Classification of the Energy Production Potential of Water Lens Solar Concentrators Using Machine Learning

Hamidreza Zarrinkafsh¹, Nasim Eslamirad², Francesco De Luca³
¹ School of Architecture, University of Tehran, Tehran, Iran
² Smart City Center of Excellence (Finest Twins) Tallinn University of Technology
³ Tallinn University of Technology
zarinkafsh@ut.ac.ir

Abstract. Assessing the potential of renewable energy sources for buildings in neighborhoods becomes a crucial task in the early planning stage. Integrating solar energy equipment into urban buildings poses many challenges, such as uncertainties and the complexity of urban built agglomeration. Due to the time-consuming solar energy potential assessment process and lack of knowledge of urban actors, a reliable framework is required to predict buildings' solar energy potential. This research presents a comprehensive machine learning data processing framework to predict output energy of Water Lenses (WL) based on buildings specifications and relationship to the neighbourhood. The research used a raw dataset consisting of 7000 sample buildings in different situations by applying 12 years of climatic conditions in Tallinn, Estonia. The results were entered into a Supervised Machine Learning process and the Gaussian Naive Bayes technique was used for classification of building features to be implemented with solar systems. Finally, the process was measured by a confusion matrix that showed 80% accuracy of ML output predictions in the urban context.

1. Introduction
Due to rapid urbanization and ever-increasing urban gas emissions, assessing the renewable energy potential for urban buildings becomes a crucial task in the early planning stage. Since energy is the most influential factor affecting other fundamental vectors of a smart city [1], predicting energy production in the long and short term is a significant part of the development of smart cities [2],[3]. However, because urban planners, architects, and public authorities often lack the knowledge to assess the renewable energy potential, a reliable framework is required to predict and estimate the energy in smart cities by machine learning [4].

Solar energy is among the cleanest renewable energy resources in terms of carbon emissions, but integrating solar technology into urban buildings poses many challenges [5], such as uncertainties on the environmental conditions, the interactions between solar energy and urban morphology [6], and the complexity of urban built agglomeration. At high latitudes like in Nordic cities, this issue is particularly challenging due to low annual solar radiation. Since the process of modeling, calculation by physics of light equations, and simulation by ray-tracing software for assessing solar system output is time-consuming and needs training, this research proposes a comprehensive data processing framework for the application of machine learning (ML) for the assessment of Water Lens solar concentrators at the scale of buildings and neighborhoods [7]. Water lenses are highly efficient solar energy systems with less environmental impact compared to other solar concentrators; they do not require expensive solar tracking systems and can benefit from direct and diffuse radiations; thus, in case of being in a neighbor building’s shadow in cities, they can still be productive. This work considers their integration in urban buildings, and the output energy is calculated by physics of light equations [8] and ray-tracing simulation in the TracePro software [9].
There are many studies on predicting solar energy by machine learning. To mention some examples, Kim et al. used a machine learning approach for solar power generation prediction based on weather data [10]. Jakubiec et al. created a method to predict city-wide electricity gains from PV combined with online mapping technologies [11]. A study used a combination of ML and geographic information systems (GIS) to estimate the rooftop solar potential for urban areas in Switzerland [12], and Sharma, Navin, et al. built a machine learning model to predict solar generation [13]. Moreover, Walch, Alina, et al. combined Machine Learning, GIS, and physical models to build a methodology to estimate the individual roof PV potential at temporal resolution [14]. All of these works focused on PV or other solar energy systems, but none assesses the potential of liquid lenses on roof top surfaces.

In recent years, some interesting approaches for utilizing liquid lenses have been shown to significantly increase efficiency and reduce the required surface area of photovoltaic (PV) modules [15]. An experimental study of water lenses for photovoltaic concentrating systems shows they can increase a wide range of solar systems’ efficiency [16] and can be employed for cogeneration systems to provide electrical and thermal energy [17]. The fundamental advantages of this type of lens are described in [18], and the surface equation of a liquid lens is introduced in [19].

The modeling results in this study were processed through an ML method to predict the maximum output energy of a rooftop WL. The building's specification, relationship to the neighborhood, and solar radiation were entered into the machine learning model and, based on learning patterns of similarly rated buildings for 12 years, the maximum solar energy potential was predicted. The algorithms of Supervised Machine Learning (SML) and various accuracy metrics were used. This research is one of the prime studies on predicting the output energy of integrated WL into urban buildings on the scale of a city in Tallinn. Urban actors can use this data to apply support planning tools, design infrastructures, and guidelines as decision support tools for the preliminary design in sustainable smart cities. The process can also be repeated for different latitudes and cities on temporal and spatial scales.

