]. A possible way to improve the OLS regression model could be adding the variables in a stepwise fashion or considering non-linear models (e.g., the negative binomial regression) to overcome some of these problems. Unfortunately, in the crime literature, we often find results based on simple OLS linear regression, which can mislead the conclusions about the causes of crime. We could further explore and check for other inconsistencies in this simple linear regression by following the several approaches proposed in the literature of linear regression [54]. Nevertheless, we now focus on an alternative approach to overcome these limitations and provide a meaningful rank of features under slight changes in the dataset. Our approach is much more flexible, requiring no transformation or data preparation to achieve good results regarding accuracy in predictions and meaning of results.

2.3. Random forest algorithm

Random forest is an ensemble learning method used for classification or regression that fits several decision trees using various sub-samples of the dataset and aggregate their individual predictions to form a final output and reduce overfitting [46–49,58]. In the random forest regression, several independent decision trees are constructed by a bootstrap sampling of the dataset. The final output is the result of a majority voting among the trees (estimators) or the average over all trees. The process of "bagging features" selects more often the best metrics to describe the data splits, and consequently, make them more important on the majority voting [52]. Unlike usual linear models, the random forest is invariant under scaling and various other transformations of the feature values. It is also robust to the inclusion of irrelevant features and produces very accurate predictions [47]. These properties of the random forest algorithm make it especially suitable for crime forecasting, because of the multicollinearity and non-linearities present in urban data.

Within the framework of random forest regression, the linear model of Eq. (1) can be rewritten as

$$H(t + \Delta t) = \frac{1}{B} \sum_{b=1}^{B} T(H(t), X_1(t), \ldots, X_n(t); \Theta_b), \tag{2}$$

where B is the number of trees, T represents each tree adjusted to data with parameters expressed by the vector Θ_b. The components of Θ_b are related to the maximum depth of the tree, splitting variables, cutting points at each node, and the terminal-node values [45].

One common question that arises when using machine learning algorithms is related to underfitting and overfitting. These behaviors appear when estimating the best trade-off that minimizes the bias and variance errors [48]. Bias is related to errors in the assumptions made by the algorithm that causes the model to miss relevant information about the features describing the data. In other words, the model is too simple to describe the data behavior, what is called underfitting. On the other hand, if we increase the complexity of the model by adding a great number of parameters, the model can become quite sensitive to the noise of the training set, causing an increase in the variance of the outputs, what is called overfitting. The random forest has two main parameters controlling the trade-off between bias and variance errors: the number of trees and the maximum depth of each tree. If a forest has only a few trees with a depth that goes to few layers of nodes, we are likely underfitting the data. However, if there are too many trees splitting the data in a lot of nodes to make the decisions, there is a chance that we are making choices based on the noise of the data rather than the true underlying behavior.

To determine the set of parameters which avoid underfitting and overfitting of data, we use the stratified k-fold cross-validation [47,59] with $k = 10$ for estimating the validation curves for a range of values of the parameter number of trees (Fig. 2A) and maximum depth (Fig. 2B). If the number of trees and maximum depth are both slightly smaller than 10, we find that the training score and validation score are small. As we increase the number of trees and maximum depth, both training and cross-validation scores increase and reach constant plateaus, which indicates that up to 10^4 trees or a maximum depth equal to 10^7 there is no overfitting.

