Applying Machine Learning and Digital Twinning for the Live Assessment of Thermal Comfort in Buildings

Mehrzad Shahinmoghadam, mehrzad.shahinmoghadam.1@ens.etsmtl.ca
Department of Construction Engineering, École de Technologie Supérieure Université du Québec, Canada

Ali Motamedi, ali.motamedi@etsmtl.ca
Department of Construction Engineering, École de Technologie Supérieure, Université du Québec, Canada

Mohamed Cheriet, mohamed.cheriet@etsmtl.ca
Department of Systems Engineering, École de Technologie Supérieure, Université du Québec, Canada

Abstract

Recent research has shown the potential of combining Building Information Modeling (BIM) and Internet of Things (IoT) to produce accurate digital twins of buildings. Such digital twins rendered within 3D visualization environments can enable highly intuitive experiences of human-building interactions. Moreover, machine learning techniques can be applied to extract new knowledge from building historical data and update the knowledge base of the digital twin. Digital twins arrayed with data-driven learning capability will be of high practical value for highly complex subjects such as thermal comfort. Accurate predictions of thermal comfort depend on numerous variables, and complex equations or simulations that are computationally expensive. This paper set out to investigate a data-driven thermal comfort prediction model and its integration within the architecture of digital twins created using BIM and IoT systems. For this purpose, a supervised learning algorithm was used to train a classification model based on the ASHRAE Global Thermal Comfort Database II. A partial implementation of the proposed system was conducted and the overall effectiveness of the method proposed for thermal comfort prediction is discussed through a comparative analysis of the results.

Keywords: Digital Twin, Machine learning, BIM, Internet of Things, Thermal comfort, Building analytics

1 Introduction

Recently, the research community in the domain of building engineering has shown a growing interest in digital twin technologies. Most recent research trends have been focusing on how to create accurate and live digital twins of buildings by integrating the rich building contextual data (3D geometry, spatial relationships, etc.) provided through Building Information Modeling (BIM), and live sensor data streaming from Internet of Things (IoT) nodes. Moreover, promising results have been reported with regards to making digital twins more intuitive and interactive using interfaces developed within interactive 3D rendering engines (e.g., WebGL). Moreover, application of machine learning algorithms over building's historical data makes it possible to add self-learning capabilities to digital twinning. This offers high value given that making predictions about a building's particular state or behavior in a real-time manner, based on experimental methods and simulation-based models is significantly challenging or inefficient from a practical standpoint. For example, to make real-time predictions of thermal comfort preferences, the data for various variables (e.g., space geometry, indoor environmental conditions, clothing insulation, metabolic rate, etc.) must be monitored and then used in complex equations or computationally expensive simulations.

Thermal comfort in buildings can be viewed as a condition of mind reflecting the level of an occupant's satisfaction with the thermal environment (ANSI/ASHRAE 2017). Various factors influence the thermal comfort of occupants inside a building. Such factors include measurable environmental conditions (e.g., air temperature and relative humidity), as well as personal and psychological factors such as, metabolic rate, level of clothing insulation, and space lighting (Grondzik & Kwok 2015). The timely and accurate prediction of the occupants' thermal preferences not only ensures their thermal comfort within building spaces, but also can lead to significant financial and energy savings. The latter can be achieved by regulating building cooling/heating strategies according to the preferences of the occupants, which were predicted based on the current state of the thermal environment within the building spaces. To this end, live data about the variables that have an influence on the occupant's thermal sensations (e.g., relative humidity, air velocity, outdoor temperature) can be collected and used in a predictive model to anticipate the expected thermal preference of the occupants for the next few minutes. However, due to the highly complex nature of the thermal comfort phenomenon and the fact that the thermal sensations expressed are directly affected by subjective judgments, effective prediction of thermal comfort remains an ongoing question to be addressed in building engineering research.

The present study investigates the application of digital twinning to enable the live assessment of thermal comfort conditions in indoor building spaces. This study builds on previous work from the authors regarding the potential of BIM-IoT integrated architectures in (Motamedi & Shahinmoghadam 2021; Shahinmoghadam & Motamedi 2019). In a more recently published work by the authors (Shahinmoghadam et al., 2021), a virtual reality-rendered digital twin for the real-time assessment of thermal comfort was created by integrating BIM and live IoT data within a game engine environment. However, the accuracy of the proposed system will decrease as the geometry of the space increases in complexity. This is due to the system's dependency on timely and accurate processing of thermal images used to calculate mean radiant temperature as a key input variable to the equations used in the PMV-PPD model (ANSI/ASHRAE 2017). To avoid such dependency on the geometric characteristics of building spaces and real-time thermal image processing results, the study reported here investigates the effectiveness of a data-driven approach as an alternative to the PMV-PPD model. In particular, methods from pattern recognition were applied to develop a thermal comfort prediction model to be included within the digital twin architecture proposed in (Shahinmoghadam et al., 2021). To achieve this, an ensemble learning method was used to develop a model capable of predicting the occupants' thermal preferences based on live IoT data and user-defined input. The ASHRAE Global Thermal Comfort Database II (Földváry Ličina et al 2018) was used to train, validate, and test the proposed model. The rest of the paper elaborates on the related background, proposed methodology, and discussion of the obtained results.

