Prediction of Thermal Comfort in Nighttime Metropolises Based on Multiple Machine Learning Models and Social Media Data

Jun Yang[1], Mengting Ge[2], Shaojiang Zhong[3], Mintai Kim[2]

[1]Virginia Tech, Virginia/USA · juny@vt.edu
[2]Virginia Tech, Virginia/USA
[3]Purdue University, Indiana/USA

Abstract: As metropolitan areas evolve into 24-hour hubs, ensuring nighttime thermal comfort has become critical for enhancing urban livability and public space usability. This study explores the use of machine learning (ML) models and social media data to predict thermal comfort in nighttime metropolitan environments. Using a 5-point thermal sensation vote (TSV) scale, over 70,000 social media posts from five diverse U.S. metropolises were analyzed. Use the LLaVA large language model for data cleaning, incorporating meteorological data retrieved via the OpenWeather API. Multiple classification and regression ML algorithms were tested for tasks. The Random Forest models demonstrated the highest performance. Social media user data can improve the accuracy of ML model predictions to a certain extent. The study also shows the ranking of various features' importance in the thermal comfort ML prediction model. The findings underscore the importance of integrating demographic and environmental data to enhance prediction accuracy and discuss the role of urban greenery in mitigating urban heat island effects. The research can provide actionable insights for urban planners, architects, and policymakers in designing thermally inclusive public spaces, promoting sustainability, and enhancing nighttime urban experiences.

Keywords: Thermal comfort, machine learning, social media data, nighttime metropolis, urban planning

1 Introduction

As metropolitan areas evolve into 24-hour hubs, the need for comfortable nighttime outdoor spaces has become critical. Urban environments support various activities at night, from social gatherings and cultural events to physical exercise and leisure activities, all of which contribute to the vitality of cities and the quality of life. Outdoor thermal comfort is crucial to ensure these spaces remain accessible and attractive, especially as more and more residents seek healthy, socially engaged, and sustainable lifestyles (CARMONA 2021, MEHTA 2013). The need for safe, comfortable nighttime environments in metropolitan areas reflects broader societal trends toward inclusive, all-weather urban planning that encourages individual and community activity (GEHL et al. 2011, JACOBS 1961).

Thermal comfort plays a central role in determining how much time people spend outdoors, how they interact with public spaces, and whether these interactions promote positive, safe experiences (HÖPPE 2002). Thermal comfort in urban areas is affected by unique environmental conditions at night, such as the urban heat island (UHI) effect, which results in temperature differences between built-up metropolitan areas and surrounding rural areas (OKE 1982). This effect can significantly affect people's perception of outdoor spaces, especially when temperatures are too high or too low during certain seasons. Urban environments are particularly prone to heat retention, leading to persistent discomfort due to dense building materials, reduced vegetation, and large areas of artificial surfaces (GAITANI et al. 2007,

Journal of Digital Landscape Architecture, 10-2025, pp. 432-442. © Wichmann Verlag, VDE VERLAG GMBH · Berlin · Offenbach. ISBN 978-3-87907-754-0, ISSN 2367-4253, e-ISSN 2511-624X, doi:10.14627/537754041.

ROTH et al. 1989). The demand for accurate nighttime thermal comfort prediction in metropolitan areas is increasingly important, not only for residents but also for tourists unfamiliar with local climate variations. Tourists often rely on weather forecasts and environmental information when planning their travel and outdoor activities, yet standard forecasts may lack the granularity needed to address nighttime comfort in dense urban settings. For individuals unfamiliar with a city's unique microclimates, understanding localized nighttime comfort levels is essential for making informed choices about where to spend time outdoors.

Current literature highlights the significance of thermal comfort in determining the quality and usability of outdoor spaces, with studies indicating that even mild discomfort can reduce the amount of time people are willing to spend in an area (SPAGNOLO & DE DEAR 2003). For example, research by Nikolopoulou and Steemers highlights that thermal comfort is one of the most influential factors in the use of outdoor spaces, especially in urban environments where reduced artificial lighting and natural shade can alter environmental perception (NIKOLOPOULOU & STEEMERS 2003). Furthermore, studies have shown that nighttime thermal comfort requires a dedicated approach, as most existing models are based on daytime conditions and do not consider the diverse microclimates or different building structures found in urban spaces at nighttime (AGHAMOLAEI et al. 2023, YIN et al. 2021).

