Landscape and Urban Planning 226 (2022) 104486

Contents lists available at ScienceDirect

Landscape and Urban Planning

journal homepage: www.elsevier.com/locate/landurbplan

Assessing the value of user-generated images of urban surroundings for house price estimation

Meixu Chen [a,b,1], Yunzhe Liu [a,c,*,2], Dani Arribas-Bel [a,3], Alex Singleton [a,4]

[a] Geographic Data Science Lab, Department of Geography and Planning, University of Liverpool, Liverpool, UK
[b] Department of Geography, University College London, London, UK
[c] Informal Cities, Oxford Martin School, University of Oxford, Oxford, UK

HIGHLIGHTS

- The contribution of user-generated images to house price estimation is minor.
- User-generated images may supplement house price estimation to capture perception.
- Flickr images are employed for perceived scene features using image recognition.
- Important housing and image features are identified and visualised.
- Random forest exceeds hedonic price model in performance and interpretability.

ARTICLE INFO

Keywords:
User-generated images
Machine learning
House price estimation
Perception
Geographic data science
Urban planning

ABSTRACT

Determinants of housing prices are particularly significant for monitoring and understanding housing prices. Traditional variables are measured through official statistics or questionnaire surveys, which are labour intensive and time-consuming. New forms of data, such as point of interest or street view imagery, have been used to extract housing location and neighbourhood features, but they cannot capture how different individuals recognised and evaluated the properties nearby, which may also be relevant in the house price estimation. Therefore, this study investigates whether user-generated images may be used to monitor and understand housing prices and how they influence real estate values. Within this context, perceived scenes features are extracted and quantified to blend with commonly used determinants of housing prices. Two machine learning algorithms, random forest and gradient boosting machines, are utilised and deployed for integration with a typical housing price modelling-hedonic price model. By comparing the performance and interpretability of different models, the relative importance of features and how they influence the estimation power of the models is visualised and analysed. The findings suggest that random forest predictions perform the best and are interpretable, with geotagged Flickr images adding 4.6% to the model's accuracy (R^2) from 61.9% to 66.5%. Although user-generated images increase minor value in house price estimation, they may be used as a supplementary data source to capture perception features for house price estimation. This could help the restructuring and optimisation of residential areas in future regional construction, planning and development.

1. Introduction

Housing affordability has become a social issue across the globe (UN-Habitat, 2020), which has brought substantial challenges to various dimensions of the social fabric, including, but not limited to, urban and social inequality (Liu et al., 2020), mobility (Causa & Pichelmann, 2020)

* Corresponding author at: Geographic Data Science Lab, Department of Geography and Planning, University of Liverpool, Liverpool, UK.
E-mail addresses: meixu@liverpool.ac.uk (M. Chen), yunzhe.liu@wrh.ox.ac.uk (Y. Liu), darribas@liverpool.ac.uk (D. Arribas-Bel), alex.singleton@liverpool.ac.uk (A. Singleton).
[1] https://orcid.org/0000-0003-2712-5551.
[2] https://orcid.org/0000-0002-7189-3323.
[3] https://orcid.org/0000-0002-6274-1619.
[4] https://orcid.org/0000-0002-2338-2334.

https://doi.org/10.1016/j.landurbplan.2022.104486
Received 12 November 2021; Received in revised form 10 May 2022; Accepted 21 May 2022
Available online 27 May 2022

and health and wellbeing (Chung et al., 2019). As a result, government agendas have prioritised monitoring and analysing housing prices to provide suggestions for house purchasers, researchers, and policymakers in urban planning and real estate, thereby building a healthy housing market (Hu et al., 2019). However, the determinants of housing prices are complicated and multidimensional; others remain uncertain (Nistor & Reianu, 2018; Wang et al., 2017; Zhang et al., 2012). The complexity of housing prices hinders policymakers from formulating and delivering equitable and sustainable planning for housing market development and urban environment improvement (Yao et al., 2018).

Determinants of housing prices can be understood from two scales. First, at the macroscale, economic bases generally influence housing prices (Wang et al., 2017), such as population, household income, and building costs at the administrative or city level (Baker et al., 2016; Cai & Lu, 2015). Second, on the microscale, surrounding environmental and social characteristics become dominant factors since urban residents are affected by common macroeconomic variables (Hu et al., 2019). Many studies have investigated the determinants of housing prices within intraurban settings, summarised into three types: structural features, location features, and neighbourhood features (Xiao et al., 2017). Structural features refer to the features related to the property itself, such as the type of property or the age of the house (Law et al., 2019). Location features reflect characteristics of the property's geographic location, such as the distance to the city centre or retail centre (Zhang & Dong, 2018). Neighbourhood features can be viewed as the availability and accessibility of several critical urban amenities or landscapes, such as landscape features, urban greenery, and health care services (Jim & Chen, 2006).

Traditionally, these influential characteristics of housing prices are obtained primarily through official statistics, proprietary listings and questionnaire surveys (Granziera & Kozicki, 2015), which are costly, labour intensive, time-consuming, and may not be readily available to the general public. With the increased availability of emerging new data and the rapid advancements in computational power, extracting valuable features from multidimensional data has become economically affordable and technically feasible (Singleton et al., 2017). Numerous studies have extracted features, particularly neighbourhood features, to analyse housing prices through new forms of data such as points of interest (POIs) collected from social media platforms or mobile sensors (Hu et al., 2019; Wu et al., 2016; Yao et al., 2018).

Emerging studies have centred on extracting perceptions to understand the driving factors of housing or rental prices, primarily through street-view imagery (Han & Lee, 2018; Hughes et al., 2016; Law et al., 2020; Yang et al., 2021). Perceptions extracted from street view images may be considered an objective reflection of the urban physical environment. However, these perceptions cannot imply what people care about or the interaction between humans and the built environment in the real world. Unlike street view imagery, user-generated geotagged photographs (i.e., Flickr imagery in this study) can directly capture people's perception of the built environment (Chen et al., 2020). These photographs are collected, derived, and shared by different and numerous individuals, reflecting their preferences on the built environment at a particular time and, in aggregate, suggesting how the city is collectively perceived. By extracting characteristics from geotagged photographs taken surrounding properties, more appealing human-perceived scenes may be identified, which we hypothesise are likely to influence property values more than other factors nearby. As such, user-generated geotagged photographs are considered to capture such potential impacts of perceived scenes representing distinctive place characteristics of the city evaluated and identified by different individuals (Zhou et al., 2014). These are relevant driving factors in housing market valuation and may help stakeholders in multiple fields. For instance, such images can be meaningful for local authorities to understand what scenes make a city more attractive (Chen et al., 2020), for government and urban planners to facilitate the revitalisation of residential neighbourhoods (Lu, 2018), or potentially as sources of inspiration for urban designers (Barkham et al., 2018).

Within this context, this study aims to explore whether user-generated geotagged images can be taken into consideration in terms of housing price estimations and how they influence real estate values. To achieve this, a Places365 convolutional neural network (CNN) trained by Zhou et al. (2018) is configured to quantify and identify perceived scenes from geotagged Flickr photographs. This model allows us to blend these perceived scenes with standard housing price features to explore whether geotagged Flickr images impact housing prices. In addition to the traditional hedonic price model (HPMs), two interpretable machine learning algorithms, random forest (RF) and gradient boosting machines (GBM), are also trained and deployed to be compared in performance and interpretation. The novelty of this research is depicted in two main aspects. On the one hand, user-generated geotagged images have feasible prospects in house price estimation. This aspect remains largely unexplored in the literature that extracts neighbourhood features of housing prices from physical POIs data rather than actual perceptions from individuals. On the other hand, RF and GBM are integrated with traditional and widely used HPMs to identify features that affect housing prices. By comparing the performance and interpretability of different models, the relative importance of features and how they influence the estimation power of the models are quantified and identified.

This research could provide helpful references for stakeholders, i.e., an additional data source (i.e., geotagged Flickr) for real estate assessors, a site selection strategy for real estate developers and better adjustment and optimisation of housing policy for the government. The remainder of the study is structured as follows. Section 2 reviews models related to housing price estimation and the use of images employed in the urban perception field. The following section describes multiple datasets collected and processed to obtain housing structural, location and perceived scene features for the estimation models. Section 4 presents the overall method framework used and each technique and procedure implemented in this study. The experimental results and discussion are reported in Section 5. Finally, Section 6 concludes the scientific and practical contributions, limitations and further works of this study.

2. Literature review

2.1. House prices models

A typical and frequently used theoretical model to identify and analyse features that influence housing value is the HPM, which has been used in various studies (Chen & Jim, 2010; Zhang & Dong, 2018; Hamilton & Morgan, 2010; Wen & Tao, 2015). This model measures how each of the potential features affects housing prices, playing a role in uncovering the intrinsic value of a single attribute based on the estimation of the marginal changes in observed prices (Rosen, 1974). For instance, Hamilton and Morgan (2010) integrated LiDAR data and geographic information science into an HPM to estimate the desire to purchase houses with beach access and view. Wen and Tao (2015) employed an HPM to examine the polycentric urban structure in determining housing prices. However, the HPMs have been criticised for their strong assumptions on the linear relation between features and prices and their inability to handle spatial heterogeneity (Anglin & Gençay, 1996; Dubé & Legros, 2014). Although alternative methods such as spatial econometrics and geographically weighted regression have been proposed to incorporate spatial effects (Choumert et al., 2014; Huang et al., 2017), they require prior knowledge and the assumption of linear relationships between attributes and housing prices and cannot address multiscale effects well (Hu et al., 2019). As a result, to overcome the above issues, more recent studies have turned to machine learning techniques in housing research. Some have compared the model performance among multiple regression approaches to determine better models for real estate price estimation (Chen et al., 2016; Hu et al., 2019; Park & Bae, 2015), and others have proposed improvements of original models or a combination of two models in housing studies (Wang et al., 2014; Yao et al., 2018; Hu et al., 2019). These works have proven the usability and advantage of machine learning methods in forecasting housing prices due to their fit for nonlinear relationships and better prediction accuracy over traditional HPMs. However, the interpretation and visualisation of machine learning results remain limited, requiring further investigation.

