Estimation of Future Vacant Housing Distribution Considering Road Environment: An Approach Using Digital Road Maps and Machine Learning

Takahito Shimizu[1], Kotaro Mizutani[2], Yuki Akiyama[3]

[1]Tokyo City University, 1-28-1, Tamazutsumi, Setagaya-ku, Tokyo, Japan - g2581624@tcu.ac.jp
[2]Tokyo City University, 1-28-1, Tamazutsumi, Setagaya-ku, Tokyo, Japan - kmizu@tcu.ac.jp
[3]Tokyo City University, 1-28-1, Tamazutsumi, Setagaya-ku, Tokyo, Japan - akiyamay@tcu.ac.jp

Keywords: Vacant House, Digital Road Map, Machine Learning, City Block, Front Road

Abstract

The increase in vacant houses has become a serious social issue in many developed countries, including Japan. Therefore, to support mid- to long-term policy planning, there is a growing need to understand future vacancy distributions. In this study, we develop a machine learning model to predict future municipal-level vacancy rates by incorporating not only demographic and building information, but also spatial indicators related to road development conditions. First, we constructed a road mesh dataset at the 500-meter grid level by aggregating physical road indicators, block rectangularity, and the proportion of buildings with front roads. We then combined these variables with data from the Population Census and the Housing and Land Survey and developed a vacancy prediction model using LightGBM. The results show that incorporating road-related indicators improves prediction accuracy. In particular, we found that municipalities with a higher density of narrow roads, more irregularly shaped blocks, and a larger proportion of buildings lacking direct road access tend to have higher vacancy rates. This study demonstrates the value of road development information, which has received limited attention in previous research, in improving vacancy prediction, suggesting that road environments can influence the spatial distribution of vacant houses. Moreover, the findings of this study contribute to the early identification of areas at risk of high vacancy and the planning of preventive measures, thereby supporting urban management through the use of smart data.

1. Introduction

1.1 Background

In recent years, the increase in vacant houses has become a serious social issue in many cities of developed countries. For example, it has been reported that there are approximately 18.6 million vacant houses in the United States and more than 11 million across Europe (Lončar and Pavić, 2020). In Japan as well, the upward trend in vacant houses is notable, reaching a record high of approximately 9 million vacant houses in 2023, accounting for 13.8% of all residential properties (Ministry of Internal Affairs and Communications, 2024). Furthermore, some projections suggest that the vacancy rate could rise to as much as 30.4% within the next decade (Nomura Research Institute, 2016; Seirin-Lee et al., 2020).

One of the primary reasons why the increase in vacant houses has become a serious issue is its wide-ranging negative impact on urban environments. For example, it has been pointed out that vacant houses can lead to the deterioration of the landscape and public safety, as well as a decline in quality of life and heightened psychological anxiety among nearby residents (Garvin et al., 2013). Therefore, local governments are required to understand the distribution of vacant houses and implement appropriate countermeasures. In response to these concerns, the “Act on Special Measures concerning Promotion of Measures for Empty Houses” was enacted in 2015, legally mandating local governments across Japan to make efforts to address the vacant house problem. In order to implement such countermeasures, it is essential to accurately grasp the distribution and status of vacant houses. However, field surveys are still the most common method for identifying vacant houses, which involve considerable labor and costs (Martin et al., 2016; Jensen, 2017; Masita and Akiyama, 2020). To more efficiently capture the spatial distribution of vacant houses, recent studies have focused on building-level vacancy prediction using statistical data (e.g., Martin et al., 2016; Akiyama et al., 2021). Martin et al. (2016) developed a method to identify vacant buildings based on attributes such as structural type, construction year, and demographic variables including household composition. Although their approach demonstrated moderate predictive accuracy, it depended on datasets that are not consistently available across regions, thereby limiting its applicability to broader geographic scales.

Additionally, other studies have aimed to predict vacancy rates over broader geographic scales, such as at the municipal level Furthermore, some research has attempted to estimate not only current vacancy rates but also those anticipated in the future (e.g., Kanamori et al., 2015; Mizutani and Akiyama, 2023). These approaches typically rely exclusively on demographic variables such as household composition, and physical attributes of buildings, such as structural type and construction year. Although these methods offer high generalizability, they do not account for spatial factors in their analyses.

In contrast to the aforementioned studies on vacant housing, research addressing real estate prices, housing reconstruction, and surrounding living environments has highlighted the significant role played by spatial factors, such as road width and connectivity. For example, Aziz et al. (2023) demonstrated that narrower roads and complex road networks can negatively affect housing prices. Souza (2009) indicated that setback distances from front roads significantly influence property values. Furthermore, it has been pointed out that densely built-up urban districts—which are common in Japan and characterized by narrow roads and older housing stock—often face difficulties in housing reconstruction (Ando et al., 1997). Thus, road characteristics not only influence housing prices and the likelihood of reconstruction but also affect the broader residential environment.