Another question is about how much data we need to train our model. This is a relevant question because we can eventually introduce more noise to the model by adding unnecessary data. To answer this question, we apply again the stratified k-fold cross-validation with $k = 10$ to estimate the learning curves for a range of values of random fractions of the dataset (Fig. 2C). We observe that the more data used for training the model, the better the cross-validated scores. However, we already have good cross-validated scores with $\approx 20\%$ of the data (about 1000 cities).
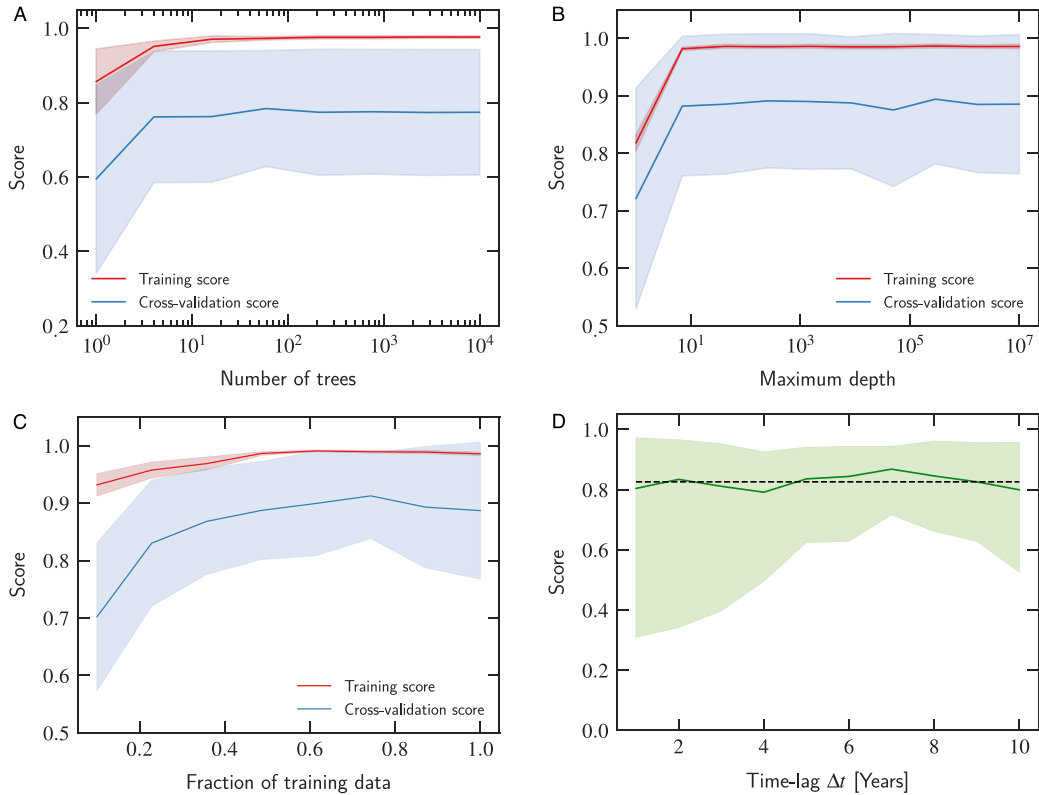
Fig. 2. Validation and learning curves for the random forest regressor. The solid lines represent the average score for the training set (red lines) and the cross-validated score (blue lines) for (A) different number of trees and (B) different maximum depth used in the model. For a forest with less than 10 trees and a maximum depth smaller than 10, we are underfitting the data. There is no overfitting up to 10^4 trees and a maximum depth of 10^7. (C) The dataset includes information of more than 5000 cities, and the learning curve suggests that good cross-validated scores are obtained with ≈20% (1000 observations) of the data to train the model. We observe that the training and cross-validated scores are practically the same for data fractions larger than ≈60%. In the previous plots, the shaded areas represent the confidence intervals defined by the standard deviation of the scores of the random forest algorithm. (D) The accuracy is not dependent on the time-lag Δt. The solid line shows the average score as a function of the time-lag Δt, and the shaded area stands for the 95% confidence interval. The dashed line is the average score (≈82%) over all different time-lags. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

2.4. Predicting crime with the random forest regressor

We have to tune the model by searching for the best combination of parameters that enhances the performance of the random forest algorithm. The previous analysis of the validation and learning curves help to build a grid of parameters to seek the combination that improves our accuracy. To find the best combination of trees and the maximum depth of the trees, we use a grid-search algorithm with the stratified k-fold cross-validation (with $k = 10$) as implemented in the Python library scikit-learn [60]. This procedure exhaustively searches for the best combination of parameters that optimizes the score from a specified grid of parameter values. For our data, we find that the best accuracy (on average) is achieved for 200 trees and maximum depth equal 100.

Having the model properly trained, we can now use the random forest regressor to make crime predictions. Randomly splitting the data into 80% for training and 20% for testing, we obtain up to 97% of accuracy in our predictions with an average adjusted−R^2 equal to 80%. Previous results using the same data and simple linear models combined with scaled-adjusted metrics were able to predict homicides with an accuracy ranging from 38% to 39% (see supplementary material of Ref. [33]). Fig. 3A shows the empirical data versus the random forest predictions for a realization (one data splitting) of the algorithm. Because the dataset is split randomly, different runs can return different outputs and different scores. Fig. 3B depicts the probability distribution (computed via kernel density estimation method) of the adjusted−R^2 for 100 different splits, where we observe that the values are concentrated around the peak at 80%.

