

## Regular Article

Lina Alhmoud\*, Ala' M. Al-Zoubi and Ibrahim Aljarah

# Solar PV power forecasting at Yarmouk University using machine learning techniques

<https://doi.org/10.1515/eng-2022-0386>

received October 01, 2022; accepted November 20, 2022

**Abstract:** Renewable energy sources are considered ubiquitous and drive the energy revolution. Energy producers suffer from inconsistent electricity generation. They often struggled with the unpredictability of the weather. Thus, making it challenging to balance supply and demand. Technologies like artificial intelligence (AI) and machine learning are effective ways to forecast, distribute, and manage renewable photovoltaic (PV) solar supplies. AI will make the energy forecasting system more connected, intelligent, reliable, and sustainable. AI can innovate how energy is used and help find solutions for decarbonizing energy systems. There are potential advantages to total energy forecasting. AI can support the growth and integration of PV solar energy. The article's main objective is to use AI to forecast the output consumed power of the Yarmouk University PV solar system in Jordan. The total actual yield is 5548.96 MW h, and the performance ratio (PR) is 95.73%. Many techniques are used to predict the consumed solar power. The random forest model obtains the best results of root mean squared error and mean absolute error are 172.07 and 68.7, respectively. This accurate prediction allows for the maximum use of solar power and the minimal use of grid power. This work guides the operators to learn trends embedded in Yarmouk University's historical data. These understood trends can be used to predict the consumption of solar power output. Thus, the control system and grid operators have advanced knowledge of the expected consumption of solar power at each hour of the day.

**Keywords:** load forecasting, neural network, solar PV system

\* **Corresponding author: Lina Alhmoud**, Department of Electrical Power Engineering, Faculty of Engineering Technology, Yarmouk University, Irbid, Jordan 21163, e-mail: lina.hmoud@yu.edu.jo  
**Ala' M. Al-Zoubi**: School of Science, Technology and Engineering Department, University of Granada, Granada, Spain, e-mail: ala.m.zoubi@gmail.com, alzoubi@correo.ugr.es  
**Ibrahim Aljarah**: King Abdullah II School for Information Technology, The University of Jordan, Amman, Jordan, e-mail: i.aljarah@ju.edu.jo

## 1 Introduction

Nowadays, a significant change in technology within the power system has changed the stable state of the power sector. The power sector is currently taking on a horizontally integrated system where generation from other alternative sources and the connection of renewable generation to the network are changing magnificently. Thus, the power sector is experiencing a paradigm shift. Electric utilities seek a stable, fair predictable, and commonly operated vertical supply chain. The power system's objective is to supply electric power to customers reliably and economically. Therefore, electric power consumption and production must be balanced continuously and instantaneously. Power system loads can be described as variability and uncertainty. The load varies throughout the day, whereas the conventional generation can often deviate from schedules. Thus, the contingencies are unexpected, and the load forecast error is also unexpected [1,2].

Good forecasting techniques are essential for increasing grid stability and reliability. Load forecasting benefits can be summarized as follows: estimating how much can be consumed by the end of the entire day, week, month, quarter, or even a year, comparing the forecasted model to establish goals, targets, and budgets, and switching the consumption to cost as preparing the budget for the next year as a financial controller. Utility companies need profound load forecasting studies using learning algorithms. Therefore, this work tries to forecast the electricity demand to provide an overview of the national electricity demand at Yarmouk University in Jordan, which serves as input to the generation development plan. The main scope is developing a suitable forecasting model of load forecasting power output with minimal error using MATLAB and Python [3].

Some renewable energy systems forecasting challenges still need to be addressed on time. So, the role of AI can be summarized by building an intelligent control center. Thus, data are generated when renewable energy forecasting is coupled with AI. This data explore a new insight into the grid operators for better control operations and forecasting. It allows energy suppliers to adjust

supply and demand intelligently. When dealing with integrated mini-grids and adding renewable energy to the main grid, this leads to difficulty in balancing the flow of energy. AI can help grid operators in this area. The AI-powered control system can also help solve quality and congestion issues. AI can provide impressive safety and reliability. Hence, AI can help the operators understand the energy consumption patterns and identify the grid's energy leakage and health, and the operator can promptly take precautions for the data. Besides, integrating AI can help renewable energy suppliers expand the market by introducing new service models and encouraging higher participation. AI can be integrated with intelligent energy storage units to provide a sustainable and reliable solution; this can help with intelligent power storage and distribution.

Accurate simulations for electric power load forecasting are essential to the operation and planning of a utility power company. It helps the company make crucial decisions, including purchasing and generating electric power, load switching, and infrastructure development. Various load forecasting methods are reported on the importance, including linear regression, exponential smoothing, stochastic processes, auto-regressive moving average models, and data mining models. Of late, artificial neural network (ANN) has widely employed load forecasting techniques. AI is considered the state-of-the-art technology for processing a large amount of data and deriving accurate insight. Therefore, it becomes a fundamental enabler in the utility to harvest data from the energy system to quickly get valuable results. AI is one of the tools in the box that can be deployed in any part of the energy system. Thus, it can fit into the process and operational chain, such as managing assets by forecasting the future's needs. For example, predicting the electricity demand for the grid and customers allows the grid operator to advance purchasing and avoid an imbalance in the transmission losses. These algorithms should include patterns and trends, prediction analysis, actionable insights, and automatic actions based on predictions. In addition to identifying patterns and trends within data analyses and forecasts, sophisticated regression analysis is done in AI. The process is started by data collection on weather, geography, demographics, demand, supply, and price. The second step is processing these data, such as through data cleaning and wrangling, exploratory data analysis, AI algorithms, data visualization, and insight. Optimization of the constraints is a critical key for data-driven, and decision-making [4].

