Evaluation of the Optimal Features and Machine Learning Algorithms for Energy Yield Forecasting of a Rural Rooftop PV Installation

Boris Evstatiev[1], Katerina Gabrovska-Evstatieva[2], Tsvetelina Kaneva[3], Nikolay Valov[4], Nicolay Mihailov[5]
Faculty of Electrical Engineering, Electronics and Automation, University of Ruse Ange Kanchev, Ruse, Bulgaria[1, 3, 4, 5]
Faculty of Natural Science and Education, University of Ruse Angel Kanchev, Ruse, Bulgaria[2]

***Abstract*—The stability and reliability of the electric grid strongly depend on the ability to schedule and forecast the energy output of all sources. Even though the share of photovoltaic installation in the energy mix is continuously increasing, they have one major drawback: their dependence on different environmental parameters, such as solar irradiance, ambient temperature, cloudiness, etc., which have a highly variable nature. Six machine learning algorithms are compared in this study, regarding their ability to forecast the power generation of a rural rooftop photovoltaic installation using different combinations of the input data. The features selected for investigation are solar radiation, ambient temperature, and wind speed, obtained from a meteorological station, as well as two additional time-based variables – the time of the day and the month of the year. During the validation and testing phases, four models performed better – artificial neural network (ANN), k-Nearest neighbor (kNN), Decision tree (DT), and Random Forest (RF), with ANN achieving the best results in all cases. The optimal combination of input data includes solar radiation, ambient temperature, wind speed, and hour of the day, though the difference with the other scenarios was small. The optimal ANN model achieved R^2, MAE, and RMSE of 0.995, 6.71 Wh, and 13.7 Wh, respectively. The results obtained in this study indicate that the yield of PV installations located in rural areas could be forecasted with high probability using a limited number of meteorological data.**

Keywords*—PV yield; forecasting; machine learning; deep learning; features; solar radiation; ambient temperature; wind speed; hour of the day

I. Introduction

Renewable energy technologies have greatly developed during the last decades, with photovoltaics holding the largest share. Many reasons exist for this, such as their low maintenance costs [1, 2], abundant availability of solar energy [3], the possibility for building integration [4], relatively easy installation, reliability [4], etc. However, one major drawback could be defined for PV installations: their strong dependence on weather conditions and especially on solar radiation, which has a highly variable nature. For this reason, the output of photovoltaic power in a stations changes wide range over time and is generally difficult to forecast. One of the options to deal with this problem is the application of energy storage systems, allowing energy charging during daylight hours and using it according to the requirements of the load profile [5, 6]. However, with the current development of energy storage technologies, this approach is still too expensive and its return on investment is too low without government incentives and subsidies [7]. Furthermore, the batteries' life expectancy is still relatively low, which requires additional investment during the PV park operation for battery replacement [8].

Reducing power uncertainties in the electric grid is a major task, which is required to ensure its energy balance. Therefore, forecasting the output of photovoltaic installations is crucial for maximizing the economic benefit and ensuring customers receive electric energy of acceptable quality and reliability [9]. The power production of PV installations depends on many environmental factors, such as solar irradiance, cloudiness, ambient temperature, wind speed [10, 11, 12], and even relative humidity and rainfall. [13, 14]. Moreover, when a photovoltaic installation is installed in an urban or suburban environment, additional factors affecting power production are introduced or enhanced, such as soiling [15, 16], shading [17, 18], panel degradation [19, 20], etc.

Different forecasting approaches exist mostly based on statistical methods (ARIMA, ARCH, GARCH) and machine learning methods, even though physical and hybrid approaches are also used [21, 22]. Machine learning has many applications in the renewable energy field, such as yield forecasting [23], condition monitoring [24], fault detection [25, 26], PV cell degradation [27], MPPT tracking [28], energy management [29], etc. When it comes to PV yield or power forecasting, different machine learning regression algorithms are used, such as Support Vector Machine (SVM), Linear regression (LR), Random Forest (RF), Regression tree (RT), etc. For example, in study [30] the global horizontal irradiance and atmospheric temperature were used, obtained from a meteorological station with a 5-minute timeframe. Several regression models were investigated, such as Gaussian process regression (GPR), LR, RT, and SVM. All models achieved similar performance, with R^2 varying between 92% and 96%, yet the RT achieved the highest score. Similarly, in the study [31] the 5 min horizontal global radiation and ambient temperature in Berlin were used as input data for a machine learning algorithm. It predicted the generated AC energy of a photovoltaic installation and achieved a coefficient of determination of 0.87.

A study for Jordan used nine features to predict the power of a PV installation: irradiation, air temperature, module temperature, as well as several time-based features – day of the week, month number, day type, week number, hour of the day and year type [32]. RF achieved the highest performance, closely followed by Bagging-REFTree. In study [33], the possibility of forecasting PV energy output in numerous regions with limited plant-specific data was investigated. The authors enriched the features by adding 1-hour-lagged meteorological data and tested different machine-learning regression methods, such as Kernel Ridge and RF. The models achieved a normalized root mean square error (NRMSE) of 3% in the case of lagged power used as input, corresponding to a 1-h time horizon.

