modelling

MDPI

Article

Two-Stage Prediction of Snowplow Dozer Operation Counts from GPS Data: A Case Study of Akita City, Japan

Akane Yamashita [1], Hiroshi Yokoyama [2] and Yoichi Kageyama [3,*]

[1] Graduate School of Engineering Science, Akita University, 1-1 Tegata Gakuen-Machi, Akita 010-8502, Japan; m8024516@s.akita-u.ac.jp
[2] Center for Information Technology and Management, Akita University, 1-1 Tegata Gakuen-Machi, Akita 010-8502, Japan; yokoyama@ie.akita-u.ac.jp
[3] Faculty of Informatics and Data Science, Akita University, 1-1 Tegata Gakuen-Machi, Akita 010-8502, Japan
* Correspondence: kageyama@ie.akita-u.ac.jp

Abstract

For effective winter road management in snow-prone regions, timely snow removal that reflects weather and traffic conditions is required. In Akita City, Japan, city hall staff measure snow depth and dispatch contracted snow removal crews only when a predefined threshold is exceeded. Consequently, dispatch decisions depend heavily on staff experience. This study demonstrates objective, experience-independent dispatching based on predicting the number of snowplow dozers in operation, thereby reducing the municipal decision burden and improving contractor efficiency. The target variable is highly imbalanced, with long non-operational periods and wide variations in the number of deployed units during snowfall events. When trained directly on such data, models tend to regress toward near-median values and face difficulty capturing operational dynamics. To address this issue, we propose a two-stage framework: firstly, a classifier predicts whether snow removal operations will occur; secondly, a regressor estimates the number of operating dozers based on the operation. We further integrate multi-year datasets to enhance generalization across diverse snow conditions. Experiments showed that the proposed approach achieved an AUPRC of 0.84 for operation occurrence and an RMSE of 1.85 for dozer-count estimation, outperforming models trained on a single year.

Keywords: snowplow; imbalanced data; GPS data; winter road maintenance

1. Introduction

In snow-prone regions, maintaining winter road infrastructure requires snow removal and clearance operations that are appropriately tailored to prevailing weather conditions and traffic patterns [1]. Owing to its long north–south extent and complex topography, the Japanese archipelago exhibits substantial regional heterogeneity in snowfall characteristics [2–5]. Accordingly, many local governments have implemented region-specific snow management strategies, and optimization efforts have progressed in line with local terrain and snowfall conditions [6]. Given the considerable public expenditure devoted to removal of snow that ultimately melts each spring, prior studies have evaluated the economic value of these operations [7–9]. Municipalities must pursue efficient and effective operations that balance fiscal constraints with the need to ensure road safety.

Akita City, located in Akita Prefecture, the most rapidly aging region in Japan, has experienced an acceleration in natural population decline, where deaths are exceeding births, compounded by the continued outmigration of younger generations [10]. Projections

Check for updates

Academic Editor: Wei Gao

Received: 13 February 2026
Revised: 24 March 2026
Accepted: 24 March 2026
Published: 29 March 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

further indicate that by 2045, the share of elderly residents will exceed that of the working-age population [11], raising concerns about future labor shortages among contracted snow removal operators as the workforce ages.

Currently, Akita City operates a dispatch system in which city hall staff measure snow depth and request snowplow deployment once a predefined threshold has been exceeded [12]. Consequently, key decisions regarding snow removal and clearance operations depend substantially on human judgment.

The system in this study was designed to support dispatch decisions by predicting the number of snowplow dozers required to reduce reliance on individual experience. The proposed system partitions Akita City into separate areas and outputs the predicted number of dozers required in the next hour for each one. This is expected to reduce the decision-making burden on municipal staff and improve the operational efficiency of the contracted snow removal companies.

A main challenge is that the target variable, the number of snowplow dozers in operation, is highly imbalanced. Non-operational periods often dominate, and when operations occur, the deployed fleet sizes vary widely. Under such conditions, models trained directly on the raw distribution tend to learn operational patterns insufficiently and produce predictions biased toward near-median values [13]. To overcome this issue, we investigated a two-stage prediction approach that first estimates whether operations will occur and then predicts the number of operating units by integrating the outputs of both stages.

In the envisioned system, the predicted number of operating snowplow dozers is not intended to completely replace that decided by city office staff but rather to be used as supplementary decision-support information. Specifically, by presenting the predicted values together with past operational records and meteorological data, the system is expected to reduce the burden on city office staff when making dispatch decisions and to help contractors anticipate upcoming operational demand.

2. Literature Review

There has been prior research on Winter Road Maintenance (WRM) to prevent road traffic functionality being reduced by snowfall. However, because operational practices, road network configurations, and meteorological conditions in cold regions differ substantially across locations, models proposed in prior studies are not necessarily transferable and cannot be directly applied to other countries or jurisdictions.

Reference [6] further identifies two salient characteristics of WRM in Japan: strong societal expectations to avoid road closures whenever possible and a low tolerance for traffic accidents and congestion during snowfall events. Consequently, WRM in Japan must satisfy stringent service-level requirements while accommodating the country's distinctive topography and climatic conditions.

Winter road conditions are highly variable and uncertain, being strongly influenced by meteorological factors, such as snowfall intensity and accumulation levels. Moreover, planning and managing snow removal and clearance operations are challenging because of multiple practical constraints, including workforce availability and limited budgets. Accordingly, WRM research spans diverse disciplines, including studies on optimizing snowplow vehicle routing and forecasting snowplow dispatch to support operational decision-making.

Representative studies on routing optimization include operational redesign approaches, such as eliminating service boundaries and relocating depots [14,15]. In addition, formulations have incorporated realistic constraints, including heterogeneous vehicle types and the replenishment of materials and resources [16–18]. Nevertheless, these studies have

several limitations. For example, key parameters such as travel speed are often treated as fixed, which may not adequately reflect field operations [14,15,17,18]. In addition, the existing models seldom explicitly account for worker proficiency or the political and contractual constraints that can challenge inter-municipal coordination [14–16].

The research on predicting snowplow dispatch includes decision-support models for real-time deployment [19–23] and forecasting approaches intended to support operational planning [24–26]. However, prior studies have faced notable challenges: performance is often sensitive to the spatial and temporal resolution of meteorological inputs, whereas forecast uncertainty is not always handled explicitly [19–22,24]; many approaches require large volumes of historical data and substantial preprocessing before deployment [23,24,26]; and models frequently require retraining for each region and vehicle type, raising concerns about generalizability to other settings [19–21,25].

This study contributes to the scholarship by providing a two-stage modeling framework that combines a classification model and a regression model to address the imbalanced nature of operational data. Using GPS-derived records from snowplow dozers in Akita City, we estimated whether operations occurred and, when they did, predicted the number of dozers in service, enabling a more reliable estimation of the number of operating snowplow dozers under highly imbalanced conditions.

3. Methodology

3.1. Data and Data Preparation

In this study, we constructed a dataset for model training and evaluation by integrating meteorological data and GPS log data from Akita City. Specifically, each dataset was organized on an hourly basis and then integrated based on the target service area. In addition, feature generation and preprocessing were performed, including missing-value handling and normalization. The meteorological data and GPS log data used in the study are described below.

3.1.1. Snow Removal Operations in Akita City

In response to heavy snowfall in the winter of 2012, Akita City formulated the Akita City Comprehensive Snow Management Basic Plan in 2013 to promote sustainable snow removal practices over the long term [12]. Subsequently, the plan was revised in 2022 to reflect challenges such as population aging and declining birth rates and broader changes in the social environment [12]. Section 3.1 describes Akita City's snow removal operations with reference to the Road Snow Removal Basic Plan [27–29], which was developed in accordance with the comprehensive plan.

During the winter season, Akita City establishes the Akita City Road Snow Removal Headquarters to coordinate road snow removal and clearance operations. Figure 1 shows how the city is partitioned into 11 service areas: Tsuchizaki–Sotoasahikawa (TS), North (NTH), North East (NE), Kawabe (KB), South East (SE), Goshono (GO), Yuwa (YW), South (STH), West (WST), Central South (CS), and Central North (CN). Next, these abbreviations are used to refer to their respective areas. Municipal staff patrol each area and measure snow depth, and snow removal activities are conducted based on these observations.