The rest of this article is arranged as follows. Section 2 summarizes the related work. Section 3 introduces

electrical load forecasting and AI. Section 4 introduces the system under study. Then, the data collection process is presented in Section 5. The proposed methodology is addressed in Section 6. Experiments and results are demonstrated in Section 7. The article finishes with the conclusion in Section 8.

## 2 Related work

The literature review states many forecasting methods. In quantitative (analytical aspects) [5,6], the forecasted values are developed using certain factors, such as mathematical models and historical data. Quantitative methods are objective in nature. They can be designed using mathematics, and they can be used when past data are available. Unlike the qualitative (social aspects) methods, the forecast values are developed depending on certain factors such as decision-making, instinct knowledge, experience, and emotion [5,6]. Qualitative methods are subjected to nature and can be used in forecasting intermediate and long-term decisions. Examples of qualitative methods are Jory of execution opinion, where the decision-making is based on a high level and expert management [7] and Delphi method, where the decision making is based on experts and staff [8]. Sales force composite decision-making is made by management based on estimates given by an individual salesperson. A consumer market method is a survey in which a survey is conducted among the target or prospective customers [9].

In associative or causal models, the forecasted variables are associated with other variables. The associative models are linear regressions where forecast-dependent variables are based on independent variables [10]. Other models are the time series model. Time series examine past data and forecast the future based on underlying patterns. Samples moving average is considered the most straightforward technique because of identifying the upstream and downstream decision-making [11]. It is estimated by dividing the sum of the last  $n$  period values and  $n$ . In addition, the exponential smoothing methods assign a decreasing weight for older values. Thus, removing any effect data might have on forecasts. In addition, they help to smooth the data by eliminating any noise or effects that the past data may have on the forecast [12].

Literature states a numerical weather forecast. It is a long-term horizon forecast that predicts weather conditions the using present conditions as inputs to the mathematical models. But this model loses accuracy at high spatial and temporal resolutions. Therefore, the

distinctive approaches on the long-time horizon include a time horizon approach, and linear models such as autoregressive integrated moving average and are used [13]. Some scholars classify the load forecasting studies into physical and statistical methods. The physical dynamical method transforms input weather data such as temperature, pressure, and surface roughness into numerical weather prediction (NWP) to create terrain-specific weather conditions. The statistical techniques use historical and real-time generation data to correct data obtained from NWP models. The NWP models are based on data assimilation, parameterization, resolution, operational information, and power conversion algorithms. In addition, persistence forecasting uses the last observation as the next forecast, and ensemble forecasting aggregates results from multiple different forecasts. In addition, load forecasting can be categorized based on point estimates, whereas deterministic forecasts can provide a false sense of certainty and probabilistic estimation. It can be beneficial if the system has a good way of using the additional information.

The forecasting techniques model can be separated into several categories based on the desired predicted output. It can be organized into two main categories: statistical and ANN. Statistical analysis is based on a sequence of observations of one or more parameters measured at a particular time. ANN uses multilayer perception, Markov chain, Fourier series, and regression methods. Furthermore, ANN can be identified as physical and hybrid. Physical models consist of mathematical equations that describe the variables' physical state, and dynamic motion. For example, it used NWP. Hybrid models are a mixture of different models with unique features to address the limitations of individual techniques.

### 3 Electrical load forecasting and AI

Electrical energy has to be generated to meet the demand. Therefore, electrical power utilities must estimate the load on their systems in advance. Estimating energy requirements and demand is crucial to effective system planning. Load predictions determine the generation capacity, transmission capacity, and distribution system. The decision-maker should use the available power plants to serve the predicted load for each hour of the coming week. This estimation is known as load forecasting. It is essential for power system planning. Thus, it starts with a forecast of anticipated future load requirements. Load forecasting is also used to establish procurement policies for construction capital energy forecasts, which are needed to

determine future fuel requirements. Thus, a good forecast reflecting the present and future trends is considered a sky to all planning.

Load forecasting (LF) is vitally essential for the electric industry. Thus, effective LF can help improve and adequately plan the power system's generation, transmission, and distribution. Therefore, accurate electric power load forecasting models are essential to the operation and planning of utility companies. LF helps an electric utility make crucial decisions, including purchasing and generating electric power, load switching, and infrastructure development. Utilities, consumers, variable generators, plant owners, and independent system operators are the ones who accrued the benefits of improved forecasting and bore the financial and reliability risks of poor forecasting. Therefore, understand how a good forecast can impact market operations and scheduling, generation, and transmission and distribution system operations. Monitoring and verification are essential components of load forecasting. Monitoring the forecasting quality measures the accuracy and how the forecast can be improved over time. It is the first step toward discovering the wrong issues early. It is also used to compare different models and determine how to get the system better financially verified.

The power system should tailor the forecasting algorithms to each customer's unique needs and context. Better load forecast decisions are obtained based on valuable information. Besides, understanding areas where forecasting improves decision-making is the first step in implementing forecasting systems. Therefore, forecast interpreting is considered a critical element of the effective implementation of decision-making. Some of the disrupters on demand forecast include but are not limited to technology and innovation, changes in customer behavior, battery storage, regulatory changes, economic status, and tariff structure. The planning approach for uncertainties in the electrical load forecasting and fast-evolving future is based on understanding the power industry's fast evolution and acknowledging electrical load vibration. In addition, consider future alternatives in the power grid and growing trends, besides understanding local and global economic events. The timely implementation of such decisions leads to the importance of network reliability and reduced equipment failures and blackouts. LF is also financially important for the energy pricing offered by the market.