Despite recognizing the importance of thermal comfort for nocturnal outdoor activities in urban environments, there are still significant gaps in our understanding of this phenomenon. Most existing studies focus on daytime conditions or natural environments, neglecting the unique variables of nocturnal metropolitan environments, such as altered air circulation, urban heat retention, and reduced solar radiation, which have a significant impact on thermal comfort (GAITANI et al. 2007, OKE 1982). Most current studies rely on meteorological data and controlled measurements, which may not capture the subjective real-time experience of urban residents. While some studies include public perception, they usually depend on post-event surveys or static data sources, lacking the immediacy and nuance of social media insights, which can reflect public perception of thermal comfort in real time, especially for spontaneous nocturnal activities (SLOAN et al. 2015). Existing thermal comfort models are usually designed for indoor or semi-outdoor spaces and have difficulty coping with the complex and changing conditions of nocturnal cities, where dense buildings, diverse microclimates, and diverse ground materials produce unpredictable thermal dynamics (NIKOLOPOULOU & STEEMERS 2003). Addressing these research gaps requires not only adapting models to account for nighttime urban conditions but also incorporating real-time subjective data from social media to provide more accurate thermal comfort predictions for nighttime metropolises.

This study aims to bridge these gaps by investigating thermal comfort in nighttime metropolitan areas. By integrating machine learning (ML) models with social media data, the research intends to develop a model to predict thermal comfort levels based on objective climatic factors and subjective public feedback. This machine learning model is specifically designed to predict thermal comfort in the unique context of urban nighttime environments. The study addresses the following research questions. Which algorithm performs best for thermal comfort prediction among the algorithms tested? Can social media user data enhance the predictive accuracy of machine learning models for nighttime thermal comfort? What specific features impact nighttime thermal comfort prediction?

2 Methodology

To achieve the research aims, the researchers considered six aspects of developing the ML prediction model. They involve data collection, data processing, algorithms, training, validation, and evaluation. The structure diagram shows the overall research design (Figure 1), and each aspect will be described in detail below.

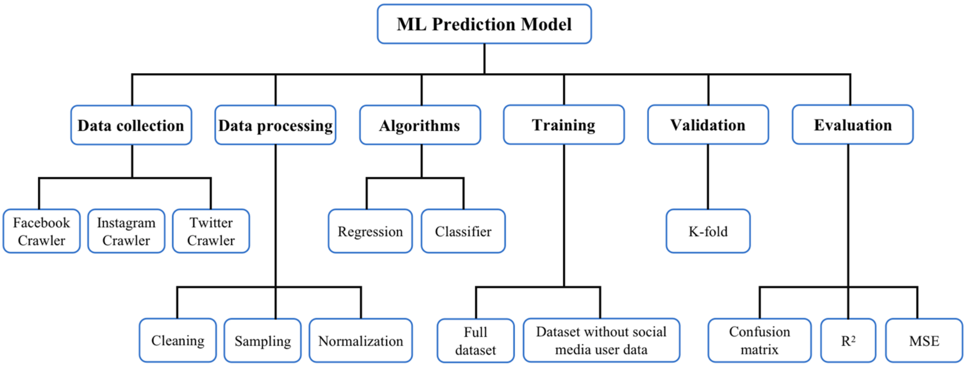


Fig. 1: Overall research design

2.1 Data Collection

Thermal sensation vote (TSV) scales are widely used in thermal comfort studies to evaluate human thermal perception (KUMAR & SHARMA 2020). These scales include the 3-point, 5-point, and 7-point options (CHEN & NG 2012), each designed to capture varying levels of thermal comfort perception. Considering the characteristics of social media data, we adopted a 5-point scale (-2 cold, -1 cool, 0 neutral, 1 warm, 2 hot) for this study. The scale is represented by thermal sensation keywords "cold", "cool", "suitable", "warm", and "hot", which guided the data collection process through the crawler.

We applied filters based on timestamps associated with social media posts to ensure that the collected data reflected nighttime thermal comfort. Posts made during the night (18:00 - 6:00) were included. Additionally, we used contextual filters to focus on phrases or content explicitly describing nighttime conditions. To ensure that the terms 'cold,' 'cool,' 'suitable,' 'warm,' and 'hot' were used in the context of thermal sensations, we applied contextual filters during the data collection process. Specifically, we focused on phrases or content that explicitly described nighttime conditions and thermal comfort. For example, posts that mentioned 'cold night' or 'cool breeze' were included, while posts that used 'cold' or 'cool' in unrelated contexts (e. g., 'cool party') were excluded. This filtering process helped ensure that the collected data accurately reflected users' perceptions of thermal comfort.