Three types of forecasting exist when it comes to photovoltaic output: short-term, medium-term, and long-term [34]. In study [35], a predictive model for PV power generation in Korea was presented, based on a recurrent neural network (RNN) and meteorological data. Four predictive features were selected: air temperature, relative humidity, solar radiation, and wind speed. Error rates of 13.8% and 13.2% were reported, respectively for the single- Long Short-Term Memory (LSTM) and multi-LSTM models. In another study, nine input parameters were used to train an artificial neural network that forecasts the output of a photovoltaic installation – global horizontal irradiance, global diffuse radiance, ambient temperature, precipitation, wind speed, air pressure, sunshine duration, relative humidity, and surface temperature [36]. The reported error rates vary between 72.64% and 0.74% for low and high insulation values, respectively. In study [37] the 24 h PV yield was forecasted using solar radiation and ambient temperature as input data. The authors proposed an artificial neural network (ANN) Multi-Layer Perceptron (MLP) model, which achieved an R^2 between 96% and 99% for sunny days and between 0.88% and 92% for cloudy days. Similarly, in [38] ultra-short PV power forecasting was investigated based on neural networks and four features – global horizontal irradiance, wind speed, ambient temperature, and relative humidity. The different models achieved an R^2 between 0.889 and 0.967 in the validation phase and between 0.910 and 0.971 in the testing phase.

Some studies have also compared the performance of the machine learning and deep learning approaches. In [39] the temperature of the PV surface, the solar irradiance, and the wind speed were used to predict the power output of a rooftop photovoltaic installation in Russia. A comparison between ANN and regression models showed the first had better performance with error rates between 0% and 30% for the different days. Similarly, in study [40] the power of a PV installation in Egypt was forecasted using three models – RF, Facebook Prophet, and LSTM. The features used are different components of the solar irradiance, wind speed at 10 m, temperature at 2 m, and sun height. The best-performing model was Prophet with R^2=0.93, which was confirmed by its mean square error (MSE) and MAE coefficients. In study [23] were used seven machine learning algorithms for short-term prediction of photovoltaic generation - extreme gradient boosting algorithm (XGB), support vector regressor (SVR), random forest (RF), classic MLP, and three LSTM-based models. Their achieved R^2 values varied from 0.90 to 0.91 for 15-minute-ahead forecasting, from 0.88 to 0.89 for 30-minute-ahead forecasting, and from 0.86 to 0.89 for 1-hour-ahead forecasting.

However, one of the key factors for forecasting the PV yield is the availability of reliable predictions for solar radiation. This task can also be implemented using either machine or deep learning. In study [41] different machine learning algorithms were evaluated in their ability to forecast solar radiation and ambient temperature, which are considered to have the highest impact on the PV output. Similarly, in study [42] predictions of the hourly solar radiation were made for time horizons h+1 and h+6 in Odeillo, France. They were based on ANN and RF models and the past solar radiations as input data. The RF algorithm achieved better results at forecasting the global horizontal irradiation with an NRMSE of 19.65% for the h+1 timeframe and 27.78% for the h+6 timeframe.

Other studies combined the forecasting of solar energy and PV power. In study [43], the MLP ANN method to forecast the PV power was used with a 10-minute discretization step. The authors investigated two scenarios – one with measured solar irradiance data and the other with predicted one. The precision of the models was estimated using different errors. The best-performing scenario achieved a 7% error for a scenario, which relies on three days of previous solar radiation data. Similarly, in study [44] two hybrid models were investigated for PV power forecasting. A statistical model for estimating solar radiation and a physical or statistical (ANN) model for estimating output power were trained. The ANN-based model achieved lower relative root mean square error, varying from 3.59% to 8.65% for 3 days ahead forecasting and from 5.25% to 11.85% for 6 days ahead forecasting.

The analysis of previous studies shows that a wide range of input data is used for forecasting photovoltaic power. Basic meteorological data, such as solar irradiance, ambient temperature, wind speed, relative humidity, rainfall, and cloudiness is used, as well as some parameters of the PV installations, such as module temperature and yield. Additional time-based features are often added, such as day of the week, month number, day type, week number, hour of the day, year type, etc. Previous authors have reported different accuracies of the existing models, which can be explained by the influence of local factors, such as shading, soiling, etc., and the chosen features. Furthermore, there isn’t an obvious winner amongst the used approaches, such as machine learning and deep learning.

The agricultural sector has a great potential for creating additional value with the help of energy from photovoltaic installations. Such applications include powering of irrigation systems [45], animal farms [46,47], etc., and are commonly rooftop mounted. Rural areas are characterized with lack of high-rise buildings and other artificial objects, which could potentially influence the energy production of photovoltaics by creating shadings. Considering the above mentioned, it is important to investigate the possibilities for precise forecasting of the output PV power under such conditions.