Load forecasting can be classified into centralized and decentralized forecasts. The system operator does centralized forecasting, enabling unit commitment and economic dispatch forecasting. Over that, it requires a mechanism to obtain data from the generator and

encourage data quality, and it allows generators consistency and reduces uncertainty at the system level. In comparison, decentralized forecasting is handled by the generator. Off-takers use it when making offers and help project operators optimize operation and maintenance, besides informing operators of the potential for transformation congestion. The centralized forecasting by the system operator, supported by the generator level forecast from the planet operator, is widely considered the best practice approach. In general, centralized and decentralized have a unique value. Moving toward centralized forecasting effectively reduces uncertainty at the system level. In addition, it helps identify the various environmental changes that occur. Thereby allowing the company to develop a mechanism by which they cope with such environmental changes. Also, forecasting involves analyzing past data, making it easy for the company to identify its weak spots and take the necessary corrective action. On the other hand, any forecasting model is used only to estimate a future event and does not assume 100% accuracy.

Load forecasting can be classified based on time intervals and horizons. Short-term load forecasting (STLF) includes intra-day adjustments, an-hour-ahead scheduling, and trading. Medium-term load forecasting (MTLF) provides estimates for a month up to a year. Long-term load forecasting (LTLF) forecasts for one year and more. STLF is the main focus of this work. It estimates the typical generation used for resources and operational and maintenance planning. Thus, it is necessary to schedule operations, control the power system, and act as inputs to power analysis functions such as load flow and contingency analysis.

AI future is a vast area that enables a machine or computer to replicate human capabilities such as object recognition, decision-making, and problem-solving. AI can help by providing tools capable of evaluating and exploring a complex and dynamic design space. The main challenge for AI is the fidelity-scalability trade-off, with methods tailored around narrowly defined applications. The primary needs are methods for exposing opportunities at the intersection of design and geographical space. Machine learning is building a more renewable electric grid and how it's starting to impact the power system and pave the way toward low emissions. Machine learning is used to boost renewable energy generation and significantly reduce the cost of photovoltaic (PV) solar systems. The latest progress in deploying AI technology is fishing solar radiation to best capture solar energy and increase energy. AI has several branches, including machine learning. Machine learning allows systems to naturally learn and improve by using their experiences to predict outcomes without being explicitly programmed. It focuses on developing computer programs that can access and leverage data

for learning. Machine learning, like a human brain, needs input to gain and learn from; however, it uses training data and knowledge graphs to comprehend entity domains and their relationships. The algorithm uses examples of direct experience or instructions as inputs to generate an estimate of a pattern in the data. Then an error function is used to evaluate the models' accuracy using those examples for comparison. Next, the model is optimized by altering weights to reduce the gap between the known sample and the model prediction. Finally, the program repeats the optimization process until the target accuracy is achieved. Machine learning utilizes three main techniques: supervised, unsupervised, and semi-supervised. Supervised learning uses training sets, which are labeled data sets. Machine learning is not based on knowledge but on data, so a lack of training or unclear and noisy data can lead to inaccurate predictions for early adopters of such technology. There may be insufficient data to make proper decisions, drastically reducing the system's effectiveness because the old system worked.

In addition, the machine-learning process is complex by itself, analyzing the data, removing data bias, and training datasets. Running machine learning models is also, a slow process that takes much time and demands colossal computation. On the other hand, it increases the reliability of the electrical grid and impressive with big players. Machine learning can be used to overcome the problem of intermittency of PV solar power plants by forecasting sunlight and airflow for solar power much better than humans using past weather data to provide accurate forecasting for renewable energy. Supply companies can plan when to produce or augment energy generation. As a result, renewable energy sources could become more reliable, affordable, and efficient. Machine learning can also be used to improve renewable energy storage and identify the optimal layout and geographical location of solar power plants from a maintenance perspective. Machine learning can be employed to collect data from sensors installed on the electric grid to detect anomalies and predict failures. Automating monitoring can monitor PV solar system health and help schedule maintenance. Technology that uses machine learning to sort the data acquired from an extensive database of what reports the idea was to reduce the variability of solar energy output, also reducing the need for excess energy storage systems by increasing the accuracy of weather forecasting as it relates to energy output.

Machine learning leads the renewable energy sector by pushing the development of faster and more powerful computing technologies. The widespread use of this technology around the globe in the coming years will contribute to a more sustainable future, and the traditional revenue companies have been losing ground as

renewables popping up everywhere. There has been more energy supply and stability variance. Machine learning could modify electricity prices depending on the vast data generated by the expanding number of smart meters and sensors used throughout the PV solar power plant. Besides, they must forecast supply and demand to balance the grid in real time and reduce downtime. The start-up neural network has been taking advantage of the edge AI platform to build smart electric meters that can take many measurements per second, allowing for fine-tuned monitoring by homeowners to automate and control how energy flows from the solar panels into the home to reduce energy bills. The internet brand of smart meters can be connected to the grid to help smart appliances monitor disruptive weather patterns that may affect energy grids, optimize energy usage, and detect energy waste or even failures. In addition, breakers are brought to the utility scale to allow microgrids to optimize numerous energy sources and storage facilities. It rapidly detects grid faults and implements dynamic pricing and demand response with real-time AI data. All this can help increase reliability, security, and sustainability.

Over that, there is a sense energy monitor installed in the electric panel that could detect improper voltage from the grid so the operator can proactively reach out to the utility to address any issues and warn the owner in case there is an appliance that draws too much current, so it serviced before failure. Thus, it alarms the electricity use on specific circuits, which has saved money. There are many ongoing machine learning and AI projects using link debate, everything from a deep learning model for forecasting foldable take energy with uncertainties to deep reinforcement learning for optimal energy management of multi-energy smartphones. It is a very active and growing industry. So, there are still several challenges that need to be overcome.

