Useful Daylight Illuminance Prediction Under Data Imbalance in an Urban Context

Ilkim Canli[1], Sinan Kalkan[1], Ipek Gursel Dino[1]
[1]Middle East Technical University
[1]{ilkim.canli|skalkan|ipekg}@metu.edu.tr

Optimal daylight illumination can aid sustainable design by improving occupants' psychological and physical health, visual and thermal comfort and decreasing electrical lighting energy usage in buildings. However, dense urban areas can result in restricted daylight access in buildings. Therefore, daylight analysis considering surrounding buildings is important for implementing daylighting strategies. Useful Daylight Illuminance (UDI) is a performance metric that can quantify the annual illuminance levels within certain illumination classes (UDIfell-short, UDIsupplementary, UDIautonomous, and UDIexceeded). UDI can be predicted using machine-learning (ML) methods. However, the calculated data is typically unevenly distributed, generally following a power-law distribution, which causes ML models to underperform for UDI classes with less data. Simulations can be utilized to increase the less dispersed data in the dataset; however, at the urban scale, the computational cost of collecting simulation data for daylighting analysis makes it difficult to augment data with simulations. To undertake this challenge, in this study, SMOTE (Synthetic Minority Oversampling Technique) was applied to augment data to increase the prediction performance of the ML model. The results showed that augmenting the data in the classes which are unevenly distributed leads to an increase in ML model prediction performance. This method shows that SMOTE can be used to increase the performance of ML models during UDI estimation at the urban scale.

Keywords: *Daylight Illumination, Machine Learning Prediction, Useful Daylight Illuminance, Data Imbalance*

INTRODUCTION

In developed and developing countries, a significant amount of the energy is used for lighting, accounting for 11% of total building energy consumption (Department of Energy, 2015). Utilizing daylighting in buildings reduces electricity consumption used for artificial lighting and the interior load of the buildings. Besides energy savings, daylighting has non-visual and visual effects on occupants. The light-dark cycle (24-h rhythms) significantly impacts health and well-being, which dominates many areas of human psychology and behavior (Webb, 2006). Daylight is also effective in users' creativity and productivity. Good daylight design can provide sufficient daylight for effective visual performance while ensuring adequate comfort. To fully realize daylight's benefits, it is important to use it strategically in buildings (Dogan and Park, 2019). However, rapid urbanization threatens sustainable solar access, reducing daylight availability in building interiors (Czachura, Gentile and Kanters, 2022). There is a need to comprehensively analyze daylighting performances of buildings located in an urban context.

Daylight performance metrics are an indicator for assessing a space's energy savings and daylight potential. The main performance metrics are 'Daylight Factor' (DF), 'Daylight Autonomy (DA)', 'Useful Daylight Illuminance (UDI)', 'Daylight Availability (DA)' and 'Annual Sunlight Exposure (ASE)'.

Several daylighting calculation methods have been proposed, including graphical and non-graphical methods, analytical formulas, scale models, computer simulations, and machine learning models (Ayoub, 2019). These methods are static, dynamic, climate-based, and machine-learning models. The static method applied Daylight Factor (DF), which was introduced by Trotter (1911) to assess daylighting. The components of DF can be calculated using graphical methods. However, the static method is limited as they do not consider different sky conditions and space orientation. Later on, Climate-Based Daylight Modeling (CBDM) was introduced based on the DC method (Mardaljevic, 2000). Using physical-based daylighting simulation tools, CBDM quantifies absolute illuminance values on the sensor grid, considering the different sky luminance distributions taken from weather data at certain temporal resolutions (Mardaljevic, 2000).

However, daylight simulations have several challenges, including the laborious process of 3D modeling and the high computational cost of simulations (Ayoub, 2020). Moreover, daylight simulations require an extensive and complex set of inputs, including weather data, building geometry, material properties of building elements, and surrounding buildings (Ayoub, 2020).