2 Background and related works

2.1 Thermal comfort prediction

The thermal comfort models prescribed by current standards such as (ANSI/ASHRAE 2017) and (ISO 7730 2005) are based on the mathematical expressions of occupant thermal sensations that have been derived from chamber experiments and field study data (Kim et al 2018). Examples of such models are Fanger's PMV model (Fanger 1970) and De Dear & Brager's adaptive model (De Dear & Brager 1998), which are among the most cited models within the existing body of knowledge. This being said, Computational Fluid Dynamics (CFD) simulations are commonly used in the design phase of buildings to evaluate thermal comfort conditions (Chiang, Wang & Huang 2012; Alizadeh & Sadrameli 2018; Chen, Xin & Liu 2020). Although such approaches can be effective during the design stage, during the operation stage of the building's lifecycle, practical issues arise when attempting to monitor thermal comfort conditions in real time. This is due to the inherent complexity of the existing models and the high cost of real-time computation for sophisticated simulations. Consequently, in practice, simplistic approaches are most often adopted, such as setback settings based on a fixed comfort temperature. Such simplistic approaches to the regularization of the building's cooling/heating operations lead to numerous situations in which building occupants feel discomfort and exhibit behaviors that are undesirable from an energy preservation perspective, such as, opening windows since they feel cold/hot while the heater/cooler is working. Hence, efficient prediction models for real-time monitoring of thermal comfort have yet to be investigated.

To address the above-mentioned issues and predict the thermal demands of occupants, researchers have been investigating various machine learning methods and developing personal comfort models (Dai et al 2017; Kim et al 2018; Luo et al 2020). In these studies, different input factors were considered to predict the thermal comfort related parameters. Depending on the input factors and the target parameters (e.g., thermal sensations, thermal preferences, etc.), different databases with varying sizes have been developed or re-used. For example, Li et al., (2018) created a dataset of facial infrared thermography to predict thermal comfort based on skin temperature data. However, access to the large volume of data in ASHRAE Global Thermal Comfort Database II (Földváry Ličina et al 2018) has provided an opportunity to more effectively investigate the potential of pattern recognition approaches to thermal comfort prediction. Given that the mentioned database is recent, the number of reported studies using it is limited. Among the recent studies (Luo et al 2020; Wang et al 2020; Zhou et al 2020), the work of Luo et al., (2020) has received particular attention within the research community. The study investigated nine different supervised machine learning algorithms to predict occupant votes for thermal sensations and preferences in 7-point and 3-point scales, respectively. The best predictions were obtained using the Random Forest algorithm with a 66.3% accuracy for 3-point thermal preference votes. However, their methodology raises two major concerns. First, the number of samples used to train and test the models was considerably smaller than the total number of samples existing in the original database. The reason for this was the high rate of missing values among the database entries. However, by investigating an effective remedial strategy to deal with the missing values, larger data sets could be used to train and test the models, which in return, could improve the generalization power of the models. Second, considering the imbalanced ratio of the samples existing in the original database for different target labels, the accuracy metric that was used in the study to evaluate the quality of predictions can be misleading. Alternatively, metrics specific to class imbalance problems could be used to more effectively account for class imbalance and accurately evaluate the prediction performance of the models.

2.2 Digital twinning and thermal comfort monitoring

Previous research has shown that BIM-IoT integrated data can be effectively used to create accurate digital twins to monitor the state and behavior of various building components and systems (Motamedi & Shahinmoghadam 2021). Furthermore, 2D/3D-rendered interactive interfaces developed for such digital twins can deliver rich visualizations and immersive experiences of user interactivity, which will be of significant value to building visual analytics applications (Motamedi et al 2014). In this light, the potential of using BIM data within game engine environments has been under investigation for a wide range of applications such as, architectural and engineering design processes (Kang & Hong 2015), indoor spatial analysis and lighting design (Motamedi et al 2017; Natephra et al 2017), and safety planning (Feng et al 2020).

