A data-driven method to minimize sensor use for dynamic shading control

Jiaqi Zhu
Elite Preparatory Academy, Somerset, New Jersey, 08873, USA
joshzhu402@gmail.com

Abstract. Dynamic facade systems are a promising strategy to balance the benefits and drawbacks of daylight penetration through the window in a built environment. A well-operated shading system should allow for sufficient daylight access while preventing glare in response to changing weather conditions. The dynamic shading system can also help with making use of daylight to reduce the lighting energy use when integrated with electric lighting control. This paper presented a data-driven approach to control the automated shading system. The control method allows the system to be responsive to outdoor weather conditions. With data from offline daylight simulation using typical meteorological year weather, three machine learning classification algorithms were applied to train predictive models for online shading control. The proposed method was validated using a climate-based simulation. It was found that all models had a satisfying prediction performance, with an accuracy of 0.85 to 0.92. The shading control could maintain the work plane illuminance for 97% to 99% of the time during the year. It was also associated with a potential of reducing electric lighting energy use by 40% to 43% compared to that of on/off lighting control.

1. Introduction

1.1. Background
In the current construction industry, the problem of increased building energy and related carbon emissions is becoming more and more serious. With the development of technology, the design and operation of buildings can be improved to achieve the goal of energy-saving and environmental protection. Such advances can make huge changes to our living environment and greatly benefit future generations. For example, we can make use of daylight to reduce the energy consumption of artificial lighting. Meanwhile, numerous studies have shown that sunlight has a positive effect on human health, wellbeing, and productivity. For instance, a recent controlled study showed that access to daylight in an office reduced eyestrain and improved office workers’ cognitive performance and satisfaction [1]. Given that people spend more than 90% of their time indoors, daylight ingress to the indoor environment is essential to create a healthy, comfortable, and satisfying environment for them. However, excessive and uncontrolled daylight from windows can cause glare, thermal discomfort, and increased building energy demand. A balance between the benefits and drawbacks of daylight ingress is required. This balance can be achieved using building automated shading systems.

1.2. Related work
Conventional automated shading systems typically use sensors such as illuminance sensors or solar radiation sensors to measure available daylight levels [2]. However, this method has its limitations. One
of the main challenges is the placement of sensors. For instance, it is usually not practical to place illuminance sensors on the work plane because they can disturb or interfere with office workers. In practice, sensors are often positioned on the ceiling, the window frame, or the sidewall. In this case, the sensor readings might not be capable of reflecting what the occupants experience at their eye level, thus failing to effectively adjust the shades. Additionally, the readings of sensors might be influenced by occupants’ activities, leading to irregular values and inappropriate operation of the shading devices. Moreover, this method could entail excessive cost and labor for the installation, calibration, maintenance, and replacement of sensors, especially in open-plan offices. Additionally, the use of multiple sensors would also be limited by considerations of aesthetics and functionality in open-plan offices. Hence, a simplified model-based control that can minimize indoor sensor use is preferred.

1.3. Contributions
To overcome the drawbacks of the sensor-based control strategies, we proposed a data-driven model predictive control method for the operation of a dynamic shading system. It can replace the use of multiple indoor sensors with one or two outdoor sensors while providing sufficient information for responsive dynamic shading control. Pre-simulated daylighting data is used to develop machine learning predictive models. With several input variables that are measured or collected in real-time, the control algorithm can predict the illuminance level on the work plane and control the shading devices accordingly. The proposed method contributes to the development of sensor-minimization control for the shading system.

2. Methodology
Figure 1 shows the overall workflow of the proposed shading control method. First, we used an offline daylight simulation to generate the dataset for machine learning model training. Second, we applied a standard machine learning pipeline to train the predictive model to estimate the work plane illuminance according to real-time weather and time input. Finally, we used the climate-based daylight simulation to verify the use of the data-driven predictive models for real-time shading control.