These insights into the relationship between housing conditions and road infrastructure suggest that poorly developed road networks around residential areas may increase the risk of properties becoming and remaining vacant. For example, in neighbourhoods characterized by narrow streets or limited connectivity, access for large vehicles and emergency services may be restricted, creating physical barriers to housing reconstruction and maintenance. These constraints can hinder renovation or sale, potentially leading to stagnation in the local housing market. In turn, reduced market demand may lower the priority of upkeep for property owners, increasing the likelihood of neglect. As a result, such homes may remain unmanaged, raising the risk of becoming vacant or remaining so over the long term. In this way, inadequate road infrastructure can impede housing renewal, depress property values, and ultimately contribute to the emergence and persistence of vacant housing.

In fact, a few studies have explicitly examined the relationship between vacant houses and road infrastructure, suggesting that buildings with poor access to roads or those facing narrow roads are more likely to become vacant (Lee et al., 2022; Park, 2019). However, these studies rely on only a narrow set of indicators related to the road environment and do not sufficiently account for features such as road width, road network density, or street block geometry. As a result, it is difficult to comprehensively assess how the surrounding road environment influences the occurrence of vacant houses. Moreover, since the analyses are confined to specific cities or districts, their findings are likely to be shaped by local urban structures and context-specific factors. Therefore, the insights gained from these studies are insufficient to determine whether the observed patterns are generalizable across broader geographic scales.

As discussed above, while previous studies have offered valuable insights into the relationship between road environments and vacant housing, direct analyses of this relationship remain limited. Consequently, the influence of road infrastructure on the occurrence of vacant houses has not been thoroughly investigated. Furthermore, although road-related information has the potential to serve as an important predictor and enhance the accuracy of vacancy distribution models, it has rarely been incorporated into predictive methods for estimating the spatial distribution of vacant houses.

1.2 Study Objective

Against this background, this study aims to develop a machine learning model that predicts municipal-level vacancy rates with higher accuracy than previously proposed models (Mizutani et al, 2025), by incorporating indicators related to road infrastructure in addition to demographic data such as those from the Population Census. In addition, the study seeks to quantify the impact of road development conditions on the occurrence of vacant houses. To this end, we first constructed a road mesh dataset by aggregating metrics such as total road length and building frontage conditions at the grid-cell level using road network data. This dataset was then integrated with physical building attributes and demographic information to create a comprehensive dataset for vacancy prediction. Subsequently, we developed a machine learning model based on this integration data, to estimate future vacancy rates across government organizations throughout Japan.

The proposed method has three distinctive features. First, it is the first study to construct a detailed dataset on road environments across Japan using comprehensive road network data. This dataset includes not only physical characteristics of roads, but also building-to-road frontage conditions and the geometric regularity of city blocks. Second, it is the first attempt to incorporate road environment variables (such as road width, density, and frontage conditions) into a nationwide model for predicting the future spatial distribution of vacant houses. This approach enables more accurate predictions than conventional models and offers new insights into the impact of road infrastructure on vacancy dynamics. Third, because similar data are available in many other countries, the proposed method is adaptable to regions outside Japan that face comparable challenges related to vacant housing.

Moreover, the data-driven approach adopted in this study, which is based on detailed nationwide road network data and statistical datasets, contributes to the advancement of evidence-based policy making (EBPM) grounded in spatial information and smart data, and supports the realization of data-driven urban management. Specifically, the method developed in this study enables local governments to make timely and continuous decisions regarding where, when, and how to implement vacant housing countermeasures. In the conventional approach, understanding the distribution of vacant houses has required field surveys, which are both time-consuming and labour-intensive. Consequently, many existing countermeasures have been reactive and ad hoc, often failing to address the root causes of vacancy. In contrast, the proposed approach allows for the efficient and ongoing identification of priority areas for intervention by integrating continuously updated road network data with demographic and household statistics. This not only enhances the effectiveness of current measures but also contributes to the development of preventive and strategic urban policies aimed at reducing future vacancy. Furthermore, if the model developed in this study reveals the extent to which road development conditions influence vacancy rates, it may offer a new perspective for reinforcing conventional countermeasures. In particular, it could enable policymakers to incorporate road environment factors—such as road width and building-to-road proximity in residential areas—which have not been sufficiently considered in previous vacancy mitigation strategies.

2. Method

2.1 Study Flow

The study flow is shown in Figure 1. This study consists of three phases. First, Phase 1 involves the creation of mesh-based data on road development conditions. Next, Phase 2 involves constructing a dataset for machine learning. Finally, Phase 3 involves building a model and predicting future vacancy rates.

In Phase 1, as a preliminary step for aggregating road-related indicators at the municipal level, this study first compiles data based on the Digital Road Map (hereinafter referred to as "DRM-DB") and other sources. Specifically, the following indicators are aggregated for each 500-meter mesh: road length categorized by width, the degree of block rectangularity, and the rate of buildings that have a front road. Although this study adopts the municipality as the spatial unit for vacancy rate prediction, the aggregation is initially performed at the mesh level in anticipation of future applications at a micro scale, such as smaller administrative areas. By preparing generalized road-related data in advance, it becomes possible to extend the analysis to finer spatial units and also reduce computational costs.