Given the metropolitan focus of our study, we aimed to cover diverse urban climates and socioeconomic conditions. Based on population rankings and climatic diversity, we selected five major U.S. metropolitan areas as study sites: New York, Chicago, Los Angeles, Phoenix, and Houston. These cities represent a broad spectrum of climate zones, including temperate continental climate, Mediterranean climate, subtropical monsoon climate, and desert climate, providing a comprehensive dataset for analysis (KOTTEK et al. 2006).

Given the sensitive nature of social media data, we took several measures to ensure the privacy and security of individuals' data. All data collected were anonymized, and no personally identifiable information (PII) was retained. We used the LLaVA large language model to analyze user profile pictures and textual descriptions, but the output was limited to estimated gender and age, without any direct linkage to individual users. Additionally, we employed blockchain-based techniques for data integrity and anonymization to further protect user privacy (ZHANG et al. 2024). These measures ensured that the data used in this study complied with ethical standards and privacy regulations.

We used web crawlers to scrape more than 150,000 social media posts from Facebook, Instagram, and Twitter platforms based on keywords as raw data. We will further process these raw data to extract relevant thermal comfort information while removing noise and unrelated data.

2.2 Data Processing

The collected raw data underwent a thorough cleaning process to ensure accuracy and relevance to the study. We used the LLaVA large language model to analyze user profile pictures and textual descriptions of thermal sensations. The model output included the user's estimated gender and age and the corresponding thermal comfort rating on a 5-point scale. Weather data corresponding to the time and location of each post was retrieved using the OpenWeather API. The API provided meteorological variables such as temperature, weather, humidity, wind speed, pressure, wind direction, and cloud.

We applied stratified random sampling to maintain balance across the 5-point thermal sensation scale and prevent bias due to over-representation of certain categories or locations. This method ensured that each thermal sensation category and city was proportionately represented in the final dataset. Posts with incomplete information or extreme outliers in weather data were excluded to enhance the dataset's reliability.

We normalized the meteorological data using min-max normalization to ensure compatibility across different ML models. This method scaled each variable to a range between 0 and 1 while preserving the relative differences among data points. Textual thermal sensation data and user gender and age were converted into one-hot encoded vectors, allowing seamless integration with numerical weather variables.

We prepared a high-quality, balanced dataset suitable for training and evaluating multiple machine learning models by cleaning, sampling, and normalizing the data. A total of 71,893 clean and reliable data points were retained for analysis. The dataset was divided into two groups: (1) Full dataset, which includes all available features, and (2) Reduced dataset, which excludes all user data while retaining meteorological data and thermal sensation labels. These two datasets will be used for training and testing in the following steps to evaluate the impact of user data on the performance of the thermal comfort prediction models.

2.3 Training and Validation

Each processed dataset was split into two subsets to train and evaluate the machine learning models: 80% for training and 20% for testing. This standard approach ensures that the models are trained on a majority portion of the data while reserving an independent set for performance evaluation. This split method is widely accepted as it balances the need for sufficient training data to build robust models and enough testing data to evaluate performance effectively. The training data was used to develop the models, while the testing data provided an unbiased assessment of how well the models generalize to unseen data.

To further validate the models and prevent overfitting, we employed k-fold cross-validation during the training process. In this method, the training data was divided into k subsets. The model was trained on k-1 folds and validated on the remaining fold, repeating the process k times so that each fold served as a validation set once. The results were then averaged to obtain a robust performance estimate.

2.4 Algorithms

To evaluate the performance of thermal comfort prediction, we tested several commonly used classifier and regression algorithms. The classifier algorithms included Adaptive Boosting (AB), Gradient Boosting Machine (GBM), K-Nearest Neighbors (KNN), Logistic Regression (LoR), Multilayer Perceptron (MLP), Naive Bayes (NB), Predicted Mean Vote (PMV), Random Forest (RF), Support Vector Machine (SVM). The regression algorithms included Decision Tree (DT), Linear Regression (LR), Random Forest (RF), and Support Vector Machine (SVM).

