Expert Systems With Applications 259 (2025) 125276

Contents lists available at ScienceDirect

Expert Systems With Applications

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

Boosting house price estimations with Multi-Head Gated Attention

Zakaria Abdellah Sellam [a,*], Cosimo Distante [a], Abdelmalik Taleb-Ahmed [b], Pier Luigi Mazzeo [a]

[a] Institute of Applied Sciences and Intelligent Systems "Eduardo Caianiello", CNR, Lecce, Italy
[b] Laboratory of IEMN, CNRS, Centrale Lille, UMR 8520, Univ. Polytechnique Hauts-de-France, Universite Polytechnique Hauts de France, F-59313, Valencienne, France

ARTICLE INFO

Keywords:
House price evaluation
Gated attention
Spatial interpolation
Spatial analysis

ABSTRACT

Evaluating house prices is crucial for various stakeholders, including homeowners, investors, and policymakers. However, traditional spatial interpolation methods have limitations in capturing the complex spatial relationships that affect property values. To address these challenges, we have developed a new method called Multi-Head Gated Attention for spatial interpolation. Our approach builds upon attention-based interpolation models and incorporates multiple attention heads and gating mechanisms to better capture spatial dependencies and contextual information. Importantly, our model produces embeddings that reduce the dimensionality of the data, enabling simpler models like linear regression to outperform complex ensembling models. We conducted extensive experiments to compare our model with baseline methods and the original attention based interpolation model. The results show a significant improvement in the accuracy of house price predictions, validating the effectiveness of our approach. This research advances the field of spatial interpolation and provides a robust tool for more precise house price evaluation. Our GitHub repository.[1] contains the data and code for all datasets, which are available for researchers and practitioners interested in replicating or building upon our work.

1. Introduction

The Real Estate sector is a cornerstone of the global economy, with house prices as a critical indicator of economic health and individual wealth. Variations in house prices can influence consumer spending and economic growth, while declines can restrict borrowing capacity and reduce investments due to diminished collateral values (Case & Shiller, 2000). The 2008 financial crisis, triggered by a housing market collapse, highlights the critical importance of accurate house price predictions for economic stability (Reinhart & Rogoff, 2010). Predicting house prices is inherently complex, involving a myriad of factors. Traditional models for house price prediction have primarily used regression techniques, considering variables like property size, age, condition, and number of rooms (Bourassa, Hoesli, & Peng, 2003). However, advancements in machine learning have transformed this landscape. Methods such as support vector machines, decision trees, and neural networks have been increasingly employed, significantly enhancing prediction accuracy (Chen, Liaw, & Breiman, 2019). Ensemble learning techniques, particularly XGBoost, have shown great promise due to their ability to handle large datasets and model complex relationships (Chen & Guestrin, 2016a; Nguyen & Nguyen, 2023; Oyedotun, Olaniyi, Oyedotun, & Akin-Ojo, 2023). To address spatial heterogeneity, Geographically Weighted Regression (GWR) and related techniques have been widely used (Fotheringham, Brunsdon, & Charlton, 2002; Huang, Cai, & Wang, 2016; Li, Claramunt, & Ray, 2018; Wang, Ni, Tenenbaum, & Li, 2018). Additionally, geostatistical methods like Kriging have been employed for spatial interpolation, providing nuanced insights into geographical data (Chung, Venkatramanan, Elzain, Selvam, & Prasanna, 2019; Kang & Ma, 2017; Paez, Scott, & Volz, 2005). Despite these advancements, these methods often struggle with capturing complex spatial relationships and are sensitive to assumptions like isotropic variability, which may not hold in diverse landscapes. Furthermore, these models can be computationally intensive and sensitive to outliers. Our study advances these traditional approaches by integrating Multi-Head and Gated Attention mechanisms with similarity calculations, providing a robust framework for house price prediction. Our work builds upon the pioneering efforts of Viana and Barbosa (2021), who utilised an attention-based spatial interpolation model. They introduced a novel use of attention mechanisms, including a Euclidean-based layer and a Geo Attention layer,

* Corresponding author.
E-mail addresses: abdellah.sellam@isasi.cnr.it (Z.A. Sellam), cosimo.distante@cnr.it (C. Distante), Abdelmalik.Taleb-Ahmed@uphf.fr (A. Taleb-Ahmed), pierluigi.mazzeo@cnr.it (P.L. Mazzeo).

[1] Final_file/tree/main/ASI-main

https://doi.org/10.1016/j.eswa.2024.125276
Received 26 January 2024; Received in revised form 10 August 2024; Accepted 29 August 2024
Available online 3 September 2024

to weigh the influence of neighbouring properties. These attention layers produced 'house embeddings,' a compact representation capturing structural and spatial features for regression models. In our model, we enhance this approach by distinctly separating Geographical and Structural Attention mechanisms. Geographical Attention focuses on spatial relationships and proximities, while Structural Attention captures intrinsic property attributes such as size, age, and condition. This dual attention mechanism, combined with a similarity calculation, ensures that the model accurately reflects properties' physical and locational characteristics. The Multi-Head Attention mechanism further allows for parallel processing, capturing various aspects of spatial relationships at different scales. The Gated Attention mechanism fine-tunes the information flow, reducing the impact of outliers and ensuring robust predictions. The superiority of our model lies in its reliance on basic geographical and structural data, making it applicable even in scenarios where multimodal data like images are unavailable. This contrasts with other recent studies that require complex and sometimes difficult-to-obtain data inputs. For instance, the study "Imbalanced Multimodal Attention Based System for Multiclass House Price Prediction" (Smith & Jones, 2023) and "Joint Gated Co Attention Based Multi-Modal Networks for Subregion House Price Prediction" (Lee & Kim, 2023) focus on incorporating multimodal data, which can be challenging to collect and label consistently. Our model's simplicity and accessibility, combined with its advanced attention mechanisms, make it a versatile and powerful tool for various real estate prediction tasks. Recent literature demonstrated substantial improvements in capturing complex spatial dependencies and reducing data dimensionality, resulting in more accurate predictions. Our approach stands out by using Multi-Head Gated Attention and integrating similarity calculations, which enhance the model's ability to discern relevant features from geographical and structural data.Our contributions to advancing real estate price prediction include:

- **Introducing a Comprehensive New Dataset:** We have created a comprehensive new dataset that includes data from different Italian cities. It combines both structural and geographical information relevant to real estate valuation. This dataset aims to offer a more complete understanding of the factors that affect property prices, making it possible to create more precise and reliable predictive models.
- **Incorporating Advanced Attention Mechanisms:** Our model uses Multi-Head and Gated Attention mechanisms, improved by a similarity-based filtering approach, to capture detailed structural and geographical contexts accurately. These attention mechanisms allow the model to concentrate on the most important features, resulting in more accurate predictions by effectively handling the high-dimensional data typically found in real estate analytics.
- **Embedding Generation with Spatial Interpolation:** We used the Multi-Head Gated Attention Spatial Interpolator model to create embeddings that improve the predictive capability of our approach. These embeddings decrease the data's complexity, enabling simpler models like linear regression to achieve performance levels similar to those of more advanced models. They significantly improve the model's predictive accuracy and efficiency by capturing intricate relationships within the data.
- **Validation Across Diverse Datasets:** We thoroughly validated our model across diverse datasets representing different geographic and economic contexts. This comprehensive testing showcases the adaptability and efficacy of our approach, emphasising its practical usefulness in real-world scenarios. The model's reliable performance across various datasets underscores its resilience and trustworthiness, positioning it as a valuable tool for precise predictive modelling in the real estate industry.

The subsequent sections of this document are structured as follows: Section 2 presents a comprehensive overview of relevant works, including literature and methodologies, that relate to house price estimation and Section 3 delves into our proposed attention network, detailing its unique features and potential benefits. In Section 4, we conduct experiments, perform data analysis, and provide a thorough evaluation of our model. Lastly, in Section 5, we draw insightful conclusions based on our experimentation, compare our approach with prior methodologies, and articulate the implications of our findings. This structure ensures a coherent and comprehensive understanding of our innovative methodology for house price prediction.

2. Related works

House price estimation is a critical activity with far-reaching implications for the real estate industry. This field has been the subject of extensive academic research, traditionally employing regression analyses that integrate multiple variables, data types, and methodologies. In this review, we explore the scholarly landscape of this subject, tracing the evolution of research methodologies and spotlighting modern advancements and emerging trends.

The Hedonic Price Theory, first introduced by Rosen in 1974 (Rosen, 1974), is the foundation for Hedonic Regression models. These models have become a crucial tool in studying house prices. The theory utilises a set of attributes, such as the number of bedrooms or bathrooms, to explain and represent a house's market value. These attributes are ranked based on their impact on a house's utility function, assuming that a market equilibrium between buyers and sellers determines the sale price. Hedonic Regression models are widely used to analyse the effects of different factors on house prices in various areas, making them a robust tool for market segmentation (Yao, Zhang, Hong, Liang, & He, 2018). Although the original Hedonic Price Theory focused mainly on the intrinsic characteristics of a house, it has evolved to account for external factors like location (Frew & Wilson, 2002). This adaptation was motivated by the realisation that solely considering a house's intrinsic attributes was insufficient for accurate price representation (Limsombunchai, Gan, & Lee, 2004). Despite its widespread use, Hedonic Regression models have faced challenges, including issues related to the stability of attribute coefficients across different locations and property types and limitations in handling non-linearity and model specification (Wang, Wen, Zhang, & Wang, 2014).

The integration of machine learning into house price prediction has been significantly accelerated by advancements in computational capabilities and the increase of data (Chaphalkar & Sandbhor, 2013). Initially, the focus was mainly on traditional machine learning algorithms such as Linear Regression (LR) (Cook, 1977). While these linear models offered computational efficiency and ease of interpretation, they were limited in capturing the high dimensional and non-linear complexities inherent in transaction price data. Researchers explored regularisation techniques like Ridge and Lasso Regression (Hoerl & Kennard, 1970; Tibshirani, 1996) to address these limitations. These methods helped mitigate overfitting and offered a more refined approach to feature selection but struggled with capturing complex, non-linear relationships. Principal Component Analysis (PCA) (Jolliffe, 1986) has also been employed for dimensionality reduction to simplify the feature space, although it has been criticised for potentially discarding crucial information. This led to the exploration of more flexible, non-linear models such as Support Vector Regression (SVR) (Drucker, Burges, Kaufman, Smola, & Vapnik, 1997) and Decision Trees (Quinlan, 1986a). Support Vector Regression (SVR) offers a solution for non-linearities through various kernel functions, while Decision Trees provide a simple yet effective approach for detecting non-linear patterns (Drucker et al., 1997; Quinlan, 1986a). However, Decision Trees are prone to overfitting. To combat this, ensemble methods like Random Forests were developed to improve model generalisation (Ho, 1995). Random

Forests combine the outcomes of many decor-related trees to minimise variance and enhance accuracy.

With advancements in computational power, the field has shifted to more sophisticated ensemble methods such as XGBoost (Chen & Guestrin, 2016b). Unlike Random Forests, XGBoost constructs trees sequentially to correct the errors made by the previous ones. This makes XGBoost particularly effective in handling diverse data structures and enhancing prediction accuracy (Pavlyshenko, 2018). These advanced ensemble models are also highly scalable and efficient, often surpassing Random Forests' performance on large datasets.

To further optimise their predictive performance, these sophisticated ensemble models are often fine-tuned using metaheuristic optimisation techniques like Particle Swarm Optimization (PSO) (Alfyatin, Febrita, Taufq, & Mahmudy, 2017; Claesen & De Moor, 2015). These optimisation techniques enable precise tuning of hyperparameters, resulting in accurate and computationally efficient models.

The latest development in house price prediction is Graph Neural Networks (GNNs) (Zhou et al., 2020), which excel in identifying spatial relationships between properties. However, GNNs can be computationally demanding and require large, well-curated datasets for practical training. Additionally, their performance can vary significantly based on the architecture and hyperparameters, which may hinder their widespread adoption. Graph Neural Networks (GNNs) have recently emerged as a prominent method for house price prediction, leveraging their ability to capture intricate spatial relationships between properties (Zhou et al., 2020). However, GNNs face several challenges, including high computational demands and the need for large, well-curated datasets, making them resource intensive (Anonymous, 2021; Piechocki & Pope, 2024). Additionally, their effectiveness can be highly dependent on the choice of architecture and hyperparameters, which may hinder widespread adoption. Furthermore, GNNs often require customised graph structures for each dataset, limiting their flexibility across different data types (Anonymous, 2021; Piechocki & Pope, 2024).

Furthermore, the domain has seen the rise of deep learning techniques. Deep Neural Networks (DNNs) (Schmidhuber, 2015) can automatically learn feature representations, eliminating the need for manual feature engineering. Although DNNs can unravel highly complex relationships in the data, they present challenges, such as the risk of overfitting and the need for substantial datasets and computational resources for practical training.

Building on these advancements, recent research has focused on integrating diverse computational models and data sources. A prime example is a groundbreaking study by Tchuente and Nyawa (2022) on the French real estate market. Utilising machine learning techniques such as Random Forest, AdaBoost (Freund & Schapire, 1997), and gradient boosting (Friedman, 2001), along with geocoding features, they analysed five years of historical real estate transactions provided by the French government. Their findings revealed that incorporating geocoding elements increased the models' predictive accuracy by over 50%.

