Article

Integration of the Machine Learning Algorithms and I-MR Statistical Process Control for Solar Energy

Yasemin Ayaz Atalan [1,*] and Abdulkadir Atalan [2,*]

1   Department of Mechanical Engineering, Yozgat Bozok University, Yozgat 66200, Turkey
2   Department of Industrial Engineering, Çanakkale Onsekiz Mart University, Çanakkale 17100, Turkey
*   Correspondence: yasemin.ayaz@bozok.edu.tr (Y.A.A.); abdulkadiratalan@gmail.com (A.A.)

Abstract: The importance of solar power generation facilities, as one of the renewable energy types, is increasing daily. This study proposes a two-way validation approach to verify the validity of the forecast data by integrating solar energy production quantity with machine learning (ML) and I-MR statistical process control (SPC) charts. The estimation data for the amount of solar energy production were obtained by using random forest (RF), linear regression (LR), gradient boosting (GB), and adaptive boost or AdaBoost (AB) algorithms from ML models. Data belonging to eight independent variables consisting of environmental and geographical factors were used. This study consists of approximately two years of data on the amount of solar energy production for 636 days. The study consisted of three stages: First, descriptive statistics and analysis of variance tests of the dependent and independent variables were performed. In the second stage of the method, estimation data for the amount of solar energy production, representing the dependent variable, were obtained from AB, RF, GB, and LR algorithms and ML models. The AB algorithm performed best among the ML models, with the lowest RMSE, MSE, and MAE values and the highest R^2 value for the forecast data. For the estimation phase of the AB algorithm, the RMSE, MSE, MAE, and R^2 values were calculated as 0.328, 0.107, 0.134, and 0.909, respectively. The RF algorithm performed worst with performance scores for the prediction data. The RMSE, MSE, MAE, and R^2 values of the RF algorithm were calculated as 0.685, 0.469, 0.503, and 0.623, respectively. In the last stage, the estimation data were tested with I-MR control charts, one of the statistical control tools. At the end of all phases, this study aimed to validate the results obtained by integrating the two techniques. Therefore, this study offers a critical perspective to demonstrate a two-way verification approach to whether a system's forecast data are under control for the future.

Keywords: solar energy; machine learning; random forest; AdaBoost; gradient boosting; linear regression; statistical process control; I-MR control chart

1. Introduction

The economic and development wealth of countries is usually measured by factors such as their energy production facilities, along with their use and accessibility. Comparing energy production methods with the technological infrastructure of countries depending on energy consumption is perceived as a fair approach [1]. Most countries use fossil fuels as their primary energy source for energy production, adversely affecting air quality [2]. The heat released by such fuels to the environment causes many adverse effects. For this reason, countries are searching for clean energy production by using the natural riches offered by nature for energy production. Solar and wind energy facilities are the first to come to mind in producing clean and renewable energy. This study discusses a case study that considers environmental factors affecting the amount of solar energy production. We analyzed the estimation data, showing that solar-based energy production that contributes to renewable energy production will be an energy source for many years. In this way, awareness of use

will be increased with the increase in solar energy production among the energy production solutions, as an alternative to the energy production obtained with fossil fuels [3].

Since solar energy is one of the clean and renewable types of energy, it is among the alternative sources of energy production and attracts significant attention from countries [4]. The most important source of this importance is the increase in the amount of electricity produced by solar energy and the decrease in the amount of fossil-based energy production, as well as being environmentally friendly [5]. Solar power generation facilities generally provide services by converting solar energy into electrical energy using photovoltaic (PV) systems. The amount of energy produced by solar energy systems is naturally significantly affected by environmental conditions [6]. For this reason, temperature, humidity, dew point, cloud coverage, altitude, visibility, pressure, and wind speed parameters, which are among the critical environmental factors, were considered in this study. By analyzing the data of these factors, it is possible to predict the amounts of energy produced by PV cells for future periods.

This study aimed to use machine learning (ML) models to estimate the amount of solar energy production. Although there is a statistical approach based on ML algorithms, these algorithms work differently than statistical applications [7]. While statistical methods generally show a mathematical approach according to the typical characteristics of the data, ML models provide prediction data by taking into account the common aspects, connections, and behaviors of the data in the datasets and briefly learning from the data [8]. In particular, ML algorithms are frequently applied by researchers to obtain estimation data on energy production [9]. The differences revealed in terms of the ML algorithms used in this field are discussed in Table 1.

Table 1. Studies of ML algorithms used to predict solar energy production.

Location	ML Algorithms	Coefficient of Determination (R^2) *	Source
Not Defined	SVM, GPR	0.98	[1]
Republic of Korea	MLF	**	[6]
Morocco	LR, RF, SVR, ANN	0.99	[10]
PV Farms	EML	**	[11]
Taiwan	MLP, RF, *k*NN, LR	0.96	[12]
PV Farms	SVR, CNN	0.54	[13]
Republic of Korea	LR, LASSO, RF, SVM, GB	**	[14]
USA	LR, MARS	0.97	[15]
Sweden, Germany	DL	0.86	[16]
Italy	DLNN	0.99	[17]
USA	AB, RF, GB, LR with SPC	0.97	This Study

* The value of the model with the highest accuracy rate is shared. ** Not available.

Studies of ML algorithms that predict solar energy production usually offer a single approach. However, in this study, a second approach, the statistical process control (SPC) method, was used to confirm the validity of the prediction data obtained from ML algorithms. The SPC technique is widely preferred in industries to monitor the parameters of processes belonging to production or service workflows [18]. Recently, among the artificial intelligence methods, ML algorithms have been used enthusiastically, especially for big data processing and analysis. This study discusses and tests ML and SPC diagrams from statistical and engineering applications with a case study on the amount of solar energy production. Since both models are based on statistical models, these two techniques are expected to work in harmony [19].

ML models perform well for large datasets [20,21]. ML models are widely preferred by researchers, especially for fields such as medicine, transportation, production, logistics, economics, and education. ML models vary according to the computer programs used [22]. There are two stages in all ML algorithms. Although the ML method relies on statistical approaches, it primarily provides predictive data by discovering standard connections between data. ML models learn from data, and then they test the data and reveal model

performances. In other words, training and testing phases are required for ML models. Training and testing stages are created by sharing a certain amount of data in the dataset. The proportion of data for the training phase is generally higher than for the testing phase. This study set the training and testing phases to 75% and 25%, respectively. Finally, in terms of obtaining prediction data, ML algorithms can be combined with many techniques, such as simulation, statistics, and optimization, making the validity of the results more robust [23]. In this study, the data of a system containing the datasets of the amount of solar energy production were analyzed by integrating ML and SPC diagrams.

Statistical approaches offer different methods in terms of data types and are used in many fields [24]. The SPC technique is also essential among statistical approaches [25]. In principle, SPC analyzes system data to test whether a system is under control [26,27]. This method changes the use of SPC diagrams depending on whether the data are continuous or discrete. Generally, Xbar-R, Xbar-S, and I-MR control charts are preferred for constant data types, while p, np, u, and c control charts are used for discrete data types. In this study, I-MR control charts were preferred, since only a dependent variable representing the amount of solar energy production was considered. One study considered the interrelationship of quality study approaches and manufacturing procedure requirements for one of the SPC charts, the Xbar-R chart, to show that every manufacturing process in a business is linked to continuous quality improvement [28]. The p control chart, one of the preferred statistical control charts for the discrete data type, has been preferred in clinical practice [29]. One study proposed an economic statistical strategy with the Xbar-R control plot for non-quality normal symmetric distributions [30].

This study aims to estimate the data of a dependent variable belonging to the continuous data type with ML algorithms and test the prediction data with SPC charts. The characteristics of studies using ML and SPC graphical methods in one study are shown in Table 2.

Table 2. Some research related to statistical control diagrams and ML algorithms.

Data For	SPC	ML Algorithms	Source
Radiology	I-MR	Not Defined	[31]
Generated	Xbar-R	AIM	[32]
Generated	Not Defined	IL, NN	[7]
Drinking Water Treatment	Not Defined	DL	[33]
Wind Turbine	Not Listed in SPC	RF, DT	[34]
Steel Production	Not Defined	EML (LR, RD, LaR, EN, SVM, KNN, RF, GBDT, LGBM, XGBoost, KRR)	[13]
Manufacturing Performance	Hotelling's T^2	RF, SVM, NB	[35]
Water Temperature	I-MR, Hotelling's T^2	SVM	[15]
Generated	I-MR	AB, GB, RF, LR	This Study

The motivation for the emergence of this study was expressed as the formation of a two-way verification mechanism of the systems that provide the prediction data. The comparison of the method used in this study with the methods used in other studies is presented in Table 2, and the autonomy of this study is shown. While the abovementioned working methods and proposed solutions to the problem are successful, these methods are only concerned with solving a particular situation. Especially since the data of SPC diagrams do not contain any tests, different approaches are needed. Therefore, this study aimed to prove the validity of the outcomes obtained by integrating the two techniques. From the perspective put forward for this study, it makes an essential contribution to easily detecting whether the systems are under control for the future processes of a system. Finally, this study uses data from a real case to demonstrate the successful implementation of real-world deployment with data from systems in different industries.

The novelty of this study will provide a double verification method instead of a one-sided verification of the forecast data of solar energy production to integrate ML and SPC

methods in solar energy production. Because ML algorithms can be used to optimize energy production by continuously monitoring and analyzing solar panel data, on the other hand, SPC methods monitor data anomalies at every stage of the production process and provide quick intervention, thus minimizing energy losses. The SPC method is needed to detect statistically significant irregularities in the forecast data by not determining the complex relationships between the ML models and the factors affecting solar energy production. The integration of ML and SPC contributes to making the accuracy of the forecast data of solar energy production more sustainable. Therefore, better energy estimates make using energy resources more effective. It is aimed in this study that the integration of ML and SPC methods in solar energy production can help the energy sector move towards a more efficient, environmentally friendly, and sustainable future. Thus, integrating these two methods is critical for forecasting solar power generation, increasing the efficiency of power plants, managing energy demands, and using resources more efficiently.

