Sustainable Energy Technologies and Assessments 44 (2021) 100923

Contents lists available at ScienceDirect

Sustainable Energy Technologies and Assessments

journal homepage: www.elsevier.com/locate/seta

Original article

Assessment of predicting hourly global solar radiation in Jordan based on Rules, Trees, Meta, Lazy and Function prediction methods

Nadia Al-Rousan [*], Hazem Al-Najjar, Osama Alomari

Department of Computer Engineering, Faculty of Engineering and Architecture, Istanbul Gelisim University, Istanbul, Turkey

A R T I C L E I N F O

Keywords:
Solar radiation
Hourly prediction
Rules
Trees
Meta
Lazy
Functions

A B S T R A C T

Nowadays, predicting solar radiation is widely increased to maximize the efficiency of solar systems globally. Meteorological data from metrological stations is used to implement the intelligent prediction systems. Unfortunately, uncertainty in the used solar variables and the selected prediction models would increase the difficulties in using intelligent models to predict solar radiation. Several studies perfectly estimated solar radiation using only time and date variables. The main objective of this study is to review different prediction methods in predicting the solar radiation of Jordan. To achieve this target, five main methods including Rules, Trees, Meta, Lazy and Function Methods are selected, and then the most important and used algorithms in each method are selected to build a prediction model. The study shows that M5Rule, Random forest, Random committee, Instance Based Learning with Parameter K and multi-layer perceptron are the best algorithms in Rules, Trees, Meta, Lazy, and Function Methods respectively. Random forest algorithm performed better than other algorithms in predicting global solar radiation. The results of the analysis found that the accuracy of prediction depends on the used category, training algorithm and variables combinations.

Introduction

Jordan imports more than 90% of its energy demands from gulf countries. It faces real challenges in saving its energy supply due to the high growth in population, thus, high growth in electricity generated capacity. However, Jordan coordinates on latitude and longitude of 30.585 N and 36.2384 E respectively [1]. Jordan is considered as a country with high daily average solar irradiance of around5.5 KWH/m^2 [2]. This lead to use the solar energy in Jordan for electricity generation, water pumping, telecommunication, and lightening.

Solar radiation in sea surface in Jordan is between 4.8 and 6.4 KWH/m^2 as shown in Fig. 1. Therefore, using solar energy to produce electricity is a hot topic. In addition, Jordan is a four-season country that can be effected by the variation in time of day, weather conditions, and the sun position across the sky [3].

Nowadays, Jordan has increased its dependence on using renewable energy to produce electricity. The total renewable energy capacity in Mega Watt (MW) is shown in Fig. 2.

As shown in Fig. 2, the total renewable energy capacity between the years 2009 and 2014 was around 17 MW/year, while it has increased exponentially between the years 2015 and 2018. The maximum capacity of 1071 MW was produced in 2018, while the solar energy was slightly used in the past. Official data of renewable capacity statistics from the International Renewable Energy Agency (IRENA) have been used to compile the data [4]. It is found that solar energy has been officially used to produce electricity since 2014.

Fig. 3 shows the amount of solar energy capacity in Jordan between the years 2009 and 2018. It reached its maximum in 2018 by producing 771 MW which is very low compared to the radiated solar energy in Jordan. In addition, the total growth of solar energy capacity in 2018 is 94.7% [5].

These statistical data showed the lack of using solar energy to produce electricity while using solar energy has been around for a decade and the achievements are still limited. On the other hand, several studies discussed and analyzed the solar radiation in Jordan, and its importance to produce electricity [6–10]. While other researches discussed the topic of using solar photovoltaic (PV) modules as a promising solution to cover the energy demands in Jordan, and focused on building and implementing efficient solar systems [11–15].

However, using solar photovoltaic to build solar tracking systems has many limitations. These limitations come from the variation in the measured solar radiation based on the time and location to take the

* Corresponding author.
E-mail addresses: nadia.rousan@yahoo.com (N. Al-Rousan), hazem_najjar@yahoo.com (H. Al-Najjar), oalomari@gelisim.edu.tr (O. Alomari).

https://doi.org/10.1016/j.seta.2020.100923
Received 11 March 2020; Received in revised form 10 November 2020; Accepted 12 November 2020
Available online 16 January 2021
2213-1388/

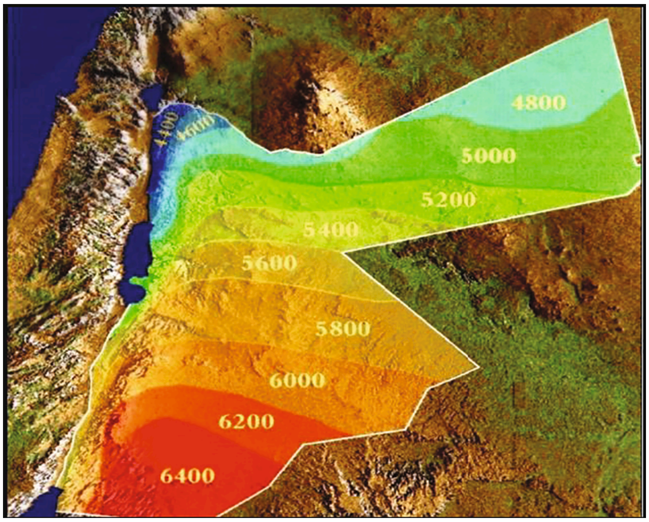

Fig. 1. Jordan Solar Radiation in WH/m^2 [3].

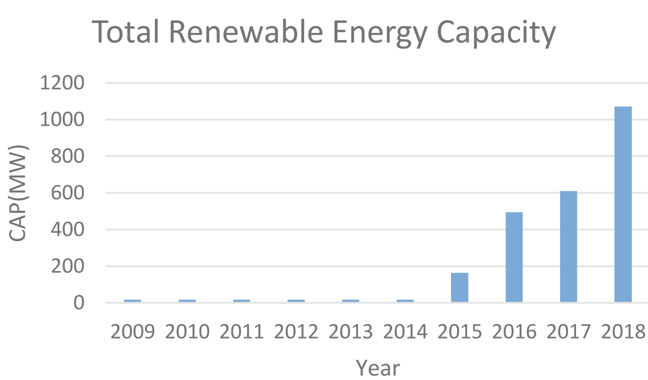

Fig. 2. Total renewable energy capacity in Jordan in MW.

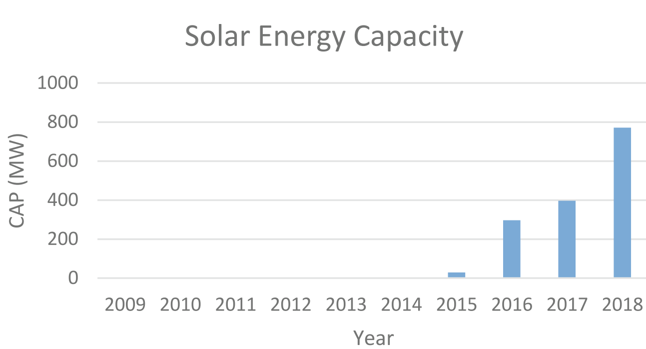

Fig. 3. Total solar energy capacity in Jordan in MW.

measurement, the randomization in direct the solar PV toward the sun, and the high cost of building and implementing solar systems. To come out with these problems, several researches conducted to determine the optimum orientation and tilt angles using measured solar radiation data [16]. On the other hand, measuring global solar radiation could be used to effectively determine the optimum directions (tilt and orientation) of solar photovoltaic to track the trajectory of the sun across the sky [17–19].

Collecting the measurements data of global solar radiation is not a naïve task. Global solar radiation data can be collected chronologically by metrological stations globally, or by using different measuring tools (i.e. pyranometer, etc.). Using these methods to measure global solar radiation in different directions, and then to determine the optimum directions to move the solar photovoltaic is a complicated process and consume time and energy. In addition, there is no rules to follow to drive the solar PV in order to obtain the maximum solar radiation. Thus, several countries developed their own stations to measure and predict the amount of solar radiation globally, and to help in building solar energy systems [20].

Moreover, several researches are conducted to use Empirical models [21,22], semi Empirical models [23,24], physical models [25], artificial intelligence models, or a combination of two or more models to estimate hourly, monthly, or yearly global solar radiation. Empirical models refers to the type of models that depends on estimating results based on empirical observations from measured metrological data (i.e. sunshine hours, air temperature, sunset hours, longitude, latitude, etc.) [22]. Semi Empirical models refer to those models that are partially empirical involving approximation, assumptions, or generalizations. Physical models refer to framework systems, physical devices, and objects that retrieve their results from measurements and experimental results (i.e. satellites, physical solar photovoltaics, etc.) [24]. Artificial intelligence models refer to computerized models that can estimate data by learning.

It is highly recommended to integrate several models together to achieve the target.

However, artificial intelligence (AI) based models are considered as high performance prediction models that can be used successfully to predict solar radiation. Artificial Intelligence learning techniques to predict the amount of global solar radiation as a first step to install solar systems [26]. This can be useful by efficiently estimate, predict, and forecast the weather conditions, the radiated power, and the optimum tilt and orientation angles [27]. It is proved that solar PV modules are strongly affected by the installation angles, and finding the optimum tilt and orientation angles could efficiently receive the maximum solar radiation [28]. On the other hand, it is found that predicting solar radiation could also maximize the generated electricity from solar energy, and would help in size photovoltaic power systems [29]. Therefore, several studies are focusing on estimating, predicting in general, or forecasting based on time series models the global solar radiation in Jordan.

Hamdan et al. [30], have used three types of artificial neural networks (ANNs) namely, feedforward neural network (FFNN), Elman neural network (Elman NN), and nonlinear autoregressive exogenous (NARX) to predict the hourly solar radiation in Amman city. Metrological data for ten years were used to train the proposed models, while another data collected in the eleventh year were used to test the proposed models. The results revealed that the three models could obtain high accuracy with predicting the solar radiation.

Badran and Dwaykat [31], have predict the monthly average daily global radiation for six major climates in Jordan. Linear regression model was used to predict the solar radiation. Comparing the estimated and experimental results proved the capability of using linear regression to predict solar radiation, and it is found that the range of the used linear regression coefficients vary from 0.7 to 0.8.

Al-Sbou et al. [20] have used seven different architectures of NARX model to predict solar radiation in Mutah city. Metrological data for daily weather condition, wind speed, and humidity were used as input variables to forecast the daily global solar radiation. The results revealed that NARX model is capable to forecast global solar radiation in Jordan.

Mohammed et al. [32] have also used NARX model to predict hourly solar radiation in Amman city. Metrological data were used to examine the proposed NARX model. By testing the proposed model, it is recommended to use NARX model to predict hourly solar radiation in Jordan.