Building upon the findings of Tchuente et al. the research conducted by Zhao, Ravi et al. (2022) represents a significant advancement in data analysis. By incorporating a multi-modal approach encompassing traffic patterns, amenities, and social emotions in the bustling city of Beijing, China, this study validated the crucial role of location-based data. Furthermore, it introduced a feature ranking mechanism that established a direct correlation between the data and its economic impact. This groundbreaking research underscores the potential of geolocated data in predicting real estate prices and highlights its transformative capabilities. Further advancing this research domain, De Nadai and Lepri (2018) delved into the economic repercussions of neighbourhood characteristics within Italian urban landscapes. Their investigative toolkit encompassed a rich array of data sources including OpenStreetMap,[2] Urban Atlas 2012, imagery from Google Street View, Italian census data,[3] alongside property tax records sourced from the "Immobiliare. it"[4] platform. Through the application of their model, they witnessed a notable 60% enhancement in nowcasting housing prices, thereby underpinning the transformative potential of leveraging rich, geolocated datasets.

Das, Ali, Li, Kang, and Sellis (2021) introduced the concept of Geospatial Network Embedding (GSNE) to address the geospatial context of neighbourhood amenities in house price predictions. Unlike traditional models, GSNE captures the influence of proximity to key points of interest (POIs) such as train stations, highly ranked schools, and shopping centres. By leveraging graph neural networks, GSNE creates embeddings of houses and various types of POIs within multipartite networks, representing relationships as edges. This method allows for understanding complex interactions between houses and POIs, offering a robust way to incorporate geospatial context into price predictions. Despite its innovations, GSNE faces several limitations. Firstly, the high computational complexity of processing large-scale data and multiple types of POIs makes it less suitable for real-time applications or environments with limited computational resources. Secondly, the model's reliance on high-quality, comprehensive geospatial data means that any incomplete or inaccurate data can significantly degrade its performance. Thirdly, scalability issues can arise when applying GSNE to larger datasets or different geographic regions, as the embedding and training processes are resource-intensive and time-consuming. Additionally, while the model performs well within the scope of the datasets used in the study, its generalisability to other regions with different market dynamics may be limited, potentially requiring specific adjustments. Finally, the complexity of interpreting the GSNE model presents challenges for stakeholders who need transparent and explainable decision-making models.

Wang, Wang, Wu, and Du (2022) introduced the Geographically Neural Network Weighted Regression (GNNWR) to enhance house price predictions by incorporating spatial heterogeneity. Unlike traditional Geographically Weighted Regression (GWR) models, GNNWR integrates neural networks to capture complex spatial relationships better, using geographical data to create weights for different properties and improve prediction accuracy. Despite its advantages, the GNNWR model has several limitations. Firstly, it requires significant computational power due to the complexity of the neural network and the large number of features, making it less suitable for environments with limited resources. Secondly, GNNWR relies on high-quality, comprehensive datasets; incomplete or inaccurate data can significantly degrade its performance. Thirdly, scalability issues may arise when applying GNNWR to larger datasets or different geographic regions, as computational demands increase with dataset size. Additionally, while GNNWR shows promising results in the Shenzhen dataset, its generalisability to other regions with different housing market dynamics remains to be determined and may require specific adjustments. Finally, the complexity of GNNWR can pose challenges for stakeholders who need transparent and explainable models for decision-making.

Kang et al. (2021) delve into house price appreciation rates, employing a multi-source extensive geo-data framework that amalgamates structural attributes, locational amenities, and visitor patterns, employing machine learning models and geographically weighted regression for accurate predictions at both micro and macro scales. Their gradient-boosting machines achieve an R-squared value of 74% at the neighbourhood scale, highlighting the effectiveness of their approach in understanding house price appreciation nuances. On a similar innovative trajectory, Wang, Chen, Su, Wang, and Huang (2021). Propel house price prediction forward by harnessing a Joint

[2] europe/italy.html.
[3] https://www.istat.it/.
[4] www.immobiliare.it.

Self-Attention Mechanism intertwined with a rich analysis of heterogeneous data, including public facilities and environmental aesthetics captured through satellite imagery. Tested in Taipei and New Taipei, this model eclipses other machine learning-based models in prediction accuracy, showcasing a lower error rate. The Spatial Transformer Network (STN) (Jaderberg, Simonyan, Zisserman, & Kavukcuoglu, 2015) and their model's novel joint self-attention mechanism intricately dissect the complex relations between different attributes impacting house prices. This work accentuates the necessity of a holistic data-rich approach and extends the versatility of the attention mechanism across various domains, setting a robust foundation for future research. In a parallel vein, Viana and Barbosa (2021) introduce a groundbreaking framework that melds the spatial essence of real estate with the structural attributes of houses. Their hybrid attention mechanism orchestrates a balanced blend between the Euclidean space of structural features and the geographic tapestry, crafting them into a unified predictive model. The inception of a house embedding vector carries through the regression analysis domain, offering a fresh lens to capture spatial dependencies. This attention-infused approach heralds a promising avenue where the convergence of spatial interpolation and machine learning unravels a richer understanding of housing market dynamics, further amplifying the potential of attention mechanisms in elucidating the multifaceted nature of house price predictions. The related work showcases a trajectory towards crafting more nuanced, robust, and insightful real estate price prediction models. These models progressively harness multi-source, geolocated data and sophisticated machine learning techniques, notably attention mechanisms. This evolution reflects a maturing field poised to address the intricate challenges inherent to urban landscapes and real estate markets.

3. Methodology

Our proposed methodology aims to create robust house embeddings by assessing the similarity between a specific house and its neighbouring properties. This approach goes beyond merely considering individual property attributes and geographical location. Instead, it encapsulates each house's local characteristics with its immediate surroundings. Unlike traditional methods, we integrate the geographical coordinates of the property to refine this embedding further, capturing the essence of its surroundings and their relation to critical landmarks or amenities.

Our approach is based on the Attention-Based Spatial Interpolation (ASI) architecture proposed by Viana and Barbosa (2021). This architecture creates geographical and Euclidean similarities and emphasises specific similar points using an attention mechanism. However, more than a simple attention head may be required to capture differentiated interrelations. For this reason, our model employs multi-head-gated attention mechanisms to optimise the extraction of these features and their interrelationships. Multi-Head Gated Attention allows the model to capture multiple contexts, such as architectural styles, proximity to amenities, and other relevant features. Concurrently, the gated attention mechanism controls the flow of information to ensure that only the most pertinent attributes are considered. This is particularly useful when there is a significant variance between the target house and its neighbours, allowing the model to focus on the most critical similarities or differences. The Euclidean Multi-Head Gated Attention layer, represented in Fig. 1(A), calculates attention weights for the structural features of neighbouring houses based on their Euclidean distance to A_i. Concurrently, the Geographical Multi-Head Gated Attention layer in Fig. 1(B) learns the spatial correlations between the n-nearest geographical neighbours of house i. The output vectors from both attention layers are concatenated with A_i and G_i and fed into a fully connected neural network, culminating in a regression layer. This architecture synthesises the influence of the neighbouring houses and the target house's attributes into a single vector, termed the "house embedding" illustrated in Fig. 1.

3.1. Background knowledge

To perform predictive analysis in real estate valuation, it is crucial to have a solid foundation of knowledge. This field employs a variety of methodologies and algorithms that are based on fundamental principles and metrics. Understanding these concepts is essential for accurately performing advanced analytical techniques. This subsection aims to clarify some of these key concepts and metrics, providing a starting point for a deeper exploration and comprehension of the subsequent methodologies and evaluations.

3.1.1. Similarity calculation

In the intricate landscape of data science, similarity is a critical underpinning for various algorithms and methodologies. This subsubsection aims to illuminate the key metrics ubiquitously employed to quantify similarity, laying the groundwork for the following analyses.

- **Euclidean Distance**: A foundational metric in geometry, Euclidean distance provides a straightforward measure of similarity by calculating the straight-line distance between two points in an Euclidean space.

$$d(P_1, P_2) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \tag{1}$$

- **Cosine Similarity**: This metric is invaluable in high-dimensional spaces, measuring the cosine of the angle between two vectors. It is especially pertinent in text analysis and natural language processing.

$$\text{Cosine Similarity} = \frac{C \cdot D}{\|C\| \times \|D\|} \tag{2}$$

- **Jaccard Index**: A set-based metric, the Jaccard Index is helpful for categorical data, quantifying the ratio of the intersection to the union of two sets.

$$J(C, D) = \frac{|C \cap D|}{|C \cup D|} \tag{3}$$

- **Identity Similarity**: This is a binary similarity measure used to ascertain whether or not two data points are identical. Unlike continuous similarity measures, the Identity Similarity scores 1 if the data points are similar and 0 if they differ. This measure is handy in scenarios requiring exact matching or where data is categorical. Mathematically, it is expressed as:

$$S(x, y) = \begin{cases} 1 & \text{if } x = y \\ 0 & \text{if } x \neq y \end{cases} \tag{4}$$

where:
$S(x, y)$ is the similarity score between data points x and y, and the score is 1 if the data points x and y are identical, and 0 if they are different.

- **Gaussian Kernel**: Also known as the Radial Basis Function (RBF) with Gaussian form, this metric is a cornerstone in non-linear data transformations. Unlike other metrics that measure distance directly, the Gaussian Kernel calculates similarity by mapping the original data points into a higher dimensional space through a Gaussian function. This allows it to capture complex, non-linear relationships between data points. Mathematically, it is expressed as:

$$K(x, y) = \exp\left(-\frac{\|x - y\|^2}{2\sigma^2}\right) \tag{5}$$

The parameter σ controls the spread of the Gaussian function, thereby influencing the similarity measure. A smaller σ will result in a narrower Gaussian function, making the similarity measure more sensitive to the distance between data points.

These metrics serve as the backbone for various algorithms and offer a nuanced understanding of how data points relate to each other in complex spaces, with the Gaussian Kernel standing out for its ability to capture non-linear relationships.

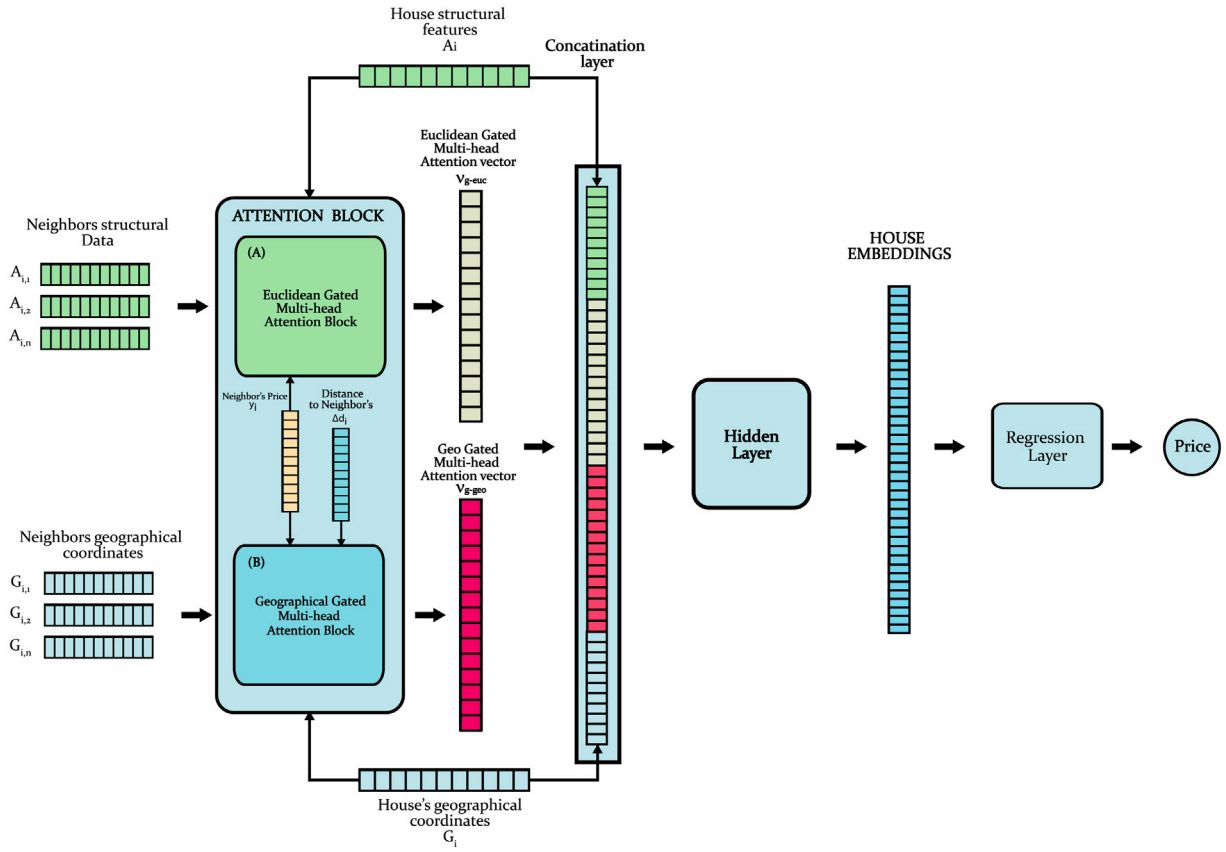

Fig. 1. Architecture representation of the multi-head gated attention-based interpolation. (A) Represent the Euclidean interpolation block based on the multi-head gated Attention. (B) Represent the geo-interpolation block based on the Multi-Head Gated attention.

3.1.2. Spatial interpolation

Spatial interpolation is a critical technique for predicting unknown values at unobserved locations based on known values at observed locations, finding applications in diverse fields such as geostatistics, environmental science, and real estate. The effectiveness of spatial interpolation is intrinsically tied to the choice of similarity measures. For instance, Euclidean distance can be employed in a straightforward approach like "inverse distance weighting" (IDW) (Shepard, 1968), where the influence of a neighbouring point on the interpolated value is inversely proportional to its Euclidean distance from the target location. On the other hand, the Gaussian Kernel (You, Pang, Cao, & Luo, 2017) offers a more nuanced approach by transforming the Euclidean distance into a measure of similarity, thereby capturing complex, non-linear spatial relationships. This is especially useful in advanced geostatistical methods like kriging (Matheron, 1969). Therefore, the choice between straightforward measures like Euclidean distance and more complex ones like the Gaussian Kernel can significantly impact the quality of spatial interpolation, exemplifying the broader applicability and importance of similarity measures in data science.