Alternatively, machine learning (ML) models can perform comprehensive daylight analysis by reducing computational costs and eliminating the need for a 3D model (Ayoub, 2020). Based on the training dataset, the models learn the relationship between inputs and output. ANN (Artificial Neural Network), DT (Decision Tree), SVM (Support Vector Machine), MLR (Multilinear Regression), and RF (Random Forest) algorithms are widely used in daylighting predictions.

Several studies utilized ML models to predict various daylighting performance metrics. Le-Thanh et al. (2022) developed an ANN model to predict four ranges of UDI in different building layouts. Ngarambe et al. (2020) compared the performances of different ML models (generalized linear models, deep neural networks, random forest, and gradient boosting) to predict daylight illuminances in indoor spaces. Nourkojouri et al. (2021) developed an ML-based framework for assessing daylight and visual comfort in early design stages, predicting UDI, SDA, mean daylight autonomy, annual sunlight exposure (ASE), and spatial visual discomfort metrics. Hin et al. (2023) utilized ANN and SVM to forecast vertical daylight illuminance indicating crucial variables for prediction. Han et al. (2021) developed an ANN model to predict UDI subclasses and Daylight Autonomy (DA) for analyzing the daylighting performance of office buildings in the early design stages.

The utilized ML models in these studies show high-performance accuracy. However, as these studies focused on a single building, the daylight data obtained or collected showed a balanced distribution. However, due to the shading from the surrounding buildings in dense urban areas, the hourly illumination level is low most of the year. Accordingly, the illuminance data shows an imbalance. Data distribution can be frequently skewed as some classes' representatives are significantly more prevalent, called 'data imbalance' (Krawczyk, 2016). In such cases, learning will be compromised as ML algorithms will be biased in favor of the majority group. In daylighting prediction, the data imbalance problem will result in the failure to make predictions in cases of very high illumination levels, which is the minority class.

To tackle the data imbalance problem, the data size of the minority class can be increased by generating new data (Ali *et al.*, 2019). The synthetic minority over-sampling technique (SMOTE) is one of the most preferred techniques for creating new data by interpolating minority class data that reside together. This study proposes an ML-based framework to predict subclasses of Useful Daylight Illuminance, considering the data imbalance problem and potential solution for handling class imbalance problems utilizing SMOTE. The novelties of the work are as follows:

- Identifying the 'class imbalance' problem in the daylighting analysis of buildings in highly shaded urban settings
- Using SMOTE to augment ML-based daylighting estimation for the first time

Our results show that oversampling techniques that generate synthetic minority class samples can be used to improve the ML-based prediction performance during daylighting illuminance assessment.

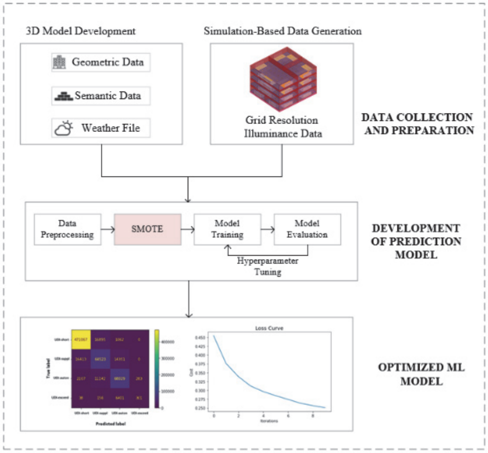

Figure 1
The methodology of the work

METHODOLOGY

This study introduces an ML-based method for predicting subclasses of Useful Daylight Illuminance (UDI) utilizing SMOTE (Synthetic Minority Oversampling Technique) to tackle the data imbalance problem. The method is based on two main steps (i) Data collection (3D model development and simulation-based data generation) and (ii) development of an ML-based model. Figure 1 shows the overall workflow of the study. One residential building with 13 living units in Ankara, Bahçelievler, is selected for modeling.

Data Collection and Preparation

This section presents the data collection and preparation steps: 3D model development and simulation-based data generation. A 3D model was developed using the Rhino environment, and simulations were conducted using the Honeybee and RADIANCE plugins for Rhino.

