



Interpreting a Multimodal Late-Fusion Model for Municipality-Level Community Bus and Demand-Responsive Bus Classification in Kyushu Area Using SHAP and Grad CAM

Raden Muhammad Azhar Ardlin Syah¹, Syuji Yoshiki², Hera Widyastuti³

¹Master's Student of Civil Engineering Department, Faculty of Civil, Planning, and Geo Engineering, at Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia

²Associate Professor of Center for Water Cycle, Marine Environment and Disaster Management at Kumamoto University, Kumamoto, Japan

³Professor of Transportation and Pavement Material Laboratory at Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia

*Corresponding Author: Raden Muhammad Azhar Ardlin Syah

E-mail: rdennngff@gmail.com



Article Info

Article history:

Received 17 August 2026

Received in revised form 2

September 2026

Accepted 28 September 2026

Keywords:

Multimodal Learning

SHAP

Grad-CAM Community Bus

Demand-Responsive Bus

Abstract

Limited public transport accessibility remains a serious issue in aging and low-density areas such as Japan. Local governments have introduced Community Bus (CB) and Demand-Responsive Bus (DRB) services, but methods for explaining the regional patterns associated with these two service types remain limited. This paper examines the interpretability of a multimodal late-fusion model that integrates tabular regional indicators and Geographic Information System (GIS)-based spatial images for municipality-level CB and DRB classification. The research used municipalities in Kyushu, Japan. A multimodal late-fusion architecture was adopted as a commonly used integration strategy for image and tabular data. Model interpretation was conducted using SHAP for the tabular branch and Grad-CAM for the image branch. Performance was summarized using Macro F1-score as the main metric. The late-fusion model achieved a mean Macro F1-score of 0.5978 and a mean accuracy of 0.6797 under the non-TTA setting. TTA was evaluated as an inference-time robustness strategy and showed slightly smaller fold-wise variability, although the non-TTA setting produced higher mean predictive scores. Therefore, the subsequent interpretation analysis was based on the non-TTA results. SHAP showed that single elder household density, lowest elevation, elder rate, highest elevation, and total population were among the most influential features. Class-specific SHAP results also showed differences between correctly classified CB and DRB municipalities. Grad-CAM indicated that the model attended to meaningful internal spatial regions, although its interpretation was more qualitative than SHAP. These results show that multimodal late fusion can provide interpretable evidence for municipality-level CB and DRB classification and support more transparent regional transport analysis.

Introduction

Limited public transportation accessibility remains a serious problem in rural and suburban areas. The problem becomes more severe in regions with declining population, low density, and rapid aging, where fixed-route services are increasingly difficult to maintain efficiently over time. Japan is a clear example of this situation because many local governments face

long-term demographic decline while still being required to maintain basic mobility within financially constrained operating conditions.

To respond to these problems, local governments in Japan have introduced alternative service types such as Community Bus (CB) and Demand-Responsive Bus (DRB). These services are intended to support local mobility where conventional fixed-route operations are no longer efficient or financially sustainable. Their implementation, however, does not follow a single universal pattern. Research on Japanese local transport planning has shown that the selection and operation of CB and DRB differ across municipalities in terms of objectives, planning logic, and evaluation criteria (Takeuchi & Yoshida, 2022; Takeuchi, Yoshida, & Inoi, 2023; Takeuchi, Yoshida, & Ozaki, 2023). This makes CB and DRB classification an important analytical issue at the municipality level.

Although CB and DRB are both used to maintain local mobility, they differ in operational structure. CB generally follows relatively fixed routes, fixed stops, and scheduled operation, while DRB offers more flexible service adjusted to user requests, service zones, or low and fluctuating demand (D. Hasegawa & Suzuki, 2011; Takeuchi et al., 2019). As a result, DRB is often considered more suitable in areas where demand is dispersed and conventional route-based services are inefficient, whereas CB is more feasible where demand is sufficiently concentrated to sustain route-based service provision (Vansteenwegen et al., 2022; Ise & Hino, 2023; Liyanage et al., 2024).

Recent research on CB and DRB has approached this issue from several directions. Some research has focused on user needs, service acceptance, and mobility behavior, showing that service effectiveness depends on both individual conditions and regional context, and that no single service type can satisfy all mobility needs equally well (M. Hasegawa et al., 2021; Nakanishi & Matsumoto, 2024; Kato, 2025). Further research has also considered integration between CB and DRB, showing that DRB can function effectively as a complementary service in transfer-based local transport systems (Nakanishi & Matsumoto, 2025). These findings clarify the practical roles of CB and DRB, but they do not yet provide a sufficiently objective framework for identifying municipality-level patterns associated with each service type.

Previous analytical work has also examined municipality-level transport conditions using structured regional indicators. This line of research has clarified important quantitative relationships between regional characteristics and transport service conditions (D. Hasegawa & Suzuki, 2011; Higashimoto et al., 2013; Hasegawa D. & Suzuki, 2019; Coutinho et al., 2020; Rich et al., 2023; Wang et al., 2023; Kim et al., 2025). These approaches remain useful, but they rely mainly on tabular variables and do not fully capture spatial patterns visible in municipal maps.

Recent progress in multimodal learning provides a way to address this gap. Research has shown that combining image and tabular data can improve representation quality when the two modalities capture complementary aspects of the same object, although the performance depends on appropriate interaction design and input representation (Atrey et al., 2010; Baltrušaitis et al., 2018; Velmurugan, 2023; Du et al., 2024; Blomqvist, 2025).

Interpretability is essential in this context because predictive performance alone is not sufficient for transport analysis. SHAP explains the contribution of tabular variables to the model prediction (Lundberg & Lee, 2017). For image inputs, Grad-CAM highlights the spatial regions receiving the strongest model attention (Selvaraju et al., 2020). These methods have become increasingly relevant in transport and geospatial modeling, including explainable mode-choice analysis and GIS-based deep learning interpretation (Xing & Sieber, 2023; Dahmen et al., 2024; Tang et al., 2024; Deressu et al., 2025).

Based on this background, this paper examines a multimodal late-fusion model for municipality-level CB and DRB classification in Kyushu, Japan. Late fusion was selected because it is a commonly used multimodal integration strategy and provides a clear structure for branch-specific interpretation (Atrey et al., 2010; Baltrušaitis et al., 2018). The model integrates tabular indicators and GIS-based spatial images, then interprets the tabular branch with SHAP and the image branch with Grad-CAM. The objective is to identify the municipality-level variables and spatial image regions most strongly associated with CB and DRB classification. The working hypothesis is that late fusion captures complementary numerical and spatial information that can be interpreted to reveal meaningful municipality-level patterns.