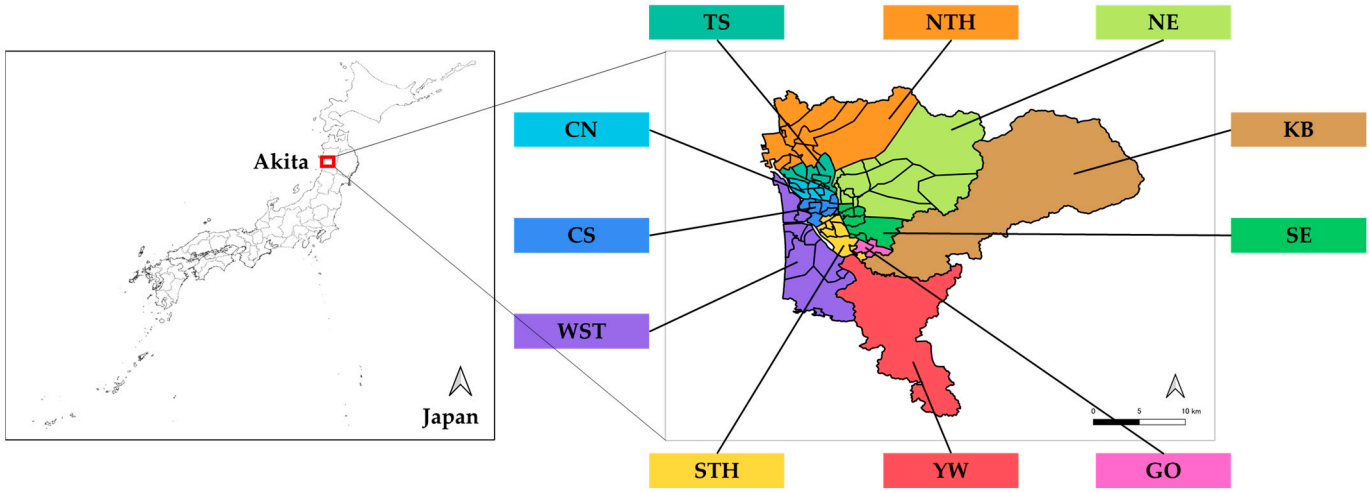


Figure 1. Map of Akita City divided into 11 service areas (the map of Japan on the left was created using a blank map from the Geospatial Information Authority of Japan [30]).

Table 1 summarizes the dispatch criteria and standard operating time windows. Dispatch is initiated when the snow depth on the road surface reaches 10 cm or more and when it is forecast to exceed 10 cm, for both roadways and sidewalks. The operating hours were set to ensure operational efficiency while minimizing disruption to general traffic and pedestrian movement. In the face of snowfall advisories, warnings, or heavy snowfall events, however, the headquarters may direct the scope and timing of operations based on the prevailing weather and road conditions, irrespective of the standard dispatch and implementation criteria.

Table 1. Snow removal dispatch criteria and operational hours.

Road Type	Dispatch Threshold (Snow Depth)	Standard Operating Hours
Carriageway	10 cm	09:00–16:00 20:00–06:00
Sidewalk	10 cm	20:00–06:00

Akita City operates a Snowplow Management System [31] that enables citizens to access the locations and operational status of snowplows via the internet. The system provides configurable display options, including selection of the display area, toggling the visibility of vehicle positions, the visualization of vehicle tracks, and filtering by date and time.

As of 1 October 2024, location-tracking terminals were installed on all snowplows (900 units) and dump trucks (1477 units), excluding small pieces of snow removal equipment.

3.1.2. GPS Data

This study derived the target variable, the number of snowplow dozers in operation, from the GPS data used in the road–snowplow operation management system described in Section 3.1.2. The target area, study period, and procedure for computing the hourly number of operating snowplow dozers are described below.

1. Target Area and Study Period

The prediction target was YW. Analyses were conducted using GPS records collected in R5 (Reiwa 5; FY2023) and R6 (Reiwa 6; FY2024), which we refer to as the R5 and R6 datasets, respectively. The study periods were as follows:

- R5 dataset: 1 November 2023–31 March 2024 (78 days);
- R6 dataset: 1 November 2024–31 March 2025 (77 days).

2. Computation of Operating Counts

Hourly operating counts were computed using the open-source GIS software QGIS [32]. The processing steps were as follows:

(i) The GPS log data were mapped onto a digital map of Akita City.
(ii) For each snowplow dozer, GPS log points/segments that fell within (overlapped with) YW were extracted.
(iii) The extracted points in (ii) were aggregated on an hourly basis.

The resulting hourly number of operating snowplow dozers was used as the target variable. It should be noted that this computation method uses, as an approximation, the number of vehicles present within the service area and does not take into account whether those vehicles were actually engaged in snow removal operations. In addition, historical operating counts were used to compute summary statistics (mean, standard deviation, first difference, sum, and exponential moving average), which were then incorporated as explanatory features. The features related to the operating counts are summarized in Table 2.

Table 2. Features related to the number of snowplow dozers in operation.

Feature	Description
L1, L3, L6, L12	Number of operating snowplow dozers 1, 3, 6, and 12 h before the prediction time
mean_L3, L6, L12	Mean operating count over the previous 3, 6, and 12 h
std_L3, L6, L12	Standard deviation of the operating count over the previous 3, 6, and 12 h
range_L12	Range (max–min) of the operating count over the previous 12 h
sum_L12	Sum of operating counts over the previous 12 h
diff3, diff6	Difference between the operating count 1 h before and 3 and 6 h before
ema6, ema12	Exponential moving average of the operating count over the previous 6 and 12 h

3.1.3. Meteorological Data

Meteorological variables were obtained from the Japan Meteorological Agency, including hourly records of weather conditions, precipitation, air temperature, snowfall, and snow depth [33], and were used as model features. The meteorology-related features used in this study are summarized in Table 3.

Because this study predicts the number of snowplow dozers in operation using GPS data collected in years with different meteorological conditions, we examined interannual differences by comparing (i) the mean air temperature and (ii) the maximum snow depth during the core winter period (December–February) for R5 and R6. Figure 2 presents time-series comparisons of mean temperature and maximum snow depth. The median mean temperature in R5 was higher than that in R6, indicating that R5 was relatively warmer overall. According to the maximum snow depth, R5 included several episodes with snow depths of approximately 30 cm, and for many days, there was a maximum snow depth of 0 cm, suggesting larger day-to-day variability in snow accumulation.

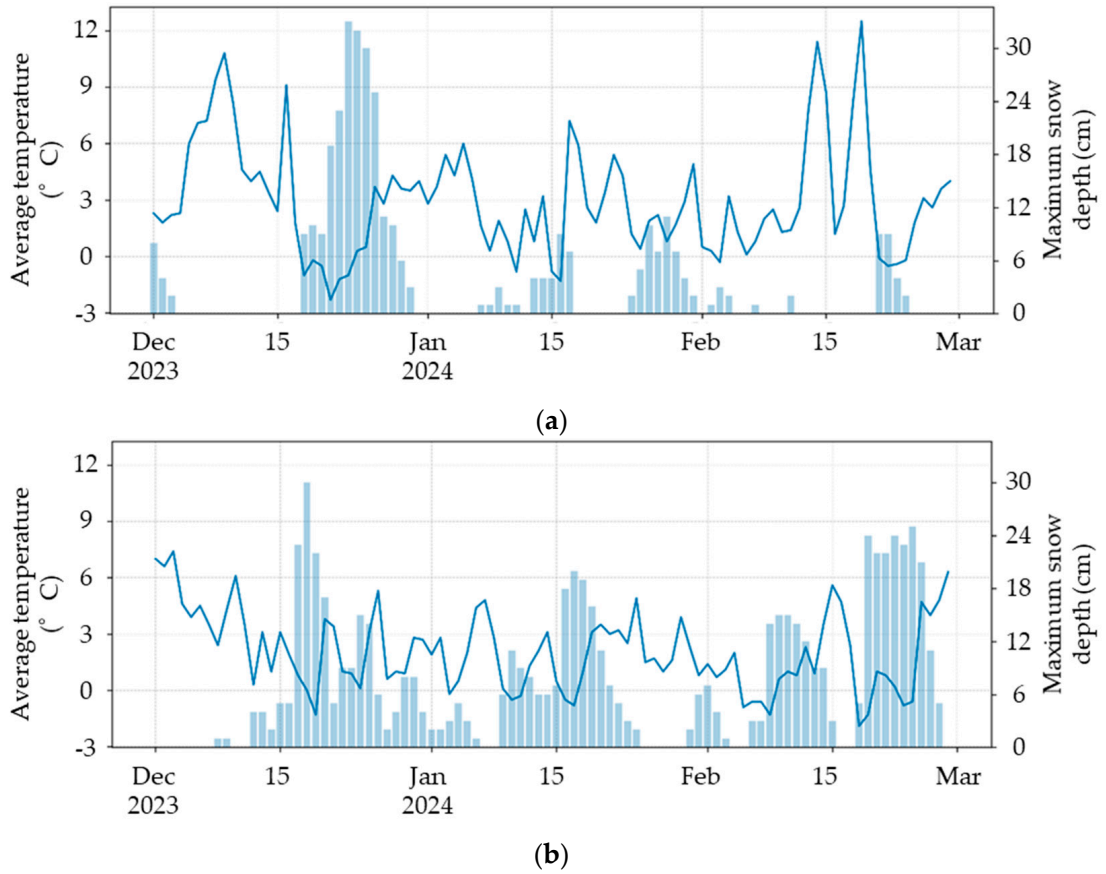


Figure 2. Time-series comparison of mean air temperature (line) and maximum snow depth (bars) for R5 and R6: (**a**) R5; (**b**) R6.