3.1.3. Attention mechanisms

Attention mechanisms (Vaswani et al., 2017) have emerged as a cornerstone in many deep learning models, predominantly in sequence-to-sequence tasks such as machine translation and speech recognition. The essence of Attention is to emulate the human ability to focus on specific segments of input data, much like how we selectively concentrate on some aspects of a visual scene or a conversation. Among the diverse attention mechanisms, Soft Attention is a mechanism that computes a weighted sum of all input values. These weights, indicative of the relevance of each input, are typically determined using a softmax function, ensuring a normalised distribution where the weights sum up to one. The continuous nature of these weights makes soft Attention inherently differentiable, rendering it particularly amenable to gradient-based optimisation techniques (Bahdanau, Cho, & Bengio, 2014). On the other hand, intricate Attention operates more selectively. Instead of distributing focus across all inputs, it zeroes in on a specific subset, effectively sidelining the others. Given its discrete selection process, traditional backpropagation struggles with optimising intricate Attention. Yet, this challenge is surmountable with techniques like the reinforce algorithm (Mnih, Heess, & Graves, 2014). The *Gated Attention* mechanism (Zhang et al., 2018) bridges the gap between these two. It adeptly amalgamates information from diverse sources and employs gating tools to ascertain the relevance of each source. This approach can be perceived as a harmonious blend of the soft and hard attention paradigms, encapsulating the strengths while mitigating their limitations (Luong, Pham, & Manning, 2015).

3.2. Attention block

The Attention Block is the computational nucleus of our architecture, designed to intricately capture the spatial relationships essential for precise house price prediction. As delineated in Fig. 1, this block comprises two main components: the Geo Multi-head Gated Attention and the Euclidean Multi-head Gated Attention. Each of these components consists of several key stages, contributing to generating their respective geo- and Euclidean-gated attention vectors. Fig. 2 elucidates the fundamental principles for calculating the Geo and Euclidean attention mechanisms. In the initial stage, represented by Fig. 2(A), the Distance Calculation Block computes the distance between the target house and its neighbours. The nature of this distance is contingent on the specific attention mechanism in play, be it Geo or Euclidean. The Similarity Calculation Block, as depicted in Fig. 2(B), transforms these distances into similarity scores. A Gaussian kernel function is employed for Geo Attention, while alternative kernel functions may

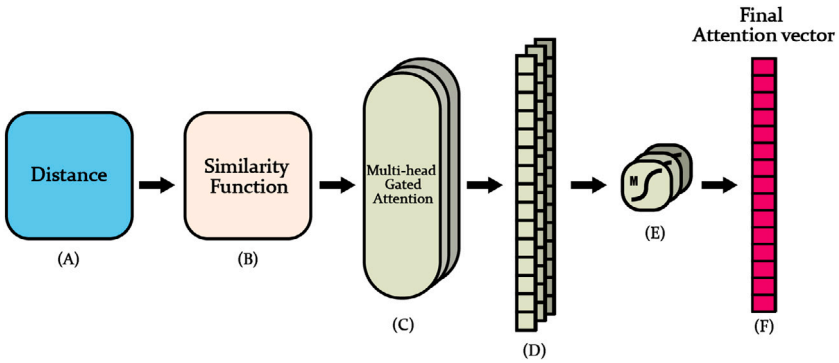

Fig. 2. Comprehensive overview of the Gated Multi-head Attention mechanism within the Attention Block. (A) depicts the initial computation of geodesic and Euclidean distances, serving as the foundation for subsequent attention calculations. (B) illustrates the Similarity Function, which transforms these foundational distances into similarity scores. (C) shows the core Multi-Head Gated Attention Block, where these similarity scores derive gated attention weights across multiple heads. (D) Highlights the Aggregated Attention Head, consolidating the gated attention weights from all heads into a singular vector. (E) represents the aggregation of multiple gated attentions for each weighted sum. (F) Indicates the Final Attention Vector.

be used for the Euclidean variant. The subsequent component is the multi-head gated Attention, illustrated in Fig. 2(C). This block leverages the similarity scores to derive attention weights, which are then gated to modulate their influence. The entire process is executed across multiple heads, capturing various facets of the spatial relationships between the target house and its neighbours. Next, the aggregated attention head, represented by Fig. 2(D), consolidates the outputs from all attention heads into a single vector. This is achieved through a weighted sum, where the weights are adaptively learned during training. If the architecture employs multiple attention mechanisms, such as Geo and Euclidean, their aggregated attention heads are combined further. Following this, Fig. 2(E) illustrates the Final Aggregation Block. Aggregated normalised gating weights are computed using a softmax function in this stage. After that, the weighted sums produced from each attention head are multiplied by these normalised weights. This aggregation is performed separately for the Geo and Euclidean attention mechanisms, resulting in their aggregated attention vectors. Finally, the vector produced from this aggregation process is the final attention vector, as depicted in Fig. 2(F). In summary, the Attention Block encapsulates the Multi-head Geo Gated Attention and the Euclidean Multi-head Gated Attention, generating their respective Geo and Euclidean Gated Attention Vectors.

3.2.1. Geo multi-head gated attention

The Geo Multi-head, Gated Attention mechanism, is designed to capture the spatial relationships between a target house and its neighbouring properties. This involves using a Gaussian kernel function to calculate geographic similarity scores between the target house and its neighbours.

Eq. (6) demonstrates how the geographic score between the target house G_i and its neighbouring house $G_{i,j}$ is computed:

$$s(G_i, G_{i,j}) = \exp\left(-\text{geo\_dist}(G_i, G_{i,j}) \times \rho\right) \tag{6}$$

The geodesic distance, $\text{geo\_dist}(G_i, G_{i,j})$, represents the shortest path over the Earth's surface, measured along the surface curvature. This distance can be calculated using the Haversine formula, which accounts for the Earth's spherical shape:

$$\text{geo\_dist}(G_i, G_{i,j}) = 2r \arcsin\left(\sqrt{\sin^2\left(\frac{\Delta\phi}{2}\right) + \cos(\phi_1)\cos(\phi_2)\sin^2\left(\frac{\Delta\lambda}{2}\right)}\right) \tag{7}$$

Here, r is the Earth's radius (mean radius = 6371 km), $\Delta\phi = \phi_2 - \phi_1$ is the difference in latitude between the two points in radians, and $\Delta\lambda = \lambda_2 - \lambda_1$ is the difference in longitude between the two points in radians. ϕ_1 and ϕ_2 are the latitudes of the two points in radians, and λ_1 and λ_2 are the longitudes of the two points in radians. Here, $\rho = \frac{\sigma^2}{2}$ is a scaling factor, where σ is a parameter controlling the width of the Gaussian kernel. The geodesic distance, $\text{geo\_dist}(G_i, G_{i,j})$, as previously defined, represents the shortest path over the Earth's surface between the target house G_i and its neighbouring house $G_{i,j}$. The vector of similarity scores L is formed by aggregating the similarity scores $s(G_i, G_{i,j})$ for all neighbours $G_{i,j}$ of the target house G_i. The equation represents the aggregation :

$$L = \sum_{j=1}^{n} s(G_i, G_{i,j}) \tag{8}$$

This vector L is then transformed into a hidden representation H' through a fully-connected layer, as described in Eq. (9):

$$H' = W' \cdot L + b' \tag{9}$$

In this equation, W' and B are the learned weights and bias terms, respectively. The attention weights a_{geo} are computed using a softmax layer, as formulated in Eq. (10):

$$a_{\text{geo}}(G_i, G_{i,j}) = \frac{\exp(H'_j)}{\sum_{j'=1}^{n} \exp(H'_{j'})} \tag{10}$$

Then, using our defined gated attention mechanism (Eq. (11)), we apply it to the attention weights:

$$\text{Gate}(x) = \sigma(W_g \cdot x + b_g) \tag{11}$$

Subsequently:

$$a'_{\text{geo}}(G_i, G_{i,j}) = \text{Gate}(a_{\text{geo}}(G_i, G_{i,j})) \odot a_{\text{geo}}(G_i, G_{i,j}) \tag{12}$$

where:

- x is the input value, in this case, the original attention weight $a_{\text{geo}}(G_i, G_{i,j})$.
- W_g represents the learned weight matrix associated with the gate.
- b_g denotes the bias term.
- σ is the sigmoid function, ensuring the output value of the gate lies in the [0,1] range.

With this, the Geo Gated Attention Vector $v_{\text{ggeo}}(G_i)$ is computed as a weighted sum of the features of the neighbouring houses using the modified attention weights a'_{geo}:

$$v_{\text{ggeo}}(G_i) = \sum_{j=1}^{n} a'_{\text{geo}}(G_i, G_{i,j})[G_{i,j} \oplus A_{i,j} \oplus \Delta d_{i,j} \oplus y_{i,j}] \tag{13}$$

In this equation, $\Delta d_{i,j}$ represents the geographic distance between house i and its neighbour j. Similarly, $y_{i,j}$ signifies the price of the

neighbour j, and $\oplus$ denotes the concatenation operation. The dimensionality of $v_{\text{geo}}(G_i)$ is derived from the summation of dimensions where $G_{i,j} \in \mathbb{R}^2$, $A_{i,j} \in \mathbb{R}^T$, $\Delta d_{i,j} \in \mathbb{R}^1$, and $y_{i,j} \in \mathbb{R}^1$. Consequently, the vector $v_{\text{geo}}(G_i)$ can be viewed as a weighted sum of vectors $G_{i,j}$, concatenated with $\Delta d_{i,j}$ and $y_{i,j}$, and weighted using the normalised geo gated attention coefficients which are determined during the training process.

3.2.2. Euclidean multi-head gated attention

The Euclidean Multi-head Gated Attention mechanism is precisely engineered to emphasise the most relevant structural similarities between a target house and its neighbouring properties. This mechanism employs the Euclidean distance to compute the similarity scores between the target house and its neighbours. The Euclidean distance between the target house A_i and a neighbouring house $A_{i,j}$ is computed as shown in Eq. (14):

$$d(A_i, A_{i,j}) = \sqrt{\sum_{p=1}^{T} (a_{i,p} - a_{i,j,p})^2} \tag{14}$$

where $d(A_i, A_{i,j})$ is the Euclidean distance indicating similarity between houses based on structural attributes, A_i represents the structural features of the target house i, $A_{i,j}$ denotes the structural features of the jth neighbouring house to i, $a_{i,p}$ and $a_{i,j,p}$ are specific structural attributes of houses i and j, respectively, and T is the total number of structural attributes considered.

After computing the Euclidean distances, we construct a vector of similarity scores L, which is then transformed into a hidden representation H through a fully connected layer, as described in Eq. (15):

$$H = W \cdot L + b \tag{15}$$

In Eq. (15), W and b are the learned weights and bias terms, respectively. The attention weights a_{euc} are computed using a softmax layer, as formulated in Eq. (16):

$$a_{\text{euc}}(A_i, A_{i,j}) = \frac{\exp(H_j)}{\sum_{j'=1}^{n} \exp(H_{j'})} \tag{16}$$

The essence of the gated attention mechanism is to refine the attention weights by introducing an additional modulation step. This modulating factor, or "gate", is typically represented as a value between 0 and 1 and is applied element-wise to the attention weights. The purpose is to amplify or diminish the original attention values based on the model's learned parameters.

Given this, the gated attention can be defined as:

$$\text{Gate}(x) = \sigma(W_g \cdot x + b_g) \tag{17}$$

where:

- x is the input value, in this case, the original attention weight $a_{\text{euc}}(A_i, A_{i,j})$.
- W_g represents the learned weight matrix associated with the gate.
- b_g denotes the bias term.
- σ is the sigmoid function, ensuring the output value of the gate lies in the [0,1] range.

Subsequently, the gated attention mechanism can be formalised as:

$$a'_{\text{euc}}(A_i, A_{i,j}) = \text{Gate}(a_{\text{euc}}(A_i, A_{i,j})) \odot a_{\text{euc}}(A_i, A_{i,j}) \tag{18}$$

Here, $\odot$ denotes element-wise multiplication. Thus, the attention weight is modulated by its gating value, allowing the model to allocate attention more selectively to houses exhibiting the most congruent features.

The Vector with Euclidean Gated Attention, denoted as $v_{\text{geuc}}(A_i)$, represents a cumulative weighted mix of attributes from the surrounding homes. This process uses the gated attention coefficients a'_{euc} and is illustrated in Eq. (19):

$$v_{\text{geuc}}(A_i) = \sum_{j=1}^{n} a'_{\text{euc}}(A_i, Ai, j) \odot [A_{i,j} \oplus y_{i,j}] \tag{19}$$

Within Eq. (19), $y_{i,j}$ defines the price of the jth neighbouring home of house i, while $\oplus$ denotes the concatenation action. The size of $v_{\text{euc}}(A_i)$ stands at $T + 1$ given that $A_{i,j}$ resides in $\mathbb{R}^T$ and $y_{i,j}$ is part of $\mathbb{R}^1$. The composition of $v_{\text{euc}}(A_i)$ involves initially multiplying the combined vector $[A_{i,j} \oplus y_{i,j}]$ for each jth neighbour of house i by its respective gated attention coefficient $a'_{\text{euc}}(A_i, A_{i,j})$, producing an individual weighted vector for every jth neighbour. An overall summation is then applied to these vectors for all n adjacent homes to house i. Consequently, the elements within $v_{\text{euc}}(A_i)$ represent a comprehensive weighted sum of the structural attributes and the valuations of the nearby homes of house i. The gated attention coefficients undergo refinement during the learning phase.