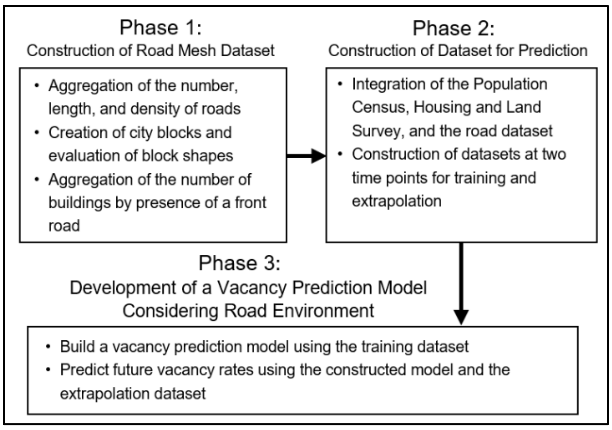

Figure 1. Study flow

In Phase 2, the mesh-level road development data created in Phase 1 is combined with data from the Population Census and the Housing and Land Survey to construct both the training dataset for model development and the extrapolation dataset for prediction.

In Phase 3, a machine learning model is developed using the training data to predict vacancy rates at the municipal level. The constructed model and the extrapolation dataset are then used to estimate future vacancy rates at the municipal level. Since this study covers all municipalities in Japan and includes large-scale validation, it is possible to build a generalized prediction model that is not dependent on any specific region.

2.2 Data

2.2.1 DRM-DB and Fundamental Geospatial Data

The data used in this study are listed in Table 1. Among the datasets shown in Table 1, the data used for aggregating road development conditions are those included in DRM-DB and the Fundamental Geospatial Data.

First, the “All Road Links” dataset included in DRM-DB was used to represent road geometry. This dataset expresses the road network as a combination of nodes and links and contains road data nationwide for roads with a width of 3.0 meters or more. Next, for representing the geographical shapes of rivers and lakes, the “Water System – Line” and “Water System – Polygon” datasets in DRM-DB were used. Rivers less than 50 meters wide are represented as lines and included in the “Water System – Line” dataset, while those 50 meters or wider are represented as polygons and included in the “Water System – Polygon” dataset. In addition, the “Railway Data” from DRM-DB was used to represent the geometry of railway lines.

For building geometries, the “Building Outlines” dataset from the Fundamental Geospatial Data was used. In this study, to target residential buildings, buildings classified as “ordinary buildings” with a footprint area of 25 square meters or more were used (Ministry of Land, Infrastructure, Transport and Tourism, 2021).

2.2.2 Population Census

The Population Census is conducted every five years to understand the actual conditions of the population and households throughout Japan, targeting all individuals and households residing in the country. The data obtained from the census include both household-level information, such as building characteristics and the presence of children within the household, and individual-level information, such as age and

Data Source	Data	Year
DRM-DB	All road links	2020
	Water system – line, polygon	
	Railway data	
Fundamental Geospatial Data	Building footprints	2020
Population Census	Population and households by gender	2015, 2020
	Population ratio by gender	
	Number of households by household size	
	Number of general households, household members, and single-person households by family type	
	Number of general households, household members, and single-person households by family type	
	Number of households by housing structure	
	Number of general households by economic status	
Housing and Land Survey	Total number of housing units	2018, 2023
	Number of vacant houses excluding rental, for-sale, and second homes (other vacant houses)	

Table 1. Data used in this study

occupation. As the types of data collected and published vary by census year, the variables used in this study were limited to those that were consistently available in both the 2015 and 2020 censuses. Additionally, since the aggregation and publication methods differ across years, certain variables were re-aggregated to ensure consistency in their format across the two time points. Furthermore, in the municipality-level data, each attribute is provided both as a total count and as a breakdown into subcategories. Therefore, for each available variable, this study calculated the proportion relative to the total count and used these ratios as features for model development. For variables where no relevant records exist in a given municipality, the value was treated as "Not applicable" and imputed as ”0”. Through these processes, the resulting features represent the proportion of households or individuals with specific attributes, which are used as explanatory variables in the model.

2.2.3 Housing and Land Survey

The Housing and Land Survey is conducted every five years to assess housing conditions and residential usage throughout Japan. The survey covers all cities, towns, and villages with a population of 15,000 or more.

This survey includes data on the number of houses with occupied households, the number of houses without occupied households, and the number of houses under construction in each municipality. Vacant houses are included among the houses without occupied households. The survey classifies vacant houses into four categories: secondary residences, rental properties, properties for sale, and vacant houses that exclude rental properties, properties for sale, and secondary residences (hereinafter referred to as “other vacant houses”).

https://doi.org/10.5194/isprs-annals-X-4-W7-2025-113-2025

Secondary residences refer to vacation houses such as resort condominiums and second homes that are not regularly occupied but are intended for temporary use within a certain period. Rental properties and properties for sale refer to unoccupied houses that are intended to be leased or sold. Although these are classified as vacant houses, they are managed by a clearly identified owner or administrator and are expected to be used or inspected as necessary within a reasonable timeframe.

In contrast, “other vacant houses” refer to houses that have been unoccupied for a long time, with no plans for reuse or circulation, and are effectively abandoned. These houses often lack a clearly identified manager, particularly when the vacancy has been prolonged. Such properties may pose risks to nearby communities, including structural deterioration or the spread of overgrown vegetation.

