AESTIMUM
CENTRO STUDI DI ESTIMO E DI ECONOMIA TERRITORIALE - Ce.S.E.T.

Firenze University Press
www.fupress.com/ceset

Citation: Oliveira, A. A. F., Reyes-Bueno, F., González, M. A. S. & Silva, E. (2024). Comparing traditional and machine learning techniques in apartments mass appraisal in Fortaleza, Brazil. *Aestimum* 85: 21-38. doi: 10.36253/aestim-15344

Received: November 15, 2023

Accepted: July 2, 2024

Published: February 14, 2025

Data Availability Statement: The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

ORCID:
AAFO: 0000-0003-3890-4219
FRB: 0000-0002-5646-0263
MASG: 0000-0002-1975-0026
EdS: 0000-0001-9724-8384

Comparing traditional and machine learning techniques in apartments mass appraisal in Fortaleza, Brazil

Antônio Augusto Ferreira de Oliveira[1], Fabián Reyes-Bueno[2], Marco Aurelio Stumpf González[3,*], Everton da Silva[4]

[1] *Municipal treasury auditor of Fortaleza, Brazil*
[2] *Facultad de Ciencias Exactas y Naturales, Universidad Técnica Particular de Loja, Loja, Ecuador*
[3] *Polytechnic School, Universidade do Vale do Rio dos Sinos, São Leopoldo, Brazil*
[4] *Geosciences Department, Universidade Federal de Santa Catarina, Florianópolis, Brazil*
E-mail: augusto.oliveira@sefin.fortaleza.ce.gov.br; frreyes@utpl.edu.ec; mgonzalez@unisinos.br; everton.silva@ufsc.br
*Corresponding author

Abstract. Mass appraisal has significant applications, such as urban planning, real estate appraisal, and property tax. Due to the challenges of analyzing massive models, they are often developed using semi-automatic assessment methods and machine learning techniques. This article explores different appraisal model methods that utilize statistics and machine learning. It also looks at incorporating spatial information to see if the chosen method can effectively capture the typical spatial dependency of the real estate market. This can help reduce the spatial autocorrelation observed in the residuals. The study compared nine machine learning methods with traditional statistical approaches using a dataset of over 43,000 apartments in Fortaleza, Brazil. The results of the machine learning algorithms were similar. The XGBoost minimized spatial autocorrelation. The easiest interpretations were with MRA, M5P, and MARS techniques. Although, these techniques had the greatest residual spatial autocorrelations. There is a trade-off between the methods, depending on whether the aim is to improve accuracy or provide a clear explanation for property taxation.

Keywords: semi-automatic assessment methods, mass appraisal techniques, machine learning.
JEL codes: O18, R33.

1. INTRODUCTION

The real estate market is a segment of the economy, and as such, the importance of traded goods is measured through the sales prices reached because of buyer and seller agreements. This market presents macro and microeconomic aspects. Macroeconomic aspects are related to government decisions, the conduct of the economy, international influences, interest rates, and national and regional economic growth, among others. Microeconomic aspects are linked to the pricing decisions of real estate agents (companies and families), and they are related to local sociocultural issues.

The price of a property is proportional to its utility, measured as a quality index. Market agents consider several elements, including the property's physical attributes, spatial context (location), and market conditions.

Physical attributes refer to the property's characteristics, such as size, number of rooms, building standard, and age. The location aspect includes accessibility and neighborhood quality and suggests the product's spatial immobility. The market condition consists of the current preferences in social and cultural terms, the economic context, and the transaction terms, such as payment method, interest rate, and time of sale.

Mass appraisal, a systematic process to value multiple properties simultaneously using standardized methods and statistical models, is crucial for efficient and consistent property valuation, especially for taxation and urban planning. Mass appraisal methods should start from a representative sample of price data for the most diverse building typologies. De Cesare et al. (2023) highlight the relevance of systematizing data collection in forming a real estate market observatory. In general data used is collected from buyers and sellers, real estate agents, internet portals, and official government information. With this data, conducting an Exploratory Spatial Data Analysis is advisable to ensure the representativeness of prices throughout the study area. After this stage, automated valuation models (AVMs) can be employed. Due to the heterogeneous nature of properties, several attributes must be considered simultaneously in the appraisal models, assuming different weights in the formation of prices for each kind of property, and it is more common to develop models for one specific segment (for land, houses, commercial properties, and so on). The hedonic pricing theory is the theoretical basis behind price modeling (Rosen, 1974; Sheppard, 1999). A hedonic price model represents the price as a function of the property attributes. Nevertheless, these attributes are not directly priced, and the relationship between attributes and property prices can be understood as indirect or implicit prices (Rosen, 1974).

In practical terms, enough data must be collected to build pricing models. Several techniques could connect a set of independent variables to a dependent variable (in this case, the market price) through an equation. The objective is to develop a numerical model explaining relationships and estimating values. In the traditional approach, coefficients are estimated through multiple regression analysis (MRA). Several conditions (assumptions) must be checked to ensure the quality of the regression model. Among them are homoscedasticity, linearity of the relationship, absence of perfect multicollinearity (especially using several explanatory variables), non-existence of serial or spatial correlation, and lack of significant, un-explicated errors (outliers). In the presence of one or more of these statistical problems, the model loses performance or is even invalidated.

2. LITERATURE REVIEW

A relevant question in the real estate market is to estimate the influence of location (a non-directly measured attribute). Consequently, it is essential to verify and control spatial dependence. A literature review was conducted by searching in SCOPUS database. We select journal papers in English, not including conferences or preprints, using the query:

("semi-automatic valuation" OR "automatic valuation methods" OR "AVM" OR "mass appraisal" OR (("property price" OR "house price" OR "housing price"))) AND ("Spatial Error Regression" OR "decision trees" OR "Multiple Adaptive Regression Splines" OR "M5 pruned" OR "ensemble decision trees" OR "Random Forest" OR "Quantile Random Forest" OR "Gradient Boosting Machine" OR "XGBoost" OR "CATBoost" OR "LightGBM" OR "Deep Learning" OR "machine learning" OR "artificial inteligence" OR "artificial intelligence")

In a first view, there was removed papers about other issues, such as energy, covid and sustainable construction, among others, resulting in a sample of 281 papers, in the 2004 – 2024 period. We selected articles with comparative studies and then selected journals with the highest IF (after JCR). These articles are cited in Table 1.

Hedonic price models based on MRA have been used for a long time (in urban markets, at least since 1970). Following the literature, some overviews on property valuation modeling indicate that there are still several shortcomings in traditional hedonic-MRA models (Wang and Li, 2019; Jayantha and Oladinrin, 2020; Geerts and De Weerdt, 2023), especially about locational attributes and the consideration of the spatial behavior of real estate market (Heyman et al., 2018; Rico-Juan and La Paz, 2021; Chen et al., 2023; Rey-Blanco, 2024). Likewise, the need to verify spatial dependence also has long been pointed out, and different alternatives, such as geographically weighted regressions (GWR), Spatial Regressions, and regression-kriging, have been proposed (Anselin, 1988; Can, 1992; Dubin, 1992; Hengl et al., 2007).

Table 1. Some research comparing the prediction performance among various models.

Authors and goals	Techniques	Data
Zurada et al. (2011) - comparison among regression, ML, and other Artificial intelligence (AI) models	Regression models: MRA, SVM-SMO (Support Vector Machines using Sequential Minimal Optimization), additive regression, M5P trees, and AI-based methods (MBR, neural networks, RBFNN)	222,000 tax assessment of residential properties in Louisville, Kentucky, using 143 variables
Antipov & Pokryshevskaya (2012) - first attempt to use Random Forest in residential estate mass appraisal.	Random Forest, MRA, CHAID, CART, KNN, Artificial Neural Networks (MLP and RBF), and Boosted Trees.	2,848 two-room apartments in Saint-Petersburg, Russia, using 17 variables.
Park & Bae (2015) - investigate improving the accuracy of machine learning techniques in housing price prediction.	C4.5, RIPPER, Naïve Bayesian e AdaBoost	5359 townhouses housing data in Fairfax County, Virginia
Reyes-Bueno et al. (2018) - evaluating less subjective methodologies	Linear regression, M5P, MARS	410 land plot sales transactions - 2003–2009 - rural sector of the Vilcabamba parish, southern Ecuador
Oliveira (2020) - comparing ML techniques for land parcels appraisal	MRA, Random Forest, and XGBoost	8,209 land sales and listing prices in Fortaleza, Brazil - 2015-2019 with 39 variables
Ho et al. (2021) - comparing ML techniques for real estate appraisal	Machine learning algorithms: GBM, RF, and SVM	about 40,000 housing transactions in 18 years, Hong Kong
Rico-Juan & La Paz (2021) - investigate the precision and non-linear relationships between housing prices and housing attributes in the real estate market.	Machine learning: AdaBoost, CatBoost, Decision Tree, Nearest Neighbours, Random Forest, and XGBRegressor. Hedonic and Quantile regression.	About 56,000 dwelling individuals sold on the observation market in 5 years in Spain.
Iban (2022) - investigates eXplainable Artificial Intelligence (XAI) methods that can be integrated with mass real estate appraisal studies.	Tree-based ML regressors, RF, XGBoost, LightGBM, and Gradient Boosting, were compared with multiple regression analysis.	1002 samples and 43 independent variables, Mersin, Turkey
Hu et al. (2022) - considering spatial autocorrelation in modeling house prices with machine learning algorithms	Linear regression and RF in four models: alone, with Local Moran's I (LM) /Local spatial autocorrelation (SA) measures; geocoding coordinate variables (x,y) and spatial eigenvectors	17,028 single-family house sales in 2016-2017 in Fairfax County, Virginia
Baur et al. (2023) - investigate multiple models with different numerical presentations and baselines, textual descriptions, and improvement analysis of the model in predicting property prices with additional features input.	Linear regression, Elastic net, Support Vector Regression, Random Forest, and a Gradient Boosting Algorithm (LightGBM)	30,218 rental apartment offers in Berlin, and 33,610 house purchase offers in Los Angeles
Hurley & Sweeney (2024) - investigate the impact of address mislabeling on predictive performance (with a more distinguished post-code)	Text Mining of Price Prediction Features and in sequence AVM, using Hedonic regression and GAM and Machine Learning Approaches: Decision Trees, K-Nearest Neighbors, and RF	5,208 property sales - from January to November 2018, Dublin, Ireland

Source: Cited Authors.