In the context of thermal comfort, real-time monitoring of the comfort conditions can be achieved by capturing the current state of the thermal comfort variables through a network of sensors and using the monitoring data as input for a prediction model. Moreover, by linking live thermal environment monitoring data to a building's contextual information provided through BIM-based workflows (e.g., 2D/3D spatial information), various rich visualizations can be created, thereby enabling an in-depth analysis of the current conditions (Marzouk & Abdelaty 2014; Chang, Dzeng & Wu 2018). Other studies have shown how game engines can be utilized to render BIM and IoT sensor data, so as to add a sense of immersion and higher levels of interactivity to the monitoring of thermal environments (Natephra & Motamedi 2019a, 2019b; Hosokawa et al 2016; Fukuda et al 2019).

In light of all the issues discussed earlier in this paper, the present study sets out to include a machine learning-based prediction model within a digital twin architecture designed for the real-time and immersive monitoring of the thermal comfort conditions in building enclosures. Based on the promising results reported in (Luo et al 2020), the Random Forest algorithm was selected. The algorithm was used to train and test a classification model, to predict 3-point thermal preferences, i.e., occupant votes described as “cooler”, “warmer”, or “no change”. The model was trained and tested using the ASHRAE Global Thermal Comfort Database II (Földváry Ličina et al 2018) and proper strategies were considered to address the shortcomings of (Luo et al 2020) mentioned in Section 2.1. Moreover, a general architecture was proposed to illustrate how the live IoT sensor data and user-defined input can be used to make live predictions of thermal preferences within a digital twin rendered using BIM geometric data. The next section elaborates on the methodology proposed to meet the mentioned objectives.

3 Methods

3.1 Model training and evaluation

3.1.1 Data preparation

To prepare the data sets for model training and testing, data cleaning and preprocessing steps were required. Hence, as the first step, non-informative columns (e.g., columns containing the same data represented using different units) were removed from the original database. Subsequently, each column with more than a 60% rate of missing values was discarded. Next, all the data entries with missing values in the target variable (i.e., thermal preference) were removed. As a result of taking the aforementioned steps, ≈76,000 samples described with 9 features were kept and these constituted the base data frame that was used in this study. Among the 9 features, 3 were of categorical type (season, building type, cooling strategy) and the rest were of numerical type (air temperature, relative humidity, air velocity, average outdoor monthly temperature, metabolic rate, clothing insulation level).

Preprocessing of numerical features consisted of removing the mean and scaling to unit variance. In this way, the numerical data for each feature was represented so as to resemble a Gaussian distribution with zero mean and unit variance. For categorical data types, one-hot encoding was used by deriving the categories based on the unique values appearing in each feature and creating a binary column for each unique value. As a result of one-hot encoding, the size of the feature vector was increased to 19. Afterwards, to preserve a maximum number of samples from the original database, mean and median imputation strategies (replacing missing values with the mean/median of each column) were considered in this work.

To account for the class imbalance issue, two re-sampling strategies were compared. This was an important step given the ratio of the samples describing each thermal preference class: “no change”: 51%, “cooler”: 32%, “warmer”: 17%. The main purpose of applying the re-sampling strategy was to prevent the decision functions of the trained classifiers from favoring the class with the larger number of samples, i.e., “no change”. The two re-sampling strategies compared in this project were random under-sampling and random over-sampling. For random under-sampling, random resampling without replacement was performed to reduce the number of samples from majority classes (“no change” and “cooler”) to the number of the samples existing for the minority class (“warmer”). For random over-sampling, the number of samples from minority classes (“cooler”, “warmer”) were increased to the number of the majority class (“no change”) samples, through random resampling with replacement.

Finally, 85% of total samples were used for model training, and the remaining 15% of the samples were kept as the test set. Sample splitting was performed in a stratified manner to account for the class ratios. Moreover, particular attention was paid to perform re-sampling only for the training set, thereby eliminating the risk of introducing the test samples to the classifier during the training phase.

3.1.2 Model tuning and performance evaluation

As mentioned previously, the Random Forest (Breiman 2001) algorithm is used in this work for the purpose of training the thermal preference prediction model. The Random Forest model was developed as an ensemble of Decision Tree classifiers that were trained via the bootstrap aggregating (bagging) method. In particular, every classification tree in the ensemble was trained using different random subsets of the training set when sampling was performed with replacement, i.e., bagging. To avoid overfitting the training data, tuning of hyper-parameters was performed to restrict the decision trees used in the Random Forest classifier from adapting themselves to the training data. For this purpose, a grid search implementation was used to search the pre-defined hyper-parameter space for the best cross-validation score. The search space considered in this project for the Random Forest hyper-parameters is described in Table 1.