2.2. The use of images in urban perceptions

Scientists have claimed that a physical or mental image is an

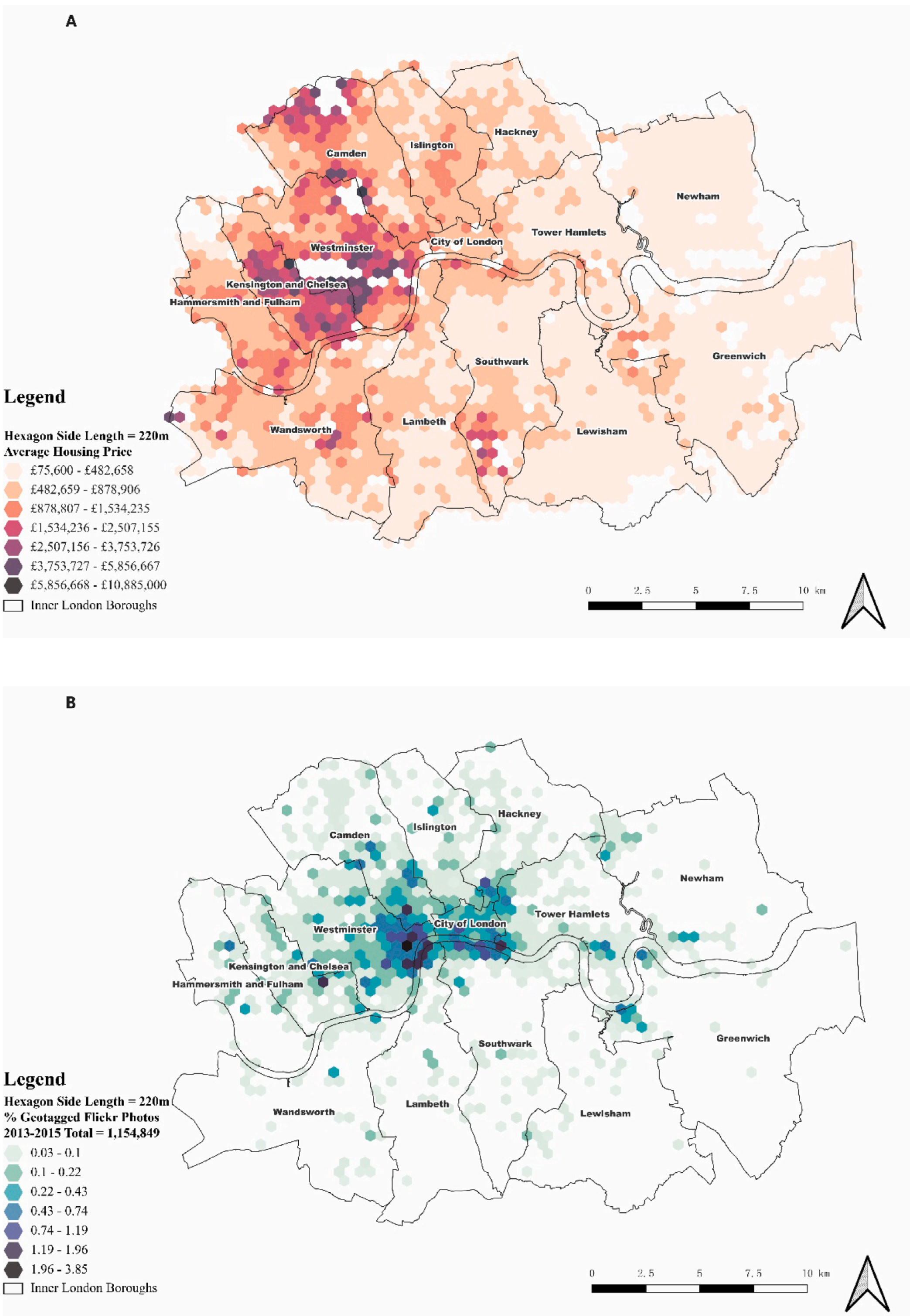

Fig. 1. The spatial distribution of (a) the average housing prices and (b) the percentage of Flickr imagery datasets at hexagonal aggregation in Inner London (2013–2015) (Note: For better visualisation, hexagons with less than 346 Flickr photographs are not plotted on the map, which does not represent there is no Flickr photograph at these areas).

intuitive and direct perspective for capturing the human perception of a city (Ittelson, 1978; Lynch, 1960). However, as previous studies primarily relied on qualitative analysis, such as visual surveys and interviews (Nasar, 1990; Scott, 1998), quantitative measurements remained limited until the technological advances in computer vision revolutionised the field a decade ago. Since then, an increasing number of researchers have utilised images in urban perception studies (Biljecki & Ito, 2021; Ibrahim et al., 2020). Some are interested in the identification of visual representations in the city (Chen et al., 2020; Comber et al., 2020; Doersch et al., 2012; Zhang et al., 2018), and others focus on quantifying perceptual characteristics of the city or on their relationships with nonvisual socioeconomic attributes, such as population density and crime rate (Arietta et al., 2014; Dubey et al., 2016; Khosla et al., 2014; Naik et al., 2014; Salesses et al., 2013). These works measure the perception of places through varying image recognition techniques, while they are unable to reflect different people's interactions and perceptions of a place because they are usually not sourced from varying individuals.

Taking the above literature into consideration, this study seeks to exploit the potential of image-based social media (i.e., Flickr photographs) to analyse housing price studies, aiming to uncover whether social media images can be used to explore the environmental impacts on property nearby and how these perceptual features affect housing values.

3. Data acquisition and processing

This study parimarily used three datasets. The first two encompass traditional housing attributes, including structural, location and neighbourhood features of housing. The third includes geotagged images collected from the social media platform Flickr, capturing scenes around properties. This study focuses on London as a case study, as it provides a reasonable degree of Flickr usage. To obtain a higher density of images and reduce spatial heterogeneity, Inner London is selected since a large volume of Flickr images (73%) fall into this area compared to Outer London. The datasets are collected from 2013 to 2015, when Flickr users were most active, based on the number of photographs uploaded. Fig. 1 shows the density distribution of the average housing price transactions (a) and the percentage of Flickr images (b) in Inner London at hexagonal aggregations, calculated at hexagon length of 220 m. The average prices in Fig. 1(a) display a decreasing pattern from the western boroughs of Kensington and Chelsea to Newham in the east. Fig. 1(b) depicts the percentage of Flickr density larger than 0.03% of total data (i.e., 346 Flickr photographs), diminishing from the highest percentage in eastern Kensington and Chelsea to the rest of Inner London.

3.1. Traditional housing features

Structural features are obtained from housing price data published by the UK HM Land Registry (HM Land Registry, 2019), which has tracked property sales in England and Wales monthly since 1995. The original dataset collected for London includes 226,332 property transactions within the 2013 to 2015 period. The dataset is subsequently cleaned based on Dong et al. (2019), which only keeps properties sold for total market values, as repossessions or buy-to-lets do not reflect real estate market values. Since all housing structural features are categorical data, we quantified them into indicator variables for each category. Additionally, a georeferencing process is used to assign spatial coordinates to all postcode addresses of houses to identify neighbourhood and perceived scene features. Furthermore, as this study only focuses on Inner London, the data outside the study area are removed. After these preprocessing steps, 137,132 property transaction records remained, with each containing property transaction price, the postcode address, the spatial coordinates, the date of transfer, the property type, whether the property is new or old, and the tenure type.

Neighbourhood and location features are collected from POIs data and open greenspace data published by the Ordnance Survey (Ordnance Survey, 2020). As demonstrated in the literature reviewed above, buyers tend to purchase houses with perfect amenities that relate to the area where it is located, such as convenient transportation, easy access to social infrastructure and access to open spaces. Therefore, a spatial query is conducted to measure the number of buses, underground stations, schools, medical (i.e., health) care, and entertainment centres within 800 m of each house and the distance from one house to the nearest amenities. Furthermore, areas of green space within 800 m of houses are also calculated following evidence of their relevance in determining housing values (Hu et al., 2019; Sirmans et al., 2005). In particular, the 800 m Euclidean distance is used as a threshold because it was considered a pedestrian and cycling-friendly distance for residents in a neighbourhood (Liu et al., 2020).

3.2. Scene (image) features

Scene features are identified from geotagged social media images collected from Flickr, an online photo-sharing community with over 90 million monthly users (Smith, 2020). Although Flickr data have some biased aspects such as possible GPS bias of image geolocation, the bias of skewed representation of landmarks and, most importantly, self-selection, their usability in identifying representative urban characteristics is powerful and has been demonstrated by many studies (Chen et al., 2019; Kisilevich et al., 2010; Seresinhe et al., 2018; Zhou et al., 2014). Furthermore, the extensive coverage of Flickr data for Inner London has enabled both iconic landmarks such as towers and bridges as well as daily life scenes such as bars and conference centres to be identified (Chen et al., 2020), further illustrating the potential usability of Flickr images in this study.

A total of 1,154,849 geotagged images from 2013 to 2015 are collected from the official Flickr API (https://www.flickr.com/services/api/) and used for image recognition, with each image including latitude, longitude and the specific date taken. A series of preprocessing steps are performed to reduce the data redundancy and dominance of the extremely active users to mitigate heterogeneity, as stated in our previous work (Chen et al., 2019). Scene features are extracted by a pre-trained Places365 convolutional neural network model, an image recognition technique designed for identifying 365 scene-related categories or places (Zhou et al., 2018). This model is used because it is trained based on the Flickr dataset and its high performance that the recognised accuracy of the top five categories approaches 85.08% (Zhou et al., 2018). Furthermore, its capability to identify scenario-based places from the built environment has gone beyond many other image recognition models, such as YOLO (You Only Look Once), trained on 20 objects that include office furniture, animals and food-related categories. Through this procedure, each image is assigned to return only five scenario-based categories with identified probabilities from high to low, where a class with higher percentage values implies more significant features of that image.