Mass valuation models accentuate these issues, which have a broader range of attributes, many properties to be assessed, and comprehensive spatial coverage (often the entire city or a region). Furthermore, some models must be rebuilt or renewed regularly. For instance, the modeling for tax purposes has an annual period.

Thus, in the case of mass appraisal, it is not productive to work with "manual" techniques, and techniques should be used with some degree of automation. The increasing use of machine learning (ML) in this field indicates promising options. Property appraisal has undergone a significant transformation in recent years with the advent of Machine Learning techniques (Ho et al., 2021). These methods have revolutionized how properties are valued, making the process faster, more accurate, and less subjective. Studies provided by Antipov and Pokryshevskaya (2012), Wang and Li (2019), Alfaro-Navarro et al. (2020), Hong et al. (2020), Al-Qawasmi (2022), Renigier-Biłozor et al. (2022), Rico-Juan and La Paz (2021), Iban (2022), Kayakuş et al. (2022), Yağmur et al. (2022), Baur et al. (2023), Belmiro et al. (2023), Bilgilioğlu and Yılmaz (2023), Gunes (2023) and Doan et al. (2024) show an overview of the potential in developing applications. In short, one can see that machine learning algorithms are more flexible, and they have lower demands for data. Table 1 shows previous studies and research which compares prediction performance between various models.

3. RESEARCH METHOD

In this paper, we compare different Machine Learning techniques for the mass appraisal of properties in Fortaleza, Brazil. According to the Brazilian Institute of Geography and Statistics (IBGE, 2022), Fortaleza is the fourth-largest city in Brazil, with 2.4 million inhabitants in 2022. Its rapid urbanization makes it an ideal location for this case study. We analyze the performance of several machine learning models, including linear regression, decision trees, random forests, and neural networks, and evaluate their effectiveness in predicting property values. In addition to performance metrics known in machine learning and evaluation institutes such as the International Association of Assessing Officers (IAAO, 2013), we tested the spatial autocorrelation among models' residuals as an indication of poor specification regarding the lack of more spatial proxies' variables to explain the observed prices. In addition, we assessed whether there were significant differences between the models using the Wilcoxon paired test. In Brazil's taxation system on property transfer, apartments are the typology that most appear in real estate transactions. Thus, the accuracy of the predictions of market values must be easy to implement, explain, and update, and regression models must be avoided in taxation. Our results provide insights into the strengths and limitations of different techniques and can inform policy decisions related to property valuation in Fortaleza and other cities facing similar challenges. We do not focus on the interpretation of explanatory variations of the model, which is the objective of another study.

The methodological process followed is shown in Figure 1.

3.1. Data collected

The data set of 43,585 records was obtained from the listed prices, and actual sales of apartments in 2017 and 2021 were recorded in the Real Estate Market Observatory maintained by the Local Finance Secretary of Fortaleza. The original dataset was randomly split into a training set with 32,688 samples and a test set with 10,897 samples with stratification by seven classes of observed prices. All chosen models used these same split datasets.

Fortaleza, the capital city of the State of Ceará, located in the northeast of Brazil, has nearly 2.4 million inhabitants, the fourth largest population in the country, and occupies an area of 312.21 km² (IBGE, 2022). It is a very touristic city, bathed by the Atlantic Ocean in the north and east with 34 km of coastline. As provided in the Constitution of the Federative Republic of Brazil (1988), tax lia-

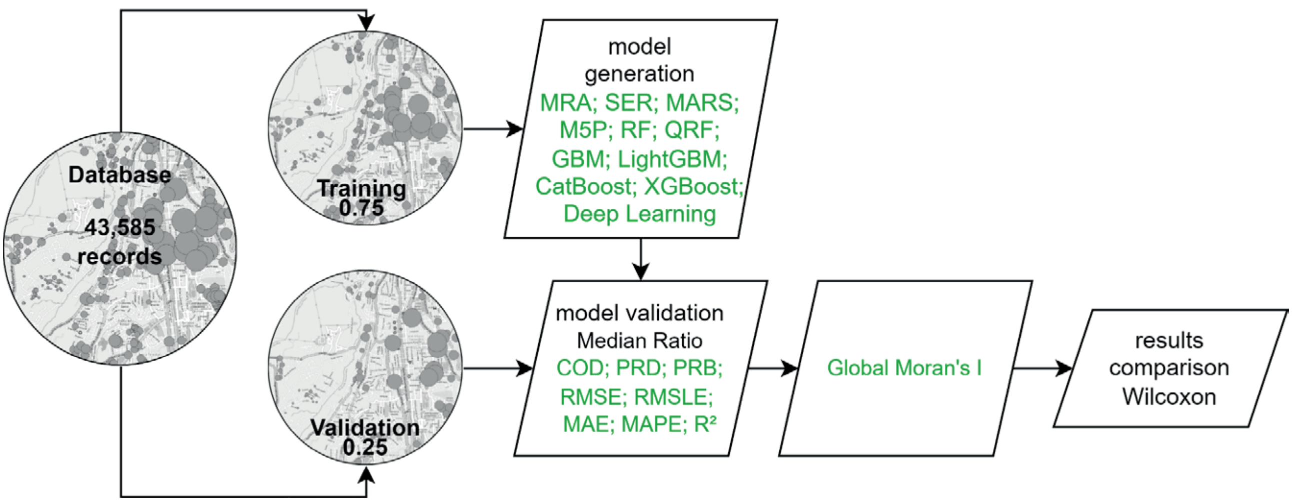

Figure 1. Graphical representation of the methodological process followed.

bility is based on the property's value, and municipalities manage real estate cadastral systems. Fortaleza is known for its georeferenced property base, an urban observatory of property values, and machine learning techniques in real estate taxation (Eguino and Erba, 2024).

The apartment buildings considered for this analysis vary widely in location and characteristics. The sample includes apartments with an average private area of 96.59m², ranging from 30m² to 883m², ranging from 1 to 7 bedrooms, and with prices per square meter ranging from US$ 209/m² to US$ 3,502/m². It includes high-end luxury apartments along the northern coast, near the hotel area, with modern amenities, security, and robust infrastructure, including shopping centers, businesses, and services. Conversely, condominiums are sparse in the southern and southwestern regions of the city, as they need more infrastructure and are generally in poorer condition. The dataset encompasses a wide range of building ages, standards, and features, including minimum lot areas and well-suited areas for leisure and swimming pools, as well as the presence of elevators. In lower-standard condominiums, the presence of these amenities is uncommon. This variability in the data ensures a more comprehensive analysis of the factors affecting apartment values in different areas of the city.

Figure 2 shows the spatial distribution of the initial sample and categorized price per square meter. As can be seen, the most expensive apartments are in the north of the city, close to the seafront, with good urban infrastructure, hotels, and public facilities. Despite an extensive coastline, prices are low in the city's eastern part. This occurs due to the need for more infrastructure, urban equipment, and high air salinity. In the geographic center of the city, there is the airport, which prevents the presence of tall buildings in its surroundings. The poorest region is in the town's southwest, with few tall buildings.