Table 3. Features related to meteorological data.

Feature	Description
Weather	Weather condition (categorical)
precipitation	Hourly precipitation (mm/h)
Temperature	Air temperature (°C)
Temperature_L1, L3, L6, L12	Air temperature 1, 3, 6, and 12 h before the prediction time
Temperature_mean_L6, L12, L24	Mean air temperature over the previous 6, 12, and 24 h
is_freezing	Indicator for air temperature ≤ 0 °C
Snowfall	Hourly snowfall (cm/h)
Snowfall_L1, L3, L6, L12	Snowfall 1, 3, 6, and 12 h before the prediction time
Snowfall_mean_L6, L12, L24	Mean snowfall over the previous 6, 12, and 24 h
Snowfall_std_L6, L12, L24	Standard deviation of snowfall over the previous 6, 12, and 24 h
Snowfall_range_L24	Range (max–min) of snowfall over the previous 24 h
Snowfall_sum_L24	Sum of snowfall over the previous 24 h
Snowfall_diff1, diff3	Difference between current snowfall and that 1 and 3 h earlier

Table 3. *Cont.*

Feature	Description
Snowfall_over_Temperature+1	Snowfall ÷ (air temperature + 1)
Snowfall_ema6, ema12	Exponential moving average of snowfall over the previous 6 and 12 h
Snowdepth	Snow depth (cm)
Snowdepth_L1, L3, L6, L12	Snow depth 1, 3, 6, and 12 h before the prediction time
Snowdepth_mean_L6, L12, L24	Mean snow depth over the previous 6, 12, and 24 h
Snowdepth_std_L6, L12, L24	Standard deviation of snow depth over the previous 6, 12, and 24 h
Snowdepth_range_L24	Range (max–min) of snow depth over the previous 24 h
Snowdepth_sum_L24	Sum of snow depth over the previous 24 h
Snowdepth_diff1, diff3	The difference between the current snow depth and that 1 and 3 h earlier
Snowdepth_over_Temperature+1	Snow depth ÷ (air temperature + 1)
Snowdepth_ema6, ema12	Exponential moving average of snow depth over the previous 6 and 12 h

These results confirm that the snowplow dozer GPS datasets used in this study were collected under different weather conditions.

3.2. Two-Stage Modeling Approach

This study improved the prediction accuracy for the number of snowplow dozers in operation, which exhibits a highly imbalanced distribution. Because the target variable is heavily concentrated at zero, a single-stage regression model tends to struggle to estimate the number of operating snowplow dozers accurately. Therefore, we propose a two-stage framework that decomposes the task into (i) predicting whether snow removal operations are occurring, and (ii) if so, predicting the number of operating dozers. It then combines the outputs of the two stages to produce the final estimate. The overall workflow of the proposed method is shown in Figure 3.

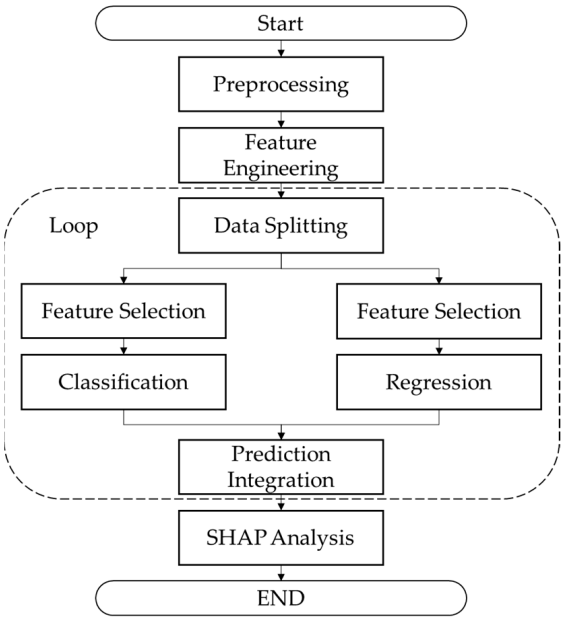


Figure 3. Workflow of the proposed method.

For both stages, we employed Light Gradient Boosting Machine (LightGBM) [34] as the learning algorithm. This choice was motivated by the weak linear associations between the

number of operating snowplow dozers and individual meteorological variables (maximum correlation coefficient: 0.34), indicating the need to model nonlinear relationships. In addition, LightGBM can accommodate imbalanced data through its tree-based structure and objective formulations, making it suitable for our setting, which involves both low pairwise correlations and a pronounced class imbalance.

3.2.1. Classification Model

The classification stage formulates the occurrence or non-occurrence of snowplow dozer operation at each time as a binary classification problem. Let Y denote the operating count. We define the binary label Z according to Equation (1), where $Z = 1$ indicates that operations are occurring and $Z = 0$ indicates that operations are not occurring. Given the explanatory variable X, the classifier estimates the probability of operation, $P(Z = 1 \mid X)$, as expressed in Equation (2).

The dataset used in this study was severely class-imbalanced, with very few samples in the active class and an imbalance ratio (IR) of 5.56. To mitigate the impact of this imbalance and, in particular, to reduce false negatives (i.e., predicting inactivity when operations actually occur), we introduced class weights into the LightGBM loss function.

$$Z = \begin{cases} 1\ (Y > 0) \\ 0\ (Y = 0) \end{cases} \tag{1}$$

$$p = P(Z = 1|X) \tag{2}$$

3.2.2. Regression Model

The regression stage estimates the number of operating units as a continuous value so long as operations are taking place. Let Y denote the operating count. Given the explanatory variables X, the regression model learns the mapping $f_{reg}(X)$ defined in Equation (3).

The outputs of the classification and regression stages are then combined to output the final prediction $\hat{Y}$ according to Equation (4). This integration weights the regression estimate based on the predicted probability of operation: when the probability is high, the regression output is emphasized; when low, its contribution is attenuated.

$$\hat{y}_{reg} = f_{reg}(X) \tag{3}$$

$$\hat{Y} = p^{\gamma}\hat{y}_{reg} \tag{4}$$

where p, $\hat{y}_{reg}$, and γ represent the probability of operation estimated by the classification model, the output of the regression model, and a hyperparameter controlling the contribution of p, respectively.

3.3. Feature Selection

We performed feature selection based on Mutual Information (MI) [35] to improve the generalization performance of both the classification and regression models and mitigate overfitting. MI quantifies the statistical dependence between an explanatory variable X and a response variable Y, as defined in Equation (5). In this study, the MI between each preprocessed feature and the target variable was estimated using a kNN (k-Nearest-Neighbor)-based estimator, and features with higher estimated MI scores were retained.

For the classification stage, MI was estimated using Mutual Information classification [36], with the target treated as a binary label. For the regression stage, MI was estimated using mutual info regression [37] after applying the transformation $log(1 + y)$ to the target variable and restricting the data to positive (operational) samples. In both cases, k was set to 1–3 depending on the size of the training set.

Based on the estimated MI scores, SelectKBest [38] was used to select the top features: $k_{cls} = \max(1, \min(43, p_{cls}))$ for classification and $k_{reg} = \max\left(1, \min\left(33, p_{reg}\right)\right)$ for regression, where p_{cls} and p_{reg} denote the number of available features after preprocessing.

$$I(X;Y) = \sum_{x \in D_X} \sum_{y \in D_Y} P_{X,Y}(x,y) \log \frac{P_{X,Y}(x,y)}{P_X(x) P_Y(y)} \tag{5}$$

where X and Y, D_X and D_Y, $P_{X,Y}(x, y)$, and $P_X(x)$ and $P_Y(y)$ represent the random variables (with X denoting a feature and Y the target variable), the sets of possible values that X and Y can take, the joint probability mass function, and the marginal probability mass functions, respectively.