3.2.3. Final aggregation block

The final aggregation stage shown in Fig. 2E involves collecting and combining the attention vectors from each head of the attention mechanism and applying the gated attention based on the normalised gates weights and biases. It is important to note that this process is unique for each attention mechanism, namely Geo and Euclidean, and it results in the formation of two separate aggregated attention vectors.

To ensure the effectiveness of the attention mechanism in both Geo and Euclidean interpolation, it is crucial to normalise the gating weights and biases using a softmax function, as shown in Eq. (20). By normalising the gating weights and biases, they fall within the range of 0 to 1, which makes them more easily interpretable.

$$\text{gate}_{\text{norm},i} = \frac{\exp(\text{Gate\_weights}_i + \text{Gate\_bias}_i)}{\sum_{j=1}^{n} \exp(\text{Gate\_weights}_j + \text{Gate\_bias}_j)} \tag{20}$$

After normalising the gating weights and biases, we perform element-wise multiplication with each attention and then aggregate them. The resulting vector that shows the aggregated gated geographic attention, denoted as $v_{\text{agg\_ggeo}}$, is presented in Eq. (21).

$$v_{\text{agg\_ggeo}} = \sum_{i=1}^{n} \text{gate}_{\text{norm, geo},i} \odot v_{\text{ggeo},i} \tag{21}$$

where $\text{gate}_{\text{norm, geo},i}$ represents the softmax-normalised gating weights and biases, and $v_{\text{geo},i}$ refers to the attention vectors from the Geo attention heads.

In a similar vein, the aggregated gated Euclidean attention vector $v_{\text{agg\_geuc}}$ is represented by Eq. (22):

$$v_{\text{agg\_geuc}} = \sum_{i=1}^{n} \text{gate}_{\text{norm, euc},i} \odot v_{\text{geuc},i} \tag{22}$$

Here, $\text{gate}_{\text{norm, euc},i}$ signifies the softmax-normalised gating weights, and $v_{\text{geuc},i}$ portrays the gated attention vectors emergent from the Euclidean attention heads.

In conclusion, the consolidated Geo attention vector $v_{\text{agg\_ggeo}}$ and the Euclidean attention vector $v_{\text{agg\_geuc}}$ are computed using an element-wise multiplication between the softmax-normalised gating weights and their corresponding attention vectors as illustrated in Fig. 2F derived from the Geo and Euclidean attention heads, respectively. This approach ensures an accurate integration of the significance associated with each feature and reflects the complex spatial relationships inherent within the Geo and Euclidean contexts.

3.3. House embeddings

Embeddings are a pivotal component in contemporary machine-learning architectures, especially in scenarios that involve manipulating high-dimensional or categorical variables. In the realm of real estate price prediction, the utility of embeddings is accentuated for the encoding of categorical attributes such as neighbourhood classifications, types of properties, and associated amenities into a continuous vector space (Smith & Doe, 2021). These continuous embeddings can capture intricate relationships between disparate categories, thereby augmenting the predictive efficacy of the machine learning model (Mikolov et al., 2013; Pennington, Socher, & Manning, 2014). The transformation from a sparse, high-dimensional feature space to a dense, lower-dimensional vector space has found applications across a multitude of domains, ranging from natural language processing to recommendation engines and graph-based machine learning algorithms (Cai, Zheng, & Chang, 2017; Devlin et al., 2018; Koren, Bell, & Volinsky, 2009; Peters et al., 2018). However, effectively utilising embeddings necessitates meticulous tuning and validation to mitigate the risk of overfitting and ensure robust generalisation on unseen data (Chiu & Korhonen, 2019). In the present study, as delineated in Fig. 1, we introduce a novel methodology for generating house embeddings. Initially, two distinct Multi-Head Gated Attention mechanisms are employed: one geographically oriented (Geo Multi-Head Gated Attention) and another focused on structural attributes (Euc Multi-Head Gated Attention). The Geo Multi-Head, Gated Attention mechanism leverages the geographical coordinates of proximate properties, while the Euc Multi-Head Gated Attention mechanism utilises the structural attributes of neighbouring properties. The vectors generated from these attention mechanisms are concatenated with the original geographical (Gi) and structural (Ai) attributes of the property. This concatenated vector is propagated through a hidden neural layer to synthesise the final house embeddings. This intricate methodology enables the model to assimilate geographical and structural nuances, enhancing its predictive capabilities.

3.4. Regression layer

For the empirical component of our study, we employed a diverse set of regression algorithms, each optimised through rigorous cross-validation techniques. The algorithms were selected based on their suitability for our dataset's specific characteristics and the computational resources at our disposal. Below is an exhaustive list of the algorithms utilised:

- **Linear Regression (LR)**: Utilised with default hyperparameters as implemented in the scikit-learn library (Pedregosa et al., 2011). This algorithm serves as a baseline model for our study.
- **Random Forest (RF)**: An ensemble of decision trees, optimised using grid search and k-fold cross-validation. Hyperparameters such as the number of trees were varied, with tests conducted for 50, 100, 200, 700, and 1000 trees (Breiman, 2001).
- **LightGBM (LGBM)**: A gradient boosting framework that uses tree-based learning algorithms. Hyperparameters including the number of trees (50, 100, 200), the number of leaves (3, 4, 5, 100, 300), and the learning rate (0.03, 0.05, 0.07, 0.1) were fine-tuned (Ke et al., 2017).
- **Extreme Gradient Boosting (XGB)**: An optimised distributed gradient boosting library, fine-tuned through cross-validation. Parameters such as minimum child weight, gamma, subsample, column sample by the tree, learning rate, and maximum depth were adjusted (Chen & Guestrin, 2016a).
- **Categorical Boosting (CatBoost)**: An algorithm specifically designed for handling categorical variables. The depth parameter was optimised, with tests conducted for depths of 8 and 10 (Prokhorenkova, Gusev, Vorobev, Dorogush, & Gulin, 2018).
- **K-Nearest Neighbors (KNN)**: A distance-based algorithm, optimised by adjusting the number of neighbours, with tests conducted for 10 and 15 neighbours (Cover & Hart, 1967).
- **Decision Tree (DT)**: A basic tree-based model, optimised by adjusting the maximum depth parameter, with tests conducted for a depth of 9 (Quinlan, 1986b).
- **Support Vector Machines (SVM)**: A kernel-based algorithm suitable for linear and non-linear problems. Parameters 'C' and 'gamma' were fine-tuned using cross-validation (Cortes & Vapnik, 1995).
- **Regression Layer (RL)**: This layer serves as the terminal component of our attention-based neural network model, generating the final housing price prediction based on the feature map (house embeddings) obtained from preceding layers.

This empirical analysis aims to comprehensively evaluate the selected algorithms, thereby elucidating the relative merits and demerits in the context of housing price prediction.

4. Experimentation

This section presents the experimentation methodology adopted for our house price prediction task, including the specifics of the dataset preparation, model implementation, training, and evaluation process.

4.1. Dataset

In the experimental section, we utilised several datasets from different cities across various parts of the world to showcase the effectiveness of our model.

1. **Italian (IT) Dataset**: We obtained our dataset of Italian (IT) properties from Immobiliare. It is a well-known real estate platform in Italy. To collect the data, we designed a web scraper that extracted information from eight different cities: Genoa, Milan, Turin, Rome, Bologna, Florence, Naples, and Palermo. We filtered the data to include only five types of properties, such as apartments and penthouses, while excluding outliers like farms, buildings, and properties under construction. This ensured that the dataset was representative and coherent. We then conducted a thorough cleaning process to eliminate outliers. This process helped us eliminate data entry errors and rare property types, resulting in a consistent dataset suitable for analysis. To enrich the dataset, we added geographical data points. We included precise longitude and latitude coordinates for each property listing and leveraged OpenStreetMap to enhance each listing with Points of Interest (POI) data. This provided more profound insights into the property's surroundings, which could be significant in assessing its value. The final IT dataset comprises 30,918 property listings spread across eight significant cities in Italy. Each listing includes 19 distinct features that capture structural attributes, such as surface area, year of construction, and geographical details.
2. **Beijing (BJ) Dataset**: This dataset consists of 28,550 real estate transactions in Beijing and is sourced from the H4M study (Zhao, Shi et al., 2022). It includes 25 features, which range from structural attributes like surface area and year of construction to geographical elements such as district location and Point Of Interest (POI) information. The features are detailed in Table 1.
3. **Kings County (KC) Dataset**: Sourced from the GitHub[5] In the repository associated with the "Attention-Based Interpolation" paper, there is a dataset representing the Kings County, USA housing market. This dataset comprises 21,650 house samples

[5] https://github.com/darniton/ASI.

Table 1
Summary of datasets.

Dataset	Price range	Number of samples	Number of features
IT (Italian)	(60 000 to 720 000) Euro	30,918	24
BJ (Beijing)	(5500 to 170 000) Yuan	28,550	26
KC (Kings county)	(75,000 to 7,700,000) Dollar	21,650	18
POA (Porto Alegre City)	(70,000 to 1,168,324) Reais	15,368	7

characterised by 19 distinct features. These features encompass structural and geographical attributes and are detailed in a separate table, Table 1.
It is important to note that the prices in this dataset are provided in a log-scaled format.

4. **Porto Alegre City (POA) Dataset**: Derived from the repository provided by Vianna and Barbosa, this dataset focuses on Brazil's Porto Alegre City housing market. It includes 15,368 house samples, each described by six features, similar to the KC dataset. The features are outlined in Table 1.
It is essential to recognise that the prices in this dataset are provided in a log-scaled format.

4.2. Model configuration

Our model was developed in a Python 3.7 environment, using TensorFlow 2.5 as the backend for the Keras framework. The model was executed on a system with an Intel Core i5-13700K CPU and an NVIDIA GeForce RTX 3070 GPU. We used cross-validation and grid search techniques for hyperparameter tuning to achieve optimal results with regression algorithms such as XGBoost and RandomForest. For our custom model, we fine-tuned the hyperparameters using a validation subset of the data to obtain the best possible embedding representation and predictive performance. The hyperparameters and their values are summarised in Table 2, and we describe each hyperparameter and its significance below.

- **n-nearest**: Specifies the number of nearest neighbours to consider. The best values were 40 for IT, 60 for KC, 60 for POA, and 30 for BJ.
- **sigma** (σ): Controls the width of the Gaussian kernel. Optimal values were 2 for IT, 2 for KC, 2 for POA, and 10 for BJ.
- **nodes**: Represents the number of nodes in the hidden layers. The best values were 60 for IT, 60 for KC, 60 for POA, and 60 for BJ.
- **Num_heads**: Specifies the number of attention heads in the model. Optimal values were 8 for IT, 8 for KC, 4 for POA, and 4 for BJ.
- **Num_geo**: Indicates the number of geographical features to consider. The best values were 30 for IT, 30 for KC, 10 for POA, and 15 for BJ.
- **Num_euc**: Represents the number of Euclidean dimensions for distance calculations. The best values were 25 for IT, 30 for KC, 15 for POA, and 15 for BJ.
- **LR (Learning Rate)**: Controls the step size during optimisation. Optimal values were 0.001 for IT, 0.008 for KC, 0.001 for POA, and 0.001 for BJ.
- **batch size**: Specifies the number of samples per batch during training. Optimal values were 32 for IT, 250 for KC, 32 for POA, and 250 for BJ.
- **act func (Activation Function)**: Either Rectified Linear Unit (ReLU) or Exponential Linear Unit (ELU) was used. ELU was optimal for all datasets.
- **hidden act function (Hidden Layer Activation Function)**: The activation function for the hidden layers was either ReLU, ELU, regression, or linear. The linear function was optimal for all datasets.
- **similarity function**: We used the Gaussian Kernel and Identity function to compute similarities between data points. The Identity function was optimal for IT and POA, while the Gaussian Kernel was optimal for KC and BJ (see Table 2).

4.2.1. Evaluation metrics

After training, the model was evaluated using standard regression metrics such as Root Mean Squared Error (RMSE) and Mean Absolute Error (MALE). These metrics serve specific purposes in assessing the model's performance:

- **RMSE (Root Mean Squared Error)**: Provides a measure of the model's prediction error, penalising more significant errors more severely than smaller ones. It is advantageous when significant errors are undesirable in the prediction task.
- **MALE (Mean Absolute Logarithmic Error)**: This metric expresses the average magnitude of the relative errors between predicted and actual values while disregarding their direction. It is beneficial when dealing with exponential growth, or underestimation is more critical than overestimation.

These metrics collectively offer a comprehensive evaluation of the model's performance in predicting house prices, allowing for the assessment of the model's accuracy and its goodness of fit to the actual data.

4.3. Results and interpretation

In our evaluation, we consider the average and best performance metrics to comprehensively view each model's capabilities. The average performance metrics are derived from 10-fold cross-validation, indicating how the model will likely perform on unseen data. It gives us a more generalisable performance measure by mitigating the risk of the model overfitting to a particular subset of the data. On the other hand, the best performance metrics are extracted using grid search techniques. These values demonstrate the optimal performance that the model can potentially achieve under ideal hyperparameter settings. Including both types of metrics allows for a balanced understanding of the model's robustness and potential for excellence. It helps identify the most consistently high-performing models and those with the capacity for superior performance under the right conditions.

