Journal of the Midwest Association for Information Systems

Volume 2023 | Issue 2 Article 2

Date: 07-31-2023

A Review of Machine Learning Approaches for Real Estate Valuation

Thomas H. Root
Drake University, tom.root@drake.edu

Troy J. Strader
Drake University, troy.strader@drake.edu

Yu-Hsiang (John) Huang
Georgia College & State University, john.huang@gcsu.edu

Abstract

Real estate managers must identify the value for properties in their current market. Traditionally, this involved simple data analysis with adjustments made based on manager's experience. Given the amount of money currently involved in these decisions, and the complexity and speed at which valuation decisions must be made, machine learning technologies provide a newer alternative for property valuation that could improve upon traditional methods. This study utilizes a systematic literature review methodology to identify published studies from the past two decades where specific machine learning technologies have been applied to the property valuation task. We develop a data, reasoning, usefulness (DRU) framework that provides a set of theoretical and practice-based criteria for a multi-faceted performance assessment for each system. This assessment provides the basis for identifying the current state of research in this domain as well as theoretical and practical implications and directions for future research.

Keywords: real estate valuation, property valuation, machine learning, systematic literature review

DOI: 10.17705/3jmwa.000082

1. Introduction

Real estate is a major component of the global economy. The correct valuation of real estate plays a key role in the economy and is crucial for active market participants who are buying, selling and developing property, as well as property and non-property owners not actively engaged in the market. Governments use valuation techniques to determine property taxes which generate revenue for roads, schools and other government projects. Rental rates for non-property owners depend upon property valuation. Capital markets use valuations to determine collateral for both commercial and residential lending. Finally, activity in the real estate market creates employment for construction workers, income for retailers, and opportunities for service providers. The correct valuation of property is a crucial component of the complex web of economic activity linked to the real estate markets. The traditional valuation approach tasks market participants with collecting a wide range of data on both the characteristics of the property to be valued and comparable properties in close proximity. The goal is to then develop a system to understand the relative importance of the different inputs in relation to the specific property under consideration and arrive at an estimate of the property's value. Individual property appraisals often combine a variety of mathematical tools with a subjective adjustment of the mathematical results. This approach can be flawed because it has the potential for wide variation in estimated values and is at least partially dependent upon the experience, knowledge, and skill of a human appraiser.

Advances in technology have allowed for utilizing a wider variety of information and incorporating increasingly sophisticated analysis of the data. These advances have improved the ability to perform mass appraisals that evaluate groups of properties utilizing standardized methods and data. Mass appraisal is often utilized to inform the setting of property taxes and also in development planning. These same techniques may also be used as a starting point for individual valuations performed by market participants buying, selling and managing real estate. It is a common adage in real estate that the three most important determinants of price are location, location and location. The preferences for different property characteristics differ across locations making adjustments for local preferences a key component of property valuation. This has traditionally limited the applicability of mass appraisal as the sole valuation technique. Large data sets and traditional statistical methods have trouble accounting for locational differences limiting the effectiveness of many mass appraisal techniques.

In many domains, decision-makers have turned to machine learning (ML) technologies that have the potential to improve upon existing decision processes because they can quickly analyze a vast amount of timely data with a wide range of problem features. The task of real estate valuation appears to provide an excellent opportunity for improvements through the adoption of machine learning applications because of the large number of potential inputs and the nonlinear nature of the relationships between the variables. In other words, machine learning technology appears to fit the task of property valuation (Goodhue & Thompson, 1995). The utilization of machine learning techniques has expanded in many fields in the past decades as computing power has increased along with the availability of massive amounts of easily accessible and timely data.

The interaction between the appropriateness or fit of a technology to perform a task and the adoption of the technology has been the subject of a large volume of academic research (see Spies, Grobbelaar, and Boths (2020), Furneaux (2012) and Marangunić and Granić (2015)). This paper combines a systematic review of the use of machine learning technology in property valuation over the last two decades with a critical evaluation of the progress made aligning the fit of the technology to the property valuation task in a manner that supports its adoption. This provides the basis for answering several related questions. First, it enables us to see which machine learning technologies have been applied to this task and what improvements they have made when compared with alternative methods. Second, the review provides the basis for an analysis of which specific machine learning techniques (for example, artificial neural networks, genetic algorithms, etc.) show the most promise for improving property valuation or which technology best fits this task. Finally, the review provides insight into whether the academic literature addresses the correct elements of the task to promote successful adoption of the technology by real estate professionals. Investigating this question may also shed light on directions for future research and the path to adoption of machine learning in other finance-related decision domains.

In the following sections we discuss several traditional property valuation methods to provide the background for this study. We then identify both theoretical and practical criteria for evaluating the performance of machine learning property valuation techniques, describe the systematic literature review methodology used in this study, and discuss the findings from our review of each of the identified published studies. We conclude with a discussion of overall findings, related managerial and theoretical implications, and directions for future research that could extend our current knowledge in the domain of machine learning technology-based property valuation.

2. Introduction to Property Valuation

The value of real estate incorporates a wide range of variables including property characteristics, geographic amenities, market conditions, and preferences of market participants. Given the wide range of potential variables and the endless permutations that are possible, the process of valuation involves an understanding of the relative importance of the different variables and their respective impact on the value of a property. A common element included in the valuation is often comparing properties with similar identifying characteristics within close proximity and incorporating data on those properties to the one being considered. Without a formal quantitative methodology, the task of property valuation would be dependent solely upon the subjective ability of the person performing the valuation to determine the correct variables to consider and their impact on the final property value. The importance of each variable will vary based upon location, potential use of the property, availability of properties with similar characteristics in the market, preferences of potential buyers, the economic environment and a range of other variables. The desire to develop more formal quantitative techniques providing more objectivity in estimating the market value of a property has produced an extensive academic literature and range of related professional practices.

There are three broad approaches to property valuation: the sales comparison approach, income approach, and cost approach. The sales comparison approach focuses on comparing the attributes of a property under consideration to similar properties that have recently sold in the local real estate market. The standard comparable model is intuitively appealing when considering residential properties where characteristics such as number of bedrooms, number of bathrooms, lot size, total square footage, etc. are common points of comparison for potential buyers. This framework also aligns well with the hedonic regressions described by Rosen (1974). Each potential individual purchaser places different weights on each identifiable characteristic. For example, one buyer might place the most emphasis on the number of bedrooms, while a different purchaser may place the most emphasis the presence of a garage. While there are individual differences across buyers, common themes emerge such as a larger number of bedrooms generally increases the value of a home. Commercial properties also include many points of comparison allowing for a similar approach. However, with commercial properties there are often additional considerations related to the income generated from the property which are not applicable to the owner-occupied residential market.

When evaluating commercial property, the income approach is often used. The income approach requires an estimate of the present value of the future cash flow generated by the property. Participants evaluate the potential future earnings that could be generated from the property, including evaluating profits from rental income, incomes from industrial use, and/or the value added to the property through the development process. Residential properties purchased as investment vehicles providing rental income would be included in this area. Valuation methods focusing on the income approach expand upon basic comparable approaches to include modelling of the property as a revenue generating asset.

The cost approach is the third traditional approach to property valuation. The cost approach starts with the value of the land then adds the cost of building a new similar structure minus depreciation. This works well for relatively new construction and relies upon a comparison approach for the initial land value portion of the calculation before incorporating construction costs.

The three broad types of comparable approaches are intuitively appealing when considering the value of an individual property since they incorporate varying preferences of market participants and also the location of the property. However, there is a strong need for a methodology that has wider application and can help produce consistent objective valuations across properties to aid in appraisals. This is highlighted by Epley (2017) who focuses on the need for consistent terminology and characterization of the output produced by computer assisted automated valuation methods (AVM). One key point made is that many models in the academic literature do not contain reference to important requirements of the appraisal process required by the Uniform Standards of Professional Appraisal Practice (USPAP), which includes a definition of market value. Epley (2017) points to parts of the appraisal process missing in most AVM techniques, such as the requirement for an inspection. These differences are also discussed in the IAAO AVM standards. The standard recommends that the user review the values produced by AVMs that are based on traditional appraisal approaches (Epley, 2017). He also points to identified differences in AVM methodologies across countries, and a lack of transparency in the techniques utilized by some commercially available AVMs due to patent filings. Ultimately Epley (2017) points to the output from the AVM as being mislabeled and not a true market value of the property, but instead a starting point to aid in the determination of a market value. A clearer understanding of the methodology utilized in the ability of different machine learning approaches to property valuation will help bridge the potential gaps between the academic literature end user, an understanding necessary to evaluate the alignment of ML techniques with the task of aiding in property valuation.