2.5. Features importance

The rank of features (urban indicators) importance is another intriguing question for a better understanding of crime. As previously discussed, usual linear regression models have found different answers to this question [16,56,57]. Here, we use
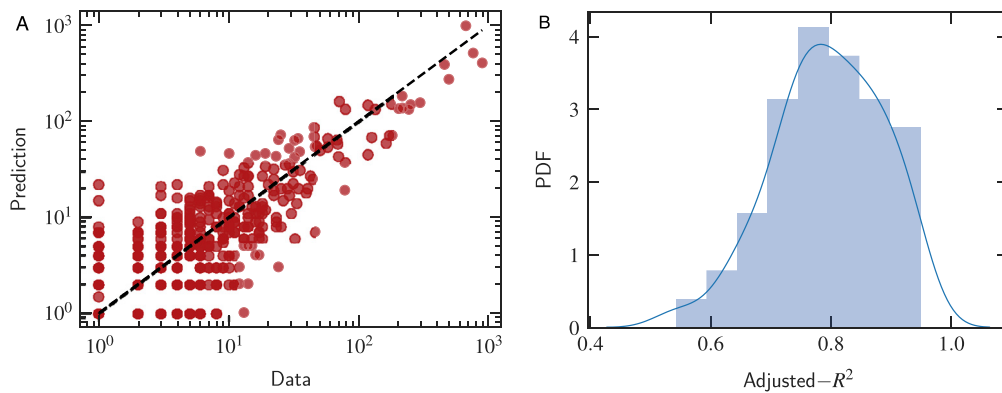
Fig. 3. Data versus prediction. (A) The red dots are the 20% of the data used for testing the model versus the predicted values obtained from the random forest regressor trained with 80% of data. (B) The distribution of the adjusted $-R^2$ obtained by using different random splits of the data (training and test). The boxes represent the histogram, and the solid line is the probability distribution function estimated via kernel density method.

the random forest algorithm to identify the most important urban indicators to predict crime and test whether this ranking of features is robust under slightly changes on the dataset.

The importance of a feature in a tree can be computed by accessing its contribution to the decision process. In each node of a single tree, the variable X_n that have the best improvement in the squared error risk is used to split the region associated with that node into two subregions [47]. Thus, Breiman et al. [61] proposed that the relative importance of a feature X_n in a tree can be calculated by the sum of the improvements overall internal nodes of the tree where X_n is used for splitting. The generalization of this measure to the random forest is the average importance of the features over all trees in the model, as implemented in the Python library scikit-learn [60].

We use this metric to calculate the importance of the features describing the number of homicides in our data. Because of the slight modifications in the dataset caused by the splits in training and test samples, the feature importance varies for each realization of the algorithm. To verify whether these changes affect the importance rank of the features, we calculate the importance of the features for 100 different samples and compute a box-plot ranked by the median importance of the outputs returned by the different samples (Fig. 4A). We check whether the differences in importance are statistically significant by computing the p-values of the t-Student test (testing if two samples have identical average values) with Bonferroni corrections (that corrects the p-values for multiple comparisons [62]). The resulting p-values are shown in Fig. 4B and ranked by the median in correspondence with the box-plot. From this figure, we note that unemployment is the most important feature to describe the crime, followed by illiteracy and male population. The next urban indicators in order of importance form groups of indistinguishable importance (squares on the diagonal matrix of Fig. 4B), i.e., the fourth most important feature is actually a group that includes female population, population, and sanitation. This group of features is followed by other three groups: child labor and homicides (at the fifth position), traffic accidents and elderly population (at the sixth position), and suicides and income (at the seventh position). The less important feature to predict crime is the GDP of cities. It is interesting to note that the number of homicides in past is a poor predictor of homicides in future. It is also worth remarking that despite some fluctuations in the rank caused by the different samples used to train the model, the ranking of features remains the same. This result demonstrates the robustness of the random forest algorithm for detecting the importance of features.