3.2. Explanatory variables

The choice of explanatory variables for a mass real estate valuation model depends significantly on the

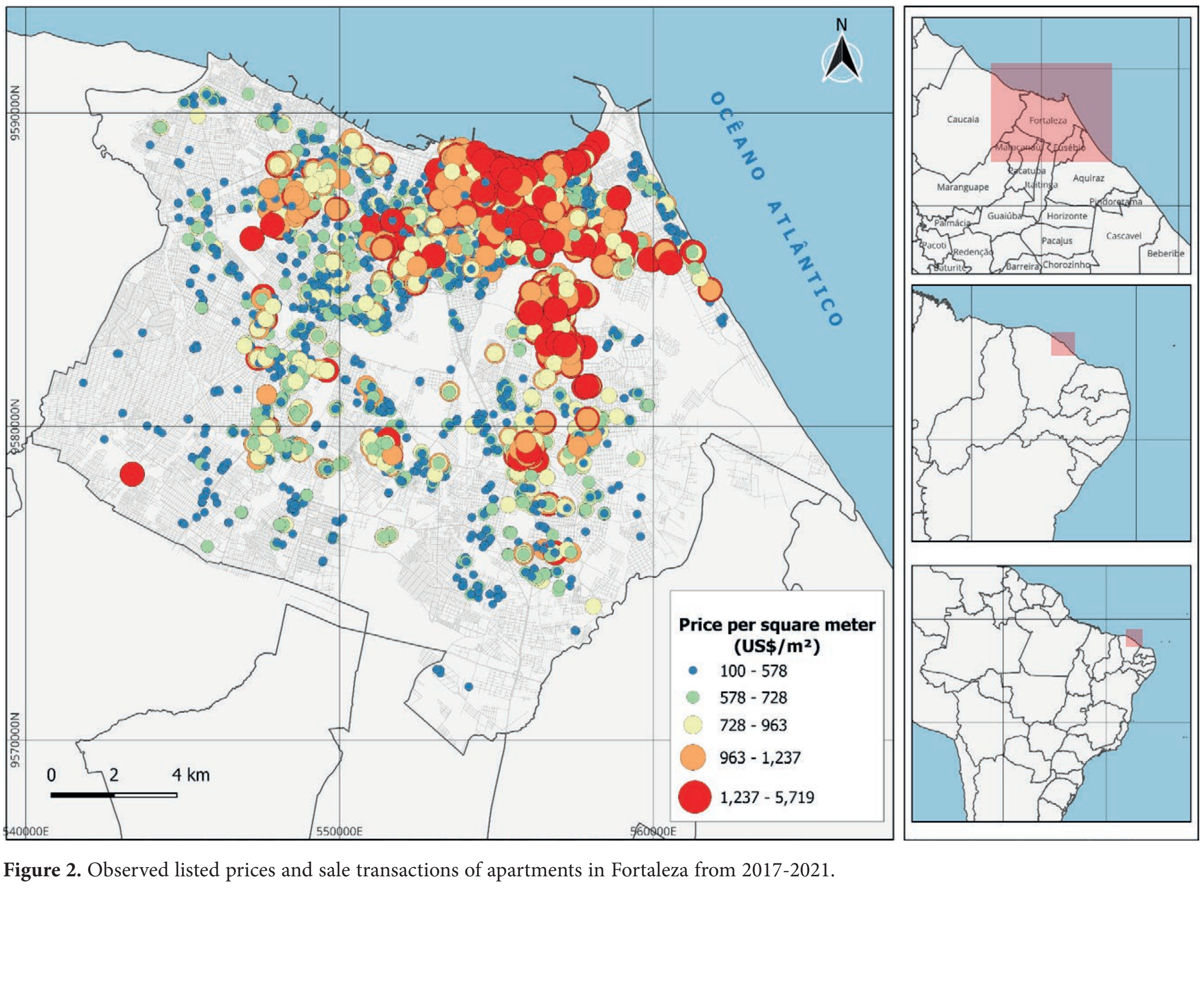

Figure 2. Observed listed prices and sale transactions of apartments in Fortaleza from 2017-2021.

researcher's prior knowledge of the market in that region under study. In addition, geographic information systems (GIS) are of fundamental importance in creating new "spatial proxies" variables that can capture the spatial dependence observed in real estate market datasets. For example, in the northern part of the city, in the Meireles neighborhood, there is an important tourist hub with a promenade for walking that is only 3 km long, offering a beautiful view of the city's coastline and high urban density in its surroundings, with little availability of vacant lots. It is where the leading hotels and high-end luxury apartments are located. Using Euclidean distance, the explanatory distance variable to this promenade ("distbm") was calculated for each sample data point. It is expected that the greater this distance, the less valuable the apartment will be. On the other hand, the city is known for its extreme income concentration, reflected in the presence of informal settlements scattered throughout its territory, with a greater prevalence in the southwest region. In some of these areas, the crime rate is higher, which can negatively impact the prices of nearby apartments. Similar to the "distbm" variable, the "distassp" variable was established to capture the negative effect on the observed prices. Using GIS techniques, a kernel density estimation raster was calculated to capture the concentration of apartment buildings in the geographic space. The aim was to establish zones of higher demand, which could increase the prices of available properties ("vertdens" variable). Other distance variables to value hubs were used: "distpv" and "distsh". The first refers to the Euclidean distance from the apartment to the nearest hub chosen from the following categories: squares, schools, universities, recreational parks, and public cultural facilities. The second refers to the Euclidean distance to the nearest shopping mall, given its significant relevance in providing products and services in the surrounding area and its positive impact on apartment prices. Other variables were chosen based on the physical attributes of the apartment or its condominium, such as the number of bedrooms, floor position, private area, building age, condominium lot area, presence of an elevator and pool, and indication of being on the top floor. The number of bathrooms did not exist in the municipal property cadastre, so it was not used. The Dependent variable was the natural logarithm from price. Table 2 shows more details about the predictor variables used in the research.

3.3. Techniques applied in this study

The authors utilize ten estimation methods to compare models with Multiple Regression Analysis (MRA) results, including Spatial Error Regression (SER), decision trees (Multiple Adaptive Regression Splines and M5 pruned), ensemble decision trees (Random Forest, Quantile Random Forest, Gradient Boosting Machine, XGBoost, CATBoost, and LightGBM) and Deep Learning. We present some basic concepts about them.

3.3.1. Multiple analysis regression and spatial regression

The multiple regression analysis (MRA), which estimates parameters using the method of ordinary least squares, was chosen as the baseline model. As this method violates some of its assumptions, especially regarding homoscedasticity and the presence of spatial autocorrelation in the residuals, it also used a Spatial Error model (SER). The matrix notation for the SER model is given by Eq. 1:

$$y = X\beta + u \tag{1}$$

Where **u** is an error vector that follows an autoregressive process, λ is the spatial autoregressive parameter, **W** is the spatial weight matrix, and ε is a vector of errors (Anselin, 1988):

$$u = \lambda W u + \varepsilon \tag{2}$$

The spatial weight matrix was a first-order, standardized contiguity queen matrix (**W**). The estimate for λ considered the heteroscedasticity with estimation by the Generalized Method of Moments (GMM) based on Arraiz et al. (2010).

Multiple regression analysis and spatial error models assume a linear relationship between variables, which does not always occur in hedonic price real estate markets. Additionally, the effectiveness of the SER model is affected by the choice of the spatial weight matrix, being computationally intensive for large datasets, and by the lack of a transparent methodology for out-of-sample prediction.

3.3.2. ML techniques

There are two main functions that can be developed by an ML application: estimate values (predictive goal) and classify after some attribute using supervised and unsupervised approaches. In some cases, both options are used. Several machine learning approaches could be used in mass appraisal applications. To name some of them, applications used in real estate applications include decision trees, such as Multiple Adaptive Regres-

Table 2. Description and descriptive statistics of the predictor variables assessed in the models.

Variable	Description	Min	Median	Mean	Max
age	Age in years	0	5.74	10.60	66.76
distpv	Nearest distance to main urban amenities in m	1	272.47	363.46	2,795.92
distassp	Distance to the nearest precarious settlement	1	224.15	282.46	1,379.06
distsh	Distance to the nearest shopping center in m	1	1,136.34	1,367.98	6,231.17
distbm	Distance to the beach (north of the city) in m	1.79	4,144.86	5,185.59	16,442.75
income	The mean income of the head of the family in Brazilian minimum wage	0.63	5.74	7.13	29.96
landarea	The total area of the land parcel in m²	143	4,580.52	6,286.20	5,7718
maxind	Maximum allowed floor area ratio (FAR)	0.47	2.5	2.32	3
parcelarea	The proportional area of a land plot in m² related to the apartment	4.5	43.46	77.80	40,012.5
privarea	Private property size in m²	30	76.66	96.58	883.32
test	Length in m of the front of the land parcel.	5.5	61	73.60	379.32
totalarea	Total area in m²	32.16	126.97	152.56	2,278.95
vertdens	Verticalization density measure. Indicates the concentration of unities in the neighborhood	0.01	0.1587	0.24	0.9982
xcen	X coordinates in UTM of the normalized geographic position	-1.06	0.28	0.17	0.98
ycen	Y coordinate in UTM of the normalized geographic position	-0.98	0.4	0.25	0.91
bedroom	Number of bedrooms	1	3	3	7
floor	Apartment' floor	0	5	7	31
garage	Number of garages	0	2	2	15
nfloors	Total number of floors in the building	2	14	13	32
elev	Presence of one or more elevators - dummy	0	1	0 or 1	1
pool	Presence of swimming pool - dummy	0	1	0 or 1	1
lastfloor	Property is on top floor position - dummy	0	0	0 or 1	1
standardi	Finishing standard - set of dummies (1-rustic, 2-proletarian, 3-economical, 4-simple, 5-medium, 6-superior, 7-fine, 8-luxury).			1-8	
$year_i$	Real estate transaction year - set of dummies for 2017-2021			0 or 1	
sale	Effective sale – dummy			0 or 1	
listing	List price – dummy: 0 = sale's price; 1 = asking price			0 or 1	

Source: Authors.

sion Splines (MARS) and M5 pruned (M5P); Deep Learning; Ensemble Decision Trees, like Random Forest (RF); Quantile Random Forest (QRF); Gradient Boosting Machine (GBM) and its optimized implementations such as XGBoost, CATBoost, and LightGBM. All of them are essential alternatives.

