AESTIMUM

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

Firenze University Press
www.fupress.com/ceset

OPEN ACCESS

Citation: Zilli, C.A., Bastos, L.C., & da Silva, L.R. (2024). Machine learning models in mass appraisal for property tax purposes: a systematic mapping study. *Aestimum* 84: 31-52. doi: 10.36253/aestim-15792

Received: February 15, 2024

Accepted: April 2, 2024

Published: August 4, 2024

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:
CAZ: 0000-0002-4926-1186
LCB: 0000-0002-8822-7634
LRS: 0000-0002-3178-814X

Machine learning models in mass appraisal for property tax purposes: a systematic mapping study

Carlos Augusto Zilli[1,*], Lia Caetano Bastos[2], Liane Ramos da Silva[2]

[1] *Federal Institute of Santa Catarina (IFSC), Florianópolis, Brazil*
[2] *Federal University of Santa Catarina (UFSC), Florianópolis, Brazil*
E-mail: carloszilli@gmail.com; lia.c.bastos@ufsc.br; liane.ramos@ufsc.br
*Corresponding author

Abstract. The use of machine learning models in mass appraisal of properties for tax purposes has been extensively investigated, generating a growing volume of primary research. This study aims to provide an overview of the machine learning techniques used in this context and analyze their accuracy. We conducted a systematic mapping study to collect studies published in the last seven years that address machine learning methods in the mass appraisal of properties. The search protocols returned 332 studies, of which 22 were selected, highlighting the frequent use of Random Forest and Gradient Boosting models in the last three years. These models, especially Random Forest, have shown predictive superiority over traditional appraisal methods. The measurement of model performance varied among the studies, making it difficult to compare results. However, it was observed that the use of machine learning techniques improves accuracy in mass property appraisals. This article advances the field by summarizing the state of the art in the use of machine learning models for mass appraisal of properties for tax purposes, describing the main models applied, providing a map that classifies, compares, and evaluates the research, and suggesting a research agenda that identifies gaps and directs future studies.

Keywords: Mass appraisal, Machine learning, Property valuations, Appraisal for property tax, Systematic mapping study.
JEL codes: C53, D83, R32.

1. INTRODUCTION

Mass property appraisal has been gaining importance owing to the large share of the real estate market in economic measures, which has become one of the development indicators in several countries (Yilmazer et al., 2020).

These appraisals play a very useful role in determining the basis for calculating taxes within the jurisdiction of municipalities, e.g., Brazil's municipal property tax (IPTU). They are also widely used for calculating indemnities and implementing urban policy instruments.

According to the International Association of Assessing Officers (IAAO, 2013), mass appraisal is the process of assessing a group of properties as of a particular date using common data, standardized methods, and statistical tests.

This group of properties is, in most cases, composed of hundreds of data that need to be collected, processed, and modeled properly to reflect, with minimal error and distortion, the behavior of the real estate market in the target region. Given this large amount of data, the use of automated assessment methods is advisable.

In the literature, several primary studies have addressed the importance of using automated techniques to carry out mass appraisals for property tax, including some machine learning methods. Despite the increasing number of primary studies, none of the published systematic mapping studies has provided, to date, a comprehensive overview of the state of research in this field. A systematic mapping encompasses a broad review of primary studies in a given field, identifying what evidence is available (Kitchenham et al., 2010).

There is a need for systematic mapping studies showing which machine learning methods are used in mass appraisals for property tax purposes and how the accuracy of these methods is checked in the respective primary studies.

The present article seeks to fill this gap and provide an overview, through a systematic mapping study, of the main machine learning techniques that have been used in mass appraisals for property tax, as well as show how these techniques are measured regarding the accuracy of their results.

The main contributions of this article are:

- an overview of the state of the art on the use of machine learning models in mass appraisal for property tax purposes;
- a description of the main machine learning models that are commonly used in the field of mass appraisal for property tax purposes;
- a systematic map that classifies, compares and assesses existing research on the use of machine learning models in mass property appraisal;
- an agenda that consists in describing research gaps and suggestions for future studies with implications for practitioners and researchers; and
- an overview of core research topics and key findings on the use of machine learning models for mass property appraisal.

The remainder of this article is structured as follows: Section 2 presents background and related work; Section 3 describes the review protocol adopted to carry out this systematic mapping study; Section 4 shows the results of the present review; Section 5 discusses the main findings and research opportunities concerning the study; Section 6 describes threats to validity and, finally, Section 7 concludes this article with suggestions for future research.

2. BACKGROUND AND RELATED WORKS

This section provides background on mass property appraisal and use of machine learning models for such purpose, providing an overview of related works, including secondary studies.

2.1. Definition of mass property appraisal

Mass and individual appraisals of property differ only in scale, because they both seek to provide an accurate assessment of the value of one or more properties (McCluskey et al., 1997). According to the authors, mass appraisal arose from the need for standardized valuations when a high number of properties have to be valued.

Mass appraisals, therefore, consist in determining the values of all properties in a region or municipality, and they play an important role in property taxation. The correct estimation of values is essential to achieve equity (the same ratio for all properties between the appraisal value and the market value) and to enable fiscal justice to occur (Uberti et al., 2018).

In this way, mass property appraisal aims to systematically determine, on a large scale, the values of properties to maintain them proportional in view of their generic location and the specific characteristics of land and respective improvements, using statistical analysis or other techniques capable of accurately estimating the value of goods (Liporoni, 2014).

2.2. Machine learning models for property appraisal

In recent years, machine learning models have been used, with some degree of success, in mass appraisals for property tax. There are several machine learning models available; however, this study will only address tree-based regression models as they clearly present more accurate predictions when compared to other models (Valier et al., 2020).

Therefore, tree-based models frequently investigated in mass property appraisals will be discussed - from the simplest ones (decision trees) to their improved versions, which are popularly known as ensemble methods. The following decision tree-based machine learning models will be addressed: Bagging (Bootstrap Aggregation), Random Forest (Breiman, 2001), AdaBoost, CatBoost, Gradient Boosting (Friedman, 2001), XGBoost and LightBoost.

There are several metrics available for evaluating the performance of machine learning models. These metrics are essential for the design, adjustment, and evaluation of models, as they seek to compare the values found for the response variable with the values predicted by the models being applied. Such comparison is performed by simplifying the results to an understandable value. The major metrics are described below:

Root-mean-square error (RMSE): Root mean square error is calculated as the square root of the mean squared differences between observed and predicted values.

Mean Square Error (MSE): Mean square error is commonly used to check the accuracy of models. Each error is squared individually, and then these squared errors are averaged.

Mean Absolute Error (MAE): The mean absolute error measures the average of the error differences between the observed and predicted values by the models without considering their direction (Islam et al., 2022).

Mean Absolute Percentage Error (MAPE): The mean absolute percentage error is the average of all percentage absolute errors, regardless of whether the error is positive or negative, providing an indication of the average size of the error, expressed as a percentage of the observed value.

Determination Coefficient (R^2): measures the goodness of fit by estimating the variation of the response variable on the basis of explanatory variables. It is a measure of the proportion of variability in one variable that is explained by the variability of the other variables.

Coefficient of Dispersion (COD): represents the average deviation, expressed as a percentage, of the assessed value of each property from the median of the assessed value divided by the observed value. Thus, the COD quantifies the extent of uniformity in appraisals by analyzing the observed variability (IAAO, 2013).

Price-Related Differential (PRD): it is an indicator that measures the degree of vertical inequality, based on systematic differences in the valuation of low and high value properties; it is suitable for large samples (IAAO, 2013).

2.3. Secondary studies on the theme

Few secondary studies have systematically analyzed the literature on specific topics regarding the use of machine learning models in mass property appraisal. Three secondary studies were identified (see Table 1): two of them are systematic reviews of the literature while one is a critical review of the literature.

These reviews addressed several topics, e.g., the use of automated methods and their results (Wang and Li, 2019), prediction accuracy using machine learning models (Valier and Micelli, 2020) and optimal models for predicting the value of properties and price indices (Ja'afar et al., 2021).

The study of Wang and Li (2019) provided a systematic review of mass appraisal models used for property tax, including works published between the years 2000 to 2018. Three main trends were identified: AI-based model, GIS-based model and mixed models, and a total of 104 articles were analyzed. Multiple linear regression models, intelligent systems, artificial neural networks, tree-based models, hierarchical modeling, cluster analysis, fuzzy set theory and reasoning-based models were reviewed. The article does not exactly focus on machine learning models for mass property appraisal, but methods are sometimes cited. One of the limitations of the study of Wang and Li is the fact that it focuses only on the Web of Science electronic database; although it reflects the trend towards this topic, such database may not contain all articles addressing mass appraisal. The article ends by citing the concept of mass appraisal 2.0, a procedure for assessing, analyzing, and testing a group of properties as of a certain date. It combines artificial

Table 1. Secondary studies on mass appraisals with machine learning models.

Year	Authors	No. of studies	Title
2019	Wang and Li	104	Mass appraisal models of real estate in the 21st century: a systematic literature review.
2020	Valier and Micelli	165	Automated models for value prediction: a critical review of the debate.
2021	Ja'afar et al.	47	Machine learning for property price prediction and price valuation: a systematic literature review

intelligence, geoinformation systems, and mixed methods for optimal modeling of spatial and non-spatial data on property.

The study of Valier and Micelli (2020) sought to identify what evidence emerges from the literature on automated assessment models. The authors critically reviewed articles that empirically investigated the effectiveness of models for property value prediction. Their review included a total of 165 studies published up to July 2019. The article advances by reviewing automated machine learning models and addresses decision trees, random forest, artificial neural networks, genetic algorithms, k-nearest neighbors and support vector machine. The results showed a certain predominance of automated machine learning models over traditional econometric models with regard to the ability to predict the market value of property. From an operational point of view, the high performance achieved in forecasting property prices makes machine learning models attractive to all traders who value, manage or trade property assets.

Finally, the study by Ja'afar et al. (2021) analyzed the use of machine learning in property appraisal to identify the best model for predicting the values of properties based on characteristics such as location, land size, number of rooms and others. For such purpose, the authors reviewed 47 studies published in the Scopus and Web of Science databases between 2009 and 2021. The authors analyzed the following models: random forest, support vector machine, gradient boosting, decision trees, principal components analysis, artificial neural networks, and k-nearest neighbors, among others. The authors reported that supervised learning is the most popular model among the reviewed articles, and random forest is the model that best predicts property value. This algorithm can easily adapt to the specificities of property data and produce accurate and effective results.