2.1. Machine learning algorithm
We used classification models for the development of the shading control method. As there are only four input variables, classification is expected to perform better than regression. Specifically, three common machine learning classification models were selected, including the k-nearest neighbors (KNN), support vector machine (SVM), and random forest (RF). They were chosen because they are: widely used in previous studies; simple to implement and allows for quick prediction. The brief introduction to each algorithm was as follows.
2.1.1. **K-nearest neighbor**
KNN is a type of supervised ML algorithm that can be used for both classification and regression predictive problems. However, it is mainly used for classification predictive problems in applications. It is a typical non-parametric method, that does not It is a non-parametric approach that does not make assumptions about the distribution of the data it is modeling. It is an algorithm that classifies data points based on the points that are most similar to it. In particular, it searches for the K known samples that are closest to the unknown data point, i.e., its K nearest neighbors. The new sample is usually classified as the majority label of the K nearest neighbors. KNN is easy to use and quick to compute, especially useful for machine learning with a low-dimension dataset. The value of K is the only factor that determines the performance of a KNN classifier.

2.1.2. **Support Vector Machine**
SVM is a machine learning method that is also much popular for its relative simplicity and flexibility for addressing a range of classification problems [3]. It can be used for both regression and classification problems. This algorithm distinctively affords balanced predictive performance, even in studies where sample sizes may be limited. The goal of an SVM classifier is to find the best hyperplane that separates the two classes. The hyperplane can be linear or non-linear depending on the type of kernel. In this research, a radial basis function kernel was applied. The regularization parameter can have a significant impact on the performance of an SVM model, which needs to be specified.

2.1.3. **Random Forest**
RF consists of a large number of individual decision trees that operate as an ensemble. Each tree in the random forest gives a class prediction. The majority of the prediction is selected as the prediction of the RF classifier. As a result, the classifier gets a more accurate and stable prediction than a decision tree. Two important parameters that influence the performance of an RF classification model include the number of trees and the maximum depth of each tree.

2.2. **Description of the building model**
The MIT reference office building was used to verify the proposed method [4]. It is an 8.5 m×3.9 m daylit open-plan office. There is a window on the south-facing façade, equipped with an external venetian blind. Of the two workstations next to the window, the one facing the east was selected to demonstrate the application of the proposed shading control method. There are no surrounding buildings or other objects that might block sunlight from the building. It was assumed that the office was located in Pittsburgh, Pennsylvania, USA (latitude 40.4N, longitude 80W).

![Figure 2. The MIT reference office [4]](image-url)
2.3. Daylight simulation
We used the software Rhinoceros to create the geometry of the building model. Rhinoceros (Rhino) is a stand-alone, commercial NURBS-based 3D modelling tool [5]. DIVA for Grasshopper was applied to perform point-in-time illuminance simulation to obtain hourly illuminance on the work plane. DIVA is an environmental analysis plugin in both Rhino and Grasshopper [6]. The simulation was conducted for the occupied hours (8 am to 6 pm) yearly. Hourly average illuminance on the work plane was obtained and used to construct the dataset for machine learning model training. The training dataset was derived from daylight simulation using typical meteorological year (TMY) weather. The simulation using historical year’s weather (the year 2018 in this case study) was also conducted to verify the presented shading control strategy.

The material properties used in the daylight simulation model are listed in Table 1.

| Material properties in the daylight simulation model |
|-----------------------------------------------|
| **Opaque Material** | **Reflectance** |
| Ceiling | 0.7 |
| Wall | 0.7 |
| Floor | 0.2 |
| Blinds | 0.8 |
| Desk | 0.5 |
| Partition | 0.5 |
| **Glazing Material** | **Transmissivity** |
| Window glaze | 0.96 |
| PC monitor | 1.39 |

2.4. Machine learning model training
The process for machine learning model training and validating is illustrated in Figure 3. It follows a standard machine learning process. The description of each step is as follows.

- **Data processing:** the hourly illuminance data was processed for training a classification model. A threshold of 2000 lux was used to label the data. Specifically, we labelled the outcome as 1 when the illuminance is higher than 2000 lux; otherwise, the outcome variable was labelled as 0. As a result, the obtained dataset was imbalanced, with more “1” cases than “0” cases. The performance of a machine learning model trained using such a dataset would lead to poor performance on the minority class, i.e., the cases with illuminance above 2000 lux. To deal with this problem, the data processing technique - oversampling the minority class was employed [7]. In particular, the simplest oversampling method that duplicates samples in the minority class was used. This is called the Synthetic Minority Oversampling Technique.