3D Model Development. Daylighting simulations will be performed on the zone level, which will be modeled as closed breps (3D shapes that define the limits of a volume) and an analysis point grid placed in this zone. To model these zone, it is first necessary to generate a building mass and then define each zone within this mass. Windows in each zone are either modeled as geometric entities (i.e., surfaces on the outside walls) or semantic data regarding window-to-wall ratios (WWR). Window visible transmittance (VT) values are set to 0.43. Daylight simulations also require climate data for utilizing global horizontal illuminance (GHI), direct normal illuminance (DNI), and diffuse horizontal illuminance (DHI). The typical meteorological data (.EPW) of Ankara, Central, was used. Figure 2 shows the selected building and its zones and surrounding buildings.

Simulation-Based Data Generation. After the 3D model was created, each unit was divided into a 2m x 2m grid to calculate hourly illuminances for each grid. In RADIANCE, ambient bounces and super-samples were set to 2 and 256, respectively to increase the accuracy of the results.
The aim is to calculate each grid point's hourly illuminance annually and convert these illuminances to UDI subclasses. Simulations were also utilized to calculate 'Sky Exposure'. This parameter indicates the visible sky ratio considering surrounding buildings as seen from the center of the windows positioned in each zone's four directions. A Sky

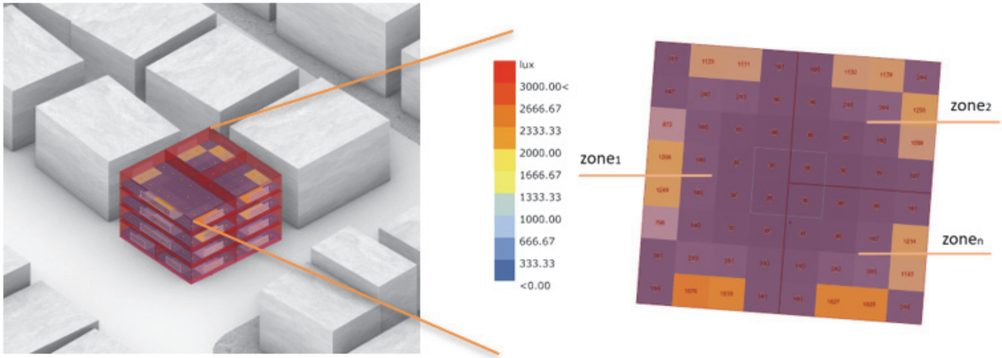

Figure 2
Simulated building and its zones

Exposure of 0 indicates that the zone is completely blocked from surroundings, while a sky exposure of 100 refers to the zone's illumination not being affected by the surroundings. To consider the effects of surrounding buildings, Sky Exposure (SE) values were given to the ML as input.

After calculating hourly illuminance values of each grid point annually using daylight simulations, input parameters for the ML model, including aspect ratio, total window/total floor area, and the distance of each grid points to the windows (wd), were recorded. Output data for the ML model is Useful Daylight Illuminance (UDI), which is calculated using the daylight simulation results. UDI is defined as the annual occurrence of illuminances throughout the work plane that is all within the range of 100/2000 lux (Nabil and Mardaljevic, 2005) or 100/3000 lux (Mardaljevic *et al.*, 2012). Alternatively, four UDI ranges were introduced by (Mardaljevic *et al.*, 2012) as (i) UDIfell-short (UDI-f) (<100 lux), (ii) UDIsupplementary (UDI-s) (≥100 lux and ≤300 lux) , and (iii) UDIautonomous (UDI-a) (>300 lux and <3000 lux) and (iv) UDIexceeded (UDI-e) (≥3000 lux). In the study, the four ranges of UDI were chosen as output classes for the ML model, as those metrics can accurately reflect the daylight status of spaces.

Development of the Prediction Model

The development of the prediction model is based on four main steps that are: (i) data preprocessing, (ii) model training, (iii) model evaluation, and (iv) hyperparameter tuning.

Data Preprocessing. For the ML model, inputs were collected from different sources. The obtained hourly illuminance results were converted to the discrete category labels (classes) based on UDI categories (given in Section Simulation-Based Data Generation). Table 1 shows the description and statistical values of inputs utilized in the ML model.