Attempting to capture the intuition underlying individual property evaluation for a group of properties has long been a goal in the literature. The use of hedonic regressions to estimate house prices was first attempted by Rosen (1974) while discussing price setting for differentiated products that can be classified by a set of observable characteristics and prices. The application of standard computational methods and more advanced statistical methods such as multiple variable regression and stepwise regressions were the focus of much of the early literature on property valuation. Pagourtzi, *et al.* (2003) termed these approaches traditional valuation techniques and classified a second set of techniques, many utilizing machine learning, as advanced valuation methods. They classify traditional valuation techniques as those that develop a function of observable property characteristics to produce a valuation. Ultimately, the traditional methods assume a deterministic relationship between the independent variables used as points of comparison across projects, similar to the hedonic regressions suggested by Rosen (1974). Thus, given a set of property characteristics, a price can be determined by incorporating the implied importance of each characteristic from the representative sample. However, this is a poor representation of the actual pricing process which includes very localized changes in the relative importance of each variable. Additionally, there are possibly variables and relationships which are not directly observable which influence the value of a property.

Advanced techniques are those that attempt to capture the thought process of market participants and incorporate inferences about relationships and/or variables which are not easily observable or quantified by the traditional models. Basic examples include consumer perception of construction quality, age of the house, importance of proximity to different types of infrastructure, changes in consumer wealth and consumer perceptions of the future economic environment. Riddel (1999) first discussed how changes in local perceptions can be the product of relying on past price changes creating a localized feedback loop. Similarly, changes in localized future market expectations can create a speculative bias that influences prices. The impact of, and relationship among, both the observable and hidden variables will vary greatly by the location of the property. Rico-Juan and Taltavull de Laz Pav (2021) reference these influences as creating non linearities in the data and can result in spatial autocorrelation in local house prices.

The advanced models described by Pagourtzi *et al.* (2003) included various machine learning systems such as artificial neural networks (ANNs) and also other approaches such as fuzzy logic, spatial analysis models, and autoregressive integrated moving average (ARIMA) techniques. The common theme underlying these techniques is that they have the potential to model more complex, nonlinear, and sometimes hidden relationships that exist between the underlying property characteristics. While early attempts to utilize advanced valuation techniques had been attempted at the time of their classification there has been large advancements in machine learning approaches over the last twenty years which are the focus of this study.

3. Identifying Machine Learning Property Valuation System Performance Criteria

The performance of a machine learning property valuation system cannot be evaluated solely based on its output. Or stated another way, how accurately it identifies a property value. Consideration must also be given to the alignment of the technology with the task, the replicability of the results, and the quality of the data. The validity of a machine learning property valuation system should include consideration of the inputs used, the process for its selection and use, and its output. In other words, are the newer machine learning approaches perceived by users as useful and easy to use, and do they effectively model the problems in this domain by matching the technology to the task? In addition, from a real-world practice perspective, do these systems require input data that is readily available, can they explain the process used to produce their valuation results, and do they follow real estate industry appraisal principles? In this section we identify an overall set of theory-based and practice-based property valuation system evaluation criteria that are related to the questions listed above.

Worzala, Lenk, & Silva (1995) provided an early exploration into the application of neural networks to real estate valuation and ultimately questioned its usefulness. Their work provides an interesting starting point for an analysis of the adoption of machine learning in real estate valuation. Using a dataset of 288 single-family residential properties in Fort Collins, CO, the authors compared two different ANN software packages (@Brain from Talon Development Corp. and NeurShell from Ward Systems Group) to a multiple regression approach. The authors ran multiple trials for each software package to determine the optimal number of hidden layers in the model and the optimal number of nodes in each hidden layer. This process resulted in computation times for each model ranging from 30 seconds to forty-five hours. The neural network approaches made a slight improvement upon a multiple regression approach in some cases but with many caveats. Sometimes when one software package would outperform the multiple regression model, the other often did not and which package was superior depended upon the sample. They also found that the packages produced internally inconsistent results when reran with the same number of hidden layers and nodes since they assign random weights at the beginning of the training for each node. In contrast, the multiple regression models would be expected to produce identical results regardless of software given the same parameters. The problems of proper model specification, parameterization, differing results with multiple applications of the same models, and inconsistent measurement error potentially extends to many machine learning techniques.

Inconsistency in the ability of technology to improve performance of a task plays a key role in inhibiting its adoption. This points to the first theory-based set of evaluation criteria. The Technology Acceptance Model (TAM) explores how both the perceived usefulness of a technology and its perceived ease of use impact the intention to adopt the technology (Davis, 1989). The model demonstrates that factors measuring usefulness such as improved quality of work, effectiveness of outcomes, and increased productivity have a large influence on an end user's intention to adopt the technology. The model combines the importance of usefulness with ease of use which revolves around factors such as the complexity of the technology and the time taken to become skillful in its use.

A second theoretical perspective on system performance evaluation is identified in the early work by Goodhue and Thompson (1995). They developed the Task Technology Fit (TTF) model which focuses on the alignment between the technology and the task it purports to improve. Their model shows that the end user's perceived benefit derived from adoption of a technology is influenced not only by the factors present in the TAM model, but also by the users' evaluation of the fit between the task and technology. Goodhue and Thompson (1995) identify eight factors influencing the evaluation of task technology fit. Four of the factors focus on elements related to the quality, availability, accessibility, and compatibility of the data. The other four factors relate to the technology being used. These factors include ease of use, the reliability of the technology, the timeliness of the output, and the ability for the technology to improve performance. Their eight factors align with and include some elements identified by TAM and emphasize the additional impact of the appropriateness of the technology. TTF concludes that performance improvement only occurs if user's attitudes allow it to be utilized and there is also a good fit between the technology and the task it is addressing.

The often-cited adage of "garbage in, garbage out" illustrates the importance of the data being input into a machine learning system. Defining high-quality data as being fit for use by data consumers, Strong, Lee, and Wang (1997) categorize the dimensions of data quality into four areas: intrinsic, accessibility, contextual and representational. Intrinsic data quality issues arise if there is a perceived issue with the reputation of the data. This often occurs when there is more subjectivity in the data or when accuracy and believability of the data is questioned due to mismatches from multiple data sources. Contextual data quality problems focus on a lack of appropriate data. Poor relevance of the data can result from incomplete or missing data, poorly measured data, or an insufficient amount of data and a lack of timeliness in the data production. Contextual issues can also result when the aggregation and integration of different data sources produces inconsistent representations adding little or value to its adoption. Accessibility data quality issues arise from security of the data, difficulties in producing the data in a useable format and the length of time taken to process or arrange the data into a useable format. Representational data quality issues focus on the ease of interpretability of the data and whether it is presented in a concise and consistent manner. There is a wide variety of data sources and approaches when investigating the use of machine learning to perform property valuation. The ability of adopters of the machine learning approach is partially dependent upon their ability to access quality data similar that utilized in academic research, making assessment of the quality of the data an important consideration in the comparison of different machine learning techniques.

Some more recent academic work appears to echo similar themes as those identified by Worzala *et al.* (1995), and help shape a broad set of practical criteria for analysis. McCluskey *et al.* (2013) discuss the progress made in the use of artificial neural networks in real estate valuation. They conclude that the "ANNs retain a 'blackbox' architecture that limits their usefulness to practitioners in the field." They provide four simple questions to more broadly evaluate the efficacy of machine learning techniques to real estate valuation.

- How simple is the approach to explain?
- How consistent is the model structure?
- How transparent is the model output?

- How explicit is the locational element?

The first three of these questions echo the discussion by Worzala *et al.* (1995). They also mirror the characteristics identified as being important to the adoption of technology by the TAM and TTF models. The final question places a focus on aligning an important element of the valuation task with the use of machine learning. These questions are related to the practice-based evaluation criteria we identified at the beginning of this section. Users prefer that a machine learning property valuation system can produce results that are explainable, consistent, and transparent.

Finally, the importance of adopting appropriate technology has not been lost on real estate professionals. The International Association of Assessing Officers (IAAO) Standard on Automated Valuation Models (AVM) (2018) stress the importance of many similar themes. Their standards provide guidelines practitioners should follow when utilizing machine learning in developing AVMs. The standards include underlying statements of principles related to model transparency, model applicability, and establishing public trust in the results (confidence for stakeholders), which mirror the desire for consistency, simplicity, and transparency of adopted machine learning techniques discussed by McCluskey *et al.* (2013). The standards additionally include detailed principles related to the validity and transparency of the data utilized as well as statistical testing and certification of the AVM which echo elements of the TTF model. Finally, the standards emphasize the use of recognized appraisal principles in designing the AVM, effectively addressing the need to align the task to the technology including an emphasis on utilizing information related to the location of the property as highlighted by McCluskey *et al.* (2013).

These theoretical and practical criteria lay an interesting foundation for the critical analysis of how the application of machine-based learning has progressed since the work of Worzala *et al.* (1995). It is natural to ask if the literature shows a maturing and refinement of approaches that helps advance the adoption of machine learning to real estate valuation and if the lessons learned may also be applied to other applications of machine learning. Using the general principles of the TAM and TTF models as a foundation, and then incorporating the areas identified by Strong *et al.* (1997), McClusky *et al.* (2013) and the 2018 IAAO standards, a broad theoretical and practical set of property valuation system performance evaluation criteria are available for assessing each study identified in the systematic literature review. A new technique does not have to perfectly match all of the criteria to be seen as an effective solution, but the more criteria it addresses the better.