In general, ML models provide better accuracy in their predictions than hedonic pricing models and their variants, as long as they are trained with a large volume of high-quality data that is representative of the population under study. They are capable of identifying complex non-linear patterns that traditional methods may not capture. They have high scalability for deployment in environments that require quick and accurate responses, such as those in governmental property taxation departments. On the other hand, such models are known as "black boxes" because they are difficult to interpret, making it impossible in some cases to directly extract the marginal contribution of each explanatory variable to the observed price under "ceteris paribus" conditions. Additionally, ML models can overfit the training data, losing the ability to generalize to new data (overfitting).

3.3.3. Decision trees

In direct words, a Decision Tree (DT) is a non-parametric supervised learning approach used in statistics and machine learning to develop classification and regression as a predictive model to conclude a set of observations by partitioning the feature space (if-else conditions). Tree models in which the target variable can assume a discrete set of values are called classification trees. In these structures, leaves are class labels, and branches represent connections of features that lead to those class labels. Decision trees where the target variable can be continuous (usually real numbers) are called regression trees. DT has the advantage of being more intuitive and easier to interpret than other machine learning models. Depending on the number of nodes and depth, each decision criterion chosen by the algorithm can be easily visualized graphically, facilitating the understanding of the entire decision-making process. However, they can easily overfit the training data, capturing noise and leading to poor performance on new data. Pruning and ensemble decision trees (e.g., Random Forest and XGBoost) are techniques necessary to mitigate this problem. The more robust algorithms described below were used in this research instead of this approach.

3.3.4. Multiple adaptive regression splines

Multiple Adaptive Regression Splines (MARS) is a decision tree that combines recursive partitioning with spline fitting to model the relationship between a set of input (predictive) variables and dependent variables (Friedman, 1991; Friedman and Roosen, 1995). Data is modeled by separate piecewise linear segments (splines) of differing slopes known as essential functions. MARS generates basis functions by searching stepwise, where an adaptive regression algorithm selects the locations of the knot (endpoints of the segment) (Reyes-Bueno et al., 2018). The regions between submarkets are continuous. Some recent studies using MARS include Al-Qawasmi (2022), Reyes-Bueno et al. (2018), and Wang and Li (2019). For the model used in MARS, the best combination of hyperparameters was obtained using the "RandomizedSearchCV" method with ten folds for cross-validation, with a maximum degree of interaction (degree) from 1 to 3, and a maximum number of terms in the pruned model ("nprune") from 1 to 100. The strategy to evaluate the performance was the lowest RMSE. The best results were obtained with a "nprune" of 21 and a degree of 1. Although MARS can capture non-linear relationships more flexibly than M5P, the resulting model still lacks the ability to capture complex interactions compared to ensemble techniques due to its reliance on local linear fits.

3.3.5. M5 pruned

M5 pruned (M5P) is a model tree algorithm developed to predict continuous variables by applying regression that can exploit the local linearity of the data. For this purpose, model trees generate subsets by choosing attributes to minimize the intra-subset variation in the class values down each branch and maximize the expected error reduction (Zurada et al., 2011). There are three significant steps when applying the M5P methodology: (1) tree construction, (2) tree pruning, and (3) tree smoothing (Reyes-Bueno et al., 2018). The regions between submarkets are often discrete. For the model used in this research, the best combination of hyperparameters was obtained using the "RandomizedSearchCV" method with ten folds for cross-validation, with a range of minimum number of instances from 0 to 1000. The strategy to evaluate the performance was the lowest RAE. The best results were 50 as the minimum number of cases per leaf. Inside subM5P generates segments and exploits the local linearity of the data within those segments, which can be a limitation in situations where the relationships between variables are not locally linear, as it does not adequately capture the true nature of the interactions between variables. Within each submarket, M5P exploits the local linearity of the data, which can be a limitation in situations where the relationships between variables are not locally linear, as it does not adequately capture the true nature of the interactions among variables.

3.3.6. Random Forests

Random forests (RF) or random decision forests are an ensemble learning method for classification, regression, and other tasks that operate by building several decision trees at training time. For classification tasks, the output of the RF is the class selected by most trees. For regression tasks, the mean or average prediction of the individual trees is returned. Random decision forests correct for decision trees' habit of overfitting to their training set. RF outperforms regular decision trees, but their accuracy is lower than gradient-boosted trees. Data characteristics, however, can affect their performance. Random Forests, like other machine learning and deep learning algorithms, are often named "black box" models because they generate reasonable predictions for a wide range of data while requiring little configuration. However, the analyst cannot understand the logic behind them.

RF is a combination of predictor trees such that each tree depends on the values of a random vector assessed independently and with the same distribution for each of these. It is a substantial modification of bagging that builds a vast collection of uncorrelated trees and averages them. The method combined the idea of applying the general technique of bootstrap aggregating (bagging) technique to develop trees and the random subset of attributes to build a collection of decision trees with controlled bias and variance. The selection of a random subset of features is an example of the random subspace method, which is a way of conducting stochastic discrimination (Breiman, 2001).

For the model used in this research, the best combination of hyperparameters was obtained by the "RandomizedSearchCV" method with 5-fold for cross-validation. The strategy to evaluate the performance was the lowest MAPE. The best results were 15 features selected when looking for the best split ("max_features") and 556 trees in the forest ("n_estimators").

3.3.7. Quantile random forests

A very known extension of RF is the Quantile Random Forest (Wang et al., 2022) which estimates not only the mean or average of the target variable in each leaf but also the entire distribution of it in terms of quantiles for all individual trees. This makes it possible to compute the empirical quantile estimates of the target distribution and even confidence intervals for the predictions. This can be useful in mass appraisal for tax purposes, where one seeks to minimize over-taxation risks. In addition to requiring more computational resources than RF, QRF needs larger datasets to accurately predict quantiles and capture the entire distribution of the variables of interest. We perform the QRF algorithm with the same hyperparameters as the RF model but take median predictions from each tree.

3.3.8. Gradient boosting machine

Gradient boosting machine (GBM) is a machine learning technique used in regression and classification models. It gives a prediction model in the form of weak prediction models, typically decision trees. The decision trees are weak learners but are trained together and sequentially, each trying to correct the error from its predecessor (gradient-boosted trees). It usually outperforms Random Forest. A gradient-boosted trees model is built in a stage-wise fashion as in other boosting methods, but it generalizes the different techniques by allowing optimization of an arbitrary differentiable loss function. Gradient boosting is typically used with decision trees (especially CART trees) of a fixed size as base learners (Friedman, 2001). XGBoost often provides higher accuracy and performance than RF and QRF due to its use of boosting, which iteratively adjusts the errors of previous models. Additionally, its implementations can leverage GPU usage, significantly improving execution speed. One of the significant challenges for researchers using this technique is the need to utilize and test numerous hyperparameters for optimal performance.

In the GBM model, the best combination of hyperparameters was obtained by the "RandomizedSearchCV" method with 10-fold for cross-validation. The strategy to evaluate the performance was the lowest RMSE. The best results were 4,984 trees, with a learning rate (shrinkage) of 0.05, a maximum depth of each tree ("interaction. depth") of 7, and a minimum number of observations in the terminal nodes ("n.minobsinnode") of 15.

Notable extensions (alternatives) of GBM include XGBoost, LightGBM, and CatBoost. Because of this, we prefer to use these algorithms instead of the traditional GBM. Again, hyperparameter tuning was performed by "RandomizedSearchCV" with 5-folds for cross-validation. Each of these algorithms had its own optimized set of hyperparameters. However, this time, a more extensive set of hyperparameter possibilities were tested, such as the maximum depth of the trees, the minimum number of samples required to split an internal node, minimum loss reduction required to make a further partition on a leaf node of the tree (gamma), etc. (Jabeur et al., 2021; Iban, 2022).

3.3.9. Deep Learning

Deep Learning (also known as structured deep Learning, hierarchical Learning, or deep machine learning) is a ramification of machine learning based on artificial neural networks that try to model high-level abstractions of data using a deep graph with multiple processing layers (whence "deep learning" composed of various linear and non-linear transformations). Deep learning algorithms transform the inputs using more layers than shallow learning algorithms. A processing unit, such as an artificial neuron, converts the signal at each layer, whose parameters are "learned" through training. Deep learning algorithms are applied to supervised, semi-supervised, and unsupervised learning tasks (Ganaie et al., 2022).

Using deep Learning for mass assessment is challenging because their results are not fully interpretable ("black-box models"). While machine learning models already have several libraries to allow the influence of each predictor variable on the target, this has yet to happen for deep learning models. Hyperparameter tuning must be done with caution so as not to cause overfitting. This research used a structure with three intermediate layers, one input layer, and another output layer. The Relu function was adopted as the activation function in the first four layers, while the last used a linear function.