Alomari et al. [33] have studied the correlation between solar radiation and solar PV power. Metrological data were used to predict the generated power after 24 h using artificial neural network. It is proved that prediction the power production is the most promising goal in Jordan to optimize the integration of solar PV modules.

On the other hand, several researches were published to predict, estimate, and forecast global solar radiation globally as well. These works have employed several intelligent models that could determine the optimum directions for solar photovoltaic, maximize the generated clean electricity, or size photovoltaic power systems.

Sun et al. [34] have investigated the capability of building a solar system based on predicting solar radiation variable in different three sites. Meteorological data and pollution index were considered to implement the proposed predictor. A Random Forest method was used as intelligent regression model. The results revealed that using an air pollution index to predict the solar radiation could improve the capability of the prediction model. The improvement ratio in fitting and predicting root mean square error is from 2.0% to 7.4% and from 9.1% to 17.0% respectively.

Ibrahim and Khatib [35] have proposed a prediction model to predict hourly solar radiation by considering meteorological data. Random forest was used to implement the proposed model too. Firefly algorithm was employed to optimize the number of trees in random forest method. To evaluate the proposed model different error functions were used. The results showed that the proposed model is better than the conventional random forest. The improvement ratio in root mean square error was 18.98%.

Lou et al. [36] have investigated the ability of building a predictive model using logistic regression to predict the horizontal sky-diffuse irradiance by considering meteorological variables. The study adopted different variables including stability index, clearness index, visibility, sunshine duration, wind speed, air temperature, mean sea level pressure, relative humidity, solar altitude, air temperature and cloud amount. The results showed that including the selected variables with the proposed logistic regression predictor could improve the prediction rate.

Ghimire et al. [37] have designed a hybrid model combined convolutional neural network with the Long Short-Term Memory Network to predict half-hour a head of global solar radiation. The results showed that the proposed hybrid model outperforms all other prediction models. This study revealed that convolution neural network could accurately be used to predict global solar radiation.

Gala et al. [38] have investigated the capability of different intelligent learning techniques in predicting global solar radiation. Support Vector Regression (SVR), Gradient Boosted Regression (GBR), Random Forest Regression (RFR), and a hybrid method from these methods were employed to predict solar radiation. The aim of propose a hybrid method is to improve the accuracy of the prediction model. The results revealed that hybrid model could accurately predict global solar radiation, and could improve the accuracy of the predictive model.

Due to previous works, it was proved that machine-learning methods are quite effective in predicting solar radiation. Besides, several researches have proposed different intelligent models to predict, forecast, or estimate several other variables (i.e. air pollution, energy, illuminance, etc.).

Bui et al. [39] have proposed a new prediction model based on M5Rules and genetic algorithm to predict the load heating energy in building. The dataset was taken from cml.ics.uci.edu website. Eight inputs were employed to predict two outputs. The results revealed that the proposed hybrid model is efficient to predict the output variables. In addition, the proposed predictive model is efficient to be used in other applications.

Queiroz et al. [40] have proposed a new prediction model to manage renewable energy production by driven multi-agent system. The study used wind and photovoltaic energy-based units as data sources, where different predictive algorithms are considered including Multi-Layer Perceptron, linear regression, M5P and M5Rules. The results found that using hybrid models performs better than using separate model.

Corani et al. [41] have investigated the ability of predicting air pollution variable using three different prediction methods including feed-forward neural networks, pruned neural networks, and lazy learning. Input data were collected from meteorological stations which contains meteorological variables including solar radiation, temperature, rain, pressure, humidity, and wind speed, where the output is three pollute gases including O_3, SO_2, PM10. The results found that feed-forward neural network is the most efficient model in predicting the air pollution quality.

Bellocchio et al. [42] have investigated the ability of predicting illuminance in different climatic conditions by using support vector machine regression technique. The dataset was collected from MeteoLab between October 2005 and October 2007. Day, year, and illuminance of one hour before, average illuminance of the day before, and average illuminance of the week before were employed as input variables. Where the output is illuminance. The results found that support vector machine is efficient in predicting the illuminance and can be used efficiently as a predictive model.

Ferlito et al. [43] have investigated the efficiency of predicting a grid-connected photovoltaic plant production using different machine learning techniques including Multiple Linear Regression, Regression Tree, Model Tree M5, Extreme Learning Machines, weighted k-Nearest Neighbors, Multivariate Adaptive Regression Spline, Support Vector Machines, Bayesian Regularized Neural Networks, and ensemble methods, as Random Forests, Cubist, and Extreme Gradient Boosting. Five variables including PV AC power, the total cloud cover, the ambient temperature, the Weather condition, and the plane of array irradiance clear sky model variables were employed as input variables. The results found that Cubist and M5 models performed better compared to other models. The prediction error of these models was very low, and the accuracy was very high compared with other models.

Kumar et al. [44] have investigated the ability of using different intelligent predictors to predict the potential of wind energy in Fiji. The study used different predictors including M5Rules, Attribute Selected Classifier, Multilayer Perceptron, Linear Regression, Additive Regress, Gaussian Processors, Stacking, LeastMedSq, SMOreg, Decision Tree, Regression By Discretization, Bagging, CV Parameter, MultiScheme, HoltWinter, ZeroR, Random Committee, Randomizable Filtered Classifier, RandomSubSpace, Vote, InputMap Classifier, Decision Stump, Random Forest, Random Tree, IBK, KStar, LWL, and RepTree. The results found that Randomizable Filtered Classifier algorithm performed better than other algorithms. It achieved the lowest prediction error with acceptable computational complexity.

Wang et al. [45] conducted a comparison study between Multilayer Perceptron (MLP), Artificial Neural Network (ANN) methods, Generalized Regression Neural Network (GRNN) and Radial Basis Neural Network (RBNN) in predicting daily global irradiation in China, by using Metrological parameters as model inputs. The results revealed that MLP and RBNN are robust and accurate in estimating solar radiation at various climatic zones in China.

Qin et al. [46] have investigated the capability of four well-regarded shortwave solar radiation (SSR), including a physically-based model (EPP), Yang's hybrid model (YHM), neural network mode (ANNM), and hourly solar radiation model (HSRM). These models were evaluated using metrological variables collected from 827 stations in China. The results demonstrated that YHM has superior performance compared to EPP, ANNM, and HSRM with daily mean RMSE of 2.414, 2.535, 2.855, and 3.645 MJm^{-2} day^{-1}, respectively.

Wang et al. [47] proposed and investigated the applicability of 97 models with new correlation coefficients in predicting daily diffuse radiation in different areas in China. The researchers classified these models into several categories based on (i.e., number of variables and periodicity of solar radiation. Moreover, the models fall in each category were also subdivided into various groups according to different input parameters. The proposed models were evaluated in several measurements, and the results revealed that the proposed model outperformed the existing models in the literature in terms of model performance.

Qin et al. [48] investigated the accuracy and applicability of 12 models for predicting daily Photosynthetically active radiation (PAR). Four out of 12 models are physically based models, and the remaining are artificial intelligence models. In this study, PAR dataset was created by means of meteorological observations at 2474 Chinese Meteorological Administration stations in China for the first time. The experiments of all models were carried out on PAR dataset. The results revealed that an optimization model that combines the Evolutionary algorithm and backpropagation neural network achieved the highest accuracy.

Zou et al. [49] used Coupled Model Intercomparison Project Phase 5 (CMIP5) to investigate the long-term variation and the spatial distribution in global solar energy. The finding indicates that there is a significantly decreased in global mean surface solar radiation between 1850 and 2005.

Feng et al. [50] have investigated the capability of 15 empirical models for estimation of daily diffuse solar radiation in diverse climate zones of China. The results demonstrated that the model embeds of second-order polynomial outperformed other studied models.

Wang et al. [51] used adaptive neuro-fuzzy inference systems (ANFIS), hybridized ANFIS with grid partition (ANFIS-GP), hybridized ANFIS with subtractive clustering (ANFIS-SC), and M5Tree methods for predicting daily global solar radiation in China. The methods are evaluated at different stations, and results showed that ANFIS models had resulted in the most accurate estimations at station 58238, while M5Tree managed to produce the highest accuracy at the station 51777.

From the previous works, several intelligent predictors are exist to predict different variables for different applications globally. Some of these variables were employed to predict solar variables and global solar radiation, while the majority of these variables were proposed and implemented for other kinds of applications.

Based on the literature, employing artificial intelligence models to predict, estimate, or forecast global solar radiation is the most promising task in Jordan and other countries globally. It can be generalized that using artificial intelligence in solar systems is the present and the future of renewable energy. Using artificial intelligence could develop novel technologies to reach the optimum production from natural resources, besides; it could help to get better management and distribution systems [52]. In addition, it was proofed that like other domains and applications (i.e. food, medicine, health, sport, communication, etc.), artificial intelligence would help in developing renewable energy and solar energy as well.

Using artificial intelligence predictors to predict solar radiation would help to efficiently develop, install, and integrate solar PVs. However, it is clear that no specific rules to follow in implementing intelligent systems to predict solar radiation. Thus, several different prediction models were selected to predict solar radiation. In addition, most of the current proposed techniques depend on testing their proposed models based on metrological data. To the best of authors' knowledge, no specific research publish to examine and evaluate using all prediction models methods together to predict the global solar radiation or to investigate of the optimum intelligent prediction models that can be used to predict solar radiation. Besides, few researches were concentrated on using intelligent learning techniques to predict global solar radiation in Jordan while several research were conducted to predict the solar radiation world wild.

The aim of this article is to investigate of using all methods of intelligent prediction techniques to predicting hourly global solar radiation. The most common used techniques from each AI-method are used along with the selected variables to predict global solar radiation in Jordan. Short-term real collected dataset from real PV module in Jordan are used for collecting input and output variables. The study adopted different prediction algorithms to evaluate their fitness and flexibility, data quantity availability, and algorithm complexity level in predicting hourly global solar radiation in Jordan. This research is to evaluate the capability of a satisfactorily predictor be built using short term information, and to find whether the higher complexity models are better than models with lower complexity in term of determination coefficient (R^2) and mean square error (MSE).

To come out with this research, five main methods including Rules, Trees, Meta, Lazy and Functions Methods are used to predict hourly global solar radiation. This research would help other researcher in the field globally to select the most suited, appropriate, efficient, and optimum predictive models for their proposed solar systems. Such research will be a core study and a reference to contribute to the field of solar systems based intelligent predictors.

This paper is organized as follows. Section 2 describes the used predicting models in the study besides to the preliminaries and definitions used. Section 3 presents the research methodology and the data collected. Section 4 reports the results and models performances. Finally, Section 5 reports the conclusion and the findings of the research.