Methods

This paper focuses on municipalities in Kyushu, Japan, where local public transport provision has been shaped by population decline, aging, and differences in settlement structure. Kyushu was selected because it contains a broad range of municipality conditions, from relatively dense urban areas to sparsely populated rural areas, making it suitable for examining municipality-level patterns associated with CB and DRB classifications. Kyushu consisted of 233 municipalities. However, only 106 municipalities were included in the labeled dataset used. A municipality was included only when its service condition could be assigned clearly to a single category, namely CB or DRB. Municipalities were excluded if they operated both service types or if they did not operate either service type. Based on this criterion, the final labeled dataset consisted of 41 CB municipalities and 65 DRB municipalities. The CB and DRB labels were collected from the official websites of the respective local governments in January 2026. In the classification model, CB was coded as 0, and DRB was coded as 1. Each municipality was represented by two input modalities: a set of tabular regional indicators and a GIS-based spatial image. The tabular data were compiled at the municipality level from official Japanese data sources and then organized into a unified dataset for multimodal analysis. All tabular features were collected in May 2026. The tabular variables used in this paper are summarized in Table 1.

Table 1. Tabular features used in research

Category	Feature	Raw Unit	Unit used	Conversion /Formula	Year	Source
Demographic / Household	Total Elder	People	People	Provided by source	2020	E-Stat
	Total Population	People	People	Provided by source	2020	E-Stat
	Single Elder Household	Household	Household	Provided by source	2020	E-Stat
	Total Household	Household	Household	Provided by source	2020	E-Stat
	Elder Rate	Ratio	Ratio	Provided by source	2020	E-Stat
	Single Elder Household Rate	Ratio	Ratio	Provided by source	2020	E-Stat
Mobility / Vehicle / Service	Car Ownership	Units	Units	Provided by source	2023	E-Stat
	Bus Share	-	Ratio	$\frac{\text{Use bus Commuters and students traveling}}{\text{}} \times 100\%$	2020	E-Stat
	Car Share	-	Ratio	$\frac{\text{Use car Commuters and students traveling}}{\text{}} \times 100\%$	2020	E-Stat
	Total Vehicle	-	Units	Car + Small two-wheeled vehicles +	2023	E-Stat

				Motorized bicycle + Two-wheeled vehicles + Three-wheeled vehicles		
	Total Bus Stop	Bus stop	Bus stop	Provided by source	2022	MLIT
	Route Length	m	Km	m/1000	2022	MLIT
Density / Built Intensity	Density	People /Km ²	People /Km ²	Provided by source	2020	E-Stat
	Household Density	-	Household /Km ²	$\frac{\text{Household}}{\text{Habitable area}}$	2020	E-Stat
	Single Elder Household Density	-	Household /Km ²	$\frac{\text{Single Elder Household}}{\text{Habitable area}}$	2020	E-Stat
	Elder Density	-	People /Km ²	$\frac{\text{Elder}}{\text{Habitable area}}$	2020	E-Stat
	Road Density	m/Km2	m/Km ²	Provided by source	2010	MLIT
Topography	Average Elevation	m	m	Provided by source	2011	MLIT
	Highest Elevation	m	m	Provided by source	2011	MLIT
	Lowest Elevation	m	m	Provided by source	2011	MLIT
Land / Area Structure	Total Area	Ha	Km ²	Ha/100	2023	E-Stat
	Habitable Area	Ha	Km ²	Ha/100	2020	E-Stat
	Residence Area	Ha	Km ²	Ha/100	2022	E-Stat

Source: D. Hasegawa & Suzuki, 2011; Higashimoto et al., 2013; Hasegawa D. & Suzuki, 2019; Coutinho et al., 2020; Rich et al., 2023; H. Wang et al., 2023; E.-J. Kim et al., 2025

The spatial image data were generated from ArcGIS using municipality-level exports based on the navigation map setting. This map type was selected because its clear contrast improves readability and supports stable visual feature extraction in CNN-based models (Sibrel et al., 2020; Wu et al., 2024). To ensure consistency, all images were exported using identical settings, including clip to feature index, 1:15,000 scale, 24-bit alpha, PNG format, uniform resolution, and 1850 mm × 1850 mm paper size. All images were first exported at 3200 × 3200 px and then resized to 640 × 640 px before being input into the model. All images were exported in January 2026. This approach follows previous research showing that higher-resolution export followed by resizing can be preferable to preparing images directly at the exact input size (T. He et al., 2019; Touvron et al., 2022). The example of the navigation map used as spatial image input is shown in Figure 1.



Figure 1. Example of Navigation Map

Model evaluation was conducted using stratified five-fold cross-validation. The 106 labeled municipalities were divided at the municipality level, with one fold used for validation and the remaining four folds used for training in each iteration. The fold assignment was stratified by class so that CB and DRB municipalities were distributed as evenly as possible across the

five folds. In addition, all images belonging to the same municipality were assigned to the same fold, ensuring that each municipality was evaluated once as unseen data and preventing leakage between training and validation subsets. The overall procedure for fold splitting and training-validation rotation is illustrated in Figure 2.

Dataset CB: 41; DRB: 65			
Validation Data CB: 9; DRB: 13	Training Data CB: 32; DRB: 52		
Training Data CB: 9; DRB: 13	Validation Data CB: 8; DRB: 13	Training Data CB: 24; DRB: 39	
Training Data CB: 17; DRB: 26		Validation Data CB: 8; DRB: 13	Training Data CB: 16; DRB: 26
Training Data CB: 25; DRB: 39		Validation Data CB: 8; DRB: 13	Training Data CB: 8; DRB: 13
Training Data CB: 33; DRB: 52			Validation Data CB: 8; DRB: 13

Figure 2. K-fold cross-validation procedure

Because the number of labeled municipalities was limited, image augmentation was applied to the training images to improve model robustness and reduce the risk of overfitting. The augmentation techniques included rotation at 90°, 180°, and 270°, together with horizontal and vertical flipping. Rotation was included because conventional CNNs are not inherently rotation-invariant, and augmentation has been widely used to improve robustness to orientation variation in image classification and aerial-image analysis (Perez & Wang, 2017; Ding et al., 2018; Quiroga et al., 2020; Han et al., 2021). Horizontal and vertical flipping were also applied as practical image-level augmentations to increase visual diversity and stabilize training (Perez & Wang, 2017). However, because geographic orientation may carry meaning in map-based data, these flipping operations were treated as robustness-oriented transformations at the image-pattern level rather than as geography-preserving transformations in a strict cartographic sense. Augmentation was applied only to the training images within each cross-validation fold. The validation images were kept unchanged during model training. The tabular data remained unchanged, and each augmented image retained the same municipality-level tabular vector and class label as the original municipality.