- **Model selection:** Four model input variables were selected: global solar radiation and direct normal radiation to characterize the sky condition, as well as altitude angle and azimuth angle to identify the sun position. These variables were selected according to previous relevant studies on using machine learning to predict daylighting [8,9]. The accuracy of the predictive model was used to determine the best model.

- **Model training:** the annual illuminance data derived from daylight simulation with TMY weather was used to train and validate the model. This process was conducted using cross-validation. The common 10-Fold Cross-Validation was used in this study. Specifically, the original dataset was split into 10 subgroups. Nine of them were used to train the model and the other subgroup was used to validate the model. This process was repeated 10 times. In this way, each of the 10 subgroups was used as the validation dataset for one time.
Figure 3. The workflow of training and validating the machine learning model

Hyperparameter tuning: in machine learning, a hyperparameter is a pre-defined parameter that configures the model training process. It needs to be initialized before model training and is not derived from the dataset. It can have a significant impact on the performance of the model. To optimize the trained machine learning model to have the best prediction, the hyperparameter of each classification model was tuned. A grid search technique was applied. The hyperparameters to be tuned and the range for each model are listed in Table 2.

Prediction: the dataset from daylight simulation with the year 2018 weather was used to test the final performance of the predictive model. The prediction accuracy was used as the metric to evaluate the performance of the predictive models.

The shading control logic: In this study, a simple on/off control was applied to adjust the venetian blind. This means the shades are either fully open or fully closed. Specifically, when the machine learning model predicts that the work plane illuminance is above 2000 lux, the building automation system will fully close the shades; otherwise, the shading devices will be kept fully open. The control algorithm is shown in Figure 4.

| Model | KNN | SVM | RF |
|-------|-----|-----|----|
| Hyperparameters | Number of neighbors $K \in [10, 50]$ | Regularization parameter $C \in [1, 20]$ | Number of trees $N \in [1, 20]$; maximum depth of each tree $d \in [1, 6]$ |
3. Result

3.1. Model accuracy
Table 3 summarizes the accuracy of different machine learning models with the tuned hyperparameters. It can be noted that all three models had a satisfying prediction accuracy, varying from 0.84 to 0.92. In particular, the RF model had the highest accuracy, followed by SVM and KNN. However, the variation was not significant.

| Model | KNN | SVM | RF |
|-------|-----|-----|----|
| **Best hyperparameter** | Number of neighbors $K = 5$ | Regularization parameter $C = 15$ | Number of trees $N = 16$; maximum depth of each tree $d = 5$ |
| **Accuracy** | 0.85 | 0.90 | 0.92 |

3.2. Annual work plan illuminance
In this study, the work plane illuminance was controlled. If it was above 2000 lux, there might be glare risk. An annual analysis showed that the proposed control method was capable of maintaining the work plane illuminance within 2000 lux for 97% to 99% of the time during the year, suggesting that the proposed method exhibited excellent capability in preventing daylight glare.

![Figure 5. The percent of the time when work illuminance is below 2000 lux during the year](image)

3.3. Lighting energy saving
The potential of the proposed shading control method to reduce electric lighting use was analysed. It was assumed that the shading control was integrated with lighting control and electric lights were switched off when the work plane illuminance was above 500 lux. The default lighting operation is an on/off control that switches the lights based on office hours, i.e., electric lights are kept on during office hours (8 am to 6 pm). As shown in Figure 6, electric lighting energy use can be reduced by 39.5% to 43.4% when lighting control is integrated with the proposed shading control strategy.
4. Discussion
This study presented a simplified model-based method for the control of an automated shading system. With the use of machine learning predictive models plus a few real-time input variables, the indoor illuminance sensor was eliminated. The evaluation of the daylight level can be more accurate than depending on physical sensors since average illuminance was used to represent the brightness on the work plane. It also avoids problems with indoor sensor placement, such as the entailed cost and disturbance to occupants, especially in office buildings. The proposed method is particularly to control shades in an open-plan office. In this case, one or two outdoor solar radiation sensors are sufficient to provide real-time data for the control of shades on facades facing different orientations.