## 4 System under study

Jordan has a vast solar energy potential because it is located within the world's solar belt, with average solar radiation varying between 5 and 7 kW h/m<sup>2</sup> as shown in Figure 1, implies a potential of at least 1,000 GW h per year. This work forecasts the total electrical load in one of the significant forefront of Middle Eastern and Jordanian universities. Yarmouk University is a public and comprehensive university. It is located in northern Jordan and was established in 1976. Table 1 shows the plant under study information. Table 2 shows a description of data

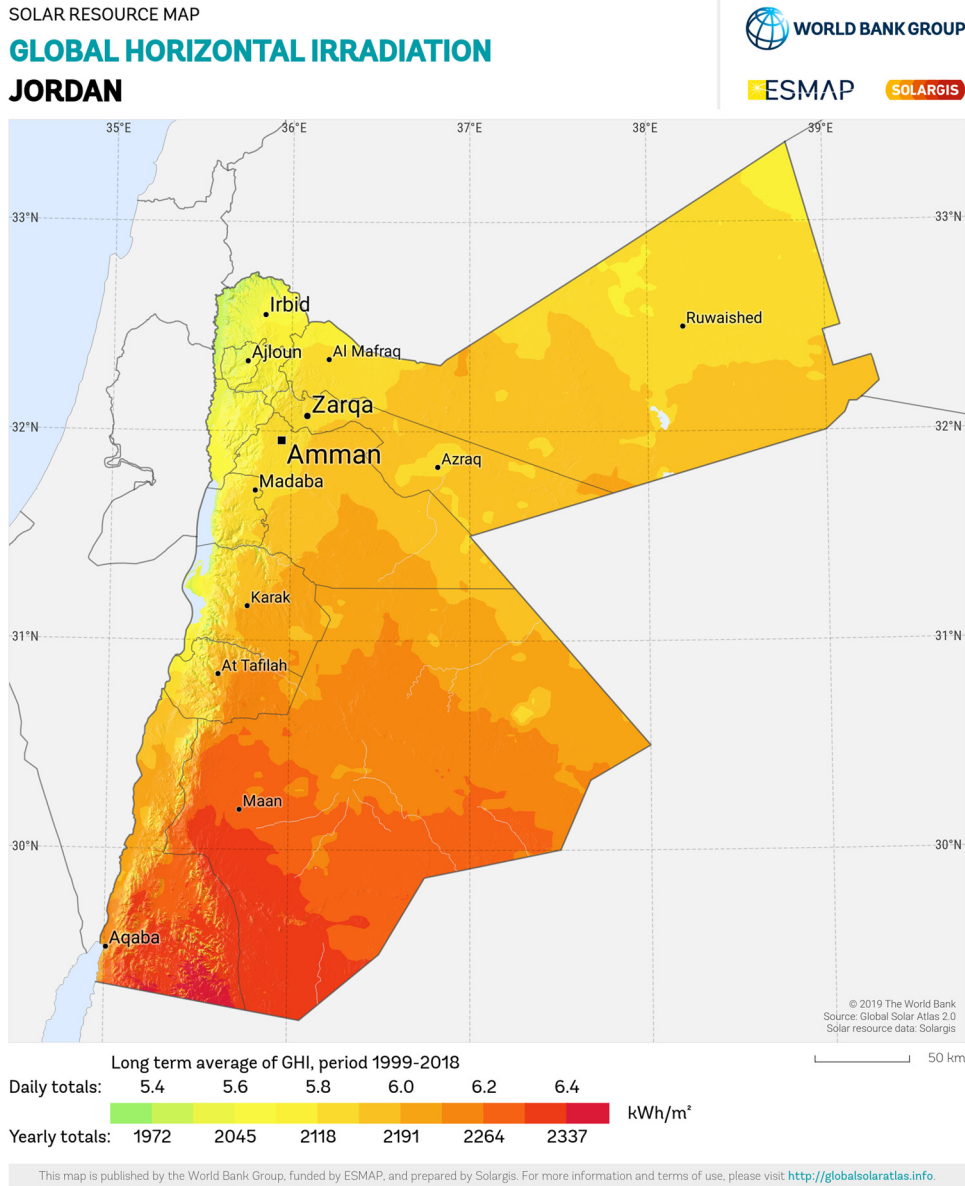
collection on the Yarmouk University solar PV plant. Figure 2 shows solar paths at the Yarmouk University location. Table 3<sup>1</sup> shows the energy estimation for the Yarmouk University solar power plant for the three contracted years, respectively. Table 4 shows Yarmouk University subdivided into 29 zones of generation. This table shows the actual yield (MW h), the contractual yield (MW h), and the period performance (PP), respectively. PP can be defined as the percentage ratio of the actual yield to the expected weather-corrected yield as equation (1). Table 5 shows the calculation for the year 2019 per month. It shows the PP, solar insolation, and PR calculations, respectively. PV solar insolation energy measures the solar system radiation collected on a particular array's surface in a reported time period (kW h/m<sup>2</sup>/month). PR which is the percentage of the relationship between actual and theoretical energy yield of a solar PV. The main distribution panel is the final distribution board that distributes the power supply to the final electrical circuit. It consists of breakers, isolators, busbars, and enclosures. Irradiance tilted is the amount of PV solar incident energy when the plan of irradiance is equal to the tilted module surface, and the nominal power is the maximum capacity of the PV solar system at standard test conditions.

$$PP (\%) = \frac{\text{Actual yield (kW h)}}{\text{Expected weather correction yield}} \cdot (1)$$

## 5 Data collection

Smart meters cause a rich and growing in energy data. Tremendous data are available more than the data in the traditional meters era. These new data resources are analyzed and transformed using machine learning techniques, data mining visualization, and statistical analysis. Transformation into these techniques increases reliability, decreases system cost, and improves environmental sustainability. The characteristics of these big data can be classified into four categories: volume (quantity of existing data), velocity (streaming data), variety (data forms such as numbers, text, and multimedia), and veracity (uncertainty of data quantity). System operators' data collection strategies can be accomplished in several ways, such as policy mandates implementation for utility-scale and distributed generation, interconnection or market requirements set by the government, utility power purchase agreements,

<sup>1</sup> Kwar Energy, Operational Report, Yarmouk University PVPP. Third operational year. KE-OM-RPT-OP-YU-012.