The combination of performance criteria designed for use across all information systems and criteria specific to the real estate valuation problem can be categorized into three areas related to the standard components of an information system: input, process and output. A summary of the reviewed criteria is presented in Table 1 utilizing three broad categories: (1) Data quality and ease of use; (2) Reasoning, alignment, and explication; and (3) Usefulness and consistency of results. We refer to this as the data, reasoning and usefulness (DRU) evaluation framework. The criteria provide a basis for analysis of the progress made on the development of machine learning property valuation systems.

	Input	**Process**	**Output**
Source	**Data Quality and Ease of Use**	**Reasoning, Alignment, Explication**	**Usefulness and Result Consistency**
Data Quality Strong, Lee & Wang (1997)	Intrinsic, Accessibility, Contextual, Representational		
Technology Acceptance Model (TAM) Davis (1989)	Perceived Ease of Use, Perceived Complexity of Technology	Time Taken to Become Skillful	Perceived Usefulness of Result, Improved Quality of Work, Effectiveness of Outcomes, Increased Productivity
Task-Technology Fit (TTF)	Quality, Availability, Accessibility and Compatibility of Data, Ease of Use	Technology Appropriateness, Technology Reliability	Ability to Improve Performance, Output Timeliness

Goodhue & Thompson (1995)			
McCluskey, McCord, Davis, Haran & McIlhatton (2013)	How Explicit is the Locational Element?	Consistency of Model Structure, How Simple is the Approach to Explain?	Result Transparency
International Association of Assessing Officers (IAAO) (2018)	Data Validity and Transparency, Locational Element	Recognized Appraisal Principles, Statistical Principles	Statistical Significance

Table 1. Summary of DRU Evaluation Framework Criteria

4. Systematic Literature Review Methodology

A systematic literature review identifies relevant research that incorporates related methodologies and then compares them based upon set of criteria to assess their strengths and weaknesses as well as identify areas for further exploration. The resulting analysis will provide directions for researchers in their utilization of machine learning techniques in real estate valuation as well as their use in other related problem domains.

The use of a systematic literature review to assess the use of machine learning has often been cited as providing insight into future research and application of machine learning techniques. Wen *et al.* (2012) reviewed empirical literature related to the application of machine learning for software effort estimation. Malhotra (2015) similarly reviewed the use of machine learning models in software fault prediction. Other examples have focused on the application of machine learning in the sciences. Mosavi, Ozturk and Chau (2018) reviewed the literature using machine learning to assist hydrologists in predicting floods and designing approaches to decrease loss of life and damage to property. Cabitza, Locoro and Banfi (2018) evaluated the use of machine learning in the design of orthopedics designed to address bond and muscle disorders. Finally, Strader *et al.* (2020) utilized a systematic literature review to investigate the ability of machine learning techniques to predict stock market performance.

The authors each independently conducted a search for peer reviewed academic literature utilizing machine learning in the real estate field. Searches were conducted using EconLit, EBSCO, and Google Scholar. The authors decided to include articles beginning in 2000 to the present to provide insight into how the use of machine learning has evolved during this time when applied to property valuation. Studies looking at other aspects of the real estate industry were excluded. For example, Zekić-Sušac, Has and Knežević (2021) used deep learning to identify predictors of the cost of energy consumption in public buildings. While the ability to forecast energy costs could provide an identifying characteristic utilized in real estate valuation, it is not a direct example of valuation. Articles evaluating broad market trends such as returns to real estate investment trusts (REITs) (Chen *et al.*, 2014) or house price indexes (Milunovich, 2019) are also not discussed. The final list of articles included in the study is not intended to be a comprehensive review of all articles on the subject, but instead a representative sample. Hardware, software, data storage, and networking capabilities prior to 2000 are substantially different from today so very early machine learning studies may not provide insights that are relevant to current and future work in this domain. Articles were then reviewed and broadly categorized by the prominent machine learning technique utilized.

5. Review of Studies Using Machine Learning Technologies for Real Estate Valuation

Machine learning encompasses a variety of intelligent system approaches where outcomes are predicted based on a set of input parameters. To assist in our evaluation of previously published machine learning-based property valuation studies, we organize the following review based on the primary machine learning method used in each study. We found that the most common machine learning techniques used in these studies included artificial neural networks, regression trees, and support vector machines. The final sub-section considers additional studies that utilized hybrid methods, multiple approaches, or other techniques. Each sub-section provides a brief description of the machine learning technique followed by a review of the related studies.

5.1 Artificial Neural Networks (ANNs)

Artificial neural networks (ANN) are computational models based on biological neural networks. In the network, nodes (neurons) are grouped together starting with an input layer and ending with an output layer. Signals are propagated through several layers of connected nodes as they learn based on examples and modify their solutions by adjusting parameter weights to reduce the level of prediction error. The use of neural networks is a good fit for the complexity of the real estate valuation problem and it is was one of the earliest ML techniques investigated.

Din, Hoesli and Bender (2001) used a geographic information system (GIS) to determine a quantitative geo-index capturing eight environmental measures such as the distance to the city center, level of quietness, and quality of view for residential real estate in Geneva, Switzerland. They developed four different scenarios all of which incorporate four standard quantitative variables such as number of bedrooms. The scenarios differ by the number and type of environmental variables included. For each year from 1978 to 1992 the authors use linear regression and ANN models to construct a price index and compare the model to a yearly price index. The authors only provide a general graphical representation of the results from the four scenarios for both the linear regression and ANN models. The limited size of their dataset prevented a thorough training, testing, and validation procedure in the ANN model. Alternatively, they deployed a back propagation technique to optimize correlation with target prices. The resulting best ANN models had two to four hidden nodes depending on the scenario. They conclude that in the ANN model the differences across scenarios are more pronounced and that the general shape of the two approaches is similar within each scenario. Without a more formal measure of performance, it is difficult to draw conclusions on the appropriateness of the ANN model and its performance from their work. The small dataset provides an example where the technique did not align well with the data quality. While the model includes locational components, this work illustrates many of criticisms of the early attempts to adopt ANN with limited data and limited ability to test for result robustness.

In contrast, Nguyen and Cripps (2001) find that ANN generally improves upon a multiple regression approach (MRA) when a large dataset is utilized and model specification is carefully considered. The authors describe a host of possible methodological issues which must be addressed when employing ANN. Included among these are the selection of the training sample, the selection of the validation set, the number of hidden layers, number of nodes in each hidden layer, and the possibility of overtraining. To investigate the importance of possible model misspecification they specify six different models across 18 different training sets. To test the performance of the model the authors look at the mean absolute percentage error (MAPE) and absolute percentage error (APE) or forecasting error (FE). The general conclusion of their work is that, as the model becomes more complex, the size of the training set must increase for the ANN model to outperform the MRA. Unfortunately, it may also decrease confidence in the use of ANNs by providing such a wide combination of training sets and models. The issues surrounding model specification may increase the perceived difficulty of use and prevent adoption of the technique.

The importance of model specification and sample size are supported by Peterson and Flanagan (2009) who also find that ANN model performance increases with larger training sets, while the performance of linear hedonic regression models does not. The authors also test for the importance of nonlinearities in the data negatively impacting the regression results and reject the null hypothesis that no missed nonlinearities exist. Ultimately ANN was found to significantly outperform linear hedonic regressions. This provides strong evidence that correctly specified ANN models have the potential to improve upon traditional regression approaches.

The correct specification of the model is a crucial element of the utilization of ANN for real estate valuation and in developing widely utilized techniques that can be adopted professionally. McCluskey *et al.* (2012) compared ANN to MRA, spatial simultaneous autoregressive models (SAR), and localized geographically weighted regressions (GWR) for the mass appraisal of properties in Northern Ireland. The authors focus on multiple measures of model performance expanding upon those used by Peterson and Flanagan (2009). McCluskey *et al.* (2012) attempt to find a balance among the four competing features discussed earlier. The authors explain that a tension exists between the simplicity of the approaches, the consistency of the model structures, the transparency of the output and the explicit nature of the locational elements. By looking at a wide range of performance criteria they found that the GWR method strikes the best balance of accuracy while accounting for consistency of results. This is in contrast to the ANN approach whose validity is retested each time the model is deployed. The GWR approach also provides more transparency in its results compared to ANN. The authors conclude by ranking each of the four techniques in each of the four desired features. Their final mean rankings place ANN models in last place with MRA and SAR ranked higher for simplicity and consistency and GWR ranking first for transparency and the explicitness of the locational element. In short, the ANN results do not provide any evidence of the importance of any included locational components of the ANN model. Their results highlight the difficulty of widespread adoption of MLA approaches for real estate valuation.