This study aims to investigate the influence of different features on the performance of machine learning and deep learning models for forecasting the yield of PV installations located in rural areas.

II. Materials and Methods

A. Data Acquisition

This study relies on two data sources: a mid-scale PV park and a dedicated meteorological station. They are located in the village of Staro Selo, near the city of Tutrakan, Bulgaria, coordinates 43°59'11''N 26°32'49''E (Fig. 1).

Fig. 1. Geographic location of the experimental facility.

The photovoltaic park is installed on the roof of a building and its total power is 68.040 kWp. It is built of 378 monocrystalline modules SPV180M-24 by Sinski PV Co., Ltd. (Wuxi, Jiangsu, China). They are characterized by 180 Wp peak power, 14.1% efficiency, 45V open-circuit voltage, and 5.3 A short-circuit current under standard testing conditions. The orientation of the PV modules has an azimuth angle of 3° and an angle of inclination of 30°. The installation also contains 9 Sunny Mini Central 6000TL single-phase grid-connected inverters by SMA Solar Technology AG (Niestetal, Germany). Their key characteristics are 97.7% efficiency, 1 MPPT with four inputs, and an MPP voltage range of 333 V to 500 V. The inverters are connected to the internet via a Sunny WebBox and all data is stored on the SunnyPortal platform with a 1-hour time step.

The meteorological data is collected using a Sunny Sensor box by SMA Solar Technology AG (Niestetal, Germany), which includes:

- A solar radiation sensor with a measuring range 0÷1500 W/m^2 and an accuracy of 8%.
- A temperature sensor with a measuring range -30÷80 °C and an accuracy of 0.5°C.
- An anemometer with a measuring range of up to 40 m/s and an accuracy of 0.5%.

It is installed near the PV panels and similarly to the inverter, transmits data to the SunnyPortal platform via the Sunny WebBox with a 1-hour time step.

B. Methodology for Data Processing and Data Analysis

In this study, we have applied a data analysis methodology, which includes the following steps (Fig. 2): data preparation, feature preparation and engineering, model optimization, and features evaluation.

1) Step 1. Data preparation: In this step, the data is extracted from the cloud platform. The four datasets are exported in Microsoft Excel format and are merged into a single file with five columns – timestamp, energy yield, solar radiation, temperature, and wind speed. During the merging process, special attention should be paid to the correspondence of the timestamps of the records.

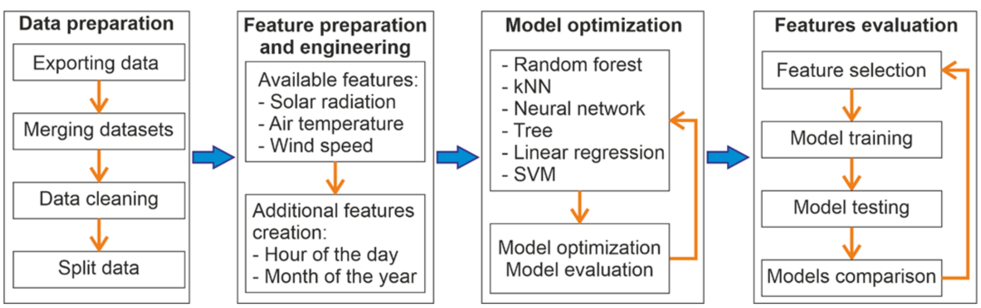


Fig. 2. Overview of the methodology used.

Next, the created dataset is analyzed for inconsistencies, such as:

- Empty or invalid values;
- The solar radiation is non-zero, while the PV yield is zero and vice-versa.

All records with such inconsistencies are removed from the dataset. Finally, the available data is divided into training/validation and testing datasets.

2) Step 2. Features preparation and engineering: In this step, the main features of the machine learning are selected. As previous authors have stated, the factors with the highest influence on the energy yield of photovoltaic installations are solar radiation, ambient temperature, and wind speed [22, 48]. Therefore, they are selected as the main features for model training. Secondary features, which are known to be correlated with solar radiation are the “month of the year” and the “hour of the day”. They are extracted from the timestamp of the datasets using Microsoft Excel’s “Month” and “Hour” functions. This way the “month of the year” feature takes values from 1 to 12 and the “hour of the day” feature takes values from 0 to 23.

3) Step 3. Model optimization: This step aims to obtain the optimal parameters of each of the selected machine-learning models, using all available features. In this study, we have chosen the Orange Data Mining v3.36 tool, developed by the University of Ljubljana (Ljubljana, Slovenia) [49]. The reason for choosing it is the wide range of available components for training regression models, evaluation and comparison, modification of the input and output data, etc.

