NTERNATIONAL JOURNAL OF CIRCUITS, SYSTEMS AND SIGNAL PROCESSING
DOI: 10.46300/9106.2020.14.88 Volume 14, 2020

An Efficient Intelligent Power Detection Method for Photovoltaic System

Ayman M Mansour, Jalal Abdallah, and Mohammad A Obeidat
Tafila Technical University,
'IS, Tafila, 66110,
Jordan

Abstract— Jordan has experienced a significant increase in both peak load and annual electricity demand within the last decade due to the growth of the economy and population. Photovoltaic (PV) system is one of the most popular renewable energy source in Jordan. PV system is highly nonlinear with unpredictable behavior since it is always subject to many external factors such as severe weather conditions, irradiance level, sheds, temperature, etc. This makes it difficult to maintain maximum power production around its operation ranges.

In this paper, an intelligent technique is used to predict and identify the working ability of the PV system under different weather factors in Tafila Technical University (TTU) in Jordan. It helps in optimizing power productions for different operation points. The PV system in Tafila with size 1 MWp PV generated 5.4 GWh since 2017. It saves about € 1.5 million in three years. A real power data from the PV system and a weather data from world weather online site of TTU location are used in this study. Decision tree technique is employed to identify the relation between the output power and weather factors. The results show that the system accuracy is 82.01% during the training phase and 93.425 % on the validation set.

Keywords—PV system, Decision Tree, weather features, Power, Real Data.

I. INTRODUCTION

Highly demand for energy makes the electrical power systems grow day by day rapidly. Thus, thinking of establishing new power sources will be the persistent need. A rapid increase in the cost of fossil fuel, which is used in conventional energy sources, makes the decision makers employ a new kind of energy sources with lower cost and environmentally friendly. Alternative Energy (AE) sources such as wind farms, photovoltaic PV technology, biomass, hydropower and many different new power alternatives are used. The geometrical region, environmental issues, natural sources and weather conditions should be taken into account before establishing any type of AE source [1].

The use of renewable energy sources adds additional complexity to power systems and makes them more challenging to the operator. The renewable power generation sources are divided into two categories: the sources, which have similar characteristics to conventional power generation facilities in that they are predictable and controllable, such as hydroelectric generation and biomass; and the variable and intermittent sources, such as wind and solar [2].

The effect of wind and solar renewable sources on power system stability is different from that of the conventional power sources. Wind and solar are depending on variable conditions such as weather conditions, as wind speed and solar irradiance, site dependence and types of generators used [3]. Such weather poses many challenges for system operators to ensure grid stability and reliability.

Jordan experienced a significant increase in both peak load and annual electricity demand within the last decade due to a strong growth of economy and population. The peak load of Jordan's electrical system in 2017 was 3282 MW and increased to be 3724 MW in 2020. The peak load for the years from 2017 to 2020 is shown in Fig. 1

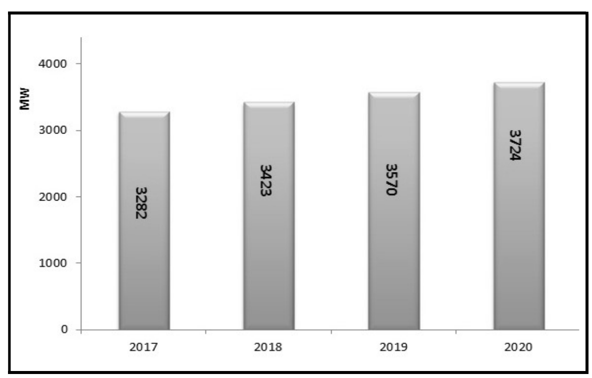

Fig. 1 National grid peak load (2017.2020) [1]

Solar and wind energy systems are one of the most prominent sources of energy, and their utilization has become increasingly popular due to modular and environmental friendly nature [2-4]. Jordan has a high solar index as shown in Fig. 2.

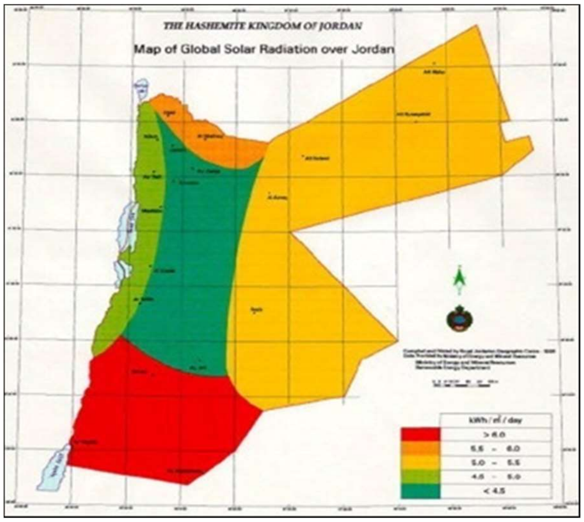

Fig. 2 Global solar radiation over Jordan [1]

Fig.3 shows the renewable (PV and Wind) energy projects distribution all over Jordan. Most of the RE sources are located in south and far away from the center. The energy is transmitted to the consumer with overhead transmission lines with different voltage levels mainly 132 kV and 400 kV [1].

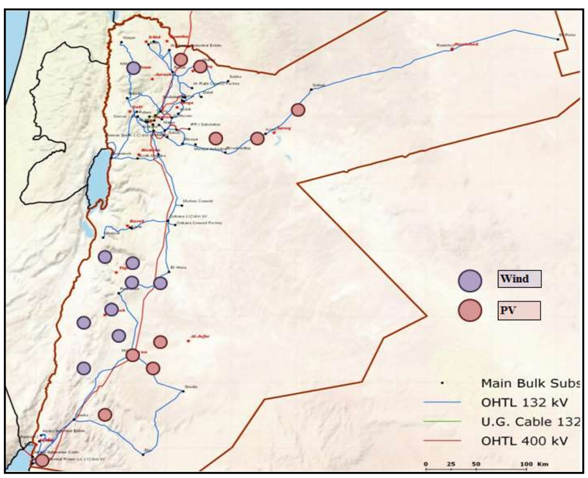

Fig. 3 PV and wind energy project distribution

Nowadays, producing Energy from solar system is very important as the conventional energy has noticeable bad impact on environment. Many methods are used to analyze, develop, design, and control the photovoltaic system. Such methods are very important especially when integrating renewable sources with conventional power [5-7].

The efficiency of a solar panel depends on three main factors: the efficiency of the model used for a particular panel, the number of photovoltaic model inside each solar cell, and the amount of sunlight that the PV panel received. After installing the PV panels, inverters are added to transfer electricity from direct current (DC) to alternating current (AC). The efficiency of a solar panel depends on the solar incident angle. When the sunlight is vertical, the efficiency of any given solar panel is at its maximum.

PV is very sensitive to high temperatures and dust, which reduces its efficiency. Solar PV is more practical for homes and commercial buildings, while PV technology is very expensive; its cost has decreased dramatically over time, especially in the past five years.

Recent researches are still going on to make PV systems more efficient. Power detection methods are widely used in the literature, but these methods fail under a rapid change in weather conditions. Therefore, many improvements have been made to these algorithms to mitigate inaccurate responses during abrupt changes in the level of sunshine.

Data mining techniques are used recently in many applications because of its benefits to develop models and to make decisions [8-15]. Some methods use artificial intelligent techniques such as artificial neural network as in [16-20]. Other methods use data mining techniques like support vector machine [21-24] and K-nearest neighbor as in [25-27]. In addition, there are some optimization techniques such as genetic algorithm and particle swarm are used to predict and improve the solar system depending on environmental factors such as the temperature, wind, and cloud [28-31].

Authors in [32] deal with similar situation, but they did their analysis of power generated from solar system without considering weather data; instead, they used artificial neural network to predict the power. They claim that the power can be forecasted without using weather data. However, the power generated from PV system is not similar to the conventional power because it is directly depending on weather factors such as temperature, wind, and cloud cover. In addition, they claim that some outliers happened in the system, and they eliminate them. This is not true in the case of PV system since each day has its own weather conditions and sometimes the climate changes suddenly. Such changes affect the power generation, which cannot be considered as outliers. The developed system in this paper solve these problems.