Figure 3 shows the distribution of samples for the UDI classes. Here, it was observed that most of the illuminance values belong to the 'UDI-short.' The dataset contains only a small amount of 'UDI-exceed' data. As seen from the plot, the dataset has a significant imbalance problem, which can cause an ML model to make poor predictions. Therefore, to tackle the data imbalance problem, SMOTE was applied.

SMOTE (Synthetic Minority Oversampling Technique). Data imbalance stemming from the skewed nature of data is a common problem impacting ML models' performances negatively (Ali et al., 2019). In data imbalance, majority classes bias the classifiers, and ML classifiers are optimized towards mostly the majority class ignoring the minority class. SMOTE is a technique utilized to handle data imbalance problems (Chawla et al., 2002). In SMOTE, the samples are increased by

Table 1
Description and statistical values of ML inputs

Name	Description	Unit	Source	Min	Max
floor	Indication of floor	-	Google Images	0	3
Aspect ratio	The ratio between the total window area and total floor area of units	-	3D model	0.54	1.09
Total Window/ Floor Area	The ratio between the total window area and total floor area of units	-	3D model	0.12	0.25
SENorth	Sky Exposure in north	%	Simulation	0	87.92
SEWest	Sky Exposure in west	%	Simulation	0	88.98
SESouth	Sky Exposure in south	%	Simulation	0	87.98
SEEast	Sky Exposure in east	%	Simulation	0	88.86
wdNorth	Distances of each grid to the windows in north	m	3D model	1.7	15.74
wdWest	Distances of each grid to the windows in west	m	3D model	1.7	10.54
wdSouth	Distances of each grid to the windows in south	m	3D model	1.7	15.74
wdEast	Distances of each grid to the windows in east	m	3D model	1.7	10.54
GHI	Global Horizontal Illuminance	lux	Climate file	0	121992

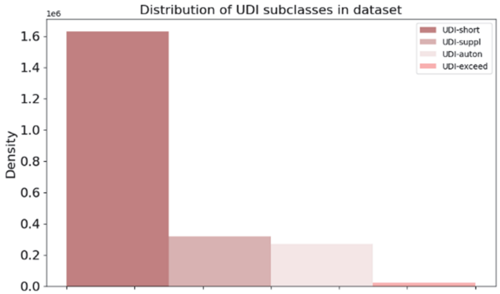

Figure 3
The distribution of samples for the UDI classes

generating new samples. This is achieved by taking each minority class sample and generating synthetic examples along the line segments joining any/all of the k minority class nearest neighbors. In the study, especially, SMOTE was applied to increase the number of data in all three classes except 'UDI-short'. SMOTE was chosen instead of other oversampling methods as, unlike random sampling, SMOTE algorithm creates synthetic data based on the similarities between existing data in minority classes (Chawla et al., 2002).

After data was augmented with SMOTE, the dataset was split between the training (0.70) and testing data (0.30). Training data were scaled using Standard Scaler from the sklearn library. Then inputs and outputs were given to the ML model.

Machine Learning Model. Artificial Neural Networks (ANNs) are inspired by how neurons in the brain work. By learning from patterns in data and generalizing the patterns for unforeseen circumstances, ANNs handle linear and non-linear functions. Multilayer Perceptron (MLP) is one of the most popular ANNs (Camacho Olmedo et al., 2018). MLP is a feed-forward neural network consisting of three layers: the input, output, and hidden layers. The input layer is where the input signal for processing is received. The hidden layers are the computational engine of the MLP. In MLP, the data moves from the input to the output layer. Backpropagation algorithms are used to train the neurons. MLPs can resolve issues that are not linearly separable because they are made to approximate any continuous function.