The lack of transparency in explaining the results and in the methodology decreases perceived ease of use and may slow adoption of ANN techniques and creates difficulty explain the reasoning behind the adoption of the technique.

In contrast, Cajias and Ertl (2018) found poor performance using a geographically weighted regression (GWR) compared to a generalized additive model (GAM) in estimating house prices in Germany. GWR expands OLS to incorporate local regressions in hopes of identifying spatially varying parameters. GAM allows covariates to take nonlinear form, consistent with the locational aspect of real estate data. They surprisingly find that GAM and OLS perform similarly and both outperform GWR. The performance differences of GWR may relate to the sample used in the two studies. McCluskey *et al*. (2012) sample was limited in size with 2694 properties in Northern Ireland. Calias and Ertl's sample encompassed 570,000 observations across 46 regions of Germany.

The importance of localized variables was confirmed by Mimis, Rovolis and Stamou (2013). They included a locational characteristic in their data for properties in Athens, Greece and found the that an ANN model outperformed an SAR model in terms of lower absolute mean errors. The authors improved upon earlier locational variables by incorporating a GIS to create variables linked to locational characteristics across a wide geographic area.

Chiarazzo *et al.* (2014) also found that the inclusion of environmental, landscaping, and proximity to industrial areas all play a key role in valuations in their dataset. These are also very localized variables and specific to the dataset and location of the properties under consideration in Italy. In contrast to the Mimis study, Chaiarazzo *et al.* (2014) includes 42 input variables and incorporates an ANN with three hidden layers with the first two containing 20 neurons each. Through a sensitivity analysis, the authors find that the most relevant features impacting property value were special features such as proximity to a beach. Their choice of a more complex and less transparent ANN model brings back concerns related to the impact of changes in specifications and their sample of only 193 records is not an appropriate fit for the MLA.

Two additional studies Yalpir (2018) and Abidoye and Chan (2017) also suffered from small sample sizes that question if an ANN approach was appropriate. Yalpir (2018) investigated three different ANN activation functions and an MRA approach and found that ANN outperformed the MRA approach. The study included an adaptive neuron activation function developed by Kahramanli and Allahverdi (2009) that had previously been used mainly in the biomedical field. The results indicate that the adaptive ANN outperformed the classical, nonadaptive, training functions in multiple measures of performance.

Abidoye and Chan (2017) found that an ANN provides reliable valuation in a developing economy using data from Nigeria. Their work used data gathered from multiple real estate firms as opposed to standardized data from a centralized database. The pooling of data created an interesting set, but might limit practitioners from attempting a similar approach. They found that the number of servants quarters was the number one determinant of the property value, which they point out is in contrast the literature from other parts of the world.

Yacim and Boshoff (2018a) attempt to improve upon model specification issues by comparing multiple training algorithms to 3,232 sales transactions in Cape Town, South Africa. When identifying the best approach, the authors build upon the work of McCluskey *et al.* (2012) and provide a ranking order of multiple measures of error, including a measure of coefficient of dispersion. The use of a ranking of multiple measures of model predication and accuracy demonstrates the need for consistency in model evaluation. Their results indicate that the back propagation training algorithm that has been widely utilized in real estate valuation research did not perform as well as the three other training algorithms explored. The other three algorithms, Conjugate Gradient (CG), Levenberg-Marguart (LM) and Powell-Beale Conjugat Gradient (PBCG) all outperformed BP. This study provides a valuable addition with their comparison of multiple ANN techniques.

5.2 Regression Trees (RTs)

A regression tree is built through an iterative process known as binary recursive partitioning that splits the data into partitions or branches and then continues splitting each partition into smaller and smaller groups. There are many different techniques for building the trees and determining the number of partitions or depth of the tree. Mullainathan and Spiess (2017) discuss the complexities associated the use of regression trees to investigate a multitude of economic problems, including an illustration of multiple techniques to real estate valuation. Following Mullainathan and Spiess (2017), a tree that results in each observation ending up as a final leaf would fit the sample data perfectly, however this tree would perform poorly in predicting prices from the out of sample data. This result illustrates the problem of overfitting which is common to regression trees. Ultimately, instead of looking for the perfect tree, the goal should be choosing the best tree conditioned upon, or regulated by, a set of criteria designed to improve the out of sample performance of the tree. Multiple iterations of the tree are then produced using partitions of the out of sample data and looking for the tree that provides the best out of sample performance for each partition. Finally, a tree is built by combining the best fit trees across the different partitions.

Central to the process summarized by Mullainathan and Spiess (2017) is the criteria utilized in improving the performance of the tree. There are a large variety of potential criteria that could impact tree performance including tree characteristics (depth of tree, number of nodes, connectivity between nodes, etc.), included independent variables, process characteristics (number of partitions, size of partitions, number of random draws, etc.), specified functional forms (minimized loss function, regression, etc.) and performance criteria (coefficient of determination, absolute percentage errors, etc.).

Mullainathan and Spiess (2017) provide the following four general classes and examples of regularization processes:

- Global /parametric predictors (Linear: subset selection, LASSO, ridge regression, Elastic net)
- Local / nonparametric predictors (decision trees, random forest, nearest neighbors, kernel regression)
- Mixed predictors (deep leaning, neural nets, convolutional neural nets, splines)
- Combined predictors (bagging: unweighted average of predictors from bootstrap draws, Boosting: linear combination of predictions of residuals, ensemble: weighted combination of different predictors)

An illustrative example is the random forest process. Random forest utilizes random choices from the possible variables to determine each split in the tree. The process of random selection occurs over a large number of sample subsets. The final result then represents an average across a large number of deep but similar trees. In contrast, gradient boosting techniques start with a shallow tree using a subset of the data. The tree is then updated based upon residuals which are used to correct for mistakes in the original tree and produce an updated tree. The updating process is repeated until no improvement is found. This requires a process for tuning the tree or selecting various attributes of the tree to optimize performance.

Antipov and Pokryshevskaya (2012) provide an early application of random forest to property valuation and a comparison of ten different tree approaches. The authors point to the benefits of random forest when dealing with common issues in property data such as its ability to deal with categorical data, missing data, and unique property characteristics. Applying two-step method to address heteroscedasticity in the data, the authors provide a broad ranking of their approaches by segmenting their results into three different buckets based on MAPE and COD. The highest performing group are techniques that produce high accuracy and low sales ratio variability. This group includes random forest, boosted trees and K nearest neighbor (KNN) (mean), with random forest performing the best. They find that chi-squared automatic interaction detection (CHAID), Exhaustive CHAID, CART and radical basis function neural network provide lower performance in both categories. Finally, they find that standard regression, KNN (median) and MLP NN provided low accuracy and high sales ratio variability.

Mullainathan and Spiess (2017) compare OLS to four different regression tree processes, a simple regression tree tuned by depth, LASSO, random forest, and an ensemble method combining LASSO and forest. The survey includes 51,808 observations from the 2011 American Housing Survey of house value with 150 independent variables including home characteristics and quality measures. The authors randomly draw a training sample of 10,000 units and evaluated performance on the remaining out of sample observations. The authors find the R^2 of the hold out sample improves from .417 for OLS to .455 for random forest and to .459 for the ensemble method. They also partition the data into quintiles and find strong relative improvement over OLS for the Ensemble in each quintile and an improvement for random forest in the lowest four quintiles. The authors provide an in-depth review of potential complications with the use of regression trees. One limitation they highlight is the tradeoff between the flexibility of the tuning process considering multiple trees and similar performance from two outcomes with divergent coefficients. This possibility stresses the importance the choice of tuning mechanism to insure it is a fit to the task of property valuation.

Kok, Koponen and Martinez-Barbosa (2017) confirm the superiority of some regression trees to OLS utilizing a large data set of multifamily properties from the commercial real estate markets in California, Texas, and Florida. The dataset is compiled from three different primary sources. The traditional approach for commercial property is to focus on the income approach utilizing the net operating income (NOI) of the property discounted at a capitalization rate. Importantly, the capitalization rate and NOI are not necessary in a regression tree methodology and the authors are able to investigate the importance of including NOI in the model. The authors compare four different sets of data. The first three sets differ in their treatment of NOI with one excluding NOI as an explanatory variable while the second includes NOI and the third includes a modeled NOI. The final set of data predicts the NOI per unit as opposed to the price. On each data set they utilize four techniques OLS, random forest, gradient boost and XG Boost. The findings indicate each of the three regression tree approaches improve upon OLS in both MdAPE and R^2. Gradient Boost and XG boost also both outperformed random forest in each data set. The inclusion of NOI is shown to be a key component with both the actual NOI and modeled NOI showed much lower absolute errors and higher R^2 in all of the AVMs. The gradient boost and XG boost methodologies showed the most improvement relative to OLS when NOI was included. Surprisingly, the Random Forest approach had a higher MdAPE and slightly lower R^2 than the OLS for the data set including NOI.