**Figure 1:** Global horizontal irradiation in Jordan.

penalties, rewards, and partnerships with metrological agencies and third-party vendors. Hence, it is required to update the interconnection standards and power purchase agreement to enable data gathering and work with national metrological institutions to improve underlying weather data or access. In addition, facilities must train the operators on metrology, interpret forecasts, and work with and support vendors. Thus, develop a smooth IT interface between forecast vendors and users. The meteorological data and the density and frequency of observations highly impact the forecasted quality. Static and dynamic data are needed to set up a load forecasting model. Static data such as plan location (latitude and longitude), installed capacity, and historical

data are used as training data. While dynamic data, such as real-time generators' availability data wind and solar) and park potential of total output power based on available resources at the farm level meteorological data, such as wind speed, irradiance, temperature, and pressure. There is no one-size-fits-all approach to collecting data, procuring, and monitoring forecasts. In addition, electricity consumption is affected by changes in population, electricity prices, extreme weather conditions, holidays, and special events such as popular football matches. Accurate load forecasting reduces the uncertainty involved in production and helps a company in planning and decision-making. The scenario for selecting the optimal set of input neurons is as follows:

**Table 1:** Yarmouk University PV solar power plant information

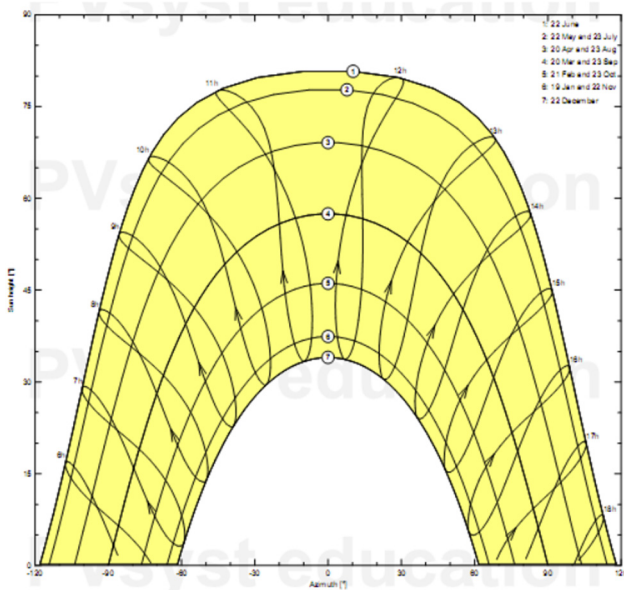
Site	Yarmouk University
Total period	June 1, 2018–Jan 31, 2021
Data source	Kawar Energy Company
Data organization	Daily data
Data resolution	15 min
Data composition	Initialization data
Field latitude	32.33
Field longitude	35.51
System capacity	3,003 kW <sub>p</sub>
Commercial operation date	Varies depending on the zone
Contractual yield (MW h)	5561.29
Actual yield (MW h)	5312.09
Expected weather correction yield (MW h)	5186.94
KPI (%)	95.52
PP (%)	102.41

**Table 3:** Monthly estimated energy (MW h) for Yarmouk University PV solar power plant for three contractual years [14]

Month	Contractual year 1	Contractual year 2	Contractual year 3
Dec	315.58	307.65	305.44
Jan	343.30	334.67	332.27
Feb	345.06	336.39	333.97
Mar	477.04	465.05	461.71
Apr	504.59	491.91	488.38
May	586.38	571.64	567.54
Jun	601.16	586.05	581.85
Jul	608.84	593.54	589.28
Aug	582.73	568.09	564.01
Sep	518.77	505.74	502.11
Oct	470.54	458.72	455.42
Nov	391.90	382.05	379.30
Yearly yield	5745.89	5601.5	5561.29

**Table 2:** Parameters of statistics for Yarmouk University PV solar power plant [14]

Number of PV plants	Period	Mean	Median	Max	Min	Standard deviation
29	1 June 2018–31 January 2021	519.67	17.01	2652.18	0	779.44

**Figure 2:** Solar paths at Yarmouk University location.

- (1) Day of the week (1/Sunday – 7/Saturday),  $L_1(k)$ ;
- (2) Month number in a year (1/January, 2/February, …, 12/December),  $L_2(k)$ .
- (3) Day type such as workday, weekend, and holiday (–1 for work day, –2 for holiday, and 0 for Friday and Saturday),  $L_3(k)$ ;

- (4) Week number in a year (1, 2, …, 52),  $L_4(k)$ .
- (5) Hour of the day (00:00/12:00 AM–23:00/11:00 PM),  $L_5(k)$ ;
- (6) Year type such as 1 for 2018, 2 for 2019, 3 for 2020, and 4 for 2021,  $L_6(k)$ ;
- (7) Irradiation ( $\text{W}/\text{m}^2$ ),  $L_7(k)$ ;
- (8) Ambient temperature ( $^{\circ}\text{C}$ ),  $L_8(k)$ ;
- (9) Module temperature ( $^{\circ}\text{C}$ ),  $L_8(k)$ ;

The output is the consumed load in kW h in a zones generation during the operational year, reflected in a zone generation, where  $L(t, d)$  is the load at time  $t$  and  $d$  is the day of the predicted day.