Before model training, the tabular data underwent consistency checking and cleaning to ensure comparability across municipalities. Missing values were handled using median imputation, with the median calculated separately from the training subset in each fold and then applied to the corresponding validation subset to avoid data leakage. Unrealistic extreme values were checked against official sources.

After cleaning, the tabular variables were standardized using the z-score transformation shown in Equation (1).

$$z = \frac{x - \mu}{\sigma}$$

In this equation, z denotes the standardized feature value, x denotes the original value of the feature, μ denotes the mean of the feature calculated from the training subset, and σ denotes the standard deviation of the feature calculated from the training subset. The resulting mean and standard deviation from each training fold were then applied to the corresponding validation fold. For the image branch, all spatial images were normalized using the standard

ImageNet mean and standard deviation values of (0.485, 0.456, 0.406) and (0.229, 0.224, 0.225), respectively.

This research was designed as a municipality-level binary classification and interpretation analysis using a multimodal late-fusion model. This paper adopts a multimodal late-fusion model that integrates tabular indicators and GIS-based spatial images within a municipality-level classification framework. The image branch used a pretrained ResNet18 backbone initialized with ImageNet weights (K. He et al., 2015). Its original fully connected layer was removed and replaced with an identity mapping so that the backbone produced a 512-dimensional image feature vector. In this research, the ResNet18 backbone was fine-tuned rather than frozen because all backbone parameters remained trainable during model optimization.

The tabular branch was implemented as a multilayer perceptron (MLP) with two hidden layers of 64 and 32 units, respectively. A ReLU activation function and a dropout rate of 0.2 followed each hidden layer. Since the tabular input consisted of 23 municipality-level features, the MLP transformed the input into a 32-dimensional tabular representation.

After feature extraction, the 512-dimensional image representation and the 32-dimensional tabular representation were concatenated, resulting in a 544-dimensional fused feature vector. This fused representation was then passed to a classifier head consisting of a fully connected layer from 544 to 128 units, followed by ReLU activation, dropout of 0.2, and a final fully connected output layer from 128 to 2 units for binary CB or DRB classification. The overall architecture is illustrated in Figure 3.

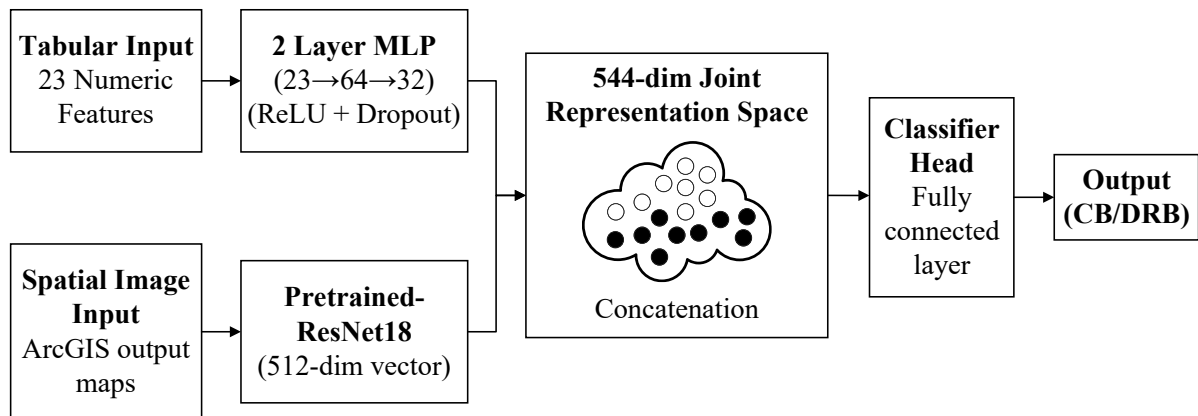


Figure 3. Multimodal late fusion architecture

A multimodal late-fusion architecture was adopted because it provides a clear and practical framework for integrating heterogeneous data types. Multimodal learning research has shown that fusion is a central challenge when different modalities capture complementary but structurally different information. This issue is especially relevant for image-tabular learning, where structured variables and image features differ fundamentally in representation and therefore require suitable interaction design (Baltrušaitis et al., 2018; Du et al., 2024). In this context, late fusion remains a widely used strategy because each modality can be processed in a separate branch before being combined, thereby preserving modality-specific feature learning while keeping the integration structure relatively simple (Joze et al., 2020).

The model was trained as a binary classifier for CB and DRB municipalities. Training used cross-entropy loss, as defined in Equation (2) (Goodfellow et al., 2016).

$$L = -\frac{1}{N} \sum_{i=1}^N \sum_{c=1}^C y_{ic} \log(p_{ic})$$

Where N is the number of samples, C is the number of classes, y_{ic} is the true label for sample i for class c , and p_{ic} represents the predicted probability for that class.

Early stopping was applied to reduce overfitting and to retain the best-performing checkpoint within each fold. The main training settings are summarized in Table 2.

Table 2. Training configuration of the multimodal late-fusion model

Parameter	Value
Optimizer	AdamW
Learning rate	5×10^{-5}
Weight decay	1×10^{-4}
Batch size	8
Maximum epochs	100
Loss function	Cross-entropy loss
Early stopping	Enabled
Early stopping patience	10
Minimum improvement (Δ)	1×10^{-4}

Source: Data Training, 2026

Model performance was evaluated using accuracy, precision, recall, and Macro F1-score, with Macro F1-score used as the main metric because it gives equal importance to both classes (Sokolova & Lapalme, 2009; Powers, 2011).

Accuracy, precision, recall, and F1-score were calculated using Equations (3)–(6), and Macro F1-score was obtained by averaging the class-wise F1-scores in Equation (7).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F1 - score = 2 \times \frac{Recall \times Precision}{Recall + Precision}$$

$$Macro F1 = \frac{1}{C} \sum_{i=1}^C F1_i$$

In these equations, TP, TN, FP, and FN denote true positives, true negatives, false positives, and false negatives, respectively, while C denotes the number of classes and $F1_i$ denotes the F1-score of class i . These metrics provide complementary views of classification performance, and Macro F1-score was used as the main metric because it gives equal weight to each class and is suitable for balanced class-wise evaluation (Van Rijsbergen, 1979; Sokolova & Lapalme, 2009; Powers, 2011).

In addition to augmentation during training, Test-Time Augmentation (TTA) was applied during model evaluation. TTA was used to improve prediction robustness by generating multiple transformed versions of the same validation image and averaging the resulting predicted probabilities. The transformations used in TTA were the same as those used during image augmentation in training, namely rotation and horizontal and vertical flipping. TTA was applied only to the image modality, while the tabular input remained unchanged.