Therefore, in this study, we focus on this specific category of vacant houses—those that are likely to require municipal intervention in the future due to lack of proper management. We define the vacancy rate as the proportion of “other vacant houses” relative to the total number of housing units. This vacancy rate is used as the target variable for the machine learning model developed in this study. By doing so, the model aims to estimate the extent to which unmanageable vacant houses, which are likely to necessitate municipal action, comprise the overall housing stock.

2.3 Construction of Road Mesh Dataset

2.3.1 Overview of Road Mesh Dataset

Road data is typically provided as line data. However, line data alone makes it difficult to capture spatial characteristics such as total road length or density, making it challenging to obtain information on which types of roads are prevalent in which areas from an areal perspective. Therefore, to use road-related indicators as features for predicting vacancy rates at the municipal level, which is the primary objective of this study, it is necessary to convert line-based road data into areal data.

To address this issue, we developed a dataset that captures road development conditions by aggregating multiple indicators at the 500-meter mesh level. The dataset was constructed using DRM-DB and the Fundamental Geospatial Data, and includes variables such as the number, total length, and density of roads by width category, block rectangularity, and the number of buildings classified by the presence or absence of a front road. Table 2 presents the list of attributes included in the road dataset.

In selecting road environmental variables, we aimed to capture spatial conditions that are assumed to affect vacancy rates. Narrow road widths may limit reconstruction and emergency response in the event of a disaster. Irregular block shapes may lead to inefficient land use and reduced resident satisfaction.
Furthermore, buildings without access to front roads may face barriers to reconstruction, such as legal or large vehicle access restrictions. The subsequent sections provide detailed explanations of the aggregation methods used for each attribute.

2.3.2 Aggregation of the Number, Length, and Density of Roads

The number of roads and total road length were calculated based on the number of road links and their lengths, categorized by width, within each mesh. Road density was then calculated by counting road endpoints located along the top, bottom, left, and right edges of each mesh, also by width category.

Data Category	Variable name	Data source
Road	Number of Roads by Width Category	DRM-DB
	Total Road Length by Width Category	
	Road Density by Width Category	
City Block	City Block Rectangularity	
	Number of Buildings by City Block Rectangularity Category	
Front Road	Number of Buildings by Front Road Presence	Fundamental Geospatial Data

Table 2. Attribute list of the road dataset

2.3.3 Creation of Block Data

The shape of city blocks significantly influences building development and walkability. Moreover, irregularly shaped blocks tend to lead to inefficient development and may also result in a lower-quality residential environment (Shpuza, 2021).
Therefore, in this study, city blocks were generated using data included in the DRM-DB, and their shape were evaluated. For block generation, road line data, railway data, and river data represented as lines were used. In general, city blocks are considered spatial units that make up urban areas and are defined as regions bounded by geographic features such as roads, railways, and rivers. Based on this concept, these three datasets were first merged into a single geometry. This process produced polygonal regions enclosed by roads, railways, and rivers, which were treated as candidate city blocks. Although many of the generated polygons could be regarded as valid city blocks, some were not appropriate. For example, polygons overlapping with lakes or those covering vast areas in mountainous regions with sparse road networks could not be considered actual city blocks. To remove such unsuitable polygons, we first excluded those that spatially intersected with area-based water body data. Then, based on the interquartile range (IQR) of polygon areas, we identified outliers as those exceeding 1.5 times the IQR and excluded them from the dataset. The remaining polygons were defined as the final set of city blocks.

2.3.4 Evaluation of Block Shape

Next, to evaluate the shape of each city block, we calculated the rectangularity, an index that quantifies how closely a polygon approximates a rectangle. Rectangularity is calculated using the following equation (1).

$$Rectangularity = \frac{Area\ of\ the\ city\ block}{Area\ of\ the\ minimum\ bounding\ rectangle} \tag{1}$$

Rectangularity takes a value between 0 and 1, with values closer to 1 indicating that the shape is more similar to a rectangle. Such blocks are considered to be more suitable for development.

2.3.5 Aggregation of Building Counts by Presence Front Roads

The presence and width of the front road for each building significantly affect its value and the likelihood of redevelopment (Miyakawa 2018). In this study, we defined a building as having a front road if the distance from the centroid of the building polygon, as provided in the Fundamental Geospatial Data, to the nearest road was less than 15 meters. If the distance was 15 meters or more, the building was considered to lack a front road. We then aggregated the number of buildings in each mesh, as well as the number of buildings classified by the presence or absence of a front road and by front road width.

2.4 Development of a Predictive Model for Future Vacant Housing Distribution

2.4.1 Overview of the Predictive Model for Future Vacant Housing Distribution

In this study, we used a dataset for vacancy prediction that combines the variables used in Mizutani et al. (2025) with the road data developed in Section 2.3. Figure 2 illustrates the process of building the model to predict vacancy rates three years into the future. First, we build a machine learning model using information on population and households from the 2015 Population Census, along with the road dataset, in order to predict vacancy rates based on the 2018 Housing and Land Survey. Next, we apply the trained model to the population, household, and road data from the 2020 Population Census. Under the assumption that the vacancy generation mechanism remains largely unchanged, this approach enables us to estimate municipal-level vacancy rates for 2023, three years after 2020. Throughout both the training and extrapolation stages, we use variables derived from the DRM-DB (2020), based on the assumption that road conditions at the municipal scale will not change significantly in the near future.

