Modelling Solar Energy Suitability in Java Island Using Remote Sensing and Machine Learning

Ivani Christin Nugraheni[1], Anjar Dimara Sakti[2], Ketut Wikantika[3]
{ivanichristin82@gmail.com[1], anjar@itb.ac.id[2], ketut.wikantika@itb.ac.id[3]}

Remote Sensing and Geographic Information Science Research Group, Faculty of Earth Sciences and Technology, Institut Teknologi Bandung, Bandung 40132, Indonesia[1 2 3], Center for Remote Sensing, Institut Teknologi Bandung, Bandung 40132, Indonesia[1 2 3]

Abstract. The use of electrical energy on the island of Java in 2022 will be 45,835.45 MW. This shows an increase in energy use of 9.8% from 2021. Seeing the trend of energy demand which is increasing every year, the supply of electrical energy is also increasing. The increase in meeting energy needs has an impact on the energy crisis so alternatives are needed in the form of renewable energy, one of which is solar energy. This study will model and analyze the suitability of solar energy potential on the island of Java by integrating remote sensing data and machine learning methods. It is hoped that this study can be used as material for consideration in managing renewable resources on the island of Java. The results section shows that the machine learning method used shows the potential suitability of solar energy in medium to high class.

Keywords: solar energy; suitability; remote sensing; machine learning; Java Island.

1 Introduction

According to Indonesia's ministry of energy and mineral resources, the use of electrical energy on Java Island in 2022 will be 45,835.45 MW. This shows an increase in energy use of 9.8% from 2021. Seeing the trend of energy demand that is increasing every year, the supply of electrical energy is also increasing. Increased fulfillment of energy needs has an impact on the energy crisis so that alternatives in the form of renewable energy are needed [1]. This energy transition requires transformation of actors and market behavior, and changes in existing regulations and policies [2]. In addition, this energy transition requires the potential to support the transformation process that occurs. In the energy transition, policy makers need data for planning, target setting, and sector policy design in the development of renewable energy resources. In addition, it is necessary to emphasize the availability and affordability of energy, such as access to sufficient energy sources including energy infrastructure. Providing the availability and affordability of renewable energy requires a method that can model the potential area and power that can be obtained from a renewable energy infrastructure. One of the methods developed for this is machine learning [3]. In recent decades, machine learning techniques have been widely applied to many fields related to database problems. Many studies have revealed that various machine learning models or algorithms can be used in

RESILIENCE 2023, November 24, Jakarta, Indonesia

DOI 10.4108/eai.24-11-2023.2346424

renewable energy prediction [4]. This is done with a collection of training data to make decisions so as to obtain the desired value of forecasting results [3]. The performance evaluation of machine learning models in renewable energy prediction is done by calculating the accuracy and efficiency of the model [5]. One of the important things in renewable energy prediction is the exploration of potential renewable energy locations. The exploration process in selecting potential locations for renewable energy can be done by various methods, one of which is by utilizing remote sensing and Geographic Information Systems (GIS). Remote sensing and GIS together can obtain information about the earth's surface that can contribute significantly to the method of developing renewable energy resources, especially in the exploration stage of potential site selection [6]. Satellite images using active sensors and passive (optical) sensors can be used effectively in estimating potential areas for solar power, wind characteristics, and estimating biomass as a renewable energy source [7], [8]. To be able to achieve the availability of data on renewable energy potential, it is necessary to develop the suitability of potential renewable energy areas, especially solar energy, by integrating machine learning methods with remote sensing. This has been developed in several specific countries such as Switzerland [9], South Korea [10], the United States [11], Morocco [12], Nigeria [13], Spain [14], and in several other countries.