The goal is to train regression models, which can forecast the photovoltaic energy yield using the available features, i.e. the output of the models should be the predicted energy. Based on the results of previous studies, the following machine-learning algorithms are selected for evaluation:

- *Decision tree (DT)* – builds regression organized as a tree structure.
- *Random forest (RF)* – works by creating numerous DTs during the training phase. Each tree is constructed using a random subset of the dataset to measure a random subset of features in each partition. It can be used for both classification and regression tasks. Overfitting is a common problem that may worsen the model performance, which is commonly dealt with by adding enough trees in the forest.
- *K-nearest neighbor (kNN)* – a supervised machine learning algorithm that can be used for classification and regression tasks. It requires more time and memory and is commonly useful with smaller datasets.
- *Artificial neural network (ANN)* – a set of algorithms designed for recognizing patterns in data. They are modeled after the structure and function of the human brain and have shown acceptable results in all spheres of science.
- *Support vector machine (SVM)* – a selective classifier formally defined by dividing the hyperplane. The SVM algorithm intends to find a hyperplane in an N-dimensional space that classifies the data points.
- *Linear regression (LR)* – computes the linear relationship between the dependent variable and one or more independent features by fitting a linear equation to observed data.

The output of the abovementioned regression models is additionally modified (if required), to make sure no energy production is forecasted during the dark hours of the day:

$$E_{pred.m} = \begin{cases} E_{pred};\ Solar\ radiation \geq 0 \\ 0;\ Solar\ radiation < 1 \end{cases} \quad (1)$$

During this step, the parameters of each model are changed and their performance is assessed with the help of a 5-fold cross-validation. Several statistical metrics are used that allow to evaluate the difference between the original and the predicted values:

- Coefficient of determination (R^2) - it takes values between 0 and 1 and shows how well a model predicts the outcome:

$$R^2 = 1 - \frac{\sum_i^n (y_i - \hat{y}_i)^2}{\sum_i^n (y_i - \bar{y})^2}, \quad (2)$$

where, y_i and $\hat{y}_i$ are the i^{th} samples of the actual and predicted variables and $\bar{y}$ is the mean of the actual values. It is known that when multiple regression models are evaluated, it is better to use the adjusted R^2 metric, which penalizes the additional features. However, this is true only when the number of records is relatively low. When the number of records is significantly higher than the number of features, R^2 and the adjusted R^2 have insignificant differences. That is why in this study the application of R^2 is considered appropriate.

- Mean square error (MSE) – measures the average squared difference between the actual and the predicted values with extra penalty to large errors:

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (3)$$

- Root mean square error (RMSE) – measures the average magnitude of the errors in the prediction and is the square root of MSE:

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

- Mean absolute error (MAE) – measures the average magnitude of the errors in the prediction and is useful when large errors should not be given extra penalty:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|. \quad (5)$$

- Non-zero mean absolute error (NZMAE) – measures the average magnitude of the errors in the prediction using only the non-zero records. This metric gives more accurate results, as excludes nighttime records, where no energy is generated and no error is expected.

This step is repeated numerous times with different parameters of each regression model until a peak R^2 value is achieved.

4) Step 4. Features evaluation: This step aims to evaluate the influence of the selected features on the performance of the trained models. The following variants are considered:

- Variant 1. Only solar radiation;
- Variant 2. Solar radiation and ambient temperature;
- Variant 3. Solar radiation, ambient temperature, and wind speed;
- Variant 4. Solar radiation, ambient temperature, wind speed, and hour of the day;
- Variant 5. Solar radiation, ambient temperature, wind speed, hour of the day, and month of the year.

For each of the abovementioned variants:

- The training dataset is modified to include only the corresponding features. This is implemented directly in the Orange Data Mining software, by selecting the necessary columns from the Files component.
- The six models are trained again with the selected features.

- The testing dataset is modified similarly to the training one.
- The trained models are applied to the testing dataset and the metrics from Step 3 are evaluated.

Next, the evaluated metrics are compared and the performance of each model with the different feature variants is obtained. The optimal variant and model are determined.

Other than the hourly generated energy, another important parameter of photovoltaic installations is the cumulative daily generated energy. Therefore, during this phase is also estimated the cumulative energy production for each of the testing days according to:

$$E_D = \sum_{i=0}^{23}(E_H), \tag{6}$$

where, E_H is the hourly energy production. The estimated daily energy productions can be compared to each other to identify problems where a model's predictions are dominantly above or below the actual values.

III. Results and Discussion

The datasets were prepared using data obtained in the period from 4 January 2020 to 20 December 2022. Previous studies have recommended the datasets for accurate PV forecasting to be at least 1 full year [38], therefore the used data conforms to this recommendation. Following the proposed methodology, four datasets were exported with 1 h timestep: specific yield in kWh/kWp, solar radiation in W/m^2, ambient temperature in °C, and wind speed in m/s. Thereafter, they were merged, and all records with missing, incomplete, or inconsistent data were removed. The data was split into training and testing datasets as follows:

- The training data includes 13781 records from 4 January 2020 to 31 July 2021;
- The testing data includes 3394 records from 1 August 2021 to 20 December 2021.

Next, the two additional features ("month of the year" and "hour of the day") were added to the datasets. A sample from the prepared training dataset is shown in Fig. 3. The first column (Timestamp) is not used as a feature but is kept as metadata for easier analysis. The last column contains the target variable (the specific energy yield, produced by the PV installation for 1 h), which should be forecasted.