2.4.2 Construction of a Dataset for Machine Learning

Each municipality in the Population Census and the Housing and Land Survey has a unique corresponding code, which enables one-to-one matching between the two datasets. In cases where a match was not possible due to municipal mergers or name changes, we adjusted the data appropriately to ensure that the resulting statistics reflect actual conditions. Next, we joined the road mesh data created in Section 2.3 by collecting the mesh IDs corresponding to each municipality and aggregating the relevant mesh-level information from the road dataset. Through this process, we integrated the Population Census, Housing and Land Survey, and road dataset at the municipal level, thereby constructing the dataset used for building the vacancy prediction model. Consequently, in addition to the variables such as age group, number of households, and family type used in Mizutani et al. (2025), we also included road environment factors listed in Table 3. The final input dataset for the vacancy rate prediction model consists of these variables.

2.4.3 Construction of the Machine Learning Model

We used LightGBM, a machine learning regression method based on gradient boosting decision trees, as proposed by Ke et al. (2017), to build our model. LightGBM combines multiple weak learners and iteratively learns from prediction errors to minimize them. The algorithm adopts a method called "leaf-wise" growth, which splits from the leaf that achieves the greatest reduction in loss. Additionally, it reduces computational cost by using only a subset of samples with small residuals during training. These characteristics allow LightGBM to achieve high prediction accuracy with shorter training times compared to other methods.

We adopted LightGBM in this study for the following reasons. First, gradient boosting decision tree methods such as LightGBM are well suited for building predictive models that capture complex real-world relationships. For example, vacancy rates may exhibit nonlinear patterns, with significant differences observed between single-person households and multi-person households. Simple regression analysis cannot adequately capture such nonlinear relationships. In contrast, LightGBM uses a decision tree–based approach, which allows the model to flexibly adapt to changes in feature values and achieve high prediction accuracy.

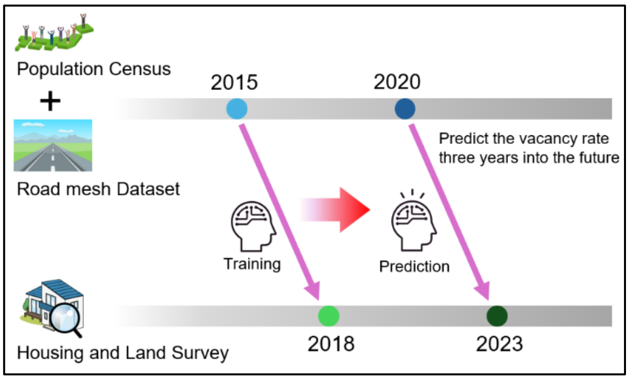

Figure 2. Overview of the vacancy prediction model

Category	Variable
Road	Total length of roads (width less than 3 meters)
	Total length of roads (width from 3 to 5.5 meters)
	Total length of roads (width from 5.5 to 13 meters)
	Total length of roads (width more than 13 meters)
	Number of roads (width less than 3 meters)
	Number of roads (width from 3 to 5.5 meters)
	Number of roads (width from 5.5 to 13 meters)
	Number of roads (width more than 13 meters)
	Road density (width less than 3 meters)
	Road density (width from 3 to 5.5 meters)
	Road density (width from 5.5 to 13 meters)
	Road density (width more than 13 meters)
Building	Number of buildings
	Proportion of buildings with a front road
City Block	Proportion of city blocks with rectangularity (0.0–0.2)
	Proportion of city blocks with rectangularity (0.2–0.4)
	Proportion of city blocks with rectangularity (0.4–0.6)
	Proportion of city blocks with rectangularity (0.6–0.8)
	Proportion of city blocks with rectangularity (0.8–1.0)

Table 3. List of Variables Used in the Model

Second, previous studies across various fields, both in Japan and internationally, have frequently reported that LightGBM delivers strong performance (e.g., Zhang et al., 2019; Takeda et al., 2022).

3. Result

3.1 Results of Road Dataset Construction

In this section, we examine part of the results from the development of the road mesh dataset. Figure 3 shows the average rectangularity for each mesh across Japan. We observe that many city blocks in the Tokyo metropolitan area exhibit high

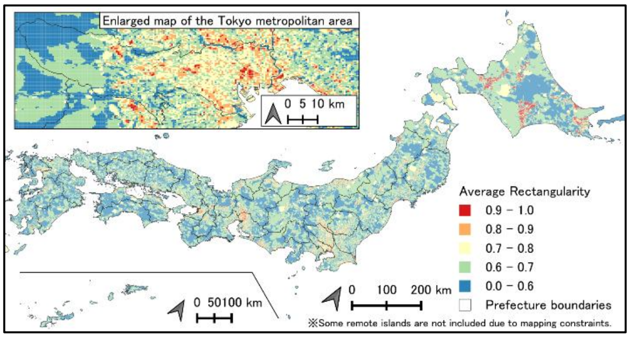

Figure 3. Average Rectangularity (500m Mesh)