Table 1. Hyper-parameter search space considered for the Random Forest classifier

Hyper-parameter description	Search space
Number of classification trees used as base estimators	[200, 500, 1000, 2000]
Maximum depth of the trees	[2, 6, 10, 15]
Minimum number of samples required to split an internal node	[4, 8, 12]
Minimum number of samples required to be at a leaf node	[2, 6, 10]

The scoring strategy considered for the cross-validation was based on “balanced accuracy” scores. The balanced accuracy metric was used since it can specifically account for the class imbalance issue by avoiding overstated performance measurements on the imbalanced data sets. The balanced accuracy scores were obtained by calculating the macro-average of recall scores (ratio of true positive predictions over the sum of the true positive and false negative predictions) per class, i.e., calculating mean of the recall scores by giving equal weight to each class.

Finally, to evaluate the quality of the predictions made by the optimal trained classifier, the test dataset was introduced to the model to predict the labels for unseen instances. Afterwards, the corresponding confusion matrix was plotted by comparing the predicted labels against the true labels. Next, regular and balanced accuracy metrics were used to quantitatively summarize the prediction performance of the model.

3.2 System architecture

Figure 1 shows how the trained thermal preference prediction model can be included within a digital twin architecture, in the form of a web service.

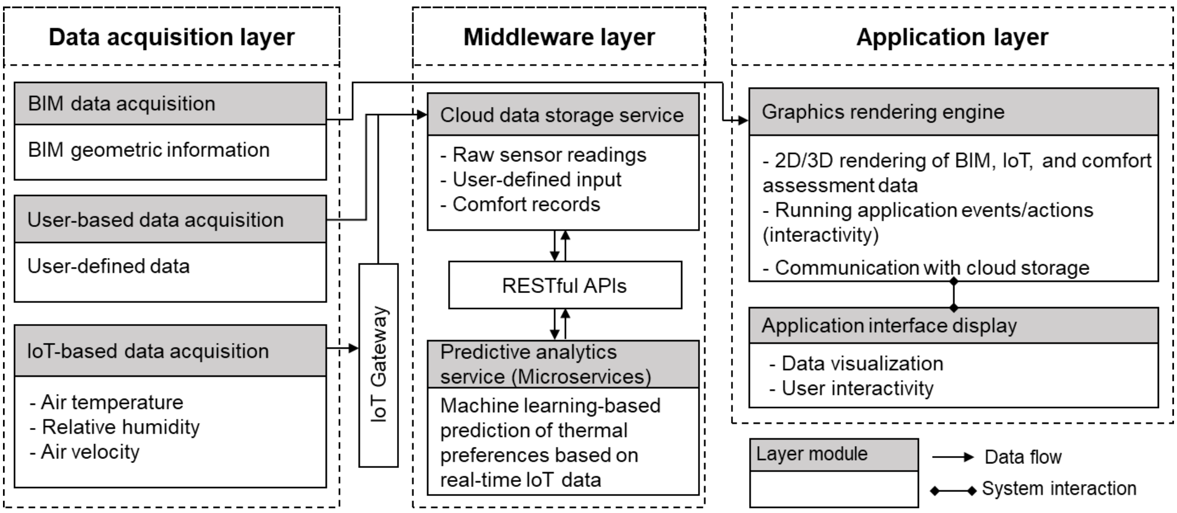


Figure 1. Proposed digital twin system architecture

A detailed explanation of the various functionalities that can be delivered by the modules included in the three layers depicted in Figure 1 can be found in (Shahinmoghadam et al., 2021). Regarding the new adaptation, i.e., integrating the "trained model" module within the previously proposed digital twin architecture, a service-oriented architecture has been considered for the sake of model deployment. In particular, subsequent to training and testing the thermal preference prediction model, it is deployed on a cloud server and the prediction functionality is requested through a web API. HTTP requests containing live IoT readings and user input data are sent to initiate the prediction function of the model deployed on the cloud server, and the prediction results are returned in the form of HTTP responses. The prediction results are then stored in the cloud to be accessed via the application layer.

4 Results and discussion

Once the preprocessing steps described in Section 3.1.1 were completed, a basic Random Forest classifier, i.e., a model with 100 trees with no parameter tuning, was used to compare the effectiveness of the two imputation strategies through a 5-fold cross-validation process. The cross-validated balanced accuracy scores obtained for the mean and median strategies were 52.23% and 52.21%, respectively. Since the mean strategy, yielded (very slightly) better results, all numerical missing values were imputed with the mean value of the corresponding columns.

