HKIE THE HONG KONG INSTITUTION OF ENGINEERS 香港工程師學會

Machine learning models for window opening status and operation behaviours in sub-tropical residential buildings

Tsz Wun Tsang, Bian Liu and Ling Tim Wong
Department of Building Environment and Energy Engineering, The Hong Kong Polytechnic University, Hong Kong, China

ABSTRACT

Using a database of environmental observations on window status, window operation behaviours, and ambient environmental conditions, this study identified environmental drivers for window operation, offering robust predictive models for window opening status and window operation behaviour. Predictive models were developed using various algorithms, including logistic regression (LR), Gaussian Naive Bayes classifiers (GNBs), decision trees (DTs), random forests (RFs), and support vector machines (SVMs). It was found that SVMs, LR and GNBs demonstrated exemplary performance in predicting a simple task such as window opening status. However, these models struggled when faced with a multiclass classification problem indicating window operation behaviour. Conversely, tree-based models effectively predicted window opening status and operation. Furthermore, by examining Gini importance (IG) scores, permutation importance (PIMP), and SHapley Additive exPlanation (SHAP) values, the study found that temperature and season were critical features in predicting window opening status, whereas the diurnal cycle exerted the strongest influence on window operation behaviour. Overall, this study systematically assessed the comparative performance of multiple machine learning approaches for predicting window status and operational behaviour, highlighting the importance of effective predictive model selection. The results contribute to advancing window behaviour modelling and improving the predictive reliability of building energy simulations.

KEYWORDS window opening status; window operation behaviours; residential buildings;machine learning; environmental drivers

CONTACT Tsz Wun Tsang hayley.tsang@polyu.edu.hk
Received 22 January 2025

1. Introduction

Maintaining optimal building performance has become increasingly challenging, with growing concerns about energy efficiency, indoor environmental quality (IEQ), and the escalating effects of climate change (Luo et al., 2018). Occupants are central in determining energy consumption patterns and their impact on the overall building performance (Jia et al., 2017; Kim et al., 2017). Understanding occupant behaviour is essential for developing strategies to improve energy efficiency and enhance IEQ.

Brager and de Dear (1998) proposed that occupants could achieve their thermal adaptation through behavioural adjustments, physiological acclimatisation, and psychological habituation or expectations. Giving occupants more control over the indoor environment can also enhance their thermal comfort. The underlying principle of environmental adaptation is based on the assumption that people respond in ways that restore comfort if they experience discomfort (Nicol & Humphreys, 2002). Specifically, actions such as window opening, the use of fans, heaters, and air conditioning, and adjusting clothing insulation have been found to contribute to improved thermal comfort (Rijal et al., 2018, 2021; KC et al., 2018; Zaki et al., 2017). This natural tendency for behavioural adjustment presents significant opportunities to modify both the perceived and actual indoor conditions.

In sub-tropical climates, window opening and closing is a prevalent strategy for thermal control. Occupants often open windows when they feel hot and desire a cooler indoor environment, and vice versa. This adaptive behaviour of adjusting window status contributes to energy savings by reducing the reliance on mechanical cooling or heating but also facilitates beneficial interaction between indoor and outdoor environments and enhances ventilation (Rijal et al., 2007). To better understand and model these adaptive behaviours, numerous studies have investigated the factors that lead to window opening and closing behaviours and have developed models to predict this behaviour (Hawila et al., 2023). In modelling window behaviours, the predominant approach revolves around analysing the window's status, specifically the frequency of window openings. This aspect is relatively more straightforward to measure, making it a popular choice for modelling and analysis purposes. Alternatively, an action-based approach by modelling the transition probability of window opening and closing can identify the specific conditions that trigger window operation, providing a different perspective on understanding window operation behaviours and contributing to our knowledge of the factors driving such behaviours (Tsang et al., 2024).

In addition to different approaches for modelling window opening behaviours, the choice of modelling techniques can also significantly impact the model's performance. Most existing studies have utilised multivariate linear and logistic regression to model window opening behaviour, incorporating various environmental and contextual parameters (Yang et al., 2022; Yao et al., 2017; Jeong et al., 2016; Rouleau et al., 2020). These regression-based models have generally demonstrated a good level of accuracy, reaching up to 75%. Recently, some researchers have explored the use of machine learning algorithms for modelling window behaviour. For example, Mo et al. (2019) employed the XGBoost algorithm to develop a predictive model for window behaviour in residential buildings in a hot summer and warm winter region of China. Their findings indicated that XGBoost offered a substantial advantage in modelling accuracy compared to the LR analysis. Similarly, Park et al. (2021) applied six machine learning algorithms, including k-nearest neighbour (KNN), RF, artificial neural networks (ANNs), classification and regression trees (CARTs), chi-squared automatic interaction detectors (CHAIDs), and SVMs, to predict window status and behaviour in Seoul and suburban areas. Other methods, such as recurrent neural networks (RNNs) and Naïve Bayes, have also been employed to develop window-opening models (Furuhashi et al., 2022; Han et al., 2020). These machine learning models have proven useful in recognising and predicting individual behavioural patterns, often providing higher accuracy and better interpretability of correlated factors than LR models (Wei et al., 2019; Pan et al., 2019; Markovic et al., 2018). However, only a limited number of studies performed feature importance analysis to identify critical predictors of window opening status, and the importance evaluation was typically restricted to simplistic indicators, such as single Gini-based scores (Park et al. 2021; Wei et al., 2019). Besides, studies which evaluated multiple machine learning models in regard to window status and behaviour prediction frequently omitted hyperparameter tuning (Park et al., 2021).

This study investigates the performance of modelling techniques for predicting window opening status and window opening behaviour. Using a dataset containing window opening patterns observed in apartment bedrooms in a sub-tropical region, the research employs a range of prominent machine learning algorithms alongside LR to model the status and operation of windows. Besides, given the limited implementation of feature importance analysis in previous studies, this study prioritised a comprehensive interpretability assessment using three complementary indicators (IG scores, PIMP, SHAP values) to explore the contribution of relative factors used to predict the window opening status and actions. Moreover, the accuracy of each predictive model, and the effects of hyperparameter tuning are investigated to address the lack of systematic hyperparameter optimisation in previous multi-model studies. The study aims to evaluate the impact of different modelling techniques on the performance of predictive models for window opening status and behaviour. Furthermore, it seeks to deepen the understanding of the environmental factors influencing window opening status and operational behaviours in residential buildings located in sub-tropical regions.