SHAP was used to quantify the contribution of each municipality-level variable to the model output in tabular branch (Lundberg & Lee, 2017). In this research, Kernel SHAP was applied using a fold-specific background dataset consisting of up to 10 tabular validation samples from the corresponding fold. SHAP values were then computed for each validation municipality individually. SHAP was examined at three levels: a global importance analysis across all validation samples, class-specific summaries for correctly classified CB and DRB municipalities, and directional summaries based on the signed SHAP values. The SHAP formulation is shown in Equation (8).

$$\phi_i = \sum_{S \subseteq F \setminus \{i\}} \frac{|S|! (|F| - |S| - 1)!}{|F|!} [f(S \cup \{i\}) - f(S)]$$

In this equation, F represents the entire set of input features, S represents the subset of features that exclude feature i , and $f(\cdot)$ represents the model output function. This formulation expresses the average marginal contribution of feature i across all possible feature subsets. The per-sample SHAP values were first obtained separately in each fold and then concatenated across all validation folds. Global and class-specific summaries were subsequently derived by averaging the absolute SHAP values, while the directional summaries were obtained from the aggregated signed SHAP values.

For the image branch, Grad-CAM was used to visualize the internal spatial regions receiving the strongest model attention (Selvaraju et al., 2020). Grad-CAM was applied using the second convolutional layer of the last residual block in ResNet18 layer4 as the target layer. Heatmaps were generated for validation samples in each fold using the corresponding best-performing checkpoint. In this paper, five representative CB heatmaps and five representative DRB heatmaps were selected across folds from correctly classified validation cases with prediction probabilities above 0.7. The Grad-CAM weighting formulation is shown in Equation (9).

$$\alpha_k^c = \frac{1}{Z} \sum_i \sum_j \frac{\partial y^c}{\partial A_{ij}^k}$$

Here, y^c denotes the output score for class c , and Z denotes the total number of spatial locations in the feature map.

The resulting class activation map is defined in Equation (10).

$$L_{\text{Grad-CAM}}^c = \text{ReLU} \left(\sum_k \alpha_k^c A^k \right)$$

Unlike the SHAP results, the Grad-CAM heatmaps were not aggregated numerically across folds. Instead, they were used as qualitative interpretive outputs, and representative examples were selected to illustrate the spatial attention patterns of the late-fusion model. The SHAP and Grad-CAM analyses were therefore used as interpretive tools rather than causal evidence. Their role in this paper was to identify which tabular indicators and spatial image regions were most strongly associated with the late-fusion model's CB and DRB predictions.

Results and Discussion

Model performance was first examined at the fold level to show the validation distribution and score variation across the five-fold cross-validation procedure. Because the reviewer requested clearer reporting of fold composition and stability, both the class distribution and the fold-wise results are presented explicitly. The validation distribution and fold-wise performance with and without TTA are presented in Table 3.

Table 3. Performance of the multimodal late-fusion model

Fold	CB in validation	DRB in validation	Accuracy (No TTA)	Macro F1 (No TTA)	Accuracy (TTA)	Macro F1 (TTA)
1	9	13	0.6364	0.4824	0.6364	0.4824
2	8	13	0.5714	0.3636	0.5714	0.3636
3	8	13	0.6667	0.637	0.6667	0.637
4	8	13	0.6667	0.6541	0.6667	0.6541
5	8	13	0.8571	0.8518	0.8095	0.7981

Source: Data Analysis, 2026

As shown in Table 3, the validation distribution remained relatively balanced across folds, with 8–9 CB municipalities and 13 DRB municipalities in each validation subset. This is consistent with the stratified five-fold cross-validation design. The same table also shows that the fold-wise results with and without TTA were identical in Folds 1–4, while a difference appeared only in Fold 5. In the non-TTA setting, the highest Macro F1-score was observed in Fold 5 (0.8518) and the lowest in Fold 2 (0.3636). In the TTA setting, the highest Macro F1-score was also observed in Fold 5, but it decreased slightly to 0.7981. These results indicate that fold-to-fold variation remained substantial, which is reasonable given the limited municipality sample and the heterogeneity of regional conditions represented in the dataset.

To evaluate the effect of TTA, the mean and standard deviation of the main metrics were compared between the non-TTA and TTA settings. Since this paper uses Macro F1-score as the primary metric, the comparison was intended to determine whether TTA improved class-balanced performance rather than accuracy alone. The mean performance comparison of the multimodal late-fusion model with and without TTA is presented in Table 4.

Table 4. Mean performance comparison of the multimodal late-fusion model with and without TTA

Metric	No TTA Mean $\pm$ SD	TTA Mean $\pm$ SD
Accuracy	0.6797 $\pm$ 0.1066	0.6701 $\pm$ 0.0871
Precision for CB	0.5810 $\pm$ 0.3717	0.5754 $\pm$ 0.3682
Recall for CB	0.4222 $\pm$ 0.3631	0.3972 $\pm$ 0.3266
F1 for CB	0.4290 $\pm$ 0.3273	0.4143 $\pm$ 0.3061
Precision for DRB	0.7200 $\pm$ 0.1267	0.7059 $\pm$ 0.1006
Recall for DRB	0.8462 $\pm$ 0.1216	0.8462 $\pm$ 0.1216
F1 for DRB	0.7665 $\pm$ 0.0657	0.7598 $\pm$ 0.0512
Macro F1	0.5978 $\pm$ 0.1853	0.5870 $\pm$ 0.1676

Source: Data Analysis, 2026

As shown in Table 4, the multimodal late-fusion model achieved a mean Macro F1-score of 0.5978 ± 0.1853 without TTA and 0.5870 ± 0.1676 with TTA. Since TTA was evaluated as an inference-time robustness strategy, these results indicate a trade-off between mean predictive score and fold-wise stability. The non-TTA setting produced slightly higher mean Macro F1-score and accuracy, whereas the TTA setting produced slightly smaller standard deviations for both metrics. A similar pattern appears in the mean accuracy, which was 0.6797 ± 0.1066 without TTA and 0.6701 ± 0.0871 with TTA. At the class level, the effect of TTA was also limited. The mean CB F1-score changed from 0.4290 ± 0.3273 without TTA to 0.4143 ± 0.3061 with TTA, while the mean DRB F1-score changed from 0.7665 ± 0.0657 to 0.7598 ± 0.0512 . These results suggest that TTA provided slightly more stable fold-wise results, but the non-TTA setting retained higher average predictive scores under the evaluation setting used in this paper.

To make the prediction outcome more explicit, the confusion matrix was also summarized for both evaluation settings. This comparison helps show how TTA changed the number of correct and incorrect classifications in each class rather than only the aggregate metrics. The confusion matrix summary of the multimodal late-fusion model with and without TTA is presented in Table 5.