2.5 Model Evaluation

To assess the performance of the machine learning models, we employed evaluation metrics tailored to the classification and regression tasks. For classifier models, we utilized the confusion matrix as the primary evaluation tool. The confusion matrix provides detailed insights into the model's performance by comparing the predicted thermal sensation categories with the actual labels. For regression models, we used the R^2 (Coefficient of Determination), which measures the proportion of variance in the dependent variable that is predictable from the independent variables. An R^2 value closer to 1 indicates better model performance and a strong fit to the data. Mean Squared Error (MSE) calculates the average of the squared differences between predicted and actual values. A lower MSE indicates more accurate predictions and less variance in the model's errors. These metrics were calculated on both the testing dataset and the results of k-fold cross-validation to ensure robust evaluation. Figure 2 shows the entire research process.

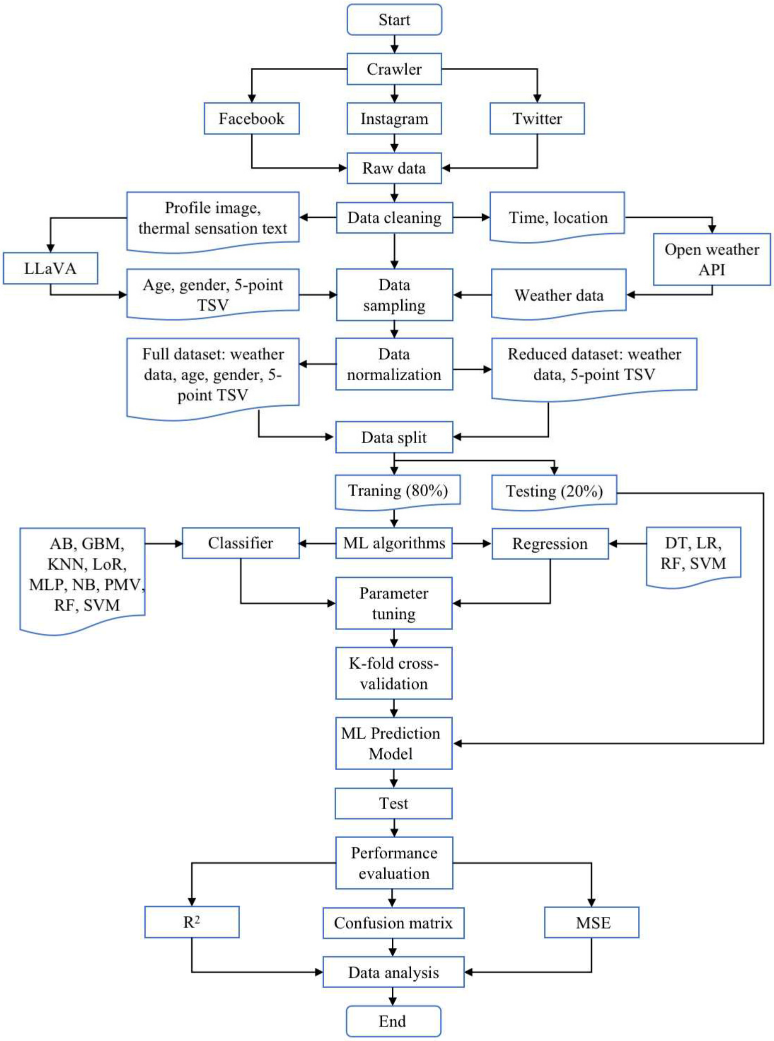


Fig. 2: Flowchart of the research

3 Results

3.1 Algorithm Performance

Figure 3 shows the accuracy of different ML regression models. Among the four commonly used algorithms, the RF regression model achieved the highest R^2 value (0.78) and the lowest MSE value (0.26), which means that it is more accurate than the other three algorithms. The second is the SVM regression model, with the R^2 value of 0.73 and the MSE value of 0.31. DT performed the worst, with an R^2 value of 0.58 and an MSE value of 0.49.