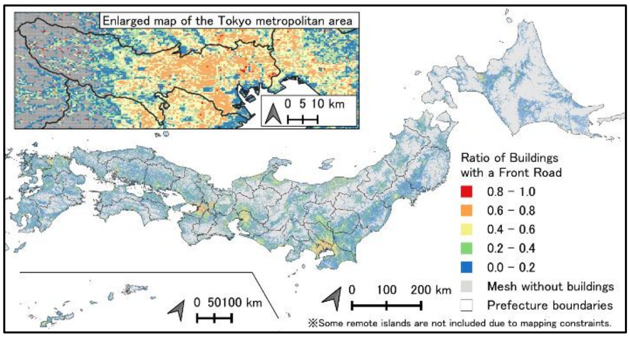

Figure 4. Ratio of Buildings with a Front Road (500m Mesh)

rectangularity. In particular, several areas within the 23 special wards of Tokyo, where large-scale redevelopment projects have taken place since the 1990s, show especially high values. We also find city blocks with high rectangularity in the central areas of major cities such as Osaka and Nagoya, where redevelopment has occurred. In contrast, suburban areas, rural regions, and mountainous zones, where narrow streets from earlier periods still remain without significant improvement, tend to have many meshes with low rectangularity.

Next, Figure 4 shows the ratio of buildings with a front road for each mesh. As with the rectangularity, it can be seen that in large cities such as Tokyo, there are many buildings with front roads and a large number of land and buildings with relatively good access by car tend to be distributed. On the other hand, in other areas, the proportion of front roads tends to be higher than in large cities and their surrounding areas. This may be due to the fact that the front roads in these areas are truly narrow, or the setback distance is greater than 15 m because the sites where buildings are located are flagpole lots.

3.2 Results of Predicting Future Vacant Housing Distribution

First, we examine the prediction accuracy. This study evaluates how much the inclusion of variables from the road dataset affects prediction performance and whether road narrowness can serve as a factor in vacancy occurrence. To do this, we compare the performance of two machine learning models: one that excludes variables from the narrow road dataset and one that includes them. For both models, we use data from the 2015 Population Census, the 2018 Housing and Land Survey, and the road dataset. We split the samples into 85% for training and 15% for testing. We evaluate prediction accuracy by calculating the generalization error based on the difference between the predicted and actual values in the test data. We also predict vacancy rates for 2023 using both models and assess their accuracy by comparing the

Presence/Absence of Road-Related Variables	Test Data		Extrapolation (Prediction for 2023)	
	R^2	MAE	R^2	MAE
Presence	0.860	0.0136	0.804	0.0172
Absence	0.849	0.0142	0.770	0.0183

Table 4. Comparison of Prediction Accuracy with and without Road Variables

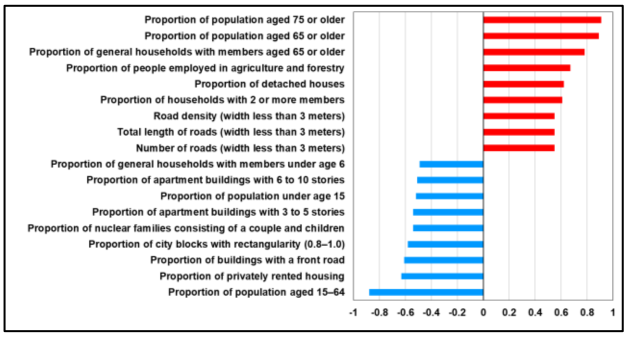

Figure 5. Top 9 Explanatory Variables Positively and Negatively Correlated with Vacancy Rates

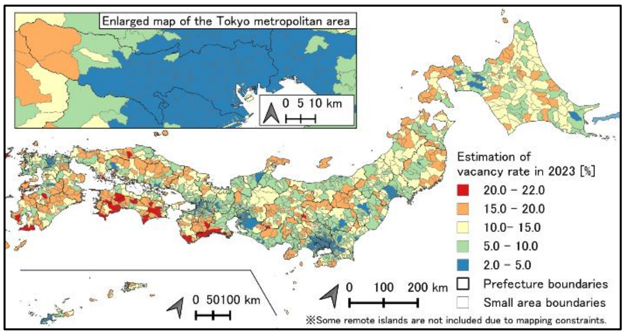

Figure 6. Estimation results of the vacancy rate in each municipality (2023)

predicted values with the actual values from the Housing and Land Survey. Table 4 presents the prediction accuracy for each model.

We confirmed that both models achieved a high level of accuracy, with the coefficient of determination (R^2) exceeding 0.8 regardless of whether the model included indicators related to road development conditions. In addition, in the extrapolation results, the R^2 value remained close to 0.8, demonstrating robust predictive performance even when applied to future data. In addition, incorporating variables derived from the road mesh dataset developed in this study led to improvement in prediction accuracy.

Next, we examine how road development conditions influence vacancy rates. Figure 5 shows the top nine explanatory variables with the highest positive and negative correlation coefficients with the predicted vacancy rates. The results indicate that a greater total length, number, and density of narrow roads with a width of less than 3 meters positively influence vacancy rates. In contrast, a higher city block rectangularity and a greater number of buildings with front roads negatively influence vacancy rates. That is, areas with a high concentration of narrow roads tend to have higher vacancy rates, while areas with better building-road connections and more regular block shapes tend to have lower vacancy rates.
values from the Housing and Land Survey. Table 4 presents the prediction accuracy for each model.