Table 5. Confusion matrix summary of the multimodal late-fusion model with and without TTA

Setting	TN (CB-CB)	FP (CB-DRB)	FN (DRB-CB)	TP (DRB-DRB)
No TTA	17	24	10	55
TTA	16	25	10	55

Source: Data Analysis, 2026

As shown in Table 5, the non-TTA setting produced 17 true negatives, 24 false positives, 10 false negatives, and 55 true positives, whereas the TTA setting produced 16 true negatives, 25 false positives, 10 false negatives, and 55 true positives. This means that TTA changed only a small portion of the prediction outcomes. The number of correctly classified DRB municipalities remained unchanged, while one correctly classified CB municipality under the non-TTA setting was classified as DRB under the TTA setting. This limited difference suggests that the model predictions were relatively stable under test-time transformations, although the averaged TTA outputs did not increase the number of correct predictions. Taken together, the results in Tables 3–5 show that TTA produced slightly lower mean predictive scores but slightly smaller fold-wise variability. Therefore, the subsequent interpretation results reported in this paper are based on the non-TTA setting, since it produced the higher mean Macro F1-score and accuracy.

Although the fold-wise results varied, the late-fusion model provided a sufficient basis for branch-specific interpretation using SHAP and Grad-CAM. The absolute SHAP values reported in this paper are numerically small because they reflect contribution magnitudes on the model output scale after standardized tabular preprocessing. For this reason, the SHAP analysis focuses on the relative ranking and directional tendency of the variables rather than on their absolute magnitude alone. SHAP values were first computed for each validation sample within each fold and were then concatenated across the five validation folds. Based on these combined results, global and class-specific summaries were subsequently derived. The SHAP results are presented in Tables 6, 7, and 8. Table 6 summarizes the global SHAP importance across all validation samples, while Tables 7 and 8 present the class-specific SHAP importance for correctly classified CB and DRB municipalities, respectively.

Table 6. Top 10 global SHAP feature importance

Rank	Feature	Mean absolute SHAP value
1	Single Elder Household Density	0.000426
2	Lowest Elevation	0.000331
3	Elder Rate	0.000262
4	Highest Elevation	0.000261
5	Total Population	0.000241
6	Elder Density	0.000212
7	Residence Area	0.000203
8	Average Elevation	0.000167
9	Single Elder Household	0.000161
10	Route Length	0.00016

Source: Data Analysis, 2026

Table 7. Top 10 SHAP feature importance for correctly classified CB municipalities

Rank	Feature	Mean absolute SHAP value
1	Single Elder Household Density	0.000736
2	Total Population	0.000332
3	Residence Area	0.000323
4	Elder Density	0.00032
5	Elder Rate	0.000308
6	Highest Elevation	0.000295
7	Car Share	0.000263
8	Density	0.000245
9	Total Elder	0.000164
10	Single Elder Household	0.000161

Source: Data Analysis, 2026

Table 8. Top 10 SHAP feature importance for correctly classified DRB municipalities

Rank	Feature	Mean absolute SHAP value
1	Single Elder Household Density	0.000382
2	Lowest Elevation	0.000376
3	Elder Density	0.000221
4	Elder Rate	0.000209
5	Highest Elevation	0.000201
6	Residence Area	0.000182
7	Total Household	0.000153
8	Habitable Area	0.000151
9	Car Ownership	0.000141
10	Average Elevation	0.000135

Source: Data Analysis, 2026

At the global level, the most influential feature was single elder household density, lowest elevation, elder rate, highest elevation, and total population, as shown in Table 6. Other important variables included elder density, residence area, average elevation, single elder household, and route length. These results show that the late-fusion model relied on a combination of household structure, demographic aging, population scale, transport-related conditions, and topographic characteristics. The global ranking therefore suggests that the model did not depend on one narrow type of variable. It drew information from several municipality-level dimensions at the same time.

The class-specific SHAP results show a more differentiated pattern. For correctly classified CB municipalities, the most important variables were single elder household density, total population, residence area, elder density, and elder rate, as reported in Table 7. This ranking places stronger emphasis on settlement concentration, population size, and residential structure. In contrast, correctly classified DRB municipalities were characterized more strongly by single elder household density, lowest elevation, elder density, elder rate, and highest elevation, as shown in Table 8. In this class, topographic variables appeared more prominently near the top of the ranking.

An important pattern is that elderly-related variables remained influential across all three SHAP summaries, including the global results as well as the correctly classified CB and DRB

subsets. Variables such as single elder household density, elder density, elder rate, and single elder household repeatedly appeared among the most important features. This suggests that the model gave substantial weight to elderly-related municipality characteristics when distinguishing between CB and DRB. Such a pattern is consistent with the broader Japanese context, where local transport planning is closely linked to population decline and population aging, especially in municipalities with a high proportion of older residents.

The SHAP direction results provide an additional layer of interpretation. The SHAP direction results are presented in Tables 9, 10, and 11. Table 9 summarizes the global SHAP direction across all validation samples, while Tables 10 and 11 present the class-specific SHAP direction for correctly classified CB and DRB municipalities, respectively.

Table 9. Top 10 global SHAP direction summary

Rank	Feature	Mean signed SHAP value	Direction
1	Single Elder Household Density	-0.000108	Toward CB
2	Residence Area	-0.000081	Toward CB
3	Single Elder Household	-0.000056	Toward CB
4	Elder Density	0.000052	Toward DRB
5	Elder Rate	-0.000046	Toward CB
6	Car Ownership	-0.000045	Toward CB
7	Total Household	-0.000041	Toward CB
8	Lowest Elevation	-0.000041	Toward CB
9	Household Density	0.000035	Toward DRB
10	Habitable Area	-0.000035	Toward CB

Source: Data Analysis, 2026

Table 10. Class-specific SHAP direction summary for correctly classified CB

Rank	Feature	Mean signed SHAP value	Direction
1	Single Elder Household Density	0.000442	Toward DRB
2	Elder Density	-0.000359	Toward CB
3	Highest Elevation	0.000171	Toward DRB
4	Elder Rate	-0.000132	Toward CB
5	Habitable Area	-0.000128	Toward CB
6	Route Length	-0.000105	Toward CB
7	Total Area	-0.000101	Toward CB
8	Total Population	-0.000074	Toward CB
9	Total Elder	0.000073	Toward DRB
10	Car Ownership	-0.000063	Toward CB

Source: Data Analysis, 2026

Table 11. Class-specific SHAP direction summary for correctly classified DRB

Rank	Feature	Mean signed SHAP value	Direction
1	Single Elder Household Density	-0.000299	Toward CB
2	Elder Density	0.000195	Toward DRB
3	Car Share	0.000102	Toward DRB
4	Lowest Elevation	-0.000083	Toward CB
5	Single Elder Household Rate	-0.000053	Toward CB

6	Total Vehicle	0.000049	Toward DRB
7	Total Elder	-0.000048	Toward CB
8	Route Length	-0.000038	Toward CB
9	Single Elder Household	-0.000037	Toward CB
10	Residence Area	0.000037	Toward DRB

Source: Data Analysis, 2026

The global direction summary, shown in Table 9, indicates that single elder household density, residence area, single elder household, elder rate, car ownership, total household, lowest elevation, and habitable area tended to contribute toward CB, whereas only two of them, which are elder density and household density, tended to contribute toward DRB. This directional pattern suggests that the model more often associated CB with household-related residential structure, land-area conditions, and several municipality-scale demographic features, while the direction toward DRB was more limited and was mainly linked to density-related indicators.