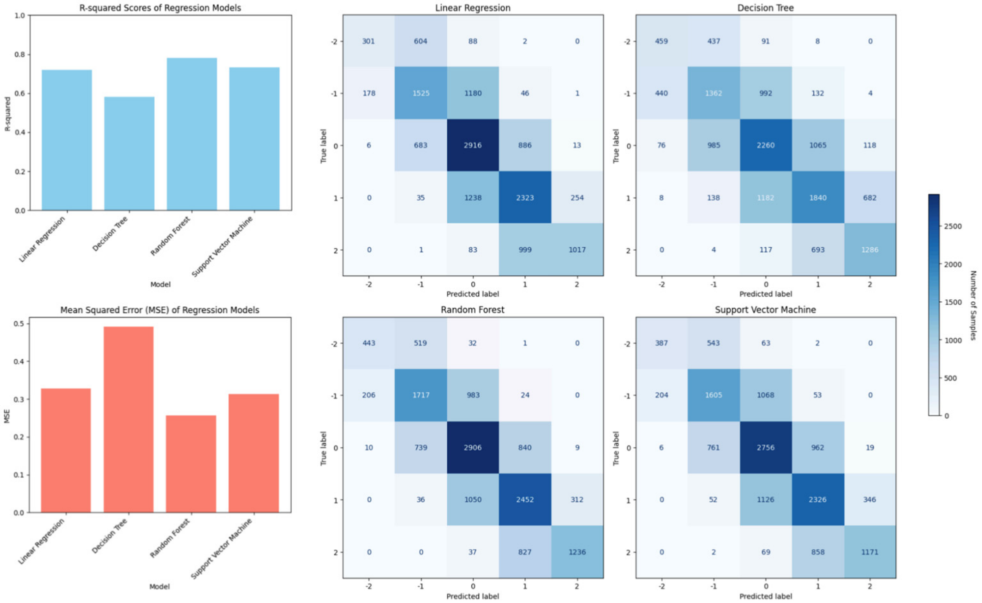


Fig. 3: The accuracy results of regression models

Figure 4 shows the accuracy of different ML classification models. Among the nine commonly used algorithms, three algorithms performed well. The RF classification model achieved the highest accuracy (0.6136). GB and MLP followed closely with accuracies of 0.6129 and 0.6041, respectively. PMV performed the worst with an accuracy of 0.3111.

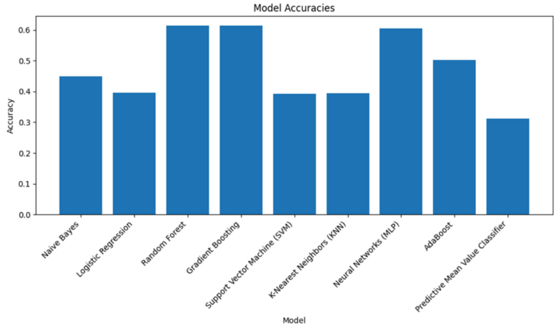
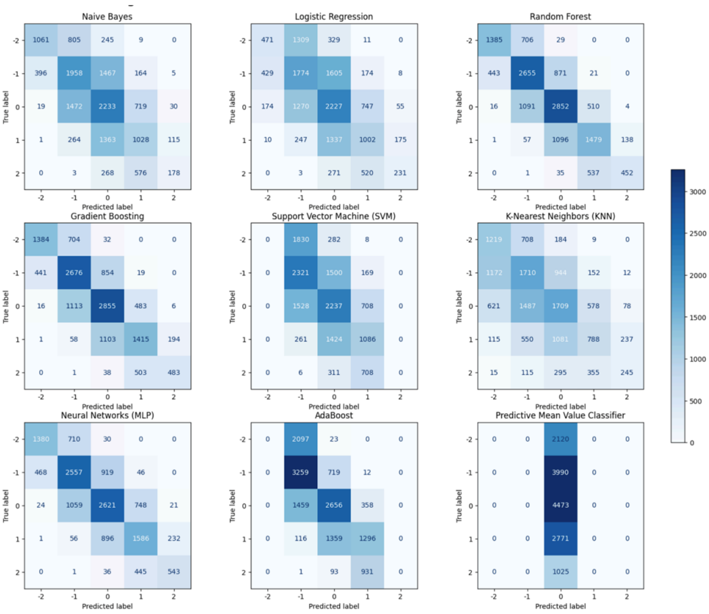


Fig. 4: The accuracy results of classifier models

From all the confusion matrix diagrams (Fig. 3, 5), we can see that the regression model has poor accuracy at the scale of [-2, -1) and good accuracy at the scale of [-1, 2]. This shows that the regression model's prediction ability for the cold part is weaker than that for other parts. The three classification models with better performance have good accuracy at the scale of [-2, 1] and poor accuracy at the scale of (1, 2]. This shows that the classification model's prediction ability for the hot part is weaker than that for other parts.