This work is organized into four essential parts: The first part of the study includes examples of using SPC diagrams and ML algorithms in the literature. Theoretical information about the research methodology and approaches is debated in the second part. The results of a numerical study using the data of input and response factors defined for this research are given in the third part. The results expressing the usage requirements of the proposed method and its importance for future studies are mentioned in the final section of the study.

2. Materials and Methods

This study tested the validity of the solar energy forecast data results, depending on the independent variables that are effective in solar energy production, by integrating SPC diagrams and ML algorithms. The data for this work were obtained from the publicly available center of the University of Illinois campus [36]. This study consisted of three stages: In the first stage, descriptive statistics and variance analysis of the dependent and independent variables of solar energy were performed for this study. GB, RF, AB, and LR models from ML algorithms formed the second phase of the study to obtain predictive data for the amount of solar energy. Finally, SPC diagrams were created to estimate the amount of solar energy, and the estimation data were compared with the actual data. The workflow diagram of the dependent and independent variable data types and method stages used in this study is shown in Figure 1.

2.1. Descriptive, Correlation, and Variance Statistics

This study considers eight independent variables and one dependent variable for the amount of solar energy (kWh). The datasets for these arguments have a numeric and continuous data type. The data of this study were collected for the period covering the 2-year data period. The independent variables of this study were cloud coverage (% range), visibility (miles), average temperature (°C) during the day, dew point (°C), relative humidity (%), wind speed (Mph), station pressure (inHg), and altimeter (inHg). These independent variables are discussed in this study to measure their effects on solar energy production, which is the dependent variable, and to express that these inputs play an essential role in the estimation data. Descriptive statistics of input and output factors are shared in Table 3.

The cloud cover (%range) variable represents the percentage of cloud cover for the 640 data points observed. The average percentage of cloud cover is 0.39, indicating that the area is usually partly cloudy. The standard deviation (0.31) indicates limited variation between observations, while the coefficient of variation (81.53) is high, indicating that the distribution is highly volatile. Skewness (-0.93) indicates that the distribution is slanted on the left, while kurtosis (-0.93) indicates that the distribution does not have extreme values. The visibility (miles) variable expresses the visibility in miles. The average viewing distance is 9.14 miles, and the distribution of these values is slightly more comprehensive, with a standard deviation of (1.41). Skewness (-2.00) is negative, which indicates that the

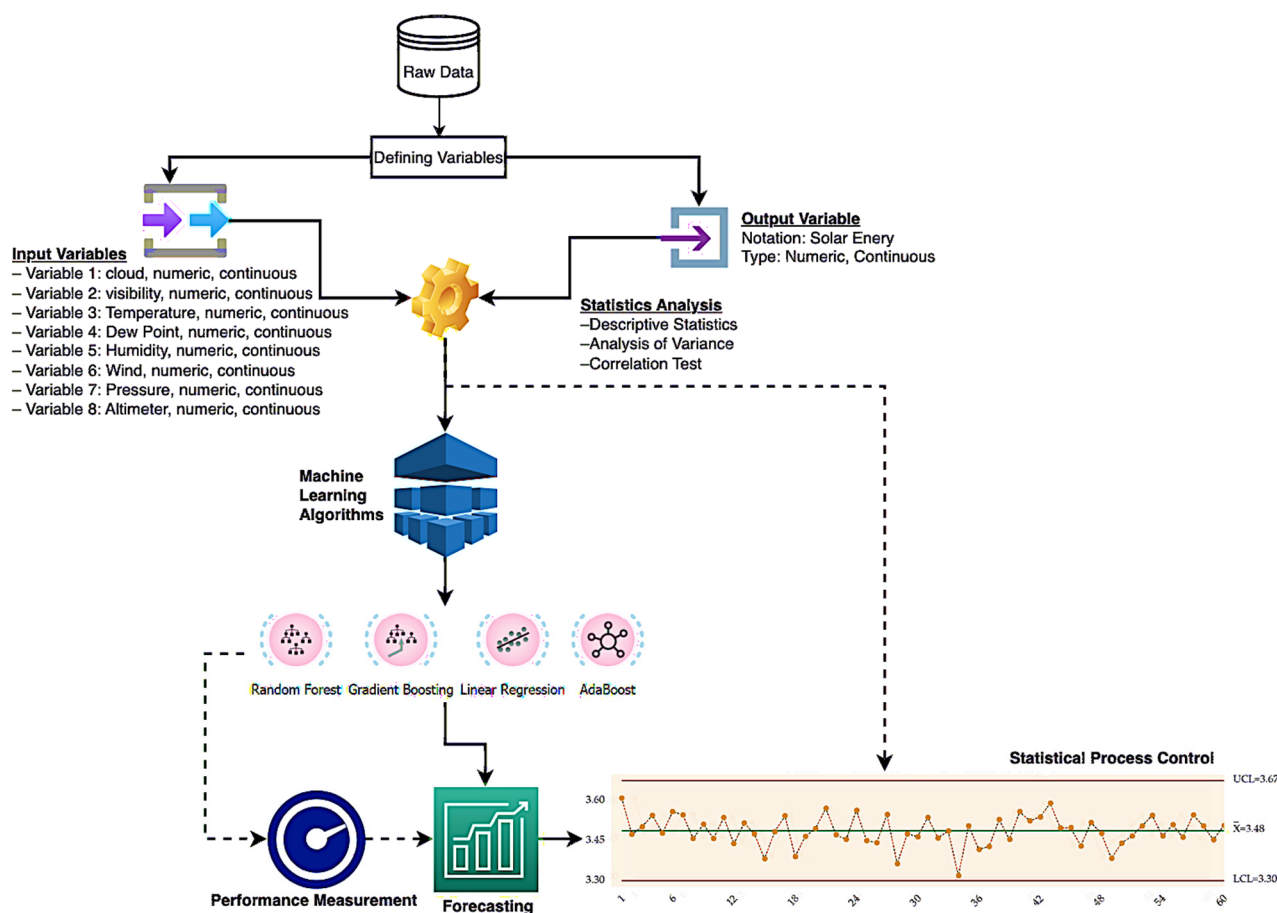distribution is slanted to the left, while kurtosis (5.94) indicates that the distribution has extreme values.

Figure 1. The workflow of the methodology developed for solar energy.

Table 3. The key results of the descriptive statistics of factors.

Variable	N	Mean	M_{se}	StDev	Var	Var_{coeff}	Min	Max	Skew	Kurt
Cloud (% range)	640	0.39	0.01	0.31	0.10	81.53	0.00	1.00	1.00	−0.93
Visibility (miles)	640	9.14	0.06	1.41	2.00	15.47	1.15	10.00	−2.00	5.94
Temperature (°C)	640	14.16	0.38	9.49	89.96	66.99	−16.06	28.18	−1.00	−0.39
Dew Point (°C)	640	9.58	0.37	9.34	87.19	97.50	−18.72	25.02	−1.00	−0.28
Humidity (%)	640	72.41	0.54	13.68	187.25	18.90	21.25	97.85	−1.00	1.07
Wind (Mph)	640	8.64	0.16	4.08	16.65	47.24	1.03	24.83	1.00	0.61
Pressure (inHg)	640	28.60	0.11	2.68	7.17	9.36	8.59	29.87	−6.00	35.07
Altimeter (inHg)	640	30.02	0.01	0.19	0.04	0.62	29.48	30.67	0.00	0.68
Energy (kWh)	640	21470	359	9095	827103	42.36	−641	45642	0.00	−0.64

The remaining sections of the table contain variables that measure weather conditions such as temperature, humidity, wind speed, and pressure. For example, the average temperature is 14.16 °C, and the data distribution is quite wide (standard deviation 9.49). Similarly, the statistical properties of other variables, such as humidity level, wind speed, and pressure, are also presented. These statistics help to understand the general trends and variability of these weather conditions and are used in analysis and decision-making processes. The average energy consumption is 21,470 kWh, and based on these data, the distribution of energy consumption appears to be relatively spread out. The standard deviation (9.095) is relatively high, indicating a wide distribution, while kurtosis (−0.64) indicates ineffective outliers.

Computing the correlation values of the input and output factors was intended to reveal the statistical dependencies between the variables [37]. As a general expression, the correlation values that can be obtained from the data types of the variables, excluding non-numeric datasets, vary between −1 and 1. The connection between the input and output factors increases as the correlation values move away from zero. However, as the correlation data approach zero, the relationship between the variables decreases statistically. The direction of the strong correlation values only refers to the positive or negative correlation between the factors. The correlation data of input and output factors are given in Table 4. The correlation values of the factors considered for this work were computed at medium or high levels.

Table 4. Correlation data of dependent and independent variables.

Feature 1	Feature 2	Correlation
Cloud	Energy	−0.988
Energy	Humidity	−0.772
Energy	Visibility	0.769
Energy	Temperature	0.700
Energy	Wind	−0.560
Dew	Energy	0.508
Altimeter	Energy	0.479
Energy	Pressure	0.470
Date	Energy	−0.301

Correlation values between variables were calculated based on Pearson analysis. In addition, the correlation values of the factors were computed considering the 95% confidence interval.

2.2. Machine Learning Algorithms

In this work, ML algorithms, a sub-approach of artificial intelligence, were used to obtain the estimation data for the amount of solar energy, which was the output variable, by considering the input factors. In the present research, estimation data of the dependent variable were obtained by using RF, AB, GB, and LR algorithms. The preferred algorithms for the prediction data of solar energy are Orange 3.35 computer programs with Python software and open access. The program model of this study using ML algorithms is visualized in Figure 2.

ML algorithms were run in two different cases to obtain the prediction data. First, analyses were carried out using the available data in the training and testing stages. Then, we tried to calculate the estimation data of the dependent variable by keeping the dependent variable data confidential. Thus, the validity of the estimation data with dual validation was tested.