Modelling the measurement of solar energy potential, especially for rooftop photovoltaics in the Swiss region, has been carried out using a machine learning approach by Assouline et al.[9]. In the study, the algorithm used to estimate the potential of rooftop solar photovoltaics in urban areas in Switzerland was to combine the Support Vector Machine (SVM) algorithm with Geographic Information Systems (GIS). The parameters used in the study were horizontal solar radiation; weather data such as temperature, cloud cover, and rainfall; and the Digital Elevation Model (DEM) in the Swiss region for both monthly and annual data. The results of the study show that on average 81% of the total ground floor area of each building corresponds to the roof area available for rooftop solar photovoltaic installations. The results of the study show that on average 81% of the total ground floor area of each building can be used as the available roof area for rooftop solar photovoltaic installations. Then there was a study on estimating renewable energy potential in Nigeria using machine learning algorithms with climate-weather forecasting by Maduabuchi et al. [13]. The study used parameters such as daily air temperature, relative humidity, atmospheric pressure, rainfall, and wind speed with data for 19 years (2004-2022) to produce 6664 high-resolution data points. The resulting data was used to build various energy networks with various parameters to find the best renewable energy electricity network settings. From the machine learning regression, a high regression correlation of 96% was obtained during the algorithm. Based on the evaluation of previous studies, machine learning algorithms can be used to predict the potential of renewable energy in a region with certain parameters whose data are dominated by remote sensing satellite data and weather/climate data. The parameters that must be considered in determining the potential of solar energy are horizontal solar radiation, elevation data, and weather-climate parameters such as temperature, rainfall, and cloud cover. Therefore, this study uses horizontal solar radiation parameters, elevation data, and weather-climate parameters from remote sensing data with machine learning methods to determine the distribution of solar energy potential on Java Island. It is hoped that this study can produce a distribution of suitability areas for solar energy on Java Island which can be used in the initial study of renewable energy exploration on Java Island in order to support the renewable energy transition as a strategic step to realize sustainable development. The integration of machine learning with solar technology aligns with nature-based solutions by optimizing efficiency, mimicking natural processes, integrating with ecosystems, contributing to climate change mitigation, adapting to change, and promoting resource conservation. This combination has the potential to create more sustainable and environmentally friendly solutions for harnessing solar energy.

2 Material and Method

2.1 Study Area

The study area for this study is Java Island. This area consists of six provinces namely DKI Jakarta, Banten, West Java. Central Java, East Java and Yogyakarta. All areas in Java Island have a tropical climate and therefore receive consistent sunlight throughout the year. According to data from the Indonesian Ministry of Energy and Mineral Resources, there is one solar energy power plant in West Java. **Figure 1** shows the scope of the study area in this research.

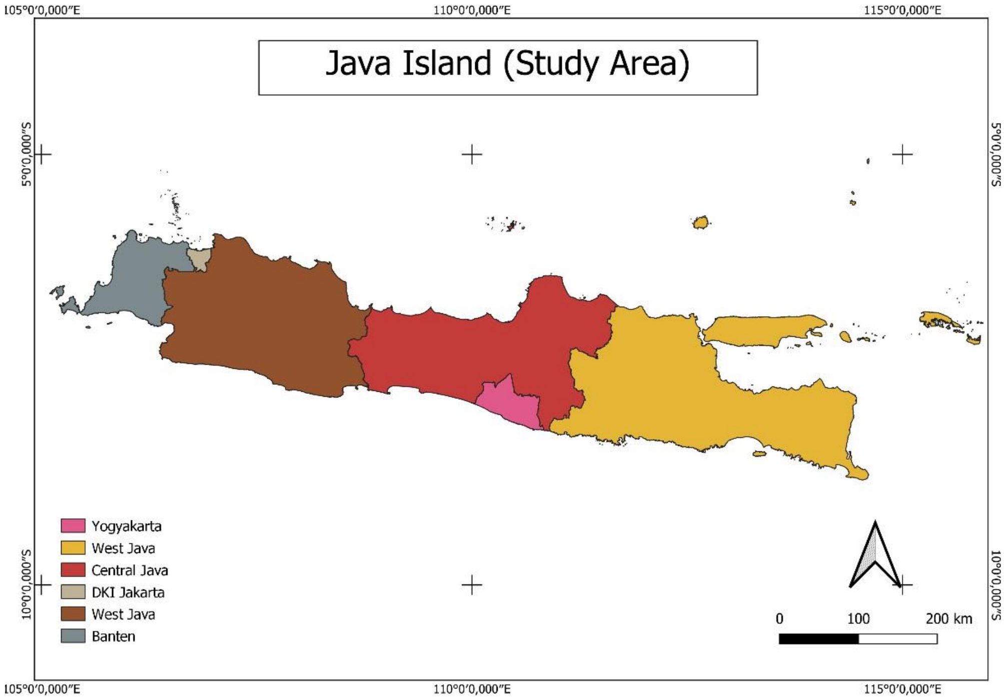

Fig. 1. Study area.