2. Methodology

In this study, a supervised machine learning (SML) method is employed to achieve a high accuracy rate in predicting the maximum output energy of the integrated WL into different models of buildings in Tallinn, Estonia accrued on June 5th.

According to the graphical representation of Figure 1, the study was organized in two main phases. The first one was modeling and obtaining the output energy of the WL in different situations. The second phase consisted of data collection, data acquisition, data pre-processing, creating the ML model, classification by the Gaussian Naive classifier [20], and finally evaluating the prediction accuracy by the Confusion Matrix measurement method. Given the nature of the data and the dynamics behind the
factors involved in the feasibility study of solar collectors, in this research, the variables and results of a former study in the Tallinn climatic zone were used. The hourly output energy of the WL was obtained through physics of light equations and ray-tracing simulation in the TracePro software, and the results present just a 6.6% difference for the period of April to September. For this reason, the study employed physics of light equations for hourly radiation records of 12 years. To define models with logical properties, data shuffling is applied.

2.1. The context of the study. The city of Tallinn (Lat. 59°26’N Lon. 24°45’E), Estonia, was selected as a case study for three reasons. Firstly, as the former study showed, at high latitudes such as that of Tallinn, water lenses are more efficient and effective than other solar collectors. Secondly, this study capitalizes on existing building morphology studies for the city of Tallinn conducted by the authors [21]. Thirdly, since Tallinn is implementing an innovative smart city concept, the results can be helpful for the development of smart energy concepts and other Nordic cities with the same latitude and climate.

Approximately 7000 sample buildings are considered with different roof slope and situation in the neighborhood. The maximum output energy of the WL was obtained based on the probability of being in a neighboring buildings’ shadow or positioning influenced by shadowing of a sloped roof. The building’s specification and relationship to the neighborhood was entered as inputs in the machine learning model and based on learning patterns of similarly rated buildings, the maximum daily output energy of the WL system was predicted.

Since ML algorithms require precision, accuracy and minimum error [22], the most trustable method for creating and evaluating an ML model is employed.

2.2. Data Acquisition. The study’s data was collected from GIS, urban analysis, solar cadastre, and weather data. The raw dataset’s labelled features directly came from a former study’s variables and results. The variables were defined in two main scales, building and neighbourhood, including zone, roof slope, buildings’ relative orientation and height to width ratio between buildings. The height to the distance ratio between buildings affects the amount of shortwave radiation absorbed by the buildings [23], and the solar energy system. Table 1 shows the association of building orientation and the height to width ratio between buildings. Each floor height is 3 m, the buildings’ height difference can vary from 0 m to 9 m, and the distance between the buildings varies from 3 m to 9 m with increments every 3 m. Roof slope is defined according to the face of the roof that is mainly exposed to the sun radiation and four kinds of roof slope have been considered as horizontal, east facing, south-facing, and west-facing. Since north-facing roofs are not exposed to much solar radiation, they were not considered.

As Figure 3 shows, the proposed date and time for evaluating the output energy of the water lens is June 5th, because this date marks the peak of solar energy generation in Tallinn. The hourly direct radiation Gb(i) and diffuse radiation Gd(i) between 2005 to 2016 for Tallinn city have been collected.
from EU Science Hub [24]. When the WL is in the shadows, just diffuse radiation impacts the output energy.

2.3. **Pre-processing**

In the pre-processing step, all numerical features in the raw dataset were converted to the categorical classes and imported in CSV format into the Python environment. Then the dataset was implemented in the Python environment using pandas, Numpy, seaborn, matplotlib, and scipy libraries. The categorical features replaced the numerical rates of radiation and WL output energy. It means that the sum of radiation (direct radiation Gb(i) and diffuse radiation Gd(i)) classified into nine groups, ranges from 1 for less than 100 W/m², 2 for 200-300 W/m², to 9 for more than 800 W/m². Similarly, the output energy of the solar system is labeled as “output 1”, “output 2”, and “output3”, respectively, for less than 300W/m², 300-600 W/m², and 600-800 W/m² results.