Given $i = 0,1, \ldots, n$ where n is the number of inputs, the quantities $\{w_i\}$ are the weights of the neuron. The output y is the predictive class of the inputs x_i, which represent features or variables. Each input feature value is multiplied by its weight $(w_i\, x_i)$ in the weighting phase and then added together $(w_0\, x_0 + w_1\, x_1 + \cdots + w_n x_n)$ and in the second step, an activation function (f; also known as a transfer function) is applied to the step. The third is the output sum of the transfer that is represented as (1) (Taud and Mas, 2018):

$$y = f(z)\ and\ z = \sum_{i=0}^{n} w_i\, x_i \qquad (1)$$

where $x_0 = 1$, w_0 is the threshold or bias, and y is the output.

Model Training and Tuning. Initially, our MLP model consists of two hidden layers with (32, 32) neurons, ‘max_iteration’ = 10 (number of epoch), activation was chosen as ‘ReLU’, and solver was chosen as ‘Adam’. ‘ReLU’ is a non-linear activation function that, if the input is positive, outputs the input directly; otherwise, it outputs zero (Sharma, Sharma and Athaiya, 2020). Solver is used for weight optimization, and ‘Adam’ refers to a stochastic gradient-based optimizer (Kingma and Ba, 2014). Following, hyperparameter tuning was performed. Five parameters were tuned, including ‘*hidden_layer_sizes*’, ‘*activation*’, ‘*solver*’, ‘*alpha*’, ‘*max_iter*’, and ‘*learning_rate_init*’. ‘*hidden_layer_sizes*’ represents the number of neurons in each hidden layer.

After hyperparameter tuning, models were evaluated based on performance metrics. Performance evaluation metrics assess how effectively ML models perform during training and testing, comparing the discrepancies between the predicted and real values (Ayoub, 2020). For classification tasks, four main metrics were utilized that are (i) accuracy, (ii) precision, (iii) recall (sensitivity), and (iv) F1score.

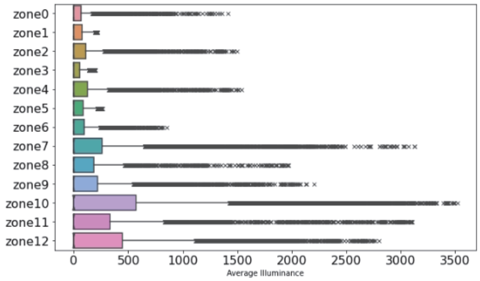

Figure 4
Average illuminances of each zones annually

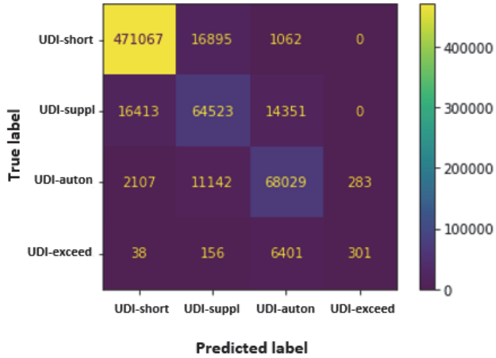

Figure 5
Baseline predictions of MLP

	precision	recall	F1-score	accuracy
UDI-short	0.96	0.96	0.96	0.90
UDI-suppl	0.7	0.68	0.69	
UDI-auton	0.76	0.83	0.79	
UDI-exceed	0.52	0.04	0.08	

Table 2
Performance metrics of baseline MLP model

RESULTS

The residential building with 13 units was simulated annually, and hourly illuminance values of each grid point in units were recorded. Figure 4 shows the average illuminance values of analysis grid points of the zones annually. Based on the Figure 4, as expected, the zones on the upper floors are more illuminated than those on the lower floors throughout the year. It has been observed that the zone, which is the most illuminated throughout the year, is illuminated from the west and south facades on the top floor and that the illumination of the surrounding buildings on these facades is not blocked. The least illuminated building throughout the year is a zone on the ground floor that receives light from the north and east facades, but the surrounding buildings significantly block the light coming to these facades. When all the zones were examined, it was observed that the zones were illuminated in a low amount for most of the year.