Subsequently, the random over-sampling and under-sampling strategies were compared using the same basic Random Forest classifier and with reference to the 5-fold cross-validated balanced accuracy scores. The corresponding scores are shown in Figure 2.

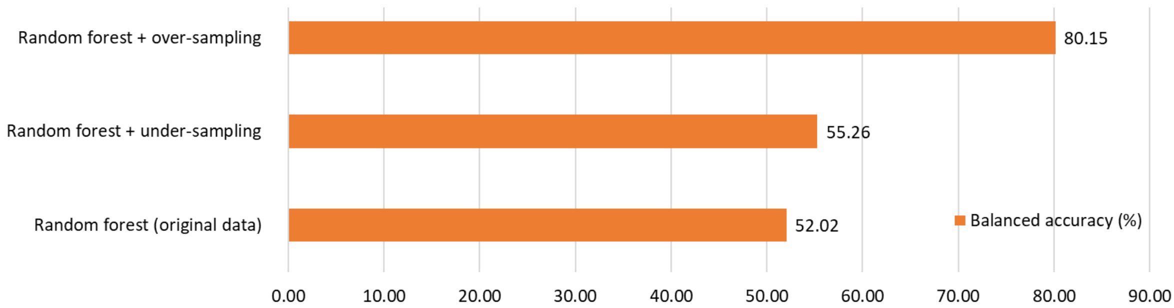


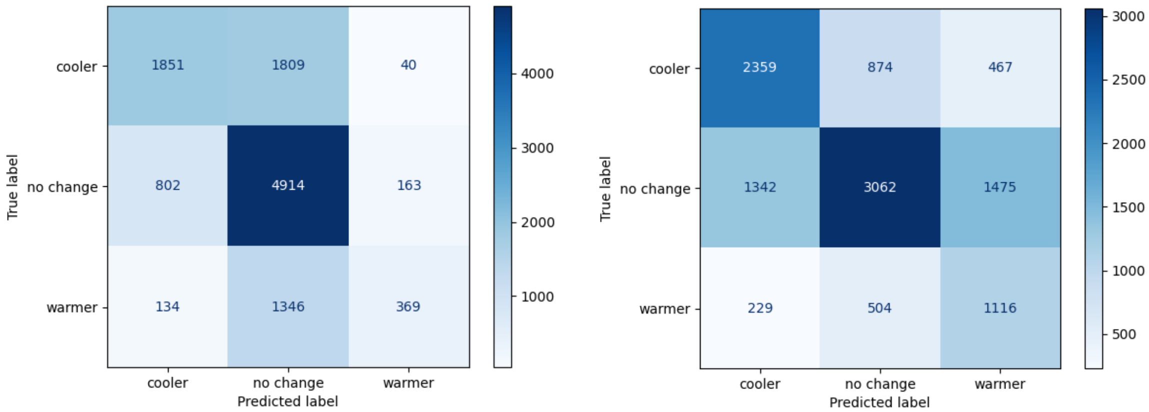
Figure 2. Comparison of the effectiveness of re-sampling strategies

As Figure 2 shows, although the balanced accuracy scores were improved after applying both re-sampling strategies, the over-sampling method yielded better results. However, it should be noted that the scores shown in Figure 2 represent the cross-validated scores, which are expected to be higher than the actual score of each model. Moreover, the 80.2% score for the over-sampling strategy should not be considered as a significant improvement at this step because the over-sampled instances in the validation folds might have been seen by the model from the training folds. However, it is safe to say that the obtained scores can be used to compare the relative effectiveness of the two strategies.

To identify the optimal hyper-parameters of the Random Forest classifier, the parameter space described in Table 1 was searched by considering all possible combinations. This was done by performing a 5-fold cross validation and calculating the balanced accuracy scores for each prospective model. The best score was obtained using 2000 decision trees, and for each tree, the maximum depth, minimum number of samples to split at internal nodes, and minimum number of samples to be used at a leaf node, were 15, 8, and 2, respectively.

After identifying the optimal hyper-parameters to be used in the Random Forest algorithm, the final classification model was developed using the training set. Subsequently, the test set was introduced to the trained model to evaluate the quality of its predictions for unseen data. To highlight the effect of the re-sampling strategy, another Random Forest classifier was trained and tested with original data. The resulting confusion matrices are shown in Figure 3.

As can be seen in Figure 3a, when no remedy was applied to account for the problem of class imbalance, the obtained decision function favored the majority class (i.e., 'no change'). The corresponding confusion matrix shows that in 1346 and 1809 cases, the 'warmer' and 'cooler' preferences were confused, respectively, with the majority class ('no change'). However, by applying the over-sampling strategy and training the model with balanced data, the number of true positive predictions were significantly increased for 'warmer' and 'cooler' classes. Moreover, the false negative predictions dropped from 1480 (1346+134) to 733 (504+229) for 'warmer' class, and from 1849 (1809+40) to 1341 (874+467) for 'cooler' class.