Preliminaries and definitions

This section explains the main mechanisms and preliminaries that employed to predict the hourly global solar radiation. Artificial intelligence models and techniques can be classified into different five categories namely, Rules Methods, Trees Methods, Meta Methods, Lazy Methods, and Function Methods. This section is divided into five main parts to discuss these basic categories and to present the most common used methods of each type.

Rules methods

Rules methods are the intelligent methods that depend on store and manipulate knowledge to interpret information based on specific sets of rules. These rules can represent the relationship between input data and actions. Several methods are considered as Rules-based methods (i.e. M5Rule, Decision Table, Fuzzy logic, etc.). This section is to present both M5Rule and Decision Table methods.

M5Rule

M5Rules is a rule-based machine learning algorithm, which is defined as a procedure to extract rules from a model tree. This model has been used in several classification and predication applications [53–55]. In classification and prediction tasks, M5Rules sta "best" leaf is transferred into a rule, and the tree is discarded. All instances that satisfied with the rule are eliminated from the dataset. This procedure is reclusively applied in the remaining instances and stop when all instances are satisfied with one or more rules. This task is basically a separate-and-conquer strategy for learning rules; however, it should be noted that M5Rules create a full model of the tree that is different than those who utilize a regular process to create a single rule. The main benefit of extracting rule from "best" leaf to diminish the risk of over-pruning. The way of generating trees in the partial decision tree (PART) is different than M5Rules. More specifically, M5Rules generates full trees, while PART generated partially explored trees. PART has greater computational efficiency, while it doesn't have an impact on the resulting rules in terms of classification and size.

Decision table

A decision table (DT) is basically computational learning algorithm modelling, which in turn advances complicated logic [56]. DT can be formulated as a complete set of decision rules, and these rules are subjected to the mutually exclusive conditional scenarios in a pre-defined problem [57]. Conventionally, standard Decision table divided into four parts: upper left part, upper right part, lower left part, and lower right part which are represented all conditions in the problem, condition space, all the possible action subjects that utilize to make decisions, and action space, respectively.

Trees methods

Trees methods are methods which use classification trees to predict a dependent variable based on one or more of independent variables. Trees methods are commonly used in data mining.

Decision stump

A decision stump is a machine learning model that relies on tree-structure hierarchy. Its one-level decision tree consists of only one internal node that has a direct connection to the terminal nodes (known leaves). According to decision stump procedure, many researchers referred to it as “1-rule” [58].

M5p

M5P is a tree based algorithm, which is extended version of M5 algorithm [59] developed by Quinlan [60]. The key advantages of model tree that it has the capability of effectively dealing with high-dimensional datasets. It’s widely known as robust techniques especially when dealing with missing data. M5P composed of four main phases. In the first phase, the data is represented as tree-structure by divided input space into different sub-spaces. The splitting criterion is used to minimize the intra-subspace variance from the root to the node. To measure the amount of variability that is required to reach that node, standard deviation is used. Creating the whole tree is obtained by means of standard deviation reduction (SDR) factor, which, in turn maximizes the expected reduction error at the node.

If the expected error for the subtree is higher than SDR for linear model in the root of sub-tree, over-training problem will be occurred. The adjacent linear models at the pruned leave can sharply discontinuities. This problem is compensated by smoothing process in the final phase. In process of construct final model at the final leaf, the models in all leaves from the leaf to the root are combined. In this regards, to compute the leaf’s prediction value, this leaf is filtered as it paths pack to the root. Then, the leaf’s predication value is combined with predicated value obtained by applying linear regression for the same leaf is done.

Random forest

The random forests algorithm (for regression and classification) is categorized as follows:

- Draw the ntree bootstrap samples, which are derived from the original data.
- For every bootstrap samples, a regression tree or a non-pruned classification is grown based on the modification, which highlights the following. At every node, perform a random sample of the predictors and select the most suitable division from the available variables instead of selecting the most suitable division from the entire predictors. (Bagging represents a particular situation that is related to random forests).
- Predict the new emerging data by collecting the predictions pertaining to the ntree (i.e. taking most of the votes for classification and average for regression).

An error rate’s prediction is acquired according to the training data as follows:

- In every bootstrap iteration, the data is not predicted in the bootstrap sample (what Breiman calls “out-of-bag”, or OOB, data) when the tree grown is being used along with the bootstrap sample.
- Collect the predictions related to the OOB (every point of data is considered out-of-bag in about 36% of the times where such predictions should be collected). Perform a calculation for the error the error rate and start calling it through the error rate’s OOB prediction.

In fact, the error rate’s OOB prediction is extremely accurate where sufficient trees are grown (otherwise, the prediction of the OOB can be biased upwards) [61].

REPTree

REPTree is considered to represent a rapid decision tree learner that creates a decision/regression tree based on the use of an information gain for forming the split-ting criterion. Additionally, it prunes by utilising a minimised-error pruning. It just sorts values that are related to numeric attributes. Unavailable values are being handled based on the use of the C4.5 method, which basically relies on the use of different fractional instances [62].

Meta methods

Meta method contains a huge number of algorithms that use different methodology like boosting, bagging methods. This section highlights only four methods that have been used by many researchers. The methods are Additive Regression, Bagging, Random Committee and Regression by Discretization.

Additive Regression (Gradient boosting)

Additive Regression, also defined as Gradient boosting, is a machine learning method that uses multiple learning algorithms to improve the performance of the prediction model, typically additive regression uses a decision trees to achieve that purpose. Boosting is used to adjust the weight of an observation by considering only the last classification, in case of wrong prediction the weight of an observation is increased. Firstly, the model is built using a stagewise method by combing weak learners into a single strong learner, then the model uses a generalization technique to optimize differentiable loss function. Leo Breiman, [63] observed the first generation of gradient boosting, then advance version of boosting is developed by Friedman [64]. Boosting predictor has many advantages including can support different loss function, can handle the interactions between observations. Unfortunately, predictor has high probability to reach overfitting in training, besides needs extra care while choosing the parameters.

Bagging

Bagging predictor is firstly introduced by Leo Breiman [65]. The main target of the model is to decrease the variance of a decision tree by creating and selecting randomly several training subsets. In which, each tree is trained by using a collection of subsets, as a result different models are generated. Average of all tree’s models’ outputs are used as an output of the bagging predictor.

Bagging predictors have many advantages that considered by many researchers like reducing the possibility of overfitting, dealing with high dimensional dataset and can work with missing data. Unfortunately, bagging predictor considers only a mean of the predictions which minimize the capability of giving the precise values.

Random committee

Random committee is an ensemble mechanism that is built from several weak predictors [66]. The weak predictor is a prediction of a base predictor (i.e., tree). Each predictor uses the same data with different random seed. The average of all predictions is considered the result of predictor. The aim of random committee is to combine different predictors together to reduce error functions like statistical, computational and representational errors.

Regression by discretization

Regression by Discretization is a conditional density estimator that uses a probability estimator [67]. The aim of this target is to quantify and visualize the uncertainty associated with continuous target prediction. Regression by discretization is a scheme that uses with any predictor on a copy of the data that has discretized target. To find the output of the perdition model, the mean class value for each discretized interval is considered as the final output.

Lazy methods

A lazy method generalizes the complete training dataset after a query is made, because of that researchers called this model as a just-in-time learning. Lazy method has many advantages compared to other machine learning algorithms, some of these advantages are lazy approach uses with small samples and when no long information is found for the system, besides lazy method does not suffer from data interference. In contrary, lazy method needs large amount of memory to store the training data and large execution time. Lazy approach is not like eager approach that carried out the generalization before observing the new instance. Therefore, using lazy approach may have benefits than machine learning approaches such as neural networks, trees and rules methods.

Locally weighted learning (LWL)

Locally weighted learning is simple in the concept and easy to apply in predicting a new data. The main concept behind LWL is to store all the training data in a memory to use them in predicting the future by grabbing similar training data and then combine the grabbed data to predict the new instances. The combination is done either by using simple regression or sophisticated operations [68]. LWL method is separated into two steps including distance function and weighting function. Distance function measures the relevance between all training data and new instance that needed to be predicted. The distance function takes the training and new point and the distance between them is returned. While, weighting function computes the weight for each distance using different kernels. Different kernels can have different output functions, besides, choosing the most appropriate kernel that can fit the training and testing data has a high influence in the performance of LWL method. Using this capability of changing kernels can decrease the computation cost of LWL method.

K-star

K-star is one of the lazy algorithms that uses the existing training data to find the nearest points to a new data by using a distance metric [69]. To define the distance metric an entropy concept is used, which is calculated by finding the mean of complexity transforming an instance into another. The complexity is calculated using two steps including mapping one instance to another instance using a finite set of transformations and mapping one instance to another instance using a program in a finite sequence of transformations which starts at (a) and terminates at (b). Furthermore, using entropy variable in defining the distance, gives a k-star method an ability to show good performance, if the dataset contains imbalance samples.

Instance based learning with Parameter k (IBK)

IBk defines as Instance-Based and k determines the analyzed number of neighbors (k) [70]. IBK method finds a regression values of new data by using the implementation of k-nearest-neighbor method. K-nearest-neighbor uses a Euclidean distance metric to define each new instance compared with existing ones, in addition for regression problem, IBK uses a distance weighting to estimate the output of a new input.

Function methods

Function methods are used to predict new values by using a function (s) that created using different weighting algorithms. The target of function method is to define a prediction model as a function of variables. Many models used a function to predict a future data including multilayer perceptron, linear regression and support vector machine regression. In this section, the main function prediction models are considered.

Support Vector Machine Regression

Support Vector Machine Regression (SVMR) is a popular machine learning model that used to predict the behavior of future using currant information. Support vector machine regression (SVMR) is used for continuous and discrete values that used kernel functions to estimate the performance. A kernel function used a non-linear mapping between input and output variables [71]. Kernel function is used to transform a non-linear decision surface to a linear equation with multiple dimensions [72]. Based on a loss function of distance measure [73], the support vector machine regression is divided into linear and non-linear support vector machine. Many studies showed that nonlinear kernel is accurate than linear kernel.

Multi-layer perceptron (MLP)

Multi-layer perceptron contains three layers including input layer, hidden layer(s), and output layer. In each layer there is a small node called neuron. Neurons are considered as independent processing units that can connect different layers together using weights [62]. MLP uses many optimization algorithms to adjust the weights between neurons to find the most appropriate weights' value that can be used to find the output of the problem more accurately. Input, hidden and output layers are used to receive input, receive the output of the input layer and receive the the outputs from the hidden layers, respectively. The output layer is responsible to find the predicted values using the following formula [74]:

$$output = f\left(\sum_{j=1}^{2} O_j w_{jk} + b_k\right) \tag{1}$$

where, K is the number of nodes in the output layer, $f(.)$ is the transfer function, w_{jk} is the weight f, and b_k is the bias. The performance of MLP depends on changing several parameters including number of inputs, number of hidden layers, number of hidden nodes, bias, weight, the used optimization algorithm, changing the performance target and the type of transfer function

Linear regression (LR)

Linear regression uses to draw a linear relationship between inputs (independent variables) and output variables (dependent variable) by fitting a predefined linear equation based on observed data. In case more than one independent variable is considered, the equation is defined as multiple linear regression.