2. Methodology

2.1. Window opening behaviour and environmental condition database

The window opening behaviour database previously collected and reported in Tsang et al. (2024) was utilised to investigate and model the window opening behaviour in bedroom environments. The pilot survey included 80 participants (26 male and 54 female respondents). Among them, 53% were aged 19–30 years, 32% were aged 31–49 years, and the rest were aged 50 years or above. Among these participants, 29 individuals (16 male and 13 female) reported regularly using window opening as a ventilation strategy in their bedrooms in all four seasons. The final dataset used in this study comprised 672 hours of valid window opening behaviour data collected from these 29 participants.

This database comprised crucial information, including self-reported window opening behaviour, air-conditioning usage habits, and outdoor environmental parameters such as air temperature (*°C*), rainfall (*mm*), relative humidity (*%*), wind speed (*km/h*), and solar radiation (*W/m²*). The data were gathered from bedrooms in Hong Kong over a one-week measurement across the four distinct seasons.

Leveraging this dataset, this study evaluates and compares various methodologies for identifying feature importance, data preprocessing techniques, and machine learning algorithms. This study aims to develop robust models capable of accurately assessing and predicting both the action of window opening or closing (window opening behaviour), and the status of a window is open or closed (window opening status).

2.2. Modelling of window opening status and window operation behaviours

2.2.1. Input features and data preprocessing

The modelling of window opening status and window operation behaviours involved binary and multiclass classification problems, considering a combination of categorical and numerical input features. Data preprocessing is important in developing robust machine learning models since it mitigates the potential impact of untreated input data on prediction outcomes (Zelaya, 2019). In this study, the numerical data underwent preprocessing utilising various techniques, including standardisation (Z-score normalisation), normalisation, robust scaling, and log transformation. Non-numeric variables were encoded into numeric form to ensure compatibility with the machine learning algorithms. To represent diurnal periodicity appropriately, the hour of day variable was transformed into cyclical sine and cosine terms. For model development, the dataset was randomly divided into training and testing subsets using a 75%–25% split. Therefore, the reported results reflect single hold-out test performance.

Table 1. Environmental features, classifications for window operation behaviour and window opening status.

Features	Observable range	Classification	Class
Temperature (°C)	13.9–33.2	Window opening status	Open
Wind speed (km/h)	0.8–34.9		Close
Relative humidity (%)	43–97	Window operation behaviours	No action
Time (in Hours)	0–24		Open from its closed position
Solar radiation (W/m^2)	0–913.4		Close from its opened position
Air quality health index (AQHI)	2–9		Adjust to a partially closed position
Rainfall (mm)	0–5.7		Adjust to a partially open position
Weekday/Weekend	Weekday Weekend		Close to switch on the air conditioner
Season	Spring Summer Autumn Winter		

Due to the heavy imbalance in the data for window operation behaviours, with a predominance of the "no action" class (95.7%), prediction models could perform well in predicting the dominant class. Still, they might perform poorly in the minority classes. To address this imbalance, the Synthetic Minority Over-sampling Technique (SMOTE), an oversampling technique, was employed to enhance the model's performance on the minority classes by creating synthetic examples through the duplication or generation of new instances (Chawla et al., 2002; He & Garcia, 2009).

2.2.2. Machine learning models

Various popular machine learning models suitable for handling mixed data were trained to assess their performance in predicting the window opening status and window operation behaviours. The models employed in this assessment comprised an LR algorithm, GNB, DT, RF, and SVM. LR assumes a linear decision boundary and utilises a logistic function to model and predict the probability of belonging to a specific class (Hosmer et al., 2013). In contrast, based on Bayesian probability, GNB predicts class probabilities by assuming conditional independence among the features (Berrar, 2018). DT models predict the outcome class through recursive partitioning of the input features (Myles et al., 2004), while RF models aggregate predictions from a collection of randomised DTs to prevent overfitting and enhance model robustness and generalisation (Breiman, 2001). DT models with maximum depths of 5, 10, and 15 were trained, along with RF models employing 10, 50, and 100 estimators.

SVM models identify an optimal hyperplane to separate data from different classes. They possess high generalisation ability, which is effective for problems involving high-dimensional feature spaces (Cortes & Vapnik, 1995). In this study, both linear and non-linear SVM models were trained. The non-linear models employed radial-basis function (RBF) kernels and polynomial kernels with 3, 4 and 5 degrees. Additionally, sigmoid kernels with gamma values of 0.1, 1, and 10 and coef0 values of -1, -0.5, 0.5, and 1 were utilised.

The five models were chosen to provide a representative comparison across the algorithmic families currently relevant to window behaviour prediction. LR was included as the conventional reference method which has been widely used for window opening prediction (Park et al., 2021). DT and RF were selected as representative tree-based methods, while an SVM was included as it is a widely used nonlinear classifier in recent machine learning studies (Park et al., 2021; Furuhashi et al., 2022). GNB was adopted as a Naïve Bayes-based probabilistic baseline that offers low computational cost and a simple structure (Furuhashi et al., 2022). Therefore, this study was designed to compare interpretable statistical models, probabilistic models, tree-based methods, and nonlinear classifiers within a unified evaluation framework.

2.2.3. Model evaluation

Precision, recall, F1 score and overall accuracy were employed to evaluate the predictive performance of the models, which are widely used in studies concerning machine learning to evaluate the performance of predictive models (Aloisio et al., 2024). Precision measures the accuracy of positive predictions, recall assesses the model's ability to identify all positive instances correctly, and the F1 score combines precision and recall into a single performance metric. Overall accuracy measures the proportion of correct predictions out of all the predictions. Equation (1) shows the calculation of the F1 score as the harmonic mean of precision and recall (Zelaya, 2019).

$$F1\ score = 2 \times \frac{recall \times precision}{recall + precision} = \frac{2TP}{2TP+FP+FN}. \qquad (1)$$

2.2.4. Feature Importance

Based on the results from model evaluation, the best performance models were selected to conduct feature importance analysis for both window-status classification (open vs. closed) and window operation behaviour prediction. Feature importance was estimated using I_G, PIMP and SHAP values, with the choice of method determined by the structural properties of the selected model.