The existing secondary studies focus on analyzing machine learning models that predict the value of property, but they do not mention mass appraisal; the only study that addresses this issue does not, in reality, fully explore the use of machine learning models, nor does it focus on appraisals for property tax. Furthermore, there is a lack of systematic mapping studies that provide a comprehensive overview of the current research landscape in this field or establish a framework for the accumulated knowledge regarding mass property appraisal through machine learning models.

For the reasons mentioned above, this article aims to fill the existing gap by providing an overview of previous research on these themes, through a systematic mapping study on the use of machine learning models in mass appraisal for property tax.

3. REVIEW PROTOCOL

While a systematic review of the literature is an important means of identifying, evaluating, interpreting, and comparing all available research relative to a specific research question, a systematic mapping study focuses on existing research rather than answering a detailed research question (Budgen et al., 2008; Petersen et al., 2008). The central objective of a systematic mapping study is to identify and classify existing evidence, without necessarily synthesizing new information (Kitchenham and Brereton, 2013; Petersen et al., 2015).

Therefore, the present study was conceived as a systematic mapping study because this type of research can deal with a wide range of areas and provide systematic procedures to identify, categorize, and analyze the existing literature (Budgen et al., 2008, Kitchenham et al., 2010; Petersen et al., 2008).

3.1. Research objective and question

Peer-reviewed journal articles will be analyzed to identify the machine learning techniques that are being used for mass appraisal for property tax and to check which of these mass appraisal techniques have provided the most accurate predictions.

For the systematic mapping study, the following general research question was formulated: *What is the state of the art of the literature regarding the use of machine learning models in mass appraisal for* property *tax?*

For the sake of clarity, this research question was broken down into other specific questions, namely:

> *RQ-01: Which machine learning models are most frequently used in research on mass appraisal for property tax?*

Today, there is still no consensus on the benefits of using machine learning techniques to perform mass appraisals for property tax, however. According to Valier and Micelli (2020), the debate around the topic confirms a greater prediction accuracy of machine learning models compared to traditional regression analysis. In this way, this research question seeks to consider all the machine learning models used in the analyzed articles and check which benchmark model is used to perform comparisons.

> *RQ-02: Which property typologies are most frequently addressed in studies on mass appraisal for property tax?*

Mass appraisals can be applied to rural properties (Uberti et al., 2018) and urban properties, including land, houses, apartments and business offices and stores (Velumani et al., 2022, Yilmazer and Kocaman, 2020, Fontoura et al., 2020, Zhang, 2015). In this sense, this research question intends to map the typologies frequently used in studies that involve the use of machine learning for mass property appraisal, as well as studies whose typology uses spatial components across the variables. This question also seeks to map the origin (area of study) of the data of each article.

RQ-03: How are machine learning techniques evaluated for accuracy in mass appraisal for property *tax?*

There are numerous statistical indicators capable of measuring the accuracy of a set of forecasts, and the choice of this indicator is not a marginal decision (Valier and Micelli, 2020). Therefore, this research question seeks to investigate the main performance indicators addressed in the articles. Additionally, based on the results reported in the respective studies, it intends to indicate the machine learning model that had the best performance among the calculated metrics.

RQ-04: What are the research trends and features of current studies on the application of machine learning to mass property appraisals?

A valuable tool for understanding the nature of a research area is the investigation of research trends and the systematic classification of existing studies (Petersen et al., 2008). In this sense, this research question intends to map the frequency of publications over time to identify research trends and seeks to categorize and aggregate existing studies to structure the target research area.

3.2. Execution of systematic mapping

This systematic mapping study consisted of three distinct steps: (*i*) search for articles, (*ii*) selection of articles, and (*iii*) data extraction, according to Petersen et al. (2008). These steps are described in further detail below.

3.2.1. Search for articles

To perform systematic mappings, many different electronic sources must be searched, because in general, a single data source is not expected to contain all relevant primary studies (Brereton et al., 2007). Therefore, an automated search was carried out in 4 different databases (DB) (see Table 2).

In the selection of electronic databases, the criteria encompassed: (i) the extensive research coverage across various disciplines provided by Web of Science and Scopus (Rodríguez et al., 2017), with the latter serving as a meta-library that compiles publications from numerous esteemed publishers, including Elsevier and Springer (Nakamura et al., 2022); (ii) the IEEE Xplore database, recognized as one of the foremost digital repositories in the field of computer engineering (Petersen et al., 2015), hosting a comprehensive collection of articles on machine learning; and (iii) the significance of Compendex as a vital interdisciplinary engineering database, cataloging a breadth of engineering journal titles and conference papers.

Table 2. Databases used in the present review.

Database	Search engine	Website
DB-01	Scopus	https://www-scopus.ez46.periodicos.capes.gov.br
DB-02	IEEE Xplore	http://ieeexplore-ieee-org.ez46.periodicos.capes.gov.br
DB-03	Web of Science	https://www-webofknowledge.ez46.periodicos.capes.gov.br
DB-04	Compendex	https://www-engineeringvillage-com.ez46.periodicos.capes.gov.br

In addition, the snowballing procedure was performed (Wohlin et al., 2012); the references of the five most cited selected articles were analyzed to identify relevant papers that were not returned during the automated search process.

Google Scholar was not selected as a database because the studies it returned tended to overlap with studies from the other databases included (Chen et al., 2010). However, the fact that the four chosen electronic databases index similar contents may reduce the possible threat to theoretical validity arising from failing to retrieve relevant studies.

Regarding type of document and time interval, the searches focused on peer-reviewed articles published in journals or in conference proceedings, from January 2015 to June 2022, when the present study was then developed.

Only publications written in English were selected, since it is the language mostly used in most international conferences and journals (Nakamura et al., 2022). It is also found that English is the predominant language in global communication; therefore, this systematic mapping study can be replicated by other researchers.

The search terms were defined using the five-step strategy proposed by Kitchenham et al. (2007). According to the author, one can develop the search terms by:

– Deriving key terms from the questions identifying population, intervention, and outcome;

- Identifying alternative spellings and synonyms for key search terms;
- Checking keywords in any relevant articles previously retrieved;
- Using the Boolean operator OR to incorporate alternative spellings and synonyms; and
- Using the Boolean AND operator to link key search terms.

Following this strategy, a generic search string was defined to connect key terms with Boolean operators, and several tests and refinements were carried out with it during the preliminary search. The following generic search string was used: *mass appraisal AND machine learning*.

Table 3 shows the set of search terms for the present study. As the search syntax is specific to each database, the search string was adapted to the specific syntax requirements of each of the four search engines.

The search string shown in Table 3 was tested several times with different combinations to reduce the number of articles that were not related to the research topic, thus ensuring a set of articles that were adequate to the objectives of this study. Table 4 shows the results of these searches.

To enhance the rigor of the automated search protocol, the investigation employed the snowballing method. This technique entails two complementary processes: backward snowballing, which involves tracing and analyzing the references cited in a primary article to uncover relevant studies, and forward snowballing, which consists of identifying subsequent publications that have cited the primary article. Such a strategy is instrumental in systematically broadening the scope of the research database. To mitigate any potential threats to the study's validity stemming from researcher bias, a secondary researcher independently conducted both the backward and forward snowballing operations. This approach yielded four new studies that were incorporated into the analysis database for the current mapping study.

Table 3. Overview of search terms and their synonyms.

Main Term	Search Terms
mass appraisal	("mass appraisal" **OR** "mass valuation" **OR** "mass assessment" **OR** "property appraisal" **OR** "property valuation" **OR** "real estate appraisal" **OR** "property tax" **OR** "land taxation" **OR** "taxes purposes")
	AND
machine learning	("machine learning" **OR** "data science" **OR** "data mining" **OR** "artificial intelligence" **OR** "ai" **OR** "computational intelligence" **OR** "automated valuation model" **OR** "avm")

Table 4. Search results in each of the search databases.

Scopus	IEEE Xplore	Web of Science	Compendex	Total
117	13	100	98	328

3.2.2. Selection of articles

Article selection criteria were defined to reduce the probability of bias and assess the relevance of the articles (Kitchenham and Charters, 2007). The article selection process returned a total of 328 + 4 = 332 publications. After this stage, screening was performed in two phases: (i) selection of relevant articles based on their metadata, namely title, abstract, keywords, year of publication, language of publication and publication type, and (ii) selection of relevant articles based on full text. The articles were selected by two researchers, working in a double-blind format using inclusion (IC) and exclusion (EC) criteria (see Table 5), as previously agreed between the researchers.

In this process, articles that met all the specified inclusion criteria were included and those that presented any exclusion criteria were discarded.

At first, all 330 retrieved articles were filtered using the EC-01, EC-02 and EC-03 exclusion criteria. Then, the remaining articles were uploaded to the software Rayyan (rayyan.ai) to detect duplicates by applying the EC-04 exclusion criterion. The remaining articles were then separated in the software Rayyan for an analysis of their metadata to identify the ones that were relevant for answering the research questions. During screening, the researchers read the title, abstract and keywords of the remaining articles and applied exclusion criteria EC-05, EC-06, EC-07 and EC-08. This analysis step was performed by the two researchers in a double-blind format. As selection procedures for the next step, the approach proposed by Petersen et al. (2015) was used; it is summarized in Graph 1.

To address potential disagreements, studies falling under conditions A, B, C, and D would be included in the research, while studies falling under the borderline condition E would undergo a joint analysis, and ultimately, studies under condition F would be definitively excluded. According to Petersen et al. (2015), studies categorized under condition D should be included since one of the researchers had no doubts regarding their inclusion in the systematic mapping, and therefore, they would need to be analyzed. Consequently, the studies falling under conditions A, B, C, and D were included in this system-

Table 5. Inclusion and exclusion criteria considered in the present review.

COD	Inclusion Criteria	
IC-01	Context	Articles that focused on machine learning methods and techniques for mass appraisal of urban property.
IC-02	Period	Articles published in 2015 and later.
IC-03	Location	Articles published in conference proceedings or in journals.
IC-04	Language	Articles published in English.
COD	**Exclusion Criteria**	
EC-01	Period	Articles published before 2015.
EC-02	Type	Items from the so-called gray literature (abstracts, books, panels, posters, editorials, short articles, reports, lectures, etc.).
EC-03	Language	Studies published in languages other than English.
EC-04	Duplicates	Articles that were duplicated, i.e., returned by more than one database.
EC-05	Reviews	Secondary studies (systematic reviews of the literature and mappings.
EC-06	Context	Articles whose abstract makes it clear that they are not related to property appraisal, even though they mentioned machine learning techniques.
EC-07	Typology	Articles whose abstract makes it clear that machine learning methods are applied for mass appraisal of rural, business or rental properties.
EC-08	Accuracy	Articles whose abstract makes it clear that they only address accurate appraisal (accurate property appraisal) and/or other studies related to appraisal engineering.
EC-09	Access:	Articles that are not available by open access, e.g., availability in the CAPES portal via the educational institution, or free availability on the Internet.
EC-10	Number of pages	Studies with five pages or less (short paper)
EC-11	Final criteria	Articles in which exclusion criteria could not be identified after reading of the title, keywords and abstract, and that were removed from the mapping after reading of the full texts, because they did not meet the inclusion criteria.