Research methodology

This section is dedicated to deal with the methodology of this research. This chapter is partitioned into three main sections. The first section describes the data samples that used to implement the proposed systems. The second section focuses on using the discussed intelligent predictors to predict hourly global solar radiation. The last section describes the performance criteria that used to evaluate the proposed systems.

Data samples

In this research, a practical dataset collected by real high efficiency 120 W polycrystalline photovoltaic module of model KC 120-1 is used. The dataset collected practically by measuring hourly global solar radiation using a Pyranometer device [12,75]. A Pyranometer device of model (Nr.8921) and factor (k = 80.4) was fixed horizontally on the surface of photovoltaic module that moves to both directions horizontally and vertically. To collect the total radiation of optimum tilt and orientation angles, several devices and instruments were connected to the solar photovoltaic module namely, a voltmeter, an ammeter, a pyranometer, and a variable load resistance (0–10 KΩ) as shown in Fig. 4 [12].

To take measurements, the adopted process involved mainly depends on moving the photovoltaic module shown in Fig. 4 in both directions, horizontally and vertically at the same time. The idea from this step is to find the optimum tilt and orientation angles that allow solar radiation to fall onto the surface of the solar panels without shading. Variable load resistance values are used to measure different voltage and current values between open circuit voltage (V_{oc}) and short circuit current

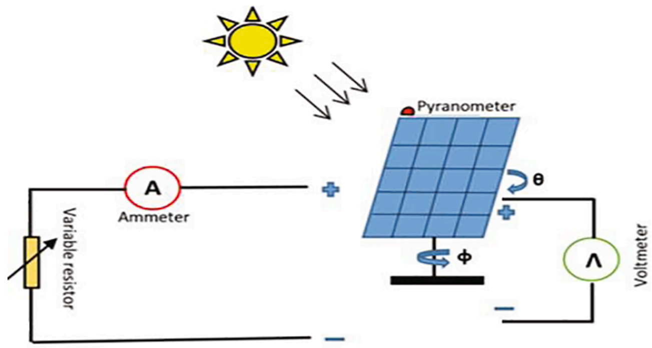

Fig. 4. Circuit diagram of solar tracking system connected to voltmeter, ammeter, and pyranometer [77].

(I_{sc}).These processes were repeated hourly from 8:00 am to 6:00 pm for different days in a month, and for different months in a year to cover the four-season conditions, cloudy, partial cloudy, and sunny days. Astronomical recommended average days for the adopted months were selected to take measurements. Average day concept is defined as a specific day for each month that can represent the whole month, and give a general idea about the measured results in that month.

By implementing the adopted processes of data collection, a dataset of time, day, month, I_{sc}, V_{oc}, current, Voltage, tilt angle, orientation angle, and the power radiation variables were recorded. For each day sample, eight observations were recorded. Meanwhile, the missed or out of range data were deleted, then the collected data were analyzed, characterized, and normalized to ensure about its capability and robustness to be used in the proposed methodology. Current-voltage characteristic curves, maximum power point (PPT) values, output power of a photovoltaic module, and power-voltage curve were calculated for each data sample. Besides, the dataset is employed to propose several intelligent prediction models [75–78].

For implementing the proposed methodology, and based on the recommendation of several research, the dataset is constituted by three variables including time, day and month, where solar radiation is selected as output variable. The data is collected for different months to cover different seasons, and the solar radiation value is recorded as W/m2. Moreover, to build prediction models, dataset was normalized, the empty values were removed, and the out of range values were fixed. Then, the dataset is randomly divided into two parts namely, training and testing. Training dataset comprises of 70% of data, while the remaining 30% of data is adopted as testing data.

Global solar radiation prediction methodology

To predict hourly global solar radiation using different AI predictors, the dataset is divided into two datasets including training and testing data. The training dataset is used to train the used predictor and then the testing dataset is used to estimate the capability of the predictor to fit a future data. The training dataset is used to build a predictor and then a testing dataset is used to validate predictor in predicting a future data. All the models are trained using different prediction models and the results of training and testing datasets are compared using different error functions. In each category that shown in Table 1, the results of the models are compared, and the best model is chosen as an optimal model for the category. Then, the results are compared together to find the most optimal model in all categories that can estimate a global solar radiation in Jordan and the conclusion is drawn based in the results analysis as shown in Fig. 5.

Performance metrics

To evaluate the performance of the selected predictors, several metrics are considered (i.e. the determination coefficient (R^2), Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Mean Bias Error (MBE), Mean Square Error (MSE), and Mean Absolute Percentage Error (MAPE) are used.

Table 1
Used methods based on the category.

Category	Method
M5P	Random Forest REPTree
Meta Methods	Additive Regression (Gradient boosting) Bagging Regression by Discretization Random Committee
Lazy Methods	locally weighted learning (LWL) Kstar Instance Based Learning with Parameter K (IBK)
Function Methods	Support Vector Machine Regression Multi-Layer Perceptron (MLP) Linear Regression (LR)

$$R^2(\text{determinationcoefficient}) = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \overline{y})^2} \quad (2)$$

$$MAE = \frac{1}{N}\sum_{i=0}^{N}|y_i - \hat{y}_i| \quad (3)$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=0}^{N}(y_i - \hat{y}_i)^2} \quad (4)$$

$$MBE = \frac{1}{N}\sum_{i=0}^{N}y_i - \hat{y}_i \quad (5)$$

$$MSE = \frac{1}{N}\sum_{i=0}^{N}(y_i - \hat{y}_i)^2 \quad (6)$$

$$MAPE = \frac{1}{N}\sum_{i=0}^{N}\frac{(y_i - \hat{y}_i)}{y_i} \quad (7)$$

where y_i , $\hat{y}_i$ and, $\overline{y}$ are solar radiation and the predicted solar radiation, the mean of solar radiation, respectively. In addition, N and i are the number of samples and the index of the data in the dataset, respectively.

Results, discussion and analysis

The results of different algorithms are reported based on the category of each algorithm. In this section, five categories are explained including Rules, Trees, Meta, Lazy and Functions Methods. Afterward, the best algorithm in each method is elected and compared together to find the best models in all categories.

Rules methods

Based on testing the trained intelligent prediction models, Decision Table and M5rule showed approximately close prediction rate and error values as shown in Table 2. Table 2 shows that Decision Table performed better than M5rule in the training process with highest R^2 and error functions, where M5rule showed better R^2 and error functions in testing process. The results revealed that using M5rule method is more efficient than using Decision Table in predicting solar radiation with lowest MSE, MAE, MBE, RMSE, and MAPE. The predicted values of solar radiation for different samples and the error of both models in testing process are shown in Fig. 6. It is clear from Fig. 6 that the radiated solar energy

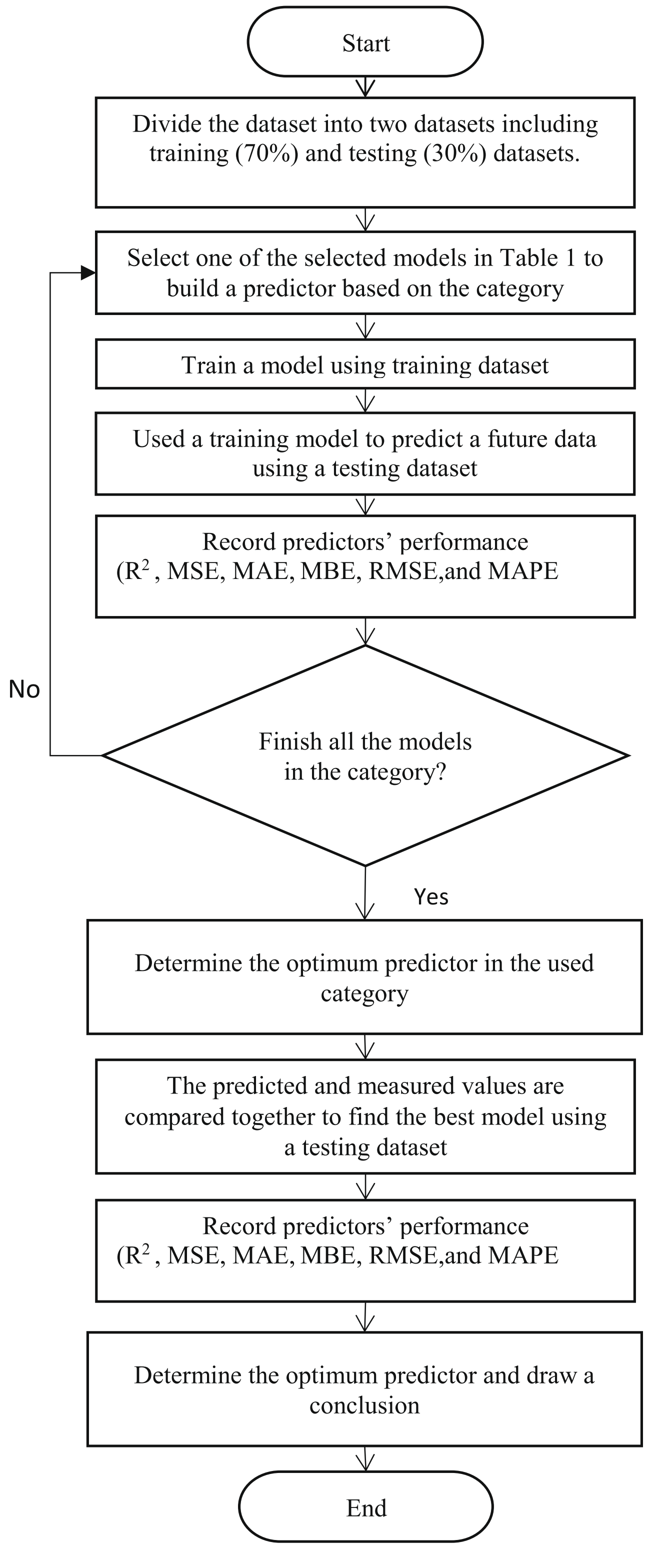

Fig. 5. Flowchart of prediction models based on the category.

values starting from low solar radiation for winter days and these values rapidly daily increased until reach its maximum values during summer. Besides, the solar radiation values normally start from low values in morning time and reach its maximum in noon time. Decision Table failed to predict this behavior for several samples, while M5rule predicted approximate close values to measured values and similar behavior with prediction rate of 94.4%. For error comparison, the error values of M5rule and Decision Table vary from one sample to another.