The I_G score is a widely recognised indicator for evaluating the relevance of features in a tree-based model. It quantifies both the frequency with which a feature i is selected for splitting within the model and its contribution to the model's overall performance in distinguishing between different classes. The calculation of the I_G involves accumulating the Gini impurity $G(\tau)$ resulting from the optimal split $\Delta G_i(\tau, T)$ for all nodes τ across all trees T in the RF, considering the feature i under evaluation (Menze et al., 2009).

$$I_G(i) = \sum_T \sum_\tau \Delta G_i(\tau, T). \tag{2}$$

PIMP addresses the bias in I_G favouring features with more categories (Strobal et al., 2007). The PIMP algorithm, as shown in full detail in Altmann et al. (2010), generates a null importances vector through repeated permutations, quantifying the relevance of features. The maximum likelihood of the measured importance on the unpermuted outcome vector is computed by fitting a probability distribution to the null importances. The resulting PIMP p-value serves as an unbiased and model-agnostic measure of feature relevance.

Based on game theory, the SHAP approach explains individual predictions by assigning contributions to each feature. SHAP values for a feature i among n features in a prediction model p are obtained using Equation (3), where S represents subsets of features excluding feature i (Lundberg & Lee, 2017).

$$\phi_i(p) = \sum_{S \subseteq N/i} \frac{|S|!(n-|S|-1)!}{n!} \big(p(S \cup i) - p(S)\big). \tag{3}$$

3. Results and discussion

3.1. Prediction of window opening status

3.1.1. Logistic regression models and Gaussian Naïve Bayes classifier

The comparison of LR models developed using different preprocessing techniques, namely standardisation, normalisation, robust scaling, and log transformation, was conducted based on the precision, recall, and F1 scores for the window open and closed status. Table 2 displays the model parameters and the evaluation metrics of the best-performing models with various machine learning algorithms.

Table 2. The model parameters and the evaluation metrics of the best-performing models with LR, DT, RF, SVM and GNB.

<table>
<tr><td></td><td></td><td colspan="2">LR</td><td colspan="2">GNB</td><td></td><td colspan="2">DT</td><td></td><td colspan="2">RF</td><td></td><td colspan="2">SVM</td></tr>
<tr><td></td><td>Preprocessing method</td><td>Normalisation</td><td>Log transformation</td><td>Log transformation</td><td>None</td><td></td><td>All methods</td><td>All except Log transformation</td><td></td><td colspan="2">All methods</td><td></td><td colspan="2">Standardisation (Z-score normalisation)</td></tr>
<tr><td></td><td>Prediction task</td><td>Status</td><td>Operation</td><td>Status</td><td>Operation</td><td></td><td>Status</td><td>Operation</td><td></td><td>Status</td><td>Operation</td><td></td><td>Status</td><td>Operation</td></tr>
<tr><td rowspan="11">Model parameters</td><td>Intercept</td><td>0.071</td><td rowspan="11">See in annotation *</td><td rowspan="11" colspan="2"></td><td rowspan="11">Maximum depth</td><td rowspan="11">5</td><td rowspan="11">15</td><td rowspan="11">Number of estimators</td><td rowspan="11">100</td><td rowspan="11">10</td><td rowspan="11"></td><td rowspan="11">RBF</td><td rowspan="11">degree-5 polynomial</td></tr>
<tr><td>Temperature</td><td>-0.125</td></tr>
<tr><td>Rainfall</td><td>-0.104</td></tr>
<tr><td>Relative humidity</td><td>0.022</td></tr>
<tr><td>Wind speed</td><td>-0.047</td></tr>
<tr><td>Solar radiation</td><td>0.001</td></tr>
<tr><td>Air quality health index (AQHI)</td><td>0.369</td></tr>
<tr><td>Hour_sin</td><td>1.133</td></tr>
<tr><td>Hour_cos</td><td>-0.052</td></tr>
<tr><td>Season</td><td>0.009</td></tr>
<tr><td>Weekday/Weekend</td><td>0.019</td></tr>
</table>

Precision														
Close	Close from its open position	0.51	0.01	0.65	0.01		0.56	0.00		0.54	0.00		0.60	0.02
	Close to switch on the air conditioner		0.03		0.03			0.05			0.05			0.04
	No action		0.98		0.97			0.96			0.96			0.97
Open	Open from its closed position	0.78	0.04	0.74	0.04		0.81	0.11		0.79	0.20		0.80	0.06
	Adjust to a partially closed position		0.01		0.01			0.33			0.00			0.02
	Adjust to a partially open position		0.00		0.00			0.00			0.00			0.01
Weighted average		0.69	0.94	0.71	0.93		0.71	0.92		0.70	0.92		0.73	0.93
Recall														
Close	Close from its open position	0.63	0.20	0.39	0.27		0.66	0.00		0.61	0.00		0.61	0.27
	Close to switch on the air conditioner		0.60		0.40			0.15			0.10			0.40
	No action		0.14		0.08			0.94			0.96			0.65
Open	Open from its closed position	0.69	0.44	0.89	0.50		0.74	0.17		0.74	0.17		0.79	0.36
	Adjust to a partially closed position		0.17		0.58			0.08			0.00			0.17
	Adjust to a partially open position		0.08		0.00			0.00			0.00			0.08
Weighted average		0.67	0.15	0.72	0.09		0.71	0.90		0.69	0.92		0.73	0.63