Divergence Analysis		**Reviewer X**		
		Include	Uncertain	Exclude
Reviewer Y	Include	A	B	D
	Uncertain	B	C	E
	Exclude	D	E	F

Graph 1. Analysis of divergences (adapted from Petersen et al., 2015).

atic mapping, and the studies falling under condition E underwent an assessment of uncertainties, and collectively, a decision was made regarding their definitive inclusion or exclusion from the systematic mapping.

All the articles approved in the previous stage were downloaded so that they could be read in full. They were downloaded directly from the database portals or through the CAPES/Brasil portal when they were not available by open access, and the EC-09 exclusion criterion was applied. After this step, all the downloaded articles were checked for number of pages according to the EC-10 exclusion criterion.

The articles returned in the previous step were then read in full for application of the EC-11 exclusion criterion. Full reading enabled the analysis of the articles in more detail than the previous reading of the title, abstract and keywords. The articles that did not meet the inclusion criteria were removed from this systematic mapping.

Table 6 shows the number of articles that were returned in each electronic database after applying each of the exclusion criteria shown in Table 5.

In the present review, the search string returned 328 articles (see Table 4): 117 from the Scopus meta-library, 13 from the IEEE Xplore database, 100 from the Web of Science database and 98 from the Compendex database. Figure 1 shows the number of articles that were excluded and that remained after application of each exclusion criteria (Table 5).

A total of 22 articles were selected for the data extraction stage: 18 by applying the exclusion criteria and 4 by applying the snowballing technique. Table 7 shows the selected articles and their authors.

The method of selecting articles using 2 researchers, ensures reliability in the decision to include or exclude a particular publication.

3.2.3. Data extraction

After completion of the article selection procedures, data extraction was started. The articles were categorized

Table 6. Number of articles returned after applying each exclusion criterion.

Database	EC-00	EC-01	EC-02	EC-03	EC-04	EC-05	EC-06	EC-07	EC-08	EC-09	EC-10	EC-11
DB-01	117	73	61	60	-	-	-	-	-	-	-	-
DB-02	13	8	8	8	-	-	-	-	-	-	-	-
DB-03	100	60	57	56	-	-	-	-	-	-	-	-
DB-04	98	41	36	36	-	-	-	-	-	-	-	-
Total	328	182	162	160	83	82	74	68	44	38	35	18

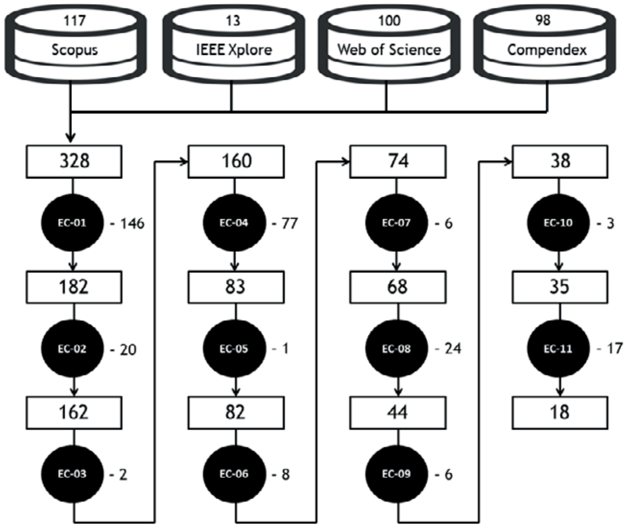

Figure 1. General structure of the answers from the data extraction form.

into different aspects (Petersen et al., 2008), since this approach is considered as a structured way to perform such a task. Therefore, the data extraction form brought together four different aspects, all related to the research questions listed in this review.

To reduce bias in the data extraction results, two researchers, based on a common understanding, performed the data extraction independently and, after the extraction process was completed, they discussed the results together and resolved conflicts to reach a consensus. Figure 2 shows the general structure of the DEF form (Data Extraction Form), which consists of four main research questions and four secondary research questions.

– **Models most frequently used in the selected articles (RQ-01)**

This research question was broken down into a major question *RQ-01: Which machine learning models were used in the study?*, whose answers could be *(a) decision trees; (b) bagging, (c) random forest; (d) adaboost; (e) gradient boosting; (f) XGBoost, (g) LightGBM* and *(h) other* and a secondary question *RQ-1.1: which benchmark model was used in the study?*, whose answers could be *(a) multiple linear regression; (b) spatial regression* and *(c) other*. When an analyzed article reported the use of more than one machine learning model (RQ-01 or RQ-1.1), a new row was added to the spreadsheet for each new model reported.

– **Most common typologies in the selected articles (RQ-02)**

This research question was also subdivided into a main question *RQ-02: What type of* property *was modeled in each study?*, whose answers could be *(a) urban land; (b) urban houses* and *(c) apartments*; and two secondary questions *RQ-2.1: What is the origin of the data set used in the study?*, whose answer would be the country of location of the data used in the study and *RQ-2.2: Did the study consider the spatial dimension when performing data modeling?*, whose answers could be *(a) did not consider it, (b) considered it as a predictor variable* and *(c) performed spatial modeling.*

– **Assessment of accuracy among selected articles (RQ-03)**

This research question was also broken down into a main question *RQ-03: Which indicator was used to evaluate the performance of the models?*, whose answers could be (a) RMSE; (b) COD; (c) PRD; (d) MSE; (e) MAE; (f) MAPE; (g) R^2 *or (h) other;* and a secondary question *RQ-3.1: Which model had the best performance in the analyzed study?*, whose answers could be *(a) decision trees; (b) bagging, (c) random forest; (d) adaboost; (e) gradient boosting; (f) XGBoost, (g) LightGBM* and *(h) other.* When the analyzed document reported more than one metric in RQ-03 or in RQ-3.1, a new row was added to the spreadsheet for each new information reported in the study.

– **Research trends and study characteristics (RQ-04)**

To assess research trends and study characteristics, data were collected regarding: *title, authors, source*

Table 7. Articles selected for the data extraction phase.

ID	Title	Authors	Year
S-01	The effect of google drive distance and duration in residential property in Sydney, Australia	Nejad, Mehrdad Ziaee; Lu, Jie; Asgari, Pooyan; Behbood, Vahid	2016
S-02	Applying dynamic Bayesian tree in property sales price estimation	Nejad, Mehrdad Ziaee; Lu, Jie; Behbood, Vahid	2017
S-03	Estimation and updating methods for hedonic valuation	Mayer, Michael; Bourassa, Steven; Hoesli, Martin; Scognamiglio, Donato	2018
S-04	An intelligent automatic valuation system for real estate based on machine learning	Niu, Jiafei; Niu, Peiqing	2019
S-05	Deep learning with XGBoost for real estate appraisal	Zhao, Yun; Chetty, Girija; Tran, Dat	2019
S-06	Sensitivity analysis of machine learning models for the mass appraisal of real estate: case study of residential units in Nicosia, Cyprus	Dimopoulos, Thomas; Bakas, Nikolaos	2019
S-07	A house price valuation based on the random forest approach: the mass appraisal of residential property in South Korea	Hong, Jengei; Choi, Heeyoul; Kim, Woo-Sung	2020
S-08	A mass appraisal assessment study using machine learning based on multiple regression and random forest	Yilmazer, Seckin; Kocaman; Sultan Kocaman	2020
S-09	Implementing a mass valuation application on interoperable land valuation data model designed as an extension of the national GDI	Aydinoglu, Arif Cagdas; Bovkir, Rabia; Colkesen, Ismail	2020
S-10	Using machine learning models and actual transaction data for predicting real estate prices	Pai, Ping-Feng; Wang, Wen-Chang	2020
S-11	Mass appraisal with a machine learning algorithm: random forest regression	Sevgen, Sibel Canaz; Aliefendioglu, Yesim	2020
S-12	Spatial prediction of housing prices in Beijing using machine learning algorithms	Yan, Ziyue; Zong, Lu	2020
S-13	A gradient boosting method for effective prediction of housing prices in complex real estate systems	Almaslukh, Bandar	2021
S-14	Developing automated valuation models for estimating property values: a comparison of global and locally weighted approaches	Doumpos, Michalis; Papastamos, Dimitrios; Andritsos, Dimitrios; Zopounidis, Constantin	2021
S-15	Predicting property prices with machine learning algorithms	Ho, Winky K.O.; Tang, Bo-Sin; Wong, Siu Wai	2021
S-16	Property mass valuation on small markets	Gnat, Sebastian	2021
S-17	A new appraisal model of second-hand housing prices in China's first-tier cities based on machine learning algorithm	Xu, Lulin; Li, Zhongwu	2021
S-18	Using machine learning to forecast residential property prices in overcoming the property overhang issue	Yee, Lim Wan; Bakar, Nur Azaliah Abu; Hassan, Noor Hafizah; Zainuddin, Norziha Megat Mohd; Yusoff, Rasimah Che Mohd; Rahim, Nor Zairah Ab	2021
S-19	Machine learning based predicting house prices using regression techniques	Manasa, J.; Gupta, Radha; Narahari, N.S.	2021
S-20	GIS & machine learning based mass appraisal of residential properties in England & Wales	Mete, Muhammed Oguzhan; Yomralioglu, Tahsin	2022
S-21	Mass appraisal as affordable public policy: open data and machine learning for mapping urban land values	Carranza, Juan Pablo; Piumetto, Mario Andres; Lucca, Carlos Maria; Silva, Everton da	2022
S-22	A comparative study of machine learning and spatial interpolation methods for predicting house prices	Kim, Jeonghyeon; Lee, Youngho; Lee, Myeong-Hun; Hong, Seong-Yun Hong	2022

(conference or journal), year of publication, author affiliation, authors' country of origin, number of study citations, abstract, keywords, name of conference or journal, place of conference, DOI code. Part of this information is extracted directly from the metadata of each article or directly from the publication's website. The number of citations for each article, up to June 2022, was collected directly on the platform semanticscholar.org.