3. **Results**

3.1. **Creating the ML model and Classification**

After importing the dataset and converting it to a data frame by the Pandas library, the classification is applied and features and target introduced as (X) and (Y), respectively. Figure 3 shows the number of the instances in each class of target in the initial dataset, as Output 1: 3742, Output 2: 1647, and Output 3: 1523 instances.

For splitting the dataset into train and test subsets, Python libraries, pandas, and sklearn were used. Figure 4 shows which size of tested and trained data allowed to realize a model of the learning process with high accuracy. As it is shown in the learning model, even just the 20 % of the test size is enough for high accuracy.

After the learning process, GaussianNB is applied to fit, predict and to test the ML classifier. Gaussian Naive Bayes was based on Bayes’ Theorem under supervised learning algorithms and strongly assumes that predictors should be independent of each other, given the class variable. Naive Bayes is a family of algorithms with higher accuracy and speed for large data points and constructs classifiers. Classifiers are models that assign class labels to problem instances, represented as vectors of feature values, where the class labels are drawn from some finite set. Here when the dataset is split into 60% train to 40% test sizes, the accuracy of the classifier is around 80%, which means a great result of the learning and classification process. Figure 6 shows how Gaussian Naive Bayes classified the study sample with about 80% accuracy with considering the target in 3 categories vs. the total solar radiation in 9 samples on June 5th.
3.2. Accuracy Measurement, Confusion Matrix

A confusion matrix is a tabular summary of the number of correct and incorrect predictions made by a classifier to measure the performance of a classification model through the calculation of performance metrics such as accuracy, precision, recall, and F1-score.

| precision | recall | f1-score | support |
|-----------|--------|----------|---------|
| 1         | 0.91   | 0.74     | 0.82    | 1503    |
| 2         | 0.68   | 0.76     | 0.72    | 642     |
| 3         | 0.73   | 0.99     | 0.84    | 620     |

| accuracy   | macro avg | f1-score | support |
|------------|------------|----------|---------|
|            | 0.78       | 0.80     | 2765    |

| weighted avg | f1-score | support |
|--------------|----------|---------|
| 0.82         | 0.80     | 2765    |

Figure 7: Confusion matrices of the study

According to the confusion matrix, from a total test dataset, from 1503 cases as “output1”, 1115 cases were recognized correctly, but the rest consisting of 224, and 164 were classified as false negatives in classes output 2 and output 3. Moreover, from 642 in “output 2” cases, 485 were correctly labeled, and about “output 3”, from 620 cases only 7 cases were labeled by mistake. Thus, the sample from each class that was correctly labeled is known as “True Positive”, while samples incorrectly labeled in the other classes would be “False Negative” or a “False Positive”.

The next step was related to finding the precision and accuracy of the measurement. The metrics for evaluating the classification used are:

- In this study accuracy measures shows 80% of the prediction is reliable data, which proves that here the classifier makes the correct prediction.
- In this study the result of precision rates are 0.91, 0.68, and 0.73 for “output 1”, “output 2”, and “output 3,” respectively. Since the ideal rate of precision is 1, in this classification as all results are near to 1, the ratio of the total number are correctly classified.
- Recall (TPR, Sensitivity) ideally shows 1. Whenever False Negative is important and trumps False Positive, it was used. Recall of outputs equaled 0.74, 0.76, and 0.99, respectively and near to the ideal case.
- In this classification, F1-Score showed 0.82, 0.72, and 0.84 for three classes of the results. Since the F1-Score decreases when the precision is increased and vice-versa, it shows good results regarding to the precision.

4. Conclusion

This research study is one of the first research studies to assess WL maximum output energy in Tallinn based on spatiotemporal variations with the presented novel machine learning approach. The research outcome is a reliable framework for preliminary solar analysis to map the solar energy potential of buildings with 80% accuracy. In this study, True Positives results of Confusion Matrix are important in evaluating the accuracy. For instance, about 91% for “output1” are correctly labelled and in the worst case, 73% for “output3” are labelled correctly. Thus, the method of the study and the classification can be implemented in similar studies to help classification of data, prediction, and data generation in different situations with more reliable outputs and resulting classes.
Acknowledgments
"This work has been supported by the European Commission through the H2020 project Finest Twins (grant No. 856602)."