2.2 Material

The data used to determine sample points in machine learning are the solar energy suitability index and solar energy power potential. The machine learning algorithm uses parameters such as Global Horizontal Irradiance (GHI) to describe sunlight radiance, night light to describe human activities at night, precipitation as a weather parameter, elevation and slope to represent surface appearance, cloud fraction as a consideration of cloud cover, and Land Surface Temperature (LST) for surface temperature parameters. Table 1 shows the data used as well as their sources and data types.

Table 1: Data used and data format.

No	Data	Product	Temporal	Type-Resolution
1	Solar energy suitability index	Solar energy suitability index [15]	2022	Raster-1000m
2	Solar energy potential	Solar energy potential [15]	2022	Raster-1000m
3	Global Horizontal Irradiance (GHI)	Global Solar Atlas	2022	Raster-250m
4	Night light	VIIRS Stray Light Corrected Nighttime Day/Night Band Composites	2022	Raster-463.83m
5	Precipitation	CHIRPS Daily: Climate Hazards Group InfraRed Precipitation With Station Data (Version 2.0 Final)	2022	Raster-5566m
6	Elevation	NASA SRTM Digital Elevation	2022	Raster-30m
7	Cloud fraction	MODIS Aqua/Terra Cloud Fraction (Monthly)	2022	Raster-5000m
8	Slope	NASA SRTM Digital Elevation	2022	Raster-30m
9	Land Surface Temperature (LST)	MOD11A1.061 Terra Land Surface Temperature and Emissivity Daily Global 1km	2022	Raster-1000m

2.3 Method

In this study, there are two main stages to determine the potential solar energy areas in Java. The first stage was conducted to determine sample points as input in the machine learning process. This stage uses the solar energy suitability index and solar energy potential. From both data, reclassify and integrate so that potential and non-potential sample points are obtained. The second stage was conducted to determine the potential areas suitable for solar energy on Java Island.

The input data used are sample points that have been obtained previously. The algorithms used are Random Forest and Gradient Tree Boost (GTB). The integration of the results of random forest and GTB is carried out with the aim of increasing the probability and sensitivity in determining the potential area of solar energy. **Figure 2** shows the stages of data processing presented in a flowchart.

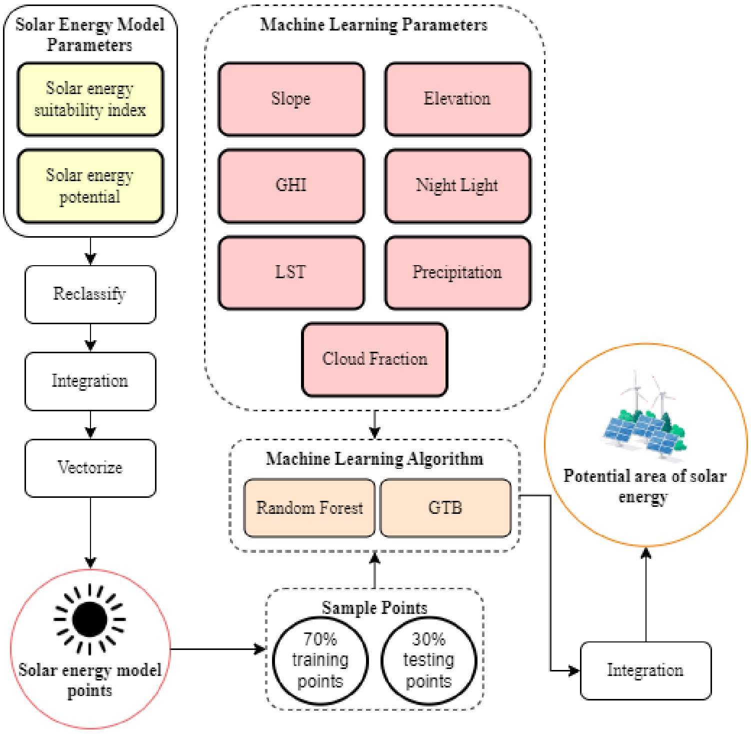

Fig. 2. Flowchart used in this research.

Random forest is a machine learning algorithm that is a collection of trees, where each tree is determined from randomly selected independent pixel values in each vector [16]. This method is a combination of Bagging and Random Sub Spaces methods [17]. GTB is also a machine learning method with a decision tree system. GTB builds a decision tree based on improvements in the tree structure under weak learning so that it can be used to correct tree errors and prevent potential overfitting of the decision tree [18]. The application of machine learning is done with several software such as QGIS and Google Earth Engine. Machine learning requires predictor parameters and sample points, which in this case represent potential and non-potential areas for solar power plants. The predictor parameters used in the machine learning process such as slope, GHI, LST, elevation, night light, precipitation, and cloud fraction are extracted from remote sensing data of the same year. The predictor parameters are included in the machine learning algorithm to see the pattern of potential and non-potential training points, so that after obtaining the pattern the algorithm is tested on the testing points entered.