To investigate the possible impacts of features encoded in images on houses, we select images georeferenced within the same 800 m distance of houses, subsequently quantifying the probability of each scene characteristic based on the corresponding space and time. However, several scene features may have little impact on housing prices, such as glaciers, icebergs, cliffs, volcanoes and tundra these countryside natural scenes, leading to higher computational costs and lower model performance. Thus, feature selection is implemented to select a subset of 365 scene features that are important and relevant to housing prices that are included in the traditional modelling framework. Given the limited computational capacity and possible multicollinearity among features (Li et al., 2017), RF feature importance is employed to select features by their built-in mean decrease impurity (MDI) function. MDI refers to the total decrease in node impurity averaged over all ensemble trees, where impurity represents a function weighted by the proportion of samples reaching that node (Pedregosa et al., 2011). The impurity measures the goodness of any node of the decision trees (i.e., variance for regression). The less impurity, the purer the node and the better the prediction accuracy (Louppe, 2014). The logic of this feature selection mechanism is that when training a tree, the more a feature decreases the impurity, the more influential the feature is. For many decision trees in an RF, the impurity decrease of each feature can be averaged across trees to compute the final importance of the variable (Breiman, 2001). To select more robust and vital features, the standard deviation of variance is set as a threshold to drop features that are lower than the value. Fig. 2 displays seven more important scene features selected (blue bars) using RF feature importance after feature selection.

Additionally, the coefficient of multiple correlation R is employed to measure the strength of the association between a given (dependent) variable and a set of (independent) variables. It ranges from 0 to 1, with 1 denoting the strongest correlation, whereas a value of 0 indicates no correlation between a combination of independent variables and the dependent variable (Cohen, 1988). This coefficient can be easily computed as the square root of the coefficient of determination of multiple linear regression. In this case, seven more important scene features are combined to build a multiple linear regression model following Equation (1) in Section 4.1 to measure the total association between seven image feature variables and one housing price variable. The computed coefficient equals 0.33, indicating a modest correlation between selected image features and house prices. Therefore, only the seven most relevant image features will be added to the house price estimation models. The average probability of these image features within 800 m of houses is computed to mitigate the impact of image density dispersion. Properties without images around were excluded, which are primarily distributed at Newham and Greenwich and account for 12% of the data coverage.

After several stages of data preprocessing, a total of 23 independent variables, including traditional housing features and scene features, are finally selected to explore whether images can be considered as an additional data source and how they influence the housing market. The overall descriptions and statistics of the three types of features are displayed in Table 1.

4. Methodology framework

A method framework (see Fig. 3) is proposed to explore two aspects of housing price estimation: (1) whether geotagged images are an efficient data source to unpack the impacts of housing prices and (2) whether machine learning ensemble methods are more promising tools in terms of performance and interpretation compared with a traditional HPM. This framework consists of three phases: feature fusion, modelling, and evaluation and interpretation.

Models will be trained on two sets of variables through feature fusion. Traditional variables include 16 basic housing structural and location features, and enriched variables include all 23 features summarised in Table 1. Three models are employed, including one HPM that serves as a baseline and two machine learning models: RF and GBM, to further train on features identified through baseline. The estimations will then be analysed and evaluated through model performance and interpretation. Model performance evaluates how well the constructed models fit the observations, and model interpretation unpacks the relationships between all independent variables and housing prices. The best model will be recognised based on prediction performance and interpretation. The remainder of this section summarises each technique and procedure involved in the framework.

4.1. Baseline hedonic price model

A parametric HPM is first used as a benchmark in our approach. The HPM assumes a linear functional form described by a group of parameters—the coefficients of independent variables (Horowitz & Lee, 2002). Although nonparametric approaches such as kernel estimates avoid the strong linear assumptions underlying the parametric methods, they have been criticised for their “curse of high dimensionality” and computational burden. Therefore, a semilogarithmic HPM is selected for its intuitive interpretation, ease of use and variation consideration (Sirmans et al., 2005; Zhang & Dong, 2018). In particular, the property transaction price is viewed as the dependent variable, and all features are independent variables. Before fitting the models, the input variables are standardised to express each variable in the same units (i.e., between 0 and 1) and thus ease interpretation. The mathematical formula of the semi-log hedonic model is displayed in Equation (1):

$$LnP = \alpha + \sum \beta kCk + \varepsilon \tag{1}$$

Ln P refers to the logarithmic form of the transaction price at the postcode level, βk is the coefficient of one housing characteristic, where k represents the number of independent variables C, α is a constant term, and ε is the random error term.

The mechanism of this model is to find the optimal coefficients for all the variables that minimise the error. The interpretation is relatively straightforward: the estimated coefficients represent the marginal change of the dependent variable when a unit increase occurs in one of the independent variables. However, this approach is highly sensitive to multicollinearity and outliers and limited to capturing nonlinear relationships and large numbers of variables (Sirmans et al., 2005). As a result, the variance inflation factor (VIF) is used to check for multicollinearity in our model trained with all 23 independent variables. VIF is a measure of collinearity and correlation among predictor variables within a multiple regression model. A rule of thumb is that if the VIF is larger than the threshold of 10, then the variable is considered highly colinear and correlated with the other variables (Kutner et al., 2004).

4.2. Machine learning methods

Both RF and GBM are ensemble machine learning methods that combine the predictions of several base estimators on a given algorithm to gain better robustness than a single estimator (Pedregosa et al., 2011). RF generates many uncorrelated decision trees based on averaging random selection of predictor variables from the training set (Breiman, 2001). GBM trains a series of models in a stagewise, additive, and sequential manner that optimises arbitrary differentiable loss functions (Friedman, 2001). Unlike RF, where each tree can be trained independently, each tree in GBM is determined by previous outputs. These two methods have been commonly used because they handle more extensive features, high accuracy performance, and robustness to skewed distributions, multicollinearity, outliers, and missing values (Pal, 2017). On the other hand, these two models have stated usability in housing price studies and higher interpretability than other machine learning models, such as neural networks (Hu et al., 2019; Yao et al., 2018). Their joint pitfalls are computationally expensive and may overfit particularly noisy datasets.

A common characteristic in machine learning methods is that they are parameterised by a range of hyperparameters, which are required to be tuned and optimised to yield an optimal model that minimises some predefined loss function (Claesen & Moor, 2015). Manual and grid searches are the most frequently used hyperparameter optimisation methods; however, they have difficulties reproducing results and suffer from too many trials for dimension exploration. (Bergstra & Bengio, 2012). Hence, random search, where each parameter setting is sampled independently from a specified distribution over the cross-validated search, is implemented due to its primarily high efficiency and less computational time. To obtain a reasonably decent set of values of the hyperparameters, either a distribution over possible and random values or a list of discrete choices can be specified for each parameter. The important parameters to adjust for RF are the number of trees, the minimum number of samples at a leaf node and the number of features for a split. In the case of GBM, the parameters are the number of boosting

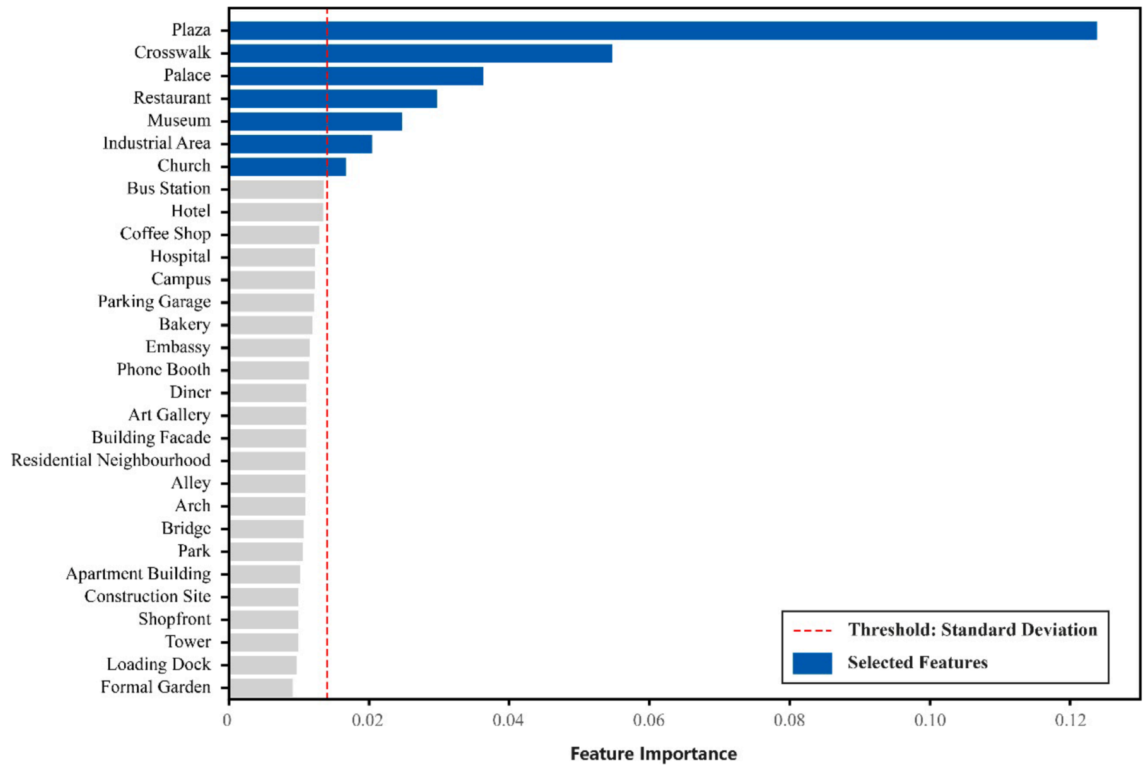

Fig. 2. The seven most relevant features selected through the feature selection process.

stages, learning rate, the minimum number of samples at a leaf node and to split the node, the maximum depth to limit the number of nodes (Pedregosa et al., 2011)[5]. Fivefold cross-validation, a typical split-train-test strategy that minimises the estimator error, is used in our random search. More details of cross-validation are explained in Section 4.3. The optimised hyperparameters of RF and GBM are shown in the footnote.