3.4. Model performance

The evaluation of results was based on specific metrics, including root mean square errors (RMSE), mean absolute errors (MAE), mean absolute percentage errors (MAPE), Coefficient of Dispersion (COD), Price Related Bias (PRB), and Price-Related Differential (PRD). While RMSE, MAE, and MAPE are commonly used measures in machine learning applications, the International Association of Assessing Officers (IAAO) suggests the other three. According to the IAAO Glossary (IAAO, 2013):

- Coefficient of Dispersion (COD): the average deviation of a group of numbers from the median expressed as a percentage of the median. The standard is a COD of 15 or less.
- Coefficient of Price Related Bias (PRB): shows the percentage by which assessment ratios change whenever values are doubled or halved. For example, a PRB of -0.03 would mean that assessment levels fall by 3% when the value doubles. The PRB should range between -0.05 and +0.05. PRBs outside the range of -0.10 to +0.10 are considered not acceptable.
- Price-Related Differential (PRD): calculated by dividing the mean by the weighted mean. The statistic has a slight bias upward. Price-related differentials above 1.03 show the assessment of regressivity; price-related differentials below 0.98 show the assessment of progressivity.

Finally, we compute the difference between predictions for each algorithm pair and apply the Wilcoxon test to see if the observed differences are statistically significant.

3.5. Measurement of spatial autocorrelation in the residuals

Spatial autocorrelation refers to the degree of similarity between the values of a variable at geographically proximate points. Tobler (1979) referred to that as "The First Law of Geography": "Everything is related to everything else, but near things are more related than distant things". The spatial autocorrelation among the independent variables and in the dependent variable violates several basic assumptions of classical regression (Anselin, 1988; Griffith, 1996). Spatial autocorrelation can imply spatially correlated residuals, which violates the assumption of independence of errors. In mass appraisal, it is widespread for specific clusters to exhibit more significant variance in residuals, violating the principle of homoscedasticity. It can also occur that certain areas have more substantial and similar residuals, which may indicate a specification error where relevant independent variables are not present in the model. The presence of spatial autocorrelation in the residuals suggests that the models used may not have fully captured the spatial structure of the data. As a result, this could lead to inaccurate inferences and an underestimation or overestimation of the impact of explanatory variables. Except for the Spatial Error Model (SER), which used its own spatial weights matrix, all other models utilized the UTM coordinates of the condominium centroid of the apartment, as well as other previously described spatial proxy variables.

Thus, in addition to the performance metrics described above, we calculated the spatial dependency that might still be present in the residuals. The expectation for a good mass appraisal would be that the selected spatial proxy variables capture all that dependency. Thus, we calculated Moran's I statistic for each set of residuals' algorithms under the null hypothesis of no spatial autocorrelation:

$$I = n\frac{\sum_{i=1}^{n} \quad \sum_{j=1}^{n} \quad w_{ij}(\varepsilon_i - \underline{\varepsilon})(\varepsilon_j - \underline{\varepsilon})}{\sum_{i=1}^{n} \quad (\varepsilon_i - \underline{\varepsilon})^2} \tag{3}$$

Where n stands for the number of observations, ε_i are the residuals of the algorithms, and w_{ij} are the elements of the spatial neighborhood matrix between observations i and j. Using the Moran scatterplot, verifying the level of spatial dependence between the residuals was also possible.

4. RESULTS AND DISCUSSION

Initially, we estimated an MRA model with all the variables in Table 1, but we kept only those statistically significant at 5%. The results from MRA and SER are in Appendix A (Table A1). Both had the natural logarithm of price as the dependent variable, resulting in a similar R-squared. However, MRA indicated the presence of heteroskedasticity in the Breusch-Pagan test. In the SER model, we observed that the spatial autoregressive parameter (λ) was statistically significant at 1%, even though the chosen spatial weights matrix could not eliminate the spatial autocorrelation among residuals (Figure 3).

The Table A2 in the study presents the results of the MARS model for property evaluation in Fortaleza, Brazil, using the natural logarithm of price as the dependent variable. Among the key findings, it is observed that being in the districtPAPICU decreases the logarithm of the price by 0.1488743 units, suggesting a significant

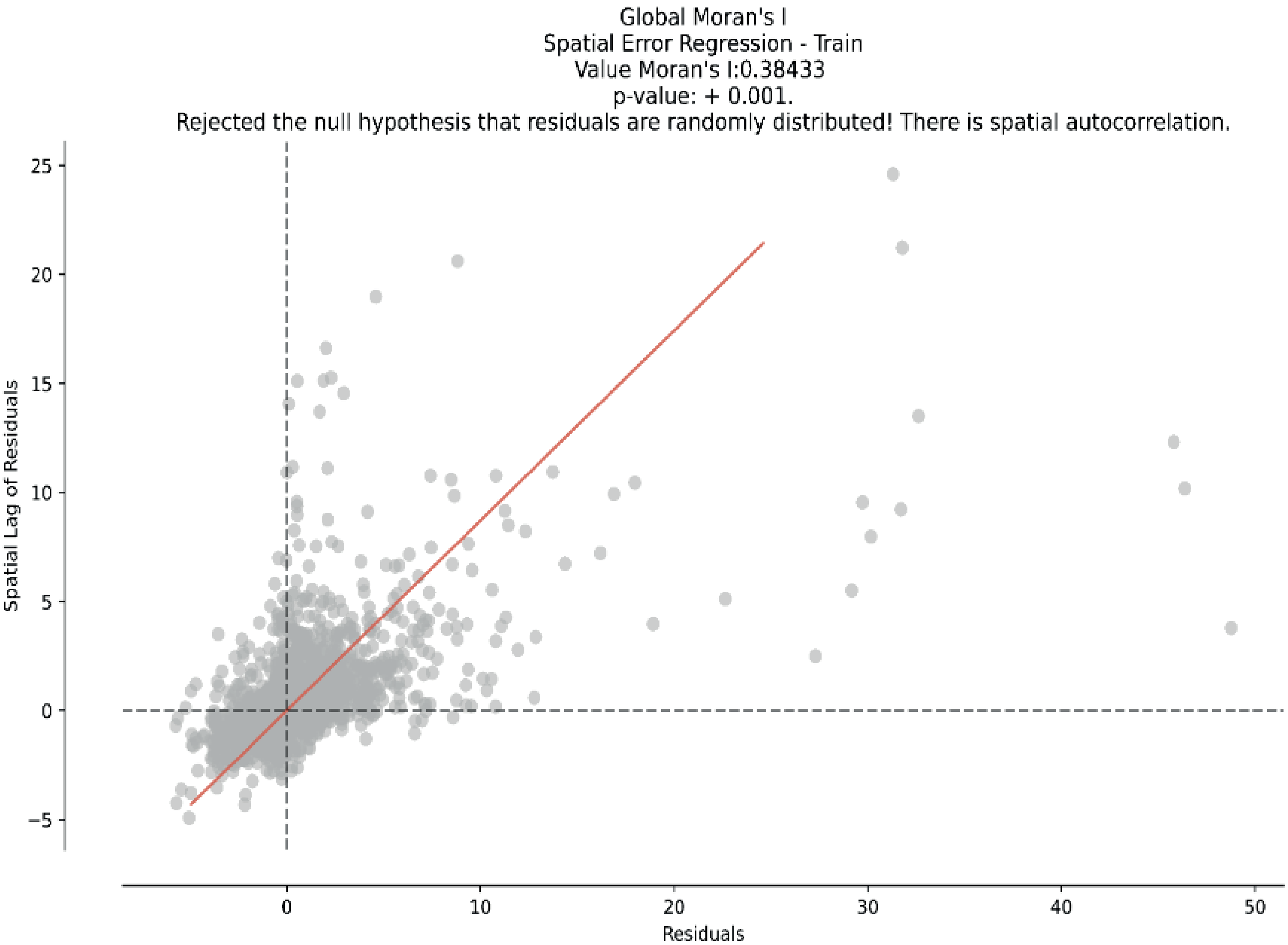

Figure 3. Spatial dependency analysis – model Spatial Error Regression (SER).

negative impact on property values in that area. The presence of a pool increases the logarithm of the price by 0.1010973 units, indicating considerable added value. Transformed variables such as h(14.68-income) and h(income-14.68) capture nonlinear effects of income on price, showing that an income less than 14.68 decreases the price by 0.0090807, while a higher income increases it by 0.0087215. Similarly, property age is reflected in h(27.56-age) and h(age-27.56), where an age less than 27.56 years increases the price by 0.0175039 units, and an older age decreases it by 0.0016117 units.

Additionally, temporal variables h(2020-year) and h(year-2020) show that transactions before 2020 decrease the price by 0.0081761 units, while transactions after 2020 increase it by 0.0691490 units. Normalized geographic coordinates also play an important role, where h(0.38-ycen) and h(ycen-0.38) indicate that a location with a Y coordinate less than 0.38 decreases the price by 0.1600824 units, and a greater Y coordinate decreases it by 0.4244964 units. These results provide a detailed view of the factors influencing apartment prices, allowing appraisers and market analysts to better understand the dynamics of the real estate market in the studied region.