Baseline Results

Initially, the MLP model was trained with the initial dataset without SMOTE. This initial dataset was imbalanced, meaning the data belonging to ‘UDI-short’ was very large, and data belonging to ‘UDI-

Table 3
Performance metrics of MLP model after SMOTE implementation

	precision	recall	F1-score	accuracy
UDI-short	0.96	0.90	0.93	0.81
UDI-suppl	0.73	0.83	0.78	
UDI-auton	0.78	0.65	0.71	
UDI-exceed	0.80	0.88	0.84	

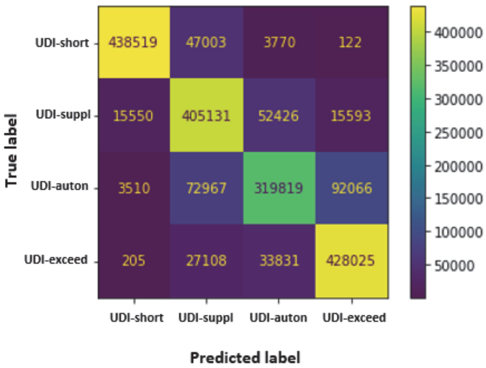

Figure 6
MLP predictions after SMOTE implementation

suppl', 'UDI-auton', and 'UDI-exceed' is comparatively scarce. In this case, the trained MLP model is expected to make the best prediction for the 'UDI-short' class, whereas the class that is expected to show the worst prediction is 'UDI-exceed' . The prediction performance of the model and the results of the training performance are given in Figure 5 and Table 2.

The results show that the MLP model performed with an overall accuracy of 0.90. However, a significant inequality was observed when the prediction performance of individual classes was examined. As expected, the MLP model predicted the most successful 'UDI-short' with 0.96 precision, recall, and F1 score values. The model showed poor estimation performance in the class with the least data (UDI-exceed), with 0.52 precision, 0.04 recall, and 0.08 F1 score. To tackle this class imbalance problem, SMOTE was applied.

The Effect of SMOTE

SMOTE was applied to increase the data of the underrepresented class in the dataset, thus increasing the MLP model's performance. It is expected that the performance of the MLP model will increase in all classes where the number of data is increased after the SMOTE application. Figure 6 and Table 3 show the prediction performances of the model in each class.

As observed from Figure 6 and Table 3, the overall accuracy of the MLP model decreased (0.81) after SMOTE application. However, when each class is examined separately, the estimation performance of classes with increased data ('UDI-suppl', 'UDI-auton', 'UDI-exceed') has increased significantly, as expected. 'UDI-exceed' estimation performance was the class that increased the most. Before the SMOTE implementation, the MLP model could make almost no estimation of 'UDI-exceed', while with the SMOTE implementation, the MLP model can make a high estimation of 'UDI-exceed' with 0.8 precision, 0.88 recall, and 0.84 F1 score value.

The only class where the model's prediction performance did not improve but slightly decreased after SMOTE implementation was 'UDI-short'. While 'UDI-short' was the most representative class in the dataset in the previous case, after the SMOTE implementation, the representation of 'UDI-short' in the dataset according to the initial state is equalized with other classes, causing the success of the MLP model to predict this class to decrease slightly. Still, 'UDI-short' was the class with the highest performance, with a precision of 0.96, a recall of 0.90, and an F1 score of 0.93.

The Effect of Hyperparameters

Table 4 shows the selected values for hyperparameters. The model with the highest performance were selected as the final model due to hyperparameter tuning, and their performances were compared.

After hyperparameter tuning, the prediction performances of all classes increased. The model's accuracy increased from 0.81 to 0.85. The numerical values that the model predicts the classes correctly due to hyperparameter tuning are given in Figure 7. Compared to the previous estimation results, 'UDI-auton' was the class in which the values were predicted correctly by the model increased the most as a result of hyperparameter tuning. The class where the number of correct guesses increased the least was 'UDI-suppl'.

Table 5 shows the performance values in estimating each class of the model as a result of hyperparameter tuning. According to the table, the class with the highest increase in recall and f1 score values was 'UDI-auton', while the precision value increased the most in 'UDI-suppl'. Overall, hyperparameter tuning has significantly increased the predictive performance of each model class.