II. 1 MWp TTU PV Power Plant

Jordan government encourages all public universities to invest in solar PV system to decrease their electricity bills. Tafila Technical University started with 1 MWp project in 2016 that cost € 1.4 million. After three years, this successful project has generated 5.4 GWH and save about € 1.5 million. The 1 MWp TTU PV power plant (Fig.4) consists of 3,876 265-Wp PV modules of type SR-P660 245-265 manufactured by Sunrise Energy Co. The modules are distributed on the rooftops of university buildings and car parks inside the university campus as shown in Fig. 4. Four transformers are used in TTU electrical system to supply the output AC power to the buildings as shown in Tables 1.

Table 1. Transformers of TTU system

Transformer number	Building Description
Transformer 1	Car Park 1 Engineering Faculty Girls Hostel Multi-Function Building Science Building Storage Hunger
Transformer 2	Business College Labs Hunger Presidency Building President Suite
Transformer 3	GYM
Transformer 4	Car Park 02 Class Room Building

The system uses 42 inverters distributed in the whole TTU campus as shown in Table 2, all are manufactured by ABB. Out of the 42 inverters 6 are used for photovoltaic car parks. The PV panels are arranged at a tilt angle of 22° and an azimuth angle of -16°.

Fig. 4 1MWp TTU PV system for rooftop

From Table 2, it's cleared that all inverters are working within the acceptable range of oversize percentage specified by ABB string configuration report. Such oversizing will not affect the inverter expected lifetime and the inverters operate normally.

Table 2. TTU PV system inverters distributed over all campus building and car parks

Building Description	Inverter Model	Number of Inverters	Inverter Power (kwp)	Produced Power (kwp) in 2019	Inverter Oversize (%)
Business College T2	TRIO 27.6 TL	2	55.2	60.927	10.375
Car Park 02 T4	TRIO 27.6 TL	2	55.2	59.442	7.685386
Car Park 1 T1	TRIO 27.6 TL	4	110.4	112.668	2.054801
Class Room Building T4	PVI-10.0-OUTD	1	10	10.825	8.258333
	TRIO 27.6 TL	4	110.4	93.706	-15.1209
Engineering Faculty T1	PVI-12.5-OUTD Universal	3	37.5	39.014	4.038
	TRIO 20.0 TL	2	40.0	41.210	3.026875
	TRIO 27.6 TL	4	110.4	114.603	3.807292
Girls Hostel T1	PVI-12.5-OUTD	2	25	28.209	12.836
GYM T3	TRIO 27.6 TL	8	220.8	206.288	-6.57224
	PVI-10.0-OUTD	1	10	9.494	-5.05083
Labs Hunger T2	PVI-12.5-OUTD	1	12.5	12.395	-0.83267
	TRIO 27.6 TL	1	27.6	22.613	-18.0658
Multi-Function Building T1	TRIO 8.5 TL	2	17	17.294	1.732843
Presidency Building T2	RIO 20.0 TL	1	20	20.824	4.123333
President Suite T2	PVI-12.5-OUTD	1	12.5	12.669	1.354667
Science Building T1	PVI-12.5-OUTD	1	12.5	11.715	-6.27733
Storage Hunger T1	PVI-12.5-OUTD	1	12.5	12.1683	-2.65333
	TRIO 27.6 TL	1	27.6	26.620	-3.54801
Total			871.9	912.693	---

The working principle of a PV plant is simple; photovoltaic cells are connected in series or parallel in order to obtain the desired current and voltage value for the PV module.

Modules are mounted on structures (typically aluminum anodized) that fix the modules either to the roof or to the ground. Ground mounted systems usually tilt the inclination angle of the modules in order to optimize the radiation input during the year. Roof mounted systems usually fix the modules parallel to the roof if the slope is sufficient for good radiation levels. On the other hand, tilted systems are also available for flat roofs. The modules are interconnected using standard electric copper cables.

Some ground-mounted systems use trackers (one or two axes) in order to follow the sun path and increase the amount of solar radiation received at the surface of the modules. The PV array is then connected to the inverter, which converts direct current (DC) into alternating current (AC). The inverter incorporates Maximum Power Point Trackers (MPPT) in order to follow the constantly changing current and voltage output of the array. The current and voltage output are changing mainly because of changing conditions like irradiance and temperature. The inverter output is then connected to a meter, which registers the amount of energy that has been fed into the grid.

Fig. 5 shows the output AC power generated from one inverter type TRIO 27.6, with generated energy 48.41 MWh from GYM building modules over one-year 2019. The energy produced by TTU PV power plant in 2019 is shown in Fig. 6.

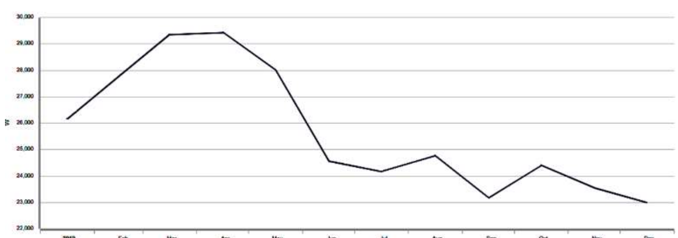

Fig. 5 Output AC power generated from one inverter of GYM building modules, 2019

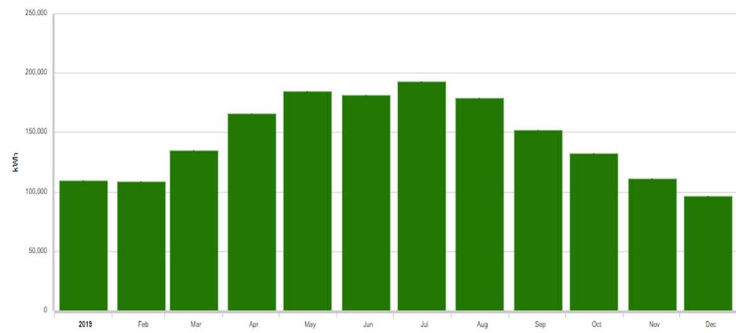

Fig. 6 Energy produced by Tafila Technical University PV power plant in 2019

III. System Design and Methodology

In this paper, output power forecasting system involves five phases: Data gathering and extraction phase, data preprocessing phase, feature selection phase, the learning phase, and the classification phase. In data gathering and extraction, the training and test set are obtained from both Tafila PV system and weather station databases. The second phase is to pre-process the extracted data, including cleaning unnecessarily information, normalizing and labeling. In selected features, dimensionality reduction of the features is being conducted. In the learning phase, the target is to build a model using a part of the data. The last step is using the remaining of the pre-processed data to test the model. A test set is used to determine the accuracy of the model. Fig. 7 shows the block diagram of the developed system phases.