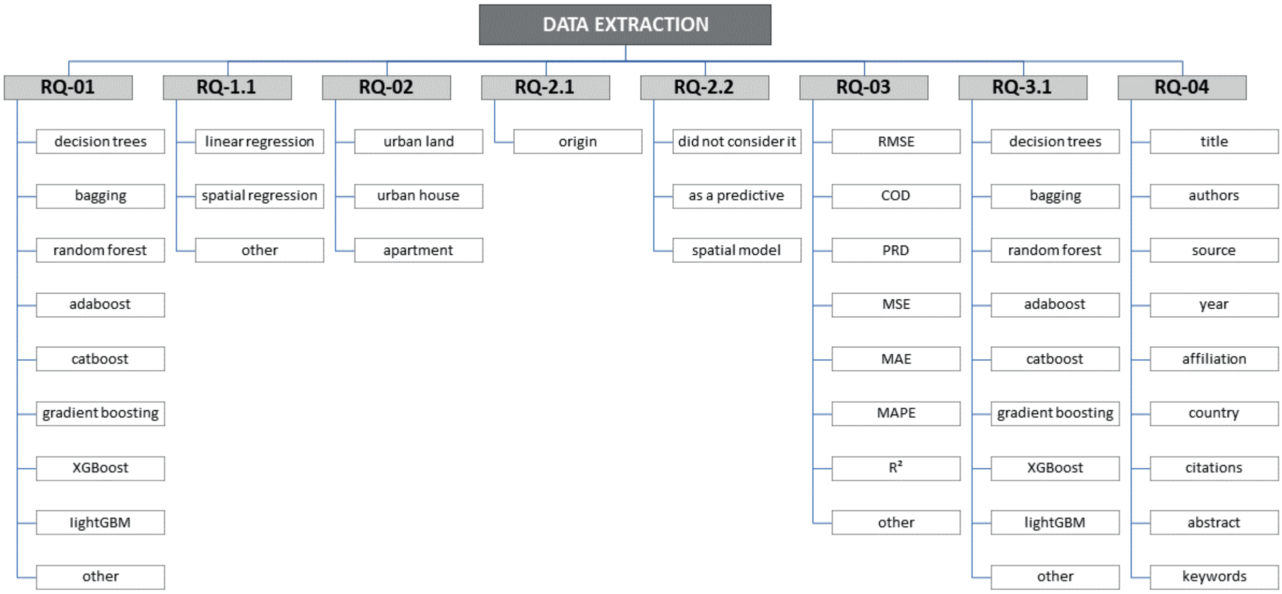

Figure 2. General structure of data extraction form (DEF) responses.

4. RESULTS

This section presents the answers to the formulated research questions (see Section 3.1). This section is organized according to the research questions.

4.1. Research Question 01 (RQ-01)

The first research question sought to identify which machine learning models, among those listed in the question itself (tree-based models), have been used in research that relates machine learning and mass property appraisal. The tree map (Figure 3) shows the absolute predominance of articles that tested the Random Forest model. There are exactly twice as many studies using this model when compared to the Gradient Boosting and XGBoost models. Among the machine learning models evaluated, AdaBoost was the least used among the analyzed studies.

The Random Forest modeI was used in 82% of the 22 selected studies; 41% performed analyses with the XGBost and Gradient Boosting methods; 18% tested the Decision Trees model; 14 % analyzed the LightGBM and Bagging models; 9% tested the AdaBoost algorithm, and only 5% of the studies tested the CatBoost algorithm.

As a sub-issue of research question 1, the benchmark model adopted by the studies was mapped to compare the results with those found by the machine learning models. Traditionally, multiple linear regression, in which the value of property is assumed to be dependent on the available characteristics, is used as a benchmark

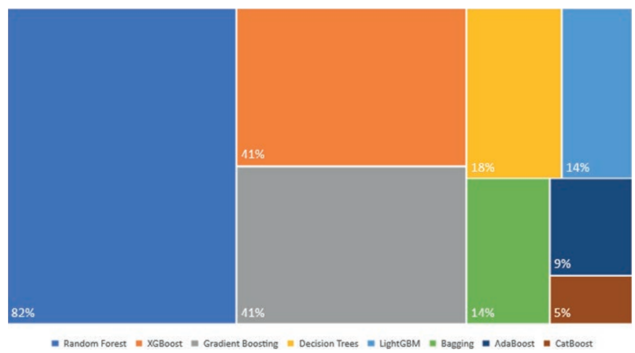

Figure 3. Tree with the machine learning models analyzed in the studies.

in property valuation (Steurer et al., 2021). The graph in Figure 4 shows these results. Most studies used multiple linear regression as a reference model. Among the 22 studies analyzed, 59% used multiple linear regression, 5% used spatial regression, and 36% used some other method of comparison to check the performance of machine learning-based models.

4.2. Research Question 02 (RQ-02)

This research question sought to map the typology of the properties modeled in each of the selected studies. Most studies used apartments for typology modeling. Among the selected studies, 59% used data on apartments; 32% on houses and 9% on urban land.

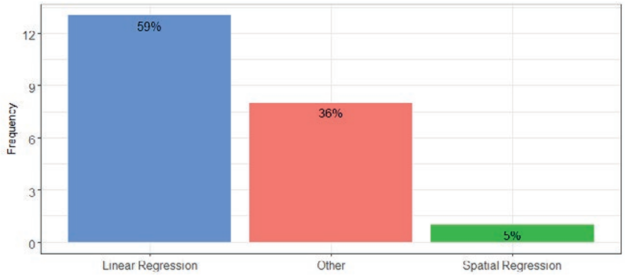

Figure 4. Model used as the main benchmark in the selected studies.

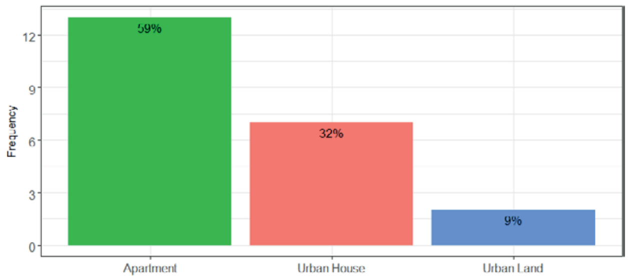

Figure 5. Typology of properties analyzed in the studies.

A research sub-issue sought to identify how the spatial dimension was explored in the studies, as shown in Figure 6. It was found that more than half of the studies used the spatial dimension only as a predictor variable. In total, 55% of the studies collected the UTM coordinates of the properties in the sample and used this information as two more predictor variables in the total set of variables. Importantly, among these 12 studies, Carranza et al. (2022) additionally calculated the Moran Global Index to verify the spatial autocorrelation of the data. It was also found that 40% of the studies disregarded the spatial dimension, i.e., they did not refer to these variables in the modeling process. Finally, 5% of the studies considered the spatial dimension by performing spatial regression. This finding about the spatiality of the data opens the possibility of carrying out research in which the spatial dimension is considered in spatial models of machine learning.

Finally, the graph in Figure 7 addresses the sub-issue aiming to identify the geographical location of the real estate data used in each analyzed study. It is observed that, in the majority of studies, the real estate data used originate from the researchers' affiliated country. However, there are exceptions, such as in the case of Carranza et al.'s study (2022), in which researchers affiliated with the University of Córdoba, Argentina, used real estate data from Fortaleza, Brazil. Additionally, another

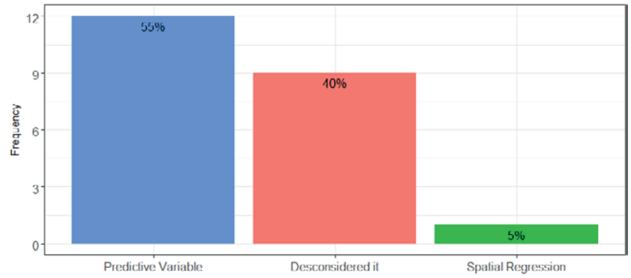

Figure 6. How the spatial dimension was considered in the studies.

interesting case was identified in which real estate data from the United Kingdom were explored by researchers affiliated with Istanbul Technical University, Turkey, as per Mete and Yomralioglu (2022).

Table 8 shows that three studies were conducted in Australia, China and Turkey, respectively. Taiwan and South Korea were the study area of two studies, each. Finally, there are several other countries that were data sources only once, e.g., Greece, Cyprus, India, Poland, Malaysia or Hong Kong. This finding indicates that there is room for researchers to investigate the behavior of machine learning models for mass property appraisal in other regions that are still little explored, or make use of new models in regions previously investigated to compare the results of both studies.

4.3. Research Question 03 (RQ-03)

This research question seeks to identify the indicators used in the studies to check the performance of the models. It was found that 82% of the studies calculated RMSE, i.e., it was the metric most often used by researchers when they wished to check the quality of machine learning models while carrying out mass property appraisals. The MAE and MAPE indicators were each calculated for 50% of the selected studies. The MSE indicator was adopted in only 10% of cases.

It should be noted that metrics such as the coefficient of dispersion (COD) and the price related differential (PRD), strongly recommended in mass appraisals for property tax by the IAAO (2013), were calculated in only 23% and 14% of the studies, respectively. Figure 8 shows the results of this analysis.