(a) Random Forest classifier trained with original data (b) Random Forest classifier trained with over-sampled data

Figure 3. Confusion matrices obtained by testing the trained Random Forest classifiers

Figure 4 illustrates in greater detail the quality of the predictions obtained from the models trained with original and over-sampled data, as well as the importance of choosing proper evaluation metrics, the models' prediction performance measurements with reference to accuracy and balanced accuracy metrics. As the figure shows, when the classifier was trained with over-sampled data, the model's prediction performance improved in terms of balanced accuracy score. The accuracy scores represented in Figure 4 are discussed below.

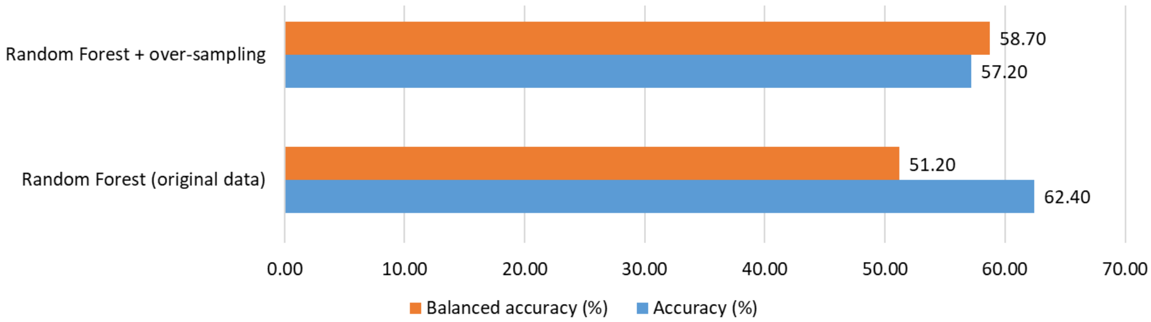


Figure 4. Summary of prediction performance measurements

Regarding the 66.3% accuracy score for 3-point thermal preference votes reported in (Luo et al 2020), two critical issues can be raised here:

First, in that study, the predictions were evaluated based on regular accuracy scores. However, as mentioned previously, accuracy scores can overstate the prediction performance when working with imbalance data sets. In this respect, an important observation can be derived from Figure 4: performance evaluation based on accuracy scores can be considerably misleading for the problem at hand. As the figure shows, when no remedial action was taken to account for class imbalance, the Random Forest model yielded a 62.4% accuracy score, while the model's balanced accuracy score was 51.2%, which is lower than the balanced accuracy score for the model trained with re-sampled data.

Second, it should be noted that the total number of samples used in this project was approximately 7 times larger than that used in (Luo et al 2020). The larger dataset used in the present work was obtained by imputing the missing values, as opposed to Luo et al.'s method that discarded samples with missing values even for a single feature. Hence, it can be concluded that the model's performance results reported here are more robust in comparison, as a larger number of test samples were used in this project to evaluate the generalization capability of the trained classifier.

As for discussing the results obtained regarding the machine learning-based prediction model, it should be noted that to select the best model parameters, the criteria used for calculating the cost of the errors should be defined beforehand. The criteria should then be used to smooth the decision function of the model (through parameter tuning). In this way, the cost of the wrong predictions made by the trained model can be minimized. In other words, it should first be defined whether the priority is given to the thermal comfort of the occupants or to other factors such as energy preservation.

Finally, although validation of the trained model against the ground truth data from the used database yielded promising results, the full implementation of the digital twin system depicted in Figure 1 and validation of its overall effectiveness is beyond the scope of the present work. However, an overview of the implementation and validation processes is given as follows. With the trained and tested prediction model at hand, live predictions can be made by deploying the trained model on a cloud server and making its prediction functionality available on the web through a designated API (deploying the developed prediction model as a microservice). For each prediction to be made by the trained model, nine feature values are required as input to the prediction function. Among these features, the values for season, building type, cooling strategy, clothing insulation level, and metabolic rate, can be defined in advance by the end-user using the main interface. Values regarding the average air temperature, relative humidity, and air velocity in the building spaces can be provided through web requests sent from the game engine development environment to the cloud storage module, which contains the most recent IoT sensor readings. As the nine feature values are provided within the game engine environment, web requests containing all the input values will be sent to the cloud server on which the thermal preference prediction model has been deployed, to initiate the prediction function. Finally, the server responses containing the corresponding predicted thermal preferences will be received in the game engine environment. To validate the overall effectiveness of the system, an experimental evaluation should be conducted by involving a diverse group of experts and asking their opinions with regards to the completeness, consistency, and usability of the proposed system.