3. Discussion and conclusions

The accurate predictions obtained through statistical learning suggest that crime is quite dependent on urban indicators. The easy interpretation and good accuracies of the random forest algorithm show that this model is an excellent solution for predicting crime and identify the importance of features, even under small perturbations on the training dataset. We cannot assert from the rank of Fig. 4A which features have a positive and negative contribution to the number of homicides. We could try to decompose the contributions of each feature in each node of the trees, and calculate an average over all nodes and trees to identify the gradients, as described by Hastie et al. [47]. However, this decomposition is problematic because features can contribute differently depending on thresholds imposed by the constraints of the sample chosen to train the model. The signals of the contributions of a particular feature can vary for different thresholds, even when the average rank of importance remains the same [61]. Thus, despite the great stability of the results, it is hard to identify whether the variables cause a positive or a negative effect on the number of homicides. While one can speculate whether the most important indicators found in our results cause crime to increase or decrease, further investigation is necessary to understand the local effects of urban metrics on crime.

Because unemployment has a very large variability at local level [28], Blumstein [63] argue that different data aggregation can lead to different conclusions on whether this indicator affects crime or not. There is also evidence supporting the idea
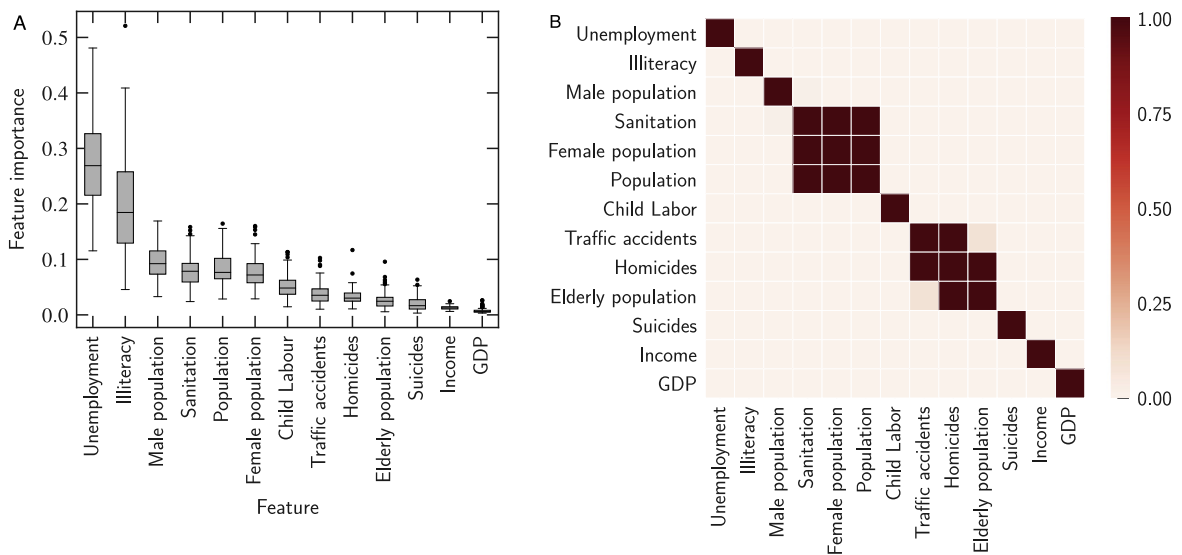
Fig. 4. The important urban metrics for predicting the crime of homicide. (A) The box-plot shows the rank the importance of urban metrics by their medians over different random sample splits. The black lines dividing the boxes represent the median, boxes represent the interquartile range, upper and lower whiskers bars are the most extreme non-outlier data points, and dots are the outliers. (B) The matrix plot shows the *p*-values of *t*-Student test comparing whether the positions in the rank are different. We identify four groups of features that are equally important in the rank (squares on the diagonal matrix).

that crime is affected by unemployment only when this indicator exceeds a given threshold [37]. This is somehow similar to what random forest does when separating the hyperplane of features by thresholds and classifying the number of homicides according to the different values of the urban indicators. The algorithm cannot indicate whether unemployment contributes positively to crime; however, it indicates that this indicator is the most important for describing crime among the set of 12 features in our dataset. Particularly, a recent work has shown that the raising of shooting in schools is related to unemployment rate across different geographic aggregation levels (national, regional and city) [64], which is in agreement with our findings.