Although demonstrated for the simple on/off control of the venation blind, the presented method can be easily extended to develop more complex shading control methods. For instance, illuminance predictive models that include the slat angle of the blind can be developed so that the shades can be controlled by adjusting its slat tilt angle. As a result, increased lighting energy saving and improved view access can be achieved, compared to the simple on/off control in this study. Besides, the proposed method can be easily applied to control other types of shading devices.

The other advantage of the method is that the machine learning models can be retrained even after the control logic has been implemented. This allows for the adaption of the system to occupants’ individual preferences. This type of adaptive control could improve occupants’ satisfaction. Accordingly, the control algorithm will be less overridden by them and its potential in reducing lighting energy use will be better utilized.

5. Conclusion and future work
In this research, a machine learning-based model predictive control method was proposed for the dynamic shading control. Three common classification algorithms were used to train the illuminance predictive model, including KNN, SVM, and RF. The obtained model was then used to control the states of the shades to prevent glare while allowing for daylight penetration. Real-time measurements of outdoor solar radiation and the sun position were required as the input variables for the model to predict illuminance condition so that the control logic can control the shades according to the prediction. The presented method was validated using climate-based daylight simulation in an open-plan office, to operate the automated shading device. It was found that the data-driven models had a satisfying prediction accuracy, capable of maintaining the work plane illuminance below 2000 lux for 97% to 99% of the time, depending on the machine learning algorithm used. When integrated with electric lighting control, the proposed shading control method had the potential in reducing lighting energy use by 39.5% to 43.4%. The result indicates that data-driven predictive models can replace the use of indoor illuminance sensors. The presented method can not only reduce the lighting energy consumption and...
labor as well as the cost associated with sensor installation but also provide sufficient natural light that benefits the occupants. It is a promising strategy for sensor-minimization shading control.

Future research to experimental validates the proposed control method is required to further support the findings of this study. Additionally, visual comfort will be better considered in the control algorithm. For instance, the glare index such as Daylight Glare Probability will be used as the constraint for visual comfort and incorporated into the control algorithm. Besides, the presented method needs to be validated to control shades facing different orientations and at various locations with different weather conditions. In particular, this study applied the method with a cloudy climate, future research to verify it with a sunny climate is required. More advanced control that can adjust the slat angle of venation blinds or the height of roller shades needs to be investigated in future studies.

Acknowledgments
At the end, for this research I would like to thank my Professor Stefan AI for his help. I would also like to thank Ms Xie for helping me complete this article. Finally, I would also like to thank all other teachers for their help and guidance in this research.

References
[1] Jamrozik, A., Clements, N., Hasan, S.S., Zhao, J., Zhang, R., Campanella, C., et al. (2019) Access to daylight and view in an office improves cognitive performance and satisfaction and reduces eyestrain: A controlled crossover study. Building and Environment, 165 (August): 106379.
[2] Karlsen, L., Heiselberg, P., Bryn, I., Johra, H. (2016) Solar shading control strategy for office buildings in cold climate. Energy and Buildings, 118: 316–328.
[3] Derek A. Pisner, David M. Schnyer (2020) Support vector machine. In: Mechelli, A., Vieira, S. (Eds.), Machine Learning, Academic Press, Cambridge. pp. 101-121.
[4] Christoph Reinhart, Alstan Jakubiec and Diego Ibarra. (2013) MIT Reference Office. http://web.mit.edu/sustainedesignlab/projects/ReferenceOffice/index.html
[5] McNeel R and A. Rhinoceros Version 4.0, Service Release 8. 2010.
[6] Reinhart, C.F., Lagios, K., Niemasz J, Jakubiec, A. DIVA for Rhino Version 2.0. 2011.
[7] Jason Brownlee. (2020) SMOTE for Imbalanced Classification with Python. https://machinelearningmastery.com/smote-oversampling-for-imbalanced-classification/
[8] Hu, J., Olbina, S. (2011) Illuminance-based slat angle selection model for automated control of split blinds. Building and Environment, 46(3):786–96.
[9] Luo, Z., Sun, C., Dong, Q. (2020) A daylight-linked shading strategy for automated blinds based on model-based control and Radial Basis Function (RBF) optimization. Building and Environment, 177(March):106854.