Kok, Koponen and Martinez-Barbosa (2017) also placed a large amount of emphasis on location. They emphasize the importance of location by including data on three broad types of information: local amenities such as restaurants, bars, music venues etc.; market data such as crime rates, vacancy rates, construction permits, etc.; and census data such as population, income, employment, etc. For each property the authors form catchment areas based upon estimated drive times of 3, 5, 10, 15 and 30 minutes. They then cross reference the catchment area for each of the sets of local characteristics. The methodology allows the authors to further investigate the importance of location and individual property characteristics. They find in all datasets that general property characteristics explain less than half of the dependent variable. In the NOI dataset, individual property characteristics explain only 24% of the price, with locational information linked to amenities, market data and census data explaining 76% of the price. While this is interesting, the most important variable for forecasting price is found to be NOI when it is included as an explanatory variable. When forecasting NOI, delinquency rates on single family homes is the most important.

Baldominos *et al.* (2018) compare a variety of ML approaches including multiple variations of each approach. The authors compare support vector regression (SVR), k-nearest neighbor (KNN), extremely randomized regression trees (an extension of random forest) and multi-layer perceptron (feed forward) ANN (MLFF-ANN). The paper focuses on high-cost residential properties in a single district of Madrid, Spain creating a fairly homogenous dataset which may limit the generalizability of their results to a wider array of properties. Locational variables were included in the dataset, emphasizing the location within the district. In comparing the best performing version of each approach, the authors report that the extremely randomized regression tree outperformed the others in terms of mean absolute error (MAE) with the top ten lowest MAE all occurring from regression trees. The regression trees also had a much smaller median absolute error (MedAE). A single layer neural network reported the worst MAE and MedAE, a result the authors attribute to possible overfitting. Interestingly the authors find that increasing the number of trees in the ensemble method has little to no impact on the MAE, while the use of bootstrapping has a significant negative impact. In the case of the MLP-ANN, the result is very sensitive to the structure with more layers and fewer units resulting in smaller MAE and MEdAE.

Mayer *et al.* (2019) compare six approaches to a dataset of 123,000 homes in Switzerland from 2005 to 2017. The authors compare OLS, ANN, robust regressions, mixed effects regression, random forest, and gradient boosting. The authors not only compare the six methods, but also compare two methodologies for updating the data set used as data is added over time. The first methodology adds data to the original data set and thus characterized as an extending window approach. The second methodology keeps the data used at a fixed interval deleting old data as new data is added and thus characterized as a moving window approach. The authors find little difference between the data selection methodology across all six of the approaches they utilize. Mayer et. al (2019) find that the gradient boosting technique performance was statistically better than the other methods for RMSE, MAE, MedAE. Gradient boosting also produces a larger portion of results within 10% and 20% of the actual transaction price.

Hu *et al.* (2019) also utilized a large number of methods in looking at housing rents in China and find very similar results among many of the techniques they deploy. The authors utilized six different measures of performance and find that two bagging based regression tree approaches performed slightly better than MLP-ANN, gradient boosting regression trees, KNN and support vector regression (SVR). The performance of the different ML approaches was very close with the exception of support vector regressions, which significantly underperformed the others.

Hong, Choi and Kim (2020) conclude that a standard random forest approach improves upon hedonic OLS regressions in their sample from South Korea. They found that 72% of the valuations were within 5% of the actual market price using random forest compared to only 15.5% when using OLS. The authors attribute part of the superior performance to their data set which was concentrated geographically and similar in structural characteristics.

The issue of location is often not addressed directly in the model adjustments, even though many authors discuss how location may impact their results. Pace and Hayunga (2020) provide a technique specifically targeting location. They investigate the ability of the information found in the residuals from OLS regressions in a spatial difference model to improve performance in the construction of a simple regression tree. They then compared the basic tree to forests using both boosting and bagging. They found an improvement from the boosting and bagging procedures incorporating both sets of residuals, with greater improvement from incorporating the OLS residuals than from those for the spatial difference model. By comparing the spatial difference and OLS models the work highlights the importance of location in real estate data. Interestingly, when reducing the entire sample to subsamples based on the 5,000 closest observations, the performance gains from bagging disappeared compared to the simple spatial difference tree. Decreasing the sample size allowed the spatial difference model to outperform the bagging methodology. The authors report that the results took over 7 hours of computing time, mainly due to the bagging procedure. This result calls into question if the adopters may view the time as decreasing the perceived ease of use of bagging and avoid adopting the technique.

When investigating rural housing price valuations Bogin and Shui (2020) echo potential issues related with perceived ease use shown by Pace and Hayunga (2020). The authors find that a random forest regression with boosting performs very similarly to a standard OLS regression in their data set. However, the boosting technique required an extra 24 hours of computing time. Contrary to other authors, Bogin *et. al* (2020) find that a standard random forest regression tree outperforms one enhanced by boosting. However, the standard random forest treatment required 48 hours of computing time and the random forest regression suffers from substantial overfitting, while their other approaches do not. The authors point to the uniqueness of their data set which encompasses 420,370 rural households as identified by the Uniform Appraisal Dataset. The authors hypothesize that including urban data, with more concentrated households in the training sample, might help with overfitting. They find that the urban data improves the performance of the random forest relative to the other approaches while decreasing the overfitting but that it also introduces a bias. Finally, the author state that their appraisal data allowed for over one hundred explanatory variables, this level of detail may not be available in public records accessed by most appraisers. Limiting the number of variables resulted in the performance gap between random forest and the other approaches to be much narrower when looking at the rural sample. The tradeoffs between computing time, performance, model specification, and samples illustrate the difficulty of generalizing ML performance to property valuation.

Alfaro-Navarro *et al.* (2020) developed an automated process to select the best tree method among those included in their study for each municipality studied in Spain. They utilize a standard regression tree along with trees based upon bagging, boosting and random forest algorithms. The process then identifies the best performing model for each of the 433 municipalities (out of 8,125 possible municipalities) which had at least 100 sample observations. The total study thus encompassed over 790,000 properties in 48 (out of 52) provinces. Their results show bagging and random forest slightly outperform boosting in terms of both the (MAPE) and coefficient of dispersion and all of the enhanced methods outperform a basic RF. Starting with segmenting the data by location, the paper provides an interesting approach to designing a more widely applicable model, which incorporates the importance of location. Additionally, their data focused on sales price and 33 property characteristics are collected from publicly available websites. This allows for a large consistent sample size utilizing publicly accessible data. The paper is notable as it provides insight into whether the same technique would produce the best results across a larger geographic region, while still including a key component linked to location. In contrast, most studies report a finding for a smaller region and provide little insight into the generalizability of the results.

Ho, Tang and Wong (2021) compare multiple ML techniques and conclude the choice between the techniques depends upon the goals and preferences of the end user. The authors compare RF, and gradient boosting machine (GBM) and support vector machine (SVM). Using eighteen years of data on 40,000 transactions in Hong Kong the authors demonstrate that both RF and GBM outperform SVM. In their data set the MSE, RMSE and MAPE are extremely close for RF and GBM and all three are higher for SVM. Similarly, the R^2 is almost identical for RF and GBM and lower for SVM. The authors caution that the results do not necessarily indicate that SVM performs poorly, to the contrary the argue that its performance is satisfactory and question if the extra computational time required for RF and GBM, especially in the case of larger data sets.

Rico-Juan and Taltavull de La Paz (2021) demonstrate that random forest techniques are capable of capturing nonlinearities that are missed by traditional hedonic regression models, specifically the nonlinearities associated with location and time. However, RF does not quantify the nonlinear relationships. They also show that quantile regression can capture and quantify the impact of a particular characteristic on the price quantile.

Deppner and Cajais (2022) take a different approach to formally address the locational issues discussed by Bogin and Shui (2020) and Pace and Hayunga (2020). The authors allow for spatial partitioning during the training of the regression tree and compare spatial and non-spatial cross validation techniques. This effectively groups training data by geographic position, placing an emphasis on the location of the property relative to others in the sample. However, they find the effectiveness of spatial partitioning depends upon the data set and goals for the predictions to be developed from its use. Their findings show that when the geographic area is very condensed, random partitioning may provide useful valuations. When the geographic area is larger with more variance in spatial dependence and density, spatial cross validation should be the preferred choice. Their work demonstrates that spatial cross validation may underestimate the predictive accuracy of the valuation, however, they argue this is preferred to an assessment of the predictive accuracy that overstates the accuracy of the valuation.

5.3 Support Vector Machines (SVMs)

SVMs use supervised learning where training examples are included in one of two categories. An SVM model represents the examples as points in a space with the goal of making the gap between the categories as wide as possible. New examples are classified based on the category in which they most likely belong.