Indicators such as COD and PRD are calculated by comparing the values predicted by the models with the values found in the market, and checking if the dispersion between these values falls within the limits established by the aforementioned standard. COD is a measure of horizontal dispersion that provides information on the standardized evaluation of the set of properties

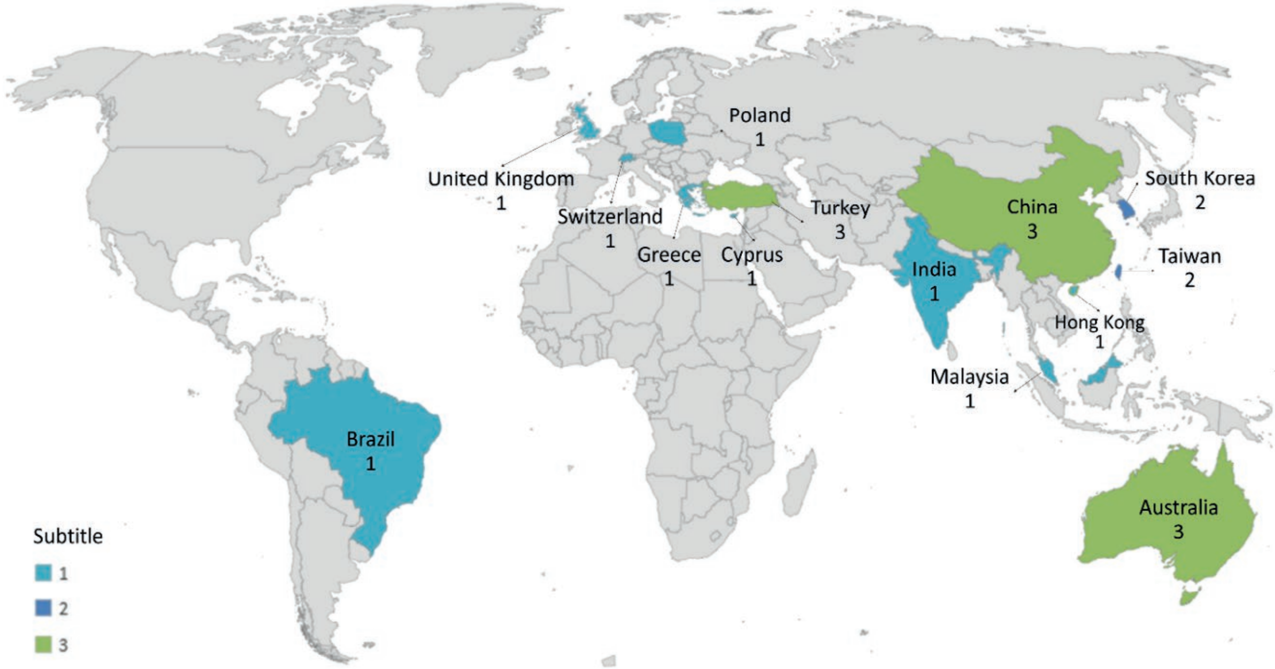

Figure 7. Number of studies and source of data in use.

Table 8. Number of studies whose data source is the country informed.

Country	Number of Studies	Country	Number of Studies
Hong Kong	1	Greece	1
Poland	1	Switzerland	1
Cyprus	1	Taiwan	2
United Kingdom	1	South Korea	2
Brazil	1	Turkey	3
Malaysia	1	China	3
India	1	Australia	3

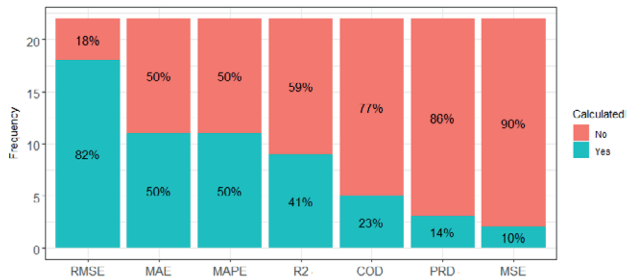

Figure 8. Performance metrics calculated in the selected articles.

while PRD is a measure to detect systematic differences in the appraisal of high- and low-value properties; it checks whether an appraisal is regressive or progressive (IAAO, 2013). These metrics are important indicators of possible inequity in the property taxation process.

A sub-issue of this research question sought to map which machine learning model had the best global performance in predicting property values. In each analyzed study, the model described by the authors was considered as the one that had the best performance or, in the absence of this information, the one that presented the best results for the set of adopted metrics.

The graph in Figure 9 shows that the Random Forest model was the one that presented the best performance in most of the selected studies. Exactly 50% of the studies reported that Random Forest is the best machine learning model for mass property appraisals. The XGBoost and Gradient Boosting models appear with the same number of citations: each was reported as the best performing model in 23% of the studies. Finally, only 4% of studies reported another machine learning model as having best performance.

To complement this research sub-issue and elucidate some important points, the graph in Figure 10 shows, among the 22 selected studies, the machine learning models checked in each selected study. The model marked with a green circle is the one that was consid-

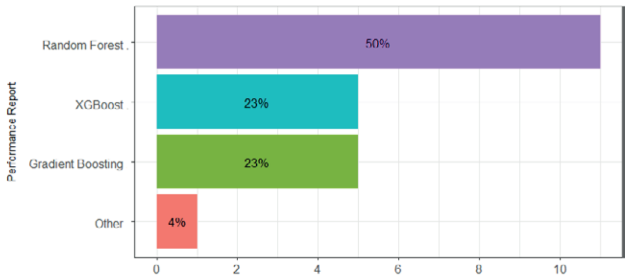

Figure 9. Models with the best performance in each analyzed study.

ered, in the respective study, as having the best global performance among all the analyzed models.

Among the 18 studies that included the Random Forest model in the analysis, this model had the best performance in 61% of them. Studies S-05, S-10, S-16 and S-19 did not test the Random Forest model in their analyses.

Another relevant aspect is that in studies S-05, S-16 and S-19, only two machine learning models were tested: XGBoost and another that is not part of the scope of models analyzed in this systematic mapping. In all these three studies, the XGBoost model showed the best performance.

Finally, it should be noted that the Gradient Boosting model performed better in most cases in which it was applied. The model appeared in 9 studies and performed better in mass property appraisals of 5 studies, which accounts for 55% of the total number of studies that used it.

4.4. Research Question 04 (RQ-04)

In order to map research trends and characteristics, an analysis was made of the set of variables found in each of the 22 selected articles. The main facts extracted from this analysis are detailed below.

4.4.1. Evolution of studies over time

The graph in Figure 11 shows that the number of published studies involving the use of machine learning models for mass property appraisal has been growing since the publication (2016) of the first study retrieved. The regression line, adjusted to the data, shows a growing trend in the studies, i.e., there is an increasing interest of the research community in this topic. Importantly, this systematic mapping considered data collected up to June, 2022, which explains the small number of studies for the trend of publications for the respective year.

Except for 2015, at least one study was published every year within the observation range. It is found that, from 2019 onwards, research on this topic began to grow at a faster rate - a fact observable through the slope of the regression line when considering the data as of the respective date.

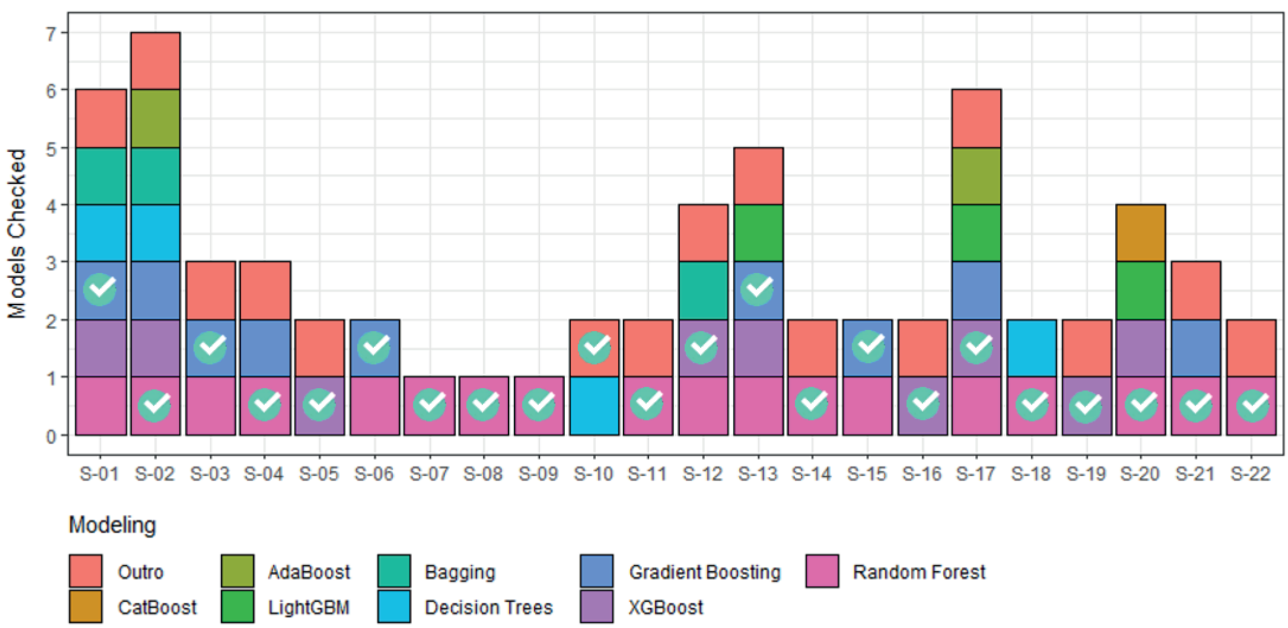

Figure 10. Models analyzed in each study and their best performance.

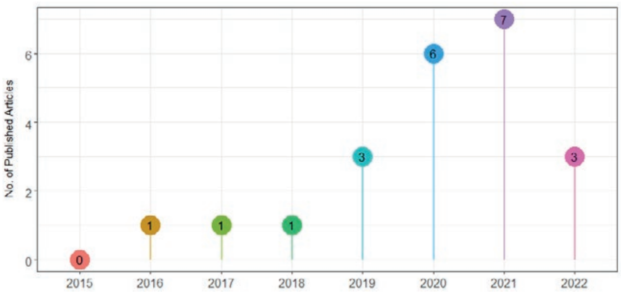

Figure 11. Evolution in the number of publications over the years.

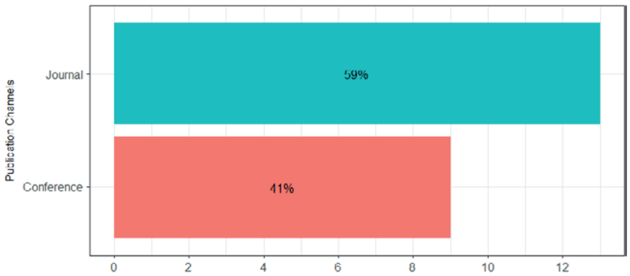

Figure 12. Number of studies by publication channel.

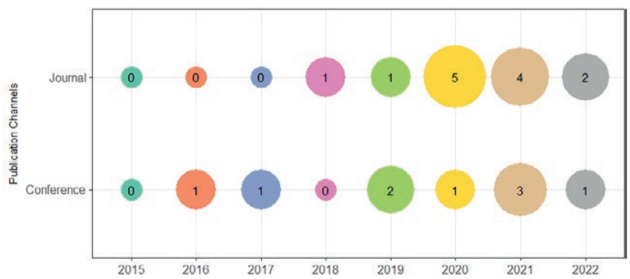

Figure 13. Evolution of studies by publication channel over time.

4.4.2. Main channels of publication of studies

The graph in Figure 12 shows that most studies were published in scientific journals. Among the 22 selected studies, 59% were published in some scientific journal and 41% in conference proceedings.