CONCLUSION AND DISCUSSION

In this study, we address the class imbalance problems that we observed during the ML-based prediction of daylighting illumination in buildings. Class imbalance occurs due to the uneven distribution of the dataset, seriously threatening prediction performance. We proposed the use of SMOTE to augment the data in minority classes to improve the performances of the MLP model. In order to demonstrate our approach, an MLP model was first developed to predict hourly UDI subclasses considering surrounding buildings. MLP classifier model was trained with simulation-based data. Although the initially trained model showed a high accuracy of 0.90, considering the estimation performance of each class, its prediction performance of the 'UDI-exceed' class (precision = 0.52, recall = 0.04, F1-score = 0.08), which was the minority class, was considerably lower. SMOTE was applied to increase the model's prediction performance on all classes. The MLP model was retrained and hyper-tuned with augmented data. Although a decrease was observed in the model's overall performance after applying SMOTE, the prediction performance increased significantly on the three classes (UDI-suppl, UDI-auton, UDI-exceed).

Table 4
Hyperparameters

Model	Description
MLP Classifier	Hidden_layer_sizes = [32,32], [32,64], [64,32]**, [64, 64]
	activation = 'tanh', 'relu'**
	solver = 'sgd', 'adam'**
	alpha = '0.0001'**, '0.1'
	max_iteration = '10', '50'**
	learning_rate_init = '0.0001, 0.001**,0.01, 0.1'

** selected final parameters

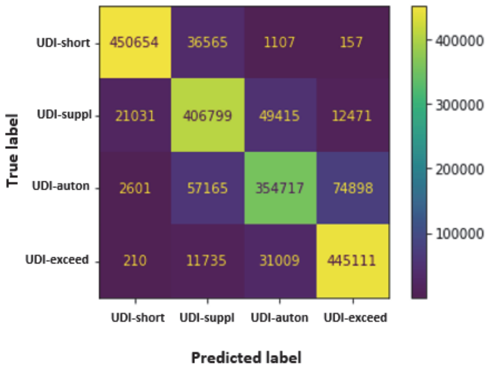

Figure 7
MLP predictions after hyperparameter tuning

Table 5
Performance metrics of MLP model after hyperparameter tuning

	precision	recall	F1-score	accuracy
UDI-short	0.95	0.92	0.94	0.85
UDI-suppl	0.79	0.83	0.81	
UDI-auton	0.81	0.72	0.77	
UDI-exceed	0.84	0.91	0.87	

This shows that evaluating the model's performance only with accuracy in classification problems may cause biased results because the majority class dominates the model. To evaluate the model's performance accurately, the predictive performance of each model class should be examined. As a result, this study showed that data imbalance problems can be effectively tackled with SMOTE during hourly daylight prediction. Our study is also the first application of SMOTE in daylighting predictions. In the future, different data augmentation techniques can be applied for other classification problems in which imbalance is observed.

ACKNOWLEDGMENTS

This study is funded by the Scientific and Technological Research Council of Turkey under grant number TUBİTAK 120M997.

REFERENCES

Ali, H. *et al.* (2019) 'Imbalance class problems in data mining: A review', *Indonesian Journal of Electrical Engineering and Computer Science*, 14(3), pp. 1552–1563. doi: 10.11591/ijeecs.v14.i3.pp1552-1563.

Ayoub, M. (2019) '100 Years of daylighting: A chronological review of daylight prediction and calculation methods', *Solar Energy*. Elsevier, 194(August), pp. 360–390. doi: 10.1016/j.solener.2019.10.072.

Ayoub, M. (2020) 'A review on machine learning algorithms to predict daylighting inside buildings', *Solar Energy*. Elsevier, 202(January), pp. 249–275. doi: 10.1016/j.solener.2020.03.104.

Camacho Olmedo, M. T. *et al.* (2018) *Geomatic Approaches for Modeling Land Change Scenarios. An Introduction*. doi: 10.1007/978-3-319-60801-3_1.