The class-specific direction summaries in Table 10 and Table 11 also reveal that the role of a variable was not always identical across CB and DRB subsets. For correctly classified CB municipalities, elder density, elder rate, habitable area, route length, total area, total population, and car ownership showed contributions toward CB, while single elder household density, highest elevation, and total elder showed direction toward DRB. For correctly classified DRB municipalities, elder density, car share, total vehicle, and residence area showed direction toward DRB, whereas single elder household density, lowest elevation, single elder household rate, total elder, route length, and single elder household showed direction toward CB. This pattern indicates that feature direction was context-dependent across the two subsets. A variable could remain influential in both classes, while its signed contribution shifted depending on whether the municipality was correctly classified as CB or DRB.

The SHAP results show that the late-fusion model relied on a combination of Demographic or Household, Density or Built Intensity, Mobility, Topography, and Land or Area Structure. Among these, single elder household density was the most consistent feature because it ranked first in the global results and also remained the top feature in both the correctly classified CB and correctly classified DRB subsets. In addition, several elderly-related variables, including elder density, elder rate, and single elder household, repeatedly appeared among the influential features. This pattern suggests that elderly-related municipality characteristics formed an important part of the numerical signals used by the model. At the same time, the direction summaries show that high importance did not always imply a fixed directional role. Some variables remained influential across multiple summaries, but their signed contributions differed between the global, correctly classified CB, and correctly classified DRB results. This indicates that the model did not interpret the two classes through the same numerical profile, even when some of the same variables remained prominent in the ranking. To support qualitative interpretation of the image branch, representative Grad-CAM heatmaps were selected from correctly classified validation cases. The selected examples included five CB municipalities and five DRB municipalities, all with prediction probabilities above 0.7, so that the interpretation focused on relatively confident model outputs. The metadata of the selected Grad-CAM examples are presented in Table 12.

Table 12. Metadata of representative Grad-CAM examples

Figure panel	Municipality	True class	Predicted class	Prediction probability
Figure 4(a)	Onojo City	CB	CB	0.783931851

Figure 4(b)	Mizumaki Town	CB	CB	0.921696246
Figure 4(c)	Oto Town	CB	CB	0.946386039
Figure 4(d)	Keisen Town	CB	CB	0.922601938
Figure 4(e)	Dazaifu City	CB	CB	0.920987189
Figure 5(a)	Karatsu City	DRB	DRB	0.999134719
Figure 5(b)	Chikuzen Town	DRB	DRB	0.972562253
Figure 5(c)	Kikai Town	DRB	DRB	0.954691529
Figure 5(d)	Shinkamigoto Town	DRB	DRB	0.866027951
Figure 5(e)	Taragi Town	DRB	DRB	0.830055714

Source: Data Analysis, 2026

The Grad-CAM results are presented in Figures 4 and 5, which show representative heatmaps for correctly classified CB and DRB municipalities, respectively.

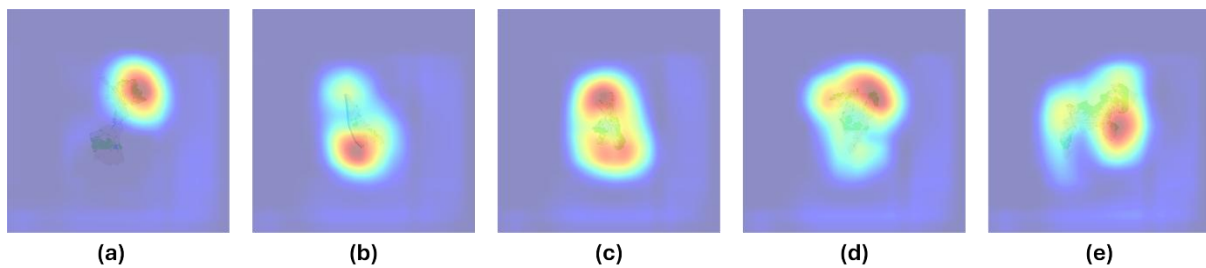


Figure 4. Representative heatmap of the CB class

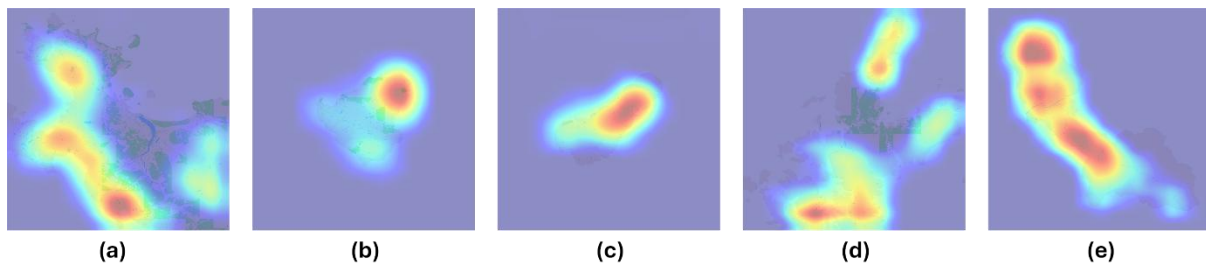


Figure 5. Representative heatmap of the DRB class

The heatmaps show that the image branch consistently attended to internal spatial regions of the municipality rather than to irrelevant background areas. In both CB and DRB examples, the highlighted regions generally overlapped with visually meaningful spatial structures, including built-up areas, settlement clusters, and road-related patterns visible in the GIS-based images. This indicates that the image branch was not responding randomly. It captured spatial configurations that were relevant to the classification process.

At the same time, the Grad-CAM results were more difficult to interpret directly than the SHAP results. SHAP provides explicit feature-level contributions for the tabular branch, whereas Grad-CAM only shows where the model attention is concentrated in the image branch. For that reason, the Grad-CAM results in this paper are treated as qualitative supporting evidence rather than as direct feature-level explanation. This limitation is important because the apparent spread of activation can also be influenced by differences in municipality size, shape, and internal geometry. For this reason, the Grad-CAM analysis was interpreted cautiously.