Wang *et al.* (2014) apply an SVM in which the parameters are determined by particle swarm optimization (PSO). The authors point to SVM as a more appropriate technique than ANN which can result in a local minimum as opposed to a global solution. The authors find that the PSO-SVM produces both a lower relative error and lower MAPE compared to a standard SVM approach. They also compare both SVM models to a back propagation neural network and find both PSO-SVM and SVM produce smaller errors than the BP neural network. While an interesting result, the authors forecast the average selling price for Chongqing, China as opposed to individual property prices. Their work does shed some light on improvements in property valuation as the focus of their method is its application to small samples, which could occur in the real estate market in many locations.

5.4 Hybrid Models, Multiple Approaches, and Additional Machine Learning Techniques

Hybrid models combine two or more machine learning techniques in an attempt to improve upon single technique performance. For example, ANN requires specification of their respective optimization process and some ensemble approaches incorporate other ML techniques in the optimization process. Included in the review of this section are specialized regressions used for training ANNs, even though the specialized regression is not technically a machine learning technique.

Kontrimas and Verikas (2011) investigate property values in the Lithuanian Republic. The authors compare results from a self-organizing map (SOM) to committees of predictors derived from OLS, SVM, and MLP-ANN approaches. The authors specify two sets of committee weights based on two different sets of locational zones, one defined by experts from the local property register center and a second approach base on the zones identified by the SOM. The authors note that the two sets were much different in how they segmented the data, but that the properties were very similar across all locations making the locational component less important. Both committee approaches showed much better outcomes than OLS, MLP or SVM on its own. The authors also looked at the number of unacceptable valuations where the predicted value differs by more than 20% from the actual sale price. Both committee weights resulted in only one unacceptable valuation respectively while the other methods resulted in 18 (SVM), 31 (OLS), and 42 (MLP) unacceptable values.

Perez-Rave, Correa-Morales and Gonzalez-Echavarria (2019) utilize a ML approach to selecting variables for two different large sets of data, one in Columbia and one in the United States. The approach of incremental sample with resampling (MINREM) designed to identify the most important variable. The authors then combine the MINREM approach with a traditional hedonic regression based upon the variables selected.

Multiple authors employ genetic algorithms to enhance other ML approaches. A genetic algorithm is an example of an evolutionary algorithm (Holland, 1992). The evolutionary process begins with a set of randomly generated problem solutions. In each iterative generation, the fitness of each solution is measured by an objective function. The solutions with higher fitness are retained (survival of the fittest) and these parent solutions are combined with other high fitness solutions to create a new generation of child solutions that retain some of the characteristics from both original solutions. The evolutionary process ends when a certain number of generations has been created or a satisfactory fitness level has been reached. Based upon the systematic literature review, it appears that GA use for real estate valuation is much less common than ANN-based valuation approaches. However, Del Guidice, De Paola and Forte (2017) present one of the few applications of GA to predict rental prices for a small sample of 64 properties located in two neighborhoods in Naples, Italy. The authors find that the GA outperform a multiple regression approach in accuracy.

Ahn *et al.* (2012) proposes the use of a ridge regression coupled with a genetic algorithm (GA-ridge) when there is no clear preference between standard MLR or ANN. Their work focuses on housing and rental indexes as opposed to individual property valuation removing any link to a localized market. The purpose of the ridge regression is to address multicollinearity among the independent variables. The GA-ridge methodology simultaneously solves for the ridge regression solution through the use GA, effectively using the ridge regression as the fitness function. Their results show that the GA-ridge outperforms MLR, ANN, and standard ridge regressions based on RMSE, MAE, and MAPE. Furthermore, they report paired t-tests of each of the reported error terms demonstrating a high level of significance in the improvement of GA ridge over the other methods.

Yacim and Boshoff (2018b) utilized particle swarm optimization (PSO) and a genetic algorithm (GA) to enhance the training of an ANN and compared the results to using back propagation neural networks BPNN as well as GA with back propagation and OLS. The purpose of the enhanced training methods is to avoid the possibility of local optimums resulting in a subpar weighting. They find an improvement in model performance utilizing both PSO and GA to train the ANN. Their data includes a neighborhood code to account for locational differences. They ultimately include 18 property characteristics from 46 possibilities. The variables that were eliminated likely exhibited multicollinearity. Their results show the hybrid models and linear regression models outperform the standard BP-ANN overall. One of the benefits of BPNN, GA, and PSO training functions is the ability to investigate the relative importance of the property attributes found from the data. The authors work provides an increase in transparency with a detailed description of methodology used and rationale for choice of methods. Sun (2019) also found that use of GA to aid in the structure of the ANN produces superior results compared to BPNN.

Pai and Wang (2020) utilize GA to select parameters for four ML models, least squares support vector regression (LSSVR), classification and regression trees (CART), general regression neural networks (GRNN) and BPNN. The authors used the average absolute percentage error in the objective function of GA to determine the specific parameters in each of the four ML approaches. They demonstrated that the parameter selection utilizing the GA procedure improved the performance of each of the four ML techniques compared to their application without GA. They also demonstrate that their LSSVR approach was superior to the other three approaches both with and without the aid of the GA technique. The authors then compare the MAPE in their studies to those of other authors. From this they conclude that the LSSVR is superior to the methods used by the other authors, with the caveat that valuation differs by many non- quantitative factors that vary by location and consumer preferences.

6. Discussion

We can make several general observations after reviewing the literature related to the different ML approaches. The review of literature related to artificial neural networks shows the need for large datasets, appropriate training samples, and very careful model specification that incorporates unique local variables associated with valuation of specific property types in specific geographic locations. A downside across all of these studies is that there does not appear to be a generalizable solution. Each study addressed property valuation for one scenario and did not work to identify generalizable techniques or problem features. The lack of transparency and perceived lack of ease of use did not improve over the twenty years of literature reviewed. Also, ANN was often applied to data sets which lack contextual quality due to their limited size. In the literature applying ANN as the primary tool, five of six studies comparing ANN to regression approaches found ANN methods outperformed the regression approach utilized.

When ANN was compared to other ML techniques, ANN was often found to be inferior to the other techniques. McCluskey *et al.* (2012) found GWR outperformed ANN when considering four different categories linked to model performance related to data, reasoning and usefulness (DRU). Even if the issues related to DRU were solved, standard neural networks were often shown to produce inferior valuations compared to other ML techniques. In five articles, a form of regression tree was shown to outperform an ANN approach (Antipov and Pokryshevskaya (2012), Ahn (2012), Baldominos, *et al*. (2018), Mayer *et al.* (2019), and Hu *et al. (*2019)). Additionally, Wang *et al.* (2014) found PSO-SVM outperformed BPANN. Multiple articles demonstrated ANN could be improved by incorporating other ML techniques to enhance ANN (Kontrimas and Verikas (2011), Wang *et al.* (2014), Yacim and Boshoff (2018b), and Sun 2019). Those papers unfortunately did not compare the enhanced neural network to regression trees, but mainly demonstrated the increased performance of the ensemble ANN to more traditional ANN approaches.

Advanced regression trees (random forest with boosting and bagging) were shown to be generally superior to standard hedonic regressions, KNN, SVR and various ANN models. Additionally, Mullainathan and Spiess (2017) found that an ensemble combination of LASSO and forest outperformed either independently and OLS. However, the ranking among different regression tree techniques is less clear. Multiple papers found that either boosting or bagging techniques outperformed more standard random forest trees but the ranking between boosting, bagging and standard RF was not consistent. Kok *et al.* (2017) and Mayer *et al.* (2019) both found boosting techniques to dominate the others in their respective studies. However, Alfro-Navarro *et al.* (2020) found bagging and standard random forest to slightly outperform boosting. Hu *et al.* (2019) found RF outperformed boosting and the extra tree regressions outperformed random forest in one sample and vice versa in a different sample. Overall, the ability of more advanced regression trees to universally outperform ANN, KNN and various more traditional regressions, makes regression trees an area of focus for future research. Similar to regression trees, SVM and the ensemble SVM approaches were shown to outperform ANN approaches when compared directly. The downside of the regression trees is the amount of computational time and effort which may decrease adoption of the techniques.

The comparisons of the various ML approaches identified some common themes when considered in the context of the DRU evaluation framework presented in Table 1. An ideal ML approach would lie at an intersection which encompasses the desired criteria of all three elements. In other words, the approach would be generalizable to a wide range of scenarios and data sources, account for the location of the property, have low barriers surrounding both its perceived ease of use and perceived ease of adoption and finally produce performance improvement justifying any computation costs associated with its adoption. This is admittedly a high bar. The literature review does however allow for some general conclusions to be drawn from the DRU framework that may help shape future research.