F1 score														
Close	Close from its open position	0.56	0.01	0.49	0.02		0.61	0.00		0.57	0.00		0.60	0.04
	Close to switch on the air conditioner		0.05		0.05			0.07			0.06			0.07
	No action		0.25		0.14			0.95			0.96			0.78
Open	Open from its closed position	0.73	0.08	0.81	0.07		0.77	0.13		0.76	0.18		0.80	0.11
	Adjust to a partially closed position		0.01		0.02			0.13			0.00			0.03
	Adjust to a partially open position		0.01		0.00			0.00			0.00			0.02
Weighted average		0.67	0.24	0.70	0.14		0.72	0.91		0.70	0.92		0.73	0.75
Accuracy														
		Logistic regression		Gaussian Naïve Bayes			Decision tree			Random forest			Support vector machine	
Overall		0.67	0.15	0.72	0.09		0.71	0.90		0.69	0.92		0.73	0.63

score(C) = 20.2792511790 + (-4.5040178776 * Temperature) + (0.4915925831 * Rainfall) + (-1.4633752741 * Relative humidity) + (-0.3716929975 * Wind speed) + (-0.0527029445 * Solar radiation) + (0.5911421873 * Air quality health index (AQHI)) + (1.7716980574 * Hour_sin) + (1.5814022589 * Hour_cos) + (0.1648011684 * Season) + (-0.6423818095 * Weekday/Weekend)

score(CA) = -64.4508286296 + (12.5713668024 * Temperature) + (-1.6567103888 * Rainfall) + (5.8638265733 * Relative humidity) + (0.2191085934 * Wind speed) + (-0.1331779612 * Solar radiation) + (-0.3396210558 * Air quality health index (AQHI)) + (-4.7469063770 * Hour_sin) + (-0.3720033270 * Hour_cos) + (-0.8632109477 * Season) + (1.0462888806 * Weekday/Weekend)

score(N) = 17.6905941357 + (-1.7120929288 * Temperature) + (0.5159857661 * Rainfall) + (-2.5266565082 * Relative humidity) + (-0.1769629650 * Wind speed) + (-0.1730391504 * Solar radiation) + (0.3804230028 * Air quality health index (AQHI)) + (0.6470118734 * Hour_sin) + (-1.3090052907 * Hour_cos) + (0.0053693577 * Season) + (0.1056022972 * Weekday/Weekend)

score(O) = 7.6904961416 + (0.3782657527 * Temperature) + (0.3238432709 * Rainfall) + (-1.1483083554 * Relative humidity) + (-0.3137355015 * Wind speed) + (0.0697429832 * Solar radiation) + (-1.3932709154 * Air quality health index (AQHI)) + (0.8212647514 * Hour_sin) + (-2.0037234546 * Hour_cos) + (-0.1494161745 * Season) + (-1.0489241870 * Weekday/Weekend)

score(OL) = 12.5518880858 + (-2.7288237306 * Temperature) + (-1.0474646029 * Rainfall) + (-0.7610636731 * Relative humidity) + (0.6103593125 * Wind speed) + (0.0554202238 * Solar radiation) + (-1.3706279644 * Air quality health index (AQHI)) + (-2.0050587344 * Hour_sin) + (-0.2456854595 * Hour_cos) + (0.3605653154 * Season) + (0.4915851836 * Weekday/Weekend)

score(OM) = 6.2385990876 + (-4.0046980181 * Temperature) + (1.3727533716 * Rainfall) + (0.0355772375 * Relative humidity) + (0.0329235581 * Wind speed) + (0.2337568491 * Solar radiation) + (2.1319547454 * Air quality health index (AQHI)) + (3.5119904292 * Hour_sin) + (2.3490152727 * Hour_cos) + (0.4818912806 * Season) + (0.0478296351 * Weekday/Weekend)*

Negative coefficients were observed for "temperature", "rainfall" and "wind speed" in the LR models. These negative coefficients indicated an inverse relationship between these features and the predicted outcome, suggesting that the likelihood of the window being open decreases as temperature, rainfall and wind speed increase.

On the contrary, positive coefficients were obtained for "relative humidity" in the LR models, indicating a positive relationship between humidity and the window opening status.

Upon examining the evaluation metrics of the LR models, it can be observed that while the regression coefficients varied across the different preprocessing methods, the overall performance of the models remained relatively consistent. The models achieved weighted F1 score averages of 0.67 and accuracies ranging from 0.66 to 0.67, indicating a reasonable balance between precision and recall and a relatively high degree of overall accuracy in the predictions.

In contrast to the LR algorithm, GNB assumes that the features are conditionally independent given the class labels, following a Gaussian distribution. When predicting the window opening status, these classifiers demonstrated nearly identical performance across various preprocessing methods. Their predictive capabilities were also comparable to those of the LR model, with weighted average F1 scores of 0.70 and overall accuracies ranging from 0.71 to 0.72.

Overall, given their similar predictive performance, GNB appears to be a simpler and computationally more efficient choice than LR models regarding understanding and modelling the window opening status.

3.1.2. Tree-based prediction models

Model evaluation metrics demonstrate that the DT and RF models employed exhibit minimal sensitivity to different preprocessing methods applied to the numerical data, highlighting the robustness of tree-based algorithms in handling various types of input features.

Regarding classification performance, both DT and RF models showcased a superior ability to classify open windows compared to closed ones. This outcome can be attributed to the higher occurrence of open window incidents within the training data, which enabled the models to capture and learn patterns representative of open window behaviour.

Specifically, a DT model with a maximum depth of 5 outperformed deeper tree structures, achieving a weighted average precision, recall, F1 score, and an overall accuracy of 0.73, 0.71, 0.72, and 0.71, respectively. These metrics indicate that a shallower tree structure proved sufficient for accurately predicting a binary window opening status classification problem. On the other hand, the RF models with 10, 50, and 100 estimators yielded comparable performance, with the best weighted average precision, recall, F1 score, and an overall accuracy of 0.70, 0.69, 0.70, and 0.69. These results suggest that increasing the estimators beyond a certain threshold did not significantly improve the predictive performance.

The findings imply that the tree-based models' relatively simple, straightforward structures with moderate accuracy and robustness could offer an optimised trade-off regarding model complexity and computational efficiency for predicting the window opening status.

3.1.3. Support vector machine

Linear SVM models performed similarly across the different transformations, while non-linear models employing RBF kernel and degree-3, 4, and 5 polynomial kernels exhibited slightly superior performance with standardisation and log transformation.

Non-linear SVM models, which utilise kernel functions to map input data into higher-dimensional feature spaces for linear separation, demonstrated slightly better predictive capabilities than linear models. The average F1 scores and accuracies for the non-linear models were 0.65 and 0.66, compared with the linear model's scores of both 0.35 and 0.42, respectively. Linear SVM models assume linear separability and identify the optimal hyperplane for separating data of different classes. In contrast, kernels in non-linear models effectively captured intricate non-linear relationships among input features, improving the accuracy when predicting window opening status (Chauhan et al., 2019).