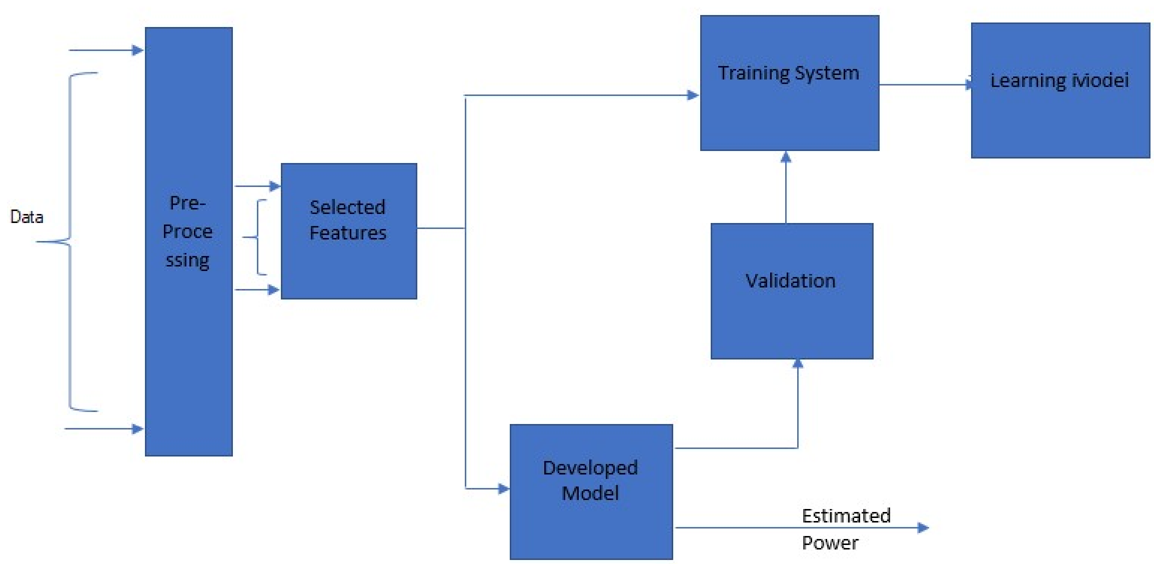

Fig. 7 Block diagram of the developed system phases

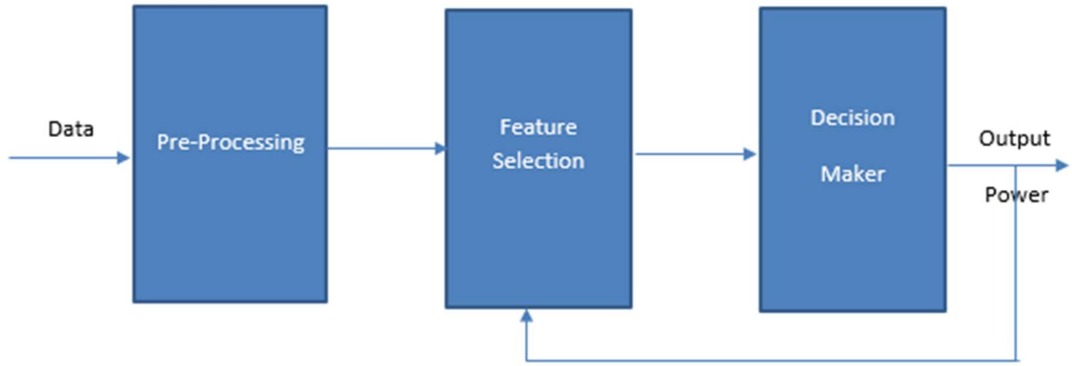

Fig. 8 Training model

The learning model (Fig. 8) is based on Decision Tree. Decision Tree classifier uses the training data to build a tree model that can be used later for classification purposes. Currently, many decision tree algorithms exist including Random Forest, Random Tree, J48, and CART. Based on the output accuracy, the features set is selected. The best set is associated with the higher output power accuracy. There are many different reduction techniques available including Principal Component Analysis (PCA), and Chi-Squared for feature selection and reduction. In this paper, PCA is used as a reduction technique [13],[15].

The extracted data from the databases includes Average Temperature, Wind speed, Humidity, Visibility, Heat index, Pressure, Weather description, cloud cover and generated power. After preprocessing phase and feature selection, six features are used. Five features are used as input features and one feature is used as output class. The six features are Average Temperature, Humidity, Pressure, cloud cover, Heat index and generated output (output class). The Decision maker is based on decision tree classifiers. When a new data values, previously unseen, is presented to the decision maker, the class is predicted based on the training instances. The Decision maker output will be the amount of output power. It has four labels; Very Low, Medium, High, Very High. Fig. 9 shows the testing model block diagram.

A decision tree partition the input space of the dataset into mutually exclusive regions by giving each region a label. The decision tree that consists of a root and internal nodes grows from a root node, by determining the best split that partition the region at internal nodes in to disjoint smaller subset and proceed down to the leaf node (terminal nodes) labeled as – Very Low, Medium, High, and Very High. In order to do the split an error function that quantifies the performance of a node t in separating data from different classes. The used error function is impurity function. The best-known impurity functions for splitting is entropy function and Gini index. By using the impurity Function φ, the impurity measure of a given node is calculated to do the splitting. Fig 10 shows the steps of decision tree methodology.

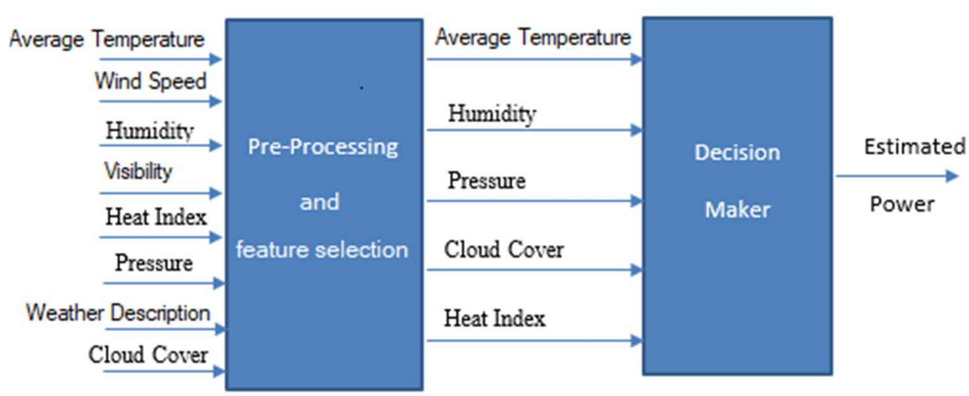

Fig. 9 Testing model

Step 1: Finding the probability of a class j at node t such that $p_j = P(j|t)$

Step 2: Finding the impurity measure of a node

$$E(t) = \varphi(p_1, p_2, \dots p_j) = -\sum_{j=1}^{J} p_j \log_2 p_j$$

Step 3: Finding the impurity measure of a tree T that is expressed as

$$E(T) = \sum_{t\in\Psi} \frac{n_t}{n} E(t)$$

Where Ψ is the set of terminal nodes in the tree T, nt = number of records at child t,

Step 4: Calculating the information gain of Parent Node P (non-leaf node, node with partition)

$$GAIN_{Split} = E(P) - E(T)$$

$$= E(P) - \sum^{k} \frac{n_i}{n} E(i)$$

Step 5: Calculating the information gain ratio

$$GainRATIO = \frac{GAIN_{Split}}{-\sum_{i=1}^{k} \frac{n_i}{n} \log_2 \frac{ni}{n}}$$

Fig. 10 Decision tree methodology

IV. Experiments and Results

The data used in this paper is extracted from two databases. The first database is Tafila PV database. This database is provided from inverter manufacture (ABB) as cloud database. The extracted data includes the power produced from the PV system from December, 23th , 2016 to February, 29st, 2020. This data includes 1164 records as shown in Fig. 11. Part of the data is shown in Table 3.

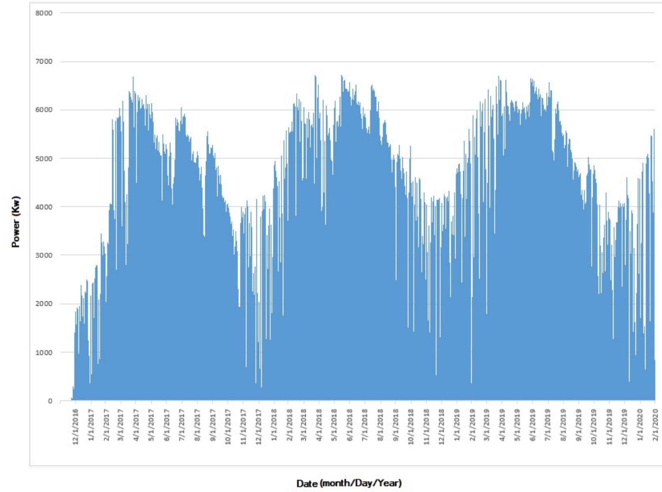

Fig. 11 Generated output power from PV system in Tafila Technical University

Table 3. Tafila PV station Database sample

Entity Id ->	10807492
Entity Name ->	Tafila Technical University
Field ->	Generated Energy
Timestamp	kilowatt-hours
2/1/2020 0:00	1707.1
2/2/2020 0:00	3255.1
2/3/2020 0:00	4732.1
2/4/2020 0:00	3962.7
2/5/2020 0:00	4807.5
2/6/2020 0:00	4915.1
2/7/2020 0:00	1398.1
2/8/2020 0:00	1532.2
2/9/2020 0:00	339.4
2/10/2020 0:00	641.1
2/11/2020 0:00	3607.5
2/12/2020 0:00	4892.3
2/13/2020 0:00	5045.7
2/14/2020 0:00	4063.4
2/15/2020 0:00	5096.5
2/16/2020 0:00	4038.3
2/17/2020 0:00	5005.4
2/18/2020 0:00	4270.8