Conclusion

This paper examined the interpretability of a multimodal late-fusion model for municipality-level classification of CB and DRB in Kyushu, Japan. The model integrated tabular regional indicators and GIS-based spatial images, and its behavior was interpreted through SHAP for the tabular branch and Grad-CAM for the image branch. Using Macro F1-score as the main metric, the model achieved moderate classification performance and provided a workable basis for interpretation.

The SHAP results showed that the model relied on a combination of Demographic or Household, Density or Built Intensity, Mobility, Topography, and Land or Area Structure. Single Elder Household Density appeared as the most consistent feature across the global ranking as well as the correctly classified CB and DRB subsets. Several other elderly-related variables, including elder density, elder rate, and single elder household, also appeared repeatedly among the influential features, indicating that elderly-related municipality characteristics were important in the model's numerical interpretation. At the same time, the direction summaries showed that influential variables did not always have the same signed role across the global, CB, and DRB results. This suggests that the model relied on overlapping but not identical numerical profiles for the two classes. The Grad-CAM results supported this interpretation by showing that the image branch attended to meaningful internal spatial regions in the GIS-based municipality images. At the same time, the Grad-CAM results were less direct than the SHAP results and were therefore interpreted as qualitative supporting evidence rather than explicit feature-level explanation.

Taken together, these findings show that a multimodal late-fusion model can provide interpretable evidence for municipality-level CB and DRB classification. The tabular branch offered clearer explanation through feature importance and direction, while the image branch contributed complementary visual evidence through spatial attention patterns. This combination makes the model useful for regional transport analysis where both structured municipality indicators and map-based spatial information matter.

Future research should expand the municipality sample, test the same interpretation framework in other regions, and examine whether other multimodal architectures or other modality produce more stable or more easily interpretable results. It would be useful to compare the model explanations with expert judgment from local transport planners so that the analytical results can be assessed against real planning practice. Future research should also compare the interpreted late-fusion model with tabular-only and image-only baselines to evaluate the relative contribution of each modality.

Acknowledgment

The author also appreciates the reviewers for their valuable comments and suggestions, which contributed to improving the quality of the manuscript.

References

- Atrey, P. K., Hossain, M. A., El Saddik, A., & Kankanhalli, M. S. (2010). Multimodal fusion for multimedia analysis: A survey. *Multimedia Systems*, 16(6), 345–379. <https://doi.org/10.1007/s00530-010-0182-0>
- Baltrušaitis, T., Ahuja, C., & Morency, L.-P. (2018). Multimodal machine learning: A survey and taxonomy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(2), 423–443. <https://doi.org/10.1109/tpami.2018.2798607>
- Blomqvist, E. (2025). *Fracture Risk Prediction Using Multimodal Neural Networks*.
- Coutinho, F. M., Van Oort, N., Christoforou, Z., Alonso-González, M. J., Cats, O., & Hoogendoorn, S. (2020). Impacts of replacing a fixed public transport line by a demand responsive transport system: Case study of a rural area in Amsterdam.

- Dahmen, V., Weikl, S., & Bogenberger, K. (2024). Interpretable Machine Learning for Mode Choice Modeling on Tracking-Based Revealed Preference Data. *Transportation Research Record: Journal of the Transportation Research Board*, 2678(11), 2075–2091. <https://doi.org/10.1177/03611981241246973>
- Deressu, T. F., Bojer, A. K., Debelee, T. G., Negera, W. G., Nadarajah, S., & Gebissa, K. W. (2025). Enhancing land use and land cover classification with deep learning-based satellite imagery segmentation. *International Journal of Applied Earth Observation and Geoinformation*, 144, 104839. <https://doi.org/10.1016/j.jag.2025.104839>
- Ding, J., Xue, N., Long, Y., Xia, G.-S., & Lu, Q. (2018). Learning RoI transformer for detecting oriented objects in aerial images. *arXiv Preprint arXiv:1812.00155*. <https://doi.org/10.48550/arXiv.1812.00155>
- Du, S., Zheng, S., Wang, Y., Bai, W., O'Regan, D. P., & Qin, C. (2024). TIP: Tabular-image pre-training for multimodal classification with incomplete data. *European Conference on Computer Vision*, 478–496. <https://doi.org/10.48550/arXiv.2407.07582>
- Goodfellow, I., Bengio, Y., Courville, A., & Bengio, Y. (2016). *Deep learning* (Vol. 1). MIT press Cambridge.
- Han, J., Ding, J., Xue, N., & Xia, G.-S. (2021). Redet: A rotation-equivariant detector for aerial object detection. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2786–2795. <https://doi.org/10.48550/arXiv.2103.07733>
- Hasegawa, D., & Suzuki, T. (2011). Urban size and density condition for the introduction of demand responsive transport. *Journal of the City Planning Institute of Japan*, 46(3), 817–822. <https://doi.org/10.11361/journalcpj.46.817>
- Hasegawa D., & Suzuki T. (2019). Evaluation of Effectiveness by Combining of Bus and Demand Responsive Transit Service Focusing Bus Route Network Configuration. *Theory and Applications of GIS*, 27(1), 1–11. <https://doi.org/10.5638/thagis.27.1>
- Hasegawa, M., Nakamura, T., Aou, Y., Suzumura, M., & Morikawa, T. (2021). Comparison Analysis Of User Service Need For Implementing Share Taxi. *Journal of Japan Society of Civil Engineers, Ser. D3 (Infrastructure Planning and Management)*, 76(5), I_1417-I_1427. https://doi.org/10.2208/jscejipm.76.5_I_1417
- He, K., Zhang, X., Ren, S., & Sun, J. (2015). *Deep Residual Learning for Image Recognition* (arXiv:1512.03385). arXiv. <https://doi.org/10.48550/arXiv.1512.03385>
- He, T., Zhang, Z., Zhang, H., Zhang, Z., Xie, J., & Li, M. (2019). Bag of Tricks for Image Classification with Convolutional Neural Networks. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 558–567. <https://doi.org/10.1109/CVPR.2019.00065>
- Higashimoto, Y., Takada, H., & Kishi, K. (2013). *Evaluation of bus route efficiency by network DEA including social priority*.
- Ise, N., & Hino, Y. (2023). A Study on Community Bus Origin-Destination Demand focused on Characteristics of Bus Stop Catchment Area: A Case Study in Kishiwada City, Osaka Prefecture. *JSTE Journal of Traffic Engineering*, 9(2), A_335-A_341. https://doi.org/10.14954/jste.9.2_A_335