A greater number of articles published in journals may indicate that the topic is becoming a more mature area of research (Uludag et al., 2022), although it is relatively recent. In general, researchers prefer to publish their articles in journals because this type of publication brings more scientific benefits.

The evolution of studies by publication channel, as seen in Figure 13, indicates a growing trend in the past four years for publications in both conference proceedings and journals. However, it is noted that in each of the last three years, the number of publications in scientific journals exceeded the number of publications in conference proceedings. This observation reaffirms, once again, that the use of machine learning in property appraisal has become a more mature field of research, with the publication of more robust and comprehensive studies in recent years.

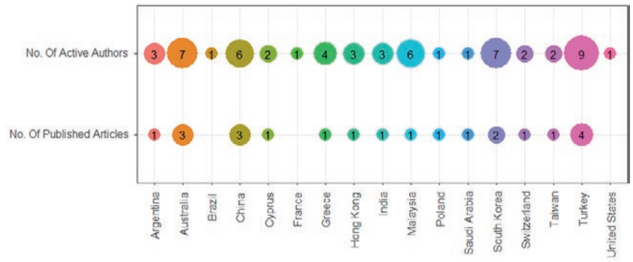

Figure 14. Countries with the most publications and active authors.

4.4.3. Most active countries in the field of studies

The graph in Figure 14 shows the most active countries in research relating machine learning and mass property appraisal. There was a total of 17 countries with active researchers, and Turkey stood out for having with the highest number of publications. It had 18.2% publications about this topic, within the analyzed time interval, and these publications were produced by 9 researchers. The analysis of the graph shows that the most active countries in publications in this area of research are Australia, China and Turkey, and the countries that concentrate the largest number of researchers are Australia, China, Malaysia, South Korea and Turkey. In Brazil, France and the United States, there are also active researchers.

This finding demonstrates that these scholars participate in international research networks, thus collaborating with the advancement of scientific research on the theme of this review. It also shows that the use of machine learning in the process of mass property appraisal is a globally relevant research topic.

4.4.4. Affiliation of researchers active in the field

Among the active authors in research related to the use of machine learning in mass property appraisal, Mehrdad Nejad, Jie Lu, and Vahid Behbood (University of Technology Sydney) stand out, each with two published works. As depicted in Figure 15, the institutions with the highest number of active researchers in this systematic mapping are observed.

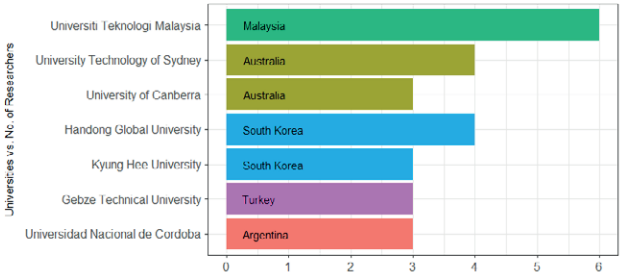

Figure 15. Universities with the highest number of active researchers.

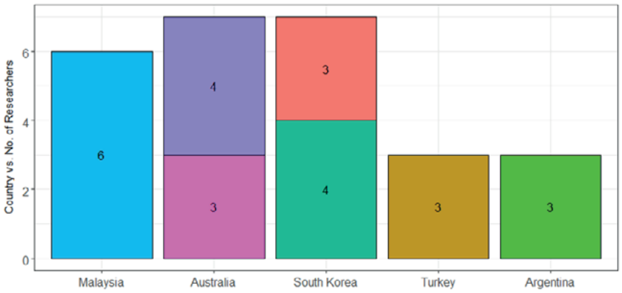

Figure 16. Countries of universities with the highest number of active researchers.

Regarding the affiliation of active researchers, it is noted that, among the 22 selected studies, all researchers are affiliated with a university. Universiti Teknologi Malaysia, University of Technology Sydney, and Handong Global University stand out as the top three institutions with the highest number of active researchers in studies related to the use of machine learning in mass property appraisal. These three universities respectively have six, four, and four researchers.

Figure 16 shows that, among the universities with the highest number of active researchers, one is located in Malaysia (six researchers), two in Australia (seven researchers), two in South Korea (seven researchers), one in Turkey (three researchers) and one in Argentina (three researchers). It should also be noted that the seven universities with the highest number of active researchers were located across virtually all continents: the Americas (Argentina), Asia (North Korea and Malaysia), Eurasia (Turkey) and Oceania (Australia). This reinforces the idea that the topic addressed in this systematic mapping is a globally relevant research topic.

4.4.5. Number of citations of selected articles

The number of citations of the studies selected for this systematic mapping, as shown in Figure 17, shows

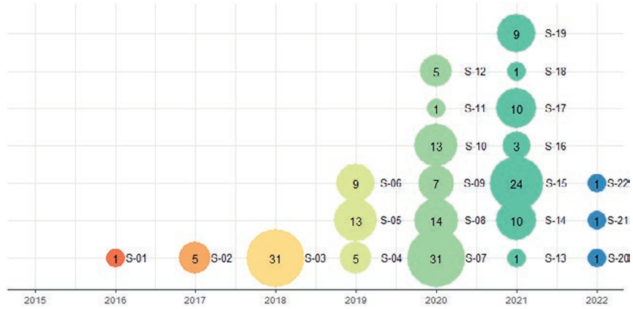

Figure 17. Number of citations of the studies considered in this review.

an increasingly relevant impact on the scientific community.

Publications S-03 by Michael Mayer et al. and S-07 by Jengei Hong et al., present the greatest number of citations according to the Semantic Scholar: 31 citations for each article, on the date this information was collected. The studies S-15 by Winky K.O. Ho et al. and S-08 by Seckin Yilmazer et al., had 24 and 14 citations each, respectively.

All four studies with the highest number of citations were published in scientific journals: S-03 in the Journal of European Real Estate Research, S-07 in the International Journal of Strategic Property Management, S-15 in the Journal of Property Research and S-08 in Land Use Policy. This reinforces the evidence that publications in scientific journals enjoy greater prestige in the research community. This finding is even more relevant when considering that in 2020 and 2021, a period in which publications in scientific journals increased, the number of citations of the studies present in this systematic mapping also increased, in comparison to conference proceedings. The scientific journal Land Use Policy received the highest number of publications among the analyzed studies: S-08 by Seckin Yilmazer et al. and S-21 by Juan P. Carranza et al.

4.4.6. Cloud of keywords cited in the studies

In scientific studies, the selection of keywords aims to facilitate the efficient retrieval of the content of a text for readers (Garcia et al., 2019). In this sense, the set of keywords of a scientific study allows other researchers to find it when they are carrying out research on that topic.

All keywords of the studies of this systematic mapping were searched, and Figure 18 and Table 9 were designed with the word cloud to show the absolute frequency of the 10 most cited keywords in the respective articles.

Figure 18. Keyword cloud of the studies considered in this review.

Table 9. Frequency of the 10 keywords most cited in the articles.

Word	Frequency
machine learning	12
mass appraisal	6
random forest	5
property valuation	4
real estate valuation	3
real estate	3
automated valuation models	3
house prices	2
gradient boosting	2
regression	2

These keywords show the strong relationship between ‘machine learning’ and ‘mass appraisal’. These were the two most cited keywords; ‘machine learning’ was cited twice as many times when compared to ‘mass appraisal’, which was the second most cited keyword. These two keywords are closely related to the theme of this systematic mapping: machine learning models for mass property appraisal.

Other words also featured prominently in the keyword cloud, e.g., the machine learning algorithms ‘random forest’ and ‘gradient boosting’, as well as words related to the real estate market, such as ‘property valuation’ and ‘real estate’. This keyword cloud presents evidence of the strong relationship between the topic addressed in this systematic mapping and the articles selected for this analysis.

5. DISCUSSION

This section shows the analysis of the results of this systematic mapping study. It also highlights issues that need to be further explored and makes some suggestions for future research.

5.1. Analysis of results

Considering the set of 332 works returned in the initial searches, including those of the snowballing process, this systematic mapping study was carried out with a final selection containing 22 studies, which sought to answer the central question of this review: *What is the state of the art of the literature regarding the use of machine learning models for mass property appraisal?* The small number of selected studies may be explained by the fact that research is still incipient regarding the use of machine learning algorithms based on regression trees for property appraisal; in addition, the focus was on studies that directly associated the use of machine learning with mass property appraisals. However, despite the small number of publications, there are gaps that can pave the way for new research opportunities and challenges that can serve as a basis for future researchers to explore the theme of the present mapping study.

It was found that *the most frequently used machine learning models in research on mass property appraisal* (**RQ-01**) are Random Forest, Gradient Boosting and XGBoost, Decision Trees, LightGBM, Bagging, Ada-

Boost and CatBoost. However, the use of the Random Forest model is predominant (S-01, S-02, S-03, S-04, S-06, S-07, S-08, S-09, S-11, S12, S-13, S-14, S-15, S-17, S18, S-20, S-21 and S-22) in research on the use of machine learning for mass property appraisals. The results of these machine learning models are, in most cases, compared with the results achieved by the multiple linear regression model (S-01, S-03, S-04, S-06, S-07, S-08 , S-12, S-14, S-16, S-17, S-18, S-19, S-21), which is often used by engineers and researchers when carrying out mass appraisals for property tax. Importantly, there was only one study (S-22) that compared the results of the machine learning models with those found by the ordinary kriging model.

The property typologies most frequently addressed in studies on mass property appraisal (**RQ-02**) are urban land (S-16 and S-21), urban houses (S-02, S-03, S-04, S-05, S-13, S-19 and S-20) and apartments (S- 01, S-06, S-07, S-08, S-09, S-10, S-11, S-12, S-14, S-15, S-17, S-18 and S-22). However, only one of these studies (S-17) considered the spatial dimension of the data, represented by spatial regression. Among these studies, 12 chose to use the UTM location coordinates of the properties as a predictor variable while another 9 studies did not use the spatial dimension in their analyses. Previous studies used data from different sources; three studies used data from properties located in Turkey (S-08, S-09 and S-11); three studies, from properties in China (S-04, S -12 and S-17), and three other studies, from properties located in Australia (S-01, S-02 and S-05). There are also data from properties located in countries such as Greece, North Korea, Switzerland, Brazil, and India.