4.3.1. Base models benchmark

Table 3 provides an exhaustive evaluation of multiple machine-learning models In an exhaustive evaluation of machine learning models on real estate datasets from Italy (IT), King's County (KC), Porto Alegre City in Brazil (POA), and Beijing (BJ), the best performance was consistently demonstrated by XGBoost (XGB). Specifically, XGB recorded the best MALE values of 0.1350 in IT, 0.1160 in KC, 0.1613 in POA, and 0.0723 in BJ. Notably, the average performance for XGB was stable and closely aligned with these best values, indicating high reliability across diverse geographic datasets. CatBoost and LightGBM also performed strongly, closely trailing XGB in each dataset. For instance, CatBoost had the best MALE values of 0.1362 in IT, 0.1131 in KC, 0.1793 in POA, and 0.0782 in BJ. LightGBM posted the best MALE deals of 0.1381 in IT, 0.1164 in KC, 0.172 in POA, and 0.0790 in BJ. The average performances of CatBoost and LightGBM were also impressively stable and nearly matched their respective best values. Conversely, Support Vector Machines (SVM) significantly underperformed, with its best MALE values being 0.4072 in IT, 0.1331 in KC, 0.2232 in POA, and a dismal 0.2234 in BJ. K-Nearest Neighbors (KNN), a traditional algorithm, also lagged, particularly in the BJ dataset, where it posted a best MALE of 0.1116. In summary, XGB takes the lead across all datasets regarding best and average performance metrics,

Table 2
The best hyperparameters that were chosen based on a grid search method to train our model on IT, KC, POA and BJ datasets.

HP values	General values	Datasets			
		IT	KC	POA	BJ
N-nearest	5, 10, 15, 60	40	60	60	30
Num_geo	20, 25, 30, 35, 40, 45, 50, 55, 60	30	30	10	15
Num_euc	20, 25, 30, 35, 40, 45, 50, 55, 60	25	30	15	15
Num_heads	1,2,4,8,12,15	8	8	4	4
Sigma(σ)	2, 5, 10, 15, 20	2	2	2	10
Nodes	5, 10, 15, 60	60	60	60	60
LR	[0.001–0.01]	0.001	0.008	0.001	0.001
Batch size	250, 300, 400, 500	32	250	32	250
Act func	Relu and ELU	ELU	ELU	ELU	ELU
Hidden act func	Relu, ELU, regression and linear	Linear	Linear	Linear	Linear
Similarity function	Identity and Gaussian kernel	Identity	Gaussian kernel	Identity	Gaussian kernel

Table 3
Benchmark the datasets on state-of-the-art machine learning models. The average value is referred to k-fold cross-validation with k = 10.

Model	IT				KC				POA				BJ			
	MALE ↓		RMSE ↓		MALE ↓		RMSE ↓		MALE ↓		RMSE ↓		MALE ↓		RMSE ↓	
	Best	Avg	Best	Avg	Best	Avg	Best	Avg	Best	Avg	Best	Avg	Best	Avg	Best	Avg
LR	0.385	0.388	76 224	76 241	0.1924	0.1925	205 460	209 330	0.2610	0.2611	152 861	153 775	0.2394	0.2396	20 551	20 558
KNN	0.247	0.248	84 637	85 163	0.1501	0.1513	174 046	175 628	0.2065	0.2078	122 521	123 113	0.1116	0.1121	12 093	12 149
DT	0.197	0.205	69 085	70 423	0.1583	0.1608	158 937	178 296	0.2163	0.2195	127 382	128 915	0.0936	0.0954	10 155	10 409
RF	0.1502	0.1508	51 774	52 147	0.1245	0.1251	133 933	136 993	0.1716	0.1731	105 183	105 975	0.0784	0.0794	8369	8475
SVM	0.4072	0.4074	128 634	128 781	0.1331	0.1336	149 265	152 675	0.2232	0.2246	126 911	128 191	0.2234	0.2237	20 652	20 668
LGBM	0.1381	0.1384	46 183	46 492	0.1164	0.1175	122 116	126 076	0.172	0.177	104 928	106 705	0.0790	0.0796	8070	8152
CatBoost	0.1362	0.1368	**45 942**	**46 233**	0.1131	0.1141	120 351	**123 077**	0.1793	0.1775	105 984	106 593	0.0782	0.0785	7995	8066
XGB	**0.1350**	**0.1358**	46 008	46 396	**0.1160**	**0.1167**	**119 479**	124 459	**0.1613**	**0.1634**	**100 212**	**101 614**	**0.0723**	**0.0744**	**7713**	**7836**

closely followed by CatBoost and LightGBM, which also show highly stable average performances. Conversely, SVM and traditional models like KNN are less effective, particularly in complex, geographically diverse datasets.

4.3.2. Experimental results for our model

Throughout our study, we conducted tests to compare our model with other models such as ASI, ANN, ASI + Multi-Head, and ASI + Gating. The dataset was divided into three subsets: 70% for training, 20% for testing, and 10% for validation. The performance metrics of our model, ASI, ANN, ASI + Multi-Head, and ASI + Gating, were compared across four different datasets from Italy (IT), King's County (KC), Porto Alegre (POA), and Beijing (BJ) and are shown in Table 4. Our model outperformed the ASI model in the IT dataset with a lower MALE of approximately 1.52% and a lower RMSE of 0.36% (0.1312 and 45,797, respectively). In the KC dataset, our model showed a remarkable 13.3% improvement in RMSE compared to the ASI model, translating to a lower RMSE of approximately 107,993 compared to ASI's 124,557. Our model also performed better than the ASI model in the POA and BJ datasets, achieving lower MALE and RMSE values. These results demonstrate the versatility and accuracy of our model across different datasets and locations. Our model incorporates Multi-Head Gated Attention mechanisms, allowing it to interpret various spatial cues and enhance predictive accuracy. The Gated Attention mechanism stood out in the MALE metric, consistently outperforming other models. Similarly, the Multi-Head Attention mechanism significantly reduced RMSE, effectively handling complex spatial relationships and minimising error rates. Additionally, compared to the ASI + Multi-Head and ASI + Gating models, our model consistently showed improvement in the IT, KC, POA, and BJ datasets. The advanced Multi-Head Gated Attention architecture was crucial in improving overall predictive accuracy across all metrics and datasets. In conclusion, our model displayed superior performance compared to the ASI, ANN, ASI + Multi-Head, and ASI + Gating models. Incorporating the advanced Multi-Head Gated Attention mechanism proved to be a critical factor in enhancing overall predictive accuracy.

4.3.3. Experimental results for embeddings generated by our model and ASI model in comparison to raw data

The benchmarking results presented in Table 5 provide a comprehensive assessment of different machine learning models using various data sources, including raw data, ASI embeddings, and the proposed Multi-Head Gated Attention Spatial Interpolator model "Ours". The analysis focuses on two main error metrics: Mean Absolute Logarithmic Error (MALE) and Root Mean Square Error (RMSE) across four regions (IT, KC, POA, BJ). When utilising raw data, XGBoost (XGB) consistently outperforms other models across all regions. For example, in the IT region, XGB achieves the lowest MALE of 0.1350 and an RMSE of 46,008, highlighting its strong performance with unprocessed data. Similarly, in the KC region, XGB records a MALE of 0.1160 and an RMSE of 119,479, again outperforming other raw data models. The introduction of ASI embeddings generally enhances the performance of the models, especially those based on gradient-boosting techniques. However, models using ASI embeddings still do not surpass the performance of XGBoost on raw data, suggesting that raw data coupled with powerful models like XGBoost can still capture significant predictive insights. The proposed "Ours" model demonstrates significant superiority over raw data and ASI embeddings across most models and regions. In the KC region, Logistic Regression (LR) with "Ours" embeddings achieves a MALE of 0.1103 and an RMSE of 106,954, outperforming both the raw data and ASI embeddings. Similarly, CatBoost with "Ours" embeddings in the IT region records a MALE of 0.1320 and an RMSE of 45,708, better than raw and ASI embeddings. Moreover, the "Ours" model consistently delivers superior results in the POA and BJ regions. For instance, in the POA region, XGBoost with "Ours" embeddings achieves a MALE of 0.1392 and an RMSE of 92,677, indicative of robust performance across different data representations.

4.3.4. Experimental results for our model house embeddings using cross-validation

In our experiment, we wanted to see how custom house embeddings generated by our Multi-Head Gated Attention model would affect the performance of various baseline machine learning models. These embeddings were created based on structural and geographical information and enhanced the feature space for algorithms like Linear

Table 4
Performance evaluation of our model against the ASI model, ANN, Multi-head only, and Gating attention only models.

Model	IT		KC		POA		BJ	
	MALE ↓	RMSE ↓	MALE ↓	RMSE ↓	MALE ↓	RMSE ↓	MALE ↓	RMSE ↓
ANN	0.197	67 835	0.2231	127 900	0.2212	125 961	0.239	19 565
ASI	0.133	46 473	0.112	124 557	0.139	93 818	0.075	7934
ASI + Multi-Head	0.135	46 347	0.113	109 302	0.138	92 073	0.75	7900
ASI + Gating	0.132	46 835	0.111	112 839	0.140	93 001	0.075	8002
Ours	**0.1312**	**45 797**	**0.110**	**107 993**	**0.136**	**92 020**	**0.073**	**7797**

Table 5
Benchmarking the embeddings of different models and the raw data on state-of-the-art machine learning models..

Data source	Method	IT		KC		POA		BJ	
		MALE	RMSE	MALE	RMSE	MALE	RMSE	MALE	RMSE
Raw	LR	0.385	76 224	0.1924	205 460	0.2610	152 861	0.2394	20 551
	SVM	0.4072	128 634	0.1331	149 265	0.2232	126 911	0.2234	20 652
	LGBM	0.1381	46 183	0.1164	122 116	0.172	104 928	0.0790	8070
	CatBoost	0.1362	45 942•	0.1131	120 351	0.1793	105 984	0.0782	7995
	XGB	**0.1350•**	46 008	**0.1160•**	119 479•	**0.1613•**	100 212•	**0.0723•**	7713•
ASI	LR	0.1543	49 081	0.1142	109 117	0.1455	94 081	0.0877	8488
	SVM	0.1497	49 431	0.1163	141 977	0.1441	95 399	0.1441	95 399
	LGBM	0.1410	47 882	0.1215	133 608	0.1502	97 131	0.0798	8123
	CatBoost	**0.1384•**	**47 088•**	0.1171	132 122	**0.1425•**	**93 562•**	0.0774	8032
	XGB	0.1417	47 932	0.1203	129 802	0.1483	94 407	0.0755	7967
Ours	LR	0.1317	45 837	**0.1103•**	**106 954•**	0.1369	91 725	**0.0732•**	**7779•**
	SVM	0.1743	58 977	**0.1103•**	107 389	**0.1357•**	**91 719•**	0.0778	8332
	LGBM	0.1324	45 885	0.1138	111 551	0.1384	92 383	0.0742	7835
	CatBoost	**0.1320•**	**45 708•**	0.1130	110 481	0.1367	91 784	0.0735	7806
	XGB	0.1324	45 961	0.1147	108 644	0.1392	92 677	0.0739	7822

Table 6
Benchmark the datasets on state-of-the-art machine learning models on the generated embeddings from our model. The average value is referred to k-fold cross-validation with k = 10.

Model	IT				KC				POA				BJ			
	MALE ↓		RMSE ↓		MALE ↓		RMSE ↓		MALE ↓		RMSE ↓		MALE ↓		RMSE ↓	
	Best	Avg	Best	Avg	Best	Avg	Best	Avg	Best	Avg	Best	Avg	Best	Avg	Best	Avg
LR	**0.1317**	**0.1318**	45 837	45 868	**0.1103**	**0.1104**	**106 954**	**107 133**	0.1369	0.1372	91 725	91 848	**0.0732**	**0.0733**	**7779**	**7786**
KNN	0.1352	0.1354	46 648	46 761	0.1208	0.1211	123 235	125 010	0.1398	0.1402	92 032	92 369	0.0767	0.0770	7980	8015
DT	0.1347	0.135	46 007	46 160	0.1353	0.1412	142 731	154 575	0.1501	0.1512	97 704	98 387	0.0752	0.0756	7853	7879
RF	0.1323	0.1325	45 921	45 995	0.1146	0.1151	112 588	115 852	0.1388	0.1396	92 623	93 243	0.0741	0.0742	7843	7864
SVM	0.1743	0.1745	58 977	59 188	0.1103	0.1103	107 389	108 700	**0.1357**	**0.1359**	**91 719**	**91 796**	0.0778	0.0779	8332	8344
LGBM	0.1324	0.1327	45 885	45 930	0.1138	0.1141	111 551	112 994	0.1384	0.1387	92 383	92 617	0.0742	0.0744	7835	7854
CatBoost	0.1320	0.1321	**45 708**	**45 752**	0.1130	0.1136	110 481	113 174	0.1367	0.1373	91 784	91 976	0.0735	0.0737	7806	7814
XGB	0.1324	0.1325	45 961	46 021	0.1147	0.1152	108 644	112 332	0.1392	0.1397	92 677	93 003	0.0739	0.0740	7822	7834

Regression, KNN, Decision Tree, Random Forest, SVM, LightGBM, CatBoost, and XGBoost. We evaluated the models' performance using four different geographical datasets: Italy (IT), King's County (KC), Porto Alegre (POA), and Beijing (BJ), and assessed the Best and Average MALE and RMSE scores.

Our results showed that our custom embeddings significantly positively impacted the predictive performance of the baseline models. For example, when the CatBoost model was augmented with our custom embeddings, it achieved the lowest RMSE score in the IT dataset at 45,708, outperforming even our original Multi-Head Gated Attention model. However, we found that the improvement magnitude was inconsistent across all datasets. The IT dataset, which combines data from various cities with significant geographical and Euclidean distances between them, showed only a modest enhancement of around 1.3% in RMSE when deploying CatBoost with custom embeddings compared to the baseline.