3 Result and Discussion

The machine learning process generates the distribution of solar energy potential areas from the solar energy potential sample points. The resulting potential area distribution is shown in blue to red color gradation. The more blue the lower the value, and the more red the higher the value. For the value of each machine learning method, the results are in the value interval 0-1, while for machine learning integration the results are in the value interval 0-2. **Figure 3** shows the results of the distribution of potential solar energy areas on the island of Java.

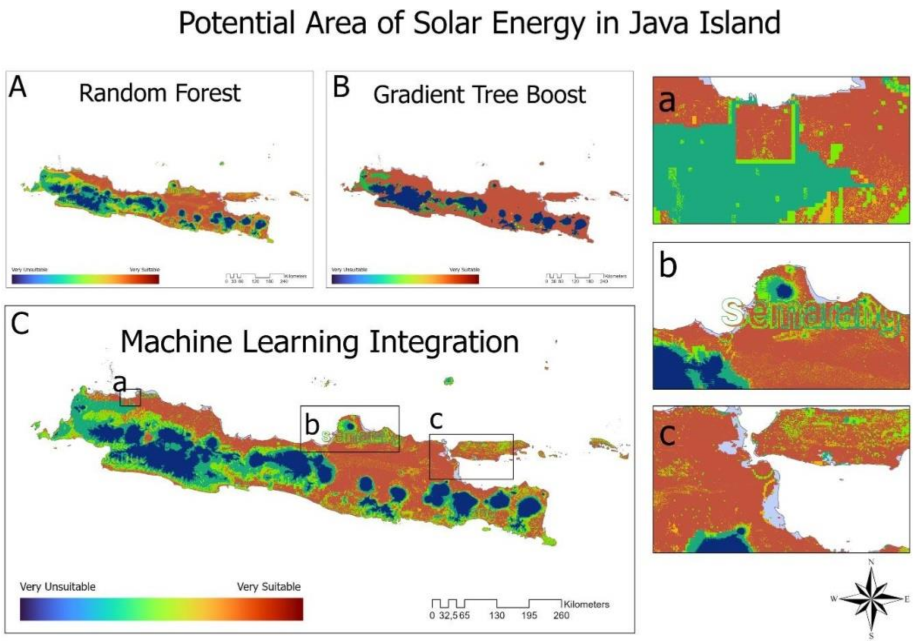

Fig. 3. Potential area of solar energy in Java Island with machine learning algorithms, such as: Random Forest (A), Gradient Tree Boost (B), and machine learning integration (C) in the DKI Jakarta (a), Semarang (b), and Surabaya (c).

In terms of visualization, both methods produce a distribution of areas that tend to be similar in pattern, no method produces a pattern that is inversely proportional to the other methods. This results in the integration of methods whose visualization also has the same pattern as the previous two methods. Areas with high suitability in the random forest results also show high conformity in the GTB results, and areas with low suitability in the random forest results also show low suitability in the GTB results.

In the random forest method, areas with high suitability have an area of 42.83%, areas with moderate suitability have an area of 17.30%, and areas with low suitability have an area of 39.87%. In the GTB method, areas with high suitability have an area of 65.59%, areas with moderate suitability have an area of 0.92%, and areas with low suitability have an area of 33.49%. In the machine learning method integration, the area with high suitability has an area of 44.72%, the area with moderate suitability has an area of 15.47%, and the area with low suitability has an area of 39.81%.

The results of the potential area distribution show that Java Island is dominated by areas with low to high potential. Both visually and spatially, areas with high potential dominate Java Island. This shows that Java Island has high potential and suitability for the use and development of solar energy power plants. Thus, the process of energy transition to renewable energy, especially with solar energy in Java Island, has the potential to be carried out.

DKI Jakarta, Semarang (Central Java), and Surabaya (West Java) areas show a red colour distribution, which means that the three cities have a high suitability for solar energy infrastructure development. This can happen because it is supported by parameters such as temperature (LST) which tends to be high, allowing high heat for solar energy sources. High GHI values indicate high sunlight in the three cities, increasing the potential for solar energy. The three cities are also located in coastal areas, allowing for high temperatures with relatively moderate cloud cover.