Among the ML models, the GB algorithm is a classification- and regression-based model that adopts an augmentation algorithm approach [38]. This algorithm trains a new model sequentially to debug and correct the previous model. Usually, this algorithm integrates weak learners with strong learners [39]. The RF algorithm is a machine learning model that incorporates the results of multiple decision trees to obtain a single result [40]. One of the most important reasons why this algorithm is preferred among ML models is that it provides flexibility for regression and classification problems [41]. The AB algorithm is an ML algorithm that adopts an incremental technique used as an ensemble method [42]. The AB model serves as a classification model by assigning high weight values to misclassified samples using samples in the dataset [43]. This ML model usually uses the SAMME—R algorithm [44]. The LR model is a supervised ML algorithm that reveals the linear relationship between more than one independent variable influencing one or more dependent variables [45]. The LR model is a statistical approach that uses univariate or multivariate linear regression depending on the number of dependent variables. This

approach creates an optimal linear equation for estimating the dependent variable data based on the independent variable data types [46].
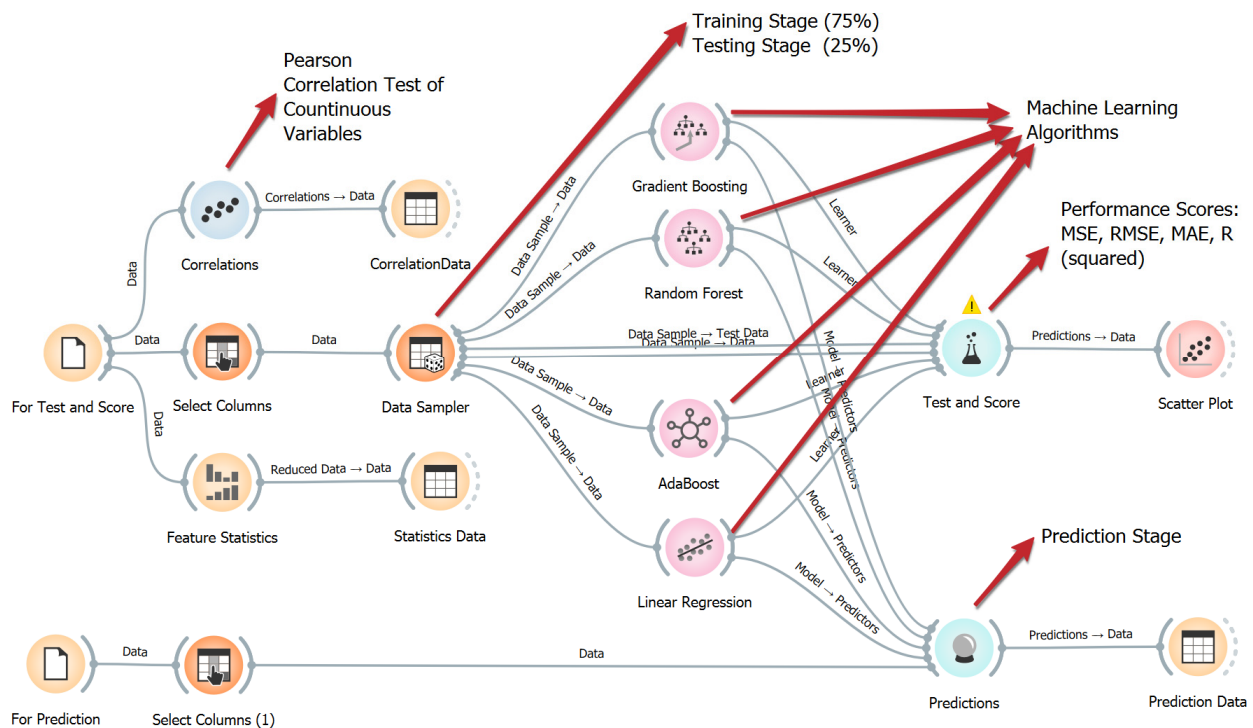

Figure 2. Screenshot of the model of ML algorithms.

The most important reason why more than one ML algorithm is preferred is to test the validity of the predicted data by comparing the performances of the models. The performances of ML models are measured by calculating the MSE (mean squared error), RMSE (root-mean-squared error), and MAE (mean absolute error) data, the margins of error, the R^2 values, and the precision coefficients. Generally, for ML to have a strong performance, it must have a coefficient of accuracy and low error values. The mathematical equations of the proposed algorithms for the performance score are given below:

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |y_i - \tilde{y}_i|^2 \tag{1}$$

$$MSE = \sum_{i=1}^{n} (y_i - \tilde{y}_i)^2 \tag{2}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (y_i - \tilde{y}_i)^2}{n}} \tag{3}$$

$$R^2 = \sum_{i=1}^{n} \left[\frac{y_i - \tilde{y}_i}{y_i - \overline{y}_i} \right]^2 \tag{4}$$

where, in the formulae above, the number of observations is indicated by n, while the estimated values are denoted by $\tilde{y}_i$ and the actual values are symbolized by y_i. The performance metrics' values of the algorithms considered were calculated, and their performances between the algorithms were compared in this study.

The above formulae are often used to evaluate the performance of forecasting models. MSE is used to measure how much predictions deviate from actual values. MSE is calculated by squaring each forecast error and taking the average of these squares. This leads to greater emphasis on significant errors and attempts to minimize these errors to achieve

statistically better results. MAE measures the absolute deviation of predictions from actual values. It takes the absolute value of each forecast error and calculates the average of these absolute values. MAE is a measure in which significant errors are not emphasized more, providing a more robust evaluation.

RMSE is the square root of MSE and has the same unit of measurement as MSE. RMSE, like MSE, highlights significant errors but is a more understandable measure of errors because it is a measure that is consistent with the original data unit. A lower RMSE means that the prediction model performs better. These three metrics are explicitly used when developing and comparing predictive models, and which metric is preferred may vary depending on the nature of the data, the requirements of the application, and the objectives of the model.

The performance of different ML algorithms, such as AB, RF, GB, and LR, in predicting data performance can vary depending on several factors. These factors are based on the characteristics of the dataset, algorithm parameters, how suitable the model is for training, and more. Some factors affecting the performance of ML algorithms are key model differences, which can be expressed as dataset complexity, simple datasets, and dataset size. As a result, which algorithm will perform best depends on the characteristics and requirements of the dataset. Ideally, trying different algorithms and tuning hyperparameters is a process that should be carried out to obtain the best results.

2.3. SPC Diagrams: I-MR Chart

In this study, SPC diagrams are proposed to test the accuracy of the results of the estimation data obtained from ML models. SPC diagrams were preferred in this study, emphasizing the testing of predictive data derived by ML of a system for the future, whether the system is under control or not.

I (individual)-MR (moving range) control diagrams were created from the SPC diagrams, and forecast data's effects on process control were followed. I-MR control diagrams are used as single observations of data for measurable variables. Using this type of diagram for data of high importance in terms of cost and time provides excellent convenience. The preferred I-MR control chart for individual measurements uses two consecutive observation ranges to estimate process variability. In I-MR control diagrams, the range of motion is defined as follows:

$$MR_i = |x_i - x_{i-1}| \tag{5}$$

where MR_i is the symbol of the moving-range value for the ith observation, x_i signifies the value of the ith datum, and x_{i-1} symbolizes the value of the $(i-1)$th datum. The I-MR control chart has three limits, which are the lower (LCL), central (CL), and upper control values (UCL). The equations of these limits for the I-chart are constructed as follows:

$$LCL_I = \bar{I} - 3 \times \left(\frac{\overline{MR}}{d_2} \right) \tag{6}$$

$$CL_I = \bar{I} \tag{7}$$

$$UCL_I = \bar{I} + 3 \times \left(\frac{\overline{MR}}{d_2} \right) \tag{8}$$

where d_2 is the constant value of the statistical control charts. The d_2 value was considered to be 2.059 in terms of 4 subgroups according to the SPC chart. The equations of the lower, central, and upper limits for the MR chart were constructed as follows:

$$LCL_{MR} = D_3 \overline{MR} \tag{9}$$

$$CL_{MR} = \overline{MR} \tag{10}$$

$$UCL_{MR} = D_4 \overline{MR} \tag{11}$$

where D_3 and D_4 are the constant values of the statistical control charts, and these values are also generated using d_2 and d_3 values. The D_3 and D_4 values were considered to be 0.000 and 2.282, respectively, in terms of 4 subgroups according to the SPC chart. The observation data should preferably be normally distributed, especially since the I and MR diagrams are sensitive to deviations from normality.

3. Results and Discussion

The effects of input factors on the output factors were experienced by performing an LR analysis of the dependent and independent factors whose descriptive statistical data were obtained for solar energy production. In addition, interactive and singular Pareto statistical significance analyses of the independent variables were performed, and their significance levels were determined. The Pareto chart expressing the statistical significance of the input variables is shown in Figure 3.
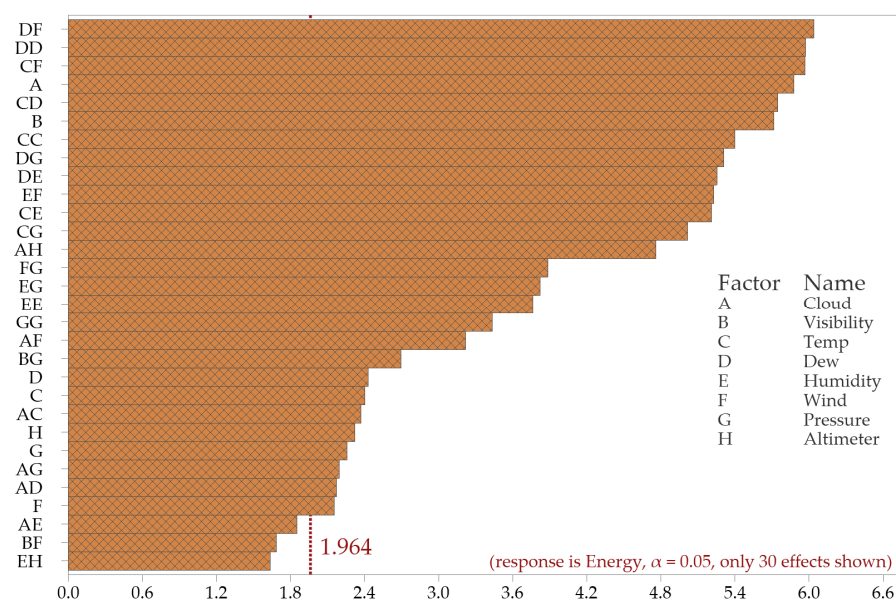

Figure 3. Statistically significant degrees of independent variables.