4.3. Model performance and interpretability

In any modelling context, validation and performance are crucial to evaluate how accurate and reliable the constructed models are (Hastie et al., 2009). A set of visualisation tools are used to validate the prediction and three popular statistical metrics, mean absolute error (MAE), mean squared error (MSE), and the coefficient of determination (R^2), to evaluate the model performance. Combining these three metrics implies the model's predictive power, and independent variables describe the variation of the observed variable. MAE and MSE compute the average absolute and squared error or loss between the predicted and the actual values, which are always positive and represent better predictions the smaller their value. R^2 is an index representing the percentage of the variance in the output explained by predictors (i.e., independent variables) in a regression model, which ranges from 0 to 1 and where larger values represent more explanatory models.

To avoid overfitting and unreliable results, cross-validation (CV) is performed to evaluate all model performances on our limited data sample. The basic approach is called k-fold CV, which divides the dataset into the number of k nonoverlapping partitions (James et al., 2013). For each k group or fold, a model is trained on k-1 of the folds, and the remaining part of the data is treated as testing data to measure the model's performance. The resulting measure is often summarised with the average values computed in the k loop. Considering the data size of this study and the computational cost, a commonly used k = 5 (Arribas-Bel et al., 2017; James et al., 2013) is configured to calculate cross-validated MSE and R^2 then the model with better performance will be recognised for interpretation.

Since RF and GBM cannot be interpreted by examining regression coefficients and significance due to their nonparametric nature, permutation importance and accumulated local effects (ALE) plots were utilised to explore the relationships between variables and the observations. These two methods can measure the relationship between input factors and the observations. Permutation importance is calculated in two steps: first, a baseline metric of the estimator is evaluated on the training dataset; second, a single feature column from the validation set is permuted, and the metric is recomputed (Breiman, 2001; Pedregosa et al., 2011). The importance is the difference between the baseline and the drop in overall metric by permuting the column. In addition to being more reliable, permutation importance can also overcome the problem that many unique values can be misleading compared to the traditional feature importance method of several ensemble methods. MSE is the metric used in this study to measure feature importance. ALE plots visually reflect how features affect the prediction of a machine learning model on average (Apley & Zhu, 2016). To estimate local effects, the feature is divided into many intervals defined by the quantiles of the feature distribution to measure their differences in the predictions. The ALE value represents the key effect of the feature at a given value compared to the average forecast centred at zero. Unlike the more popular partial dependence plots (PDPs), which display the marginal effect of one or two features on a machine learning prediction model, ALE plots are faster, unbiased and a more interpretable tool (Molnar, 2019). This is because PDPs can significantly bias the estimated feature effect if features are correlated, which is the case in our study.

[5] The values for hyper parameters that we use include:RF: n_estimators = 200, min_samples_leaf = 2, max_features= 'auto', max_depth = 30;GBM: n_estimators = 350, learning_rate= 0.1, min_samples_split= 25, min_samples_leaf= 50, and max_depth = 10.

Table 1
Descriptions and statistics of three types of variables for housing prices.

Categories	Variables	Descriptions	Mean
Structural features	type_F	Dummy variables, 1 if the property type is flat	0.785
	type_S	Dummy variables, 1 if the property type is semi-detached	0.025
	type_T	Dummy variables, 1 if the property type is terraced	0.183
	new_Y	Dummy variables, 1 if the property is newly built	0.128
	tenure_L	Dummy variables, 1 if the tenure is Leasehold	0.795
Location (neighbourhood) features	bus_num	Number of bus or coach stations within 800 m distance	0.031
	sub_num	Number of underground stations within 800 m distance	0.219
	lei_num	Number of leisure or sports centres within 800 m distance	0.157
	mel_num	Number of medical care centres within 800 m distance	0.187
	sch_num	Number of primary schools within 800 m distance	2.165
	bus_dis	Distance to the nearest bus and coach station	2.174
	sub_dis	Distance to the nearest underground station	1.477
	lei_dis	Distance to the nearest leisure or sports centre	0.875
	med_dis	Distance to the nearest medical care centre	0.918
	sch_dis	Distance to the nearest primary school	0.240
	park_area	Coverage of parks and gardens within 800 m distance	0.048
Scene features (within 800 m distance of houses)	plaza	Mean probability of images classified as plaza	0.013
	crosswalk	Mean probability of images classified as crosswalk	0.008
	palace	Mean probability of images classified as palace	0.002
	restaurant	Mean probability of images classified as restaurant	0.005
	museum	Mean probability of images classified as museum	0.006
	ind_area	Mean probability of images classified as industrial area	0.008
	church	Mean probability of images classified as church	0.002

5. Results and discussions

5.1. Model performance

Fig. 4 visualises the actual and predicted logarithmic house prices of HPM, RF and GBM between traditional and enriched variables of house price estimation. Fig. 4(a) shows scatter plots of six sets of models, and Fig. 4(b) displays their density distribution based on kernel density estimation that uses a continuous probability density curve in more dimensions. Overall, the R^2 of all models improve slightly when scene features derived from images are considered into the models. The predicted values are undervalued to some extent compared to the observed values for all models, particularly for two HPMs that predicted values obviously deviate from actual values regardless of whether image features are added. The larger HPMs gaps in Fig. 4(b) are most likely due to the influence of multicollinearity of the models, as discussed in Section 4.1. The predictions of machine learning models in blue and orange in both Fig. 4(a) and (b) appear to fit the observations better than HPM. RF and GBM approach actual log prices more when scene features are added to independent variables, indicating that models trained RF and GBM fit our data well and have better performance than both predictions of HPMs regardless of whether images are added to the variables.

Next, the cross-validated MAE in the units of logarithmic house prices, MSE in the square of logarithmic house prices and R^2 are calculated to reflect generalisation performance, as shown in Table 2. The overall performance of the three models improved with slightly higher R^2, lower MSE and MAE when image attributes were considered. The performance of HPM is inferior to RF and GBM, as shown by larger MSE, MAE and smaller R^2; in contrast, the RF model shows better accuracy and robustness, with the highest R^2 and smallest MSE and MAE, where the entire 23 input variables could explain 66.5% of the variance in the observation. This table illustrates the superiority and flexibility of the two machine learning models due to their less uncertainty (lower MSE and MAE) and greater accuracy (higher R^2). Although the improvements in prediction are not highly significant with additional perceived scene features, the results also imply that user-generated images of urban surroundings may be used as a supplementary data source in house price estimation.

Since RF with all 23 variables performs better than others, we further create a residual map at hexagonal aggregation, shown in Fig. 5, to visualise the difference in logarithmic house prices between observations and estimations to gain intuitive insight into the spatial

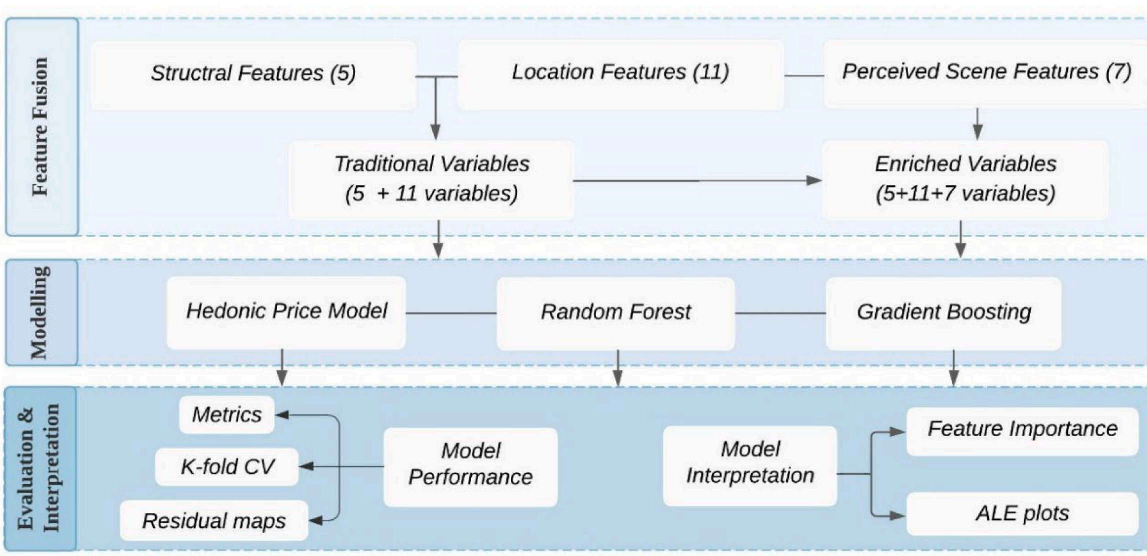

Fig. 3. Overall methodological framework.