4.4. Discussion

We discovered that the unique spatial complexities inherent in each dataset could impact the effectiveness of the custom embeddings. For instance, in the KC dataset, CatBoost with custom embeddings demonstrated significant gains over its baseline, whereas, in IT, the improvements were more restrained. We also found that even simpler models like Linear Regression could benefit substantially from the enriched feature space the embeddings provide. In the IT dataset, the best MALE improved by approximately 65.8%, the average MALE improved by approximately 66.0%, the best RMSE improved by approximately 39.8%, and the average RMSE improved by approximately 39.8%. In the CatBoost model for the IT dataset, the best MALE improved by approximately 2.4%, and the average MALE improved by approximately 2.9%. The best RMSE improved by approximately 0.5%, and the average RMSE improved by approximately 1.0%. This indicates a positive trend in reducing MALE and RMSE values, which is crucial for achieving better model performance in predictive tasks like house price prediction. In the KC dataset, the implementation of custom embeddings reflected varying degrees of improvement across different machine-learning models. The CatBoost model illustrated an enhancement in the best MALE value by approximately 5%, although the average MALE value experienced a minor deterioration by approximately 0.44%. On the brighter side, a more noticeable improvement was observed in the RMSE values, where the best RMSE value improved by approximately 8.20%, and the average RMSE value improved by approximately 8.04%. The POA dataset manifested a significant leap in performance metrics upon integrating custom embeddings. Specifically, the CatBoost model, when augmented with custom embeddings, demonstrated a robust improvement in both MALE and RMSE

values. The best MALE value improved by an impressive margin of approximately 23.77%, while the average MALE value improved by approximately 22.66%. Concurrently, the RMSE metrics also exhibited substantial enhancements, with the best RMSE value improving by approximately 13.40%, and the average RMSE value improving by approximately 13.45%. In the BJ dataset, we observed that models trained on embeddings generally perform better on average values, reflecting a more consistent performance across varying data points. However, the best values achieved in MALE and RMSE metrics were slightly better when models were trained on original data. This suggests that while embeddings generally enhance model performance, there might be specific instances or datasets where traditional feature sets could yield better or comparable results.

Building upon our previous results in Tables 3, 4 and 6 our model, based on Multi-Head Gated Attention, consistently outperforms the baseline models across multiple datasets. This superiority is particularly noteworthy as the model excels in spatial interpolation tasks and enhances the performance of other state-of-the-art machine learning models when its embeddings are used. One of the key advantages of our model over the attention-based interpolation model is the ability to capture multiple contexts from each head and control the flow of the information so that it will consider the most similar neighbours through the use of Multi-Head Gated Attention.

4.4.1. Abelation study

In this research, an ablation study was conducted to rigorously evaluate the impact of different model configurations on predictive performance across several diverse datasets: IT (data from 8 cities in Italy), KC (King County, USA), POA (Porto Alegre, Brazil), and BJ (Beijing, China). The study compares the effectiveness of using roftend attention embeddings (approximated by the ASI), and multi-head gated attention (MHGA) embeddings (as represented by the "Ours"). Additionally, the study explores the individual contributions of Multi-Head and Gating Attention mechanisms when applied separately and their combined effect in the MHGA model. Raw Data vs. Single-Head Attention Embeddings Raw data represents unprocessed features, which often include noise and irrelevant information in real-world datasets. This baseline performance highlights the inherent difficulties in making predictions without any feature refinement or focusing on relevant aspects of the data. For instance, the MALE values observed with raw data, such as 0.1350 in the IT dataset and 0.1613 in the POA dataset, reflect the challenges of handling diverse urban data from multiple cities with varying characteristics. The high RMSE values further underscore the inefficiencies of relying on raw data, which often contains a complex mix of noise and patterns that are difficult for models to disentangle. The comparison between raw data and single-head attention embeddings reveals the critical importance of basic feature engineering and selective focus mechanisms in machine learning. Single-head attention, as implemented in the ASI model, enables the model to concentrate on specific aspects of the data, resulting in moderate improvements in predictive accuracy. However, the MALE values for the ASI model, such as 0.1384 in IT and 0.1425 in POA, indicate that while single-head attention can capture certain relationships within the data, it cannot fully model the complexities present in datasets representing multiple cities. This limitation highlights the challenges of single-perspective approaches in urban analytics, where multi-dimensional data is the norm. Including Multi-Head or Gating Attention mechanisms within the ASI model leads to significant performance improvements, as evidenced in Table 4. For example, the Multi-Head mechanism enhances the model's ability to capture complex patterns, improving MALE and RMSE values in the KC dataset. Similarly, the Gating mechanism improves the model's focus on relevant information, as demonstrated in the IT dataset. Combining both mechanisms in the MHGA model yields the most substantial performance gains across all datasets. The Multi-Head mechanism enables the model to consider multiple facets of the data simultaneously, which benefits datasets with diverse and complex patterns. On the other hand, the Gating mechanism improves model performance by selectively filtering the most informative features. Combining both mechanisms in the MHGA model results in superior performance across all datasets, highlighting the synergistic effect of multiple attention heads with a gating mechanism. Introducing the Multi-Head Gated Attention (MHGA) mechanism significantly improves the model's capacity to understand and predict complex patterns within the data. By combining multiple attention heads with a gating mechanism, MHGA produces more refined and informative embeddings, outperforming other model configurations across various datasets. The comparison between single-head and multi-head gated attention embeddings demonstrates the effectiveness of advanced attention mechanisms in predictive modelling, particularly in datasets with diverse urban environments. The MHGA embeddings consistently deliver the best performance across all tested datasets, showcasing their robustness and generalisability. The ablation study confirms that Multi-Head Gated Attention embeddings provide the best performance compared to other configurations, emphasising their effectiveness in capturing the complex, multi-dimensional relationships inherent in urban environments. These findings suggest that advanced attention mechanisms like MHGA are essential for achieving high accuracy and robustness in predictive modelling of complex, heterogeneous datasets, with significant implications for urban development, policy-making, and decision-making contexts.

4.4.2. Comparison of predictive performance: XGBoost model on raw data vs. our model across datasets

The violin plots in Fig. 3 provide a detailed comparison of the actual house price predictions from the XGBoost model trained on raw data, and forecasts from our Multihead Gated Attention Spatial Interpolation (MGASI) model across four distinct datasets: IT, BJ, KC, and POA. Each subplot, labelled (A) through (D), represents the performance on a specific dataset, enabling a clear and comprehensive evaluation of the model's predictive accuracy and ability to capture the variability in the data.

In the IT dataset Fig. 3A, which includes data from 8 cities, shows a mean of 266,067.63 and a median of 239,000.00, with a standard deviation of 150,185.41, indicating moderate variability. The XGBoost model trained on raw data closely aligns with these values, achieving a mean of 266,485.18 and a median of 242,684.66. However, the slightly reduced standard deviation of 141,220.41 suggests that the XGBoost model does not fully capture the variability in the data. In contrast, our model achieves a mean of 265,289.10 and a median of 239,667.55, with a standard deviation of 143,203.24, closer to the actual data's spread. This suggests that while both models perform well, our model better captures the inherent variability, making it more robust for predictive tasks.

In the BJ dataset Fig. 3B, the actual data presents a mean of 66,282.44 and a median of 61,920.50, with a standard deviation of 26,026.37. The XGBoost model's predictions show a mean of 66,110.85 and a median of 63,005.94, with a slightly lower standard deviation of 24,322.90, indicating a slight limitation in capturing the true variability of the data. However, our model matches the mean (66,396.96) and median (62,985.21) and maintains a higher standard deviation of 24,634.84, demonstrating a better representation of the data's distribution. Although the XGBoost model shows strong performance in certain parts of the dataset, particularly in central tendency, our model's overall performance across the entire dataset, especially in capturing variability, is superior. The KC dataset Fig. 3C exhibits significant variability, with a high standard deviation of 369,439.92. The XGBoost model, with a mean of 537,451.61 and a median of 456,826.44, shows a lower standard deviation (344,344.83), suggesting that it underestimates the variability present in the actual data. In contrast, our model aligns closely with the actual data's mean (542,832.32) and median (454,579.30), achieving a standard deviation 356,626.95. This indicates that while the XGBoost model may perform slightly better

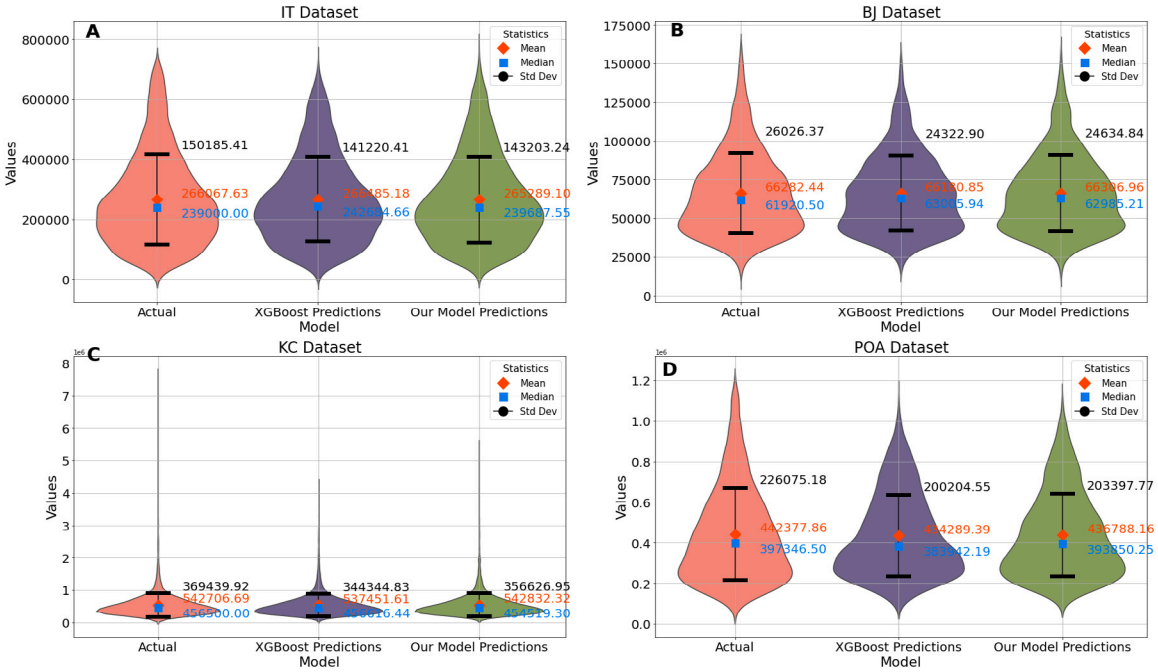

Fig. 3. Violin plots comparing the distribution of actual values, predictions from the XGBoost model trained on raw data, and predictions from our Multihead Gated Attention Spatial Interpolation (MGASI) model across four datasets: (A) IT, (B) BJ, (C) KC, and (D) POA. Each plot visualises each dataset's central tendency (mean and median) and variability (standard deviation), with distinct markers to differentiate between models. A diamond, the median by a square, and the standard deviation by error bars with circles at the ends represent the mean. The letters in the top-left corners correspond to the sections in the discussion where a detailed analysis of each dataset is provided.

in specific areas, particularly in central tendency, our model more effectively captures the data's overall spread, making it a better choice for comprehensive predictive modelling. Finally, in the POA dataset Fig. 3D, characterised by moderate variability and unique feature interactions, shows a mean of 442,377.86 and a median of 397,346.50, with a substantial standard deviation of 226,075.18. The XGBoost model, with a mean of 434,289.39 and a median of 383,942.19, shows a reduced standard deviation of 200,204.55, suggesting an oversimplification of the data's distribution. Our model, however, achieves a mean of 436,788.16, a median of 393,850.25, and a standard deviation of 203,397.77, which is closer to the actual data's variability. This underscores the effectiveness of our model in capturing the complex patterns in the data, aided by the attention mechanisms' ability to handle intricate dependencies. In summary, while the XGBoost model demonstrates more robust performance in specific parts of the datasets, particularly in central tendency metrics, our model consistently shows better average performance across all datasets. Our model's superior ability to capture both the central tendencies and the inherent variability of the actual data demonstrates its robustness and reliability in predictive tasks. The enhanced performance of our model is a direct result of the multi-head gated attention mechanism, which improves its ability to capture complex, multidimensional patterns in the data, making it an essential tool for high-precision predictive modelling.