Table 4. Weather station database sample

date	maxtempC	mintempC	avgtempC	totalprecipMM	windspeedMiles	windspee	sunhour	winddirdegree	winddir16point	weatherCode	weatherDesc	humidity	visibilityKm	pressureM	pressureInches	cloudcover	HeatIndexC	DewPointC	WindChil	WindGustKmph	FeelsLikeC	sunrise	sunset	moonrise	moonset
7/1/2008	36	20	28	0	12	20	11	306	NW	113	Sunny	41	10	1006	30	0	28	11	28	12	FeelsLikeC	4:38 AM	6:45 PM	2:21 AM	5:22 PM
7/2/2008	36	19	27	0	12	20	11	285	WNW	113	Sunny	46	10	1006	30	0	27	13	27	15	FeelsLikeC	4:38 AM	6:44 PM	3:26 AM	6:27 PM
7/3/2008	36	20	27	0	13	21	11	289	WNW	113	Sunny	50	10	1006	30	1	27	14	27	13	FeelsLikeC	4:39 AM	6:44 PM	4:36 AM	7:23 PM
7/4/2008	35	19	27	0	12	19	11	320	NW	113	Sunny	48	10	1007	30	0	27	13	27	13	FeelsLikeC	4:39 AM	6:44 PM	5:50 AM	8:09 PM
7/5/2008	35	20	27	0	14	22	11	297	WNW	113	Sunny	45	10	1008	30	0	27	12	27	14	FeelsLikeC	4:39 AM	6:44 PM	7:00 AM	8:47 PM
7/6/2008	35	19	26	0	13	21	11	276	W	113	Sunny	50	10	1008	30	1	26	13	26	15	FeelsLikeC	4:40 AM	6:44 PM	8:07 AM	9:21 PM
7/7/2008	35	18	27	0	13	21	11	278	W	113	Sunny	52	10	1007	30	0	27	14	27	14	FeelsLikeC	4:40 AM	6:44 PM	9:11 AM	9:51 PM
7/8/2008	35	19	27	0	13	21	11	268	W	113	Sunny	49	10	1006	30	0	27	13	27	14	FeelsLikeC	4:41 AM	6:44 PM	10:10 AM	10:19 PM
7/9/2008	36	19	28	0	11	18	11	303	WNW	113	Sunny	44	10	1006	30	0	27	11	28	13	FeelsLikeC	4:41 AM	6:43 PM	11:08 AM	10:47 PM
7/10/2008	38	20	29	0	11	17	11	215	SW	113	Sunny	30	10	1006	30	0	28	6	29	13	FeelsLikeC	4:42 AM	6:43 PM	12:04 PM	11:16 PM
7/11/2008	38	21	29	0	11	18	11	278	W	113	Sunny	29	10	1006	30	0	29	6	29	12	FeelsLikeC	4:42 AM	6:43 PM	1:01 PM	11:48 PM
7/12/2008	36	20	28	0	11	17	11	249	WSW	113	Sunny	38	10	1006	30	0	28	10	28	13	FeelsLikeC	4:43 AM	6:43 PM	1:58 PM	No moons
7/13/2008	37	20	28	0	12	20	11	297	WNW	113	Sunny	42	10	1007	30	0	28	11	28	14	FeelsLikeC	4:44 AM	6:42 PM	2:55 PM	12:22 AM
7/14/2008	37	19	28	0	14	22	11	278	W	113	Sunny	46	10	1010	30	0	27	13	28	13	FeelsLikeC	4:44 AM	6:42 PM	3:52 PM	1:03 AM
7/15/2008	37	20	28	0	14	22	11	277	W	113	Sunny	50	10	1009	30	0	28	15	28	15	FeelsLikeC	4:45 AM	6:42 PM	4:45 PM	1:48 AM
7/16/2008	37	19	28	0	11	18	11	294	WNW	113	Sunny	42	10	1006	30	0	27	12	28	12	FeelsLikeC	4:45 AM	6:41 PM	5:35 PM	2:39 AM
7/17/2008	38	20	29	0	13	21	11	290	WNW	113	Sunny	39	10	1005	30	0	28	11	29	12	FeelsLikeC	4:46 AM	6:41 PM	6:19 PM	3:35 AM
7/18/2008	36	19	28	0	11	18	11	309	NW	113	Sunny	52	10	1007	30	0	27	15	28	12	FeelsLikeC	4:46 AM	6:40 PM	6:58 PM	4:33 AM
7/19/2008	38	19	28	0	12	19	11	314	NW	113	Sunny	46	10	1007	30	0	28	12	28	11	FeelsLikeC	4:47 AM	6:40 PM	7:33 PM	5:33 AM
7/20/2008	37	19	28	0	13	21	11	303	WNW	113	Sunny	47	10	1007	30	0	28	14	28	14	FeelsLikeC	4:48 AM	6:40 PM	8:05 PM	6:32 AM
7/21/2008	38	21	29	0	12	20	11	204	SSW	113	Sunny	43	10	1007	30	0	29	12	29	13	FeelsLikeC	4:48 AM	6:39 PM	8:34 PM	7:32 AM
7/22/2008	38	20	29	0	11	18	11	304	WNW	113	Sunny	33	10	1006	30	0	29	8	29	12	FeelsLikeC	4:49 AM	6:38 PM	9:03 PM	8:31 AM
7/23/2008	37	20	28	0	12	20	11	288	WNW	113	Sunny	43	10	1007	30	0	28	12	28	13	FeelsLikeC	4:49 AM	6:38 PM	9:33 PM	9:31 AM
7/24/2008	37	19	28	0	13	22	11	293	WNW	113	Sunny	50	10	1006	30	0	28	14	28	15	FeelsLikeC	4:50 AM	6:37 PM	10:04 PM	10:32 AM
7/25/2008	35	20	27	0	13	21	11	302	WNW	113	Sunny	57	10	1007	30	1	28	16	27	15	FeelsLikeC	4:51 AM	6:37 PM	10:39 PM	11:37 AM
7/26/2008	33	19	25	0	12	19	11	296	WNW	113	Sunny	58	9	1008	30	18	25	15	25	15	FeelsLikeC	4:51 AM	6:36 PM	11:21 PM	12:44 PM
7/27/2008	32	18	25	0	10	17	11	301	WNW	113	Sunny	57	10	1008	30	2	25	14	25	13	FeelsLikeC	4:52 AM	6:35 PM	No moonı	1:54 PM
7/28/2008	32	17	25	0	12	19	11	294	WNW	113	Sunny	59	9	1008	30	3	25	14	25	13	FeelsLikeC	4:52 AM	6:35 PM	12:09 AM	3:04 PM
7/29/2008	33	18	25	0	12	19	11	288	WNW	113	Sunny	59	8	1007	30	15	26	15	25	13	FeelsLikeC	4:53 AM	6:34 PM	1:08 AM	4:10 PM
7/30/2008	34	17	26	0	12	19	11	287	WNW	113	Sunny	56	9	1007	30	0	26	14	26	13	FeelsLikeC	4:54 AM	6:33 PM	2:14 AM	5:09 PM
7/31/2008	35	17	26	0	11	18	11	299	WNW	113	Sunny	56	9	1006	30	0	26	15	26	12	FeelsLikeC	4:54 AM	6:33 PM	3:26 AM	5:59 PM
8/1/2008	37	19	28	0	12	19	11	295	WNW	113	Sunny	50	9	1007	30	0	28	14	28	12	FeelsLikeC	4:55 AM	6:32 PM	4:37 AM	6:41 PM
8/2/2008	37	19	28	0	11	18	11	293	WNW	113	Sunny	43	10	1007	30	0	28	12	28	14	FeelsLikeC	4:56 AM	6:31 PM	5:47 AM	7:17 PM
8/3/2008	36	19	28	0	11	18	11	291	WNW	113	Sunny	47	10	1007	30	0	28	14	28	13	FeelsLikeC	4:56 AM	6:30 PM	6:52 AM	7:49 PM
8/4/2008	36	20	27	0	12	19	11	299	WNW	113	Sunny	50	10	1006	30	0	27	13	27	14	FeelsLikeC	4:57 AM	6:29 PM	7:55 AM	8:18 PM
8/5/2008	35	19	26	0	12	19	11	300	WNW	113	Sunny	56	9	1006	30	13	26	14	26	14	FeelsLikeC	4:58 AM	6:29 PM	8:54 AM	8:46 PM
8/6/2008	35	18	27	0	10	16	11	297	WNW	113	Sunny	49	10	1005	30	0	26	12	27	11	FeelsLikeC	4:58 AM	6:28 PM	9:53 AM	9:16 PM
8/7/2008	35	19	27	0	11	18	11	276	W	113	Sunny	46	10	1004	30	0	27	12	27	11	FeelsLikeC	4:59 AM	6:27 PM	10:50 AM	9:47 PM
8/8/2008	35	19	27	0	12	20	11	279	W	113	Sunny	41	10	1005	30	0	27	10	27	13	FeelsLikeC	4:59 AM	6:26 PM	11:49 AM	10:20 PM
8/9/2008	35	20	27	0	13	21	11	269	W	113	Sunny	45	10	1005	30	1	27	12	27	14	FeelsLikeC	5:00 AM	6:25 PM	12:46 PM	10:59 PM