Naive Bayes

	-2	-1	0	1	2
-2	1061	805	245	9	0
-1	396	1958	1467	164	5
0	19	1472	2233	719	30
1	1	264	1363	1028	115
2	0	3	268	576	178

Logistic Regression

	-2	-1	0	1	2
-2	471	1309	329	11	0
-1	429	1774	1605	174	8
0	174	1270	2227	747	55
1	10	247	1337	1002	175
2	0	3	271	520	231

Random Forest

	-2	-1	0	1	2
-2	1385	706	29	0	0
-1	443	2655	871	21	0
0	16	1091	2852	510	4
1	1	57	1096	1479	138
2	0	1	35	537	452

Gradient Boosting

	-2	-1	0	1	2
-2	1384	704	32	0	0
-1	441	2676	854	19	0
0	16	1113	2855	483	6
1	1	58	1103	1415	194
2	0	1	38	503	483

Support Vector Machine (SVM)

	-2	-1	0	1	2
-2	0	1830	282	8	0
-1	0	2321	1500	169	0
0	0	1528	2237	708	0
1	0	261	1424	1086	0
2	0	6	311	708	0

K-Nearest Neighbors (KNN)

	-2	-1	0	1	2
-2	1219	708	184	9	0
-1	1172	1710	944	152	12
0	621	1487	1709	578	78
1	115	550	1081	788	237
2	15	115	295	355	245

Neural Networks (MLP)

	-2	-1	0	1	2
-2	1380	710	30	0	0
-1	468	2557	919	46	0
0	24	1059	2621	748	21
1	1	56	896	1586	232
2	0	1	36	445	543

AdaBoost

	-2	-1	0	1	2
-2	0	2097	23	0	0
-1	0	3259	719	12	0
0	0	1459	2656	358	0
1	0	116	1359	1296	0
2	0	1	93	931	0

Predictive Mean Value Classifier

	-2	-1	0	1	2
-2	0	0	2120	0	0
-1	0	0	3990	0	0
0	0	0	4473	0	0
1	0	0	2771	0	0
2	0	0	1025	0	0

Fig. 5: The confusion matrix diagrams of classifier models

3.2 Effect of Dataset

Figure 6 shows the accuracy of the full dataset and the reduced dataset in the four regression models. The MSE value of the reduced dataset has increased compared to the full dataset, ranging from 0.01 to 0.05, and the R^2 value has decreased, ranging from 0.01 to 0.04. This shows that the full dataset can achieve higher prediction accuracy than the reduced dataset in the same regression algorithm.

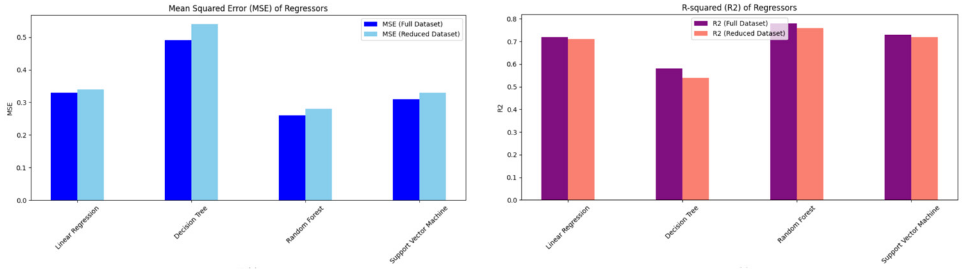


Fig. 6: Comparison of the accuracy of regression models on different datasets

3.3 Features Importance

We selected the RF algorithm with the highest accuracy for feature importance analysis. Figure 7 shows the importance of the ten features in the RF algorithm. The larger the importance coefficient, the more important the feature. Among them, temperature is the most important feature, with a coefficient of 0.241, followed by weather type (0.1389). Gender is the least important feature, with a coefficient of 0.019, followed by wind direction (0.05).