4. Experiment

4.1. Experimental Setup

We constructed two models and compared their predictive performances: (i) a model trained on the R6 dataset (R6 model) and (ii) a model trained on the combined R5 and R6 datasets (R5 + R6 model). As described in Section 4.2, R5 and R6 exhibit substantially different meteorological conditions, which in turn yield markedly different distributions of the number of operating snowplow dozers in the test data. Therefore, models trained solely on R5 data were excluded from the comparison.

4.1.1. Time-Series Cross-Validation

A key difficulty in evaluating time-series prediction models is that conventional random cross-validation can lead to information leakage by mixing past and future observations [39]. To avoid this issue, we adopted rolling-origin Forward Validation (roFV), which preserves the temporal order. roFV expands the training window by moving the prediction origin forward along the time axis and evaluates performance on a fixed-length test window immediately following each origin [39]. Repeating this procedure over multiple origins provided a more reliable estimate of the generalization performance than a single train–test split.

In this study, we specified 10 prediction origins and evaluated model performance at each origin using a fixed-length evaluation window.

4.1.2. Evaluation Metrics

The model performance was evaluated using Area Under the Precision–Recall Curve (AUPRC); Root Mean Squared Error—dozers (RMSE), i.e., RMSE computed on operational samples only (RMSE_{pos}; dozers); and Mean Absolute Error—dozers (MAE), as defined in Equations (6)–(9). The AUPRC is suitable for imbalanced settings with many negative examples because precision–recall metrics do not depend on true negatives and therefore better reflect the performance of the minority (active) class [40]. The RMSE quantifies the discrepancy between the predicted and observed numbers of operating snowplow dozers, whereas RMSE_{pos} evaluates this discrepancy when operations occurred. MAE measures the mean absolute deviation of predictions from the observed values and is generally less sensitive to outliers than RMSE.

$$AUPRC = \int_0^1 Precision(r)dr \tag{6}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2} \tag{7}$$

$$RMSE_{pos} = \sqrt{\frac{1}{N_{pos}} \sum_{i \in \{z_i > 0\}} (y_i - \hat{y}_i)^2} \tag{8}$$

$$MAE = \frac{1}{N} \sum_{i=1}^{N} |y_i - \hat{y}_i| \tag{9}$$

where r, $\hat{y}_i$, N, $z_i \in \{0, 1\}$, and N_{pos} represent recall, the observed value for sample i, the predicted value for sample i, the number of samples, the binary label indicating whether operations occur for sample i, and the number of samples satisfying $z_i = 1$, respectively.

4.1.3. Feature Importance Calculation

After model training, we computed SHapley Additive exPlanations (SHAP) values, as defined in Equation (10), to quantify the contribution of each feature to the predictions [41]. SHAP provides feature-attribution scores that indicate how individual features influence the model output. In this study, SHAP was used to visualize and interpret the contributions of (i) features derived from the number of operating snowplow dozers and (ii) meteorological features, thereby facilitating the interpretation of the trained models and identifying the variables most influential in prediction performance.

$$\varphi_i(x) = \sum_{S \subseteq F \setminus \{i\}} \frac{|S|!(M - |S| - 1)!}{M!} [v(S \cup \{i\}; x) - v(S; x)] \tag{10}$$

where $\varphi_i(x)$, x, M, $F = \{1, 2, \ldots, M\}$, i, S, $|S|!(M - |S| - 1)!/M!$, $v(S; x)$, and $v(S \cup \{i\}; x) - v(S; x)$ represent the contribution (SHAP value) of the feature i for sample x, the sample to be explained, the number of features, the set of all features, the index of the feature whose contribution is evaluated, a subset of features, the weighting term, the model output when only the features in S are known, and the marginal contribution of adding a feature i, respectively.

4.2. Results and Discussion

Table 4 compares the evaluation metrics of the R6 and R5 + R6 models. As highlighted in red, the R5 + R6 model outperformed the R6 model on all metrics except MAE. In particular, RMSEpos improved by 15.33%, indicating a substantial reduction in prediction error for the number of operating snowplow dozers during periods when operations occurred. In addition, AUPRC increased by 7.21% to above 0.80, suggesting improved detection of the minority (active) class under the imbalanced setting, with fewer missed-operation cases.

Table 4. Features related to meteorological data.

Performance Indicator	R6 Model	R5 + R6 Model	Percentage Change (%)
AUPRC	0.7890	0.8458	+7.21
RMSE	2.0106	1.8514	−7.92
RMSEpos	4.4537	3.7710	−15.33
MAE	0.8196	0.8883	+8.39

Figure 4 presents the PR curve for the R5 + R6 model. The precision remained close to 0.90 over a wide range of recall values, indicating a low false-positive rate for the active class. Moreover, precision does not deteriorate markedly even when recall increases to approximately 0.80, implying strong discriminative capability. Overall, the R5 + R6

model achieves a favorable balance between high precision and high recall across a broad operating regime.

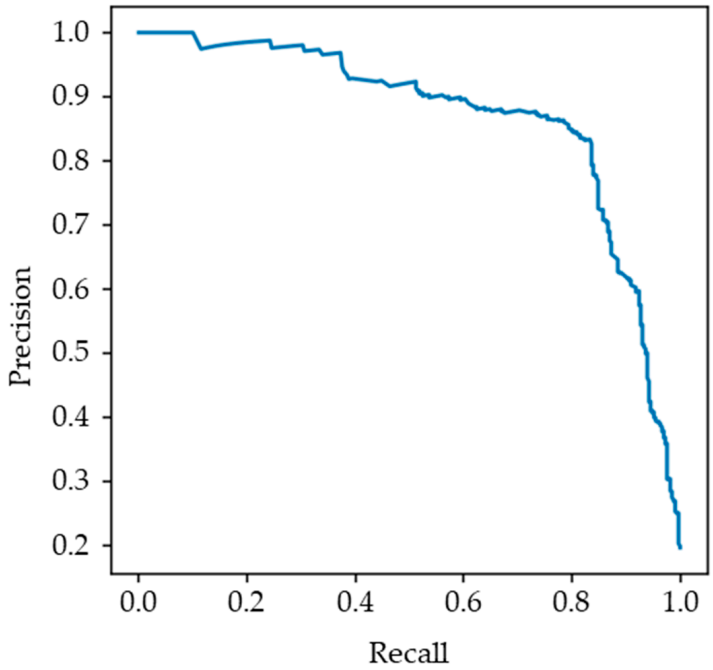


Figure 4. PR curve for the R5 + R6 model.

The time-series comparison between the observed and predicted values for the R5 + R6 model in Figure 5 shows that the model captures overall operational trends. However, one remaining limitation involves underestimation during high-operation periods, defined here as times when the operating count is 15 or higher.

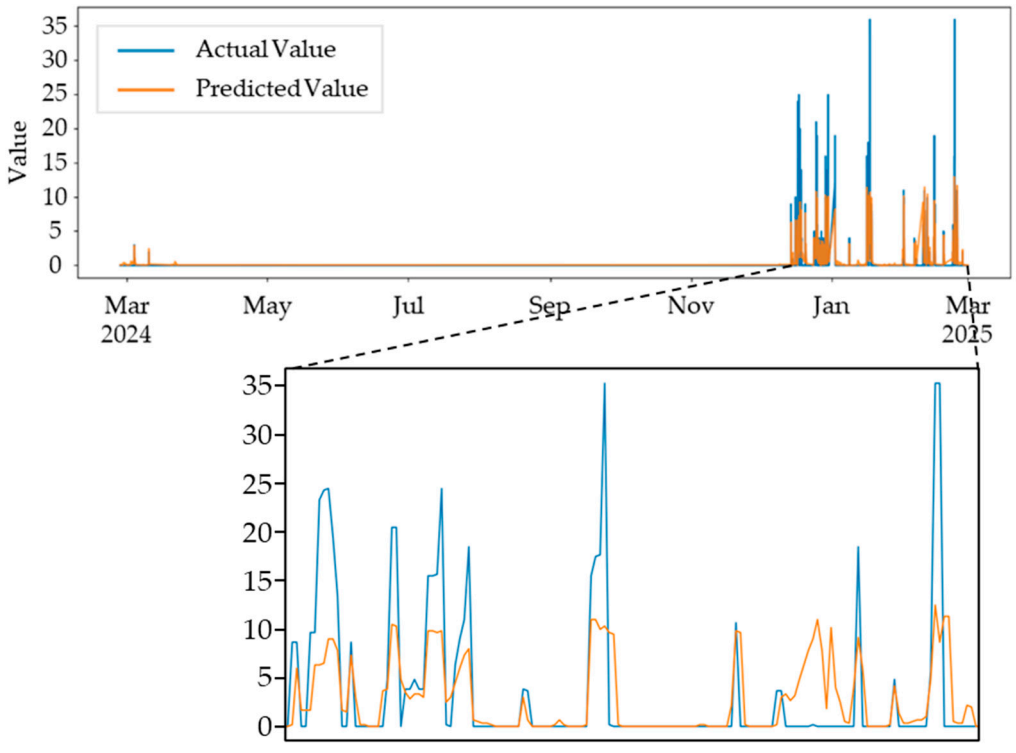


Figure 5. Time-series comparison of actual and predicted operating counts for R5 + R6 model.