5 Conclusion

This study investigated a machine learning-based thermal preference prediction model to be included within the architecture of a digital twin system. The digital twin uses BIM and IoT data to provide a real-time and immersive assessment of the thermal comfort conditions in building spaces. Using a large public database, an ensemble learning method (Random Forest) was used to train a model capable of predicting building occupant thermal preferences in a 3-point scale (warmer, cooler, no change). The study of (Luo et al 2020) was considered as a base reference to discuss the obtained results.

In a broader context, the study showed how extracting hidden patterns from accumulated operational data and updating the digital twin's knowledge base with the newly discovered patterns can enable a building digital twin to continuously evolve. The study proposed a general architecture and implementation scheme to integrate the developed data-driven thermal preference prediction model to a digital twin application, which makes use of BIM and live IoT data for thermal environment monitoring and immersive visualization purposes. In future work, the authors will investigate a full implementation of the system and validate the overall effectiveness of the proposed digital twin architecture.

References

Alizadeh, M & Sadrameli, SM 2018, 'Numerical modeling and optimization of thermal comfort in building: Central composite design and CFD simulation', *Energy and Buildings*, vol. 164, pp. 187–202.

ANSI/ASHRAE 2017, 'Standard 55–2017 Thermal Environmental Conditions for Human Occupancy', *American Society of Heating, Refrigerating and Air Conditioning Engineers: Atlanta, GA, USA*.

Breiman, L 2001, 'Random Forests', *Machine Learning*, vol. 45, no. 1, pp. 5–32.

Chang, KM, Dzeng, RJ & Wu, YJ 2018, 'An automated IoT visualization BIM platform for decision support in facilities management', *Applied Sciences*, vol. 8, no. 7, p. 1086.

Chen, Z, Xin, J & Liu, P 2020, 'Air quality and thermal comfort analysis of kitchen environment with CFD simulation and experimental calibration', *Building and Environment*, vol. 172, p. 106691.

Chiang, W-H, Wang, C-Y & Huang, J-S 2012, 'Evaluation of cooling ceiling and mechanical ventilation systems on thermal comfort using CFD study in an office for subtropical region', *Building and Environment*, vol. 48, pp. 113–127.

Dai, C, Zhang, H, Arens, E & Lian, Z 2017, 'Machine learning approaches to predict thermal demands using skin temperatures: Steady-state conditions', *Building and Environment*, vol. 114, pp. 1–10.

De Dear, R & Brager, GS 1998, 'Developing an adaptive model of thermal comfort and preference',.

Fanger, PO 1970, *Thermal comfort: Analysis and applications in environmental engineering*, McGraw-Hill, New York.

Feng, Z, González, VA, Amor, R, Spearpoint, M, Thomas, J, Sacks, R, Lovreglio, R & Cabrera-Guerrero, G 2020, 'An immersive virtual reality serious game to enhance earthquake behavioral responses and post-earthquake evacuation preparedness in buildings', *Advanced Engineering Informatics*, vol. 45, p. 101118.

Földváry Ličina, V, Cheung, T, Zhang, H, de Dear, R, Parkinson, T, Arens, E, Chun, C, Schiavon, S, Luo, M, Brager, G, Li, P, Kaam, S, Adebamowo, MA, Andamon, MM, Babich, F, Bouden, C, Bukovianska, H, Candido, C, Cao, B, Carlucci, S, Cheong, DKW, Choi, J-H, Cook, M, Cropper, P, Deuble, M, Heidari, S, Indraganti, M, Jin, Q, Kim, H, Kim, J, Konis, K, Singh, MK, Kwok, A, Lamberts, R, Loveday, D, Langevin, J, Manu, S, Moosmann, C, Nicol, F, Ooka, R, Oseland, NA, Pagliano, L, Petráš, D, Rawal, R, Romero, R, Rijal, HB, Sekhar, C, Schweiker, M, Tartarini, F, Tanabe, S, Tham, KW, Teli, D, Toftum, J, Toledo, L, Tsuzuki, K, De Vecchi, R, Wagner, A, Wang, Z, Wallbaum, H, Webb, L, Yang, L, Zhu, Y, Zhai, Y, Zhang, Y & Zhou, X 2018, 'Development of the ASHRAE Global Thermal Comfort Database II', *Building and Environment*, vol. 142, pp. 502–512.