The Pareto chart of the independent variables expresses the absolute values of the standardized effects that consider the most significant or most minor effects of the variables on the dependent variable. It needs a threshold line (i.e., statistical significance level) to show the effect sizes of the input factors on the output factor. In this work, the reference value providing the threshold line of the Pareto chart was calculated as 1.964. Dew and wind were the most influential variables in solar energy production. While the factor with the most minor effect was wind, cloud–humidity, visibility–wind, and humidity–altimeter variable interactions stood out. Even if a single variable is ineffective on the output variable, statistically, the interaction of the same variable with another variable can be effective for the dependent variable. For this reason, statistically independent variables should be analyzed individually and interactively.

GB, RF, AB, and LR algorithms from ML models were used to obtain solar energy production prediction data. For the training and testing phases of these models, 75%/25% slicing was performed. The information about the data selected from the real data for the testing stage is shown in Figure 4.
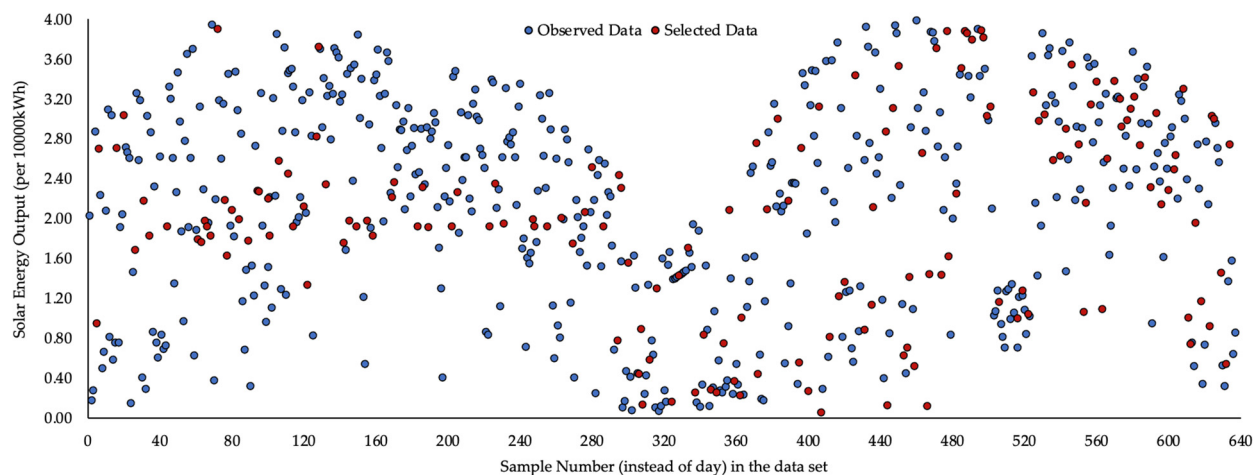
Figure 4. The selected data for the test phases in the ML algorithms.

Regression analyses of eight independent variables with numerical and continuous data types were performed according to the estimation data of the ML algorithms, and the statistical significance levels were tested. The statistical significance levels of the ML algorithms are given in Table 5. The cloud (0.001), temperature (0.030), dew (0.051), humidity (0.048), and pressure (0.001) variables were statistically effective on the actual solar energy amount data. However, the LR algorithm provided only estimation data where all variables influence solar energy. While the altimeter variable was effective on the prediction data based on the LR algorithm, its effect decreased in all other algorithms. The cloud variable had a significant impact on forecast data based on the LR (0.005), RF (0.001), GB (0.002), and AB (0.001) algorithms. Like the cloud variable, the pressure variable was effective on the forecast data of all ML algorithms (0.001 for LR, 0.003 for RF, 0.002 for GB, and 0.001 for AB). As a result, when a variable was not effective on any algorithm, it was effective on forecast data based on another algorithm. For this reason, this study used it for statistical and estimation analyses, considering all independent variables. The extended statistical results of the regression analysis of the input factors are included in Appendix A of the present study.

**Table 5.** Analysis of variance of input and output variables.

Source	Actual	LR	RF	GB	AB
Regression	0.001	0.001	0.001	0.001	0.001
Cloud	0.001	0.005	0.001	0.002	0.001
Visibility	0.969	0.012	0.031	0.007	0.476
Temperature	0.030	0.001	0.260	0.035	0.689
Dew	0.051	0.001	0.037	0.342	0.822
Humidity	0.048	0.101	0.001	0.002	0.012
Wind	0.275	0.001	0.882	0.456	0.220
Pressure	0.001	0.001	0.003	0.002	0.001
Altimeter	0.845	0.006	0.532	0.302	0.875

This study created a dual validation method to confirm the validity of the estimation of the solar energy amount obtained from the ML models. For this reason, the control chart technique was used to prove the validity of the forecast data and to test whether the forecast data obtained were under control. The number of subgroups was determined to be four when creating the I-MR control chart for the amount of solar energy—the output variable in this study. Two sources of variation emerged in the size subgroups ($n > 1$) in the I-MR control charts. These were classified as between subgroups and within subgroups in the I-MR control charts. The standard deviation values determined between and within the subgroups for the I-MR control chart created for this work are given in Table 6.

Table 6. I-MR-R/S standard deviations of actual target data.

Between	0.291249
Within	0.913997
Between/Within	0.959279

I-MR control charts were created using the real and prediction data of the dependent variable—the amount of solar energy. With these graphs, we analyzed whether a system was under control or not. For this reason, the system created with the ML algorithms was controlled by creating I-MR control charts to test whether the estimation data for the amount of solar energy production were under control. According to the subgroup chart from the I-MR control charts, the data considered in this study were outside the limits of the 11th and 12th data. According to these results, a system with these data is assumed to be out of control. However, it was found that the dataset in which accurate data were handled according to MR and standard deviation charts was under control. The I-MR control charts for the amount of solar energy, which is the output variable, are visualized in Figure 5.
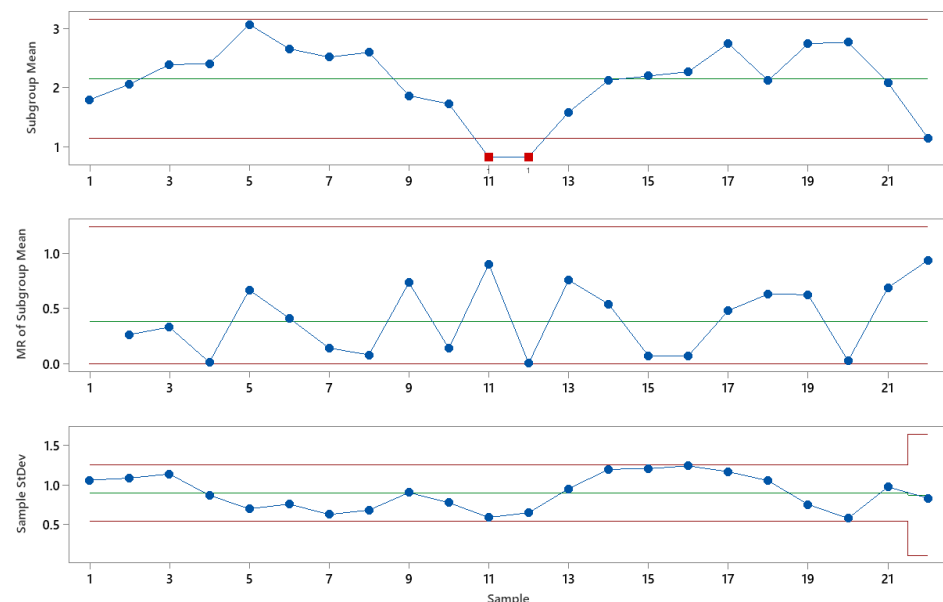
Figure 5. The I-MR control charts of energy data.

This work used the ML models AB, RF, and GB, along with LR models, to obtain estimation data of the dependent variable—the amount of solar power generation—using eight independent variables. ML algorithms usually have two phases: training and testing phases. In addition, the prediction phase and ML models involve a three-step process. The training, testing, and estimation stages' RMSE, MSE, MAE, and R^2 values were computed using the Orange 3.35 computer program. The results of the performance metrics for the ML techniques based on the testing, training, and forecast phases are given in Table 7.

The AB model, one of the preferred ML algorithms, performed best in obtaining the estimation data of the dependent variable representing the amount of solar energy production. However, the LR algorithm for the training and testing stages and the RF algorithm for the prediction phase gave poor performances. The RMSE, MSE, MAE, and R^2 values of the AB model were computed as 0.001, 0.034, 0.007, and 0.977, respectively. For the estimation stage of the RF algorithm, the RMSE, MSE, MAE, and R^2 values were computed as 0469, 0.685, 0.503, and 0.623, respectively. The mean RMSE, MSE, MAE, and R^2 values for all three phases of the LR, GB, RF, and AB models were calculated as 0.228, 0.419, 0.299, and 0.813, respectively. The suitability of using these results and the results

of each ML algorithm used for the prediction data was verified. A comparison of forecast data obtained by the AB, GB, RF, and LR models with real data is presented in Figure 6.

Table 7. The results of performance measures of ML models for testing, training, and prediction stages.