The parameters used in machine learning for potential solar energy points are seven parameters, namely slope, LST, GHI, elevation, night light, precipitation, cloud fraction, and night light. With these seven parameters, a correlation is made between parameters both with other parameters and with themselves. The result of the parameter correlation with itself will show a value of 1 because of course it will be strongly correlated.

The correlation matrix of the solar energy potential point machine learning parameters shows the correlation value between different parameters. The diagonal of this matrix is one because each parameter has a strong correlation with itself. The value of one also indicates that the parameters have the exact same value, so if the value of one does not appear on the diagonal of the matrix then the two parameters have the same or similar variables. In the correlation matrix of the parameters of the solar energy potential point machine learning, there are no values close to one, so each parameter is not the same or similar parameter.

Parameter values that are inversely proportional such as LST with slope, elevation with LST, GHI with slope, and GHI with elevation. LST shows the value of surface temperature, which will decrease with increasing altitude and slope. Therefore, the correlation of LST parameters with slope, and LST with elevation is inversely proportional. While values that are directly proportional such as elevation with slope, elevation with GHI, and GHI with LST. Elevation value shows the height of a place. Generally, if the elevation value is high, the slope is also high.

Apart from the correlation matrix, the ROC graph for each machine learning method applied to the points of the solar energy potential model was also determined. This graph contains four values of the method used. The value of this graph presents TPR (true-positive-rate) for the y-axis value, and TNR (true-negative-rate) for the x- axis value. The ROC graphs for the solar energy potential points show similar shapes to each other. This can be due to the optimal distribution of points and appropriate parameters. The value of this graph shows the magnitude of the resulting AUC value. This AUC value is one of the determinants of the quality of the resulting classification.

The AUC value generated by random forest is 0.9749, and the AUC value generated by GTB is 0.9746. Apart from the correlation matrix, the ROC graph for each machine learning method applied to the points of the solar energy potential model was also determined. This graph contains four values of the method used. The value of this graph presents TPR (true-positive-rate) for the y-axis value, and TNR (true-negative-rate) for the x-axis value [19]. The ROC graphs for the solar energy potential points show similar shapes to each other. This can be due to the optimal distribution of points and appropriate parameters [19]. The value of this graph shows the magnitude of the resulting AUC value. This AUC value is one of the determinants of the quality of the resulting classification [20]. The AUC value generated by random forest is 0.9749, and the AUC value generated by GTB is 0.9746. It shows that both machine learning methods used have classification at the excellent classification level. Thus, showing the quality of modelling that is well classified [19].

Apart from the ROC and AUC values, there are variables that play a role in the method. Each of these variables has a different role. Each importance value has a different value scale. Each method has a different parameter importance, but all parameters play an important role in determining the potential solar energy area. Random forest and GTB show the variable with the highest value, namely elevation. The Random Forest method shows the three parameters that have the highest importance are elevation, GHI, and LST, respectively. The GTB method shows the three parameters that have the highest importance are elevation, GHI, and slope.

In general, the results show that elevation, LST, GHI, and slope parameters have an important role in determining the class of existing solar energy areas. Elevation is related to the height that affects the placement of the ideal or optimal location of solar energy power plants. LST is related to the surface temperature that affects the temperature of the location of solar energy power plants, one of which is related to the temperature of solar irradiation. GHI relates to the amount of sunlight that reaches the surface, thus affecting the amount of solar energy that can be radiated to the surface. Slope is related to the slope of the slope which affects the optimal placement of solar energy power plant locations so that they can last a long time.

4 Conclusion

The use of remote sensing data in the machine learning process can be used efficiently in modelling the suitability of potential areas for solar energy. This modelling can produce areas with varying degrees of suitability depending on the parameters used. In the case of Java Island, areas with medium to high suitability or potential are dominated. To be able to further optimize the work of machine learning in energy modelling can use more diverse parameters and more evenly distributed sample points.