In contrast, models employing sigmoid kernels exhibited poor performance when predicting window opening status. Most of these models failed to indicate samples from the minority class during testing, resulting in ill-defined precision and F1 scores. Some models with gamma values of 0.1 were trainable and testable, but their performance remained poor, yielding weighted average F1 scores ranging from 0.17 to 0.60 and overall accuracies ranging from 0.34 to 0.63. The findings indicate their limitations within this specific prediction task.

3.1.4. Model comparison

Among the evaluated window opening status prediction models, a SVM with an RBF kernel and Z-score normalisation preprocessing achieved the highest average F1 score (0.73), followed by DT (F1 = 0.72), GNB (F1 = 0.72), RF (F1 = 0.70) and LR (F1 = 0.67). For comparison, Furuhashi et al. (2022) reported F1 scores of 0.77, 0.78, 0.81, and 0.79 for SVM, DT, RF, and LR, respectively. Similarly, these findings indicate that a range of machine learning algorithms can achieve comparably strong predictive performance for window opening status classification.

3.1.5. Feature importance analysis

PIMP was employed to assess the feature importance of the RBF SVM model. The results are presented in Table 3. PIMP values quantify the extent of the average change in this SVM model when a particular feature is randomly shuffled, repeated 20 times. The ranking shows that "Temperature" and "Season" emerged as the most important features based on PIMP analysis. "Relative humidity" and "Solar radiation" were located in a middle tier, but their effects are much smaller than "Temperature" and "Season". Besides, "AQHI", "Rainfall", "Time-of-day", "Wind speed" and "Weekday/Weekend" have limited or negligible influence. Overall, window status appears to be governed primarily by seasonal and thermal factors, not by occupancy schedule proxies or short-term temporal patterns.

Table 3. The ranking of features of the RBF SVM model along with PIMP values.

Rank	Feature	PIMP _ mean	PIMP _ std
1	Season	0.108024	0.00842
2	Temperature	0.06356	0.006856
3	Relative humidity	0.015137	0.003605
4	Solar radiation	0.010242	0.002234
5	Air quality health index (AQHI)	0.005771	0.001674
6	Rainfall	0.003595	0.00162
7	Hour_cos	0.00215	0.001144
8	Wind speed	0.001925	0.002552
9	Hour_sin	0.00164	0.00158
10	Weekday/Weekend	0.000791	0.001095

3.2. Prediction of window operation behaviours

3.2.1. Logistic regression models and Gaussian Naïve Bayes classifier

Multinomial LR models for predicting window operation behaviours were developed and evaluated. It is noteworthy that despite consistent results being observed across different preprocessing techniques, the prediction performances of the multinomial LR models were found to be poor, as indicated by the weighted average precisions ranging from 0.93 to 0.94, recalls from 0.11 to 0.15, F1 scores from 0.16 to 0.24, and accuracies from 0.11 to 0.15.

Similar to the LR model, GNB exhibited poor performance in classifying the multiclass window operation behaviours. Under different data transformation methods, the average weighted F1 scores and overall accuracies ranged from 0.13 to 0.14 and 0.08 to 0.09, respectively. These results indicate that both models struggled to classify various window operation behaviours accurately.

3.2.2. Tree-based prediction models

The optimal hyperparameter settings differed between the two tree-based modelling approaches. A DT with deeper structures performed significantly better than those with simpler structures when dealing with a multiclass classification problem involving window operation behaviours. Conversely, RF models performed best with a relatively small ensemble size, where a configuration with lower estimators outperformed models with larger numbers of trees. The most effective configurations were therefore a DT with a maximum depth of 15 (F1 = 0.91) and an RF with 10 estimators (F1 = 0.92).

Overall, it is essential to highlight the consistency and robustness of tree-based models in addressing both binary window opening status and multiclass classification problems of window operation behaviours. However, it is also crucial to emphasise the significance of testing various optimal model parameters to achieve the best predictive performance while maintaining computational efficiency.

3.2.3. Support vector machine

The linear SVM model demonstrated slightly superior performance with Z-score normalisation as a preprocessing method to predict window operation behaviours. However, the classification performance for different actions was still considered poor. On the other hand, SVM models with RBF kernels exhibited improved performance when numerical data were standardised. The best SVM model for this task utilised a degree-5 polynomial kernel with log transformation, which achieved a weighted average precision of 0.93, recall of 0.63, F1 score of 0.75 and overall accuracy of 0.63.

SVM models employing sigmoid kernels could not effectively and accurately address this multiclass classification task. The trained models demonstrated significant inadequacy in predicting the window operation behaviours, with average F1 scores ranging from 0.00 to 0.65 and overall accuracy ranging from 0.01 to 0.51.

These findings highlight the impact and constraints of different approaches and kernel functions employed in SVM models when predicting window operation behaviours. Compared to alternative machine learning algorithms, SVM models exhibited a higher sensitivity to various preprocessing techniques for data manipulation. Considering non-linear relationships among input features, it is evident that RBF and polynomial kernels outperformed other options in comprehending and modelling the above tasks.

3.2.4. Model comparison

In terms of window operation behaviours, the performance of each model was inconsistent. RF exhibited the highest F1 score at 0.92, followed by the DT (F1 score = 0.90), SVM (F1 score = 0.75), LR (F1 score = 0.24) and GNB (F1 score = 0.14). These findings emphasize the critical role of model selection, as predictive effectiveness depends on both the structure of the input variables and the specific behavioural outcome being modelled. However, the dataset displayed a substantial class imbalance in the distribution of window operation behaviours. Consequently, the "No action" class achieved disproportionately high predictive performance relative to the other categories. In the best performing RF model, the F1 score for "No action" reached 0.96, compared with 0.18 for "Open from its closed position" and 0.06 for "Close to switch on the air conditioner", while the remaining three action categories yielded F1 scores of 0. This imbalance in classification performance underscores the critical influence of class distribution on model evaluation in multiclass behavioural prediction tasks.