The second database is the weather station database, World Weather Online database. World Weather Online is dedicated to provide global weather forecast and weather content for websites, businesses and the travel industry. This database covers approximately 3 million cities/towns worldwide and its weather forecast is trusted and used by a wide variety of companies and organizations from SME's to large corporate clients. The weather station currently operates two high-tech weather data centers; one situated in Denmark and the other in Germany. The station delivers reliable and accurate weather information for any geo-point in the world. The used weather model is run along with other metrological models like European Centre for Medium-Range Weather Forecasts, World Meteorological Organization, NASA weather satellite imagery, NOAA GFS2 model and JMA model for research and training purposes. It deliver the most accurate weather forecast possible. The extracted data from weather station includes many features such as maxtempC, mintempC, avgtempC, totalprecipMM, windspeedKmph, sunhour winddirdegree, weatherDesc, humidity, visibilityKm pressureMB, cloudcover, heatIndexC, dewPointC, windChillC, windGustKmph, feelsLikeC, sunrise, sunset, moonrise, and moonset. A sample from the weather database is shown in Table 4.

In this paper, just eight weather features are used. These features avgtempC, windspeedKmph, weatherDesc, humidity, visibilityKm pressureMB, cloudcover, and HeatIndexC. The other features has been eliminated.

Fig. 12, Fig. 13, Fig. 14 and Fig. 15 show examples of the extracted features. It shows the weather data from July 1st , 2008 to Februry, 12, 2020 (4232 records) for Tafila Technical University location where the PV system is located.

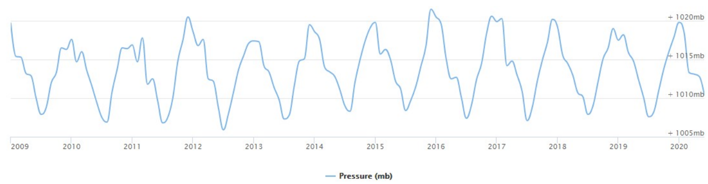

Fig. 12 Pressure data

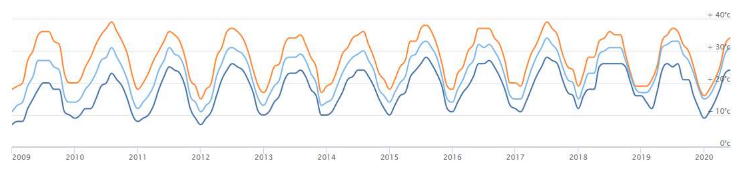

Fig. 13 Temperature data

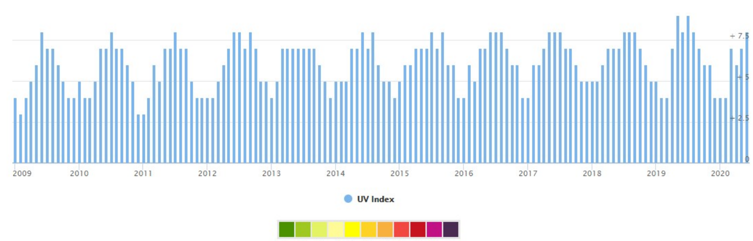

Fig. 14 UV Index data

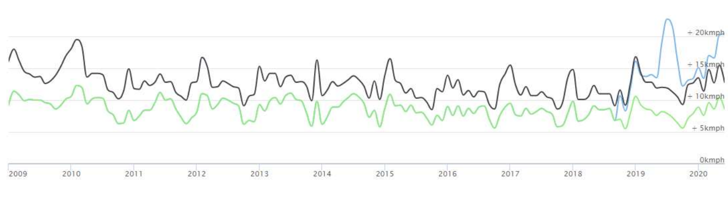

Fig. 15 Wind data

Table 5. Bended database (Weather database and PV system database)

1: date Nominal	2: Avg temp Numeric	3: wind speed Numeric	4: sun hour Numeric	5: weather Descp Nominal	6: humidity Numeric	7: visibility Km Numeric	8: pressure MB Numeric	9: cloud cover Numeric	10: Heat Index C Numeric	11: **Power out** Nominal
2/28/2017	22.0	17.0	11.0	Partly cloudy	24.0	10.0	1011.0	13.0	22.0	very low
3/1/2017	19.0	10.0	11.0	Sunny	49.0	10.0	1010.0	9.0	19.0	very low
3/2/2017	18.0	12.0	11.0	Sunny	52.0	10.0	1015.0	3.0	18.0	very low
3/3/2017	17.0	16.0	11.0	Sunny	52.0	10.0	1016.0	3.0	17.0	very low
3/4/2017	17.0	9.0	11.0	Sunny	50.0	10.0	1018.0	1.0	18.0	meduim
3/5/2017	18.0	9.0	11.0	Sunny	42.0	10.0	1018.0	0.0	18.0	meduim
3/6/2017	20.0	12.0	11.0	Sunny	29.0	10.0	1015.0	1.0	20.0	meduim
3/7/2017	19.0	12.0	11.0	Sunny	43.0	10.0	1013.0	0.0	19.0	meduim
3/8/2017	19.0	6.0	11.0	Sunny	34.0	10.0	1015.0	0.0	18.0	meduim
3/9/2017	23.0	11.0	11.0	Sunny	20.0	10.0	1012.0	0.0	22.0	high
3/10/2017	20.0	11.0	11.0	Sunny	46.0	10.0	1008.0	2.0	20.0	high
3/11/2017	18.0	27.0	11.0	Sunny	39.0	10.0	1010.0	7.0	19.0	meduim
3/12/2017	21.0	31.0	11.0	Sunny	30.0	10.0	1012.0	3.0	20.0	meduim
3/13/2017	18.0	14.0	11.0	Sunny	42.0	10.0	1018.0	4.0	18.0	high
3/14/2017	17.0	9.0	11.0	Sunny	44.0	10.0	1019.0	0.0	17.0	very high
3/15/2017	17.0	21.0	11.0	Sunny	47.0	10.0	1015.0	6.0	17.0	very high
3/16/2017	16.0	11.0	11.0	Sunny	50.0	10.0	1016.0	3.0	16.0	meduim
3/17/2017	19.0	8.0	11.0	Sunny	35.0	10.0	1014.0	17.0	19.0	meduim
3/18/2017	21.0	9.0	11.0	Sunny	26.0	10.0	1008.0	25.0	21.0	meduim
3/19/2017	17.0	17.0	11.0	Sunny	51.0	10.0	1013.0	4.0	17.0	meduim
3/20/2017	16.0	10.0	11.0	Sunny	48.0	10.0	1017.0	3.0	16.0	very high
3/21/2017	19.0	8.0	11.0	Sunny	32.0	10.0	1015.0	22.0	19.0	meduim
3/22/2017	20.0	20.0	11.0	Sunny	32.0	10.0	1010.0	23.0	20.0	meduim
3/23/2017	20.0	9.0	11.0	Sunny	36.0	10.0	1010.0	11.0	20.0	high
3/24/2017	18.0	10.0	11.0	Sunny	40.0	10.0	1015.0	1.0	18.0	very high
3/25/2017	19.0	11.0	11.0	Sunny	41.0	10.0	1016.0	1.0	19.0	very high
3/26/2017	20.0	9.0	11.0	Sunny	41.0	10.0	1017.0	1.0	20.0	very high
3/27/2017	20.0	10.0	11.0	Sunny	37.0	10.0	1016.0	19.0	20.0	high
3/28/2017	20.0	9.0	11.0	Sunny	29.0	10.0	1016.0	16.0	20.0	very high

A normalization is applied to the output power values in order to label them. The power values are normalized to values between zero and one. Each generated output power is assigned as a level that belongs to one of the four levels, i.e., The levels are shown below:

- Level 1: value from 0.00 to 0.39 represents Very Low generated power amount.
- Level 2: value from 0.40 to 0.59 represents Medium Similarity generated power amount.
- Level 3: value from 0.60 to 0.79 represents High Similarity generated power amount.
- Level 4: value from 0.80 to 1.0 represents Very High Similarity generated power amount.