References

[1] T. Mitrova and Y. Melnikov, “Energy transition in Russia,” *Energy Transit.*, vol. 3, no. 1–2, pp. 73–80, Dec. 2019, doi: 10.1007/s41825-019-00016-8.
[2] J. Blazquez, R. Fuentes, and B. Manzano, “On some economic principles of the energy transition,” *Energy Policy*, vol. 147, p. 111807, Dec. 2020, doi: 10.1016/j.enpol.2020.111807.
[3] J.-P. Lai, Y.-M. Chang, C.-H. Chen, and P.-F. Pai, “A Survey of Machine Learning Models in Renewable Energy Predictions,” *Appl. Sci.*, vol. 10, no. 17, p. 5975, Aug. 2020, doi: 10.3390/app10175975.
[4] A. G. Olabi, “Renewable energy and energy storage systems,” *Energy*, vol. 136, pp. 1–6, Oct. 2017, doi: 10.1016/j.energy.2017.07.054.
[5] A. Zendehboudi, M. A. Baseer, and R. Saidur, “Application of support vector machine models for forecasting solar and wind energy resources: A review,” *J. Clean. Prod.*, vol. 199, pp. 272–285, Oct. 2018, doi: 10.1016/j.jclepro.2018.07.164.
[6] Avtar *et al.*, “Exploring Renewable Energy Resources Using Remote Sensing and GIS—A Review,” *Resources*, vol. 8, no. 3, p. 149, Aug. 2019, doi: 10.3390/resources8030149.

[7] J. A. Jakubiec and C. F. Reinhart, "A method for predicting city-wide electricity gains from photovoltaic panels based on LiDAR and GIS data combined with hourly Daysim simulations," *Sol. Energy*, vol. 93, pp. 127–143, Jul. 2013, doi: 10.1016/j.solener.2013.03.022.

[8] N. Lukač, G. Štumberger, and B. Žalik, "Wind resource assessment using airborne LiDAR data and smoothed particle hydrodynamics," *Environ. Model. Softw.*, vol. 95, pp. 1–12, Sep. 2017, doi: 10.1016/j.envsoft.2017.05.006.

[9] D. Assouline, N. Mohajeri, and J.-L. Scartezzini, "Quantifying rooftop photovoltaic solar energy potential: A machine learning approach," *Sol. Energy*, vol. 141, pp. 278–296, Jan. 2017, doi: 10.1016/j.solener.2016.11.045.

[10] P. W. Khan, Y.-C. Byun, S.-J. Lee, D.-H. Kang, J.-Y. Kang, and H.-S. Park, "Machine Learning-Based Approach to Predict Energy Consumption of Renewable and Nonrenewable Power Sources," *Energies*, vol. 13, no. 18, p. 4870, Sep. 2020, doi: 10.3390/en13184870.

[11] Y. Lin *et al.*, "Revenue prediction for integrated renewable energy and energy storage system using machine learning techniques," *J. Energy Storage*, vol. 50, p. 104123, Jun. 2022, doi: 10.1016/j.est.2022.104123.

[12] I. Jebli, F.-Z. Belouadha, M. I. Kabbaj, and A. Tilioua, "Prediction of solar energy guided by pearson correlation using machine learning," *Energy*, vol. 224, p. 120109, Jun. 2021, doi: 10.1016/j.energy.2021.120109.

[13] C. Maduabuchi *et al.*, "Renewable Energy Potential Estimation Using Climatic-Weather-Forecasting Machine Learning Algorithms," *Energies*, vol. 16, no. 4, p. 1603, Feb. 2023, doi: 10.3390/en16041603.

[14] S. Salcedo-Sanz, L. Cornejo-Bueno, L. Prieto, D. Paredes, and R. García-Herrera, "Feature selection in machine learning prediction systems for renewable energy applications," *Renew. Sustain. Energy Rev.*, vol. 90, pp. 728–741, Jul. 2018, doi: 10.1016/j.rser.2018.04.008.

[15] A. D. Sakti *et al.*, "Spatial integration framework of solar, wind, and hydropower energy potential in Southeast Asia," *Sci. Rep.*, vol. 13, no. 1, p. 340, Jan. 2023, doi: 10.1038/s41598-022-25570-y.

[16] L. Breiman, "Random Forest," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001, doi: 10.1023/A:1010933404324.

[17] L. Breiman, "Bagging predictors," *Mach. Learn.*, vol. 24, no. 2, pp. 123–140, Aug. 1996, doi: 10.1007/BF00058655.

[18] A. Natekin and A. Knoll, "Gradient boosting machines, a tutorial," *Front. Neurorobotics*, vol. 7, 2013, doi: 10.3389/fnbot.2013.00021.

[19] F. Gorunescu, *Data mining: concepts, models and techniques*. in Intelligent systems reference library, no. 12. Berlin, Heilelberg: Springer-Verlag, 2011.

[20] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, *Classification And Regression Trees*, 1st ed. Routledge, 1984. doi: 10.1201/9781315139470.