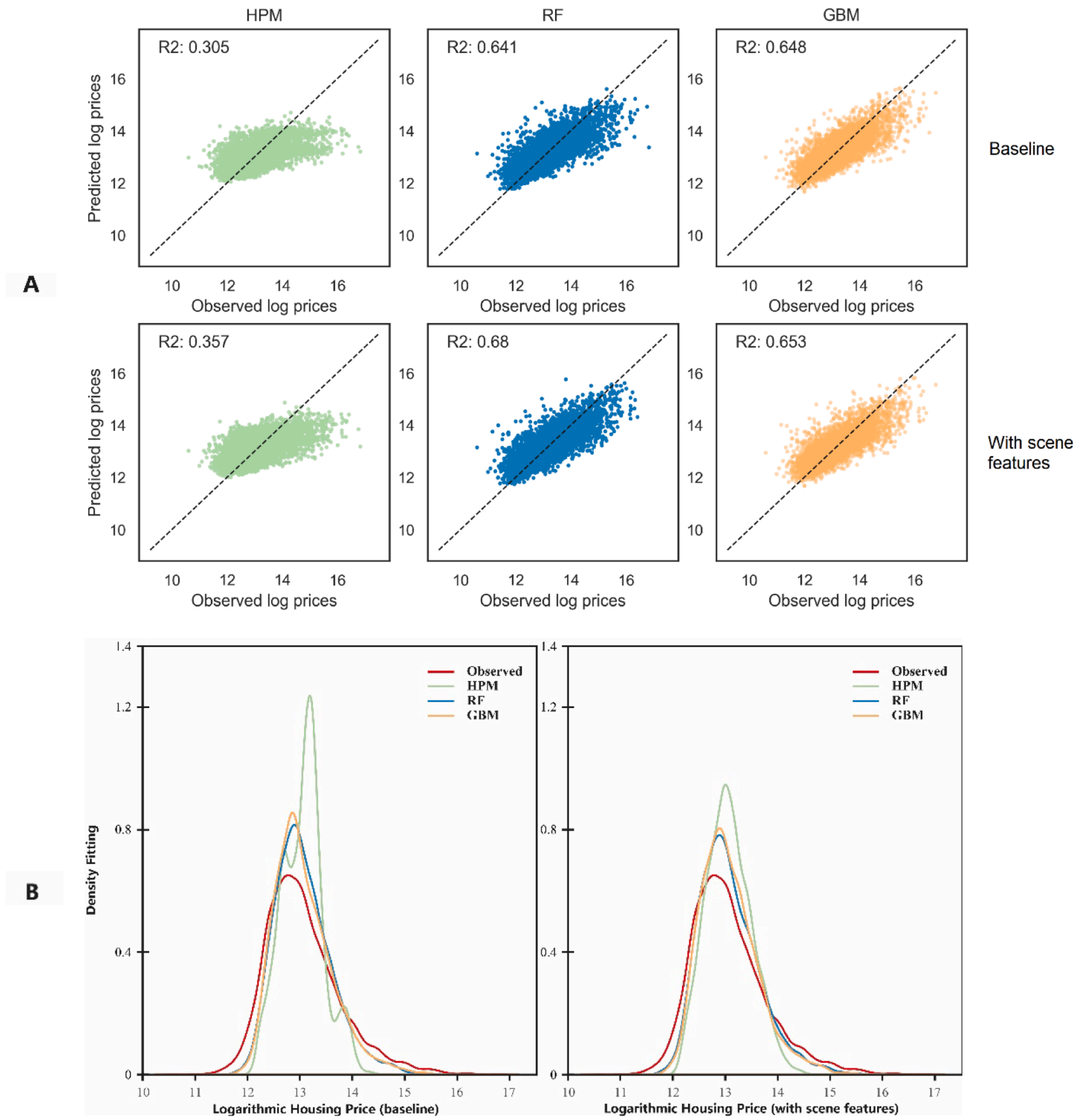

Fig. 4. Visualisation of model fitting at (a) actual and predicted logarithmic housing prices and (b) density distribution of all models, with traditional and enriched variables.

Table 2
Accuracy and error scores for various models and attributes.

Metrics	Housing attributes only			Housing attributes + Image attributes		
	HPM	RF	GBM	HPM	RF	GBM
R^2	0.305	0.619	0.616	0.356	0.665	0.635
MSE	0.355	0.193	0.189	0.363	0.169	0.185
MAE	0.453	0.315	0.320	0.439	0.293	0.313

distribution of the errors. The blue-green and red–orange dots represent higher and lower estimations than the actual log house prices. The overall residuals of housing prices in Inner London fluctuate around 0 (i.e., yellow dots). At the same time, properties distributed in the western areas of Inner London (Boroughs of Kensington and Chelsea, Westminster and Camden) have lower estimations than the actual prices, implying that the errors located in these areas are difficult to explain by our regression model. A possible explanation for this error is that the average house prices of these three boroughs are far higher than those of the other areas in Inner London, as shown in Fig. 1(a).

5.2. Model interpretation

Before looking into the interpretability of RF, which has the best performance, it would be good to see how the baseline linear HPM behaves, in other words, the magnitude of all 23 independent variables. This procedure could help us enhance the reliability of the interpretation of RF. The effect sizes (coefficient), the significance of estimated variables (p-values) within the 95% confidence interval, and VIF are displayed in Table 3. The larger the coefficient values, the more marginal the changes in the outcome associated with a unit increase in each determinant. The p-value less than 0.05 represents that the variable is statistically significant for the model.

Overall, traditional housing features such as the type of house (i.e., flat or terraced), the distance to the nearest subway station and whether the tenure type is leasehold or freehold (type_F, type_T, sub_dis, tenure_L) affect housing prices much more than others and are also statistically significant. Most coefficients and significance remain stable regardless of whether scene features are considered. Only a minor reduction of impacts on the location-type variables was obtained by

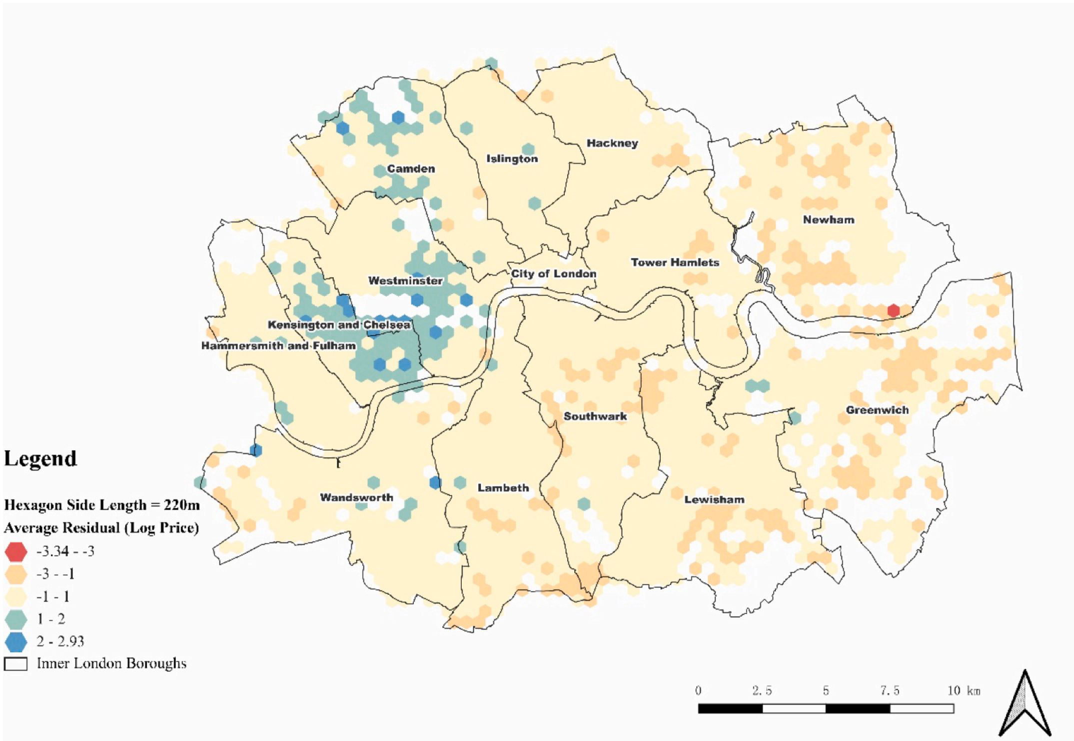

Fig. 5. Spatial distribution of residuals of actual and predicted log house prices.

Table 3
Standardised coefficients of the baseline HPM model with different numbers of variables.

Features	Traditional housing features (16)		With additional image features (23)		
	Coefficient	P-value	Coefficient	P-value	VIF
intercept	13.075	0.000	13.075	0.000	
type_F	−0.315	0.000	−0.337	0.000	35.420
type_S	−0.034	0.000	−0.036	0.000	4.396
type_T	−0.174	0.000	−0.184	0.000	21.481
new_Y	0.019	0.000	0.019	0.000	1.059
tenure_L	−0.141	0.000	−0.141	0.000	12.596
bus_num	−0.001	0.872	−0.003	0.315	1.209
sub_num	−0.074	0.000	−0.093	0.000	1.860
lei_num	0.010	0.013	0.011	0.005	1.777
medi_num	−0.023	0.000	−0.011	0.003	1.764
sch_num	−0.004	0.271	−0.003	0.447	1.653
bus_dis	−0.028	0.000	−0.013	0.000	1.317
sub_dis	−0.361	0.000	−0.316	0.000	2.078
lei_dis	0.042	0.000	0.041	0.000	1.899
med_dis	−0.070	0.000	−0.031	0.000	1.930
sch_dis	0.050	0.000	0.049	0.000	1.566
park_area%	−0.018	0.000	−0.016	0.000	1.038
plaza			0.064	0.000	1.217
crosswalk			0.053	0.000	1.147
palace			0.060	0.000	1.209
restaurant			0.054	0.000	1.080
museum			0.038	0.000	1.155
industrial_area			−0.005	0.070	1.091
church			0.048	0.000	1.208

introducing scene variables. In addition to the number of subway stations, the number of other POIs, including schools, health centres, retail centres and bus stations, has little impact on house prices within a 800 m distance. In terms of the scene features, plazas, crosswalks, palaces, restaurants, museums and churches have more influence than most neighbourhood features identified from POI data and significant explanatory power in predicting housing prices. Only the industrial scene has no association with house prices based on the coefficient and the p-value. Particularly, the flat and terraced type and the shortest distance to subway stations and health centres have negative relationships with housing prices; the higher the values of these variables are, the greater the decrease in housing prices. Conversely, most of the scene variables have positive effects on the estimation, which conforms to prior knowledge that the more attractive the scenery and robust infrastructure around a house is, the higher the price of the house.