Appendix B (Table B) shows the selected performance metrics for comparing the various models. The median ratio evaluates the overall appraisal level, considering the ratio between observed and predicted values. Values less than one indicate that the predicted values are lower than the observed ones. On the other hand, values greater than one indicate that the predicted values are more significant than the observed ones. For this metric, the results were similar for all models. As for the coefficient of dispersion (COD), the Random Forest and XGBoost models showed the best results (8,98%), indicating better uniformity in their predictions. The MARS, MRA, Deep Learning, and SER models had the worst results on this measure. The price-related differential (PRD) is a measure of the vertical iniquity. It

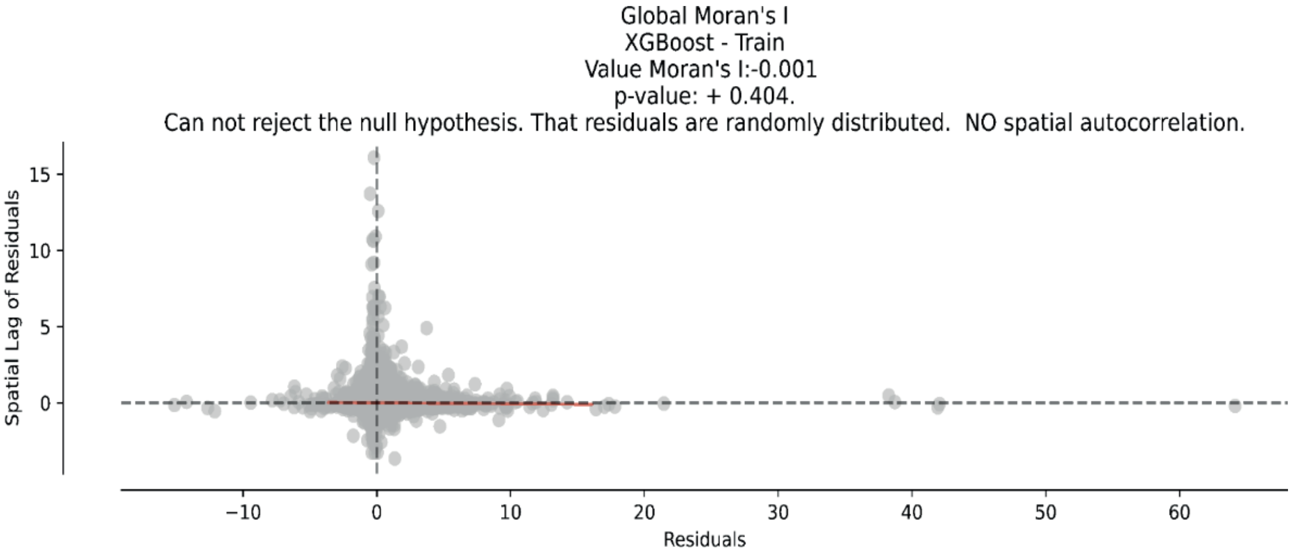

Figure 4. Spatial dependency analysis – model XGBoost.

measures whether high and low-value properties have the same appraisal. IAAO (2013) suggests that its value should be between 0.98 and 1.13. If it is outside this range, the model is regressive. That is, properties with higher observed prices had their predicted values lower than properties with lower prices. The presence of regressivity in mass valuation models is the worst possible situation for property taxation. Regressivity occurred in MRA, MARS, and Deep Learning. Another measure of vertical uniformity is the coefficient of price-related bias (PRB). By this measure, all models showed regressivity, although they are between -0.05 and +0.05, as recommended by the IAAO.

As previously mentioned, metrics like RMSE, RMSLE, MAE, MAPE, and R-squared are commonly measured in machine learning applications to assess the accuracy of machine learning models. They are also in Table B and complete the first part of the comparative study. GBM resulted in the best RMSE, followed by LightGBM and XGBoost, while MRA, MARS, and M5P were the worst. Root Mean Squared Log Error (RMSLE) is a better metric than RMSE in the presence of outliers. M5P, MARS, and MRA still show the worst results, while GBM, CATBoost, and XGBoost offer the best results. In contrast to the RMSE, mean absolute error (MAE), individual differences between observed and predicted prices are weighted equally in the average. Therefore, MAE is also less sensitive to outliers. XGBoost, RF, and CatBoost were the best, while MRA, MARS, and M5P were the worst. XGBoost, RF, and QRF had the best metrics in Mean absolute percentage error (MAPE), while MARS, MRA, and M5P were the worst. RF, GBM, LightGBM, and XGBoost had the same and the best R-squared (0.96), while MRA had a very discrepant value concerning the others, of only 0.88.

The last column in Table A1 has Global Moran's I statistic used to measure spatial autocorrelation among residuals. Moran's I statistic is interpreted against a reference distribution under the null hypothesis of complete spatial randomness. The spatial weights matrix used was queen contiguity order 1 for all models. The null hypothesis was not rejected only for the XGBoost model, meaning there is no spatial autocorrelation in the residuals (Figure 4).

The MARS and M5P models showed higher statistics with significance for spatial autocorrelation at 1%. It might suggest that the regression analysis did not consider another spatial proxy. The SER model also shows an inadequate specification of the chosen contiguity matrix. We understand that ensemble decision tree-based machine learning models are more capable of correcting spatial autocorrelation by taking advantage of the proxy variables used, randomly selecting subsets of the original training data through bootstrapped sampling and combining predictions from multiple decision trees. This became clearer with the XGBoost model, the only one to show no spatial autocorrelation in its residuals with p-value=0.404 (Figure 4). In a way, this would even be expected since gradient boosting machine algorithms are based on the iterative use of new models that try to minimize the errors of previous models, which would explain the low Moran's I index.

Finally, it is essential to know if statistical differences among performance metrics are presented in Appendix C (Table C). It shows the paired Wilcoxon test performed on the residuals of each model. It is observed that the MRA model, chosen as a baseline, had statistically similar residuals with the MARS, QRF, GBM, and Deep Learning models by the Paired Wilcoxon test. Notably, the XGBoost model was the only one to reject the null hypothesis that there is no significant difference between residuals at 5% of significance with all other models.

5. CONCLUSIONS

The study compared traditional and machine learning (simple and ensemble) techniques for mass property appraisal in Fortaleza. More than 43,000 data points for apartment values were evaluated using eleven techniques with varying levels of complexity. The initial set of independent variables selected was 36, including location and intrinsic property variables. The performance of the models was analyzed using the same subset of the data for validation.

Pointing to a possibility that outperformed the others could have been more evident. There is an equilibrium among the performance results of the different machine learning algorithms. However, the XGBoost model showed slightly better performance in several aspects, such as minimizing spatial autocorrelation and achieving statistically significant differences in residuals compared to other models. The MRA, M5P, or MARS techniques have good interpretability, although they have been shown to have a more spatial autocorrelation among residuals than the other methods. Their results highlight the importance of considering heteroscedasticity, spatial autocorrelation, regressivity, and better adjustments for property valuation. The spatial error model (SER) showed a high spatial correlation in the residuals, possibly due to the poor specification of the spatial weighting matrix; in addition, its use for mass evaluation needs to be improved by the need to predict out-of-sample data.

In Brazil, property tax is based on the property's market value, and municipalities manage it through real estate cadastral systems. Machine learning and other AVM (automated valuation models) are highly scalable in environments requiring quick and accurate responses, such as government property taxation departments. When it comes to interpreting the model, traditional techniques like MRA (multiple regression analysis), SER (structural equation modeling), M5P, or MARS (multivariate adaptive regression splines) are more understandable and transparent for people and for defending decisions in a tribunal. However, these techniques have been shown to have more spatial autocorrelation among residuals than machine learning models and result in regressive predictions. While machine learning models offer better performance and less spatial autocorrelation, the transparency and interpretability of classical techniques make them valuable for practical applications in property taxation, where clear communication and defense of valuation decisions are crucial.

Lastly, it is essential to underscore the significance of mass property assessment for land policies, particularly those related to urban spaces that profoundly impact the real estate market. Specific urban policy instruments can help mitigate the detrimental effects of real estate speculation, such as gentrification and housing deficits. Therefore, advancing research on mass property evaluation is essential to equip municipal administrators with efficient mechanisms to aid in land administration.

In future work, the authors plan to employ Explainable Artificial Intelligence techniques that aim to simplify the complexity of property taxation for end users. These new models will capture asymmetry in observed prices and other unobservable and non-linear effects between attributes. In essence, this study indicates the need to delve deeper into the search for more interpretable, less regressive models that are scalable for implementation in property taxation in large cities, where the heterogeneity of properties is more pronounced, with large clusters showing spatial autocorrelation. These aspects will help improve mass property valuation estimates and enhance taxpayer acceptance of the taxation itself.

REFERENCES

Al-Gawasmi, J. (2022). Machine Learning Applications in Real Estate: Critical Review of Recent Development. In Maglogiannis, I., Iliadis, L., Macintyre, J., & Cortez, P. (Eds). Artificial Intelligence Applications and Innovations. AIAI 2022. IFIP Advances in Information and Communication Technology, vol 647. Cham, Springer. https://link.springer.com/chapter/10.1007/978-3-031-08337-2_20.

Alfaro-Navarro, J. L., Cano, E. L., Alfaro-Cortés, E., García, N., Gámez, M., & Larraz, B. (2020). A fully automated adjustment of ensemble methods in machine learning for modeling complex real estate systems. *Complexity*, 2020, 5287263, 1–12. https://doi.org/10.1155/2020/5287263.

Anselin, L. (1988). *Spatial econometrics: methods and models*. Dordrecht, Springer Science Business Media. https://doi.org/10.1007/978-94-015-7799-1.

Antipov, E. A., & Pokryshevskaya, E. B. (2012). Mass appraisal of residential apartments: an application of random forest for valuation and a cart-based approach for model diagnostics. *Expert Systems with Applications*, 39(2), 1772–1778. https://doi.org/10.1016/j.eswa.2011.08.077.