The data component of DRU focused on the quality of the data (including the locational element) and the perceived ease of use of the technology. Regardless of the ML approach utilized, the vast majority of studies attempted to account for location of the property. In many cases the dataset utilized was implicitly accomplishing this task by limiting the valuations to a localized data set. However, there were attempts to identified the influence of location in more diverse geographic data sets. The lager issue which persists across the twenty years relates to the perceived ease of use of the technology and its alignment with the chosen data. This was especially apparent with the advanced regression models and hybrid approaches. Many of the same concerns that were expressed with early attempts to adopt ANN resurfaced with the advanced regression and hybrid approaches. The concerns questioning ease of use largely disappeared in the case of basic ANN approaches. Future research and advances in computing power may similarly be able to lessen the impact of perceived costs associated with ease of use related with the advanced RT and hybrid approaches.

The criteria summarized in the DRU framework highlight the need to consider if the benefit derived from a technique justifies the additional computing power and technical expertise required for its implementation. Weighing the impact of these tradeoff is an important component that leads to adoption in both the TAM and TTF frameworks. For studies using multi-technique approaches (two or more machine learning techniques combined into a hybrid system), these systems may marginally improve performance, but at some point, they become so complex that they are not practically usable. While the studies reviewed reported improvements in performance, they generally failed to investigate if the small increase in aggregate performance would translate into economic performance for the end users. The technique needs to be understood at a level that allows the end user to have confidence in the perceived benefit of adopting the technique. Extending the results to demonstrate the economic benefit received by end users from small gains in average performance would help decrease questions associated with the effectiveness of the outcomes and provide motivation to adopt the more complex models. Additionally, the articles reviewed often blur the line between a market value as defined by the professional appraisal standards and the output of the MP technique as discussed by Epley (2017).

Finally, the review does provide some additional directions for future research. It is obvious that future work will continue to combine multiple machine learning methods to further improve property valuation performance while ideally offering improvements that simplify the adoption of the technique. The goal should be to identify solutions that are generalizable to a wider range of property valuation scenarios. Work could also extend to related real estate management decisions. Future studies could address the problem of property portfolio management. Which machine learning techniques would best be able to identify an optimal set of properties that would produce the best return with lower risk? This is a very common financial market question, but has not been addressed to the same extent in the real estate market. Genetic algorithms may be a good fit with this task because it can represent various sets of properties and then identify the characteristics that produce the best final solution.

Another question is which machine learning techniques are best suited to identify the price direction (up or down) for a real estate market. Researchers could consider SVMs as a potential machine learning technology that may best match this task. Another related item to note is that human real estate managers will continue to have an important place in the market as they use the results from the machine learning systems as a decision support tool. The best results will come from a scenario where the relative advantages of the system and the real estate manager are combined. Studies should consider machine learning as a decision support tool and focus on identifying the relative advantages of technology and people when valuing properties. This is especially important in identifying the relative importance of specific property characteristics as an input in the valuation process. One of the main sources of nonlinearity is the dynamic and evolving nature of the property characteristics and how well they match the current demands of purchasers. Much of the real estate professional's value comes in spotting those trends in real time, a difficult task for any quantitative approach needing a large number of data points. An interesting question is whether ML can identify societal trends which may aid the real estate professional in real time adjustments to market prices.

A systematic literature review methodology has some limitations. One of the primary limitations is that machine learning applications that were not successful are rarely published so there is a bias toward techniques that showed some improvement in the unique real estate market that was being studied. Even with that bias we fill the literature review produced useful insights to be considered in future adoptions of ML techniques to property valuation.

7. Conclusions

This study identified and reviewed published studies from the two decades (starting in 2000) where machine learning was used for property valuation. The purpose was to identify which machine learning techniques have been used, develop theoretical and practical criteria that can be used to provide a broad assessment of system performance, identify common findings across the studies, and conclude with a set of theoretical and practical implications and future research directions.

The primary overall conclusion that arises from this review is that machine learning can improve the property valuation process and price identification accuracy. There is no one perfect machine learning technique in this domain, but continued work can further refine the methods used and ultimately produce solutions that meet most, if not all, of the system evaluation criteria identified in this study. From a research standpoint, it is apparent that both TAM and TTF provide a theoretical basis for assessing machine learning use for property valuation. Effective property valuation systems must be perceived as useful (with the technology matching the task), and also perceived as easy to use. From a practice perspective, future property valuation systems must focus on computational simplicity and utilization of easily accessible datasets.

8. References

Abidoye, R. B., & Chan, A. P. (2017). Modelling property values in Nigeria using artificial neural network. *Journal of Property Research*, *34*(1), 36-53. https://doi.org/10.1080/09599916.2017.1286366

Ahn, J. J., Byun, H. W., Oh, K. J., & Kim, T. Y. (2012). Using ridge regression with genetic algorithm to enhance real estate appraisal forecasting. *Expert Systems with Applications*, *39*(9), 8369-8379. https://doi.org/10.1016/j.eswa.2012.01.183

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 modelling complex real estate systems. *Complexity*, *2020*. https://doi.org/10.1155/2020/5287263

Antipov, E. and Pokryshevskaya, E. (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, 1772-1887. https://doi.org/10.1016/j.eswa.2011.08.077

Baldominos, A., Blanco, I., Moreno, A. J., Iturrarte, R., Bernárdez, Ó., & Afonso, C. (2018). Identifying real estate opportunities using machine learning. *Applied Sciences*, *8*(11), 2321. https://doi.org/10.3390/app8112321

Bogin, A. & Shui, J. (2020). Appraisal accuracy and automated valuation models in rural areas. *Journal of Real Estate Finance and Economics,* 60, 40-52. https://doi.org/10.1007/s11146-019-09712-0

Cabitza, F., Locoro, A., & Banfi, G. (2018). Machine learning in orthopedics: A literature review. *Frontiers in Bioengineering and Biotechnology*, *6*, 75. https://doi.org/10.3389/fbioe.2018.00075

Cajias, M. and Ertl, S (2018). Spatial effects and non-linearity in hedonic modeling: Will large data sets change our assumptions? *Journal of Property Investment & Finance*, 36(1), 32-49. https://doi.org/10.1108/JPIF-10-2016-0080

Chen, J. H., Chang, T. T., Ho, C. R., & Diaz, J. F. (2014). Grey relational analysis and neural network forecasting of REIT returns. *Quantitative Finance*, *14*(11), 2033-2044. https://doi.org/10.1080/14697688.2013.816765

Chiarazzo, V., Caggiani, L., Marinelli, M., & Ottomanelli, M. (2014). A neural network based model for real estate price estimation considering environmental quality of property location. *Transportation Research Procedia*, *3*, 810-817. https://doi.org/10.1016/j.trpro.2014.10.067

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 319-340. https://doi.org/10.2307/249008

Del Giudice, V., De Paola, P., & Forte, F. (2017). Genetic algorithms for real estate market analysis. In *Computational Science and Its Applications–ICCSA 2017: 17th International Conference, Trieste, Italy, July 3-6, 2017, Proceedings, Part III 17* (pp. 467-477). Springer International Publishing.

Deppner, J. & Cajias, M. (2022). Accounting for spatial autocorrelation in algorithm-driven hedonic models: A spatial cross validation approach. *The Journal of Real Estate Finance and Economics*, 1-39. https://doi.org/10.1007/s11146-022-09915-y

Din, A., Hoesli, M., & Bender, A. (2001). Environmental variables and real estate prices. *Urban studies*, *38*(11), 1989-2000. https://doi.org/10.1080%2F00420980120080899

Epley E. (2017). The Need to Reference Automatic Valuation Models to the Valuation Process. *Journal of Real Estate Literature*, *25(1).* 237-252. https://www.jstor.org/stable/26391908

Furneaux, B. (2012). Task-technology fit theory: A survey and synopsis of the literature. *Information Systems Theory: Explaining and Predicting Our Digital Society, Vol. 1*, 87-106.

Goodhue, D. L., & Thompson, R. L. (1995). Task-technology fit and individual performance. *MIS Quarterly*, *19*(2), 213-236. https://doi.org/10.2307/249689

Ho, W., Tang, B. &Wong, S. (2021). Predicting property prices with machine learning algorithms. *Journal of Property Research*, 38(1), 48-70. https://doi.org/10.1080/09599916.2020.1832558

Holland, J. H. (1992). Genetic algorithms. *Scientific American*, *267*(1), 66-73.