The calculated VIFs of the property type of flat (35.420), property type of terraced (21.481) and tenure type of leasehold (12.596) are greater than 10, indicating that these three housing variables are highly correlated with each other. Multicollinearity is not a problem for nonparametric tree-based methods. Hence, we then turn to the interpretation of RF trained on all features. Fig. 6 plots the permutation importance scores computed, where the nine red bars and three green bars represent more important traditional and image features to the prediction that are larger than the median of importance. It is evident that the distance to the nearest subway stations within a 800 m distance contributes the most to the predictive power of the model, and the other four accessibility variables (i.e., the shortest distance to a property) and the coverage of parks also have important effects on the estimations. The terraced flat and tenure type of leasehold is far more critical than the other housing structural features. Relative significant perceived scene features are palaces, plazas and crosswalks, conforming to common knowledge that attractiveness and accessibility have clear impacts on house prices. However, whether the property is new or old and is the semi-detached type has almost no association with its housing price. In addition, the number of different POIs and infrastructures within a 800 m distance of the property proves less relevant to the estimation. The possible reason for the above outcomes is that a tiny fraction of transactions has records for these features during the study period, such as the valid values for the degree of new or old and the type of semi-detached property only accounting for 10% and 3% of the data, consequently, hardly contribute much to the predictive power of the model.

The overall interpretation of the RF model is similar to that of the benchmark HPM except that the RF model captures more significance in

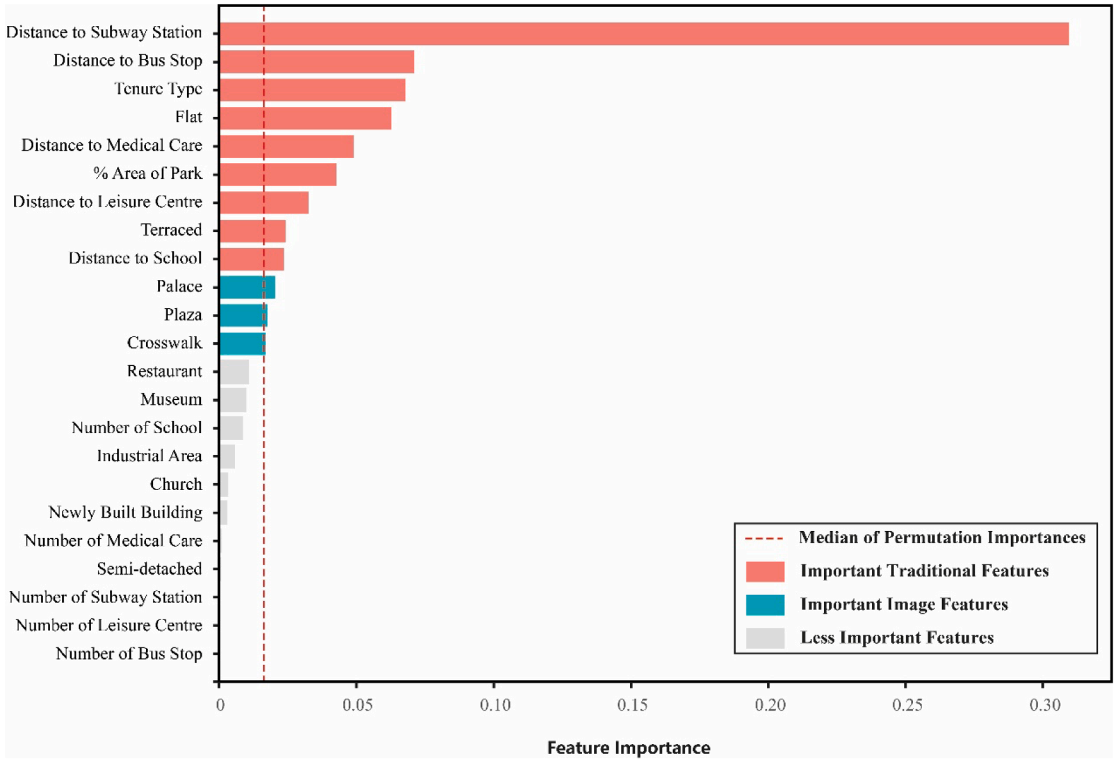

Fig. 6. Permutation feature importance based on different input variables.

service accessibility and the coverage of green parks. The more important input variables are convenient transportation, accessibility of essential social infrastructure, the flat property type, the terrace property type, the tenure type and a few perceived scenes on housing prices. The results show that in addition to the conventional influence features of housing prices, how people interact with the surrounding environment of properties also impacts housing markets. Compared with the neighbourhood features identified through POI data, the image-based perceived scene features highlight the significance of the attractiveness of specific local amenities and places to housing prices. It is the core merit of considering perceived scene features into housing prices, helping the restructuring and optimisation of residential areas in future regional construction, planning and development.

Fig. 7 displays the ALE plots of the most relevant variables from the structural, location, and perceived scene features: the flat, the shortest distance to the subway station and perceived palace scenes to explore the relationships between variables and estimations further. ALE plots of baseline HPM are also included for comparison. The horizontal and vertical axes represent the range of variables and accumulated local effect values, respectively, and the zero value implies the average prediction effect. The overall patterns of ALE plots for both baseline HPM and RF are consistent. The features property type of flat and the shortest distance to subway stations have negative relationships with the observation, while the perceived scene feature of a palace is positively associated. The differences between the two models are first linear and nonlinear relations, and second, the average prediction of HPM changed more than RF with the same increasing values of features. Specifically, the average prediction decreases with the rising value of the property type of flat, but it flattens out until 0.5 for the rest; the longer the distance to the nearest subway station, the lower the prediction power. The perceived scene feature of a palace has a strong positive effect on the prediction for HPM while a weak positive impact on the prediction for RF, which corresponds to the relatively low importance of palace for the model prediction in Fig. 6.

The results suggest that a shorter distance to subway stations, less opportunity to live in flat property types, and more potential to live next to attractive scenes such as palaces can significantly increase local real estate prices. The findings could be informative for policymakers to formulate equitable housing policies and help urban governance based on the physical environment and the popularity in terms of how people perceive and interact with a neighbourhood. Additionally, dynamic housing price changes are significant for local planners, for example, to develop a healthy housing market through neighbourhood public service configuration and distribution and more affordable homes to a broader population.

6. Conclusions

This paper explores the potential of geotagged social media images for monitoring housing prices and the superiority and flexibility of using machine learning methods to understand the impacts of various features on the housing market. Multiple datasets are employed to extract three types of elements: structural, location and scene attributes. With these, two machine learning methods, RF and GBM, are used and compared to the traditional house price estimation model HPM. The results illustrate that RF proved to be the best model based on performance, and it is also as interpretable as HPMs through a series of visualisations. In summary, the empirical results indicate that scene features extracted from geotagged user-generated images could add minor value to house price estimation in a sense. Properties surrounded by well-equipped amenities and natural scenes tend to be considered more attractive and have a higher value.

The main contributions of this research are twofold. On the one hand, we uncovered the possible potential of user-generated social media images in house price estimation. Although the marginal improvement on the model performance, user-generated images could be used as a supplementary data source for house price estimation if human perception is considered. This filled in current research gaps that neighbourhood features identified from POIs and street view data were unable to capture how people experienced and interacted with the

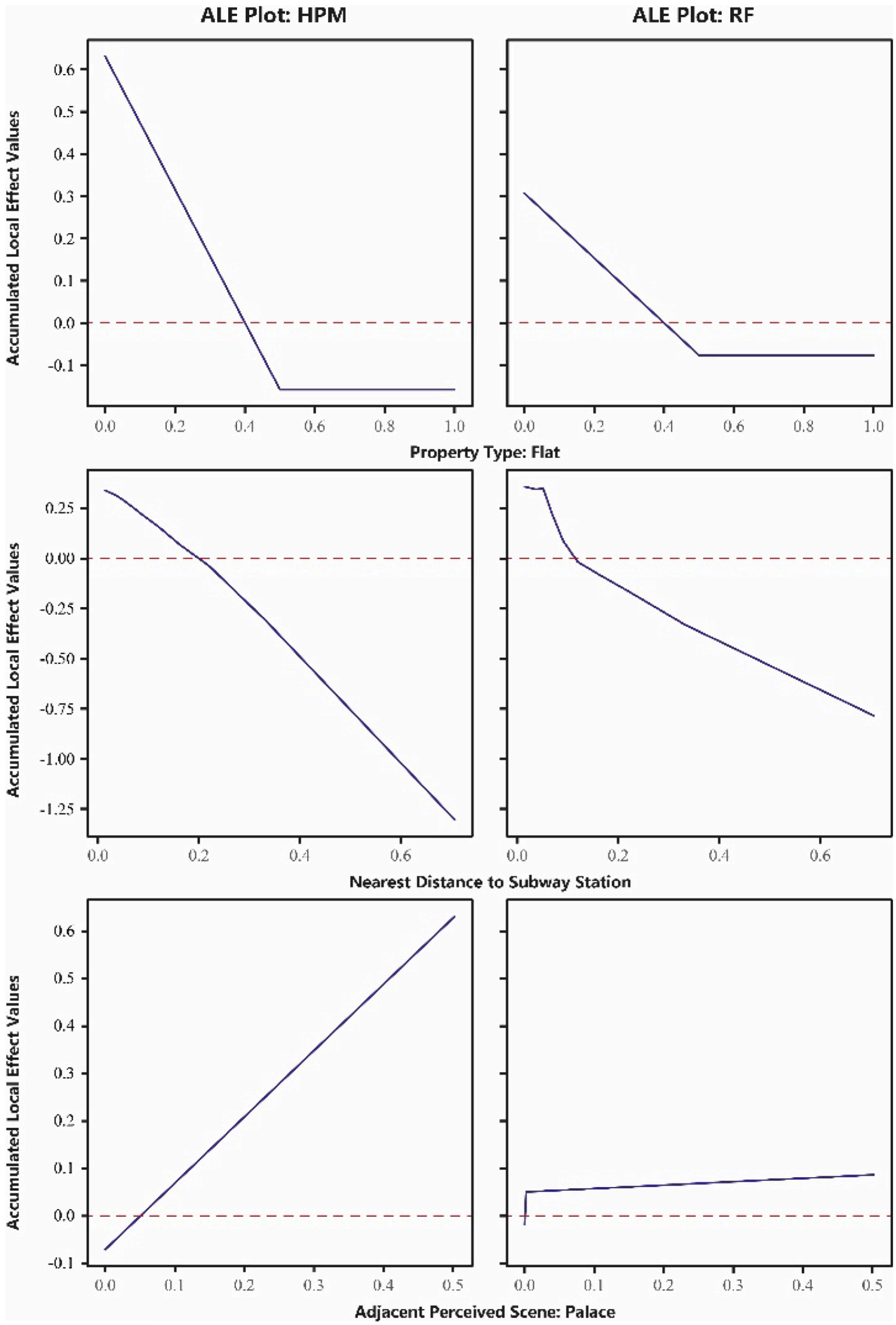

Fig. 7. Accumulated local effects plots for partial representative features.