It was found that *machine learning techniques are assessed, with respect to the accuracy of machine learning modeling (***RQ-03**) by several indicators. The most frequent in the studies selected in this systematic mapping were RMSE, COD, PRD, MSE, MAE, MAPE and R^2. Among these indicators, both RMSE and MAE are the most regularly used in model assessment studies; however, it cannot be argued that RMSE outperforms MAE, or vice versa; instead, a combination of metrics, including, but certainly not limited to, RMSEs and MAEs, is often required to assess model performance (Chai and Draxler, 2014). According to Bicak (2021), RMSE and MAE have informative value; therefore, it is advisable to use both. The RMSE indicator, calculated by using the square root of the mean squared differences between the observed and predicted values, is the one that appears in the vast majority of studies (S-01, S-02, S-03, S-04, S-06, S-08, S-09, S-11, S-12, S-13, S-15, S-16, S-17, S-18, S-19, S-20, S- 21 and S-22), whereas the MAE indicator, which measures the average of the error differences between the observed and predicted values by the models without considering their direction (Islam et al., 2022), was calculated by a smaller number of studies (S-02, S-03, S-05, S-06, S-09, S-10, S-12, S-13, S-18, S-20 and S-22). Indicators such as COD (S-06, S-07, S-08, S-09 and S-21) and PRD (S-08, S-09 and S-21), strongly indicated for analyzing the quality of mass assessments (IAAO, 2013), were calculated by few studies. It was also found that among the machine learning models used for mass property appraisal in the 22 analyzed studies, the Gradient Boosting model was cited as the one that presented the best global accuracy in 5 studies (S-01, S -03, S-06, S-13 and S-15); the XGBoost model was cited in another 5 studies (S-05, S-12, S-16, S-17 and S-19) and, finally, the Random Forest model was cited in 11 studies (S-02, S-04, S-07, S-08, S-09, S-11, S-14, S-18, S-20, S-21 and S-22).

5.1. Suggestions for further research

Based on the contributions of this systematic mapping study, the following suggestions for future research can be made:

- **(i) Conducting research that makes combined use of machine learning models and geostatistics**: No studies were found that combined geostatistics with machine learning for mass property appraisals. There are studies combining these two techniques in other areas; for example, Su et al. (2020) examined the combination of these techniques to estimate biomass in Chinese forests; however, in the area of mass appraisals, there are no studies to date.
- **(ii) Exploring feature engineering for selection of relevant variables and comparison with traditional modeling:** among the studies analyzed, some used feature importance for modeling purposes. However, there were no studies that demonstrate the gain in accuracy when comparing the results of modeling with and without the application of feature importance.
- **(iii) Developing a method of mass property appraisal using spatial random forest regression:** the analyzed studies demonstrated the use of the spatial dimension in the form of inclusion of a new predictor variable; however, no studies were found that actually performed the spatial random forest regression process (Benito, 2021). This study demonstrated that the random forest model shows good predictive performance even when using many covariates with nonlinear relationships, while the spatial regression model shows good predictive performance when using many records that are spatially autocorrelated. Thus, the application of the spatial random forest

regression model can be an interesting strategy to explore.

(iv) Checking the accuracy of machine learning models for property appraisals in other regions: among the analyzed studies, few of them have explored the use of machine learning techniques in urban land (S-16 and S-21), and few countries have explored these techniques in mass appraisals. In Brazil, for example, there is only one study that used machine learning models for mass appraisal of urban land in the city of Aracajú. Studies with data on land from other Brazilian municipalities or even new typologies, such as houses and apartments, are an alternative for comparing the effectiveness of machine learning models in mass appraisals for property tax.

6. THREATS TO VALIDITY

This systematic mapping study was conducted following a rigorous methodology, with special attention to the selection and analysis of published studies. Although this methodology is widely employed by various authors, it does have some limitations. The results observed in this research may be affected by threats to validity, despite attempts to mitigate them throughout the stages of this systematic study. For example:

(i) Article selection bias: to minimize this threat, both the protocol and the execution process were reviewed by experienced researchers. To further mitigate the article selection bias, a set of criteria was created, as presented in Section 3.2.2; it sought to ensure that the most relevant publications were found by search engines. For this process, the most important terms related to the use of machine learning in mass property appraisals were selected, and a generic search string was designed. The focus was on studies published in conference proceedings or in scientific journals. The objective was, therefore, to determine the state of the art of high-quality, peer-reviewed scientific articles that followed strict publication guidelines.

(ii) Incomplete searches: owing to the exclusion criteria adopted, this systematic mapping study may not have reached exactly all the studies on the topic, which may affect the completeness of the present study. For example, by creating an exclusion criterion that eliminates all studies that were not in English, relevant research studies published in different languages, such as Portuguese, were ignored. To mitigate this risk, studies were carried out in electronic databases commonly used in the engineering area and which contain a large number of indexed journals and conference proceedings.

(iii) Data extraction bias: the accuracy of the results of a systematic mapping study can be strongly affected by researcher bias in data extraction. To mitigate the impact of potential researcher bias on data extraction, two researchers specified a list of items to be extracted and reached a consensus on the understanding of each of these specified items. The set of primary studies selected in this systematic mapping was then distributed to the two researchers and they both carried out, independently, the extraction of data from all studies. Discrepancies arising from the data extraction process were resolved together, through a consensus meeting; after reanalysis, the two researchers decided on the correct information of the extracted data.

7. CONCLUSIONS

The use of machine learning models in mass property appraisal has been progressing towards becoming a mature research area, as evidenced by the growing number of publications on the subject in scientific journals and conferences in recent years. This trend has led to a gradual increase in the body of knowledge on the topic. However, to date, no systematic literature mapping has been identified that systematically identifies and analyzes the state of the art in this research area. This study sought to fill this gap and provide an overview of the latest research using tree-based machine learning models in mass property appraisal for tax purposes.

Delving into the realm of mass property appraisal through the use of machine learning models revealed a simultaneously intriguing and complex landscape. This study, by meticulously investigating the available literature, not only highlights the popularity and efficiency of certain models, such as Random Forest and Gradient Boosting, but also points to the urgent need for the standardization and rigorous application of these technologies. Despite the revolutionary promises of precision and efficiency in property valuation for tax purposes, significant challenges remain to be overcome.

For starters, the wide range of criteria used to measure the accuracy of these models reveals the absence of a consensus or standardized system that allows for direct comparisons. While this situation showcases the richness of methodological approaches, it may complicate the clear presentation of results and, ultimately, the adoption of these technologies by tax authorities and appraisers.

Moreover, the geographical concentration of research in certain regions suggests the influence of local factors, both in terms of available data and real estate market idiosyncrasies, on the effectiveness of machine learning models. This raises questions about the universal applicability of these solutions and emphasizes the importance of future research that takes into account diversity in mass property appraisal for tax purposes.

Although models like Random Forest are notable for their robustness and accuracy, it is crucial to remember that technology alone is not a panacea. The success of these models is directly dependent on the quality and comprehensiveness of the data they are fed. Therefore, the importance of rigorously collecting, processing, and analyzing data cannot be underestimated. Furthermore, in conducting mass appraisals, it is essential to consider ethical and social justice issues, especially in relation to tax equity, an important aspect that should not be overlooked.

This study unveils a field rich in opportunities for research and innovation at the intersection of machine learning and mass property appraisal for tax purposes. The critical approach adopted here does not seek to diminish the transformative value of these technologies, but rather to underscore the complexity and responsibilities involved in their implementation. The future, filled with possibilities and challenges, demands ongoing collaboration between academics, professionals, and policy makers to ensure that technological advancements promote fairer, transparent, and effective appraisal practices.

ACKNOWLEDGMENTS

The present study was conducted with the support of the Coordination for the Improvement of Higher Education Personnel - Brazil (CAPES) - Funding Code 001, the Support Fund for Maintenance and Development of Higher Education (FUMDES), the Federal University of Santa Catarina (UFSC), the Graduate Program in Engineering and Knowledge Management (PPGEGC), and the Federal Institute of Santa Catarina (IFSC), Brazil.

REFERENCES

Almaslukh, B. (2021). A gradient boosting method for effective prediction of housing prices in complex real estate systems. In *International Conference on Technologies and Applications of Artificial Intelligence (TAAI)*, 217–222. https://doi.org/10.1109/taai51410.2020.00047

Aydinoglu, A. C., Bovkir, R., & Colkesen, I. (2020). Implementing a mass valuation application on interoperable land valuation data model designed as an extension of the national GDI. *Survey Review,* 53, 349–365. https://doi.org/10.1080/00396265.2020.1771967

Benito, B. M. (2021). SpatialRF: Easy Spatial Regression with Random Forest. R package version 1.1.3. https://doi.org/10.5281/zenodo.5525027

Bicak, D. (2021). Geographical random forest model evaluation in agricultural drought assessment. Master thesis. Charles University, Faculty of Science, Department of Applied Geoinformatics and Cartography. Prague, 2021.

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

Brereton, P., Kitchenham, B. A., Budgen, D., Turner, M., & Khalil, M. (2007). Lessons from applying the systematic literature review process within the software engineering domain. *Journal of Systems and Software,* 80, 571–583. https://doi.org/10.1016/j.jss.2006.07.009

Budgen, D., Turner, M., Brereton, P., & Kitchenham, B. A. (2008). Using mapping studies in software engineering. In *Proceedings of Psychology of Programming Interest Group* (*PPIG), 8, 195–204.* https://ppig.org/files/2008-PPIG-20th-budgen.pdf

Carranza, J. P., Piumetto, M. A., Lucca, C. M., & da Silva, E. (2022). Mass appraisal as affordable public policy: open data and machine learning for mapping urban land values. *Land Use Policy*, 119, 106211. https://doi.org/10.1016/j.landusepol.2022.106211

Chai, T., & Draxler, R. R. (2014). Root mean square error (RMSE) or mean absolute error (MAE). *Geoscientific Model Development Discussions,* 7, 1525–1534. https://doi.org/10.5194/gmd-7-1247-2014

Chen, L., Babar, M. A., & Zhang, H. (2010). Towards an Evidence-Based Understanding of Electronic Data Sources. In Proceedings of the 14th International Conference on Evaluation and Assessment in Software Engineering (EASE), British Computer Society, Swinton, pp 135–138. https://doi.org/10.14236/ewic/ease2010.17

Dimopoulos, T., & Bakas, N. P. (2019). Sensitivity Analysis of Machine Learning Models for the Mass Appraisal of Real Estate. Case Study of Residential Units in Nicosia, Cyprus. *Remote Sensing,* 11(24), 3047. https://doi.org/10.3390/rs11243047

Doumpos, M., Papastamos, D., Andritsos, D. A., & Zopounidis, C. (2021). Developing automated valuation models for estimating property values: a comparison of global and locally weighted approaches. *Annals of Operations Research*, 306, 415–433. https://doi.org/10.1007/s10479-020-03556-1

Fontoura Júnior, C. F., Uberti, M. S., & Tachibana, V. M. (2020). Mass appraisal of apartment through geographically weighted regression. *Boletim de Ciências Geodésicas*, 26, e2020005. https://doi.org/10.1590/s1982-21702020000200005

Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of Statistics*, 29, 1189–1232. https://doi.org/10.1214/aos/1013203451

Garcia, D. C. F, Gattaz, C. C., & Gattaz, N. C. (2019). A relevância do título, do resumo e de palavras-chave para a escrita de artigos científicos. *Revista de Administração Contemporânea*, 23(3), 1–9. https://doi.org/10.1590/1982-7849rac2019190178.