Finally, Figure 6 shows the estimated vacancy rate for each municipality in 2023. As seen in the map, vacancy rates tend to be lower in major metropolitan areas such as Tokyo (shown in the enlarged map in the upper left), Osaka, and Nagoya, and in their surrounding suburbs. In contrast, higher vacancy rates are observed in mountainous regions and in prefectures located in the western part of Japan , where depopulation is more severe.

3.3 Discussion

First, the improved prediction accuracy obtained by incorporating indicators related to road development conditions into the model suggests that road conditions around buildings can serve as valuable information for predicting the distribution and occurrence of vacant houses. Furthermore, the correlations shown in Figure 5 indicate that vacancy rates are primarily influenced by a high proportion of elderly residents, a greater presence of narrow roads, irregular block shapes, and limited building access to front roads. This result supports the hypothesis proposed in this study that road development conditions affect vacancy rates, highlighting their important role in understanding vacancy patterns across municipalities. It is also consistent with the observations obtained from interviews with local governments, as mentioned in Section 1.1.

On the other hand, incorporating variables from the road mesh dataset into the model did not significantly improve prediction accuracy. This is likely because, as shown in Figure 5, population-related variables, especially the proportion of elderly residents within each municipality, explain much of the variation in the number of vacant houses. Nevertheless, these variables enhance the interpretability of the model by quantitatively identifying specific spatial factors—such as narrow roads and irregular block shapes—that are associated with higher vacancy rates. This provides valuable insights for policymakers and local governments, as it highlights the role of road environment conditions in shaping vacant housing trends and informs the prioritization of intervention areas.

In this study, we used municipalities as the spatial unit of analysis, which represents a relatively macro scale. However, as the results in Section 2.6.1 show, spatial factors such as road width, building-road connections, and block shapes vary greatly at the micro scale, such as the level of individual buildings. These factors influence residential environments and property values. Therefore, when estimating vacancy at finer spatial scales in future research, such as at the small-area or building level, incorporating information from the road mesh dataset as explanatory variables can help more accurately represent local locational characteristics and residential conditions. This may lead to improved prediction accuracy.

4. Conclusion

In this study, we aimed to improve the accuracy of future vacancy distribution estimation and clarify how road development conditions affect vacancy rates. Previous models mainly relied on municipal-level population and household attributes from the Population Census as explanatory variables. In contrast, by incorporating the newly constructed road mesh dataset into the model, we enabled the learning of features that represent the road conditions of each municipality. As a result, we made it possible to predict nationwide vacancy rates while accounting for road development conditions, and we observed a slight improvement in prediction accuracy compared to conventional models. Furthermore, municipalities with longer and denser networks of narrow roads tended to have higher vacancy rates, while those with more regular block shapes and a greater number of buildings with front roads tended to have lower vacancy rates.

We plan to further improve prediction accuracy by enhancing the explanatory variables, for example by incorporating temporal changes in variables derived from the road mesh dataset and adding information on narrower roads. We also aim to refine the model so that it can estimate vacancy rates at finer spatial scales, such as sub-municipal areas. This refinement may deepen the insights gained in this study, as discussed in Section 3.3.

The findings of this study support evidence-based policy making (EBPM) and promote efficient urban management using smart data. By visualizing areas with high or increasing numbers of vacant houses, regions requiring priority interventions can be identified. The study also reveals how road development conditions affect vacancy, offering a new perspective beyond conventional demographic factors. In particular, areas with dense narrow roads, irregular blocks, and poor building access tend to have higher vacancy risks. Improving such road networks—e.g., by widening streets or reorganizing blocks—can help reduce future vacancies and guide strategic urban planning.

In addition, many countries maintain population censuses and road-related data similar to those used in this study. Therefore, the proposed approach may be applicable beyond Japan, particularly in East Asia and other regions where vacant housing is a pressing issue. To apply this method effectively in diverse urban contexts, however, several considerations are necessary. For instance, the concept of building frontage may be less clearly defined in some developing countries due to informal address systems and unregulated land use. Moreover, vacancy in such areas may be influenced more by socioeconomic factors—such as local insecurity, poverty, or rapid urban migration—than by physical access constraints. To ensure accurate forecasting and effective policy application, it is important to tailor the model to reflect the dominant local drivers of vacancy.

Acknowledgements

This study was supported by a grant from the Japan Digital Road Map Association (DRM Association) in 2025. The DRM-DB used in this study was provided by the DRM Association. We express our sincere appreciation for their contribution to this study.