The second most important feature for describing crime is illiteracy, which has been associated with violence by other works [65]. Previous works on scaled-adjusted metrics have also shown that the levels of illiteracy are correlated to the number of homicides [33,37]. A report made by the Canadian Police further shows that people with low literacy skills are less likely to involve in group activities than those with higher literacy skills [66]. Consequently, low literate people often feel isolated and vulnerable, making them more likely to involve in violence and crime [66]. The male population is the third most important feature for describing homicides and has also been linked to high levels of violence [33,37,67]. As discussed by Hesketh and Xing [67], the surplus of male population increases the marginalization in society and is linked to antisocial behavior and violence. A similar result was reported in Ref. [37], where it was found that cities with more man than the expected (in terms of the urban scaling) usually have more crimes. Similarly, our results suggest that male population plays an important role in defining the levels of crime in cities. Notice that together unemployment, illiteracy, and male population are responsible for explaining 78% of the variance in our dataset.

It is worth mentioning that previous results based on the same data and simple linear models combined with scaled-adjusted metrics were able to predict homicides with accuracy ranging from 38% to 39% (see supplementary material of Ref. [33]), a much lower score compared to the ones obtained with the random forest approach. However, more sophisticated methods could improve even more this accuracy, and this could be the target of future investigations on crime. The literature of crime still lacks a full comparison with other methods (such as non-linear models, Bayesian regression, and other machine learning techniques), a work that can unveil new properties of the data as well as improve the accuracy of predictions.

Finally, we believe that the application of machine learning for identifying urban indicators that correlate with crime helps to settle the discussion about whether an indicator is important or not for describing a particular crime type. The results of our analysis can further be used as a guide for building other crime models and may help policymakers in the seek of better strategies for reducing crime. Indeed, our results indicate that unemployment and illiteracy levels play an important role in defining the number of homicides in Brazil.