## 6 Methodology

The base of forecasting methodology is the components used in the model, applications of the forecast, and challenges experienced. The appeal of fossil fuels is that they can deliver power at all times of the day and can be turned on and off. Solar energy faces an intermittency issue that is not continuously available for power conversion. It fluctuates on cloudy days. Thus, the sun hours contributed significantly to prediction accuracy. Forecast

**Table 4:** Actual yield, contractual yield, and PR for Yarmouk University PV solar power plant during the third operational year (2019), which are reflected in all zones generation [14]

No.	Zone	Yield (MW h)	Contractual yield (MW h)	PP (%)
1	Administration and registration (AA15)	268.9	285.5	94
2	Sport collage (AA3)	53.9	56.4	96
3	GYM (sport2) (AA13)	110.5	91.6	121
4	Khwarimi (AA5)	191.9	179.0	107
5	Earth science (AA10)	122.5	127.7	96
6	Canteen (AA7)	54.9	57.0	96
7	Education (AB1)	292.2	305.7	96
8	Languages (CC4)	54.7	51.6	106
9	New Hijjawi (CC3)	165.0	164.4	100
10	Hijjawi Amara (CC3)	89.6	91.4	98
11	New law (CD7)	166.9	149.4	112
12	Pharma 1 (BD1)	139.7	143.0	98
13	Deanship (AD1)	205.1	226.6	91
14	Pharma 2 (BD1)	276.8	291.5	95
15	Halls (BC1)	465.2	487.3	95
16	Archeology (CD4)	163.8	173.2	95
17	Islamic college (CB6)	203.7	217.1	94
18	Hijjawi collage (CC3)	221.8	227.6	97
19	Literature (BA2)	163.3	166.9	98
20	Art (BA1)	276.1	283.3	97
21	Presidency (AA1)	111.8	110.1	102
22	Conferences (AA2)	157.8	159.2	99
23	Science (AD6)	360.0	385.1	93
24	President home (HC1)	21.4	25.1	85
25	New economic (FB1)	263.2	253.4	104
26	Male school (DC1)	43.9	38.1	115
27	Female school (DD1)	210.9	226.8	93
28	Medicine (BB1)	393.2	399.3	98
29	Maintenance (CB2)	55.1	56.8	97

errors related to power systems are not gaussian, and the prediction intervals are not symmetric. Therefore, that implies that tools built directly on this assumption are far from optimal.

Furthermore, large-scale integration for renewable forecasting on temporal and spatial resolutions is typically needed. A good metrological forecast is fundamental to obtaining a solar power forecast. Sometimes, it can be challenging to obtain a good forecast or, for instance, a split between direct and directional-dependent diffuse short-wave solar radiation. The tool provider for renewable forecasting collaborates well with various metrological and mid-forecast providers. Today's med forecasts are typically integrated parts of a good forecast. The use of technologies like AI is found to be the best

**Table 5:** Monthly actual yield, contractual yield, PR and solar insolation for Yarmouk University PV solar power plant during the third operational year (2019) [14]

Month	Actual yield (MW h)	Contractual yield (MW h)	PP (%)	Solar insolation (kW h/m <sup>2</sup> )
Jan	364.09	334.67	108.79	NA
Feb	314.91	336.39	93.61	NA
Mar	381.35	465.05	83.00	NA
Apr	478.05	491.91	97.18	NA
May	558.70	571.64	97.74	208.14
Jun	538.22	586.05	91.84	197.77
Jul	583.32	593.54	589.28	222.29
Aug	553.99	568.09	97.52	214.97
Sep	496.16	505.74	98.11	193.84
Oct	430.39	458.72	93.83	NA
Nov	375.32	382.05	98.24	143.55
Dec	296.14	305.44	96.95	NA
Total year	5370.64	5599.29	96	1180.56

combination. The best models depend on the amount of historical data correlation with explanatory variables.

The study aims to forecast the solar system's power based on historical information. This information comprises a set of features impacting the gathered power energy. Consequently, the system requires a training set of rich data to extract the inherited knowledge in models. Thus, the architecture of the proposed approach is depicted in Figure 3. As shown, the system has two types of features, such as energy-based and time-based, as indicated in the diagram. The energy-based features depend merely on the amount of aggregated energy and surrounding temperatures. On the other hand, the time-based features include the days, weeks, years, and hours of deliberating the energy in the system. The features and the aggregated power are integrated into a dataset to train machine learning (ML) models.

We employ a stratified ten-fold cross-validation technique during the evaluation phase. In a sequential procedure, this approach chooses 10% of the dataset for testing and 90% of the dataset for training (the other nine folds), as illustrated in Figure 3. In each procedure, we create a classifier model and assess its performance. After that, the average performance of all folds is represented. Using such a technique, we ensure that the entire dataset is involved in the training and testing phases, thereby reducing the risk of over-fitting. This problem occurs when the model correctly classifies all training data but fails to fit the test sets.

**XGBoost algorithm:** It is a gradient boosting-based decision-tree-based ensemble machine learning algorithm



[15]. Gradient boosting is a machine learning technique for regression and classification problems that generate a prediction model from an ensemble of weak prediction models, usually decision trees. It is a high-speed and high-performance implementation of gradient-boosted decision trees.

**Random Forest algorithm:** It consists of numerous decision trees that use two fundamental concepts [16]: (i) random sampling of training data points when creating the trees and (ii) random subsets of features examined when splitting the nodes. A decision tree is one of the predictive models where the target variable can take a discrete or continuous set of values. The leaves indicate class labels, whereas the branches represent feature combinations that lead to those labels.

**Bagging algorithm:** It is a bootstrap aggregate a learning approach that combines numerous base models to produce a single best predictive model [17]. The REF-Tree model is used as the base model in this article since it produces better results on the dataset. Each model is trained independently, with the results combined using an average process. The primary focus of bagging is to achieve less variance than any model has individually.

**Linear Regression algorithm:** It is a supervised machine learning approach that uses a polynomial slope-intercept form to connect the independent and dependent variables, with the projected output being a continuous number [18]. In particular, the algorithm learns the best fit of the line between the inputs (independent variables) and the outputs (dependent variables) by reducing the error estimated from the predicted and actual targets.