3.2.5. Feature importance

Across all the window operation behaviour predictive models, RF with 10 estimators performed the best (F1 = 0.92). The IG, PIMP and SHAP values were employed to assess the feature importance of this RF model; the results are presented in Table 4 and Figure 1. Across the three indicators, time-related variables emerged as the dominant determinants of action prediction. Both IG and global mean absolute SHAP ranked "Hour_sin" first, "Hour_cos" second and "Temperature" third. The second tier was formed by "Solar radiation", "AQHI" and "RH", while "Rainfall" was consistently the weakest predictor. It seems that the window operation behaviour was mainly driven by the diurnal cycle, with thermal conditions providing the next most influential predictors.

Table 4. The ranking of features of the RF model along with their IG scores, PIMP values and SHAP values.

Rank	I_G		PIMP			SHAP	
				Mean	Std		
1	Hour_sin	0.168	Hour_sin	0.033	0.004	Hour_sin	0.067
2	Hour_cos	0.155	Hour_cos	0.033	0.005	Hour_cos	0.057
3	Temperature	0.146	Season	0.018	0.005	Temperature	0.052
4	Relative humidity	0.117	Solar radiation	0.007	0.010	Solar radiation	0.036
5	Solar radiation	0.109	Weekday/Weekend	0.004	0.007	Air quality health index (AQHI)	0.034
6	Wind speed	0.095	Temperature	0.003	0.009	Relative humidity	0.031
7	Air quality health index (AQHI)	0.094	Air quality health index (AQHI)	0.002	0.009	Season	0.023
8	Season	0.062	Rainfall	0.001	0.004	Wind speed	0.022
9	Weekday/Weekend	0.034	Wind speed	-0.015	0.010	Weekday/Weekend	0.013
10	Rainfall	0.020	Relative humidity	-0.015	0.012	Rainfall	0.008

The SHAP analysis reveals the relationship between feature values and the probability of various window operation behaviours. Figure 1 depicts the beeswarm plot, ranked by their average absolute SHAP magnitude within each class. Points to the right increase the model output for that class, whereas points to the left decrease it; red denotes relatively high feature values and blue denotes relatively low feature values.

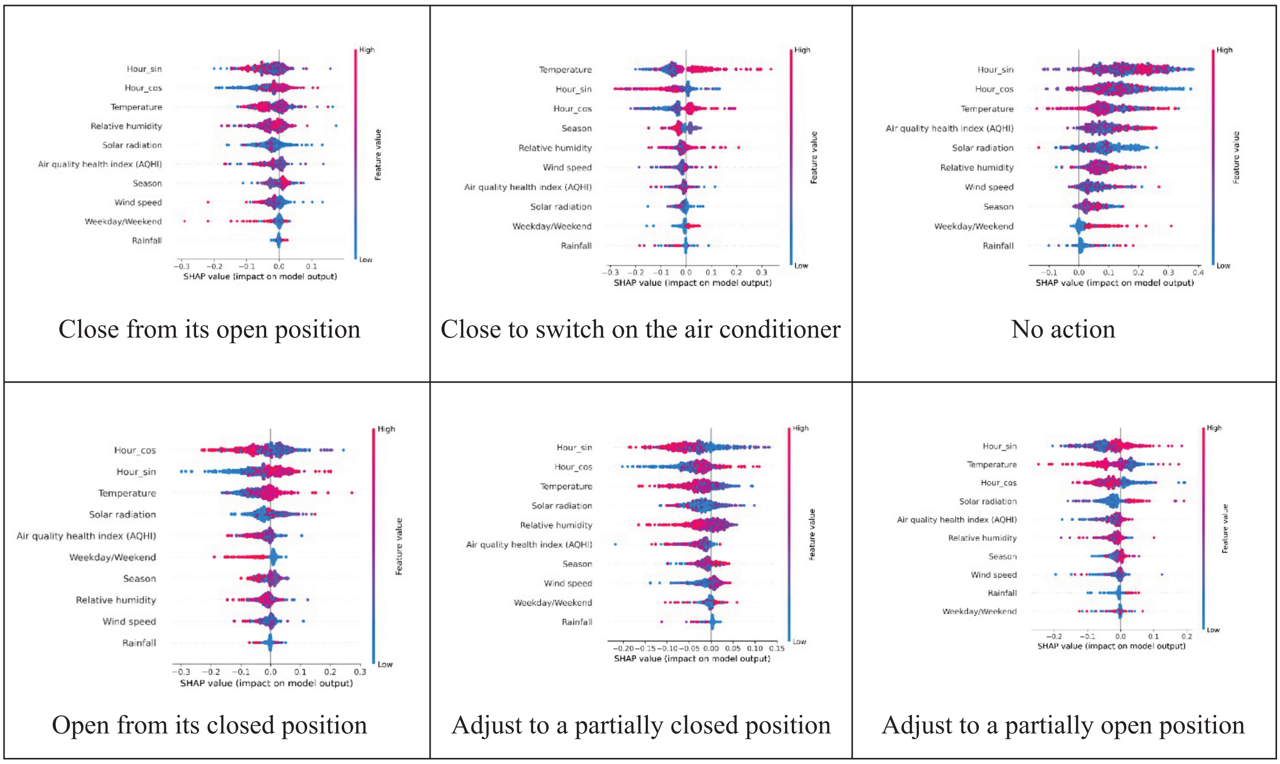

Figure 1. Beeswarm plot ranked of the RF model by mean absolute SHAP values.

The PIMP result showed that "Hour_sin" and "Hour_cos" produced the largest performance losses after shuffling, while "Season" ranked third. In contrast, "Temperature" exhibited only a lower PIMP level despite being highly ranked by both I_G and SHAP. This discrepancy may indicate that part of the "Temperature" predictive signal overlaps with time and seasonal information.

The action-specific SHAP results showed marked heterogeneity across different actions. The "No action" class displayed relatively broad SHAP spreads, indicating that "No action" occupied the most clearly separable regions of the predictor space. A further point is the consistently minor role of "Rainfall". It ranked last in both I_G and SHAP. This may suggest that, within this dataset and modelling framework, rainfall did not provide major incremental information. The same general argument applied to "Weekday/Weekend", which only showed moderate influence in a limited subset of action classes.