The superior performance of the R5 + R6 model relative to the R6 model is primarily attributable to the inclusion of multi-year data. By increasing the volume of operational samples, the model can learn operational patterns under a wider range of meteorological conditions. Nevertheless, to further improve accuracy during high-operation periods, additional training data including instances with 15 or more operating snowplow dozers are required.

In addition, the R5 + R6 dataset contains an approximately seven-month gap between the final observation in R5 and the first observation in R6. Consequently, the model may have a limited ability to represent seasonal transitions and predict the number of operating snowplow dozers at the beginning of the snow season. Therefore, improving early-season prediction is likely to require training over a longer, more continuous time span that covers the full winter period.

Figure 6 shows the feature contributions of the R5 + R6 model. SHAP analysis indicated that the most influential feature was the operating count one hour earlier, suggesting that short-term temporal dependence strongly drives the predicted values. In contrast, among the meteorological features, snowfall-related variables such as Snowfall_ema6, Snowfall_std_L12, and Snowfall_ema12 were ranked relatively high, indicating that snowfall-related features were comparatively important within the meteorological variables. In particular, for Snowfall_ema6, higher feature values tended to be associated with positive SHAP values, suggesting that when snowfall continued over the preceding several hours, the predicted number of operating snow-plow dozers was more likely to increase. However, the distribution range of the SHAP values was smaller than that of the historical operating-count features such as L1 and diff6. In other words, although meteorological features, including snowfall-related variables, constituted one of the background factors influencing dispatch decisions, historical operating counts had stronger explanatory power in the actual prediction. This characteristic is considered to be one of the factors contributing to the underestimation observed during high-operation periods.

Finally, the "number of operating snowplow dozers" in this study was derived from GPS data by counting the vehicles whose logs overlapped each service area. This approach cannot distinguish between vehicles merely traveling and those actively removing snow. Consequently, it may overestimate the true operational activity. Future studies should refine the operating-count definition by incorporating vehicle speed and movement patterns to determine whether a vehicle is performing snow removal operations rather than simply being in transit.

4.3. Model Generalizability Analysis

In this study, a model was developed to predict the number of snowplow dozers in operation for YW. To examine the generalizability of the proposed approach, we analyzed inter-area operational characteristics within Akita City and assessed the feasibility of applying the model to other service areas.

We characterized inter-area operational similarity along two dimensions: (i) consistency in operational status (i.e., whether operations occur) and (ii) variability in the number of snowplow dozers in operation. Accordingly, this section quantifies these characteristics across 11 service areas. Consistency in operational status was evaluated using the Jaccard coefficient [42], whereas similarity in operating-count variability was assessed using Pearson's product–moment correlation coefficient [43].

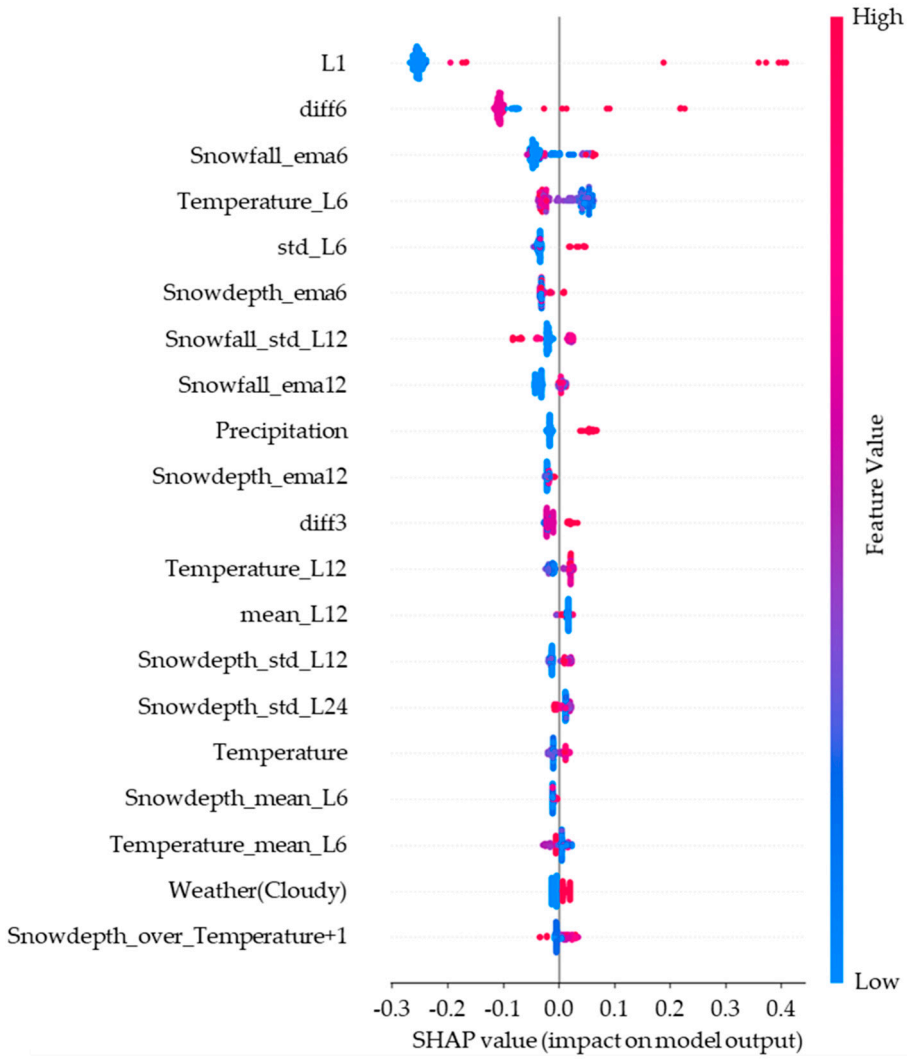


Figure 6. Feature importance in the R5 + R6 model based on SHAP values (features are ordered by importance, measured as means $|SHAP|$).

4.3.1. Accuracy of Operation Status Prediction

The data analyzed in this study were highly imbalanced, with non-operational instances dominating. Therefore, we quantified the inter-area agreement in operational status using the Jaccard coefficient, which emphasizes agreement on operational (positive) instances. The Jaccard coefficient $J(A, B)$ measures the proportion of shared elements between two sets relative to their union and is defined as Equation (11). It ranges from 0 to 1, with values closer to 1 indicating greater similarity.

$$J(A,\ B) = \frac{|A \cap B|}{|A \cup B|} \tag{11}$$

where $|A \cap B|$ and $|A \cup B|$ represent the number of elements common to A and B and the number of elements in the union of A and B, respectively.

The inter-area agreement was evaluated at the event level. An event is defined as a contiguous time interval during which at least one area is operational. For each event, an area is considered a participant if it is operational at any time within the interval. Participation is encoded as a binary indicator (0/1). The pairwise Jaccard coefficient between areas was then computed using the number of events in which both areas participated as the intersection and the number of events in which at least one of the two areas participated as the union.

Figure 7 shows the inter-area Jaccard coefficients for the R5 and R6 datasets. Overall, the operational status consistency was higher in R5 than in R6, with mean Jaccard coefficients of 0.52 and 0.37, respectively. In R5, high agreement was observed among CN, GO, STH, and WST, whereas KB and YW showed low agreement with other areas (0.25–0.33). In R6, NTH and TS exhibited the highest agreement, whereas the number of high-agreement area pairs observed in R5 decreased. Moreover, KB showed very low agreement with most areas (0.10–0.18), suggesting a stronger tendency toward area-specific operational behavior in R6 than in R5.