Trees methods

Decision Stump algorithm showed the worst performance in both training and testing processes as shown in Fig. 7 and Table 3. While Random Forest showed the highest performance compared to other algorithms in both training and testing processes. In contrast, the performance of M5D and RepTree algorithms is very close, where both algorithms could highly predict the values of global solar radiation. Besides, they achieved median error values compare to Decision Stump algorithm. Comparing the four algorithms together, the results revealed that predicting solar radiation using Random Forest is more efficient than using M5D, RepTree and Decision Stump algorithms. Fig. 7 shows the predicted solar radiation for different samples and the error of both models in testing process. It is clear from Fig. 7 that Decision Stump algorithm failed to predict the value of solar radiation because it only predicted constant values for most of data samples. In addition, RepTree failed to predict the values of solar radiation in several samples, while both M5P and Random Forest algorithms could successfully predicted the amount of radiated energy starting from low solar radiation for winter days and these values rapidly daily increased until reach its maximum values during summer as usual behavior.

Meta methods

All chosen predictors including Additive Regression, Bagging, Random Committee, and Regression by Discretization showed high performance with high R^2 and low error functions. As shown in Table 4, Additive Regression algorithm obtained the lowest R^2 values in both training and testing processes, while Random Committee obtained the highest R^2 values in training process, where Bagging algorithm obtained the highest R^2 in testing process with relatively low difference with Random Committee algorithm. On the other hand, Random Committee algorithm achieved the lowest error functions in both training and testing processes. As a result, Random Committee algorithm showed more reliable values and better performance, furthermore, Random Committee is robust than Additive Regression, Bagging and Regression by Discretization methods in predicting the values of solar radiation. Fig. 8 shows the predicted values of solar radiation for different samples and the error of all models in testing process. As shown in Fig. 8, the four algorithms of Additive Regression, Bagging, Random Committee and Regression by discretization could successfully predict the radiated power values with relatively close behavior. While error figure shows almost close variation in error values for all algorithms. This supports that the performance for each used algorithm are close to the performance of other algorithms.

Lazy methods

Using Lazy methods including LWL, Kstar and IBK to predict a solar radiation showed a variation in the performance from one model to another. LWL predictor achieved the worst R^2 in the training process, where Kstar achieved the lowest R^2 in the testing process. While IBK predictor showed the highest R^2 for both training and testing processes as shown in Table 5. In testing process, the performance of Kstar is not acceptable, since Kstar predictor showed overfitting which leads to reject Kstar results. In addition, IBK obtained the optimum performance with zero error functions in training process, where for testing process, IBK outperformed other predictors with RMSE around 33 which is considered as low error function compared to other predictors. Therefore, the results indicated that using IBK algorithm is better than using both LWL and Kstar algorithms. Hence, no overfitting detected and the performance functions are better and stable compared to LWL and Kstar. As shown in Fig. 9, it is clear that there is a problem in the predicted values of solar radiation in both LWL and Kstar algorithms. LWL failed to

Table 2
Performance of M5Rule and Decision Table compared to measured solar radiation

Dataset	Model	R^2	MSE	MAE	MBE	RMSE	MAPE
Training	M5Rule	0.9580	1079	25.4973	0.3749	32.8532	0.0431
	Decision Table	0.9627	960	24.2057	0.0015	30.9765	0.0318
Testing	M5Rule	0.9440	1458	30.2252	−5.9269	38.1858	0.0506
	Decision Table	0.9413	1494	29.7043	−9.7546	38.6542	0.0054

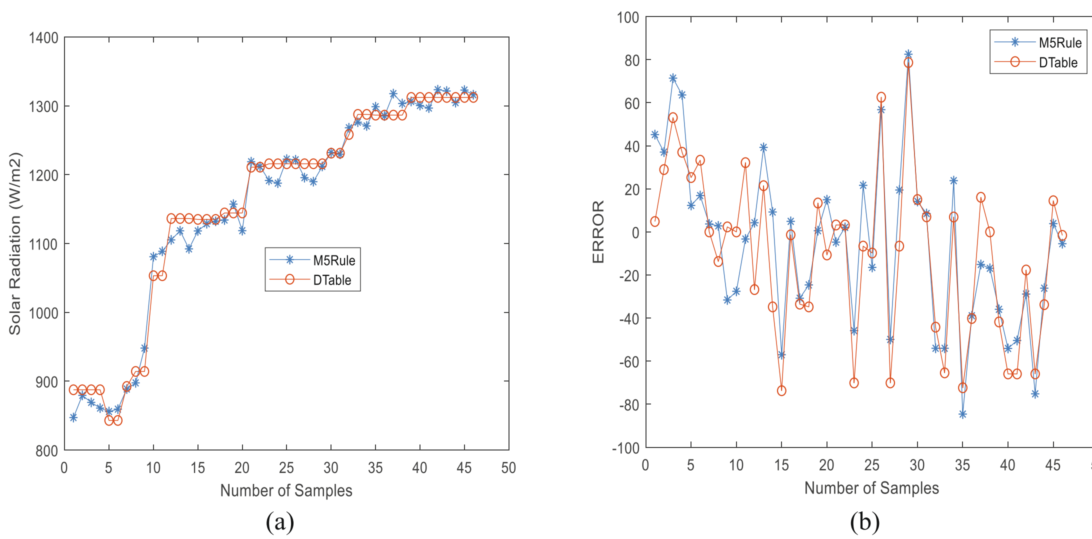

Fig. 6. The performance analysis of testing process based on M5Rule and Decision Table Rules (Dtable) (a) Predicted values (b) Error functions.

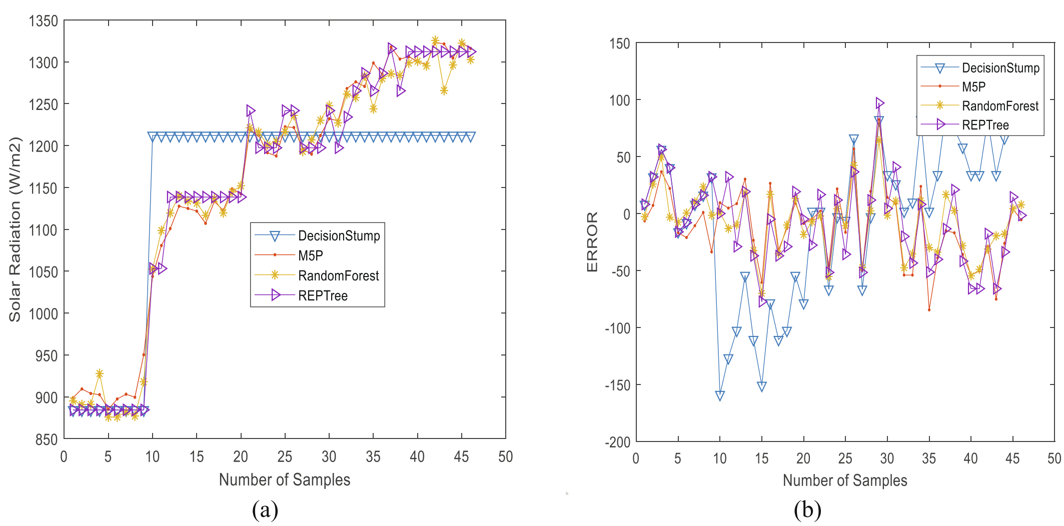

Fig. 7. The performance analysis of testing process based on Decision Stump, M5P, Random forest and RepTree methods (a) predicted values (b) Error functions.

Table 3
Performance of Decision Stump, M5P, Random forest and RepTree compared to measured solar radiation.

Dataset	Model	R^2	MSE	MAE	MBE	RMSE	MAPE
Training	Decision Stump	0.78803	5449	59.91528	−0.02254	73.81674	0.0256
	M5P	0.96286	1007	24.60841	−2.93347	31.73728	0.0313
	Random Forest	0.99453	146	9.43836	0.21114	12.09173	0.0120
	REPTree	0.96315	947	23.24687	0.00150	30.77760	0.0367
Testing	Decision Stump	0.75476	4779	54.77687	−1.62897	69.12749	0.0090
	M5P	0.94585	1236	27.63663	−10.75263	35.15359	0.0073
	Random Forest	0.96370	826	21.75170	−7.64674	28.73501	0.0030
	REPTree	0.94239	1388	30.38246	−7.51041	37.25966	0.0090

Table 4
Performance of AdditiveRegression, Bagging, Random Committee and Regression by discretization compared to measured solar radiation.

Dataset	Model	R^2	MSE	MAE	MBE	RMSE	MAPE
Training	Additive Regression	0.9602	1056	25.6837	−0.0053	32.4914	0.0097
	Bagging	0.9734	684	20.1158	0.1435	26.1610	0.0364
	Random Committee	0.9999	3	1.1564	−0.0008	1.8130	0
	Regression By Discretization	0.9714	736	19.5966	0.0030	27.1245	0.0288
Testing	Additive Regression	0.9331	1670	34.2119	−6.7816	40.8622	0.0267
	Bagging	0.9538	1117	26.1807	−7.6118	33.4264	0.0089
	Random Committee	0.9526	1079	24.9537	−8.8038	32.8472	0.0101
	Regression By Discretization	0.9382	1447	30.0172	−9.0398	38.0418	0.0058

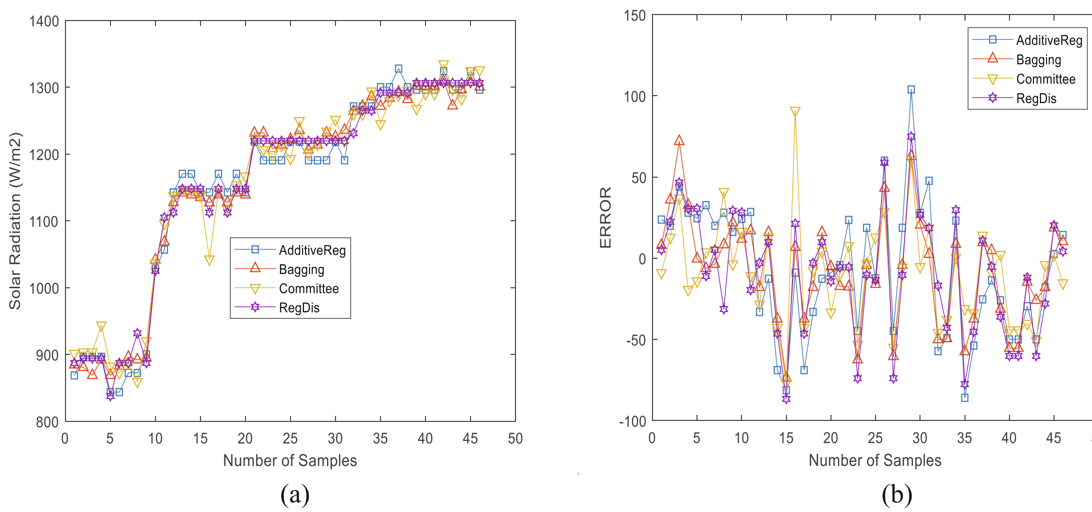

Fig. 8. The performance analysis of testing process based Additive Regression, Bagging, Random Committee and Regression by discretization (a) datasets using predicted values (b) Error functions.