Overall, the results indicate that window operation behaviour is governed first by routine daily timing, second by thermal conditions and only thereafter by secondary environmental modifiers. The strongest and most stable signal in the model was the diurnal cycle, while rainfall and weekday status added little marginal predictive power.

3.3. Model application

In terms of application, the window opening status models have immediate value for building performance practice. Since the binary task achieved comparatively stable and strong predictive performance across several algorithms, with the best result obtained by the RBF SVM model, these models can be integrated into residential building energy simulation to provide more realistic estimates of whether windows are open or closed at a given time. In contrast, the window operation behaviour models are more relevant to event-based behavioural modelling, as they aim to capture the timing and type of window actions. However, because the multiclass dataset was strongly imbalanced and minority actions were predicted less reliably, these models are currently more appropriate for behavioural pattern analysis and simulation support. Additionally, the feature importance analysis results suggest that practical implementation may be feasible with relatively accessible inputs, since window status was mainly associated with temperature and season, while window operation behaviour was driven primarily by diurnal timing and temperature. These findings indicate that the proposed models have immediate value for improving occupant behaviour representation in residential building simulations and may also provide a useful foundation for adaptive ventilation control strategies.

4. Conclusion

Understanding window opening status and occupant behaviours has garnered significant research attention due to the potential for improving building performance while reducing the reliance on mechanical methods for thermal comfort and air quality improvement. Machine learning models have been developed to model window opening behaviour in residential buildings in recent years. However, limited efforts have been made to evaluate the importance of model features and enhancing model performance through hyperparameter optimisation. This study aimed to bridge this gap by investigating the environmental drivers influencing window status and opening behaviours in apartment bedrooms in a sub-tropical region. Using a comprehensive database comprising hourly environmental observations on window status, window operation behaviours, and ambient environmental conditions collected over a year in Hong Kong, mathematical methods such as I_G, PIMP, and SHAP were employed to explore the significance of factors in determining window opening status.

Regarding predictive models, stochastic behavioural models were developed using various algorithms, including LR, GNB, tree-based prediction models and SVMs. The results demonstrate that while algorithms such as LR and GNB effectively predict window opening status, they struggle to classify various window operation behaviours accurately. The SVM model performed best in window status prediction but only showed a moderate accuracy in window operation behaviour prediction. On the other hand, tree-based models exhibited a good balance between model complexity and computational efficiency. These models provided moderate accuracy and robustness in predicting both window opening status and window operation behaviours. However, the presence of class imbalance in the dataset adversely affected the predictive performance for minority classes across all models, underscoring the importance of balanced survey design and data collection strategies.

The feature importance analysis revealed that "Temperature" and "Season" were the most critical features in determining window status. Additionally, the diurnal cycle was identified as the dominant predictor of window operation behaviour, while temperature represented the second most influential driver.

In terms of generalisation, this study suggests that the relative importance of thermal seasonal drivers for window opening status and diurnal timing for window operation behaviour may extend to similar residential environments in sub-tropical climates. Nevertheless, because the models were developed using a single regional dataset, extrapolation to other climates, dwelling types, and occupant populations should be treated with caution. The status models can be used to improve window status assumptions in building energy simulations, while the action models offer value in identifying likely transition periods and informing behaviour-aware control logic.

Overall, this study goes beyond simple model benchmarking by jointly comparing multiple algorithms, testing preprocessing and hyperparameter choices to reveal both predictive performance and the underlying environmental drivers of residential window status and behaviour.

Acknowledgement

This work was jointly supported by PolyU Internal funding (project nos. WZ2N, WZ3R, 1-BDWZ, and WZ9M).

Notes on contributors

Dr Tsz Wun Tsang is a Research Assistant Professor in the Department of Building Environment and Energy Engineering at The Hong Kong Polytechnic University. Her research interests focus on indoor environmental quality, thermal comfort, indoor air quality, indoor sensing and IoT, indoor bioaerosol transmission, building energy conservation, and carbon neutrality in the built environment.

Mr Bian Liu is a Research Assistant in the Department of Building Environment and Energy Engineering. He holds a B.Sc. in Civil Engineering from Shenzhen University, China, and an M.Sc. in Environmental Design of Buildings from the Welsh School of Architecture at Cardiff University, UK. Currently, he is pursuing a Ph.D. with a focus on indoor environmental quality.

Dr. Ling Tim Wong is an Associate Professor in the Department of Building Environment and Energy Engineering at The Hong Kong Polytechnic University. Dr. Wong is a Chartered Engineer and a member of HKIE and CIBSE. He also served as a court member during 2016-2021 and a council member during 2022-2023 for Hong Kong Baptist University.

References

[1] Aloisio, A., Rosso, M.M., Di Battista, L., Quaranta, G., (2024). Machine-learning-aided regional post-seismic usability prediction of buildings: 2016–2017 Central Italy earthquakes. Journal of Building Engineering, 91, 109526.

[2] Altmann, A., Toloşi, L., Sander, O., & Lengauer, T. (2010). Permutation importance: a corrected feature importance measure. Bioinformatics, 26(10), 1340-1347.

[3] Berrar, D. (2018). Bayes' theorem and naive Bayes classifier. Encyclopedia of bioinformatics and computational biology: ABC of bioinformatics, 403, 412.

[4] Brager, G. S., & de Dear, R. J. (1998). Thermal adaptation in the built environment: a literature review. Energy and Buildings, 27(1), 83-96. https://doi.org/Doi 10.1016/S0378-7788(97)00053-4

[5] Breiman, L. (2001). Random forests. Machine learning, 45, 5-32.

[6] Chauhan, V. K., Dahiya, K., & Sharma, A. (2019). Problem formulations and solvers in linear SVM: a review. Artificial Intelligence Review, 52(2), 803-855.

[7] Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: synthetic minority over-sampling technique. Journal of artificial intelligence research, 16, 321-357.

[8] Cortes, C., & Vapnik, V. (1995). Support-vector networks. Machine learning, 20, 273-297.

[9] Furuhashi, K., Nakaya, T., & Maeda, Y. (2022). Prediction of Occupant Behavior toward Natural Ventilation in Japanese Dwellings: Machine Learning Models and Feature Selection. Energies, 15(16).