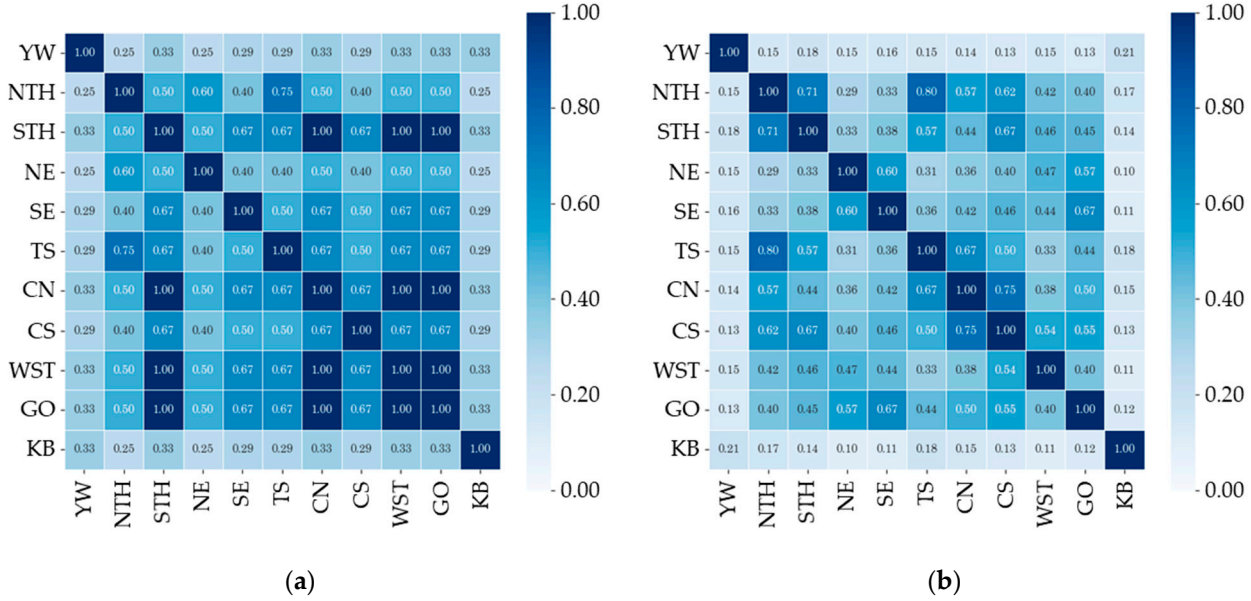


Figure 7. Heatmap of inter-area Jaccard coefficients for operational status: (**a**) R5; (**b**) R6.

We performed hierarchical clustering based on Jaccard distance to identify groups of areas with similar operational statuses. The Jaccard distance $d(A, B)$, defined in Equation (12), converts the Jaccard coefficient into a distance metric. The distances between clusters were computed using the agglomerative (average-linkage) method, which defines the inter-cluster distance as the mean of all pairwise distances between the elements in the two clusters and iteratively merges the two clusters with the smallest distance, as presented in Equation (13).

$$d(A,\ B) = 1 - J(A,\ B) \tag{12}$$

$$d(C_p,\ C_q) = \frac{1}{|C_p||C_q|} \sum_{A \in C_p} \sum_{B \in C_q} d(A, B) \tag{13}$$

where C_p and C_q,$|C_p|$ and $|C_q|$, and $A \in C_p$ and $B \in C_q$ represent clusters, the numbers of elements in each cluster, and elements belonging to each cluster, respectively.

Figure 8 shows the resulting dendrograms. The maximum linkage distance was 0.70 for R5 and 0.86 for R6, indicating lower inter-area agreement for operational status in R6. In both years, the NTH and TS merged at a small distance, reflecting a consistently high agreement. In contrast, KB and YW remained far from other clusters in both years, indicating low agreement with other areas. In addition, cluster composition changed between years: in R5, GO clustered with STH and CN, whereas in R6, it merged at a smaller distance with NE and SE. Further, STH clustered with NTH and TS, and CN clustered with CS. These results suggest that the grouping of areas with high operational consistency was reorganized from R5 to R6.

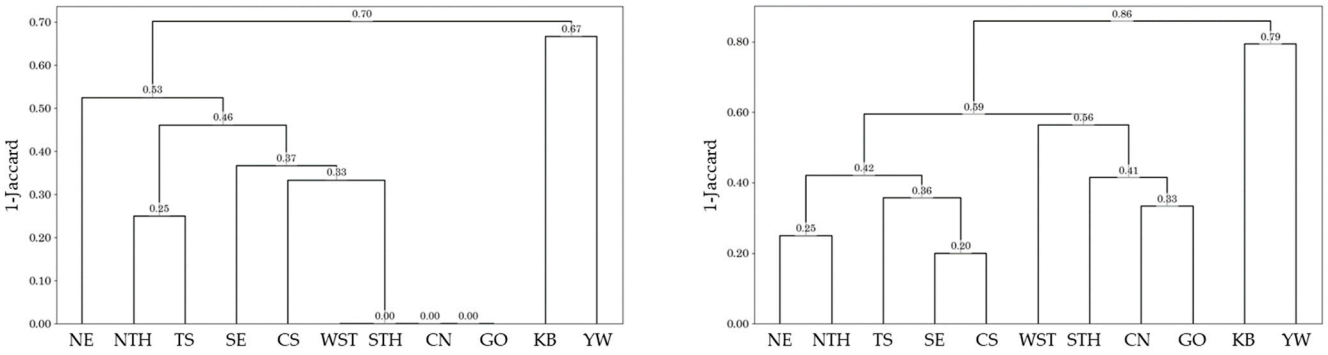


Figure 8. Hierarchical clustering of service areas based on the Jaccard distance (1—Jaccard coefficient) for operational status: (**a**) R5; (**b**) R6.

4.3.2. Variability in the Number of Snowplow Dozers in Operation

We used Pearson's product–moment correlation coefficient to quantify the similarity in the variability of the number of operating snowplow dozers across areas. Pearson's correlation coefficient measures the strength and direction of the linear relationship between two variables and is defined in Equation (14). The correlation magnitudes were interpreted using Evans' guidelines for correlation strength [44].

$$r = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \overline{y})^2}} \tag{14}$$

where n, x_i and y_i, and $\overline{x}$ and $\overline{y}$ represent the number of samples, the values of variables x and y for the i-th sample, and the sample means of x and y, respectively.

Figure 9 shows the inter-area correlation coefficients for the R5 and R6 datasets. A small number of area pairs, such as STH–CN and STH–GO, exhibited "moderate" correlations. In contrast, most other pairs showed "weak" or "very weak" correlations, indicating that only limited combinations of areas shared moderately similar variability in the number of operating snowplow dozers.

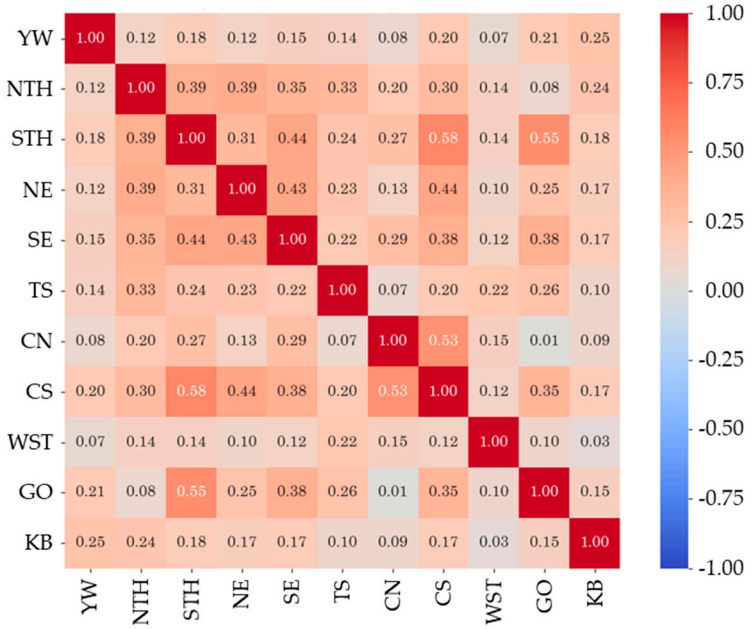


Figure 9. Heatmap of inter-area Pearson correlation coefficients of operating counts.