4.4.3. Spatial and structural analysis

The present study introduces a Multi-Head Gated Attention model that exhibits superior performance compared to baseline models when applied to various datasets, including IT and POA. This model utilises distinct weights and biases within each attention head to capture various contextual relationships within the data, showcasing its exceptional capabilities in spatial interpolation tasks. This approach provides a more comprehensive understanding of the underlying spatial dynamics. Our model's multi-head gated attention mechanism exceeds traditional singular attention approaches by integrating various spatial and structural features from the data. This integration is essential as it moderates the influence of outliers, which is expected in a vast and diverse metropolis like Beijing, where extreme data points can skew the analysis. By employing this sophisticated mechanism, the model ensures the delivery of accurate and nuanced house price predictions that genuinely reflect the complex intricacies of Beijing's housing market, setting a new benchmark for robustness and reliability. The box plots in Fig. 4(a,b) effectively illustrate each dataset's spatial and structural features. Specifically, Fig. 4(b) reveals that Kings County (KC) has a compact urban form, indicated by a median geodesic distance of just under 0.65 km, which is also supported by a low median normalised Euclidean distance shown in Fig. 4(a), highlighting high structural homogeneity among houses. In contrast, Beijing (BJ) portrays a more dispersed housing structure with a median geodesic distance of approximately 0.45 km, as indicated in Fig. 4(b), and a median normalised Euclidean distance of roughly 0.150, as shown in Fig. 4(a). These distances indicate a significant variation in structural features, suggesting a housing landscape that includes densely packed urban areas and more spread-out suburban or peri-urban zones. The Italian (IT) region demonstrates a median geodesic distance of around 0.50 km, reflecting less uniformity and greater architectural diversity, as further evidenced by a median normalised Euclidean distance of around 0.110. Moving to Porto Alegre (POA), the dataset displays a distinctive spatial composition, with a median geodesic distance that suggests moderately dense housing and a median normalised Euclidean distance of approximately 0.100. This places POA in a unique position between the densely packed environment of KC and the varied spatial arrangements of BJ and IT. The moderate variation in POA's housing structures signifies an urban design that merges densely built areas with open suburban spaces, reflecting its rich historical development and cultural diversity. Employing the multi-head gated attention mechanism for the POA dataset allows for an in-depth exploration of the city's complex architectural styles and spatial dynamics. When juxtaposed with the consistent architecture of KC and the diverse spatial distributions of BJ and IT, our model's multifaceted approach yields a deep understanding of the nuances within POA's urban clusters and the distinctive nature of its rural homes. As a result, our model stands out as a sophisticated and precise analytical tool, uniquely equipped to navigate and predict the

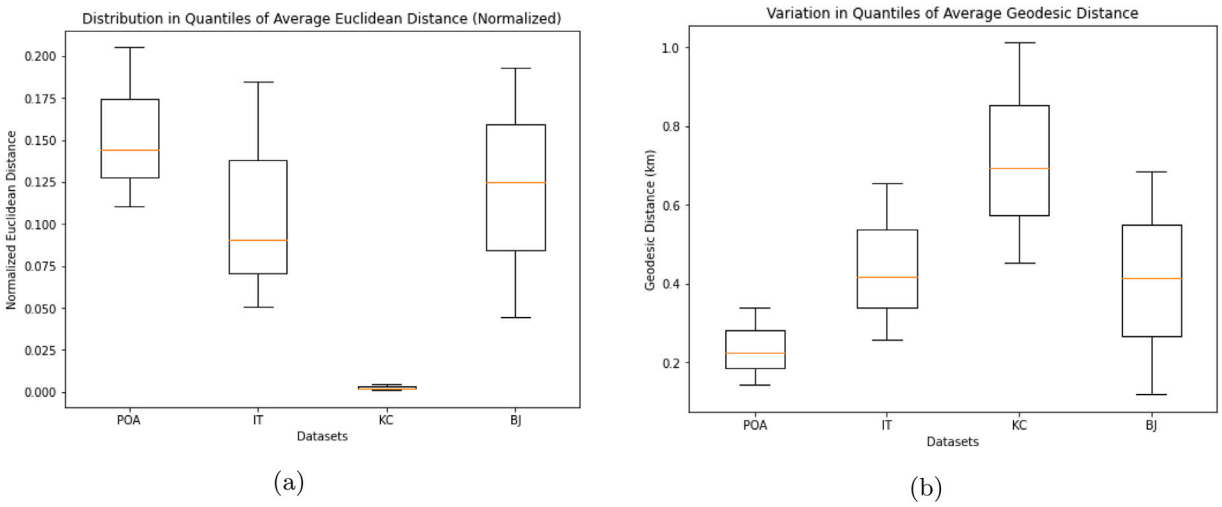

Fig. 4. Analysis of Geodesic and Euclidean Distances Among the 60 Nearest Houses Across Datasets a Highlights the variation in quantiles of the average geodesic distance (in km) for the 60 nearest houses across the four datasets, reflecting the spatial proximity of residences. b Represents the distribution in quantiles of the average normalised Euclidean distance for the 60 nearest houses, taking into account the structural features of the houses. Min–max normalisation was employed to standardise the distance values due to the diverse attributes of the houses in each dataset.

intricate dynamics of the housing market with extraordinary accuracy and insight.

The improvements highlighted in Table 4 emphasises the progress made by our model compared to the ASI model. Our model achieved improvements of 1.35% and 1.46% in MALE and RMSE, respectively, for the IT dataset, 1.79% and 13.34% for the KC dataset, 2.16% and 1.92% for the POA dataset, and 2.67% and 1.73% for the BJ dataset. These results demonstrate the superiority of our model across different datasets and spatial configurations. The multi-head gated attention mechanism played a significant role in achieving these improvements. It captures diverse contextual relationships within the data by leveraging weights and biases in each head, especially when dealing with regions with a more varied architectural landscape and pronounced geographical diversity. The improvements in the KC dataset are significant, as it has a high degree of architectural uniformity. However, the model could still capture minute differences and nuances, leading to a 13.34% improvement in RMSE. For the BJ dataset, which has a more dispersed housing layout and a vast spatial range, the model achieved a 2.67% improvement in MALE and a 1.73% improvement in RMSE, highlighting the model's ability to accurately capture the essence of each area despite the considerable differences in spatial dynamics and architectural styles. The advances in the IT dataset were also noteworthy, with the model achieving a 1.35% improvement in MALE and a 1.46% improvement in RMSE despite the unique spatial layout of the region compared to KC. These results demonstrate the robustness and reliability of our model in providing accurate predictions against the ASI model across diverse datasets and spatial configurations.

4.4.4. Embeddings performance

In Table 6, we present a comparative analysis of our model embeddings against the benchmarks outlined in Table 3. Additionally, the results from the regression layer of our model are presented in Table Table 4. The results underline the substantial advancements made by our model and the generated embeddings. Rigorous evaluations across various validation sets demonstrate the superior performance of our model in handling complex spatial datasets. Furthermore, the efficiency of the generated embeddings emphasises our model's role in reducing data complexity so that simple models like linear regression can outperform ensembling models.

In the IT dataset, our model achieved a Mean Absolute Logarithmic Error (MALE) of 0.1312 and a Root Mean Square Error (RMSE) of 45,797. These results represent a 2.89% improvement in MALE and a 0.46% improvement in RMSE over the best baseline model, XGBoost, which recorded a MALE of 0.1350 and an RMSE of 46,008.

Furthermore, the embeddings in our model outperformed the regression layer of our model and the base benchmarking in terms of RMSE, with the Catboosting model achieving the best result of 45,708. This indicates a slight improvement over our model's performance.

These results can be attributed to the challenging nature of predicting housing prices accurately in this dataset, where various factors come into play. Our model's success suggests that its embeddings effectively capture the price variations associated with the diverse housing landscape, as evident from the wide distribution of Euclidean distances in Fig. 4(a). This distribution reflects the influence of different cities in one dataset, especially Italian towns, which exhibit various housing structures from the south to the north of Italy.

Transitioning to the KC dataset, our model displayed a MALE of 0.110 and an RMSE of 107 993. This corresponds to a percentage improvement of 2.81% and 10.27% in MALE and RMSE, respectively, compared to the best baseline model, CatBoost. CatBoost had a MALE of 0.1131 and an RMSE of 120 351. However, the embeddings seem to mark the best results over our model, and the base benchmarking with 0.1103 MALE value and 106 954 RMSE scored in the linear regression model shows an improvement in comparison to our model in both metrics, further emphasising the power of our generated embeddings.

Furthermore, the significant improvement observed in the Kings County (KC) dataset demonstrates our model's enhanced capability in dense housing and architectural uniformity regions. Our model boosts the prediction accuracy for the most relevant houses and creates diverse contextual frameworks that underscore the interrelationships between houses, even in areas of uniformity. Additionally, creating embeddings encapsulating these relationships further improves the model's performance.

Our model exhibits exceptional performance on the Porto Alegre (POA) dataset, achieving the lowest Mean Absolute Logarithmic Error (MALE) at 0.136 and Root Mean Square Error (RMSE) at 92,020. This performance surpasses the XGBoost baseline's MALE of 0.1613 and RMSE of 100,212, indicating a 15.67% improvement in MALE and an 8.17% improvement in RMSE. The model's superior embeddings are instrumental in this achievement, effectively streamlining intricate urban data for linear regression without losing essential details, as indicated in Tables 6 and Table 4. Fig. 4(a,b) potentially reveal the spatial complexity of POA, with its moderately dense urban fabric intertwined with suburban and rural patches. Despite this diversity posing challenges for predictive models, our embeddings adeptly encode these

complexities, effectively representing the multifaceted housing styles and values within POA. Our model's predictive precision stems from its algorithmic sophistication and nuanced understanding of the region's unique urban tapestry.

Lastly, base benchmarking for the Beijing (BJ) dataset performs better than our model's regression layer. However, our embeddings demonstrate better results, suggesting they are more generalised than the base benchmarking outcomes. The embeddings score the best MALE of 0.072 and the best RMSE of 7713, compared to 0.073 and 0.0732 MALE and 7797 and 7779 RMSE with our model and linear regression model using our embeddings, respectively. Our embeddings' average values from cross-validation are 0.0733 MALE and 7786 RMSE, while the base benchmarking average values are 0.074 MALE and 7836 RMSE, showing a close similarity to the embeddings.

Examining the housing market in Beijing presents several challenges, including managing diverse and often extreme data points typical of a large metropolis. The median distance to the nearest 60 homes in Beijing, as depicted in Fig. 4(b), is approximately 0.45 km, highlighting an extensive and varied housing layout. The city's diverse architectural styles add another layer of complexity to the dataset. Our model, equipped with a Multi-Head Gated Attention mechanism, is adept at handling these challenges. This mechanism effectively regulates the influence of outliers, ensuring a nuanced and accurate representation of Beijing's housing landscape. The embeddings generated by our model are particularly noteworthy for their ability to generalise across Beijing's diverse housing market. While the base benchmarking results provide valuable insights, our model's embeddings capture a broader range of intricacies, ensuring they are statistically sound and meaningfully representative of the real-world scenario.

This quantitative comparison highlights the considerable enhancements of our model. The marked performance uplift in the BJ dataset accentuates our model's potential in real estate price prediction tasks. Additionally, the comparative analysis with the original attention-based interpolation model by Viana and Barbosa (2021) on the KC and POA datasets further amplifies the strengths of our model. Our model's ability to efficaciously reduce data dimensionality while retaining crucial information has led to significant improvements in MALE and RMSE across all datasets. This proficiency in compressing high-dimensional data into more digestible forms has enabled algorithms like linear regression to compete and outperform complex ensemble models like LightGBM, CatBoost, and XGBoost.

5. Conclusion

This study significantly advances house price prediction by introducing a novel dataset focused on the Italian housing market and applying innovative spatial interpolation techniques. One of the critical contributions of our research is the development and implementation of the Multi-Head Gated Attention Interpolator. This model addresses a notable gap in applying attention mechanisms within house price prediction, particularly in non-time series datasets. Our Multi-Head Gated Attention Interpolator substantially improved prediction accuracy compared to traditional and original attention-based interpolation models. This improvement underscores the untapped potential of attention mechanisms in capturing complex spatial relationships. The model's ability to capture diverse geographical and structural contexts while filtering out irrelevant data using gated attention ensures robust predictions by reducing the impact of outliers. Creating embeddings using the Multi-Head Gated Attention Interpolator significantly boosts the results of state-of-the-art models such as XGBoost. These embeddings enhance the feature representation, allowing models like XGBoost to achieve higher prediction accuracy by leveraging the enriched spatial and structural information captured by the attention mechanism. This integration of advanced embedding techniques has improved prediction accuracy and the robustness and generalisability of the models across different datasets. Introducing a dataset focused on the Italian housing market enriches existing resources and provides a unique landscape for testing new methodologies. This dataset includes comprehensive geographical and structural data for accurate house price prediction. Our model effectively captures complex spatial relationships by utilising attention mechanisms, a capability critical for understanding how different geographical and structural contexts influence house prices. Unlike models such as Graph Neural Networks (GNNs), which require the creation of specific relational graphs for each dataset, our Multi-Head Gated Attention Interpolator can generalise across different cities without needing specific relational studies. This flexibility significantly enhances the model's applicability and reduces the need for extensive preprocessing and customisation. By filtering irrelevant data and focusing on relevant features, the Multi-Head Gated Attention Interpolator minimises the impact of outliers, leading to more robust and reliable predictions. Despite its advanced capabilities, our model's reliance on primary geographical and structural data makes it applicable even in scenarios with limited computational resources. This ensures the model can be widely adopted without requiring extensive computational infrastructure. While the Gated Attention mechanism helps mitigate the impact of outliers, ensuring robust predictions, the model's performance can be challenged in regions with low housing density, such as rural areas. In these areas, the sparse data availability can affect the model's accuracy, highlighting the need for additional data sources or alternative modelling strategies to maintain high prediction performance. We propose several avenues to further enhance the model's capabilities and address its limitations. Incorporating satellite imagery and interior and exterior photographs of properties will provide a more comprehensive view of the factors influencing house prices. This integration will enhance the predictive capabilities of our model but also presents challenges related to data preprocessing and harmonisation that need to be addressed. Integrating the Kolmogorov-Arnold network could offer a robust framework for capturing nonlinear dependencies in the data, further improving the model's performance. This network is particularly effective in dealing with complex, nonlinear relationships often present in housing data. Implementing an intelligent radius mechanism to dynamically determine each property's optimal number of neighbours could enhance the model's precision. This mechanism would adaptively select neighbours based on property-specific characteristics and spatial distribution, improving the relevance and accuracy of predictions. Further studies are needed to validate the model's applicability to different geographical regions. Thorough testing across diverse datasets will ensure the model's generalisability and robustness. By combining these advanced methodologies, we aim to create a more robust and versatile model capable of delivering superior performance across diverse scenarios and datasets. This research contributes to house price prediction and opens new avenues for future research, paving the way for more accurate and reliable predictive models in real estate markets. The findings from this study have significant implications for house price prediction. By demonstrating the effectiveness of attention mechanisms, specifically the Multi-Head Gated Attention Interpolator, in capturing complex spatial relationships, we provide a new direction for future research. Integrating embeddings created through advanced attention mechanisms with state-of-the-art models like XGBoost sets a new benchmark for prediction accuracy and robustness. Additionally, the model's ability to generalise across different cities without requiring specific relational studies, as is necessary with models like Graph Neural Networks, highlights its versatility and practical applicability. The proposed future directions also emphasise the potential for further enhancements, making house price prediction models more comprehensive and adaptable to various scenarios.