physical environment. By including urban scene features extracted by various citizens, the impacts of the popularity of scenes on the housing market are revealed to a certain extent. On the other hand, the methodological framework that integrates baseline HPM with two machine learning algorithms is insensitive to multicollinearity and proven to have more accurate performance and equivalent interpretability, avoiding the usual black-box problems attributed to common machine learning algorithms. This is also applicable to other traditional empirical-statistical methods that consider spatial heterogeneity, such as geographically weighted regression, which has a high dependence on prior knowledge and relatively poor capability of addressing multiscale effects (Hu et al., 2019).

From the perspective of developing a sustainable city, in addition to traditional datasets, stakeholders may also consider user-generated images as a supplementary dataset when assessing the housing transaction market. This data source can capture residents' interactions with the urban environment, reflecting their interests and perceptions of urban scenes. The patterns may be helpful to real estate developers for early-stage site selection of residential buildings. Living environments with good amenities, such as convenient transportation, accessibility of green space, recreational places, and distinctive scenes, such as plazas, palaces and crosswalks, are relatively important factors in house price estimations. Furthermore, the government could pay more attention to the adjustment and design of housing development based on various facilities and surrounding urban features. It could assist in improving the vitality of the area surrounding a property, which subsequently influences people's willingness to buy that property.

Despite the contributions, it should be noted that the limitations of Flickr data in representation remain unresolved. First, like other user-generated social media data, the Flickr platform is being used in a self-selection process (Goodchild, 2007), implying that the number of users does not represent all age groups or genders. This has been demonstrated in a recent survey that the dominant age groups for these platforms are teenagers or middle-aged males (Barnhart, 2021). As a result, the perception features extracted from Flickr were based mainly on specific population groups that were not sufficiently representative. Second, the limitation of spatial heterogeneity and lack of data in a few neighbourhoods cannot be overlooked. It implies that more photographs existed of tourist attractions and landmarks as shown in Fig. 1b, suggesting our models are dominantly built on houses with more iconic scenes surroundings rather than all houses evenly in Inner London. Even if a series of data processing steps have been taken in Section 3.2, the underlying data bias of spatial heterogeneity in Flickr is still included. To minimise this limitation, combining other user-generated imagery data to extract features, such as Google Photos and Instagram, might be possible. The data coverage and representation could be possibly improved by linking diverse datasets through aggregated spatial and temporal scales.

This research could be extended and improved in a few ways. First, other image recognition methods, such as image segmentation, can be used to extract more precise scene features for our housing price estimation model. Second, additional datasets could be used to capture possible impact factors on real estate prices, such as the data including more housing structural features such as the size or the number of bedrooms for a single house, interior visual images (Ahmed & Moustafa, 2016) or street views imagery. Furthermore, more cities could be included to compare the relationships and differences. For example, do identical impact factors influence local housing prices and are there any distinctive scene features for each city? Moreover, the time dimension could be further considered to unpack the dynamic impacts of housing prices, helping monitor the changes in the housing market and regulating housing prices over time.

References

Ahmed, E. H., & Moustafa, M. (2016). House price estimation from visual and textual features. *IJCCI 2016 – Proceedings of the 8th International Joint Conference on Computational Intelligence*, *3*(Ijcci), 62–68. https://doi.org/10.5220/0006040700620068.

Anglin, P. M., & Gençay, R. (1996). Semiparametric estimation of a hedonic price function. *Journal of Applied Econometrics, 11*(6), 633–648. https://doi.org/10.1002/(SICI)1099-1255(199611)11:6<633::AID-JAE414>3.0.CO;2-T

Apley, D. W., & Zhu, J. (2016). *Visualizing the Effects of Predictor Variables in Black Box Supervised Learning Models*. http://arxiv.org/abs/1612.08468.

Arietta, S. M., Efros, A. A., Ramamoorthi, R., & Agrawala, M. (2014). City forensics: Using visual elements to predict non-visual city attributes. *IEEE Transactions on Visualization and Computer Graphics*. https://doi.org/10.1109/TVCG.2014.2346446

Arribas-Bel, D., Patino, J. E., & Duque, J. C. (2017). Remote sensing-based measurement of Living Environment Deprivation: Improving classical approaches with machine learning. *PLoS ONE*. https://doi.org/10.1371/journal.pone.0176684

Baker, E., Bentley, R., Lester, L., & Beer, A. (2016). Housing affordability and residential mobility as drivers of locational inequality. *Applied Geography*. https://doi.org/10.1016/j.apgeog.2016.05.007

Barkham, R., Bokhari, S., & Saiz, A. (2018). *Urban Big Data: City Management and Real Estate Markets*. https://mitcre.mit.edu/wp-content/uploads/2018/01/URBAN-DATA-AND-REAL-ESTATE-JAN-2018-1.pdf.

Barnhart, B. (2021). *Social media demographics to inform your brand's strategy in 2021*. https://sproutsocial.com/insights/new-social-media-demographics/.

Bergstra, J., & Bengio, Y. (2012). Random search for hyper-parameter optimization. *Journal of Machine Learning Research*.

Biljecki, F., & Ito, K. (2021). Street view imagery in urban analytics and GIS: A review. *Landscape and Urban Planning*. https://doi.org/10.1016/j.landurbplan.2021.104217

Breiman, L. (2001). Random forests. *Machine Learning*. https://doi.org/10.1023/A:1010933404324

Cai, W., & Lu, X. (2015). Housing affordability: Beyond the income and price terms, using China as a case study. *Habitat International*. https://doi.org/10.1016/j.habitatint.2015.01.021

Causa, O., & Pichelmann, J. (2020). Should I stay or should I go? Housing and residential mobility across OECD countries. *Economics Department Working Papers*.

Chen, M., Arribas-Bel, D., & Singleton, A. (2019). Understanding the dynamics of urban areas of interest through volunteered geographic information. *Journal of Geographical Systems*. https://doi.org/10.1007/s10109-018-0284-3

Chen, M., Arribas-Bel, D., & Singleton, A. (2020). Quantifying the characteristics of the local urban environment through geotagged flickr photographs and image recognition. *ISPRS International Journal of Geo-Information, 9*(4). https://doi.org/10.3390/ijgi9040264

Chen, W. Y., & Jim, C. Y. (2010). Amenities and disamenities: A hedonic analysis of the heterogeneous urban landscape in Shenzhen (China). *Geographical Journal*. https://doi.org/10.1111/j.1475-4959.2010.00358.x

Chen, Y., Liu, X., Li, X., Liu, Y., & Xu, X. (2016). Mapping the fine-scale spatial pattern of housing rent in the metropolitan area by using online rental listings and ensemble learning. *Applied Geography*. https://doi.org/10.1016/j.apgeog.2016.08.011

Choumert, J., Stage, J., & Uwera, C. (2014). Access to water as determinant of rental values: A housing hedonic analysis in Rwanda. *Journal of Housing Economics*. https://doi.org/10.1016/j.jhe.2014.08.001

Chung, R. Y. N., Chung, G. K. K., Gordon, D., Mak, J. K. L., Zhang, L. F., Chan, D., … Wong, S. Y. S. (2019). Housing affordability effects on physical and mental health: Household survey in a population with the world's greatest housing affordability stress. *Journal of Epidemiology and Community Health*. https://doi.org/10.1136/jech-2019-212286

Claesen, M., & De Moor, B. (2015). *Hyperparameter Search in Machine Learning*. https://arxiv.org/abs/1502.02127.

Cohen, J. (1988). Statistical Power Analysis for the Behavioural Science (2nd Edition). *In Statistical Power Anaylsis for the Behavioral Sciences*.

Comber, S., Arribas-Bel, D., Singleton, A., & Dolega, L. (2020). Using convolutional autoencoders to extract visual features of leisure and retail environments. *Landscape and Urban Planning*. https://doi.org/10.1016/j.landurbplan.2020.103887

Doersch, C., Singh, S., Gupta, A., Sivic, J., & Efros, A. A. (2012). What makes paris look like Paris? *ACM Transactions on Graphics. doi, 10*(1145/2185520), 2185597.

Dong, G., Wolf, L., Alexiou, A., & Arribas-Bel, D. (2019). Inferring neighbourhood quality with property transaction records by using a locally adaptive spatial multi-level model. *Computers, Environment and Urban Systems.*. https://doi.org/10.1016/j.compenvurbsys.2018.09.003

Dubé, J., & Legros, D. (2014). Spatial econometrics and the hedonic pricing model: What about the temporal dimension? *Journal of Property Research*. https://doi.org/10.1080/09599916.2014.913655

Dubey, A., Naik, N., Parikh, D., Raskar, R., & Hidalgo, C. A. (2016). Deep learning the city: Quantifying urban perception at a global scale. In *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. https://doi.org/10.1007/978-3-319-46448-0_12

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics*. https://doi.org/10.2307/2699986

Goodchild, M. F. (2007). Citizens as sensors: The world of volunteered geography. *In GeoJournal*. https://doi.org/10.1007/s10708-007-9111-y

Granziera, E., & Kozicki, S. (2015). House price dynamics: Fundamentals and expectations. *Journal of Economic Dynamics and Control*. https://doi.org/10.1016/j.jedc.2015.09.003

Hamilton, S. E., & Morgan, A. (2010). Integrating lidar, GIS and hedonic price modeling to measure amenity values in urban beach residential property markets. *Computers, Environment and Urban Systems.*. https://doi.org/10.1016/j.compenvurbsys.2009.10.007

Han, J. M., & Lee, N. (2018). Holistic visual data representation for built environment assessment. *International Journal of Sustainable Development and Planning*. https://doi.org/10.2495/SDP-V13-N4-516-527

Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning. In *Springer Series in Statistics*. https://doi.org/10.1007/b94608.