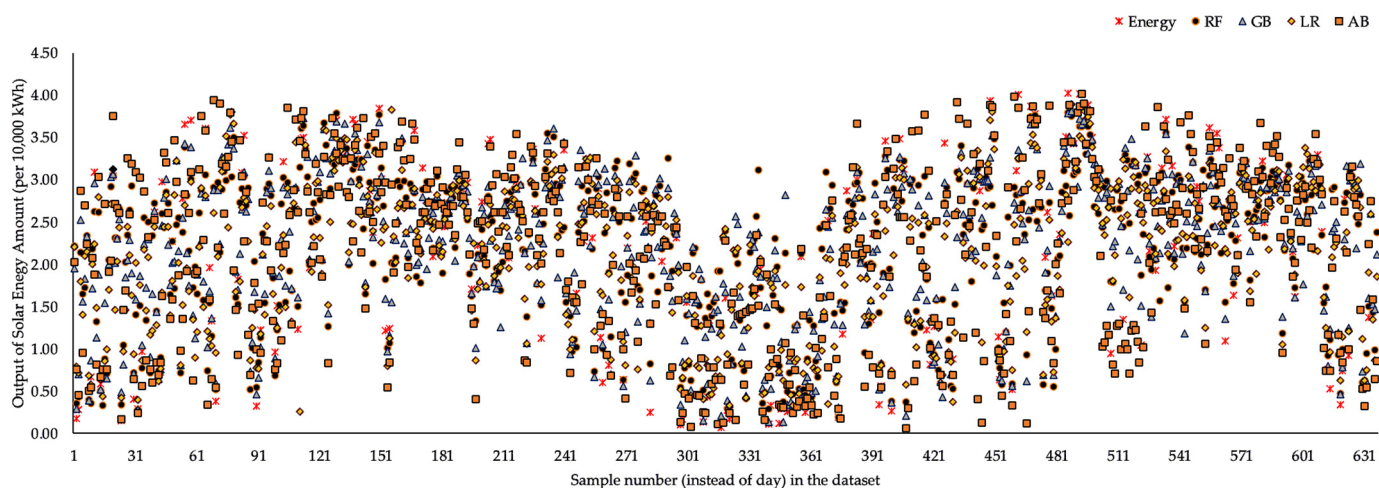| Model	MSE	RMSE	MAE	R^2	Stages
LR	0.381	0.617	0.473	0.683	
GB	0.113	0.337	0.247	0.906	Train
RF	0.048	0.220	0.165	0.960	
AB	0.001	0.034	0.007	0.977	
LR	0.444	0.666	0.507	0.619	
GB	0.182	0.426	0.320	0.844	Test
RF	0.126	0.355	0.260	0.892	
AB	0.002	0.042	0.010	0.978	
LR	0.441	0.664	0.503	0.624	
GB	0.421	0.649	0.465	0.743	Prediction
RF	0.469	0.685	0.503	0.623	
AB	0.107	0.328	0.134	0.909	

Figure 6. The comparison of actual and generated data by LR, AB, RF, and GB.

In the SPC technique, selecting subgroups in the datasets is statistically significant. Because a subgroup selection method minimizes deviations in datasets for SPC charts, I-MR control charts were obtained by forming four subgroups of the estimation data for the amount of solar energy calculated with the AB, GB, RF, and LR algorithms in this study. The standard deviation data of the within-group and between-group models for the I-MR control charts created for the ML models are given in Table 8. All of the ML models calculated the standard deviation data equally within and between groups. This situation is interpreted as meaning that the estimation data obtained by the ML algorithms for the amount of solar energy production are close to one another. Still, data relative to the actual data were obtained. The intragroup standard deviation data in the I-MR-R/S charts for the LR, RF, AB, and GB models from the ML models were computed as 0.6908, 0.7665, 0.8135, and 0.7595, respectively. The between-group standard deviation data in the I-MR-R/S charts for the LR, RF, AB, and GB models were computed as 0.3116, 0.2809, 0.3059, and 0.2925, respectively. The mean values of the standard deviation values obtained within and between groups for the ML algorithms were calculated as 0.7578, 0.8164, 0.8691, and 0.8139, respectively.

Table 8. I-MR-R/S standard deviations for ML algorithms of target data.

Models	LR	RF	AB	GB
Between	0.3116	0.2809	0.3059	0.2925
Within	0.6908	0.7665	0.8135	0.7595
Between/Within	0.7578	0.8164	0.8691	0.8139

X-bar, MR-bar, and R-bar control charts were created for each algorithm to test whether a system was under control by obtaining data on the amount of solar energy production, representing the dependent variable, using LR, AB, GB, and RF algorithms from the ML models. UCL, CL, and LCL values were calculated for each control chart. I-MR diagrams of the AB, GB, RF, and LR models are presented in Figures 7–10.
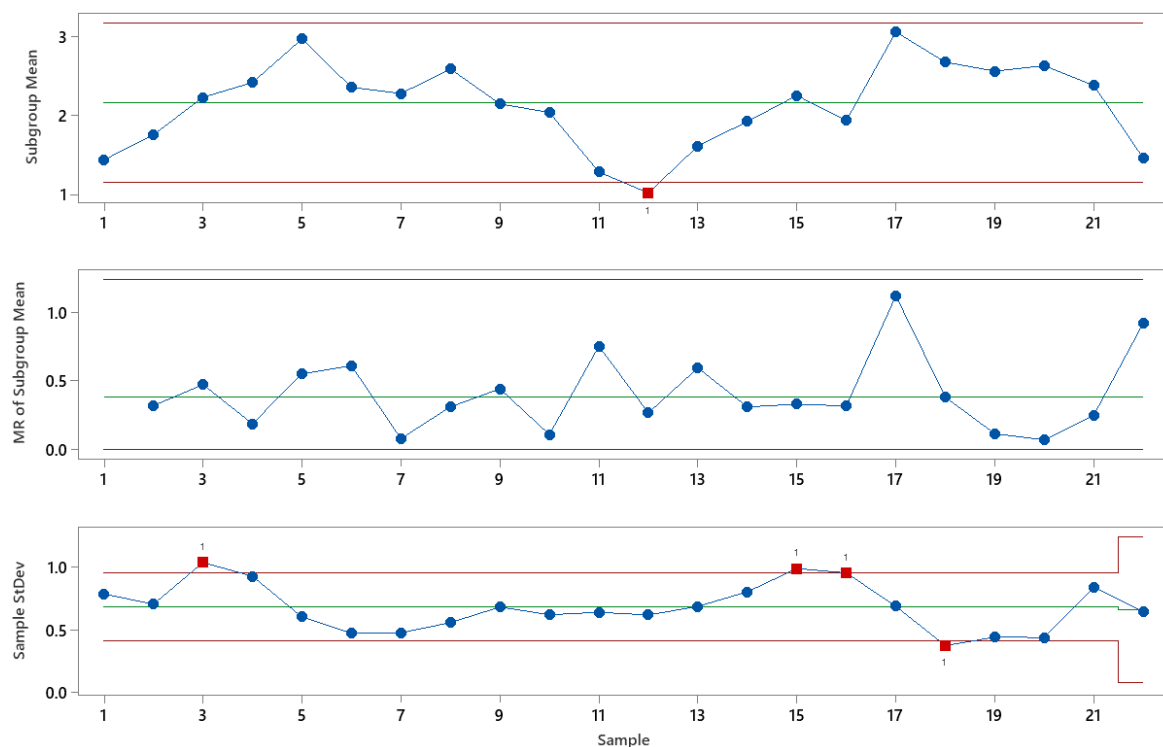
Figure 7. I-MR-R/S (between/within) chart of LR.

In all of the ML models, the prediction data for the amount of solar energy were within the control limits. Although the 12th and 49th data were out of control in the I-MR control charts obtained using actual data, only one datum was out of control in the control charts of the forecast data obtained with the RF, GB, and LR algorithms. For the control charts created with the prediction data based on ML models, the data of 636 days of solar energy production, including four subgroups, were considered. The UCL, CL, and LCL values for each control plot of the GB, AB, LR, and RF models are given in Table 9.