Timestamp	Month of the year	Hour of the day	Solar rad, W/m2	Tavg, °C	Wind speed m/s	specific_yield, Wh/kWp
05.01.2020 06:00	1	6	1.86	4.8	2.89	0
05.01.2020 07:00	1	7	1.97	4.28	4.15	0
05.01.2020 08:00	1	8	1.83	3.72	5.37	0
05.01.2020 09:00	1	9	2.2	3.09	3.76	0
05.01.2020 10:00	1	10	14.8	2.63	0.56	8
05.01.2020 11:00	1	11	25.81	2.06	0.23	17
05.01.2020 12:00	1	12	93.64	2.78	0	79
05.01.2020 13:00	1	13	120.53	3.11	0.2	105
05.01.2020 14:00	1	14	183.53	4.31	1.27	158

Fig. 3. A sample from the prepared training data.

Next, the training and testing procedure were implemented in Orange Data Mining, as shown in Fig. 4. According to step 3 of the methodology, the optimal parameters of the six machine learning algorithms were obtained experimentally using all available features so that their R2 values were as close to 1 as possible. Their optimal parameters are summarized in Table I.

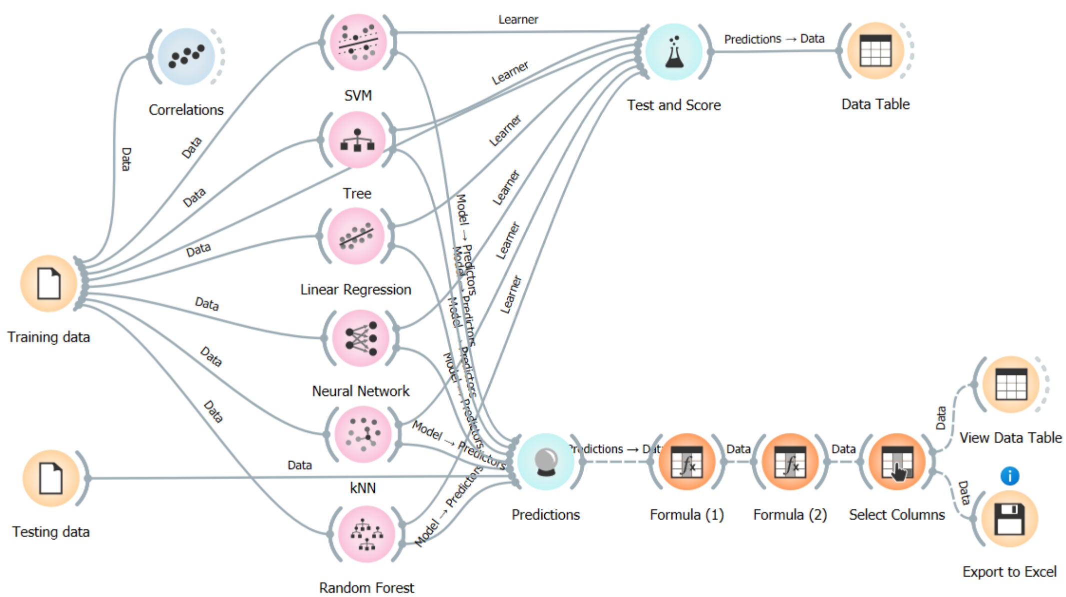


Fig. 4. Implementation of the training and testing methodology in orange data mining.

TABLE I. PARAMETERS OF THE OPTIMAL MODELS

Model	Parameters
Random forest	Number of trees: 6 Number of attributes considered at each split: not checked Replicable training: checked Balance class distribution: not checked Limit depth of individual trees: not checked Do not split subsets smaller than: not checked
K nearest neighbor	Number of neighbors: 5 Metric: Euclidean Weight: Uniform
Artificial neural network	Neurons in hidden layers: 20 Activation: ReLu Solver: L-BFGS-B Regularization, α=0: Maximal number of iterations: 300 Replicable training: checked
Linear regression	Fit intercept: checked Regularization: No regularization
Decision tree	Induce binary tree: not checked Min number of instances in leaves: 12 Do not split subsets smaller than: not checked Limit the maximal tree depth to: not checked Stop when majority reaches: 95%
Support vector machine	SVM type: SVM Cost (C): 1.40 Regression loss epsilon (ε): 0.10 Kernel: Linear Numerical tolerance: 0.0010 Iteration limit: 20 000

The model is validated using 5-fold cross-validation, which means that 20% of the records are randomly chosen for validation and the remaining 80% are used for training. The results from the models' training and validation are summarized in Table II, ordered decreasingly by their R^2 value. The best-performing algorithm is the ANN, RMSE, and MAE, with R^2 respectively 0.995, 18.3, and 7.5.