References

[1] Mutule, A., J. Teremranova, and N. Antoskovs. "Smart city Through a flexible approach to smart energy." Latvian Journal of Physics and Technical Sciences 55.1 (2018): 3.
[2] Luo, Haidong, et al. "A short-term energy prediction system based on edge computing for smart city." Future Generation Computer Systems 101 (2019): 444-457.
[3] Khan, Shahzad, et al. "Artificial intelligence framework for smart city microgrids: State of the art, challenges, and opportunities." 2018 Third International Conference on Fog and Mobile Edge Computing (FMEC). IEEE, 2018.
[4] Ertugrul ÖF, Kaya Y. Smart city planning by estimating energy efficiency of buildings by extreme learning machine. In international smart grid congress and fair 2016 (pp. 1-5). IEEE
[5] Kabir, E.; Kumar, P.; Kumar, S.; Adelodun, A.A.; Kim, K. Solar energy: Potential and future prospects. Renew. Sustain. Energy Rev. 2018, 82, 894–900
[6] Good CS, Lobaccaro G, Hårklau S. Optimization of solar energy potential for buildings in urban areas – a Norwegian case study. Energy Procedia 2014;58:166–71.
[7] H.Zarrinkafsh, “Cable Structure Aided Water Lens For Solar Concentration”, IR 95246, Validated by Iranian Research Organization for Science and Technology (IROST), IRIPO, 2017
[8] Antón, I., Pachón, D., Sala, G., Characterization of optical collectors for concentration photovoltaic applications. Prog. Photovolt. Res. Appl. 2003.11, 387–405. http://dx.doi.org/10.1002/pip.502.
[9] Pavlovic, Sasa, Velimir Stefanovic, and Milorad Bojic. "Optical simulation of a solar parabolic collector using ray-tracing software TracePro." Hong Kong, China (2014): 211-218.
[10] Kim, Seul-Gi, Jae-Yoon Jung, and Min Kyu Sim. "A two-step approach to solar power generation prediction based on weather data using machine learning." Sustainability 11.5 (2019): 1501.
[11] Jakubiec, J. Alstan, and Christoph F. Reinhart. "A method for predicting city-wide electricity gains from photovoltaic panels based on LiDAR and GIS data combined with hourly Daysim simulations." Solar Energy 93 (2013): 127-143.
[12] Assouline, Dan, Nahid Mohajeri, and Jean-Louis Scartezzini. "Quantifying rooftop photovoltaic solar energy potential: A machine learning approach." Solar Energy 141 (2017): 278-296.
[13] Sharma, Navin, et al. "Predicting solar generation from weather forecasts using machine learning." 2011 IEEE international conference on smart grid communications (SmartGridComm). IEEE, 2011.
[14] Welch, Alina, et al. "Big data mining for the estimation of hourly rooftop photovoltaic potential and its uncertainty." Applied Energy 262 (2020): 114404.
[15] Peters, Sascha, and Diana Drewes. "8. Renewable Energy And Energy Production." Materials in Progress. Birkhäuser, 2019. 220-250.
[16] Qudsi, Dania. "Experimental Study of Water Lens Photovoltaic Concentrating System." (2019).
[17] Ryman, A., and G. Bastian. "Small-Scale Solar Cogeneration Systems." Journal of Clean Energy Technologies 6.5 (2018).
[18] Mondol, A. S., B. Vogel, and G. Bastian. "Large scale water lens for solar concentration." Optics express 23.11 (2015): A692-A708.
[19] N. Sugiuera and S. Morita, “Variable-focus liquid-filled optical lens,” Appl. Opt. 32, (1993).
[20] Voyant, Cyril, et al. "Machine learning methods for solar radiation forecasting: A review." Renewable Energy 105 (2017): 569-582.
[21] Esalamirad, N.; DeLuca, F.; Sakary Lylykangas, K.; (2021). Building Morphology Influence on Outdoor Comfort in Urban Environments. An Analysis through Machine Learning. Conference: 12th annual Symposium on Simulation for Architecture and Urban Design (SimAUD 2021)
[22] Esalamirad, N., et al. "Thermal comfort prediction by applying supervised machine learning in green sidewalks of Tehran," Smart and Sustainable Built Environment (2020).
[23] De Luca, Francesco, Emanuele Naboni, and Gabriele Lobaccaro. "Tall buildings cluster form rationalization in a Nordic climate by factoring in indoor-outdoor comfort and energy," Energy and Buildings 238 (2021): 110831.
[24] https://ec.europa.eu/jrc/en/PVGIS/tools/hourly-radiation