Figure 10 shows a correlation network constructed from these coefficients using a threshold of 0.4. Areas linked in the network, indicating similar count-operating variability, are geographically contiguous. This suggests that the service areas in the central part of Akita City exhibit relatively similar operational characteristics, whereas on the riverside, TS, WST, YW, and NTH display distinct operating patterns that are more independent of the central areas.

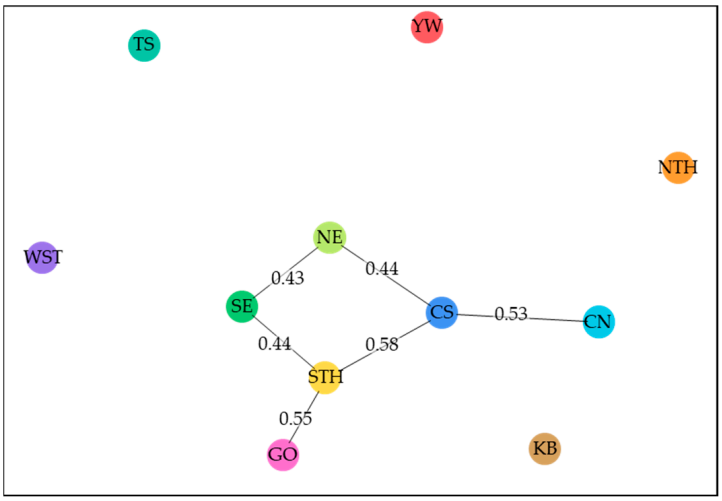


Figure 10. Correlation network of operating counts between service areas ($|r| \geq 0.4$).

4.3.3. Comparison of Results

Table 5 summarizes the evaluation metrics for the 11 service areas in Akita City. AUPRC exceeded 0.70 in all areas, and NTH, NE, and CN (highlighted in red) achieved higher AUPRC values than YW. RMSE was approximately 1.85 for YW, whereas all other areas recorded values above 2.00. $RMSE_{pos}$ indicates substantially larger errors during operational periods: approximately 3.77 for YW, 66.87 for NTH, and 55.62 for STH. MAE was approximately 0.88 for YW, whereas the other areas ranged from approximately 0.96 to 5.55. In several areas, MAE was notably smaller than RMSE, consistent with its lower sensitivity to outliers.

Table 5. Comparison of evaluation metrics across the 11 service areas in Akita City (mean values).

Area Name	AUPRC	RMSE	RMSEpos	MAE
YW	0.8458	1.8514	3.7710	0.8883
NTH	0.8859	14.2299	66.8704	4.9830
STH	0.8194	15.8217	55.6263	5.5513
NE	0.8471	14.2955	27.7614	4.3692
SE	0.7806	6.7033	22.8100	2.4878
TS	0.7175	14.8749	71.5417	5.0940
CN	0.9141	7.6036	45.6763	2.6249
CS	0.8399	7.6516	34.8931	2.7131
WST	0.7096	6.4916	21.8381	2.3202
GO	0.7316	8.9888	31.6411	2.8093
KB	0.8341	2.1534	4.4222	0.9622

Overall, KB achieved a performance comparable to YW across the evaluation metrics, whereas the remaining nine areas generally exhibited a weaker performance compared to

YW. This degradation is likely attributable to differences in operational characteristics and data distributions between areas, as discussed above. The degradation in performance is considered to reflect the differences in operational characteristics and data distributions among the service areas described above. The model constructed in this study was trained using operating-count data from YW, where operational patterns differ from those in the other areas. Therefore, differences in road structure, traffic conditions, the composition of priority routes, and the actual operational practices used for dispatch decisions among the areas likely affected the predictive performance. In particular, YW and KB receive more snowfall than the other areas and, although they formally follow the same dispatch criteria, snow removal in these areas is carried out based on the judgment of contractors rather than city office staff. Therefore, to improve predictive accuracy, it is necessary not only to construct models for groups of areas with similar operational behavior, but also to design models that explicitly incorporate area-specific characteristics such as road structure and traffic conditions.

4.4. Limitations and Future Work

In this study, a two-stage model was constructed to predict the number of snowplow dozers in operation in Akita City. Although the proposed method has demonstrated a certain level of effectiveness, it also has several limitations. The limitations of this study and corresponding directions for future work are described below.

1. Definition of the Target Variable

The target variable used in this study was defined approximately by aggregating the number of vehicles present within the target service area from GPS logs, and it did not distinguish between vehicles that were actually engaged in snow removal operations and those that were simply in transit. This means that the target variable may not accurately reflect the actual operational status. Therefore, future work should define the target variable more precisely by incorporating additional information such as vehicle speed, back-and-forth movement patterns, and travel routes to distinguish active operations from simple movement.

2. Limitations in predictive accuracy during high-operation periods

This study primarily focused on comparing single- and multi-year training; however, comparisons with other methods specifically designed for imbalanced data remain insufficient. In the operating-count data analyzed in this study, zero values accounted for the majority of observations, while high-operation cases were statistically rare. As a result, the prediction model tended to be trained in a way that reduced overall error, which led to underestimation during high-operation periods. Therefore, future work should compare the proposed method with alternative approaches to handling imbalanced data, such as oversampling methods and the supplementation of high-operation cases using generative models.

5. Conclusions

The system in this study can assist in snowplow dispatch decisions in Akita City, where the population is aging. Our investigation into the prediction of snowplow dozer operating counts demonstrated that models trained on integrated multi-year datasets achieved a better performance than those trained on data from a single year. In addition, a two-stage framework combining classification and regression was able to capture operational trends, yielding an MAE of approximately 0.8.

Inter-area analyses of operational characteristics further suggested that the variability in the number of operating snowplow dozers differs between the central area of Akita City

and other service areas and that the composition of area groups exhibiting high agreement in operational status is not necessarily consistent across years. Therefore, for citywide predictions, it is necessary to construct models for groups of areas with similar operational characteristics while explicitly accounting for inter-area differences in data distribution.

Nevertheless, the data used in this study are limited to two snow seasons and the proposed approach has not yet achieved accurate prediction during high-operation periods (i.e., when the number of operating snowplow dozers is 15 or more) or at the beginning of the snow season. Future work is required to further improve the prediction accuracy by incorporating additional seasons under diverse snowfall conditions and extending the observation period to include longer time spans.

Author Contributions: Methodology, A.Y.; writing—original draft preparation, A.Y.; writing—review and editing, Y.K. and H.Y.; supervision, Y.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data presented in this study are available upon request from the corresponding author due to privacy concerns regarding the study participants.

Acknowledgments: The authors thank Akita City Hall for providing the snowplow GPS data used in this study. During the preparation of this manuscript, the authors used ChatGPT-5.2 to translate the text from Japanese, their native language, into English. The authors have reviewed and edited the output and take full responsibility for the content of this article.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

WRM	Winter Road Maintenance
kNN	k-Nearest-Neighbor
MI	Mutual Information
roFV	Rolling-Origin Forward Validation
AUPRC	Area Under the Precision–Recall Curve
RMSE	Root Mean Squared Error
MAE	Mean Absolute Error
SHAP	SHapley Additive exPlanations

References

1. Sakai, T.; Kuriyama, H. An Evaluation Method for Road Snow Removal Costs. *Seppyo J. Jpn. Soc. Snow Ice* **1993**, *55*, 327–334. [CrossRef]
2. Yano, K. Snow Cover in Japan. *J. Jpn. Soc. Hydrol. Water Resour.* **1993**, *6*, 48–57. [CrossRef]
3. Ishizaka, M. Regional Characteristics of Winter Climate and Snowfall in Japan. *Seppyo J. Jpn. Soc. Snow Ice* **1996**, *58*, 329–338. [CrossRef]
4. Ishii, H.; Suzuki, K. Regional Characteristics of Variation of Snowfall in Japan. *J. Jpn. Assoc. Hydrol. Sci.* **2011**, *41*, 27–37. [CrossRef]
5. Daimaru, H.; Onodera, H.; Matsuoka, H.; Kitada, M.; Oohara, H.; Higuchi, H. Snow type distribution in the coldest season, between Akita and Miyako, in the northern part of Tohoku District, Japan. *Seppyo J. Jpn. Soc. Snow Ice* **1993**, *55*, 207–212. [CrossRef]
6. Perrier, N.; Langevin, A.; Campbell, J.F. A Survey of Models and Algorithms for Winter Road Maintenance. Part I: System Design for Spreading and Plowing. *Comput. Oper. Res.* **2006**, *33*, 209–238. [CrossRef]
7. Kamimura, Y. An Evaluation of Economic Criteria for Snow-Plowing Operation. *Seppyo J. Jpn. Soc. Snow Ice* **1998**, *60*, 25–35. [CrossRef]
8. Hanbali, R.M. Economic Impact of Winter Road Maintenance on Road Users. *Transp. Res. Rec. J. Transp. Res. Board* **1994**, *1442*, 151–161.
9. Hayashiyama, Y.; Tanabe, S. The Economic Evaluation of the Winter Road Service Level by the Conjoint Analysis: Effectiveness of the Orthogonal Main Effect Design. *Jpn. J. City Plan.* **2002**, *19*, 47–54. [CrossRef]