Arraiz I., Drukker D., Kelejian H., & Prucha I. (2010). A spatial Cliff-Ord-type model with heteroskedastic innovations: small and large sample results. *Journal of Regional Science*, 50(2), 592–614. https://doi.org/10.1111/J.1467-9787.2009.00618.X.

Baur, K., Rosenfelder, M., & Lutz, B. (2023). Automated real estate valuation with machine learning models using property descriptions. *Expert Systems with Applications*, 213(C), 119147. https://doi.org/10.1016/j.eswa.2022.119147.

Belmiro, C., Neto, R. D. M. S., Barros, A., & Ospina, R. (2023). Understanding the land use intensity of residential buildings in Brazil: An ensemble machine learning approach. *Habitat International*, 139, 102896. https://doi.org/10.1016/j.habitatint.2023.102896.

Bilgilioğlu, S. S., & Yılmaz, H. M. (2023). Comparison of different machine learning models for mass appraisal of real estate. *Survey Review*, 55(388), 32–43. https://doi.org/10.1080/00396265.2021.1996799

Brazil (1988). Constitution of the Federative Republic of Brazil. Available at: https://legis.senado.leg.br/norma/579494/publicacao/33296461 (accessed 26 June 2024).

Brazilian Institute of Geography and Statistics - IBGE (2022). *Panorama Censo 2022*. Available at: https://censo2022.ibge.gov.br/panorama/ (accessed 26 June 2024).

Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1), 5–32. https://doi.org/10.1023/A:1010933404324.

Can, A. (1992). Specification and estimation of hedonic housing price models. *Regional Science and Urban Economics*, 22(3), 453–474. https://doi.org/10.1016/0166-0462(92)90039-4.

Chen, Y., Jiao, J., & Farahi, A. (2023). Disparities in affecting factors of housing price: A machine learning approach to the effects of housing status, public transit, and density factors on single-family housing price. *Cities*, 140, 104432. https://doi.org/10.1016/j.cities.2023.104432

De Cesare, C. M., Silva, E., & Silva, L. R. (2023). Avaliação de Imóveis. In Silva, E. (Ed.). *Cadastro territorial multifinalitário aplicado à gestão municipal*, 1ª ed., Cap. 5, pp. 101–126. Florianópolis, Universidade Federal de Santa Catarina. Available at: https://geografia.blog.br/pdf/2023ctmagmufsc.pdf#page=99 (accessed 26 June 2024).

Doan, Q. C., Chen, C., He, S., & Zhang, X. (2024). How urban air quality affects land values: Exploring non-linear and threshold mechanism using explainable artificial intelligence. *Journal of Cleaner Production*, 434, 140340. https://doi.org/10.1016/j.jclepro.2023.140340

Dubin, R.A. (1992). Spatial autocorrelation and neighborhood quality. *Regional Science and Urban Economics*, 22(3): 433–452.

Eguino, H., & Erba, D. (2024). *Mercado inmobiliario e impuesto predial: aplicaciones de técnicas de valuación masiva.* Washington, DC, Banco Interamericano de Desarrollo. http://dx.doi.org/10.18235/0005488

Friedman, J. H. (1991). Multivariate adaptive regression splines. *The Annals of Statistics*, 19(1), 1–67. https://www.jstor.org/stable/2241837.

Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *The Annals of Statistics*, 29(5), 1189–1232. https://www.jstor.org/stable/2699986..

Friedman, J. H., & Roosen, C. B. (1995). An introduction to multivariate adaptive regression splines. *Statistical Methods in Medical Research*, 4(3), 197–217. https://doi.org/10.1177/096228029500400303.

Ganaie, M. A., Hu, M., Malik, A. K., Tanveer, M., & Suganthan, P. N. (2022). Ensemble deep learning: a review. *Engineering Applications of Artificial Intelligence*, 115, 105151. https://doi.org/10.1016/j.engappai.2022.105151

Geerts, M., & De Weerdt, J. (2023). A survey of methods and input data types for house price prediction. *ISPRS International Journal of Geo-Information*, 12(5), 200. https://doi.org/10.3390/ijgi12050200

Griffith, D. A. (1996). Spatial Autocorrelation and Eigenfunctions of the Geographic Weights Matrix Accompanying Georeferenced Data. *The Canadian Geographer / Le Géographe Canadien*, 40(4), 351–367. https://doi.org/10.1111/j.1541-0064.1996.tb00445.x

Gunes, T. (2023). Model agnostic interpretable machine learning for residential property valuation. *Survey Review*, 1–16. https://doi.org/10.1080/00396265.2023.2293366

Hengl, T., Heuvelink, G. B., & Rossiter, D. G. (2007). About regression-kriging: from equations to case studies. *Computers Geosciences*, 33(10): 1301–1315. https://doi.org/10.1016/j.cageo.2007.05.001.

Heyman, A. V., Law, S., & Berghauser P. M. (2018). How is Location Measured in Housing Valuation? A Systematic Review of Accessibility Specifications in Hedonic Price Models. *Urban Science*, 3(1), 3. https://doi.org/10.3390/urbansci3010003.

Ho, W. K., Tang, B. S., & Wong, S. W. (2021). Predicting property prices with machine learning algorithms.

Journal of Property Research, 38(1), 48–70. https://doi.org/10.1080/09599916.2020.1832558.

Hong, J., Choi, H., & Kim, W. S. (2020). A house price valuation based on the random forest approach: the mass appraisal of residential property in South Korea. *International Journal of Strategic Property Management*, 24(3), 140–152. https://doi.org/10.3846/ijspm.2020.11544.

Hu, L., Chun, Y., & Griffith, D. A. (2022). Incorporating spatial autocorrelation into house sale price prediction using random forest model. *Transactions in GIS*, 26(5), 2123–2144. https://doi.org/10.1111/tgis.12931.

Hurley, A.K. & Sweeney, J. (2024). Irish property price estimation using a flexible geo-spatial smoothing approach: what is the impact of an address? *Journal of Real Estate Finance and Economics*, 68, 355–393. https://doi.org/10.1007/s11146-022-09888-y.

Iban, M. C. (2022) An explainable model for the mass appraisal of residences: the application of tree-based Machine Learning algorithms and interpretation of value determinants, *Habitat International*, 128, 102660. https://doi.org/10.1016/j.habitatint.2022.102660.

International Association of Assessing Officers (2013). IAAO Glossary for Property Appraisal and Assessment, 2nd ed. Available at: https://www.iaao.org/media/Glossary_Ed2_Web/IAAO_GLOSSARY_2015.pdf (accessed 26 June 2024).

Jabeur, S. B., Gharib, C., Mefteh-Wali, S., & Arfi, B. W. (2021). CatBoost model and artificial intelligence techniques for corporate failure prediction. *Technological Forecasting and Social Change*, 166, 120658. https://doi.org/10.1016/j.techfore.2021.120658.

Jayantha, W. M., & Oladinrin, O. T. (2020). Bibliometric analysis of hedonic price model using Citespace. *International Journal of Housing Markets and Analysis*, 13(2), 357–371. https://doi.org/10.1108/ijhma-04-2019-0044.

Kayakuş, M., Terzioğlu, M., & Yetiz, F. (2022). Forecasting housing prices in Turkey by machine learning methods. *Aestimum*, 80, 33–44. https://doi.org/10.36253/aestim-12320.

Oliveira, A. A. F. (2020). Avaliação em massa com modelos de aprendizado de máquina aplicados aos terrenos urbanos do Município de Fortaleza (Mass evaluation with machine learning models applied to urban land in the Municipality of Fortaleza). Master's Thesis in Public Sector Economics. Federal University of Ceará, Fortaleza, Brazil. Available at: http://www.repositorio.ufc.br/handle/riufc/53263 (accessed 26 June 2024).

Park, B., & Bae, J. K. (2015). Using machine learning algorithms for housing price prediction: the case of Fairfax County, Virginia housing data. *Expert Systems with Applications*, 42(6), 2928–2934. https://doi.org/10.1016/j.eswa.2014.11.040.

Renigier-Biłozor, M., Źróbek, S., Walacik, M., Borst, R., Grover, R., & D'Amato, M. (2022). International acceptance of automated modern tools uses must-have for sustainable real estate market development. *Land Use Policy*, 113, 105876. https://doi.org/10.1016/j.landusepol.2021.105876.

Rey-Blanco, D., Zofío, J. L., & González-Arias, J. (2024). Improving hedonic housing price models by integrating optimal accessibility indices into regression and random forest analyses. *Expert Systems with Applications*, 235, 121059. https://doi.org/10.1016/j.eswa.2023.121059

Reyes-Bueno, F., García-Samaniego, J. M., & Sánchez-Rodríguez, A. (2018). Large-scale simultaneous market segment definition and mass appraisal using decision tree learning for fiscal purposes. *Land Use Policy*, 79, 116–122. https://doi.org/10.1016/j.landusepol.2018.08.012

Rico-Juan, J. R., & La Paz, P. (2021). Machine learning with explainability or spatial hedonics tools? An analysis of the asking prices in the housing market in Alicante, Spain. *Expert Systems with Applications*, 171, 114590. https://doi.org/10.1016/j.eswa.2021.114590.