References

Akiyama, Y., Baba, H., Ono, Y., Takaoka, H., 2021: Sophisticat ion of monitoring method for spatial distribution of vacant hous es using machine learning. Journal of Architecture and Planning (Transactions of AIJ), 86(786), 2136–2146. https://doi.org/10.3 130/aija.86.2136

Ando, M., Kouda, M., Sakamoto, S., 1997: A study on the diffic ulty of housing reconstruction in small sites facing on narrow ro ads of densely built-up areas with wooden structures — A case s tudy of West-Suma area struck by the Hanshin earthquake. Jour nal of the City Planning Institute of Japan, 32, 751–756. https:// doi.org/10.11361/journalcpij.32.751

Aziz, A., Anwar, M. M., Abdo, H. G., Almohamad, H., Al Dugh airi, A. A., Al-Mutiry, M., 2023: Proximity to Neighborhood Se rvices and Property Values in Urban Area: An Evaluation throu gh the Hedonic Pricing Model. Land, 12(4). https://doi.org/10.3 390/land12040859

Garvin, E., Branas, C., Keddem, S., Sellman, J., Cannuscio, C., 2013: More than just an eyesore: Local insights and solutions on vacant land and urban health. Journal of Urban Health, 90(3), 4 12–426. https://doi.org/10.1007/s11524-012-9782-7

Jensen, J. O., 2017: Vacant houses in Denmark: Problems, locali zation and initiatives. ENHR Conference 2017, Tirana, Albania.

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., Liu, T.-Y., 2017: LightGBM: A Highly Efficient Gradient B oosting Decision Tree. Advances in Neural Information Processi ng Systems, 30, 3149–3157. https://dl.acm.org/doi/10.5555/329 4996.3295074

Mizutani, K., Akiyama, Y., Baba, H., 2025. Estimation of future vacancy rates throughout Japan using government statistics. Ur ban and Regional Planning Review, 12, 25–43. https://doi.org/1 0.14398/urpr.12.25

Lee, J., Newman, G., Lee, C., 2022: Predicting Detached Housin g Vacancy: A Multilevel Analysis. Sustainability, 14(2). https:// doi.org/10.3390/su14020922

Lončar, S., Pavić, D., 2020: Areas with abandoned and vacant p roperties in Croatia: A plea for recognition, research, policies an d the development of regeneration strategies. Etnološka Tribina, 50(43), 197–237. https://doi.org/10.15378/1848-9540.2020.43. 09

Martin, H., Oduro, I., García-Cobián Richter, F., Urban, A. H., Whitaker, S. D., 2016: Predictive Modeling of Surveyed Propert y Conditions and Vacancy. Federal Reserve Bank of Cleveland, Working Paper Series, WP 16-37. https://doi.org/10.26509/frbc-wp-201637

Mashita, M., Akiyama, Y., 2020: The recent trend of housing va cancy studies in Japan. Geographical Space, 13(1), 1–26. https:// doi.org/10.24586/jags.13.1_1
Ministry of Internal Affairs and Communications, 2024: Prelimi nary Summary of the Housing and Land Statistics Survey for 20 23.

Miyagawa, T., Asami, Y., Hino, K., Taima, M., Usui, H., 2018: Relationship between the occurrence of building renewal and re sidential environments in Tokyo Wards Area. Journal of the Cit y Planning Institute of Japan, 53(3), 1485–1490.

Mizutani, K., Akiyama, Y., 2023: Estimation of future vacancy r ates throughout Japan using government statistical information. Proceedings of the 2023 Asian Conference on Remote Sensing (ACRS 2023), Paper No. ACRS2023263, 316–324.

Nomura Research Institute, 2016: News Release. https://www.n ri.com/content/900036619.pdf (18 April 2025).

Pan, J., Dong, L., 2021: Spatial Identification of Housing Vacan cy in China. Chinese Geographical Science, 31(2), 359–375. htt ps://doi.org/10.1007/s11769-020-1171-7

Park, J. Il., 2019: A multilevel model approach for assessing the effects of house and neighborhood characteristics on housing va cancy: A case of Daegu, South Korea. Sustainability, 11(9). http s://doi.org/10.3390/su11092515

Seirin-Lee, S., Nomata, M., Mukunoki, M., 2020: Mathematical modeling and regionality-based optimal policy to reduce empty houses, Akiya, in Japan. Japan Journal of Industrial and Applie d Mathematics, 37(2), 365–382. https://doi.org/10.1007/s13160-019-00400-3

Shpuza, E., 2021: Block shape in historic cities: A classification based on compactness and fragmentation. ISUF 2020 Virtual C onference Proceedings, Vol. 1. https://doi.org/10.26051/0D-9N1 5-6Q4F

Souza, M. T., 2011: The effect of land use regulation on housing price and informality: A model applied to Curitiba, Brazil. SSR N Working Paper, Lincoln Institute of Land Policy. http://ssrn.c om/abstract=1762658

Takeda, N., Furuya, T., Akiyama, Y., 2022: Development of esti mation method for building structure using open data and statisti cs. International Geoscience and Remote Sensing Symposium (I GARSS), 2022-July, 2438–2441. https://doi.org/10.1109/IGAR SS46834.2022.9883801

Zhang, J., Mucs, D., Norinder, U., Svensson, F., 2019: LightGB M: An effective and scalable algorithm for prediction of chemic al toxicity – Application to the Tox21 and mutagenicity datasets. Journal of Chemical Information and Modeling, 59(10), 4150–4 158. https://pubs.acs.org/doi/10.1021/acs.jcim.9b00633