Gnat, S. (2021). Property Mass Valuation on Small Markets. *Land,* 10(4), 388. https://doi.org/10.3390/land10040388

Ho, W., Tang, B., & Wong, S. W. (2020). Predicting property prices with machine learning algorithms. *Journal of Property Research,* 38, 48–70. https://doi.org/10.1080/09599916.2020.1832558

Hong, J., Choi, H., & Kim, W. (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

IAAO (2013). *Standard on Ratio Studies*. International Association of Assessing Officers. Kansas City, IAAO.

IEEE (2022). The Impact of Technology in 2022 and Beyond: IEEE Global Survey. Available at: https://transmitter.ieee.org/impact-of-technology-2022/ (accessed 23 May 2022)

Islam, M. D., Li, B., Lee, C., & Wang, X. (2022). Incorporating spatial information in machine learning: The Moran eigenvector spatial filter approach. *Transactions in GIS,* 26, 902–922. https://doi.org/10.1111/tgis.12894

Ja'afar, N.S., Mohamad, J., & Ismail, S. (2021). Machine learning for property price prediction and price valuation: a systematic literature review. *Planning Malaysia*, 19(3), 411–422. https://doi.org/10.21837/pm.v19i17.1018

Kim, J., Lee, Y., Lee, M., & Hong, S. (2022). A comparative study of machine learning and spatial interpolation methods for predicting house prices. *Sustainability*, 14(15), 9056. https://doi.org/10.3390/su14159056

Kitchenham, B. A., & Brereton, P. (2013). A systematic review of systematic review process research in software engineering. *Information and Software Technology*, 55, 2049–2075. https://doi.org/10.1016/j.infsof.2013.07.010

Kitchenham, B. A., & Charters, S. (2007). Guidelines for performing systematic literature reviews. In Software Engineering, Technical Report EBSE- 2007-01, 1-57. Departament of Computer Science Keele University, Keele.

Kitchenham, B. A., Brereton, P., & Budgen, D. (2010). The educational value of mapping studies of software engineering literature. In *Proceedings of the 32nd ACM/IEEE International Conference on Software Engineering, Volume 1*, 589–598. https://doi.org/10.1145/1806799.1806887

Kitchenham, B. A., Mendes, E., & Travassos, G. H. (2007). Cross versus within-company cost estimation studies: a systematic review. *IEEE Transactions on Software Engineering,* 33, 316-329. https://doi.org/10.1109/tse.2007.1001

Liporoni, A. S. (2014). *Avaliação em massa com ênfase em planta de valores.* In IBAPE/SP Engenharia de Avaliações, 2. ed., 2, 664.

Manasa, J., Gupta, R., & Narahari, N. (2021). Machine learning based predicting house prices using regression techniques. In *Proceedings of the 2nd International Conference on Innovative Mechanisms for Industry Applications (ICIMIA 2020)*, 624–630. https://doi.org/10.1109/icimia48430.2020.9074952

Mayer, M., Bourassa, S.C., Hoesli, M. & Scognamiglio, D. (2019). Estimation and updating methods for hedonic valuation. *Journal of European Real Estate Research*, 12(1), 134–150. https://doi.org/10.1108/JERER-08-2018-0035

McCluskey, W., Deddis, W. G., Mannis, A., McBurney, D., & Borst, R. A. (1997). Interactive application of computer assisted mass appraisal and geographic information systems. *Journal of Property Valuation and Investment,* 15, 448–465. https://doi.org/10.1108/14635789710189227

Mete, M. O., & Yomralioglu, T. (2022). GIS & machine learning based mass appraisal of residential properties in England & Wales. In *30th Annual Geographical Information Science Research UK (GISRUK)*. https://doi.org/10.5281/zenodo.6410120

Nakamura, W. T., Oliveira, E., Oliveira, E. H., Redmiles, D. F., & Conte, T. (2022). What factors affect the UX in mobile apps? A systematic mapping study on the analysis of app store reviews. *Journal of Systems and Software,* 193, 111462. https://doi.org/10.1016/j.jss.2022.111462

Nejad, M. Z., Lu, J., & Behbood, V. (2017). Applying dynamic Bayesian tree in property sales price estimation. In *2017 12th International Conference on Intelligent Systems and Knowledge Engineering (ISKE)*, 1–6. https://doi.org/10.1109/ISKE.2017.8258810

Nejad, M. Z., Lu, J., Asgari, P., & Behbood, V. (2016). The effect of google drive distance and duration in residential property in Sydney, Australia. In *World Scientific Proceedings Series on Computer Engineering and Information Science*. https://doi.org/10.1142/9789813146976_0101

Niu, J., & Niu, P. (2019). An intelligent automatic valuation system for real estate based on machine learning. In *Proceedings of the International Conference on Artificial Intelligence, Information Processing and Cloud Computing (AIIPCC19), 1–6*. https://doi.org/10.1145/3371425.3371454

Pai, P., & Wang, W. (2020). Using machine learning models and actual transaction data for predicting real estate prices. *Applied Sciences*, 10(17), 5832. https://doi.org/10.3390/app10175832

Petersen, K., Feldt, R., Mujtaba, S., & Mattsson, M. (2008). Systematic Mapping Studies in Software Engineering. In *12th International Conference on Evaluation and Assessment in Software Engineering* (EASE). BCS Learning & Development. https://doi.org/10.14236/EWIC/EASE2008.8

Petersen, K., Vakkalanka, S., & Kuzniarz, L. (2015). Guidelines for conducting systematic mapping studies in software engineering: an update. *Information and Software Technology*, 64, 1–18. https://doi.org/10.1016/j.infsof.2015.03.007

Rodríguez, P., Haghighatkhah, A., Lwakatare, L.E., Teppola, S., Suomalainen, T., Eskeli, J., Karvonen, T., Kuvaja, P., Verner, J. M., & Oivo, M. (2017). Continuous deployment of software intensive products and services: a systematic mapping study. *Journal of Systems and Software,* 123, 263–291. https://doi.org/10.1016/j.jss.2015.12.015

Sevgen, S. C., & Aliefendioglu, Y. (2020). Mass apprasial with a machine learning algorithm: random forest regression. *Journal of Information Technologies*, 13(3), 301–311 . https://doi.org/10.17671/gazibtd.555784

Steurer, M., Hill, R., & Pfeifer, N. (2021). Metrics for evaluating the performance of machine learning based automated valuation models. *Journal of Property Research,* 38, 99–129. https://doi.org/10.1080/09599916.2020.1858937

Su, H., Shen, W., Wang, J., Ali, A., & Li, M. (2020). Machine learning and geostatistical approaches for estimating aboveground biomass in Chinese subtropical forests. *Forest Ecosystems,* 7, 1–20. https://doi.org/10.1186/s40663-020-00276-7

Uberti, M. S., Antunes, M. A., Debiasi, P., & Tassinari, W. D. (2018). Mass appraisal of farmland using classical econometrics and spatial modeling. *Land Use Policy*, 72, 161–170. https://doi.org/10.1016/j.landusepol.2017.12.044

Uludag, O., Philipp, P., Putta, A., Paasivaara, M., Lassenius, C., & Matthes, F. (2022). Revealing the state of the art of large-scale agile development research: a systematic mapping study. *Journal of Systems and Software*, 194, 111473. https://doi.org/10.1016/j.jss.2022.111473

Valier, A., & Micelli, E. (2020). Automated models for value prediction: a critical review of the debate. *Valori e Valutazioni*, 24, 151–162. https://www.researchgate.net/publication/348962087

Velumani, P., Priyadharshini, B., Mukilan, K., & Shanmugapriya (2022). A mass appraisal assessment study of land values using spatial analysis and multiple regression analysis model (MRA). *Materials Today: Proceedings,* 66, 2614–2625. https://doi.org/10.1016/j.matpr.2022.07.224

Wang, 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

Wohlin, C., Runeson, P., Höst, M., Ohlsson, M. C., & Regnell, B. (2012). *Experimentation in Software Engineering*. Heidelberg, Springer Science & Business Media. https://doi.org/10.1007/978-3-642-29044-2

Xu, L., & Li, Z. (2021). A new appraisal model of second-hand housing prices in China's first-tier cities based on machine learning algorithms. *Computational Economics,* 57, 617–637. https://doi.org/10.1007/s10614-020-09973-5

Yan, Z., & Zong, L. (2020). Spatial prediction of housing prices in beijing using machine learning algorithms. In *Proceedings of the 2020 4th High Performance Computing and Cluster Technologies Conference & 2020 3rd International Conference on Big Data and Artificial Intelligence,* 64–71. https://doi.org/10.1145/3409501.3409543

Yee, L.W., Bakar, N. A., Hassan, N. H., Zainuddin, N. M., Yusoff, R. C., & Rahim, N. Z. (2021). using machine learning to forecast residential property prices in overcoming the property overhang issue. In *2021 IEEE International Conference on Artificial Intelligence in Engineering and Technology (IICAIET)*, 1–6. https://doi.org/10.1109/iicaiet51634.2021.9573830

Yilmazer, S., & Kocaman, S. (2020). A mass appraisal assessment study using machine learning based on multiple regression and random forest. *Land Use Policy*, 99, 104889. https://doi.org/10.1016/j.landusepol.2020.104889

Zhang, R., Du, Q., Geng, J., Liu, B., & Huang, Y. (2015). An improved spatial error model for the mass appraisal of commercial real estate based on spatial analysis: Shenzhen as a case study. *Habitat International,* 46, 196–205. https://doi.org/10.1016/j.habitatint.2014.12.001

Zhao, Y., Chetty, G., & Tran, D. T. (2019). Deep Learning with XGBoost for Real Estate Appraisal. In *2019 IEEE Symposium Series on Computational Intelligence (SSCI)*, 1396–1401. https://doi.org/10.1109/SSCI44817.2019.9002790