Table 5
Performance of LWL, Kstar and IBK compared to measured solar radiation.

Dataset	Model	R^2	MSE	MAE	MBE	RMSE	MAPE
Training	LWL	0.8404	4146	52.5050	3.5000	64.3914	0.0178
	Kstar	0.9998	6	1.5892	−0.0293	2.4231	0.0006
	IBK	1.0000	0.0000	0.0000	0.0000	0.0000	0
Testing	LWL	0.8145	3565	48.0879	2.0048	59.7117	0.0136
	Kstar	0.8070	4315	42.9249	−25.1390	65.6863	0.0091
	IBK	0.9535	1106	25.6756	−10.2597	33.2596	0.0180

predict the majority of the collected solar radiation, besides, Kstar algorithm showed instability in predicting the correct behavior. This stability problem is clearly shown in the error figure. The results revealed that the error of using IBK algorithm is stable, thus IBK is capable to be used in predicting solar radiation values.

Function methods

SVMR and LR algorithms showed very close performance with almost similar R^2 and different performance values as shown in Table 6. SVMR algorithm obtained the lowest R^2 in training process, while both LR and SVMR algorithms obtained the same value of R^2 in testing process. For performance functions, SVMR obtained better performance than LR in all performance metrics in testing process. On the other hand, MLP using one hidden layer has the optimum R^2 values in both training and testing processes, besides it obtained the lowest performance error functions compared to other algorithms. MAE. MBE and RMSE are 1146, 27, −12 and 34 respectively. It is clear from Fig. 10, the three methods could successfully predict the values of solar radiation, as shown in figure, SVMR and LR showed identical behavior for both predicted output and error figures. In contrast, MLP showed the lowest error values and it is stable to predict solar radiation values. Therefore, MLP algorithm can be adopted as the optimum predictor compared to SVMR and LR algorithms in predicting a solar radiation values.

Results comparison and discussion

In this section, the optimum predictors from different categories are compared together to investigate of optimum algorithm to predict the hourly global solar radiation. M5rule, Random Forest, Random committee, IBK, and MLP algorithms could efficiently predict the solar radiation values with relatively high R^2. All of the selected algorithms could predict more than 94% of actual data in testing process as shown in Table 7. However, for error functions showed huge variations using different error functions. M5rule and MLP showed the worst performance results compared to other predictors for R^2 values and error functions. IBK and Random committee showed approximately similar performance for R^2 and error functions. This result can be supported by Fig. 11. It is clear from the figure, limited variation exist between the selected algorithms, besides, all algorithms showed almost similar

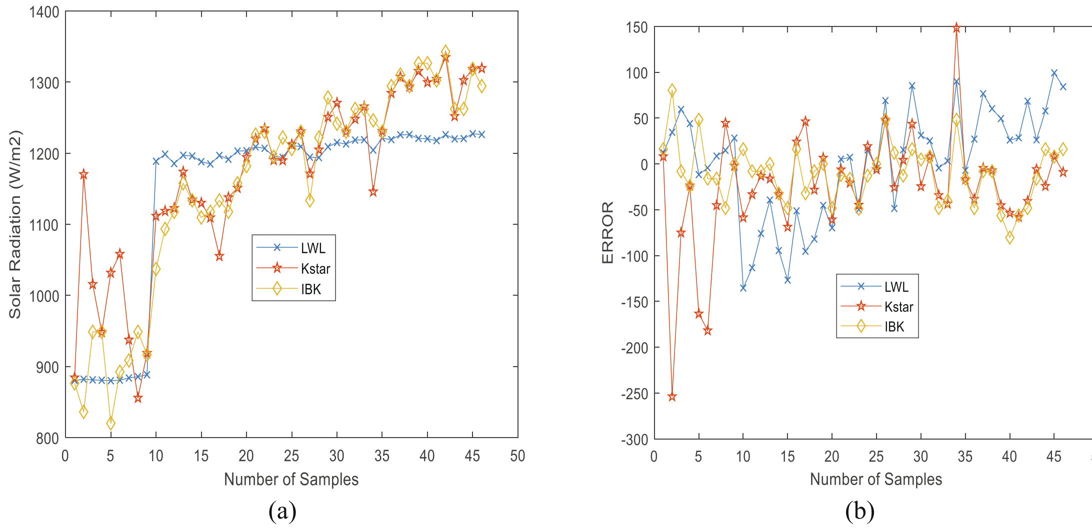

Fig. 9. The performance analysis of testing (a,b) datasets using predicted values (a,c) and Error functions (b,d) based LWL, Kstar and IBK.

Table 6
Performance of MLP, SVMR and LR compared to measured solar radiation.

Dataset	Model	R^2	MSE	MAE	MBE	RMSE	MAPE
Training	MLP	0.9618	1146	27.2038	−12.2298	33.8521	0.0129
	SVMR	0.8558	3729	48.5323	3.2859	61.0635	0.0607
	LR	0.8595	3612	49.6714	−0.0008	60.1032	0.0670
Testing	MLP	0.9513	1559	32.5585	−20.1245	39.4872	0.0001
	SVMR	0.8477	3639	48.5179	−0.9980	60.3252	0.0418
	LR	0.8477	3925	51.3424	−4.5251	62.6489	0.0434

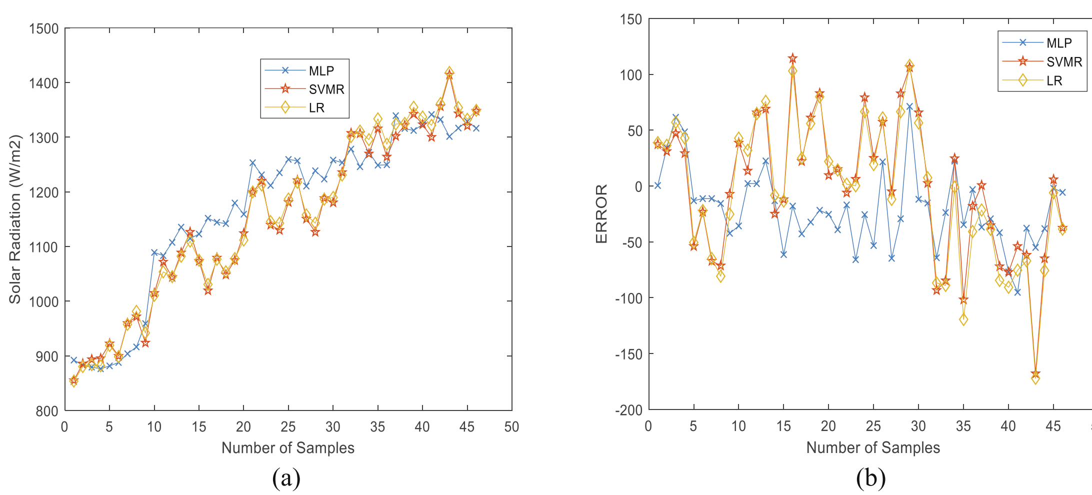

Fig. 10. The performance analysis of testing (a,b) datasets using predicted values (a,c) and Error functions (b,d) based MLP, SVMR and LR.

Table 7
Performance of M5Rule, Random Forest, Random Committee, IBK and MLP compared to measured solar radiation using testing dataset.

Models	R^2	MSE	MAE	MBE	RMSE	MAPE
M5Rule	0.9440	1458	30.2252	−5.9269	38.1858	0.0431
Random Forest	0.96370	826	21.75170	−7.64674	28.73501	0.0030
Random Committee	0.9526	1079	24.9537	−8.8038	32.8472	0.0101
IBK	0.9535	1106	25.6756	−10.2597	33.2596	0.0180
MLP	0.9513	1559	32.5585	−20.1245	39.4872	0.0001

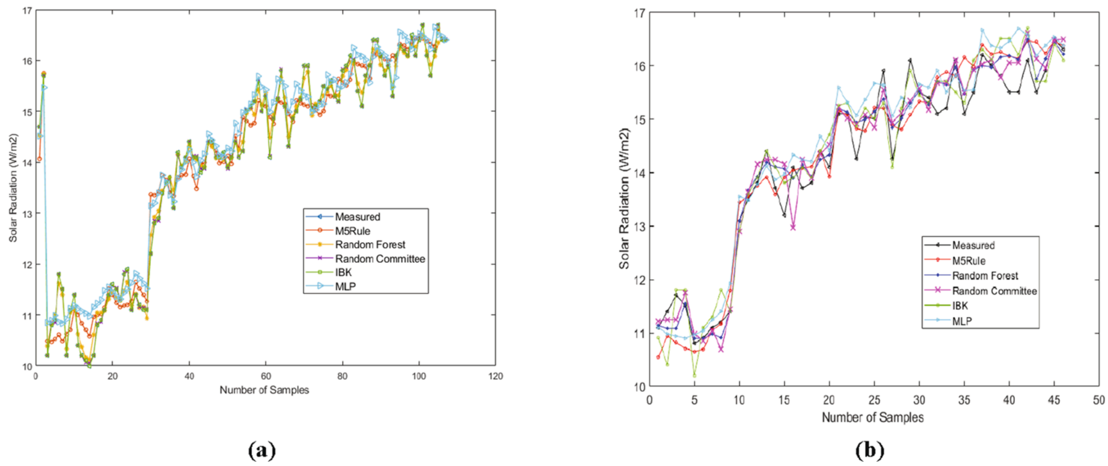

Fig. 11. The predicted solar radiation values of training (a) and testing (b) datasets based on measured M5Rule, Random Forest, Random Committee, IBK and MLP.

behavior. It is clear that all predictors are capable to predict solar radiation values. However, Random forest algorithm achieved the highest R^2 and the lowest MSE, MAE, RMSE, and MAPE.

The results of this research are in line with [34,35] that investigated of using Random Forest to predict solar radiation values. Both proposed models proved the capability and the robustness of using Random forest technique to predict solar radiation values. Random forest model showed high performance results and could improve the MSE by 17% and 18.98%, respectively.