Acknowledgments

References

[1] C. Castellano, S. Fortunato, V. Loreto, Statistical physics of social dynamics, Rev. Modern Phys. 81 (2009) 591.
[2] A. Pentland, Social physics: how good ideas spread — the lessons from a new science, in: EBL-Schweitzer, Scribe Publications Pty Limited, 2014.
[3] R. Pastor-Satorras, C. Castellano, P. Van Mieghem, A. Vespignani, Epidemic processes in complex networks, Rev. Modern Phys. 87 (2015) 925.
[4] M.R. D'Orsogna, M. Perc, Statistical physics of crime: A review, Phys. Life Rev. 12 (2015) 1–21.
[5] H.V. Ribeiro, L.G. Alves, A.F. Martins, E.K. Lenzi, M. Perc, The dynamical structure of political corruption networks, J. Complex Netw. cny002 (2018) 1–15.
[6] Z. Wang, C.T. Bauch, S. Bhattacharyya, A. d'Onofrio, P. Manfredi, M. Perc, N. Perra, M. Salathé, D. Zhao, Statistical physics of vaccination, Phys. Rep. 664 (2016) 1–113.
[7] M. Perc, J.J. Jordan, D.G. Rand, Z. Wang, S. Boccaletti, A. Szolnoki, Statistical physics of human cooperation, Phys. Rep. 687 (2017) 1–51.
[8] A. Quetlet, Physique Sociale, Bachelier, Paris, 1869.
[9] S. Galam, Sociophysics: A Physicist's Modeling of Psycho-Political Phenomena, Springer Science & Business Media, 2012.
[10] R. Conte, N. Gilbert, G. Bonelli, C. Cioffi-Revilla, G. Deffuant, J. Kertesz, V. Loreto, S. Moat, J.-P. Nadal, A. Sanchez, et al., Manifesto of computational social science, Eur. Phys. J. Spec. Top. 214 (2012) 325.
[11] J. Carrasquilla, R.G. Melko, Machine learning phases of matter, Nat. Phys. (2017).
[12] G. Carleo, M. Troyer, Solving the quantum many-body problem with artificial neural networks, Science 355 (2017) 602–606.
[13] E.P. van Nieuwenburg, Y.-H. Liu, S.D. Huber, Learning phase transitions by confusion, Nat. Phys. 13 (2017) 435–439.
[14] K. Ch'ng, J. Carrasquilla, R.G. Melko, E. Khatami, Machine learning phases of strongly correlated fermions, Phys. Rev. X 7 (2017) 031038.
[15] L. Zdeborová, Machine learning: New tool in the box, Nat. Phys. 13 (2017) 420–421.
[16] M.B. Gordon, J.R. Iglesias, V. Semeshenko, J.-P. Nadal, Crime and punishment: the economic burden of impunity, Eur. Phys. J. B 68 (2009) 133–144.
[17] M.R. Kamaluddin, N. Shariff, A. Othman, K.H. Ismail, G.A.M. Saat, Linking psychological traits with criminal behaviour: A review, ASEAN J. Psychiatry 16 (2015) 13–25.
[18] M.R. Gottfredson, T. Hirschi, A General Theory of Crime, Stanford University Press, 1990.
[19] J.L. Gamble, J.J. Hess, Temperature and violent crime in dallas, texas: relationships and implications of climate change, Western J. Emergency Med. 13 (2012) 239.
[20] S.M. Hsiang, M. Burke, E. Miguel, Quantifying the influence of climate on human conflict, Science 341 (2013) 1235367.
[21] M.B. Short, M.R. D'Orsogna, V.B. Pasour, G.E. Tita, P.J. Brantingham, A.L. Bertozzi, L.B. Chayes, A statistical model of criminal behavior, Math. Models Methods Appl. Sci. 18 (2008) 1249–1267.
[22] L.G.A. Alves, E.K. Lenzi, R.S. Mendes, H.V. Ribeiro, Spatial correlations, clustering and percolation-like transitions in homicide crimes, Europhys. Lett. 111 (2015) 18002.
[23] G.S. Becker, Crime and punishment: An economic approach, in: The Economic Dimensions of Crime, Springer, 1968, pp. 13–68.
[24] I. Ehrlich, The deterrent effect of capital punishment: A question of life and death, 1973.
[25] J.Q. Wilson, G.L. Kelling, Broken windows, Atlantic Monthly 249 (1982) 29–38.
[26] E.L. Glaeser, B. Sacerdote, J.A. Scheinkman, Crime and social interactions, Q. J. Econ. 111 (1996) 507–548.
[27] M.B. Gordon, A random walk in the literature on criminality: A partial and critical view on some statistical analyses and modelling approaches, European J. Appl. Math. 21 (2010) 283–306.
[28] S.D. Levitt, Alternative strategies for identifying the link between unemployment and crime, J. Quant. Criminol. 17 (2001) 377–390.
[29] W. Spelman, Specifying the relationship between crime and prisons, J. Quant. Criminol. 24 (2008) 149–178.
[30] M.D. Maltz, J. Targonski, A note on the use of county-level ucr data, J. Quant. Criminol. 18 (2002) 297–318.
[31] S. Raphael, R. Winter-Ebmer, Identifying the effect of unemployment on crime, J. Law Econ. 44 (2001) 259–283.
[32] M. Kelly, Inequality and crime, Rev. Econ. Stat. 82 (2000) 530–539.
[33] L.G.A. Alves, R.S. Mendes, E.K. Lenzi, H.V. Ribeiro, Scale-adjusted metrics for predicting the evolution of urban indicators and quantifying the performance of cities, PLoS One 10 (2015) e0134862.
[34] L.M. Bettencourt, J. Lobo, D. Helbing, C. Kühnert, G.B. West, Growth, innovation, scaling, and the pace of life in cities, Proc. Natl. Acad. Sci. 104 (2007) 7301–7306.
[35] L.M. Bettencourt, J. Lobo, D. Strumsky, G.B. West, Urban scaling and its deviations: Revealing the structure of wealth, innovation and crime across cities, PLoS One 5 (2010) e13541.
[36] L.G.A. Alves, H.V. Ribeiro, R.S. Mendes, Scaling laws in the dynamics of crime growth rate, Physica A 392 (2013) 2672–2679.
[37] L.G.A. Alves, H.V. Ribeiro, E.K. Lenzi, R.S. Mendes, Distance to the scaling law: a useful approach for unveiling relationships between crime and urban metrics, PLoS One 8 (2013b) e69580.
[38] L.G.A. Alves, H.V. Ribeiro, E.K. Lenzi, R.S. Mendes, Empirical analysis on the connection between power-law distributions and allometries for urban indicators, Physica A 409 (2014) 175–182.
[39] Q.S. Hanley, D. Lewis, H.V. Ribeiro, Rural to urban population density scaling of crime and property transactions in english and welsh parliamentary constituencies, PLoS One 11 (2016) e0149546.
[40] J.C. Leitao, J.M. Miotto, M. Gerlach, E.G. Altmann, Is this scaling nonlinear? Roy. Soc. Open Sci. 3 (2016) 150649.
[41] L.G.A. Alves, H.V. Ribeiro, F.A. Rodrigues, The role of city size and urban metrics on crime modeling, in: Understanding Crime Through Science - Interdisciplinary Analysis and Modeling of Criminal Activities, 2018, ArXiv preprint arXiv:1712.02817.
[42] M. Marsili, Y.-C. Zhang, Interacting individuals leading to zipf's law, Phys. Rev. Lett. 80 (1998) 2741.
[43] R. Davidson, J.G. MacKinnon, Estimation and inference in econometrics, in: JSTOR, 1993.
[44] H.-W. Kang, H.-B. Kang, Prediction of crime occurrence from multi-modal data using deep learning, PLoS One 12 (2017) e0176244.
[45] L. Breiman, Statistical modeling: The two cultures, Qual. Control Appl. Stat. 48 (2003) 81–82.
[46] L. Breiman, Random forests, Mach. Learn. 45 (2001) 5–32.
[47] T. Hastie, R. Tibshirani, J. Friedman, The elements of statistical learning: Data mining, inference, and prediction, in: Springer Series in Statistics, Springer, New York, 2013.
[48] G. James, D. Witten, T. Hastie, R. Tibshirani, An introduction to statistical learning: with applications in R, in: Springer Texts in Statistics, Springer, New York, 2014.
[49] G. Death, K.E. Fabricius, Classification and regression trees: a powerful yet simple technique for ecological data analysis, Ecology 81 (2000) 3178–3192.
[50] T.K. Ho, Random decision forests, document analysis and recognition, in: Proceedings of the Third International Conference on, Vol. 1, IEEE, 1995, pp. 278–282.
[51] Y. Amit, D. Geman, Shape quantization and recognition with randomized trees, Neural Comput. 9 (1997) 1545–1588.
[52] T.K. Ho, The random subspace method for constructing decision forests, IEEE Trans. Pattern Anal. Mach. Intell. 20 (1998) 832–844.
[53] Brazil's Public healthcare System (SUS), Department of Data Processing (DATASUS), 2017. (Accessed 1 June 2017).
[54] J. Wooldrige, Introductory Econometrics: A Modern Approach, South-Western College Pub, 2003.