CRediT authorship contribution statement

Zakaria Abdellah Sellam: Conceptualization, Methodology, Literature search, Data analysis, Writing – original draft. **Cosimo Distante:** Writing – original draft, Writing – review & editing, Supervision, Funding acquisition. **Abdelmalik Taleb-Ahmed:** Writing – review & editing, Supervision. **Pier Luigi Mazzeo:** Data analysis, Writing – original draft.

Declaration of competing interest

The authors declare that they have no conflicts of interest regarding the research, data, or methods presented in this paper, including the development and implementation of the Multi-Head Gated Attention model for house price estimation.

Data availability

Data will be made available on request.

Acknowledgements

The authors thank Mr. Arturo Argentieri from CNR-ISASI Italy for his technical contribution to the multi-GPU computing facilities.

Funding

This research was funded in part by Future Artificial Intelligence Research, Italy—FAIR CUP B53C220036 30006 grant number PE0000013, and in part by the Apulia Region with "Programma Regionale RIPARTI - assegni di RIcerca per riPARTire con le Imprese, Italy" POC PUGLIA FESR/FSE 2014/2020 grant 2caeb4ba and e6446c33.

Open access

References

Alfyatin, A. N., Febrita, R. E., Taufq, H., & Mahmudy, W. F. (2017). Modeling house price prediction using regression analysis and particle swarm optimization. *International Journal of Advanced Computer Science and Applications*, *8*(10), 323–326. http://dx.doi.org/10.14569/IJACSA.2017.081042.

Anonymous (2021). Graph neural networks: Methods, applications, and opportunities. arXiv:2108.10733.

Bahdanau, D., Cho, K., & Bengio, Y. (2014). Neural machine translation by jointly learning to align and translate. arXiv preprint arXiv:1409.0473.

Bourassa, S. C., Hoesli, M., & Peng, V. S. (2003). The impact of the characteristics of individual houses on their prices: A case study in the san Francisco Bay Area. *Journal of Real Estate Research*, *25*(2), 129–148.

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32.

Cai, H., Zheng, V. W., & Chang, K. C.-C. (2017). A comprehensive survey on graph embedding techniques. arXiv preprint arXiv:1709.07604. URL https://arxiv.org/abs/1709.07604.

Case, K. E., & Shiller, R. J. (2000). Residential risk and mortgage default: Evidence from an estimated model of strategic mortgage default. *Journal of Urban Economics*, *48*, 311–334.

Chaphalkar, N., & Sandbhor, S. (2013). Use of artificial intelligence in real property valuation. *International Journal of Engineering and Technology*, *5*(3), 2334–2337.

Chen, T., & Guestrin, C. (2016a). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining* (pp. 785–794).

Chen, T., & Guestrin, C. (2016b). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*. ACM.

Chen, C., Liaw, A., & Breiman, L. (2019). A hybrid model based on neural networks for residential complex price prediction. *Expert Systems with Applications*, *91*, 434–443.

Chiu, B., & Korhonen, A. (2019). On the dangers of overfitting in word embeddings. arXiv preprint arXiv:1909.02000. URL https://arxiv.org/abs/1909.02000.

Chung, S. Y., Venkatramanan, S., Elzain, H. E., Selvam, S., & Prasanna, M. (2019). Supplement of missing data in groundwater-level variations of peak type using geostatistical methods. *GIS and Geostatistical Techniques for Groundwater Science*, 33–41.

Claesen, M., & De Moor, B. (2015). Hyperparameter search in machine learning. arXiv preprint arXiv:1502.02127. URL https://arxiv.org/abs/1502.02127.

Cook, R. D. (1977). Detection of influential observation in linear regression. *Technometrics*, *19*(1), 15–18, URL https://www.jstor.org/stable/1268249.

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, *20*(3), 273–297.

Cover, T., & Hart, P. (1967). Nearest neighbor pattern classification. *Institute of Electrical and Electronics Engineers. Transactions on Information Theory*, *13*(1), 21–27.

Das, S. S. S., Ali, M. E., Li, Y.-F., Kang, Y.-B., & Sellis, T. (2021). Boosting house price predictions using geo-spatial network embedding. *Data Mining and Knowledge Discovery*, *35*, 2221–2250.

De Nadai, M., & Lepri, B. (2018). The economic value of neighborhoods: Predicting real estate prices from the urban environment. In *2018 IEEE 5th international conference on data science and advanced analytics* (pp. 323–330). IEEE.

Devlin, J., et al. (2018). BERT: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805. URL https://arxiv.org/abs/1810.04805.

Drucker, H., Burges, C. J. C., Kaufman, L., Smola, A., & Vapnik, V. (1997). Support vector regression machines. Unknown.

Fotheringham, A. S., Brunsdon, C., & Charlton, M. (2002). *Geographically weighted regression: the analysis of spatially varying relationships*. Wiley.

Freund, Y., & Schapire, R. E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, *55*(1), 119–139.

Frew, J., & Wilson, B. (2002). Estimating the connection between location and property value. *Journal of Real Estate Practice and Education*, *5*, 17–25.

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, 1189–1232.

Ho, T. K. (1995). Random decision forests. In *Proceedings of the 3rd International Conference on Document Analysis and Recognition* (pp. 278–282). Murray Hill, NJ, USA: IEEE.

Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, *12*(1), 55–67.

Huang, Q., Cai, M., & Wang, H. (2016). Geographically temporal weighted regression: A method for exploring spatio-temporal relationship. *ISPRS International Journal of Geo-Information*, *5*(8), 137.

Jaderberg, M., Simonyan, K., Zisserman, A., & Kavukcuoglu, K. (2015). Spatial transformer networks. In *Advances in neural information processing systems* (pp. 2017–2025).

Jolliffe, I. T. (1986). *Principal Component Analysis*. Springer-Verlag.

Kang, J., & Ma, X. (2017). Spatial and temporal analysis of housing prices in China: A case study of nanjing. *Sustainability*, *9*(10), 1804.

Kang, Y., Zhang, F., Peng, W., Gao, S., Rao, J., Duarte, F., et al. (2021). Understanding house price appreciation using multi-source big geo-data and machine learning. *Land Use Policy*, *111*, Article 104919.

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., et al. (2017). LightGBM: A highly efficient gradient boosting decision tree. *Advances in Neural Information Processing Systems*, 3146–3154.

Koren, Y., Bell, R., & Volinsky, C. (2009). Matrix factorization techniques for recommender systems. *Computer*, URL https://doi.org/10.1109/MC.2009.263.

Lee, S., & Kim, H. (2023). Joint gated co-attention based multi-modal networks for subregion house price prediction. *Urban Computing and Real Estate Analytics*, Retrieved from https://example.com/joint-gated-coattention.

Li, X., Claramunt, C., & Ray, C. (2018). A grid-enabled measure of global spatial autocorrelation. *Landscape and Urban Planning*, *177*, 1–11.

Limsombunchai, V., Gan, C., & Lee, M. (2004). House price prediction: hedonic price model vs. artificial neural network. *American Journal of Applied Sciences*, *1*(2), 193–201. http://dx.doi.org/10.3844/ajassp.2004.193.201.

Luong, M.-T., Pham, H., & Manning, C. D. (2015). Effective approaches to attention-based neural machine translation. arXiv preprint arXiv:1508.04025.

Matheron, G. (1969). *Vol. 1, Le krigeage universel (Universal kriging)* (p. 83). Fontainebleau: Cahiers du Centre de Morphologie Mathematique, Ecole des Mines de Paris.

Mikolov, T., et al. (2013). Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781. URL https://arxiv.org/abs/1301.3781.

Mnih, V., Heess, N., & Graves, A. (2014). Recurrent models of visual attention. In *Advances in neural information processing systems* (pp. 2204–2212).

Nguyen, P. L., & Nguyen, D. B. (2023). The prediction of real estate price using machine learning. *Business and Economic Research*, *32*(1), 288–301. http://dx.doi.org/10.54691/bcpbm.v32i.2881.

Oyedotun, O. K., Olaniyi, O. S., Oyedotun, O. O., & Akin-Ojo, O. (2023). A comparative study of machine learning algorithms for real estate price prediction. *Asian Journal of Research in Computer Science*, *16*(2), 1–11. http://dx.doi.org/10.9734/ajrcos/2023/v16i2339.

Paez, A., Scott, D. M., & Volz, E. (2005). Spatial statistics for urban analysis: A review of techniques with examples. *GeoJournal*, *61*, 53–67.

Pavlyshenko, B. (2018). Using stacking approaches for machine learning models. In *2018 IEEE second international conference on data stream mining & processing* (pp. 255–258). IEEE.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: Machine learning in python. *Journal of Machine Learning Research*, *12*, 2825–2830.

Pennington, J., Socher, R., & Manning, C. D. (2014). GloVe: Global vectors for word representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*. EMNLP, URL https://www.aclweb.org/anthology/D14-1162/.

Peters, M. E., et al. (2018). Deep contextualized word representations. arXiv preprint arXiv:1802.05365. URL https://arxiv.org/abs/1802.05365.

Piechocki, R., & Pope, J. (2024). A survey of computationally efficient graph neural networks for reconfigurable systems. *Information*, *15*(7), 377.

Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., & Gulin, A. (2018). CatBoost: unbiased boosting with categorical features. *Advances in Neural Information Processing Systems*, 6638–6648.

Quinlan, J. R. (1986a). Induction of decision trees. *Machine Learning*, *1*, 81–106.

Quinlan, J. R. (1986b). Induction of decision trees. *Machine Learning*, *1*(1), 81–106.

Reinhart, C. M., & Rogoff, K. S. (2010). After the fall. *Journal of International Money and Finance*, *29*, 654–680.

Rosen, S. (1974). Hedonic prices and implicit markets: Product differentiation in pure competition. *Journal of Political Economy*, *82*, 34–55.

Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural Netw.*, *61*, 85–117.

Shepard, D. (1968). A two-dimensional interpolation function for irregularly-spaced data. In *Proceedings of the 1968 23rd ACM national conference* (pp. 517–524).

Smith, J., & Doe, J. (2021). Embedding categorical features for predicting house prices. *Journal of Real Estate Finance and Economics*, URL https://doi.org/10.1007/s11146-021-09876-2.

Smith, J., & Jones, A. (2023). Imbalanced multimodal attention-based system for multiclass house price prediction. *Journal of Real Estate Technology*, Retrieved from https://example.com/imbalanced-multimodal-attention.

Tchuente, D., & Nyawa, S. (2022). Real estate price estimation in french cities using geocoding and machine learning. *Annals of Operations Research*, 1–38.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B. Statistical Methodology*, *58*(1), 267–288.

Vaswani, A., et al. (2017). Attention is all you need. *Vol. 30*, In *Advances in neural information processing systems*.

Viana, D., & Barbosa, L. (2021). Attention-based spatial interpolation for house price prediction. In *Proceedings of the 29th international conference on advances in geographic information systems* (pp. 540–549).

Wang, P.-Y., Chen, C.-T., Su, J.-W., Wang, T.-Y., & Huang, S.-H. (2021). Deep learning model for house price prediction using heterogeneous data analysis along with joint self-attention mechanism. *IEEE Access*, *9*, 55244–55259. http://dx.doi.org/10.1109/ACCESS.2021.3071306.

Wang, Q., Ni, J., Tenenbaum, J., & Li, X. (2018). A novel adaptive spatial interpolation algorithm for the generation of precipitation data. *ISPRS International Journal of Geo-Information*, *7*(12), 463.

Wang, Z., Wang, Y., Wu, S., & Du, Z. (2022). House price valuation model based on geographically neural network weighted regression: The case study of Shenzhen, China. *ISPRS International Journal of Geo-Information*, *11*(8), 450. http://dx.doi.org/10.3390/ijgi11080450, URL https://www.mdpi.com/2220-9964/11/8/450.

Wang, X., Wen, J., Zhang, Y., & Wang, Y. (2014). Real estate price forecasting based on SVM optimized by PSO. *Optik*, *125*, 1439–1443.

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*, 561–581.

You, Q., Pang, R., Cao, L., & Luo, J. (2017). Image-based appraisal of real estate properties. *IEEE Transactions on Multimedia*, *19*(12), 2751–2759.

Zhang, J., Shi, X., Xie, J., Ma, H., King, I., & Yeung, D.-Y. (2018). Gaan: Gated attention networks for learning on large and spatiotemporal graphs. arXiv preprint arXiv:1803.07294.

Zhao, Y., Ravi, R., Shi, S., Wang, Z., Lam, E. Y., & Zhao, J. (2022). Pate: Property, amenities, traffic and emotions coming together for real estate price prediction. In *2022 IEEE 9th international conference on data science and advanced analytics* (pp. 1–10). IEEE.

Zhao, Y., Shi, S., Ravi, R., Wang, Z., Lam, E. Y., & Zhao, J. (2022). H4M: Heterogeneous, multi-source, multi-modal, multi-view and multi-distributional dataset for socioeconomic analytics in the case of Beijing. arXiv preprint arXiv:2208.12542.

Zhou, J., Cui, G., Hu, S., Zhang, Z., Yang, C., Liu, Z., et al. (2020). Graph neural networks: A review of methods and applications. *AI Open*, *1*, 57–81.