TABLE II. COMPARISON OF THE MODELS' VALIDATION PERFORMANCE

Model	MSE	RMSE	MAE	R^2
ANN	336	18.3	7.50	0.995
RF	383	19.6	6.94	0.994
kNN	390	19.8	7.31	0.994
DT	411	20.3	7.74	0.994
LR	952	30.9	18.7	0.985
SVM	1041	32.3	15.5	0.984

The order of the next three models is disputable for the following reasons:

- The RF has the second-best R^2 equal to 0.994; however, its MAE (6.94) is the lowest. In other words, if we choose the optimal model based on its MAE then the RF model performs slightly better than the trained ANN.
- The kNN model has the same R^2 as RF, and its MAE (7.31) is also lower than ANN's.
- The DT has the same R^2 as RF and kNN, and almost the same MAE (7.74).

In general ANN, RF, kNN, and DT perform almost equally well in our study. The other two models (LR and SVM) perform slightly worse, though their R^2 values are still impressive – 0.985 and 0.984, respectively. However, their MAE metrics are more than twice as bad (18.7 and 15.5, respectively), which means their forecasts contain more errors. This is also indicated by their RMSE metrics (30.9 and 32.3, respectively), which penalize large errors.

Next, according to the developed methodology, the performance of the models was evaluated for the different feature variants. For each one the corresponding features were selected from the training and testing datasets and the models were retrained and reevaluated with the testing dataset. The results from Variants 1, 2, 3, 4, and 5 are summarized in Table III, Table IV, Table V, Table VI, and Table VII, respectively.

TABLE III. RESULTS FROM THE TESTING OF VARIANT 1 (ONLY SOLAR RADIATION)

Model	MSE	RMSE	MAE	NZMAE	R^2
ANN	567	23.8	10.2	21.2	0.990
DT	638	25.3	10.8	22.6	0.989
kNN	691	26.3	11.3	23.7	0.988
RF	838	29.0	12.6	26.6	0.986
LR	985	31.1	13.8	28.8	0.983
SVM	1021	32.0	13.8	28.9	0.983

TABLE IV. RESULTS FROM THE TESTING OF VARIANT 2 (SOLAR RADIATION AND AMBIENT TEMPERATURE)

Model	MSE	RMSE	MAE	NZMAE	R^2
ANN	323	18.0	6.41	13.4	0.994
kNN	352	18.8	6.72	14.3	0.994
DT	347	18.6	6.83	14.3	0.994
RF	406	20.1	7.15	15.2	0.993
LR	930	30.0	13.5	28.3	0.984
SVM	1022	32.0	13.9	28.9	0.983

TABLE V. RESULTS FROM THE TESTING OF VARIANT 3 (SOLAR RADIATION, AMBIENT TEMPERATURE, AND WINDSPEED)

Model	MSE	RMSE	MAE	NZMAE	R^2
ANN	314	17.7	6.12	12.8	0.995
kNN	358	18.9	6.61	14.1	0.994
DT	355	18.8	6.90	14.5	0.994
RF	383	19.6	6.84	14.5	0.993
LR	930	29.9	13.60	28.4	0.984
SVM	1025	32.0	13.86	28.9	0.983

TABLE VI. RESULTS FROM THE TESTING OF VARIANT 4 (SOLAR RADIATION, AMBIENT TEMPERATURE, WINDSPEED, AND HOUR OF THE DAY)

Model	MSE	RMSE	MAE	MZMAE	R^2
ANN	299	17.3	5.71	11.9	0.995
kNN	357	18.9	6.65	14.2	0.994
DT	358	18.9	6.93	14.6	0.994
RF	372	19.3	6.56	14.1	0.993
LR	930	30.5	13.6	28.4	0.984
SVM	1022	32.0	13.9	28.9	0.983

TABLE VII. RESULTS FROM THE TESTING OF VARIANT 5 (SOLAR RADIATION, AMBIENT TEMPERATURE, WINDSPEED, HOUR OF THE DAY, AND MONTH OF THE YEAR)

Model	MSE	RMSE	MAE	NZMAE	R^2
ANN	335	18.3	6.44	13.4	0.994
kNN	367	19.2	6.65	13.7	0.994
DT	373	19.3	6.93	14.0	0.994
RF	436	20.9	6.56	14.1	0.993
LR	936	30.6	16.2	28.4	0.984
SVM	1023	32.0	14.0	28.9	0.983

If we take a look at the obtained coefficients of determination, several things can be noticed:

- The ANN models have the best performance in all five variants of the input features with R2 ranging between 0.990 and 0.995;
- The RF, kNN and DT models return very close results in all cases with R2 between 0.986 and 0.994;
- The SVM and the LR models have the worst performance in all cases, although it is not significantly worse. They are practically the same in all five variants; i.e., if these algorithms are selected, solar radiation can be used as the only feature.