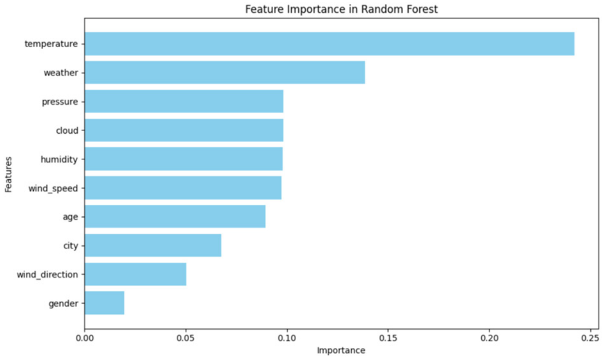


Fig. 7: Feature importance analysis in RF algorithm

4 Discussion

The results highlight the effectiveness of machine learning models in predicting nighttime thermal comfort. Among the tested ML models, the random forest regression model and the classification model showed the highest accuracy, which indicates that it has the potential to predict the thermal comfort index reliably and can robustly handle the complex relationship between urban meteorological variables and human thermal perception. However, the overall accuracy of the classic PMV model for 3-point TSV prediction is only 42.5% (LUO et al. 2020), which is lower than many well-performing ML algorithms tested. For further development of more reliable nighttime thermal comfort prediction models, it is possible to consider combining the better-performing parts of the regression model and the classification model to improve the accuracy of the prediction model.

The comparison between the full and reduced datasets underscores the importance of users' data in improving prediction accuracy. Although these variables showed lower individual importance coefficients, they collectively contributed to a more precise prediction model. Techniques for feature-specific enhancement inspired by heterogeneous memristive models may guide future refinements (ZHONG 2019). This finding underscores the value of integrating personal and environmental data to create customized thermal comfort solutions. For architects, planners, and urbanists, predicting thermal comfort in specific locations enables informed decision-making in designing inclusive public spaces. For example, understanding thermal comfort variations can help optimize the placement of shaded seating areas, water features, and cooling or heating infrastructure in urban parks.

Urban heat plays a critical role in shaping nighttime thermal comfort. The higher temperatures retained in urban areas due to the UHI effect can exacerbate nighttime discomfort, particularly during warmer seasons. Cities can significantly improve thermal comfort by mitigating UHI effects through strategies like urban greenery. For instance, trees and vegetation provide shade during the day and lower nighttime temperatures through evapotranspiration and reduced heat retention in built surfaces. Our study emphasizes the importance of thermal comfort prediction in integrating urban greenery planning to reduce the UHI. For example, planners can prioritize tree-lined streets, green roofs, and vertical gardens in areas identified by predictive models as having poor nighttime thermal comfort. These strategies can draw from concepts like dynamic system-based optimizations for urban comfort (XU et al. 2024).

These measures will enhance the quality of life and contribute to climate resilience, an increasingly critical consideration in urban planning. These strategies align with landscape architecture principles, where design interventions balance human comfort with environmental sustainability.

While our study provides valuable findings, there are several limitations. First, the reliance on social media data introduces potential biases, as the demographics of social media users may not fully represent the general population. Individual differences, such as age, sex, and physiological traits, can influence thermal perception. To improve the accuracy of thermal comfort prediction, future studies should consider these factors and psychological factors like mood and activity level. Second, this study's prediction model based on social media data is only based on the TSV 5-point scale. There is room for improvement in further developing a prediction model corresponding to the 7-point scale used in the thermal comfort TSV. Third, the study focuses on five U.S. cities, which limits its generalizability to other regions with different cultural, social, or climatic contexts. Future research should address these limitations by incorporating diverse data sources, such as field surveys or sensor-based thermal comfort measurements, to complement social media data. Additionally, expanding the study to include cities from other countries and climate zones would enhance the global applicability of the findings.

5 Conclusion and Outlook

This study demonstrates the potential of machine learning models in accurately predicting nighttime thermal comfort in urban environments. Among the tested models, the Random Forest algorithm, both in regression and classification forms, achieved the highest performance, underscoring its robustness in handling the complex interplay of urban meteorological variables and human thermal perception. The findings also highlight the importance of incorporating demographic and environmental data to enhance prediction accuracy. Social media user data can improve the accuracy of ML model predictions to a certain extent. The study also shows the ranking of various features' importance in the thermal comfort ML prediction model. By integrating targeted interventions, these insights can guide urban planners and architects in designing more inclusive and thermally comfortable public spaces. For instance, the findings can inform the placement of shaded seating areas, water features, and cooling or heating infrastructure in urban parks. Specifically, in areas identified by the predictive models as having poor nighttime thermal comfort, planners can prioritize the installation of tree-lined streets, green roofs, and vertical gardens.