In contrast to [43,44], which have investigated the ability of using different intelligent predictors to estimate the values of solar radiation (i.e. Linear regression, Regression Tree, M5Tree, Machine Learning, weighted k-Nearest Neighbors, Adaptive Regression Spline, SVMR, Bayesian Regularized Neural Networks, Random Forests, Cubist, Extreme Gradient Boosting, M5Rules, Attribute Selected Classifier, Multilayer Perceptron, Additive Regress, Gaussian Processors, Stacking, LeastMedSq, SMOreg, Decision Tree, Regression By Discretization, Bagging, CV Parameter, Multi Scheme, Holt Winter, ZeroR, Random Committee, Randomizable Filtered Classifier, Random Sub Space, Vote, Input Map Classifier, Decision Stump, Random Tree, IBK, KStar, LWL, and RepTree). It is found that Cubist and M5 models performed better compared to other models with relative very low error and high accuracy [38]. Besides, Randomizable Filtered Classifier algorithm achieved high performance, low error compared to major predictors, and an acceptable computational complexity [39].

In this study, the results revealed that all of the optimum techniques can be used to predict a solar radiation successfully. Therefore, based on the requirements of solar radiation prediction systems, designers of solar predictors can choose which model is most appropriate for their designed solar system.

Following the recommended predictors to estimate and predict solar radiation would help investment sector, researchers, and workers in the field to employ more appropriate intelligent models that can achieve accurate estimation for solar radiation values. Thus, many companies and homes' owners can target the highest solar radiation days and hours to store energies. Besides, it gives an alert before the end of sunny days.

Conclusion

In this paper, several prediction algorithms from five categories including Rules, Trees, Meta, Lazy and Functions Methods are used to predict hourly global solar radiation in Jordan. Two algorithms of Rules methods (i.e. Decision Table and M5rule), four algorithms from Trees methods (i.e. Decision Stump, M5P, Random Forest, and REPTree), four algorithms from Meta Methods (i.e. Additive Regression, Bagging, Random Committee, and Regression by Discretization), three algorithms from Lazy methods (i.e. LWL, Kstar, and IBK), and three algorithms from Function methods (i.e. MLP, SVMR, and LR) are examined and evaluated to check their capability to predict hourly global solar radiation. To find the optimum prediction model, the original dataset is divided into two datasets including training and testing datasets with splitting ratio equal to 70% and 30% of original dataset, respectively. After that, for each category the results of training and testing are compared together to find the best algorithm in each category. It is found that M5rule, Random Forest, Bagging, IBK, and MLP algorithms are the optimum algorithms for Rules, Trees, Meta, Lazy and Functions categories, respectively.

The optimum algorithms in the five categories are compared to each other to find the most appropriate algorithm that can efficiently obtain the highest prediction rate and lowest error function besides acceptable complexity. The results found that random forest is the most acceptable model in overall models. The results revealed that not all models that achieved high fitting rate can achieve high prediction rate. Therefore, with this research we would recommend that random forest algorithm is feasible to construct a reliable solar radiation model for short term information. Moreover, using extra history information should further increase the prediction accuracy of other algorithms.

However, the dataset that used in this research is for Jordan which is a four-season country. This would bound this research to be tested and evaluated in the country of Jordan or its neighbors which have the same environmental conditions. Testing and evaluating this research for other countries requires collecting specific datasets for these countries. A future trend is to evaluate the proposed models in other countries that have different seasonal conditions. This is to ensure about the capability of the proposed models for other countries or by using different practical data that collected in other environmental conditions. Besides, to use other intelligent predictors that wildly recommended in solar energy field (i.e. deep learning) and other hybrid techniques. Another future trend is to investigate of optimum models among Empirical, semi Empirical, physical and artificial intelligence models to predict and estimate global solar radiation values in different sites.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] Alrwashdeh S, Alsaraireh M, Saraireh A. Solar radiation map of Jordan governorates. Int J Eng Technol 2018;7(3):1664–7. https://doi.org/10.14419/ijet.v7i3.15557.

[2] Duffie JA, Beckman WA, Blair N. Solar engineering of thermal processes, photovoltaics and wind. John Wiley & Sons; 2020.
[3] Al-Soud S, Hrayshat SA. 50 MW concentrating solar power plant for Jordan. J Clean Prod 2009;17(6):625–35. https://doi.org/10.1016/j.jclepro.2008.11.002.
[4] Adrian W, Javier E, Sonia R, Samah E, Iana A. Renewable Capacity Statistics. International Renewable Energy Agency (ARENA). https://www.irena.org/-/media/Files/IRENA/Agency/Publication/2019/Jul/IRENA_Renewable_energy_statistics_2019.pdf; 2019.
[5] Petroleum B. BP statistical review of world energy. https://www.bp.com/content/dam/bp/business-sites/en/global/corporate/pdfs/energy-economics/statistical-review/bp-stats-review-2019-full-report.pdf.
[6] Al Tarabsheh A, Ababne M. Analysis of solar radiation in Jordan. Jordan J Mech Ind Eng 2010;4(6):733–7. https://eis.hu.edu.jo/deanshipfiles/pub106382737.pdf.
[7] Akash A, Mohsen S. Energy analysis of Jordan's rural residential sector. Energy Convers Manage 1999;40(12):1251–8. https://doi.org/10.1016/S0196-8904(99)00034-5.
[8] Jaber O, Elkarmi F, Alasis E, Kostas A. Employment of renewable energy in Jordan: current status, SWOT and problem analysis. Renew Sustain Energy Rev 2015;49:490–9. https://doi.org/10.1016/j.rser.2015.04.050.
[9] Alrwashdeh S, Alsaraireh M. Investigation of solar radiation distribution over three zones north, middle and south of Jordan. Int J Eng Technol 1999;7(4):5047–50. https://doi.org/10.14419/ijet.v7i4.27174.
[10] Alrwashdeh S. Assessment of photovoltaic energy production at different locations in Jordan. Int J Renew Energy Res 2018;8(2):798–804.
[11] Al Shafeey M, Harb M. Photovoltaic as a promising solution for peak demands and energy cost reduction in Jordan. In: 2018 International renewable energy congress. IEEE; 2018. p. 1–4.
[12] Nadia R, Al-Rousan M, Shareiah A, Hazem N. Choosing the efficient tracking method for real time tracking system in Jordan and it's neighbors to get maximum gained power based on experimental data. In: 2012 International conference on renewable energy research and applications. IEEE. p. 1–6.
[13] Alrwashdeh S. Comparison among solar panel arrays production with a different operating temperatures in Amman-Jordan. Int J Mech Eng Technol 2018;9(6):420–9. http://www.iaeme.com/IJMET/issues.asp?JType=IJMET&VType=9&IType=6.
[14] Al-Najideen I, Alrwashdeh S. Design of a solar photovoltaic system to cover the electricity demand for the faculty of Engineering-Mu'tah University in Jordan. Resour-Effic Technol 2017;3(4):440–5. https://doi.org/10.1016/j.reffit.2017.04.005.
[15] Dar-Mousa N, Makhamreh Z. Analysis of the pattern of energy consumptions and its impact on urban environmental sustainability in Jordan: Amman City as a case study. Energy, Sustain Soc J 2019;9(1):15. https://doi.org/10.1186/s13705-019-0197-0.
[16] Li H, Lam N. Determining the optimum tilt angle and orientation for solar energy collection based on measured solar radiance data. Int J Photoenergy 2007;8(1):1–6. https://doi.org/10.1155/2007/85402.
[17] Khoo S, Nobre A, Malhotra R, Yang D, Rüther R, Reindl T, et al. Optimal orientation and tilt angle for maximizing in-plane solar irradiation for PV applications in Singapore. IEEE J Photovolt 2013;4(2):647–53. https://doi.org/10.1109/JPHOTOV.2013.2292743.
[18] Lave M, Kleissl J. Optimum fixed orientations and benefits of tracking for capturing solar radiation in the continental United States. J Renew Energy 2011;36(3):1145–52. https://doi.org/10.1016/j.renene. 2010.07.032.
[19] Yadav K, Chandel S. Tilt angle optimization to maximize incident solar radiation: A review. Renew Sust Energy Rev 2013;23:503–13. https://doi.org/10.1016/j.rser.2013.02.027.
[20] Al-Sbou A, Alawasa M. Nonlinear autoregressive recurrent neural network model for solar radiation prediction. Int J Appl Eng Res 2017;12(14):4518–27. https://www.ripublication.com/ijaer17/ijaerv12n14_59.pdf.
[21] Ayodele R, Ogunjuyigbe O. Performance assessment of empirical models for prediction of daily and monthly average global solar radiation: the case study of Ibadan, Nigeria. Int J Ambient Energy 2017;38(8):803–13. https://doi.org/10.1080/01430750.2016.1222961.
[22] Quansah E, Amekudzi K, Preko K, Aryee J, Boakye R, Boli D, et al. Empirical models for estimating global solar radiation over the Ashanti region of Ghana. J Solar Energy 2014;2014:1–6. https://doi.org/10.1155/2014/897970.
[23] Pattarapanitchai S, Janjai S. A semi-empirical model for estimating diffuse solar irradiance under a clear sky condition for a tropical environment. Procedia Eng 2012;32:421–6. https://doi.org/10.1016/j.proeng.2012.01.1288.
[24] Korres N, Tzivanidis C. Development of two new semi-empirical formulas for estimation of solar absorptance in circular cavity receivers. Therm Sci Eng Progress 2019;10:147–53. https://doi.org/10.1016/j.tsep.2019.01.017.
[25] Cornejo-Bueno L, Casanova-Mateo C, Sanz-Justo J, Salcedo-Sanz S. Machine learning regressors for solar radiation estimation from satellite data. Sol Energy 2019;183:768–75. https://doi.org/10.1016/j.solener.2019.03.079.
[26] Nadia R, Isa M, Desa M. Advances in solar photovoltaic tracking systems: a review. Renew Sustain Energy Rev 2018;82:2548–69. https://doi.org/10.1016/j.rser.2017.09.077.
[27] Belu R. Artificial intelligence techniques for solar energy and photovoltaic applications. Handbook of research on solar energy systems and technologies. IGI Global; 2013.
[28] Mousavi S, Hizam H, Gomes C. Estimation of hourly, daily and monthly global solar radiation on inclined surfaces: models re-visited. Energies 2017;10(1):134–62. https://doi.org/10.3390/en10010134.
[29] Demirtas M, Yesilbudak M, Sagiroglu S, Colak I. Prediction of solar radiation using meteorological data. In: 2012 International conference on renewable energy research and applications (ICRERA). IEEE.
[30] Hamdan A, Abdelhafez E, Ghnaimat O. Prediction of Hourly Solar Radiation in Amman-Jordan by Using Artificial Neural Networks. Int J Therm Environ Eng 2017;14(2):103–8. https://doi.org/10.5383/ijtee.14.02.003.
[31] Badran A, Dwaykat F. Prediction of solar radiation for the major climates of Jordan: a regression model. J Ecol Eng 2018;19(2):24–38. https://doi.org/10.12911/22998993/81240.
[32] Mohammed B, Hamdan A, Abdelhafez A, Shaheen W. Hourly solar radiation prediction based on nonlinear autoregressive exogenous (Narx) neural network. Jordan J Mech Ind Eng 2013;7(1):11–8. http://jjmie.hu.edu.jo/files/vol7n1/2-JJMIE-130-13-1%20mod-21-2013.pdf.
[33] Alomari H, Adeeb J, Younis O. Solar photovoltaic power forecasting in jordan using artificial neural networks. Int J Electr Comput Eng 2018;8(1):497–504. https://doi.org/10.11591/ijece.v8i1.pp497-504.
[34] Sun H, Gui D, Yan B, Liu Y, Liao W, Zhu Y, et al. Assessing the potential of random forest method for estimating solar radiation using air pollution index. Energy Convers Manage 2016;119(1):121–9. https://doi.org/10.1016/j.enconman.2016.04.051.
[35] Ibrahim A, Khatib T. A novel hybrid model for hourly global solar radiation prediction using random forests technique and firefly algorithm. Energy Convers Manage 2017;138(1):413–25. https://doi.org/10.1016/j.enconman.2017.02.006.
[36] Lou S, Li H, Lam C, Cha W. Prediction of diffuse solar irradiance using machine learning and multivariable regression. Appl Energy 2016;181(1):367–74. https://doi.org/10.1016/j.apenergy.2016.08.093.
[37] Ghimire S, Deo C, Raj N, Mi J. Deep solar radiation forecasting with convolutional neural network and long short-term memory network algorithms. Appl Energy 2019;253(1):113541. https://doi.org/10.1016/j.apenergy.2019.113541.
[38] Gala Y, Fernández Á, Díaz J, Dorronsoro R. Hybrid machine learning forecasting of solar radiation values. Neurocomputing 2015;176(1):48–59. https://doi.org/10.1016/j.neucom.2015.02.078.
[39] Bui N, Moayedi H, Rashid A. Developing a predictive method based on optimized M5Rules–GA predicting heating load of an energy-efficient building system. Eng Comput 2020;36(1):931–40. https://doi.org/10.1007/s00366-019-00739-8.
[40] Queiroz J, Leitão P, Dias A. Predictive data analysis driven multi-agent system approach for electrical micro grids management. In: 2016 IEEE 25th international symposium on industrial electronics (ISIE), vol. 25, no. 1. 2016. p. 738–43. https://doi.org/ 10.1109/ISIE.2016.7744981.
[41] Corani G. Air quality prediction in Milan: feed-forward neural networks, pruned neural networks and lazy learning. Ecol Model 2005;185(2–4):513–29. https://doi.org/10.1016/j.ecolmodel.2005.01.008.
[42] Bellocchio F, Ferrari S, Lazzaroni M, Cristaldi L, Rossi M, Poli T, Paolini R. Illuminance prediction through SVM regression. In: 2011 Workshop on environmental energy and structural monitoring systems. IEEE. p. 1–5.
[43] Ferlito S, Adinolfi G, Graditi G. Comparative analysis of data-driven methods online and offline trained to the forecasting of grid-connected photovoltaic plant production. Appl Energy 2017;205(1):116–29. https://doi.org/10.1016/j.apenergy.2017.07.124.
[44] Kumar A, Ali S. Prospects of wind energy production in the western Fiji—An empirical study using machine learning forecasting algorithms. In: 2017 Australasian universities power engineering conference. IEEE. p. 1–5.
[45] Jha K, Bilalovic J, Jha A, Patel N, Zhang H. Renewable energy: Present research and future scope of Artificial Intelligence. Renew Sustain Energy Rev 2017;77(1):297–317. https://doi.org/10.1016/j.rser.2017.04.018.
[46] El-Bendary N, Elhariri E, Hazman M, Saleh SM, Hassanien AE. Cultivation-time recommender system based on climatic conditions for newly reclaimed lands in Egypt. Proc Comput Sci 2016;96(1):110–9. https://doi.org/10.1016/j.procs.2016.08.109.
[47] Bayzid SM, Mohamed Y, Al-Hussein M. Prediction of maintenance cost for road construction equipment: a case study. Can J Civ Eng 2016;43(1):480–92. https://doi.org/10.1139/cjce-2014-0500.
[48] Sharma R, Kumar S, Maheshwari R. Comparative analysis of classification techniques in data mining using different datasets. Int J Comput Sci Mob Comput 2015;4(12):125–34. https://www.ijcsmc.com/docs/papers/December2015/V4I12201528.pdf.
[49] Witten H, Frank E. Data mining: practical machine learning tools and techniques with Java implementations. Acm Sigmod Record 2002;31(1):76–7. https://doi.org/10.1145/507338.507355.
[50] Wang L, Kisi O, Zounemat-Kermani M, Salazar A, Zhu Z, Gong W. Solar radiation prediction using different techniques: model evaluation and comparison. Renew Sustain Energy Rev 2016;61:384–97. https://doi.org/10.1016/j.rser.2016.04.024.
[51] Qin W, Wang L, Lin A, Zhang M, Xia X, Hu B, et al. Comparison of deterministic and data-driven models for solar radiation estimation in China. Renew Sustain Energy Rev 2018;81:579–94. https://doi.org/10.1016/j.rser.2017.08.037.
[52] Wang L, Lu Y, Zou L, Feng L, Wei J, Qin W, et al. Prediction of diffuse solar radiation based on multiple variables in China. Renew Sustain Energy Rev 2018;103(1):151–216. https://doi.org/10.1016/j.rser.2018.12.029.
[53] Qin W, Wang L, Zhang M, Niu Z, Luo M, Lin A, et al. First effort at constructing a high-density photosynthetically active radiation dataset during 1961–2014 in China. J Clim 2019;32(10):2761–80. https://doi.org/10.1175/JCLI-D-18-0590.1.
[54] Zou L, Wang L, Li J, Lu Y, Gong W, Niu Y. Global surface solar radiation and photovoltaic power from Coupled Model Intercomparison Project Phase 5 climate models. J Clean Prod 2019;224(1):304–24. https://doi.org/10.1016/j.jclepro.2019.03.268.