The UCL values of the X-bar control charts of the estimation data for the amount of solar energy production obtained according to the ML algorithms were calculated as 3.662, 3.667, 3.659, and 3.669 for the LR, AB, RF, and GB models, respectively. The same graph calculated the CL values as 3.485, 3.483, 3.484, and 3.485 for the LR, AB, RF, and GB models, respectively. The LCL values of the estimation data for the amount of solar energy production according to the ML algorithms for the X-bar graph were calculated as 3.307, 3.299, 3.308, and 3.301 for the LR, AB, RF, and GB models, respectively. According to the X-bar control charts, this has the smallest limit range (0.351). According to the RF model, the limit ranges of the AB and GB algorithms are the same (0.368), but the limit ranges of these algorithms were calculated as high.
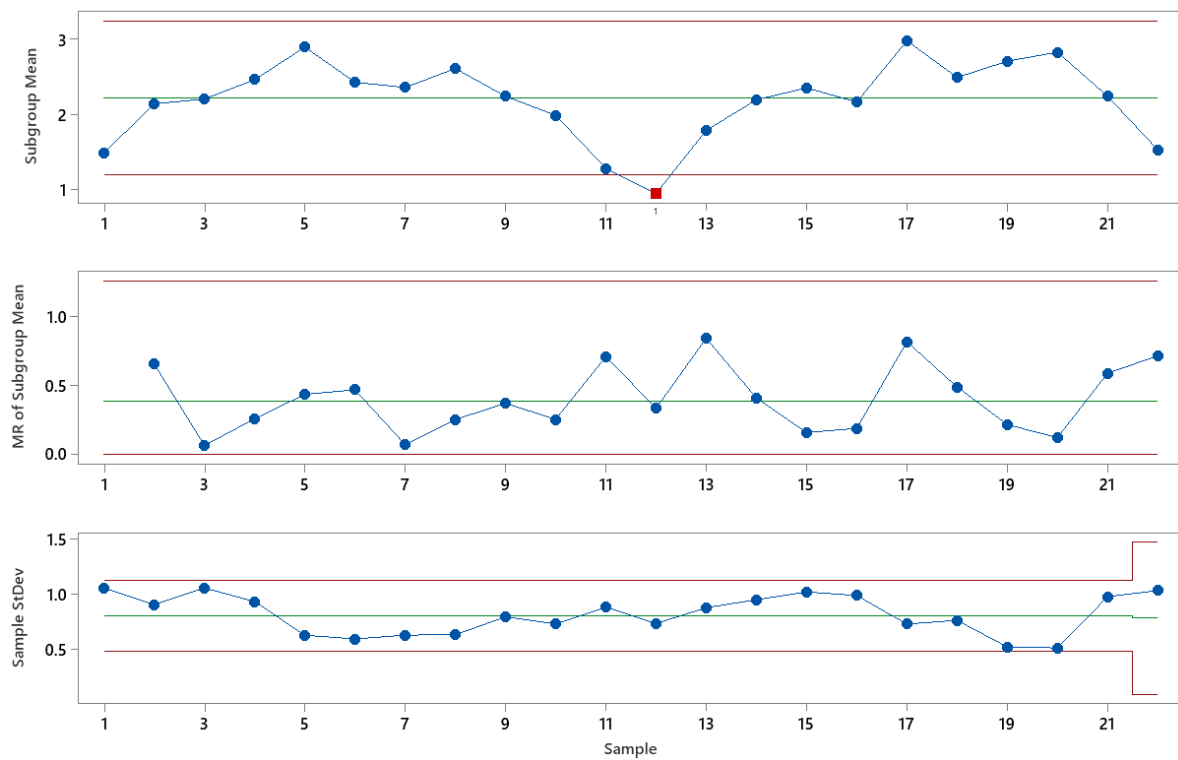
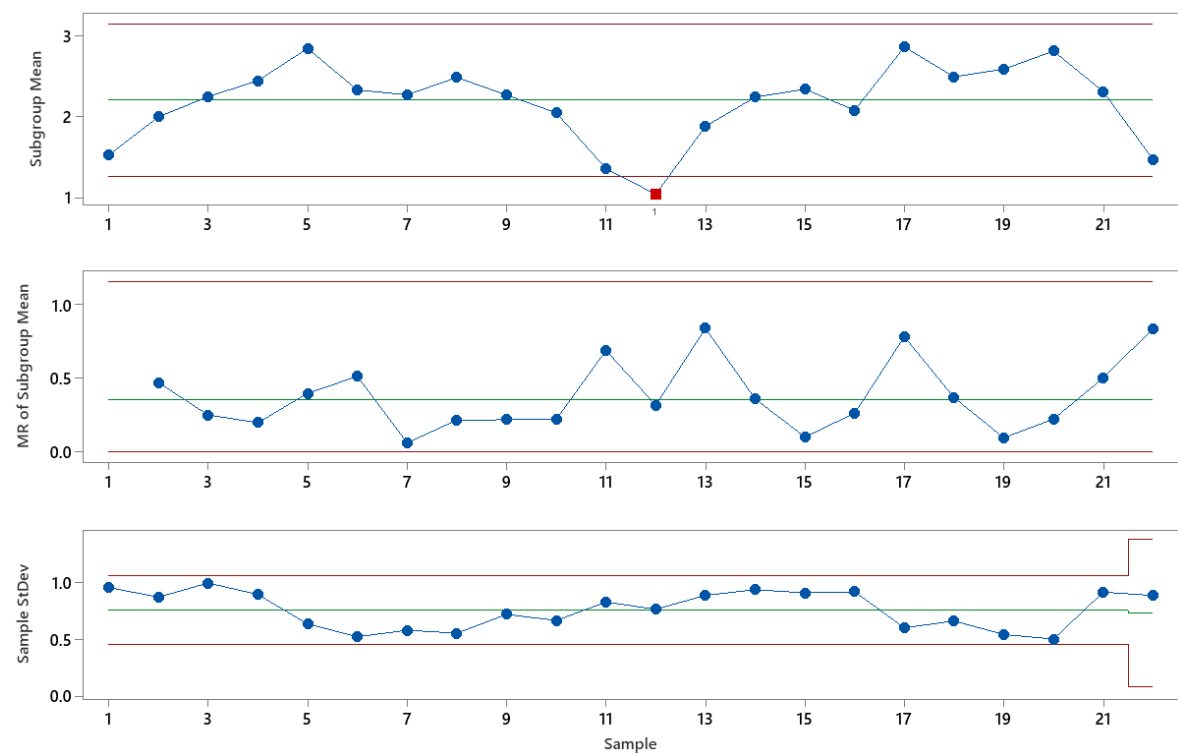**Figure 8.** I-MR-R/S (between/within) chart of AB.
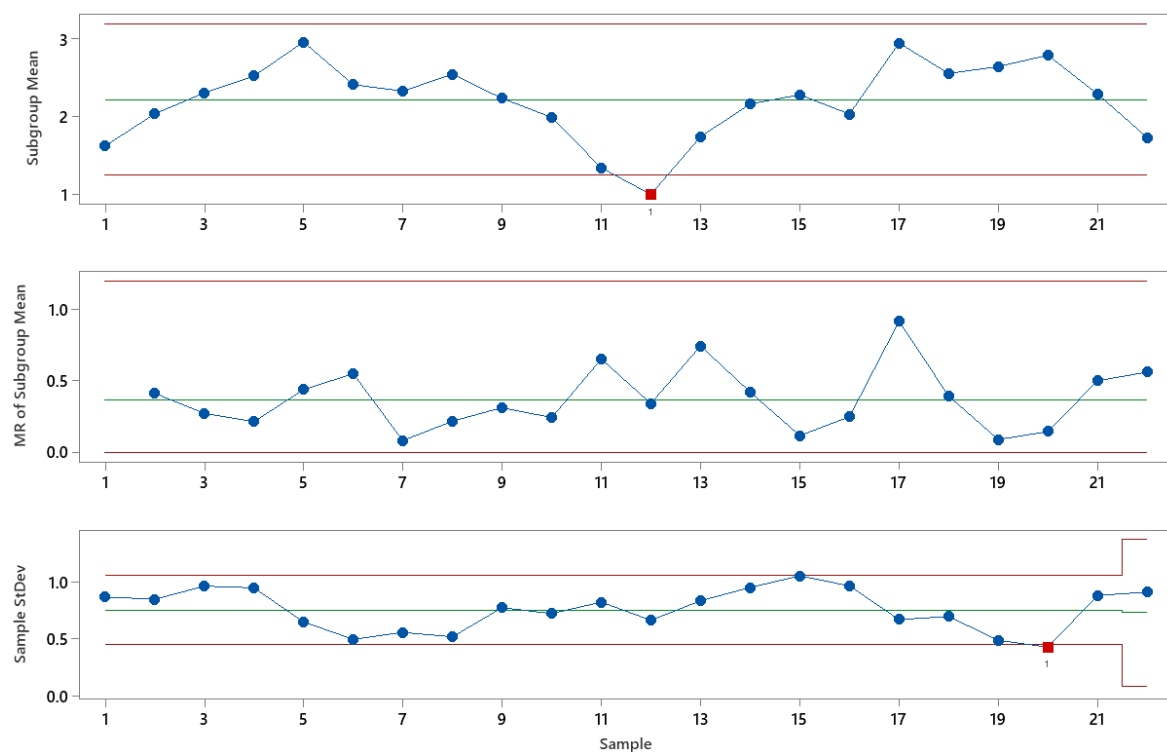

Figure 9. I-MR-R/S (between/within) chart of RF.

Figure 10. I-MR (between/within) chart of GB.

Table 9. The UCL, CL, and LCL values of the ML algorithms.

Model	Chart	UCL	CL	LCL	Point	Control
LR	X-bar	3.662	3.485	3.307	1	Out
	MR-bar	0.218	0.067	0.000	0	In
	S-bar	0.563	0.247	0.000	4	Out
AB	X-bar	3.667	3.483	3.299	1	Out
	MR-bar	0.267	0.069	0.000	0	In
	S-bar	0.573	0.251	0.000	0	In
RF	X-bar	3.659	3.484	3.308	1	Out
	MR-bar	0.216	0.066	0.000	0	In
	S-bar	0.546	0.247	0.000	0	In
GB	X-bar	3.669	3.485	3.301	1	Out
	MR-bar	0.226	0.069	0.000	0	In
	S-bar	0.577	0.253	0.000	1	Out

The LCL values of the MR-bar and R-bar graphs created with the estimation data for the amount of solar energy production based on ML algorithms were calculated as 0. In general, if the LCL values of the control process charts are negative, the LCL breakpoint is accepted as 0. The LCL values of the MR-bar and R-bar control charts created for the LR, AB, RF, and GB algorithms were accepted as 0 because they were negative.

The UCL values of the MR-bar control charts created for the LR, AB, RF, and GB models were computed as 0.218, 0.267, 0.216, and 0.226, respectively. The CL values of the MR-rod control charts were calculated as 0.067, 0.069, 0.066, and 0.069 for the LR, AB, RF, and GB algorithms of the ML models, respectively. The minimum limit range for the RF model was obtained according to the MR-rod control charts (0.216). Regarding the highest limit range, the limit range of the MR-bar graph of the AB algorithm was calculated as 0.267.

Based on the R-bar control charts, the UCL values generated for the LR, AB, RF, and GB algorithms were calculated as 0.563, 0.573, 0.546, and 0.577, respectively. The CL values

of the MR-bar control charts were calculated as 0.247, 0.251, 0.247, and 0.253 for the same ML models, respectively. The minimum limit range for the RF model was obtained (0.546) according to the MR-rod control charts. Regarding the highest limit range, the limit range of the MR-bar graph of the GB algorithm was calculated as 0.577.

Generally, integrating ML with any statistical method shows the accuracy of the results to be significant in terms of validity, although the statistical methods used in this study and existing studies in the literature differ. A study statistically integrating the DOE and ML approaches presented a hybrid model [47]. In another study, correlation analysis was performed to determine the input parameters to estimate the amount of solar energy production using ML algorithms [10]. Khan and Zeiler analyzed the prediction results obtained from ML algorithms using descriptive statistics, and as a result, they emphasized that a 10–12% improvement in R^2 values was shown in their study [48]. In another study, researchers integrated advanced statistical methods and ML algorithms to obtain forecast data for solar energy production by predicting weather parameters 24 h ahead [49].