HM Land Registry. (2019). *HM Land Registry Price Paid Data*. https://data.gov.uk/dataset/4c9b7641-cf73-4fd9-869a-4bfeed6d440e/hm-land-registry-price-paid-data.

Horowitz, J., & Lee, S. (2002). Semiparametric methods in applied econometrics: Do the models fit the data? *Statistical Modeling*. https://doi.org/10.1191/1471082x02st024oa

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

Huang, Z., Chen, R., Xu, D., & Zhou, W. (2017). Spatial and hedonic analysis of housing prices in Shanghai. *Habitat International*. https://doi.org/10.1016/j.habitatint.2017.07.002

Hughes, C., Zagheni, E., Abel, G. J., Wiśniowski, A., Sorichetta4, A., Weber, I., & Tatem, A. J. (2016). *Inferring Migrations: Traditional Methods and New Approaches based on Mobile Phone, Social Media, and other Big Data Feasibility study on Inferring (labour) mobility and migration in the European Union from big data and social media data* (Issue August). https://doi.org/10.2767/61617.

Ibrahim, M. R., Haworth, J., & Cheng, T. (2020). Understanding cities with machine eyes: A review of deep computer vision in urban analytics. *Cities*, *96*(October 2019). https://doi.org/10.1016/j.cities.2019.102481.

Ittelson, W. H. (1978). Environmental Perception and Urban Experience. *Environment and Behavior*. https://doi.org/10.1177/001391657810200

James, G., Witten, D., Hastie, T., & Tibishirani, R. (2013). An Introduction to Statistical Learning with Applications in R (older version). *Springer Texts in Statistics*.

Jim, C. Y., & Chen, W. Y. (2006). Impacts of urban environmental elements on residential housing prices in Guangzhou (China). *Landscape and Urban Planning*. https://doi.org/10.1016/j.landurbplan.2005.12.003

Khosla, A., An, B., Lim, J. J., & Torralba, A. (2014). Looking beyond the visible scene. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. https://doi.org/10.1109/CVPR.2014.474

Kisilevich, S., Krstajic, M., Keim, D., Andrienko, N., & Andrienko, G. (2010). Event-based analysis of people's activities and behavior using Flickr and Panoramio geotagged photo collections. In *Proceedings of the International Conference on Information Visualisation*. https://doi.org/10.1109/IV.2010.94

Kutner, M., Nachtsheim, C., & Neter, J. (2004). *Applied Linear Regression Models* (4th ed.). McGraw-Hill Irwin.

Law, S., Paige, B., & Russell, C. (2019). Take a look around: Using street view and satellite images to estimate house prices. *ACM Transactions on Intelligent Systems and Technology, 10*(5). https://doi.org/10.1145/3342240

Law, S., Seresinhe, C. I., Shen, Y., & Gutierrez-Roig, M. (2020). Street-Frontage-Net: Urban image classification using deep convolutional neural networks. *International Journal of Geographical Information Science, 34*(4), 681–707. https://doi.org/10.1080/13658816.2018.1555832

Li, J., Cheng, K., Wang, S., Morstatter, F., Trevino, R. P., Tang, J., & Liu, H. (2017). Feature selection: A data perspective. *ACM Computing Surveys*. https://doi.org/10.1145/3136625

Liu, Y., Singleton, A., & Arribas-Bel, D. (2020). Considering context and dynamics: A classification of transit-orientated development for New York City. *Journal of Transport Geography*. https://doi.org/10.1016/j.jtrangeo.2020.102711

Louppe, G. (2014). Understanding Random Forests from theory to practice [University of Liège]. *Cornell University Library*. https://arxiv.org/pdf/1407.7502.pdf.

Lu, Y. (2018). The association of urban greenness and walking behavior: Using google street view and deep learning techniques to estimate residents' exposure to urban greenness. *International Journal of Environmental Research and Public Health*. https://doi.org/10.3390/ijerph15081576

Lynch, K. (1960). *The Image of the City*. The MIT Press.

Molnar, C. (2019). 5.3 Accumulated Local Effects (ALE) Plot. In *Interpretable Machine Learning- A Guide for Making Black Box Models Explainable*. https://christophm.github.io/interpretable-ml-book/ale.html.

Naik, N., Philipoom, J., Raskar, R., & Hidalgo, C. (2014). Streetscore-predicting the perceived safety of one million streetscapes. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*. https://doi.org/10.1109/CVPRW.2014.121

Nasar, J. L. (1990). The Evaluative Image of the City. *Journal of the American Planning Association*. https://doi.org/10.1080/01944369008975742

Nistor, A., & Reianu, D. (2018). Determinants of housing prices: Evidence from Ontario cities, 2001–2011. *International Journal of Housing Markets and Analysis*. https://doi.org/10.1108/IJHMA-08-2017-0078

Ordnance Survey. (2020). *Order OS OpenData*. https://www.ordnancesurvey.co.uk/opendatadownload/products.html.

Pal, R. (2017). Overview of predictive modeling based on genomic characterizations. *Predictive Modeling of Drug Sensitivity*. https://doi.org/10.1016/b978-0-12-805274-7.00006-3

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

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., … Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research.*

Rosen, S. (1974). Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition Sherwin Rosen the. *Tetrahedron Letters*. https://doi.org/10.1016/S0040-4039(00)85403-9

Salesses, P., Schechtner, K., & Hidalgo, C. A. (2013). The Collaborative Image of The City: Mapping the Inequality of Urban Perception. *PLoS ONE, 8*(7). https://doi.org/10.1371/journal.pone.0068400

Scott, J. C. (1998). Seeing like a state: how certain schemes to improve the human condition have failed. In *Yale University Press.* https://doi.org/10.5860/choice.36-1224.

Seresinhe, C. I., Moat, H. S., & Preis, T. (2018). Quantifying scenic areas using crowdsourced data. *Environment and Planning B: Urban Analytics and City Science, 45*(3), 567–582. https://doi.org/10.1177/0265813516687302

Singleton, A. D., Spielman, S., & Folch, D. (2017). *Urban Analytics*. SAGE Publications.

Sirmans, G. S., Macpherson, D. A., & Zietz, E. N. (2005). The composition of hedonic pricing models. *Journal of Real Estate Literature*.

Smith, C. (2021). 20 Interesting flickr States and Facts (2021)| By the Numbers. https://expandedramblings.com/index.php/flickr-stats/.

UN-Habitat. (2020). *Addressing the Housing Affordability Challenge: A Shared Responsibility*. https://unhabitat.org/addressing-the-housing-affordability-challenge-a-shared-responsibility.

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

Wang, Y., Wang, S., Li, G., Zhang, H., Jin, L., Su, Y., & Wu, K. (2017). Identifying the determinants of housing prices in China using spatial regression and the geographical detector technique. *Applied Geography*. https://doi.org/10.1016/j.apgeog.2016.12.003

Wen, H., & Tao, Y. (2015). Polycentric urban structure and housing price in the transitional China: Evidence from Hangzhou. *Habitat International*. https://doi.org/10.1016/j.habitatint.2014.11.006

Wu, C., Ye, X., Ren, F., Wan, Y., Ning, P., & Du, Q. (2016). Spatial and social media data analytics of housing prices in Shenzhen, China. *PLoS ONE*. https://doi.org/10.1371/journal.pone.0164553.

Xiao, Y., Chen, X., Li, Q., Yu, X., Chen, J., & Guo, J. (2017). Exploring determinants of housing prices in Beijing: An enhanced hedonic regression with open access POI data. *ISPRS International Journal of Geo-Information*. https://doi.org/10.3390/ijgi6110358

Yang, J., Rong, H., Kang, Y., Zhang, F., & Chegut, A. (2021). The financial impact of street-level greenery on New York commercial buildings. *Landscape and Urban Planning*. https://doi.org/10.1016/j.landurbplan.2021.104162

Yao, Y., Zhang, J., Hong, Y., Liang, H., & He, J. (2018). Mapping fine-scale urban housing prices by fusing remotely sensed imagery and social media data. *Transactions in GIS, 22*(2), 561–581. https://doi.org/10.1111/tgis.12330

Zhang, F., Zhang, D., Liu, Y., & Lin, H. (2018). Representing place locales using scene elements. *Computers, Environment and Urban Systems, 71*(May), 153–164. https://doi.org/10.1016/j.compenvurbsys.2018.05.005

Zhang, Y., Hua, X., & Zhao, L. (2012). Exploring determinants of housing prices: A case study of Chinese experience in 1999–2010. *Economic Modelling*. https://doi.org/10.1016/j.econmod.2012.06.025

Zhang, Y., & Dong, R. (2018). Impacts of street-visible greenery on housing prices: Evidence from a hedonic price model and a massive street view image dataset in Beijing. *ISPRS International Journal of Geo-Information*. https://doi.org/10.3390/ijgi7030104

Zhou, B., Lapedriza, A., Khosla, A., Oliva, A., & Torralba, A. (2018). Places: A 10 Million Image Database for Scene Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 40*(6), 1452–1464. https://doi.org/10.1109/TPAMI.2017.2723009

Zhou, B., Liu, L., Oliva, A., & Torralba, A. (2014). Recognizing city identity via attribute analysis of geo-tagged images. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. https://doi.org/10.1007/978-3-319-10578-9_34