**Support Vector Machine algorithm:** It is a supervised machine learning technique that can be used to solve issues in classification and regression [19]. Each data item is represented in the SVM method as a point in  $n$ -dimensional space (where  $n$  is the number of features or inputs), with the value of each feature being the value of a certain coordinate. Following that, classification is carried out by locating the hyperplane that separates the two target classes.

Particle swarm optimization (PSO) is a nature-inspired method to solve tough and challenging problems in computation. They often exist in nature as birds, insects, and fish swarms. These swarms exist in nature as controversial as AI-driven weapons. The benefits of these swarms include the fact that predators may not attack swarms of animals. Because the groups are more extensive, louder, and more intimidating than a single animal, and the predator can sometimes be confused as to which animal to concentrate on and hunt because there are too many animals. A single animal may be easy to spot against the background. Thus, the swarm forms the entire background-the many eyes hypothesis. Finally, swarms may also allow quicker movement. Therefore, riders line up behind each other and avoid air resistance. Swarm intelligence can be defined as the collective behavior of decentralized and self-organized systems, natural or artificial. Decentralized means that there is no controlling entity in this swarm. Thus, the entire swamp collectively organizes itself to complete a particular task collaboratively as social learning.

Hence, PSO is a metaheuristic based on flocking swarm behavior. It solves complex optimization problems and

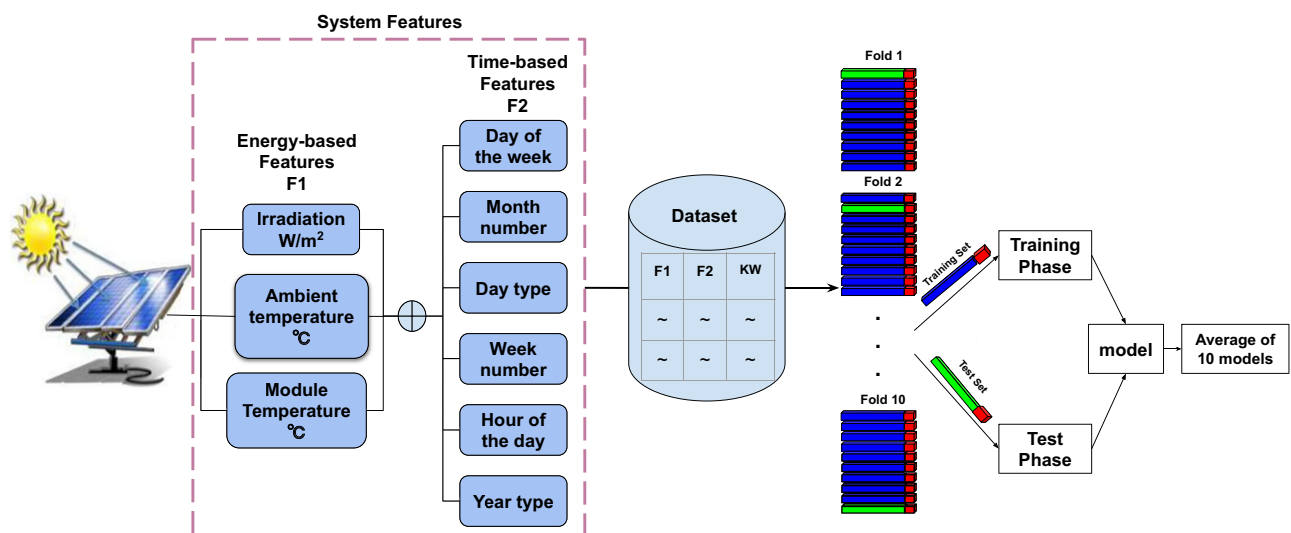


Figure 3: The architecture of the proposed methodology.

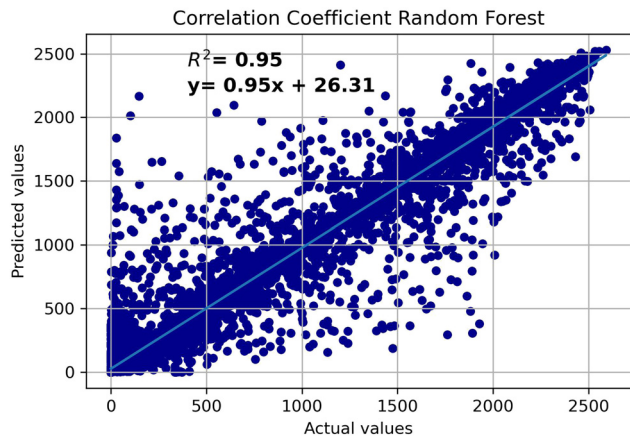
maintains a population of solutions regarding a quantity measure. It depends on inertia, experience, and societal swarm influence to guide the search for better solutions. A particle is a candidate solution, and improvements are made by moving the particles around in the search space. Position and velocity are influenced by each particle’s best-known position, which is also updated by the better position found by other particles in each iteration. PSO uses fewer resources than traditional optimization algorithms. It can search large spaces of candidates’ solutions. It does not use the gradient of the problem being optimized, like classic optimization methods, so this does not require the problem to be differentiable. So, there is no guarantee that an optimal solution will be found. The main applications of PSO are nonlinear optimization problems, neural training networks, and power system forecasting.