This study has some limitations. First, solar energy production data, which represent only one dependent variable, were used in the data used for ML and SPC. Determining the number of subgroups in the dependent variable data for control charts can result in changes in the number of subgroups and control chart limits. Another limitation is that a variable with a categorical data type was not used among the response or input factors. Since the preferred dependent variable data type for ML algorithms is continuous and numeric, the algorithms must calculate F1 (i.e., the harmonic mean of precision and recall), ROC (receiver operating characteristic) curves, recall, precision, etc., and performance scores cannot be calculated. Finally, as a limit, the structural and material parameters of the PV cells used for solar energy production were considered to be fixed, without any changes. As a result of the changes to be made in the PV cells, there may be a change in the amount of energy produced. As a result, ML algorithms should be used in integration with the SPC technique to analyze whether a system is in control for the future. This study highlights the need to make a concrete decision about the future of a system by obtaining I-MR control charts based on predictive data of machine learning.

Integrating ML and SPC methods has excellent potential for improving industrial and business processes, but some difficulties and problems may arise with combining these two methods. First, data requirements can complicate the integration process. While ML algorithms usually require an extensive and high-quality dataset, SPC can rely on fewer data, so data collection and cleaning can be a significant problem. Also, incompatibilities and conflicts may arise, since these two approaches have different mathematical foundations. Second, difficulties in model training and updating can affect the integration process. ML models should be updated regularly because business processes can change over time. SPC methods can be more static, so how to integrate these two approaches on an ongoing basis can be an issue. It is also essential to know how updates are integrated into business processes and how data sources are managed. Generally, businesses can expect fast results from the integration of ML and SPC. Still, results can take time due to the complexity of these processes and the many variables that need to be optimized.

In this study, some concerns were highlighted when integrating the ML and SPC methods to predict the amount of solar energy production and to test it under control. High-quality data are needed for ML and SPC. Solar power generation data can include many variables, such as weather conditions, panel performance, and energy consumption. These data must be sensitive and accurate. Problems like lack of data, noise, and inaccurate measurements can negatively affect model predictions and process control. For this reason, some limits were applied to the preferred variables for this study. While integrating ML and SPC into solar power generation can bring many benefits, it can also come with challenges and problems.

4. Conclusions and Future Perspective

SPC diagrams are often used to test whether systems created for the manufacturing or service industries are under control. In statistical process diagrams, different charts are used according to whether the data are continuous or discrete. This study discusses eight other independent variables with numerical and continuous data types and a dependent variable representing the amount of solar energy production. In this study, the I-MR control chart was preferred because the dataset for the amount of solar energy production (the dependent variable) has a continuous quantitative data type. The dependent variable for the I-MR charts was evaluated in four subgroups.

This work sought to integrate ML and SPC graphing techniques to analyze predictive data to test whether a system would be under control in the future. It tested whether the system was under control for the future by integrating AB, RF, GB, and LR models from ML models and I-MR control diagrams from SPC diagrams. The accuracy of the control of a system was compared with the actual data by analyzing the forecast data from ML models in the I-MR control charts. In conclusion, this study suggests that valuable results can be obtained by integrating ML models with I-MR control charts. An approach has been proposed by creating a two-way validation approach to verify the validity of the results obtained by combining these two methods. A case study was carried out to show that this approach works correctly by considering the factors affecting solar energy production.

For this study, it was preferred that the dependent variable data type for the SPC charts and ML algorithms be continuous. This study is thought to help calculate performance values such as F1, recall, precision, and receiver operating characteristic (ROC), which are other performance measurement parameters of ML algorithms, especially by using variables with categorical data types. In addition, with the approach proposed in this study, it would be possible to perform *n*, *np*, *u*, and *c* techniques from dependent variable SPC charts with a discrete data type.

Abbreviations

AIM	Abductory induction mechanism	ML	Machine learning
AB	AdaBoost (adaptive boosting)	Max	Maximum value
ANN	Artificial neural network	MAE	Mean absolute error
CL	Central limit	MSE	Mean squared error
Var_{coeff}	Coefficient of variation	Mph	Miles per hour
CNN	Convolutional neural network	Min	Minimum value
DT	Decision tree	MR	Moving range
DL	Deep learning	MARS	Multivariate adaptive regression splines
DLNN	Deep learning neural network	NB	Naïve Bayes
°C	Degrees Celsius	NN	Neural network
DWT	Drinking water treatment	%	Percentage
EN	Elastic net	PV	Photovoltaic
EML	Ensemble machine learning	R^2	Coefficient of determination
XGBoost	Extreme gradient boosting	RF	Random forest

EML	Ensemble machine learning	R^2	Coefficient of determination
XGBoost	Extreme gradient boosting	RF	Random forest
GPR	Gaussian process regression	ROC	Receiver operating characteristic
GB	Gradient boosting	RR	Ridge regression
GBDT	Gradient boosting decision tree	RMSE	Root-mean-squared error
InHg	Inch of mercury	Mean	Sample mean
I	Individual	N	Sample size
IL	Inductive learning	Skew	Skewness
kNN	k-Nearest neighbor	StDev	Standard deviation
KRR	Kernel ridge regression	M_{se}	Standard error of the mean
kWh	Kilowatt hours	SPC	Statistical process control
Kurt	Kurtosis	SVM	Support-vector machine
LASSO	Least absolute shrinkage and selection operator	SVR	Support-vector regressor
LGBM	Light gradient boosting machine	UCL	Upper control limit
LR	Linear regression	Var	Variance
LCL	Lower control limit		

Appendix A

Table A1. The extended statistical results of the regression analysis of the independent variables.

Variables	Coefficient	SE of Coeff.	T-Value	*p*-Value
Cloud	−3104	2390	−5.880	0.000
Visibility	−12,326	2097	5.720	0.000
Temp	14,428	2524	2.400	0.017
Dew	54,696	22,766	−2.430	0.015
Humidity	−56,014	23,065	1.350	0.179
Wind	12,300	9138	−2.160	0.031
Pressure	−3846	1783	2.260	0.024
Altimeter	17,015	7533	−2.320	0.021
Cloud × Cloud	−4611	1986	−1.460	0.144
Visibility × Visibility	−601	411	0.800	0.426
Temp × Temp	934	1174	5.400	0.000
Dew × Dew	350,957	65,011	5.970	0.000
Humidity × Humidity	414,596	69,416	3.760	0.000
Wind × Wind	62,734	16,672	0.950	0.344
Pressure × Pressure	751	793	3.440	0.001
Altimeter × Altimeter	27,139	7900	0.130	0.895
Cloud × Visibility	142	1069	1.070	0.286
Cloud × Temp	1145	1071	−2.370	0.018
Cloud × Dew	−25,886	10,920	2.170	0.030
Cloud × Humidity	24,423	11,231	−1.850	0.065
Cloud × Wind	−9020	4871	−3.220	0.001
Cloud × Pressure	−2568	798	2.200	0.028
Cloud × Altimeter	8928	4065	−4.760	0.000
Visibility × Temp	−3802	799	1.180	0.240
Visibility × Dew	30,164	25,654	−1.220	0.223
Visibility × Humidity	−31,761	26,009	0.760	0.446
Visibility × Wind	7747	10,150	1.690	0.092
Visibility × Pressure	2455	1456	−2.700	0.007
Visibility × Altimeter	−23,154	8584	1.260	0.206
Temp × Dew	1985	1569	−5.750	0.000
Temp × Humidity	−768,914	133,804	5.210	0.000
Temp × Wind	334,280	64,179	5.970	0.000
Temp × Pressure	93118	15,606	−5.020	0.000
Temp × Altimeter	−197,378	39,348	1.190	0.234
Dew × Humidity	20,298	17,022	−5.250	0.000
Dew × Wind	−348,713	66,390	−6.040	0.000
Dew × Pressure	−96,458	15,973	5.310	0.000
Dew × Altimeter	215,750	40,650	−1.240	0.215
Humidity × Wind	−21,474	17,314	5.230	0.000
Humidity × Pressure	37,660	7205	−3.820	0.000
Humidity × Altimeter	−86,214	22,546	1.640	0.102
Wind × Pressure	12,842	7844	−3.890	0.000
Wind × Altimeter	−20,460	5266	0.330	0.743
Pressure × Altimeter	458	1399	−0.570	0.567

Abbreviation: Coeff., coefficient; SE of Coeff., standard error of coefficient.