All models in all variants achieved excellent coefficients of determination, ranging between 0.983 and 0.995. At first glance, the last statement means that there is not any significant difference between the six algorithms. That is why a closer look should be taken at the other metrics. For Variant 1 (Table III) the ANN model achieved an MAE of 10.2 Wh/kWh/h, which means that for the investigated testing period (3394 hours or approximately five months) the expected cumulative error is 34.6 kWh/kWp. However, if only the non-zero records are accounted for, as no error is expected during the dark hours of the day, the NZMAE metric is 21.2 Wh/kWh/h, i.e. approximately twice as high as MAE. For the worst-performing model (SVM) the MAE and NZMAE are 13.8 Wh/kWp/h and 28.0 Wh/kWhp/h, respectively, corresponding to a cumulative error of 46.837 kWh/kWp.

For Variant 2, ANN's MAE and NZMAE reach 6.41 Wh/kWp/h and 13.4 Wh/kWp/h, respectively (i.e., a cumulative error of 21.8 kWh/kWp), and for Variant 3 – 6.12 Wh/kWp/h and 12.8 Wh/kWp/h, respectively (a cumulative error of 20.8 kWh/kWp). The best performance was achieved for Variant 4, where ANN's MAE and NZMAE reached 5.71 Wh/kWp/h and 11.9 Wh/kWp/h (a cumulative error of 19.4 kWh/kWp), while for Variant 5 the score was slightly worse.

If the RMSE metric is analyzed, which adds a penalty to higher errors, once again the optimal value is achieved by the ANN model with Variant 4 – 17.3 Wh/kWp/h and the lowest by the SVM model, which is 32.0 Wh/kWp/h for all five variants.

For a better understanding of the precision of the trained models, the worst-case (Variant 1) and best-case (Variant 4) models are further compared. In Fig. 5 statistics about the total daily absolute error of the models for Variant 1 is presented. The minimal daily errors for the different models vary between 0.6 (SVM) and 4.0 (kNN) Wh/kWp/Day. The maximum daily errors vary between 523 (ANN) and 711 (LR) Wh/kWp/Day. The average daily error is the lowest for DT (169.973 Wh/kWp/Day and the highest for SVM (218.508 Wh/kWp/Day). The cumulative daily error for the investigated period is the lowest for DT (24136 Wh/kWh) and the highest for SVM (31028.2 Wh/kWh).

Similarly, in Fig. 6 the total daily absolute errors for Variant 4 (best-case) are presented. It is interesting to notice that the maximal daily errors are higher in this situation and vary between 596 (kNN) and 767 (RF) Wh/kWp/Day. Nevertheless, the cumulative errors for the investigated period are significantly lower for all algorithms except SVM and LR. The lowest cumulative error was achieved by ANN (10813 Wh/kWp) and the highest again by SVM (31045.4 Wh/kWp). Similarly, the lowest average daily error was achieved by ANN (76.82 Wh/kWp/Day) and the highest again by SVM (218.63 Wh/kWp/Day).

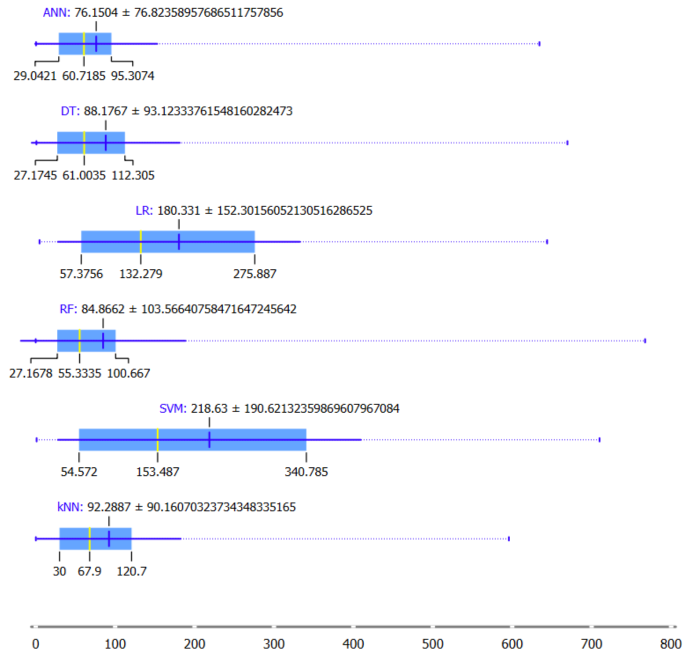


Fig. 5. Cumulative absolute daily errors of the 6 models for Variant 4 of the selected features: dark blue vertical line – mean value; thin blue – standard deviation; yellow line – the median; blue highlighted area – the values between the first and the third quartile.