Fukuda, T, Yokoi, K, Yabuki, N & Motamedi, A 2019, 'An indoor thermal environment design system for renovation using augmented reality', *Journal of Computational Design and Engineering*, vol. 6, no. 2, pp. 179–188.

Grondzik, WT & Kwok, AG 2015, *Mechanical and Electrical Equipment for Buildings*, Wiley & Sons, New Jersey.

Hosokawa, M, Fukuda, T, Yabuki, N, Michikawa, T & Motamedi, A 2016, 'Integrating CFD and VR for indoor thermal environment design feedback', in *CAADRIA 2016, 21st International Conference on Computer-Aided Architectural Design Research in Asia - Living Systems and Micro-Utopias: Towards Continuous Designing*, Hong Kong, pp. 663–672.

ISO 7730 2005, *Ergonomics of the thermal environment - Analytical determination and interpretation of thermal comfort using calculation of the PMV and PPD indices and local thermal comfort criteria*, British Standards Institution, Geneva, Swizerland.

Kang, TW & Hong, CH 2015, 'A study on software architecture for effective BIM/GIS-based facility management data integration', *Automation in Construction*, vol. 54, pp. 25–38.

Kim, J, Zhou, Y, Schiavon, S, Raftery, P & Brager, G 2018, 'Personal comfort models: Predicting individuals' thermal preference using occupant heating and cooling behavior and machine learning', *Building and Environment*, vol. 129, pp. 96–106.

Li, D, Menassa, CC & Kamat, VR 2018, 'Non-intrusive interpretation of human thermal comfort through analysis of facial infrared thermography', *Energy and Buildings*, vol. 176, pp. 246–261.

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*, vol. 210, p. 109776.

Marzouk, M & Abdelaty, A 2014, 'Monitoring thermal comfort in subways using building information modeling', *Energy and Buildings*, vol. 84, pp. 252–257.

Motamedi, A, Hammad, A & Asen, Y 2014, 'Knowledge-assisted BIM-based visual analytics for failure root cause detection in facilities management', Automation in Construction, vol. 43, pp. 73–83.

Motamedi, A & Shahinmoghadam, M 2021, 'BIM-IoT-integrated Architectures as the Backbone of Cognitive Buildings: Current State and Future Directions', CRC Press, pp. 45–68.

Motamedi, A, Wang, Z, Yabuki, N, Fukuda, T & Michikawa, T 2017, 'Signage visibility analysis and optimization system using BIM-enabled virtual reality (VR) environments', *Advanced Engineering Informatics*, vol. 32, pp. 248–262.

Natephra, W & Motamedi, A 2019a, 'BIM-based Live Sensor Data Visualization Using Virtual Reality for Monitoring Indoor Conditions', in *24th Annual Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA 2019)*.

Natephra, W & Motamedi, A 2019b, 'Live Data Visualization of IoT Sensors Using Augmented Reality (AR) and BIM', in *2019 Proceedings of the 36th ISARC, Banff, Alberta, Canada*, pp. 632–638.

Natephra, W, Motamedi, A, Fukuda, T & Yabuki, N 2017, 'Integrating building information modeling and virtual reality development engines for building indoor lighting design', *Visualization in Engineering*, vol. 5, no. 1, p. 19.

Shahinmoghadam, M & Motamedi, A 2019, 'Review of BIM-centered IoT deployment: State of the Art, Opportunities, and Challenges', in *2019 Proceedings of the 36th ISARC, Banff, Alberta, Canada*, pp. 1268–1275.

Shahinmoghadam, M, Natephra, W & Motamedi, A 2021, 'BIM- and IoT-based virtual reality tool for real-time thermal comfort assessment in building enclosures', Building and Environment, vol. 199, p. 107905.

Tang, S, Shelden, DR, Eastman, CM, Pishdad-Bozorgi, P & Gao, X 2019, 'A review of building information modeling (BIM) and the internet of things (IoT) devices integration: Present status and future trends', *Automation in Construction*, vol. 101, pp. 127–139.

Wang, Z, Wang, J, He, Y, Liu, Y, Lin, B & Hong, T 2020, 'Dimension analysis of subjective thermal comfort metrics based on ASHRAE Global Thermal Comfort Database using machine learning', *Journal of Building Engineering*, vol. 29, p. 101120.

Zhou, X, Xu, L, Zhang, J, Niu, B, Luo, M, Zhou, G & Zhang, X 2020, 'Data-driven thermal comfort model via support vector machine algorithms: Insights from ASHRAE RP-884 database', *Energy and Buildings*, vol. 211, p. 109795.