References

1. Zazoum, B. Solar photovoltaic power prediction using different machine learning methods. *Energy Rep.* **2022**, *8*, 19–25. [CrossRef]
2. Ghose, M.K. Climate change and energy demands in India: Making better use of coal resources. *Environ. Qual. Manag.* **2012**, *22*, 59–73. [CrossRef]
3. Teke, A.; Yıldırım, H.B.; Çelik, Ö. Evaluation and performance comparison of different models for the estimation of solar radiation. *Renew. Sustain. Energy Rev.* **2015**, *50*, 1097–1107. [CrossRef]
4. Hagumimana, N.; Zheng, J.; Asemota, G.N.O.; Niyonteze, J.D.D.; Nsengiyumva, W.; Nduwamungu, A.; Bimenyimana, S. Concentrated Solar Power and Photovoltaic Systems: A New Approach to Boost Sustainable Energy for All (Se4all) in Rwanda. *Int. J. Photoenergy* **2021**, *2021*, 5515513. [CrossRef]
5. Nordell, B. Thermal pollution causes global warming. *Glob. Planet. Chang.* **2003**, *38*, 305–312. [CrossRef]
6. Chung, M.H. Estimating Solar Insolation and Power Generation of Photovoltaic Systems Using Previous Day Weather Data. *Adv. Civ. Eng.* **2020**, *2020*, 8701368. [CrossRef]
7. Kang, B.-S.; Park, S.-C. Integrated machine learning approaches for complementing statistical process control procedures. *Decis. Support Syst.* **2000**, *29*, 59–72. [CrossRef]
8. Atalan, A.; Şahin, H.; Atalan, Y.A. Integration of Machine Learning Algorithms and Discrete-Event Simulation for the Cost of Healthcare Resources. *Healthcare* **2022**, *10*, 1920. [CrossRef]
9. Aksoy, B.; Selbaş, R. Estimation of Wind Turbine Energy Production Value by Using Machine Learning Algorithms and Development of Implementation Program. *Energy Sources Part A Recover. Util. Environ. Eff.* **2021**, *43*, 692–704. [CrossRef]
10. Jebli, I.; Belouadha, F.-Z.; Kabbaj, M.I.; Tilioua, A. Prediction of solar energy guided by pearson correlation using machine learning. *Energy* **2021**, *224*, 120109. [CrossRef]
11. Vennila, C.; Titus, A.; Sudha, T.S.; Sreenivasulu, U.; Reddy, N.P.R.; Jamal, K.; Lakshmaiah, D.; Jagadeesh, P.; Belay, A. Forecasting Solar Energy Production Using Machine Learning. *Int. J. Photoenergy* **2022**, *2022*, 7797488. [CrossRef]
12. Wei, C.-C. Predictions of Surface Solar Radiation on Tilted Solar Panels using Machine Learning Models: A Case Study of Tainan City, Taiwan. *Energies* **2017**, *10*, 1660. [CrossRef]
13. Li, F.; Wu, J.; Dong, F.; Lin, J.; Sun, G.; Chen, H.; Shen, J. Ensemble Machine Learning Systems for the Estimation of Steel Quality Control. In Proceedings of the 2018 IEEE International Conference on Big Data (Big Data), Seattle, WA, USA, 10–13 December 2018; pp. 2245–2252. [CrossRef]
14. Kim, G.Y.; Han, D.S.; Lee, Z. Solar Panel Tilt Angle Optimization Using Machine Learning Model: A Case Study of Daegu City, South Korea. *Energies* **2020**, *13*, 529. [CrossRef]
15. Chou, S.-H.; Chang, S.; Tsai, T.-R.; Lin, D.K.J.; Xia, Y.; Lin, Y.-S. Implementation of statistical process control framework with machine learning on waveform profiles with no gold standard reference. *Comput. Ind. Eng.* **2020**, *142*, 106325. [CrossRef]
16. Frimane, Â.; Johansson, R.; Munkhammar, J.; Lingfors, D.; Lindahl, J. Identifying small decentralized solar systems in aerial images using deep learning. *Sol. Energy* **2023**, *262*, 111822. [CrossRef]
17. Mellit, A.; Pavan, A.M.; Lughi, V. Deep learning neural networks for short-term photovoltaic power forecasting. *Renew. Energy* **2021**, *172*, 276–288. [CrossRef]
18. Abdel-Motaleb, H. Statistical Process Control. *Cut. Tool Eng.* **2022**, *74*, 32–35.
19. Atalan, A. Forecasting drinking milk price based on economic, social, and environmental factors using machine learning algorithms. *Agribusiness* **2023**, *39*, 214–241. [CrossRef]
20. Rigatti, S.J. Random Forest. *J. Insur. Med.* **2017**, *47*, 31–39. [CrossRef]
21. López-Martínez, F.; Núñez-Valdez, E.R.; García-Díaz, V.; Bursac, Z. A Case Study for a Big Data and Machine Learning Platform to Improve Medical Decision Support in Population Health Management. *Algorithms* **2020**, *13*, 102. [CrossRef]
22. Fuentes, S.; Gonzalez Viejo, C.; Cullen, B.; Tongson, E.; Chauhan, S.S.; Dunshea, F.R. Artificial Intelligence Applied to a Robotic Dairy Farm to Model Milk Productivity and Quality based on Cow Data and Daily Environmental Parameters. *Sensors* **2020**, *20*, 2975. [CrossRef]
23. Schwendicke, F.; Samek, W.; Krois, J. Artificial Intelligence in Dentistry: Chances and Challenges. *J. Dent. Res.* **2020**, *99*, 769–774. [CrossRef] [PubMed]
24. Atalan, A.; Atalan, Y.A. Analysis of the Impact of Air Transportation on the Spread of the COVID-19 Pandemic. In *Challenges and Opportunities for Transportation Services in the Post-COVID-19 Era*; Catenazzo, G., Ed.; IGI Global: Hershey, PA, USA, 2022; pp. 68–87. [CrossRef]
25. Dönmez, C.Ç.; Atalan, A. Developing Statistical Optimization Models for Urban Competitiveness Index: Under the Boundaries of Econophysics Approach. *Complexity* **2019**, *2019*, 4053970. [CrossRef]
26. Montgomery, D.C. *Introduction to Statistical Quality Control*, 6th ed.; John Wiley & Sons: Hoboken, NJ, USA, 2009.
27. Novak, S.; Djordjevic, N. Information system for evaluation of healthcare expenditure and health monitoring. *Phys. A Stat. Mech. its Appl.* **2019**, *520*, 72–80. [CrossRef]
28. Burlikowska, D.M. Using control charts X-R in monitoring a chosen production process. *J. Achiev. Mater. Manuf. Eng.* **2011**, *49*, 487–498.
29. Duclos, A.; Voirin, N. The p-control chart: A tool for care improvement. *Int. J. Qual. Health Care* **2010**, *22*, 402–407. [CrossRef] [PubMed]

30. Veljkovic, K.; Elfaghihe, H.; Jevremovic, V. Economic Statistical Design of X Bar Control Chart for Non-Normal Symmetric Distribution of Quality Characteristic. *Filomat* **2015**, *29*, 2325–2338. [CrossRef]

31. Benitez, G.B.; Fogliatto, F.S.; Faccin, C.S.; Dora, J.M.; Torres, F.S. Productivity evaluation of radiologists interpreting computed tomography scans using statistical process control charts. *Clin. Imaging* **2021**, *77*, 135–141. [CrossRef]

32. Shewhart, M. Interpreting statistical process control (SPC) charts using machine learning and expert system techniques. In Proceedings of the IEEE 1992 National Aerospace and Electronics Conference@m_NAECON 1992, Dayton, OH, USA, 18–22 May 1992; pp. 1001–1006. [CrossRef]

33. Li, L.; Rong, S.; Wang, R.; Yu, S. Recent advances in artificial intelligence and machine learning for nonlinear relationship analysis and process control in drinking water treatment: A review. *Chem. Eng. J.* **2021**, *405*, 126673. [CrossRef]

34. Hsu, J.-Y.; Wang, Y.-F.; Lin, K.-C.; Chen, M.-Y.; Hsu, J.H.-Y. Wind Turbine Fault Diagnosis and Predictive Maintenance Through Statistical Process Control and Machine Learning. *IEEE Access* **2020**, *8*, 23427–23439. [CrossRef]

35. Khoza, S.C.; Grobler, J. *Comparing Machine Learning and Statistical Process Control for Predicting Manufacturing Performance BT—Progress in Artificial Intelligence*; Moura Oliveira, P., Novais, P., Reis, L.P., Eds.; Springer International Publishing: Cham, Switzerland, 2019; pp. 108–119.

36. Kuzmiakova, A.; Colas, G.; McKeehan, A. Short-Term Memory Solar Energy Forecasting at University of Illinois. 2017. Available online: http://cs229.stanford.edu/proj2017/final-reports/5244273.pdf (accessed on 8 August 2023).

37. Atalan, A.; Dönmez, C.Ç.; Ayaz Atalan, Y. Yüksek-Eğitimli Uzman Hemşire İstihdamı ile Acil Servis Kalitesinin Yükseltilmesi için Simülasyon Uygulaması: Türkiye Sağlık Sistemi. *Marmara Fen Bilim. Derg.* **2018**, *30*, 318–338. [CrossRef]

38. Bhavsar, S.; Pitchumani, R. A novel machine learning based identification of potential adopter of rooftop solar photovoltaics. *Appl. Energy* **2021**, *286*, 116503. [CrossRef]

39. Wang, J.; Li, P.; Ran, R.; Che, Y.; Zhou, Y. A Short-Term Photovoltaic Power Prediction Model Based on the Gradient Boost Decision Tree. *Appl. Sci.* **2018**, *8*, 689. [CrossRef]

40. Naghibi, S.A.; Pourghasemi, H.R.; Dixon, B. GIS-based groundwater potential mapping using boosted regression tree, classification and regression tree, and random forest machine learning models in Iran. *Environ. Monit. Assess.* **2016**, *188*, 44. [CrossRef]

41. Islam, S.; Amin, S.H. Prediction of probable backorder scenarios in the supply chain using Distributed Random Forest and Gradient Boosting Machine learning techniques. *J. Big Data* **2020**, *7*, 65. [CrossRef]

42. Chefrour, A. Incremental supervised learning: Algorithms and applications in pattern recognition. *Evol. Intell.* **2019**, *12*, 97–112. [CrossRef]

43. Li, K.; Zhou, G.; Zhai, J.; Li, F.; Shao, M. Improved PSO_AdaBoost Ensemble Algorithm for Imbalanced Data. *Sensors* **2019**, *19*, 1476. [CrossRef]

44. Feng, X. Research of Sentiment Analysis Based on Adaboost Algorithm. In Proceedings of the 2019 International Conference on Machine Learning, Big Data and Business Intelligence (MLBDBI), Taiyuan, China, 8–10 November 2019; pp. 279–282. [CrossRef]

45. Kavitha, S.; Varuna, S.; Ramya, R. A comparative analysis on linear regression and support vector regression. In Proceedings of the 2016 Online International Conference on Green Engineering and Technologies (IC-GET), Coimbatore, India, 19 November 2016; pp. 1–5. [CrossRef]

46. Ashrafi, Z.; Ebrahimi, H.; Khosravi, A.; Navidian, A.; Ghajar, A. The Relationship Between Quality of Work Life and Burnout: A Linear Regression Structural-Equation Modeling. *Health Scope* **2018**, *7*, e68266. [CrossRef]

47. AlKandari, M.; Ahmad, I. Solar power generation forecasting using ensemble approach based on deep learning and statistical methods. *Appl. Comput. Inform.* **2020**; *ahead-of-print*. [CrossRef]

48. Khan, W.; Walker, S.; Zeiler, W. Improved solar photovoltaic energy generation forecast using deep learning-based ensemble stacking approach. *Energy* **2022**, *240*, 122812. [CrossRef]

49. Chen, C.; Duan, S.; Cai, T.; Liu, B. Online 24-h solar power forecasting based on weather type classification using artificial neural network. *Sol. Energy* **2011**, *85*, 2856–2870. [CrossRef]