[55] B. Efron, R. Tibshirani, An Introduction to the Bootstrap, in: CRC Monographs on Statistics & Applied Probability, Chapman & Hall/, 1994.

[56] H. Entorf, H. Spengler, Socioeconomic and demographic factors of crime in germany: Evidence from panel data of the german states, Int. Rev. Law Econ. 20 (2000) 75–106.

[57] P. Fajnzylber, D. Lederman, N. Loayza, What causes violent crime? Eur. Econ. Rev. 46 (2002) 1323–1357.

[58] L. Breiman, Bagging predictors, Mach. Learn. 24 (1996) 123–140.

[59] R. Kohavi, et al., A study of cross-validation and bootstrap for accuracy estimation and model selection, in: Ijcai, Vol. 14, Stanford, CA, pp. 1137–1145.

[60] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, E. Duchesnay, Scikit-learn: Machine learning in Python, J. Mach. Learn. Res. 12 (2011) 2825–2830.

[61] L. Breiman, J. Friedman, C.J. Stone, R.A. Olshen, Classification and Regression Trees, CRC press, 1984.

[62] G.M. Ruper Jr., Simultaneous Statistical Inference, Springer Science & Business Media, 2012.

[63] A. Blumstein, Crime modeling, Oper. Res. 50 (2002) 16–24.

[64] A. Pah, J. Hagan, A. Jennings, A. Jain, K. Albrecht, A. Hockenberry, L. Amaral, Economic insecurity and the rise in gun violence at us schools, Nat. Hum. Behav. 1 (2017) 0040.

[65] T.C. Davis, R.S. Byrd, C.L. Arnold, P. Auinger, J.A. Bocchini, Low literacy and violence among adolescents in a summer sports program, J. Adolescent Health 24 (1999) 403–411.

[66] Literacy and policing project of the Canadian association of chiefs of police, Literacy awareness resource manual for police, 2008 (Accessed 1 June 2017).

[67] T. Hesketh, Z.W. Xing, Abnormal sex ratios in human populations: causes and consequences, Proc. Natl. Acad. Sci. 103 (2006) 13271–13275.