The blended database has been constructed from the two databases in order to train and test the decision maker model after assigning the power output level for each record. Table 5 shows some records from the bended database.

Fig. 16 shows the histogram of five feature with a projection of output power label presented as colors.

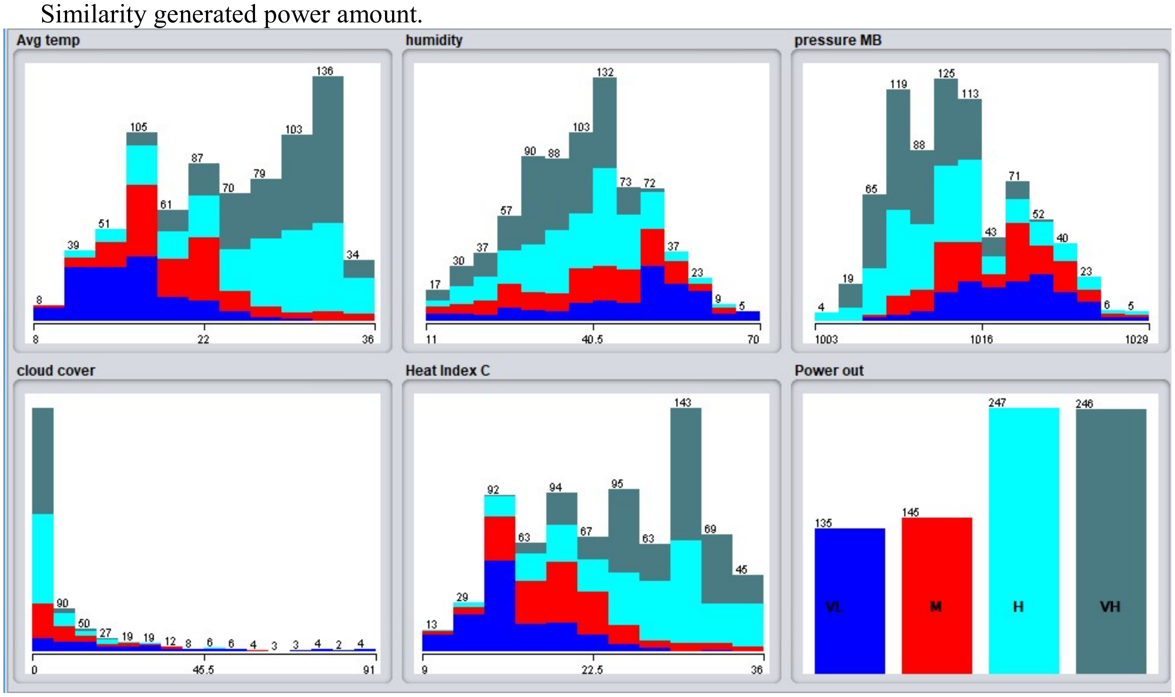

Fig. 16 Input data histogram

In this study, WEKA (Waikato Environment for Knowledge Analysis) is used to construct decision trees according to the training set with the standard algorithm J48. Weka is a popular suite of machine learning software written in Java, developed at the University of Waikato, New Zealand. Weka is a free software tool available under the GNU General Public License. It contains a collection of visualization tools and algorithms for data analysis and predictive modelling that support data pre-processing, clustering, classification, regression, visualization, and feature selection. Weka has a powerful Graphical User Interface that supports its functionality.

Once the features are extracted and grouped into a feature vector, classification takes place, where the output power values are classified in one of the following classes: Very Low, Medium, High, Very High. J48 (C4.5) is an algorithm used to generate a decision tree. It has been developed by John Ross Quinlan [33]. C4.5 is an extension of Quinlan's earlier ID3 algorithm. The decision trees generated by C4.5 is normally used for classification. It has been ranked first in the Top 10 Algorithms in Data Mining algorithm [34].

Java was selected as the development language, and J2SDK version 1.6.0_22 was used for the Java running environment. Access database 2013 is adopted for the development of database. Java database connectivity (JDBC), an application-programming interface for the Java programming language, is used to access the database. JDBC could wrap a structured query language (SQL) statement, send it to the database, and retrieve the desire.

Part of extracted data from January, 1st, 2017 to February 12, 2020 (774 records excluding 2018 in order to be used later for validation) has been used for training and testing the decision maker module. The confusion matrix for each class (Very Low, Medium, High Very High) is constructed. The confusion matrix has the form shown in Table 6.

Table 6. Confusion matrix

	Very Low	Medium	High	Very High	
Very Low	a	b	c	D	Sum_r1
Medium	e	f	g	H	Sum_r2
High	i	j	k	L	Sum_r3
Very High	m	n	o	P	Sum_r4
Sum	Sum_c1	Sum_c2	Sum_c3	Sum_c4	Total

10-fold cross validation is used in this paper. In K-fold the training set will be randomly splitted into K sets that have approximately the same size. Then the Decision Tree will be trained using (K-2) subset. One of the two remaining subsets will be used for validation and the last for testing. This process will be repeated K times, while a different subset is used for testing and validation.

The performance measurements used for this paper are recall, precision, classifier F1 rating and accuracy. They are defined as follows:

Recall (R) is the ratio of the relevant data among the retrieved. Precision (P) is the ratio of the accurate data among the retrieved data. Their formulas are given as follow:

$\boldsymbol{Recall(R)} = \frac{T_P}{T_P+F_N}$ if TP+FN > 0, otherwise undefined.

The recall of a class "Very Low" is defined as:

$$R_{Very\ Low} = \frac{a}{Sum\_C1}$$

$\boldsymbol{Precision(P)} = \frac{T_P}{T_P+F_P}$ if TP+FN > 0, otherwise undefined.

The precision of a class "Very Low" is defined as:

$$P_{Very\ Low} = \frac{a}{Sum\_r1}$$

Classifier F1 rating is the harmonic mean of the classifier recall and the precision. It is given as

$$\boldsymbol{F_1 = \frac{2 * P * R}{P + R}}$$

where R represents the recall, , and P represents the precision

Accuracy, which indicates the fraction of correctly classified samples among all the samples, obtained by:

$$\boldsymbol{Accuracy = \frac{T_P + T_N}{T_P + T_N + F_P + F_N}}$$

$$= \frac{a + f + k + P}{Total}$$

Kappa statistic is used to give the agreement between developed model and the actual PV system. A Kappa score ranges between 1 which shows full agreement and 0 which shows no agreement. Table 7 shows the resulted power output evaluation results.

Table 7. Overall performance results (training and testing set).

	Decision Tree
Correctly Classified Instances	634
Correctly Classified Percentage	82.0181 %
Incorrectly Classified Instances	139
Incorrectly Classified Percentage	17.9819 %
Kappa statistic	0.7547
Mean absolute error	0.1356
Root mean squared error	0.2604
Relative absolute error	37.1043 %
Root relative squared error	60.9173 %
Total Number of Instances	773

Another performance indicated by confusion matrix is shown in Table 8. This confusion matrix is built based on data testing.

We constructed the confusion matrix for each class (Very Low, Medium, High, Very High).

Table 8. Confusion matrix (training and testing set).

Output Power	**Real System**				
Model		Very Low	Medium	High	Very High
	Very Low	128	3	4	0
	Medium	19	112	11	3
	High	6	9	182	50
	Very High	5	8	21	212

The performance measurements result is shown in Table 9.