Chawla, N. V. *et al.* (2002) 'SMOTE: Synthetic Minority Over-sampling Technique', *Journal of Artificial Intelligence Research*, 16(Sept. 28), pp. 321–357. Available at: https://arxiv.org/pdf/1106.1813.pdf%0Ahttp://www.snopes.com/horrors/insects/telamonia.asp.

Czachura, A., Gentile, N. and Kanters, J. (2022) 'Identifying Potential Indicators of Neighbourhood Solar Access in Urban Planning'.

Department of Energy, U. (2015) *AN ASSESSMENT OF ENERGY TECHNOLOGIES AND RESEARCH OPPORTUNITIES Chapter 5: Increasing Efficiency of Building Systems and Technologies*.

Dogan, T. and Park, Y. C. (2019) 'A critical review of daylighting metrics for residential architecture and a new metric for cold and temperate climates', *Lighting Research and Technology*, 51(2), pp. 206–230. doi: 10.1177/1477153518755561.

Han, Y., Shen, L. and Sun, C. (2021) 'Developing a parametric morphable annual daylight prediction model with improved generalization capability for the early stages of office building design', *Building and Environment*. Elsevier Ltd, 200(April), p. 107932. doi: 10.1016/j.buildenv.2021.107932.

Hin, D., Li, W. and Imuetinyan, E. (2023) 'Predicting Vertical Daylight Illuminance Data From Measured Solar Irradiance : A Machine Learning-Based Luminous Efficacy Approach', 145(June), pp. 1–16. doi: 10.1115/1.4055915.

Kingma, D. P. and Ba, J. (2014) 'Adam: A Method for Stochastic Optimization'. Available at: http://arxiv.org/abs/1412.6980.

Krawczyk, B. (2016) 'Learning from imbalanced data: open challenges and future directions', *Progress in Artificial Intelligence*. Springer Verlag, pp. 221–232. doi: 10.1007/s13748-016-0094-0.

Le-Thanh, L. *et al.* (2022) 'Machine learning-based real-time daylight analysis in buildings', *Journal of Building Engineering*. Elsevier Ltd, 52(November 2021), p. 104374. doi: 10.1016/j.jobe.2022.104374.

Mardaljevic, J. (2000) 'Simulation of annual daylighting profiles for internal illuminance', *Lighting Research & Technology*, 32(3), pp. 111–118. doi: 10.1177/096032710003200302.

Mardaljevic, J. *et al.* (2012) 'DAYLIGHTING METRICS : IS THERE A RELATION BETWEEN USEFUL DAYLIGHT ILLUMINANCE AND DAYLIGHT GLARE PROBABILITY', *Proceedings of the Building Simulation and Optimization Conference BSO12*. Available at:

https://infoscience.epfl.ch/record/179939?ln=en.
Nabil, A. and Mardaljevic, J. (2005) 'Useful daylight illuminance: A new paradigm for assessing daylight in buildings', *Lighting Research and Technology*, 37(1), pp. 41–59. doi: 10.1191/1365782805li128oa.
Ngarambe, J. *et al.* (2020) 'Comparative performance of machine learning algorithms in the prediction of indoor daylight illuminances', *Sustainability (Switzerland)*, 12(11), pp. 1–23. doi: 10.3390/su12114471.
Nourkojouri, H. *et al.* (2021) 'Development of a Machine-Learning Framework for Overall Daylight and Visual Comfort Assessment in Early Design Stages', 8, pp. 270–283. doi: 10.15627/jd.2021.21.
Sharma, Siddharth, Sharma, Simone and Athaiya, A. (2020) *ACTIVATION FUNCTIONS IN NEURAL NETWORKS*, *International Journal of Engineering Applied Sciences and Technology*. Available at: http://www.ijeast.com.
Taud, H. and Mas, J. F. (2018) 'Multilayer Perceptron (MLP) Neural networks', pp. 451–455.
Trotter, A. (1911) 'Illumination; its distribution and measurement', *Nature*, (88), pp. 72–73.
Webb, A. R. (2006) 'Considerations for lighting in the built environment: Non-visual effects of light', *Energy and Buildings*, 38(7), pp. 721–727. doi: 10.1016/j.enbuild.2006.03.004.