[10] Han, M. J., May, R., Zhang, X. X., Wang, X. R., Pan, S., Yan, D., & Jin, Y. (2020). A novel reinforcement learning method for improving occupant comfort via window opening and closing. Sustainable Cities and Society, 61.

[11] Hawila, A. A. W., Diallo, T. M., & Collignan, B. (2023). Occupants' window opening behavior in office building: A review of influencing factors, modeling approaches and verification of literature models. Building and Environment, 110525.

[12] He, H., & Garcia, E. A. (2009). Learning from imbalanced data. IEEE Transactions on knowledge and data engineering, 21(9), 1263-1284.

[13] Hosmer Jr, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). Applied logistic regression (Vol. 398). John Wiley & Sons.

[14] Jeong, B., Jeong, J. W., & Park, J. S. (2016). Occupant behavior regarding the manual control of windows in residential buildings. Energy and buildings, 127, 206-216.

[15] Jia, M., Srinivasan, R. S., & Raheem, A. A. (2017). From occupancy to occupant behavior: An analytical survey of data acquisition technologies, modeling methodologies and simulation coupling mechanisms for building energy efficiency. Renewable and Sustainable Energy Reviews, 68, 525-540.

[16] Jones, R. V., Fuertes, A., Gregori, E., & Giretti, A. (2017). Stochastic behavioural models of occupants' main bedroom window operation for UK residential buildings. Building and Environment, 118, 144-158.

[17] Kc, R., Rijal, H. B., Shukuya, M., & Yoshida, K. (2018). An in-situ study on occupants' behaviors for adaptive thermal comfort in a Japanese HEMS condominium. Journal of Building Engineering, 19, 402-411.

[18] Kim, J., Hong, T., Jeong, J., Koo, C., & Kong, M. (2017). An integrated psychological response score of the occupants based on their activities and the indoor environmental quality condition changes. Building and Environment, 123, 66-77.

[19] Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. Advances in neural information processing systems, 30.

[20] Luo, M. H., Wang, Z., Ke, K., Cao, B., Zhai, Y. C., & Zhou, X. (2018). Human metabolic rate and thermal comfort in buildings: The problem and challenge. Building and Environment, 131, 44-52.

[21] Markovic, R., Grintal, E., Wölki, D., Frisch, J., & van Treeck, C. (2018). Window opening model using deep learning methods. Building and Environment, 145, 319-329.

[22] Menze, B. H., Kelm, B. M., Masuch, R., Himmelreich, U., Bachert, P., Petrich, W., & Hamprecht, F. A. (2009). A comparison of random forest and its Gini importance with standard chemometric methods for the feature selection and classification of spectral data. BMC bioinformatics, 10, 1-16.

[23] Mo, H., Sun, H. J., Liu, J. J., & Wei, S. (2019). Developing window behavior models for residential buildings using XGBoost algorithm. Energy and Buildings, 205.

[24] Myles, A. J., Feudale, R. N., Liu, Y., Woody, N. A., & Brown, S. D. (2004). An introduction to decision tree modeling. Journal of Chemometrics: A Journal of the Chemometrics Society, 18(6), 275-285.

[25] Nicol, J. F., & Humphreys, M. A. (2002). Adaptive thermal comfort and sustainable thermal standards for buildings. Energy and Buildings, 34(6), 563-572.

[26] Pan, S., Han, Y., Wei, S., Wei, Y., Xia, L., Xie, L., ... & Yu, W. (2019). A model based on Gauss Distribution for predicting window behavior in building. Building and Environment, 149, 210-219.

[27] Park, J., Jeong, B., Chae, Y. T., & Jeong, J. W. (2021). Machine learning algorithms for predicting occupants' behaviour in the manual control of windows for cross-ventilation in homes. Indoor and Built Environment, 30(8), 1106-1123.

[28] Rijal, H. B. (2021). Thermal adaptation of buildings and people for energy saving in extreme cold climate of Nepal. Energy and Buildings, 230.

[29] Rijal, H. B., Humphreys, M. A., & Nicol, J. F. (2018). Development of a window opening algorithm based on adaptive thermal comfort to predict occupant behavior in Japanese dwellings. Japan Architectural Review, 1(3), 310-321.

[30] Rijal, H. B., Tuohy, P., Humphreys, M. A., Nicol, J. F., Samuel, A., & Clarke, J. (2007). Using results from field surveys to predict the effect of open windows on thermal comfort and energy use in buildings. Energy and Buildings, 39(7), 823-836. https://doi.org/10.1016/j.enbuild.2007.02.003

[31] Rouleau, J., & Gosselin, L. (2020). Probabilistic window opening model considering occupant behavior diversity: A data-driven case study of Canadian residential buildings. Energy, 195, 116981.

[32] Strobl, C., Boulesteix, A. L., & Zeileis, A. Hothorn, és T. (2007). Bias in random forest variable importance measures: illustrations, sources and a solution. BMC Bioinformatics, 8, 25.

[33] Tsang, T. W., Wong, L. T., Lung, H. T., & Mui, K. W. (2024). Stochastic behavioral models of bedroom window operation in sub-tropical residential buildings. Building and Environment, 262, 111784.

[34] Wei, Y., Yu, H., Pan, S., Xia, L., Xie, J., Wang, X., ... & Li, Q. (2019). Comparison of different window behavior modeling approaches during transition season in Beijing, China. Building and Environment, 157, 1-15.

[35] Yang, X., Liu, J., Meng, Q., Wei, Y., Lei, Y., Wu, M., ... & Lian, Y. (2022). Window Opening Behavior of Residential Buildings during the Transitional Season in China's Xi'an. Complexity, 2022.

[36] Yao, M., & Zhao, B. (2017). Window opening behavior of occupants in residential buildings in Beijing. Building and Environment, 124, 441-449.

[37] Zaki, S. A., Damiati, S. A., Rijal, H. B., Hagishima, A., & Abd Razak, A. (2017). Adaptive thermal comfort in university classrooms in Malaysia and Japan. Building and Environment, 122, 294-306.

[38] Zelaya, C. V. G. (2019, April). Towards explaining the effects of data preprocessing on machine learning. In 2019 IEEE 35th international conference on data engineering (ICDE) (pp. 2086-2090). IEEE.