Despite its promising findings, the study also reveals areas for future research and improvement. Although using social media data as a thermal comfort prediction dataset is innovative, it also has certain limitations. Future work should focus on integrating various data sources, such as field measurements and global datasets, to verify and expand the model's applicability. Additionally, developing ML prediction models based on social media data for the 7-point TSV scale could further enhance the reliability of thermal comfort assessments. By addressing these limitations and building on the findings, this research paves the way for more comprehensive and practical applications in urban planning, helping to create thermally inclusive and sustainable urban environments worldwide.

References

AGHAMOLAEI, R., AZIZI, M. M., AMINZADEH, B. & O'DONNELL, J. (2023), A comprehensive review of outdoor thermal comfort in urban areas: Effective parameters and approaches. Energy & Environment, 34 (6), 2204-2227.

CARMONA, M. (2021), Public places urban spaces: The dimensions of urban design. Routledge.

CHEN, L. & NG, E. (2012), Outdoor thermal comfort and outdoor activities: A review of research in the past decade. Cities, 29 (2), 118-125.

GAITANI, N., MIHALAKAKOU, G. & SANTAMOURIS, M. (2007), On the use of bioclimatic architecture principles in order to improve thermal comfort conditions in outdoor spaces. Building and environment, 42 (1), 317-324.

GEHL, J., SVARRE, B. B. & RISOM, J. (2011), Cities for people. Planning News, 37 (4), 6-8.

HÖPPE, P. (2002), Different aspects of assessing indoor and outdoor thermal comfort. Energy and Buildings, 34 (6), 661-665.

JACOBS, J. (1961), Jane Jacobs. The Death and Life of Great American Cities, 21 (1), 13-25.

KOTTEK, M., GRIESER, J., BECK, C., RUDOLF, B. & RUBEL, F. (2006), World map of the Köppen-Geiger climate classification updated.

KUMAR, P. & SHARMA, A. (2020), Study on importance, procedure, and scope of outdoor thermal comfort – A review. Sustainable Cities and Society, 61, 102297.

LUO, M., XIE, J., YAN, Y., KE, Z., YU, P., WANG, Z. & ZHANG, J. (2020), Comparing machine learning algorithms in predicting thermal sensation using ASHRAE Comfort Database II. Energy and Buildings, 210, 109776.

MEHTA, V. (2013), The street: a quintessential social public space. Routledge.

NIKOLOPOULOU, M. & STEEMERS, K. (2003), Thermal comfort and psychological adaptation as a guide for designing urban spaces. Energy and Buildings, 35 (1), 95-101.

OKE, T. R. (1982), The energetic basis of the urban heat island. Quarterly journal of the royal meteorological society, 108 (455), 1-24.

ROTH, M., OKE, T. R. & EMERY, W. J. (1989), Satellite-derived urban heat islands from three coastal cities and the utilization of such data in urban climatology. International Journal of Remote Sensing, 10 (11), 1699-1720.

SLOAN, L., MORGAN, J., BURNAP, P. & WILLIAMS, M. (2015), Who tweets? Deriving the demographic characteristics of age, occupation and social class from Twitter user meta-data. PloS one, 10 (3), e0115545.

SPAGNOLO, J. & DE DEAR, R. (2003), A field study of thermal comfort in outdoor and semi-outdoor environments in subtropical Sydney Australia. Building and environment, 38 (5), 721-738.

XU, J., ZHANG, X. & ZHONG, S. (2024), Hidden complex multistable dynamical analysis and FPGA implementation of integer-fractional order memristive-memcapacitive chaotic system. Physica Scripta, 99 (12), 125248.

YIN, Q., CAO, Y. & SUN, C. (2021), Research on outdoor thermal comfort of high-density urban center in severe cold area. Building and Environment, 200, 107938.

ZHANG, X., CHEN, X., LIU, S. & ZHONG, S. (2024), Anonymous Authentication and Information Sharing Scheme Based on Blockchain and Zero Knowledge Proof for VANETs. IEEE Transactions on Vehicular Technology.

ZHONG, S. (2019), Heterogeneous memristive models design and its application in information security. Computers, Materials & Continua, 60 (2), 465-479.