- Joze, H. R. V., Shaban, A., Iuzzolino, M. L., & Koishida, K. (2020). MMTM: Multimodal transfer module for CNN fusion. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 13289–13299. <https://doi.org/10.48550/arXiv.1911.08670>
- Kato, H. (2025). Acceptance and usage of demand-responsive transport among people in poor health: Evidence from Senboku New Town. *Transportation Research Interdisciplinary Perspectives*, 34, 101632. <https://doi.org/https://doi.org/10.1016/j.trip.2025.101632>
- Kim, E.-J., Oh, D., & Kim, H. (2025). *How demand-responsive transit enhances public transit equity: A spatial causal inference approach*. SSRN. <https://doi.org/10.2139/ssrn.5119770>
- Liyanage, S., Dia, H., Duncan, G., & Abduljabbar, R. (2024). Evaluation of the Impacts of On-Demand Bus Services Using Traffic Simulation. *Sustainability*, 16(19), 8477. <https://doi.org/10.3390/su16198477>
- Lundberg, S. M., & Lee, S.-I. (2017). A Unified Approach to Interpreting Model Predictions. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, & R. Garnett (Eds.), *Advances in Neural Information Processing Systems* (Vol. 30). Curran Associates, Inc. https://proceedings.neurips.cc/paper_files/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf
- Nakanishi, T., & Matsumoto, Y. (2024). Interregional comparison of various effects by DRT focusing on individual attributes and regional characteristics. *Proceedings of the City Planning Institute of Japan, Chubu Branch*, 35, 27–32. https://doi.org/10.11361/cpijchubu.35.0_27
- Nakanishi, T., & Matsumoto, Y. (2025). Understanding Transfer Intentions and Areas Suitable for Introducing Transfer-based DRT. *Journal of the Eastern Asia Society for Transportation Studies*, 16, PP3937. <https://doi.org/10.11175/easts.16.PP3937>
- Perez, L., & Wang, J. (2017). The effectiveness of data augmentation in image classification using deep learning. *arXiv Preprint arXiv:1712.04621*. <https://doi.org/10.48550/arXiv.1712.04621>
- Powers, D. (2011). Evaluation: From Precision, Recall and F-Measure to ROC, Informedness, Markedness & Correlation. *Journal of Machine Learning Technologies*, 2(1), 37–63.
- Quiroga, F., Ronchetti, F., Lanzarini, L., & Bariviera, A. F. (2020). Revisiting Data Augmentation for Rotational Invariance in Convolutional Neural Networks. In J. C. Ferrer-Comalat, S. Linares-Mustarós, J. M. Merigó, & J. Kacprzyk (Eds.), *Modelling and Simulation in Management Sciences* (pp. 127–141). Springer International Publishing. https://doi.org/10.1007/978-3-030-15413-4_10
- Rich, J., Seshadri, R., Jomeh, A. J., & Clausen, S. R. (2023). Fixed routing or demand-responsive? Agent-based modelling of autonomous first and last mile services in light-rail systems. *Transportation Research Part A: Policy and Practice*, 173, 103676. <https://doi.org/10.1016/j.tra.2023.103676>
- Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2020). Grad-CAM: Visual explanations from deep networks via gradient-based localization. *International Journal of Computer Vision*, 128(2), 336–359. <https://doi.org/10.1007/s11263-019-01228-7>

- Sibrel, S. C., Rathore, R., Lessard, L., & Schloss, K. B. (2020). The relation between color and spatial structure for interpreting colormap data visualizations. *Journal of Vision*, 20(12), 7–7.
- Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information Processing & Management*, 45(4), 427–437. <https://doi.org/10.1016/j.ipm.2009.03.002>
- Takeuchi, R., Iwamoto, T., Takaku, M., Nakamura, T., & Ozaki, M. (2019). *Research on the introduction situation of various local public transportation (Report No.153)*. Policy Research Institute for Land, Infrastructure, Transport and Tourism. https://www.mlit.go.jp/pri/english/houkoku/gaiyou/english_kkk153.html?
- Takeuchi, R., & Yoshida, I. (2022). A study on the difference of the introduction situation and planning process on community bus and demand responsive transport in local governments. *Journal of Japan Society of Civil Engineers, Ser. D3 (Infrastructure Planning and Management)*, 78(6), II_613-II_622. https://doi.org/10.2208/jscejipm.78.6_II_613
- Takeuchi, R., Yoshida, I., & Inoi, H. (2023). Perspective And Business Continuity Influence Factors For Transportation Passenger Service With Private Car By Actual Situation Of Operation Body (Paratransit). *Japanese Journal of JSCE*, 79(20), 23–20026. <https://doi.org/10.2208/jscej.23-20026>
- Takeuchi, R., Yoshida, I., & Ozaki, M. (2023). Research On The Factors Influencing The Effectiveness Of The Introduction Of Demand Transport And Community Buses By Local Authorities. *Journal of Japan Society of Civil Engineers, Ser. D3 (Infrastructure Planning and Management)*, 78(5), I_627-I_637. https://doi.org/10.2208/jscejipm.78.5_I_627
- Tang, L., Tang, C., Fu, Q., & Ma, C. (2024). Predicting Travel Mode Choice With a Robust Neural Network and Shapley Additive Explanations Analysis. *IET Intelligent Transport Systems*, 18(7), 1339–1354. <https://doi.org/10.1049/itr2.12514>
- Touvron, H., Vedaldi, A., Douze, M., & Jégou, H. (2022). *Fixing the train-test resolution discrepancy* (arXiv:1906.06423). arXiv. <https://doi.org/10.48550/arXiv.1906.06423>
- Van Rijsbergen, C. (1979). Information retrieval: Theory and practice. *Proceedings of the Joint IBM/University of Newcastle upon Tyne Seminar on Data Base Systems*, 79, 1–14.
- Vansteenkoven, P., Melis, L., Aktaş, D., Montenegro, B. D. G., Sartori Vieira, F., & Sörensen, K. (2022). A survey on demand-responsive public bus systems. *Transportation Research Part C: Emerging Technologies*, 137, 103573. <https://doi.org/10.1016/j.trc.2022.103573>
- Velmurugan, P. (2023). *Species Distribution Modelling: A Multimodal Learning Approach*.
- Wang, H., Li, J., Wang, P., Teng, J., & Loo, B. P. Y. (2023). Adaptability analysis methods of demand responsive transit: A review and future directions. *Transport Reviews*, 43(4), 676–697. <https://doi.org/10.1080/01441647.2023.2165574>
- Wu, M., Sun, Y., & Xia, X. (2024). Computational assessment of the aesthetic quality of map colors. *Transactions in GIS*, 28(6), 1556–1576. <https://doi.org/https://doi.org/10.1111/tgis.13206>
- Xing, J., & Sieber, R. (2023). The Challenges of Integrating Explainable Artificial Intelligence into GeoAI. *Transactions in GIS*, 27(3), 626–645. <https://doi.org/10.1111/tgis.13045>