Rosen, S. (1974). Hedonic prices and implicit markets: product differentiation in pure competition. *Journal of Political Economy*, 82(1), 34–55. https://www.jstor.org/stable/1830899.

Sheppard, S. (1999). Hedonic analysis of housing markets, Chapter 41, 1595–1635. In Cheshire, P., and Mills, E. S. (Eds). *Handbook of Regional and Urban Economics*, 3. https://doi.org/10.1016/S1574-0080(99)80010-8.

Tobler, W. R. (1979). Cellular geography. In Gale, S., & Olsson, G. (Eds.). *Philosophy in geography*. Dordrecht, D Reidel, pp. 519–536.

Waang, D., & Li, V. J. (2019). Mass appraisal models of real estate in the 21st century: a systematic literature review. *Sustainability*, 11(24), 7006. https://doi.org/10.3390/su11247006.

Wang, D., Wang, P., Wang, C., & Wang, P. (2022). Calibrating probabilistic predictions of quantile regression forests with conformal predictive systems. *Pattern Recognition Letters*, 156, 81–87. https://doi.org/10.1016/j.patrec.2022.02.003.

Yağmur, A., Kayakuş, M., & Terzioğlu, M. (2022). House price prediction modeling using machine learning techniques: a comparative study. *Aestimum*, 81, 39–51. https://doi.org/10.36253/aestim-13703.

Zurada, J., Levitan, A., & Guan, J. (2011). A comparison of regression and artificial intelligence methods in a mass appraisal context. *Journal of Real Estate Research*, 33(3), 349–388. https://doi.org/10.1080/10835547.2011.12091311.

APPENDIX A. RESULTS OF MRA, SER, AND MARS MODELS.

Table A1. Regression Results for MRA and SER models

Variable	Coefficients		Variable	Coefficients	
	MRA (OLS)	SER (GMM)		MRA (OLS)	SER (GMM)
intercept	8.8663 *	8.8529 *	bedroom	0.0084 (0.0015)	0.0163 *
parcelarea	-1.24E-05 *	-1.26E-05 *	garage	0.0317 *	0.0305 *
income	0.0120 *	0.0118 *	ln_privarea	0.3680 *	0.3670 *
vertdens	0.1830 *	0.1919 *	ln_totalarea	0.3021 *	0.3061 *
distpv	-8.40E-06 (0.0137)	-1.09E-05 (0.1556)	lastfloor	0.0689 *	0.0799 *
distassp	4.88E-05 *	3.93E-05 *	floor	0.0065 *	0.0065 *
distbm	-1.34E-05 *	-1.49E-05 *	age	-0.0097 *	-0.0098 *
elev	0.0372 *	0.0488 *	a2018	0.0095 *	0.0026 (0.3697)
pool	0.1085 *	0.0938 *	a2019	0.0075 (0.0335)	0.0003 (0.9242)
nfloors	0.0104 *	0.0103 *	a2020	0.0230 *	0.0187 *
standard3	0.0989 *	0.1833 *	a2021	0.0792 *	0.0719 *
standard4	0.3631 *	0.3979 *	sale	-0.0626 *	-0.0759 *
standard5	0.5058 *	0.5163 *	listing	0.0528 *	0.0407 *
standard6	0.6890 *	0.6667 *	Lambda (λ)	n/a	0.6389 *
standard7	0.8583 *	0.8574 *	R-squared	0.9386	0.9380
standard8	1.1789 *	1.1219 *	Observations	32,688	32,688

Source: Authors. Notes: The dependent variable is the natural logarithm from price. *Shows P-values < 0.01; otherwise, values in parentheses are the P-values.

Table A2. Regression Results for MARS model

Variable	Coefficients	Variable	Coefficients	Variable	Coefficients
Intercept	7.7533639	h(14.68-income)	-0.0090807	h(1-floor)	-0.0329661
districtPAPICU	-0.1488743	h(income-14.68)	0.0087215	h(floor-1)	0.0058125
pool	0.1010973	h(0.0908-vertdens)	-1.8229510	h(27.56-age)	0.0175039
of	0.1154500	h(vertdens-0.0908)	0.1308400	h(age-27.56)	-0.0016117
h(2020-year)	-0.0081761	h(distbm-125.04)	0.0032095	h(4.8820-ln_privarea)	-0.4324477
h(year-2020)	0.0691490	h(1708.96-distbm)	0.0033438	h(ln_privarea-4.8820)	0.6268375
h(0.38-ycen)	-0.1600824	h(distbm-1708.96)	-0.0032261	h(4.40769-ln_areaed)	-0.2502487
h(ycen-0.38)	-0.4244964	h(18-pvtp)	-0.0235340	h(ln_areaed-4.40769)	0.2851514
h(53.31-cotat)	-0.0051410	h(2-garage)	-0.0236326		
h(cotat-53.31)	-0.0000217	h(garage-2)	0.0693027		

Source: Authors.

APPENDIX B. PRELIMINARY RESULTS OF DEVELOPED MODELS IN THE TEST SET

Table B. Regression Results for MRA and SER models (p-values between parentheses)

#	Model	Median Ratio	COD (%)	PRD	PRD Status	PRB	PRB Status	RMSE	RMSLE	MAE	MAPE (%)	R^2	Global Moran's I* (P)
1	MRA	1.00	13.89	1.038	Regressivity	-0.019	Regressivity between +/- 5%	197,400.28	0.18	77,416.86	13,90	0.88	0.37474 (0.001)
2	SER	1.00	11.01	1.025	Normal	-0.011	Regressivity between +/- 5%	137,965.35	0.15	60,292.53	11.02	0.94	0.38433 (0.001)
3	MARS	1.00	14.19	1.045	Regressivity	-0.022	Regressivity between +/- 5%	177,807.97	0.19	75,590.82	14.20	0.90	0.32921 (0.001)
4	M5P	0.99	13.78	1.026	Normal	-0.007	Regressivity between +/- 5%	164,919.99	0.20	71,138.19	13.67	0.92	0.28474 (0.001)
5	RF	1.01	8.98	1.019	Normal	-0.011	Regressivity between +/- 5%	121,314.74	0.13	48,558.11	9.05	0.96	-0.0097 (0.010)
6	QRF	1.00	9.18	1.018	Normal	-0.009	Regressivity between +/- 5%	129,928.75	0.14	49,908.67	9.18	0.95	-0.01903 (0.001)
7	GBM	1.00	9.24	1.018	Normal	-0.011	Regressivity between +/- 5%	113,640.60	0.13	50,229.93	9.25	0.96	-0.01252 (0.001)
8	LightGBM	1.01	10.52	1.022	Normal	-0.016	Regressivity between +/- 5%	114,551.80	0.14	54,026.65	10.60	0.96	0.00919 (0.006)
9	CatBoost	1.01	9.15	1.020	Normal	-0.013	Regressivity between +/- 5%	127,548.82	0.13	48,832.36	9.21	0.95	-0.02712 (0.001)
10	XGBoost	1.00	8.98	1.020	Normal	-0.013	Regressivity between +/- 5%	116,633.77	0.13	48,542.72	8.96	0.96	-0.001 (0.404)
11	Deep Learning	1.01	12.09	1.051	Regressivity	-0.045	Regressivity between +/- 5%	138,462.94	0.17	61,889.17	12.25	0.94	0.15635 (0.001)

Source: Authors. Notes: *Global Moran's I was performed over the training sample; values in parentheses show P-values.

APPENDIX C – WILCOXON TEST

Table C. Wilcoxon signed-rank test on residuals (p-values in parentheses).

	MRA	SER	MARS	M5P	RF	QRF	GBM	LightGBM	CatBoost	XGBoost	Deep Learning
MRA		28654871 (0.002)	29287516.5 (0.222)	10764694 (0.000)	28273101 (0.000)	29387804 (0.359)	29621056 (0.836)	28776431.5 (0.005)	28924854 (0.020)	28080357 (0.000)	29606772.5 (0.802)
SER			29152309 (0.102)	8112396 (0.000)	26330696 (0.000)	28932765 (0.021)	29264854 (0.197)	26362893 (0.000)	26818224 (0.000)	28868465 (0.012)	28811277 (0.008)
MARS				12006409.5 (0.000)	29140611 (0.095)	29377956 (0.344)	29292579.5 (0.228)	29551878.5 (0.677)	29140611 (0.095)	28617181.5 (0.000)	29509893 (0.586)
M5P					4378069 (0.000)	6370087.5 (0.000)	6920394 (0.000)	5432615 (0.000)	4378069 (0.000)	5506569.5 (0.000)	8206475 (0.000)
RF						24311550 (0.000)	28713742 (0.003)	28938424 (0.022)	28360664 (0.0000)	19623564.5 (0.000)	27151928 (0.0000)
QRF							29681109.5 (0.981)	2779046 (0.000)	27826249 (0.000)	27070630.5 (0.000)	29682344 (0.984)
GBM								29051187.5 (0.052)	29067581 (0.056)	28791992 (0.006)	29447044.5 (0.4619)
LightGBM									29497222 (0.559)	24482096.5 (0.000)	26983269.5 (0.000)
CatBoost										1644228 (0.000)	5852011 (0.000)
XGBoost											28031575 (0.000)
Deep Learning											

Note: Bolded values represent significance with p-value < 0.01.