10. Information and Statistics Division, Planning and Finance Department, Akita City. *Population of Akita City by Age Group and District (Estimates Based on the 2020 Population Census)*; Akita City: Akita, Japan, 2025.
11. Akita City. *Akita City Population Vision*; Akita City: Akita, Japan, 2021.
12. Akita City. *Akita City Comprehensive Snow Management Basic Plan*; Akita City: Akita, Japan, 2022.
13. Yang, Y.; Zha, K.; Chen, Y.; Wang, H.; Katabi, D. Delving into Deep Imbalanced Regression. In Proceedings of the 38th International Conference on Machine Learning (ICML 2021), Virtual, 18–24 July 2021; Volume 139, pp. 11842–11851.
14. Nguyen, P.H.D.; Tran, D. Constraint-Based Snowplow Optimization Model for Winter Maintenance Operations. *Transp. Res. Part A Policy Pract.* **2024**, *179*, 103911. [CrossRef]
15. Xu, S.; Wu, M.; Kwon, T.J.; Perchanok, M. Optimization Models for Snowplow Routes and Depot Locations: A Real-World Implementation. *Transp. Res. Rec. J. Transp. Res. Board* **2022**, *2676*, 199–211. [CrossRef]
16. Kinable, J.; van Hoeve, W.-J.; Smith, S.F. Optimization Models for a Real-World Snow Plow Routing Problem. In *Integration of AI and OR Techniques in Constraint Programming (CPAIOR 2016)*; Quimper, C.-G., Ed.; Lecture Notes in Computer Science; Springer: Cham, Switzerland, 2016; Volume 9676, pp. 229–245. [CrossRef]
17. Kinable, J.; van Hoeve, W.-J.; Smith, S.F. Snow Plow Route Optimization: A Constraint Programming Approach. *IISE Trans.* **2021**, *53*, 685–703. [CrossRef]
18. Rasul, A.; Seo, J.; Xu, S.; Kwon, T.J.; MacLean, J.; Brown, C. Optimization of Snowplow Routes for Real-World Conditions. *Sustainability* **2022**, *14*, 13130. [CrossRef]
19. The Weather Network Commercial Services. *Maintenance Decision Support System (MDSS) Field Trial—Final Report*; Pelmorex Media Inc.: Oakville, ON, Canada, 2012.
20. Maintenance Decision Support System (MDSS). Available online: https://ral.ucar.edu/solutions/products/maintenance-decision-support-system-mdss (accessed on 16 January 2026).
21. Pisano, P.A.; Nelson, G.G. Advanced Decision Support for Winter Road Maintenance: FHWA Documentation of Requirements for Intelligent Transportation Systems. *Transp. Res. Rec. J. Transp. Res. Board* **2001**, *1741*, 129–136. [CrossRef]
22. Fu, L.; Trudel, M.; Kim, V. Optimizing Winter Road Maintenance Operations under Real-Time Information. *Eur. J. Oper. Res.* **2009**, *196*, 332–341. [CrossRef]
23. Okura, H.; Yokoyama, S.; Yamashita, T.; Kawamura, H. Analysis of Weather Information and Road Surface Images for Snow Removal Dispatch Prediction. In Proceedings of the INTELLI 2024: The Thirteenth International Conference on Intelligent Systems and Applications, Athens, Greece, 10–14 March 2024; pp. 17–24.
24. Kršmanc, R.; Slak, A.Š.; Demšar, J. Statistical Approach for Forecasting Road Surface Temperature. *Met. Apps.* **2013**, *20*, 439–446. [CrossRef]
25. Ahabchane, C.; Trépanier, M.; Langevin, A. Street-Segment-Based Salt and Abrasive Prediction for Winter Maintenance Using Machine Learning and GIS. *Trans. GIS* **2019**, *23*, 48–69. [CrossRef]
26. Mohammadi, N.; Klein-Paste, A.; Lysbakken, K.R. Simulating Winter Maintenance Efforts: A Multi-Linear Regression Model. *Cold Reg. Sci. Technol.* **2024**, *227*, 104307. [CrossRef]
27. Construction Department, Akita City. *Basic Plan for Road Snow Removal and Clearance (FY2023)*; Akita City: Akita, Japan, 2023.
28. Construction Department, Akita City. *Basic Plan for Road Snow Removal and Clearance (FY2024)*; Akita City: Akita, Japan, 2024.
29. Construction Department, Akita City. *Basic Plan for Road Snow Removal and Clearance (FY2025)*; Akita City: Akita, Japan, 2025.
30. Geospatial Information Authority of Japan. Available online: https://www.gsi.go.jp/kikakuchousei/kikakuchousei40182.html (accessed on 16 January 2026).
31. Snow Removal Vehicle Management System. Available online: https://akitashi-jyosetsu.ndsoft.jp/public/map/system/?mode=view (accessed on 16 January 2026).
32. QGIS Development Team. *QGIS*, version 3.34. Geographic Information System. QGIS: Zurich, Switzerland. Available online: https://qgis.org/ (accessed on 16 January 2026).
33. Past Weather Data Search. Available online: https://www.data.jma.go.jp/stats/etrn/index.php (accessed on 16 January 2026).
34. Ke, G.; Meng, Q.; Finley, T.; Wang, T.; Chen, W.; Ma, W.; Ye, Q.; Liu, T.-Y. LightGBM: A Highly Efficient Gradient Boosting Decision Tree. In Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA, 4–9 December 2017; pp. 3149–3157.
35. Shannon, C.E. A Mathematical Theory of Communication. *Bell Syst. Tech. J.* **1948**, *27*, 379–423. [CrossRef]
36. Scikit-Learn Developers. sklearn.feature_selection.mutual_info_classif. Available online: https://scikit-learn.org/stable/modules/generated/sklearn.feature_selection.mutual_info_classif.html (accessed on 16 January 2026).
37. Scikit-Learn Developers. sklearn.feature_selection.mutual_info_regression. Available online: https://scikit-learn.org/stable/modules/generated/sklearn.feature_selection.mutual_info_regression.html (accessed on 16 January 2026).
38. Scikit-Learn Developers. sklearn.feature_selection.SelectKBest. Available online: https://scikit-learn.org/stable/modules/generated/sklearn.feature_selection.SelectKBest.html (accessed on 16 January 2026).

39. Schnaubelt, M. *A Comparison of Machine Learning Model Validation Schemes for Non-Stationary Time Series Data*; FAU Discussion Papers in Economics, No. 11/2019; FAU: Erlangen, Germany, 2019.
40. Davis, J.; Goadrich, M. The Relationship Between Precision-Recall and ROC Curves. In Proceedings of the 23rd International Conference on Machine Learning (ICML 2006), Pittsburgh, PA, USA, 25–29 June 2006; pp. 233–240. [CrossRef]
41. Lundberg, S.M.; Lee, S.-I. A Unified Approach to Interpreting Model Predictions. In Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA, 4–9 December 2017; pp. 4768–4777.
42. Jaccard, P. Étude Comparative de la Distribution Florale dans une Portion des Alpes et du Jura. *Bull. Soc. Vaudoise Sci. Nat.* **1901**, *37*, 547–579. [CrossRef]
43. Pearson, K. Mathematical Contributions to the Theory of Evolution—III. Regression, Heredity, and Panmixia. *Philos. Trans. A Math. Phys. Eng. Sci.* **1896**, *187*, 253–318. [CrossRef]
44. Schober, P.; Boer, C.; Schwarte, L.A. Correlation Coefficients: Appropriate Use and Interpretation. *Anesth. Analg.* **2018**, *126*, 1763–1768. [CrossRef] [PubMed]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.