Table 9. Performance measurements

TP Rate	FP Rate	Precision	Recall	F-Measure	Class
0.948	0.04	0.810	0.948	0.874	very low
0.772	0.03	0.848	0.772	0.809	Medium
0.737	0.06	0.835	0.737	0.783	High
0.862	0.10	0.800	0.862	0.830	very high

There is excellent agreement if the Kappa coefficient is greater than 0.75, poor agreement for Kappa coefficient less than 0.4, and fair to good agreement for kappa coefficient between 0.40 and 0.75. In this paper, Kappa coefficient is 0.75, which shows excellent agreement.

As the number of features in database was eight features, we used PCA dimensionality reduction method, and we selected around 63% of the features (5 features) with the highest importance. The Accuracy percentage for each number of features is shown in Table 10.

Table 10. Feature selection

Number of Features	Accuracy (%)
2	64.32%
3	71.53%
4	73.75%
5	82.01%
6	82.01%
7	82.01%
8	82.01%

Using another 365 cases from January,1st, 2018 to December 31st 2018 from the original extracted databases not previously used in the training or used in cross validation. Table 11 shows 44 records of these cases as an example with both real system output level value and the model output. Fig. 17 shows a part of the learned decision tree associated with these records which is triggered to make power output decision.

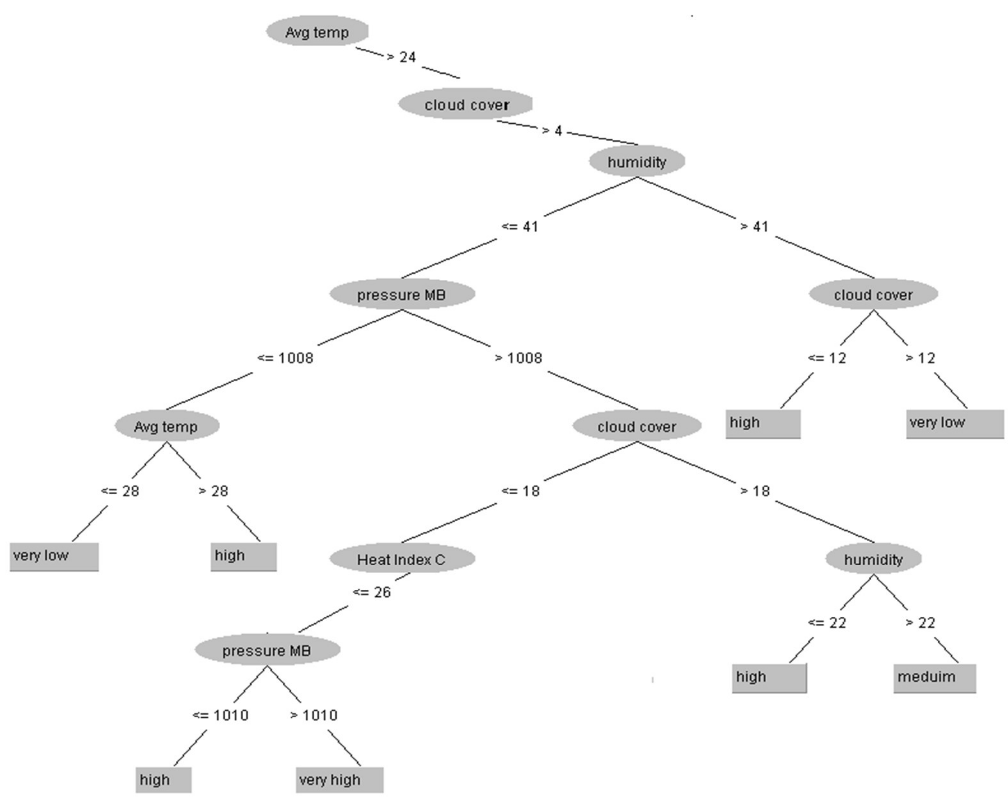

Fig. 17 Part of decision tree used in validation step

Table 11. Sample of validation data set

ID	Average Temperature	Humidity	Pressure	Cloud Cover	Heat Index	Power out (real System)	Power out (Model)
1	36	30	1006	5	36	medium	medium
2	35	26	1010	5	35	medium	medium
3	32	45	1012	5	33	medium	medium
4	26	44	1014	8	27	medium	high
5	27	24	1010	19	26	medium	medium
6	25	24	1017	25	24	medium	medium
7	25	27	1011	35	24	medium	medium
8	30	24	1008	6	29	medium	medium
9	28	41	1006	7	28	medium	medium
10	27	17	1011	15	26	medium	medium
11	34	29	1011	9	33	high	high
12	32	37	1016	5	32	high	high
13	30	48	1016	5	31	high	high
14	32	34	1015	13	31	high	high
15	31	21	1006	11	30	high	high
16	28	45	1015	7	29	high	high
17	29	43	1014	7	29	high	high
18	31	25	1008	12	29	high	high
19	28	47	1013	8	28	high	high
20	29	15	1008	42	28	high	high
21	29	25	1008	35	28	high	high
22	29	29	1008	19	28	high	high
23	29	28	1007	10	28	high	high
24	27	48	1005	6	28	high	high
25	28	29	1015	8	27	high	high
26	26	52	1014	9	27	high	high
27	26	54	1013	19	27	high	very low
28	28	22	1010	27	27	high	high
29	25	51	1015	10	26	high	high
30	27	22	1014	24	26	high	high
31	27	25	1010	33	26	high	high
32	26	18	1010	16	25	high	high
33	25	33	1009	16	25	high	high
34	25	21	1011	67	24	high	high
35	34	15	1014	6	32	high	high
36	33	26	1013	16	32	high	high
37	30	23	1012	31	29	high	high
38	26	29	1012	5	25	very high	Very high
39	25	32	1012	5	25	very high	very high
40	25	17	1012	10	24	very high	very high
41	25	31	1012	6	24	very high	very high
42	30	47	1015	14	32	very high	very high
43	28	34	1008	7	28	very high	very high
44	27	42	1008	20	27	very high	very high

The prediction result by the system is highly matched by the actual PV system generation. The achieved results are shown in Table 12 and Table 13. These coefficients suggest excellent agreement between the real system and the developed model.

Table 12. Confusion matrix (validation set)

Output Power	Real System				
Model		Very Low	Medium	High	Very High
	Very Low	88	2	0	1
	Medium	4	77	1	2
	High	1	3	68	5
	Very High	2	2	1	108

Table 13. Overall performance results (validation set)

Precision	Recall	F-Measure	Class
0.967	0.926	0.946	very low
0.916	0.916	0.916	Medium
0.883	0.971	0.924	high
0.955	0.931	0.942	very high
Correctly Classified Percentage			93.425 %
Kappa statistic			0.912
Total Number of Instances			365

V. Acknowledgment

The authors wish to express their thanks for the support of Tafila Technical University. The authors very much appreciate the support by Kawar Energy Company and Electricity Distribution Company (EDCO).

VI. Conclusion

In this paper, a new model of 1 MWp TTU PV system is analyzed using intelligent decision classifier. The developed system gives the relation between the generated output power and different weather conditions. The model of the system is built based on real output power and weather databases. The results show high accuracy and precision. In the future work different intelligent techniques will be used to increase system accuracy and precision.

References

[1] A. Al-Azzam and Amani, ‘The Grid (Opportunity vs. Challenges)’, Jordanian Forum Renew. Energy, 2014.

[2] V. Khare, S. Nema, and P. Baredar, ‘Solar-wind hybrid renewable energy system: A review’, Renew. Sustain. Energy Rev., vol. 58, pp. 23–33, 2016.

[3] V. Vittal and T. G. Heydt, ‘Impact of large scale PV generation’, pp. 1–1, 2014.

[4] R. Shah, N. Mithulananthan, A. Sode-Yome, and K. Y. Lee, ‘Impact of large-scale PV penetration on power system oscillatory stability’, IEEE PES Gen. Meet. PES 2010, pp. 1–7, 2010.

[5] P. S. Panel and P. Affairs, ‘Integrating Renewable Electricity on the Grid A Report by the APS Panel on Public Affairs’, Leadership, pp. 1–38, 2010.

[6] G. Balaban, G. C. Lazaroiu, V. Dumbrava, and C. A. Sima, ‘Analysing Renewable Energy Source Impacts on Power System National Network Code / Romania’, 2017.