## 7 Experiment and results

We conducted several experiments to validate predicting the future consumption of solar power generated from solar panels. ML techniques receive the input dataset and extract the inherited knowledge within a huge amount of records by using a mathematical pattern. As discussed, each ML model maintains its own unique strategy to reduce the error prediction between the actual and the model’s output, resulting in varied outputs on distinct datasets. Table 6 provides the results of ML models using the five evaluation metrics. The Random Forest model obtains a superior outcome compared to other techniques with a CC value of 0.9751, which is approximately identical to the Bagging\_REFTree model’s result of 0.9729. These results indicate that the tree-based approaches outperform the other models. These models split the training set into

**Table 6:** Error forecasting results for the techniques used

Algorithm	Root mean squared error	Mean absolute error
Random forest	172.97	68.7
Bagging-REFTree	180.137	74.365
PSO-XGB	191.983	82.122
Multi-verse optimizer (MVO)-eXtreme gradient boosting (XGB)	256.144	153.630
GA-XGB	398.350	238.735
WOA-XGB	284.903	163.988
Linear regression	556.420	393.480
SVM regression	359.921	618.687



**Figure 4:** The scattered plot between the actual and predicted values of the Random Forest model results.

different sub-training sets and build sub-models into a merged model that reduces bias toward specific records or features. Hence, the error obtained from one sub-tree is handled by other sub-trees. The best results of the forecasting is shown in Figure 4. The figure shows the actual and predicted values of the Random Forest model results.

## 8 Conclusion

PV solar power plant forecasting is needed to prioritize investment strategies, inform collective impacts, and evaluate technology innovations. Also, this type of forecasting depends on localized factors and evolving technology. PV solar power forecasting gives the ability to predict solar energy output and let grids function better under variable conditions. It provides a way for the grid operators to predict and balance energy generation and consumption. Thus, the main goal is to develop a streamlined surrogate model for assessing plants. PV solar power forecasting is integrated into energy management systems. It is increasingly valuable to electrical energy system operators. In this work, several experiments are conducted to validate predicting the future consumed PV solar power plant in Yarmouk University. These techniques are Random Forest, Bagging-REFTree, PSO-XGB, MVO-XGB, linear regression, and SVM regression. The Random Forest model obtains a superior result in comparison to other techniques.

**Acknowledgment:** The authors would like to express special thanks to Yarmouk University’s engineering, production, and maintenance departments for providing access to the research data.

**Conflict of interest:** Authors state no conflict of interest.

## References

- [1] Alhmoud L, Nawafleh Q. Short-term load forecasting for Jordan's power system using neural network based different. In: 2019 IEEE International Conference on Environment and Electrical Engineering and 2019 IEEE Industrial and Commercial Power Systems Europe (EEEIC/I & CPS Europe). Genova, Italy: IEEE; 2019 Jun 11. p. 1–6.
- [2] Alhmoud L, AbuKhurma R, Al-Zoubi AM, Aljarah I. A real-time electrical load forecasting in Jordan using an enhanced evolutionary feedforward neural network. *Sensors*. 2021 Sep 17;21(18):6240.
- [3] Alhmoud L, Nawafleh Q. Short-term load forecasting for Jordan power system based on NARX-Elman neural network and ARMA model. *IEEE Can J Electr Comput Eng*. 2021 Jul 15;44(3):356–63.
- [4] Alasali F, Nusair K, Alhmoud L, Zarour E. Impact of the Covid-19 pandemic on electricity demand and load forecasting. *Sustainability*. 2021 Jan 29;13(3):1435.
- [5] Goldstone JA. Using quantitative and qualitative models to forecast instability. Washington, DC: United States Institute of Peace; 2008.
- [6] vom Scheidt F, Medinová H, Ludwig N, Richter B, Staudt P, Weinhardt C. Data analytics in the electricity sector—a quantitative and qualitative literature review. *Energy and AI*. 2020 Aug 1;1:100009.
- [7] Salkuti SR. A survey of big data and machine learning. *Int J Electr Comput Eng* (2088–8708). 2020 Feb 15;10(1):575–80.
- [8] Barrios M, Guilera G, Nuño L, Gómez-Benito J. Consensus in the Delphi method: What makes a decision change? *Technol Forecast Soc Change*. 2021 Feb 1;163:120484.
- [9] Sari T. Responsive Demand Management in the Era of Digitization. In: *Strategic Outlook for Innovative Work Behaviours*. Cham: Springer; 2020. p. 275–91.
- [10] Cai R, Chen J, Li Z, Chen W, Zhang K, Ye J, et al. Time series domain adaptation via sparse associative structure alignment. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 35, Issue 8; 2021 May 18. p. 6859–67.
- [11] Lai Y, Dzombak DA. Use of the autoregressive integrated moving average (ARIMA) model to forecast near-term regional temperature and precipitation. *Weather Forecast*. 2020 Jun;35(3):959–76.
- [12] Smyl S. A hybrid method of exponential smoothing and recurrent neural networks for time series forecasting. *Int J Forecast*. 2020 Jan 1;36(1):75–85.
- [13] Alzahrani SI, Aljamaan IA, Al-Fakih EA. Forecasting the spread of the COVID-19 pandemic in Saudi Arabia using ARIMA prediction model under current public health interventions. *J Infect Public Health*. 2020 Jul 1;13(7):914–9.
- [14] Kawar Energy Company. Annual Report for PV Solar Pwer Plant for Yarmouk University; Jan 30, 2021.
- [15] Chen T, He T, Benesty M, Khotilovich V, Tang Y, Cho H, et al. Xgboost: extreme gradient boosting. R package version 0.4-2. 2015 Aug 1;1(4):1–4.
- [16] Chaudhary A, Kolhe S, Kamal R. An improved random forest classifier for multi-class classification. *Inf Process Agric*. 2016 Dec 1;3(4):215–22.
- [17] Tüysüzöğlü GÖ, Birant D. Enhanced bagging (eBagging): A novel approach for ensemble learning. *Int Arab J Inf Technol*. 2020;17(4):1–17.
- [18] Ottaviani FM, De Marco A. Multiple linear regression model for improved project cost forecasting. *Procedia Comput Sci*. 2022 Jan 1;196:808–15.
- [19] Velásquez RM. Support vector machine and tree models for oil and Kraft degradation in power transformers. *Eng Fail Anal*. 2021 Sep 1;127:105488.