[55] Feng L, Lin A, Wang L, Qin W, Gong W. Evaluation of sunshine-based models for predicting diffuse solar radiation in China. Renew Sustain Energy Rev 2018;94(1): 168–82. https://doi.org/10.1016/j.rser.2018.06.009.
[56] Wang L, Kisi O, Zounemat-Kermani M, Zhu Z, Gong W, Niu Z, et al. Prediction of solar radiation in China using different adaptive neuro-fuzzy methods and M5 model tree. Int J Climatol 2017;37(3):1141–55. https://doi.org/10.1002/joc.4762.
[57] Witlox F, Antrop M, Bogaert P, De Maeyer P, Derudder B, Neutens T, et al. Introducing functional classification theory to land use planning by means of decision tables. Decis Support Syst 2009;46(4):875–81. https://doi.org/10.1016/j.dss.2008.12.001.
[58] Banu R. A Role of decision tree classification data mining technique in diagnosing thyroid disease. J Comput Inf Sci Eng 2016;4(11):64–70.
[59] Wang Y, Witten H. Induction of model trees for predicting continuous classes; 1996.
[60] Quinlan R. Learning with continuous classes. In: 5th Australian joint conference on artificial intelligence, vol. 92, no. 1; 1992. p. 343–8. https://doi.org/10.1.1.34.885.
[61] Breiman L. Manual on setting up, using, and understanding random forests. 3d version. USA: Statistics Department University of California Berkeley; 2002.
[62] Zhao Y, Zhang Y. Comparison of decision tree methods for finding active objects. Adv Space Res 2008;41(12):1955–9. https://doi.org/10.1016/j.asr.2007.07.020.
[63] Breiman L. Arcing the edge. Technical Report 486, Statistics Department, University of California at Berkeley; 1997. https://statistics.berkeley.edu/sites/default/files/tech-reports/486.pdf.
[64] Friedman H. Stochastic gradient boosting. Comput Stat Data Anal 2002;38(4): 367–78. https://doi.org/10.1016/S0167-9473(01)00065-2.
[65] Breiman L. Bagging predictors. Mach Learn 1996;24(2):123–40. https://www.stat.berkeley.edu/~breiman/bagging.pdf.
[66] Niranjan A, Nutan H, Nitish A, Shenoy D, Venugopal R. ERCR TV: ensemble of random committee and random tree for efficient anomaly classification using voting. In: 2018 3rd international conference for convergence in technology. IEEE. p. 1–5.
[67] Frank E, Bouckaert R. Conditional density estimation with class probability estimators. In: 2009 Asian conference on machine learning. Springer. p. 65–81.
[68] Atkeson G, Moore W, Schaal S. Locally weighted learning. Lazy Learn 1997.
[69] Aha W, Kibler D. Albert KInstance-based Learning Algorithms. Mach Learn 1991;6 (1):37–66. https://doi.org/10.1007/BF00153759.
[70] Keleş K. Breast cancer prediction and detection using data mining classification algorithms: a comparative study. Tehnički vjesnik 2019;26(1):149–55. https://doi.org/10.17559/TV-20180417102943.
[71] Wu H, Tzeng H, Lin H. A Novel hybrid genetic algorithm for kernel function and parameter optimization in support vector regression. Expert Syst Appl 2009;36(3): 4725–35. https://doi.org/10.1016/j.eswa.2008.06.046.
[72] Üstün B, Melssen J, Buydens M. Facilitating the application of support vector regression by using a universal Pearson VII function based kernel. Chemom Intell Lab Syst 2006;81(1):29–40. https://doi.org/10.1016/j.chemolab.2005.09.003.
[73] Gunn R. Support vector machines for classification and regression. ISIS Techn Rep 1998;14(1):5–16. https://doi.org/10.1039/B918972F.
[74] Landwehr N, Hall M, Frank E. Logistic model trees. Mach Learn 2005;59(1): 161–205. https://doi.org/10.1007/s10994-005-0466-3.
[75] Zhang j, Gupta C. Neural networks for RF and microwave design. Artech House; 2000.
[76] Nadia, A. R., Al-Rousan, M., & Shareiah, A. A fuzzy logic model of a tracking system for solar panels in northern Jordan based on experimental data. In: 2012 International conference on renewable energy research and applications (ICRERA). IEEE. p. 1–6.
[77] AL-Rousan A, Isa M, Desa M. Efficient single and dual axis solar tracking system controllers based on adaptive neural fuzzy inference system. J King Saud Univ-Eng Sci 2020;32(7):459–69. https://doi.org/10.1016/j.jksues.2020.04.004.
[78] Al-Rousan A, Isa M, Desa M. Correlation analysis and MLP/CMLP for optimum variables to predict orientation and tilt angles in intelligent solar tracking systems. Int J Energy Res 2020. [in press]. 10.1002/er.5676.