[7] Obeidat, M.A. and Al-Sarayrah, R.M., “The Impact of Renewable Energy Integration on Stability of the Jordanian National Grid,” International Journal of Computer Applications, 975, p.8887, 2019.

[8] Ayman M. Mansour, Murad M. Alaqtash, Mohammad Obeidat “Intelligent Classifiers of EEG Signals for Epilepsy Detection,” WSEAS Transactions on Signal Processing, Volume 15, 2019.

[9] Jafar.Abu Khait, Ayman M Mansour and Mohammad Obeidat, “Classification based on Gaussian-kernel Support Vector Machine with Adaptive Fuzzy Inference System,” PRZEGLĄD ELEKTROTECHNICZNY. Vol 5, pp 16-24, 2018.

[10] B. Hawashin et al., "Efficient Texture Classification Using Independent Component Analysis," 2019 IEEE Jordan International Joint Conference on Electrical Engineering and Information Technology (JEEIT), Amman, Jordan, pp. 544-547, 2019.

[11] Ayman M. Mansour, “Decision Tree-Based Expert System for Adverse Drug Reaction Detection using Fuzzy Logic and Genetic Algorithm,” International Journal of Advanced Computer Research (IJACR), Vol 8(36), 2018.

[12] Mohammad A. Obeidat and Ayman M. Mansour, “EEG Based Epilepsy Diagnosis System using Reconstruction Phase Space and Naïve Bayes Classifier,” WSEAS Transactions on Circuits and Systems, Volume 17, 2018.

[13] Ayman M. Mansour, “Texture Classification using Naïve Bayes Classifier,” International Journal of Computer Science and Network Security (IJCSNS), vol. 18, no. 1, January 2018

[14] Bilal Hawashin, Ayman M. Mansour and Shadi Aljawarneh, “An Efficient Feature Selection Method for Arabic Text Classification,” International Journal of Computer Applications (IJCA), vol. 83, no. 17, pp. 1-6, December 2013.

[15] D.A. Al Nadi and Ayman Mansour, "Independent Component Analysis (ICA) for texture classification", 5th International Multi-Conference on Signals and Devices, IEEE SSD, 2008.
[16] Almonacid, F. J. M. F., Rus, C., Hontoria, L., & Munoz, F. J. "Characterisation of PV CIS module by artificial neural networks. A comparative study with other methods," Renewable Energy, 35(5), 973-980,2010.
[17] Garud, Kunal Sandip, Simon Jayaraj, and Moo-Yeon Lee. "A review on modeling of solar photovoltaic systems using artificial neural networks, fuzzy logic, genetic algorithm and hybrid models." International Journal of Energy Research, 2020.
[18] Almonacid, Florencia, et al. "Review of techniques based on artificial neural networks for the electrical characterization of concentrator photovoltaic technology." Renewable and Sustainable Energy Reviews 75 : 938-953, 2017.
[19] Rodríguez, Fermín, et al. "Predicting solar energy generation through artificial neural networks using weather forecasts for microgrid control." Renewable energy 126: 855-864,2018.
[20] Monteiro, Raul VA, et al. "Estimating photovoltaic power generation: performance analysis of artificial neural networks, Support Vector Machine and Kalman filter." Electric Power Systems Research 143:643-656, 2017.
[21] Kazem, Hussein A., Jabar H. Yousif, and Miqdam T. Chaichan. "Modeling of daily solar energy system prediction using support vector machine for Oman." International Journal of Applied Engineering Research 11.20: 10166-10172, 2016.
[22] Zendehboudi, Alireza, M. A. Baseer, and R. Saidur. "Application of support vector machine models for forecasting solar and wind energy resources: A review." Journal of cleaner production 199: 272-285, 2018.
[23] Bae, Kuk Yeol, Han Seung Jang, and Dan Keun Sung. "Hourly solar irradiance prediction based on support vector machine and its error analysis." IEEE Transactions on Power Systems 32.2: 935-945, 2016.
[24] Mojumder, Juwel Chandra, et al. "Application of support vector machine for prediction of electrical and thermal performance in PV/T system." Energy and Buildings 111: 267-277,2016.
[25] Liu, Zhao, and Ziang Zhang. "Solar forecasting by K-Nearest Neighbors method with weather classification and physical model." 2016 North American Power Symposium (NAPS). IEEE, 2016.
[26] Kartini, Unit Three, and Chao Rong Chen. "Short term forecasting of global solar irradiance by k-nearest neighbor multilayer backpropagation learning neural network algorithm." Proceedings of the International Conference on Graphics and Signal Processing. 2017.
[27] Valgaev, Oleg, Friederich Kupzog, and Hartmut Schmeck. "Building power demand forecasting using K-nearest neighbours model–practical application in Smart City Demo Aspern project." CIRED-Open Access Proceedings Journal 2017.1: 1601-1604, 2017.
[28] Mercado, Katheryn Donado, Jamer Jiménez, and M. Christian G. Quintero. "Hybrid renewable energy system based on intelligent optimization techniques." 2016 IEEE International Conference on Renewable Energy Research and Applications (ICRERA). IEEE, 2016.
[29] Ebrahimi, S. Mohammadreza, et al. "Parameters identification of PV solar cells and modules using flexible particle swarm optimization algorithm." Energy 179: 358-372, 2019.
[30] Rajan, Naik Aashay, et al. "Solar PV array reconfiguration using the concept of standard deviation and genetic algorithm." Energy Procedia 117: 1062-1069, 2017.
[31] Kumari, P. Ashwini, and P. Geethanjali. "Adaptive genetic algorithm based multi-objective optimization for photovoltaic cell design parameter extraction." Energy Procedia 117: 432-441, 2017.
[32] Ordiano, Jorge Ángel González, et al. "Photovoltaic power forecasting using simple data-driven models without weather data." Computer Science-Research and Development 32.1-2: 237-246, 2017.
[33] Quinlan JR. "C4. 5: programs for machine learning," Elsevier; 2014.
[34] Wu X, Kumar V, Quinlan JR, Ghosh J, Yang Q, Motoda H, et al. "Top 10 algorithms in data mining," Knowledge and Information Systems volume 14, pp 1–37, 2008.

Dr. Ayman M Mansour received his Ph.D. degree in Electrical Engineering from Wayne State University in 2012. Dr. Mansour received his M.Sc degree in Electrical Engineering from University of Jordan, Jordan, in 2006 and his B.Sc degree in Electrical and Electronics Engineering from University of Sharjah, UAE, in 2004. He graduated top of his class in both Bachelor and Master. Currently, Dr. Mansour is an Associate Professor in the Department of Communication and Computer Engineering, Tafila Technical University, Jordan. He is also the director of the Energy Research Center at Tafila Technical University. His areas of research include Communication Systems, Multi-agent Systems, Fuzzy Systems, Data Mining and Intelligent Systems. He conducted several researches in his area of interest. Dr. Mansour is a member of IEEE, Michigan Society of Professional Engineers, IEEE Honor Society (HKN), Society of Automotive Engineers (SAE), Tau Beta Pi Honor Society, Sigma Xi and Golden Key Honor Society.
.

Prof, Jalal M. Abdallah was born in Amman, Jordan. He received his M.S.1994 and Ph.D.1998 in electrical engineering from Vinitsa Stat Technical University, Ukraine. Since 2005, he has joined Tafila Technical University where he is currently a Professor at the electrical Engineering Department and the vice president of academic affairs. Dr. Jalal Abdallah main research interests are power system analysis, monitoring, and diagnostic of high-voltage power transformers.

DOI: 10.46300/9106.2020.14.88 Volume 14, 2020

Dr. Mohammed Obeidat Received his PhD in Electrical Engineering from Wayne State University in 2013, M.Sc degree in Electrical Engineering was from Yarmouk University , Jordan in 2006, and B.Sc degree in Electrical Engineering from Jordan University of Science & Technology, Jordan in 1999. He is an Associate professor in power and mechatronics department at Tafila Technical University. He is a member of IEEE, Tau Beta PI Honor Society and Golden Key Honor Society. He was given the honor to be a Sigma Xi member from the Board of Governor, in 2012. He demonstrated excellent research and academic abilities as well as professional potentials.

Dr. Obeidat has published several journal papers. Dr. Obeidat research of interest in the field of intelligent control systems, renewable energy, intelligent systems, and mechatronics