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.2020.11455

Hu, L., He, S., Han, Z., Xiao, H., Su, S., Weng, M., Cia, Z., (2019) Monitoring housing rental prices based on social media: An integrated approach of machine learning algorithms and hedonic modeling to inform equitable house policies. *Land Use Policy*, 82, 657-673. https://doi.org/10.1016/j.landusepol.2018.12.030

International Association of Assessing Officers. (2018). *Standard on automated valuation models (AVMs).* https://www.iaao.org/media/standards/Standard_on_Automated_Valuation_Models.pdf

Kahramanli, H., & Allahverdi, N. (2009). Rule extraction from trained adaptive neural networks using artificial immune systems. *Expert Systems with Applications*, *36*(2), 1513-1522. https://doi.org/10.1016/j.eswa.2007.11.024

Kok, N., Koponen, E. L., & Martínez-Barbosa, C. A. (2017). Big data in real estate? From manual appraisal to automated valuation. *The Journal of Portfolio Management*, *43*(6), 202-211. https://doi.org/10.3905/jpm.2017.43.6.202

Kontrimas, V., & Verikas, A. (2011). The mass appraisal of the real estate by computational intelligence. *Applied Soft Computing, 11*(1), 443-448. https://doi.org/10.1016/j.asoc.2009.12.003

Malhotra, R. (2015). A systematic review of machine learning techniques for software fault prediction. *Applied Soft Computing, 27,* 504-518. https://doi.org/10.1016/j.asoc.2014.11.023

Marangunić, N., & Granić, A. (2015). Technology acceptance model: a literature review from 1986 to 2013. *Universal Access in Information Society, 14,* 81–95. https://doi.org/10.1007/s10209-014-0348-1

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

McCluskey, W.J., McCord, M., Davis, P. T., Haran, M., & McIlhatton, D. (2013). Prediction accuracy in mass appraisal: a comparison of modern approaches. *Journal of Property Research*, *30*(4), 239-265. https://doi.org/10.1080/09599916.2013.781204

McCluskey, W., Davis, P., Haran, M., McCord, M., & McIlhatton, D. (2012). The potential of artificial neural networks in mass appraisal: the case revisited. *Journal of Financial Management of Property and Construction*, *17*(3), 274-292. https://doi.org/10.1108/13664381211274371

Milunvoich, G. (2019). Forecasting Australia's real house price index: A comparison of time series and machine learning methods. *Journal of Forecasting*, 39, 1098-1118. https://doi.org/10.1002/for.2678

Mimis, A., Rovolis, A., & Stamou, M. (2013). Property valuation with artificial neural network: The case of Athens. *Journal of Property Research*, *30*(2), 128-143. https://doi.org/10.1080/09599916.2012.755558

Mosavi, A., Ozturk, P., & Chau, K. W. (2018). Flood prediction using machine learning models: Literature review. *Water*, *10*(11), 1536. https://doi.org/10.3390/w10111536

Mullainathan, S. and Spiess J. (2017). Machine learning: An applied econometric approach. *Journal of Economic Perspectives, 31*(2), 87-106. https://doi.org/10.1257/jep.31.2.87

Nguyen, N., & Cripps, A. (2001). Predicting housing value: A comparison of multiple regression analysis and artificial neural networks. *Journal of Real Estate Research*, *22*(3), 313-336.

Pace, R. K., & Hayunga, D. (2020). Examining the information content of residuals from hedonic and spatial models using forests and trees. *Journal of Real Estate Finance & Economics, 60*, 170-180. https://doi.org/10.1007/s11146-019-09724-w

Pai, P. F., & Wang, W. C. (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

Pagourtzi, E., Assimakopoulos, V., Hatzichristos, T., & French, N. (2003). Real estate appraisal: A review of valuation methods. *Journal of Property Investment & Finance*, *21*(4), 383-401. https://doi.org/10.1108/14635780310483656

Peterson, S., & Flanagan, A. (2009). Neural network hedonic pricing models in mass real estate appraisal. *Journal of Real Estate Research, 31*(2), 147-164.

Pérez-Rave, J., Correa-Morales, J. C., & González-Echavarría, F. (2019). A machine learning approach to big data regression analysis of real estate prices for inferential and predictive purposes. *Journal of Property Research*, 36(1), 59–96. https://doi.org/10.1080/09599916.2019.1587489

Rico-Juan, J. & Taltavull de Laz Pav, 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,* 1-14. https://doi.org/10.1016/j.eswa.2021.114590

Riddel, M. (1999). Fundamentals, feedback trading, and housing market speculation: Evidence from California. *Journal of Housing Economics, 8*, 272-284

Rosen, S. (1974). Hedonic prices and implicit markets: Product differentiation in pure competition. *Journal of Political Economy***,** 82(1), 34-55. https://doi.org/10.1086/260169.

Spies, R., Grobbelaar, S., & Botha, A. (2020). A scoping review of the application of the task-technology fit theory. In *Responsible Design, Implementation and Use of Information and Communication Technology: 19th IFIP WG 6.11 Conference on e-Business, e-Services, and e-Society, I3E 2020, Skukuza, South Africa, April 6–8, 2020, Proceedings, Part I 19* (pp. 397-408). Springer International Publishing.

Strader, T. J., Rozycki, J. J., Root, T. H., & Huang, Y. H. J. (2020). Machine learning stock market prediction studies: Review and research directions. *Journal of International Technology and Information Management*, 28(4), 63-83.

Strong, D. M., Lee, Y. W., & Wang, R. Y. (1997). Data quality in context. *Communications of the ACM*, *40*(5), 103-110.

Sun, Y. (2019). Real estate evaluation model based on genetic algorithm optimized neural network. *Data Science Journal*, *18*(1), 36. http://doi.org/10.5334/dsj-2019-036

Wang, X., Wen, J., Zhang, Y., & Wang, Y. (2014). Real estate price forecasting based on SVM optimized by PSO. *Optik*, *125*(3), 1439-1443. https://doi.org/10.1016/j.ijleo.2013.09.017

Wen, J., Li, S., Lin, Z., Hu, Y., & Huang, C. (2012). Systematic literature review of machine learning based software development effort estimation models. *Information and Software Technology*, 54(1), 41-59. https://doi.org/10.1016/j.infsof.2011.09.002

Worzala, E., Lenk, M., & Silva, A. (1995). An exploration of neural networks and its application to real estate valuation. *Journal of Real Estate Research*, *10*(2), 185-201. https://www.jstor.org/stable/24881837

Yacim, J. A., & Boshoff, D. G. B. (2018a). Impact of artificial neural networks training algorithms on accurate prediction of property values. *Journal of Real Estate Research*, *40*(3), 375-418. https://doi.org/10.1080/10835547.2018.12091505

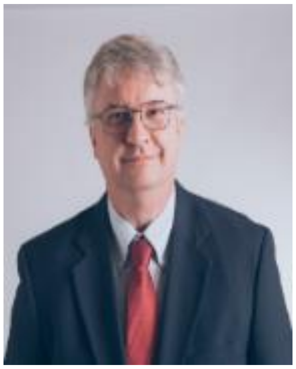
Yacim, J. A., & Boshoff, D. G. B. (2018b). Combining BP with PSO algorithms in weights optimisation and ANNs training for mass appraisal of properties. *International Journal of Housing Markets and Analysis*, *11*(2), 290-314. https://doi.org/10.1108/IJHMA-02-2017-0021

Yalpır, Ş. (2018). Enhancement of parcel valuation with adaptive artificial neural network modeling. *Artificial Intelligence Review*, *49*(3), 393-405. https://doi.org/10.1007/s10462-016-9531-5

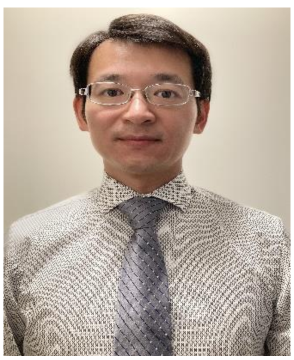
Zekić-Sušac, M., Has, A., & Knežević, M. (2021). Predicting energy cost of public buildings by artificial neural networks, CART, and random forest. *Neurocomputing*, *439*, 223-233. https://doi.org/10.1016/j.neucom.2020.01.124

Author Biographies

Thomas Root holds four degrees including a Ph.D. in Economics from the University of Kansas. He has been teaching finance at Drake University since 1999. He has received the Graduate Teaching Award and the College of Business Research Award. He published in many academic journals such as the *Quarterly Journal of Accounting and Finance; Journal of Business and Behavioral Sciences; International Business: Research, Teaching and Practice; International Research Journal of Applied Finance; Journal of the Academy of Business Education; Energy Economics; Journal of Financial Education*; and *International Review of Financial Analysis*.



Troy J. Strader is the Aliber Distinguished Professor of Information Systems and Research Fellow in the Zimpleman College of Business at Drake University. He received his Ph.D. in Business Administration (Information Systems) from the University of Illinois at Urbana Champaign in 1997. Dr. Strader has edited three books and his research has been published in *Decision Support Systems, Electronic Markets, Journal of Information Technology, Journal of the Association for Information Systems, International Journal of Electronic Commerce, Communications of the ACM, Electronic Commerce Research, European Journal of Information Systems* and other academic and practice publications. He is the founding Editor-in-Chief for the Drake Management Review.



Dr. Yu-Hsiang (John) Huang is an Assistant Professor of Management Information Systems at Georgia College & State University. Dr. Huang received his Ph.D. in Business Administration (Information Systems) from the University of Illinois at Urbana-Champaign. His research interests include IT Strategy & IT Governance, Financial Technology, and Risk Management. His work has been published in several scholarly journals such as *Information & Management, Information Technology and Management, Information Resources Management Journal,* and *Technovation.*