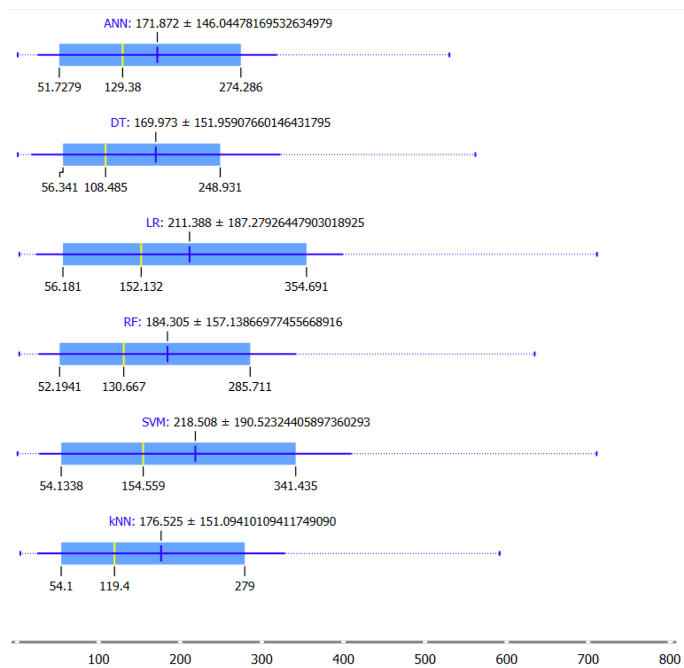
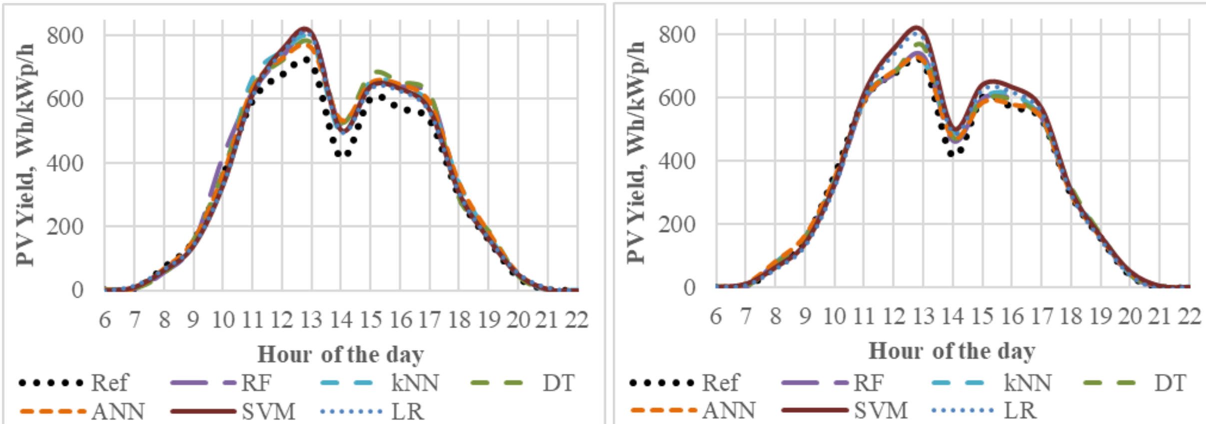


Fig. 6. Cumulative absolute daily errors of the 6 models for Variant 1 of the selected features: dark blue vertical line – mean value; thin blue – standard deviation; yellow line – the median; blue highlighted area – the values between the first and the third quartile.

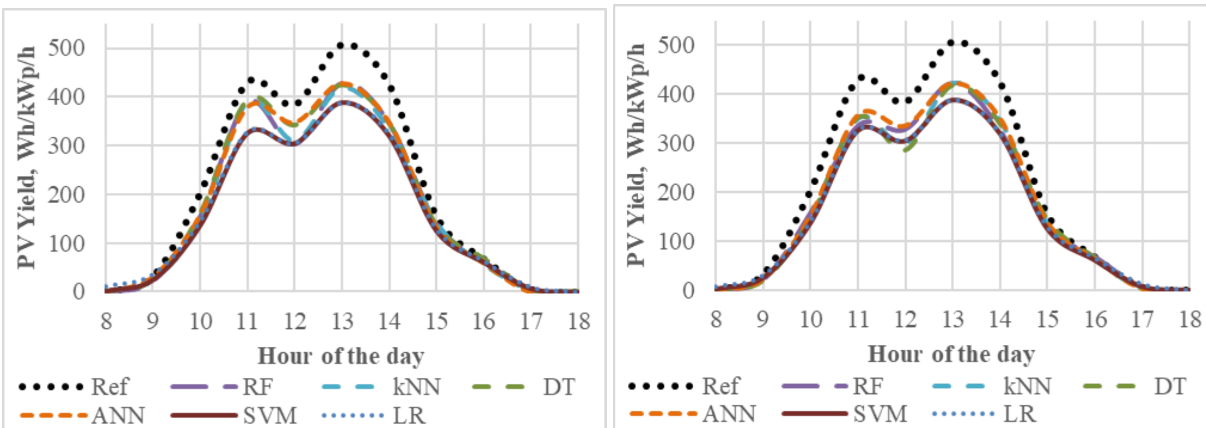
Furthermore, the following examples of the actual and predicted hourly PV yields with high errors are demonstrated:

- Example 1: Hourly data forecasts of one of the days with the worst cumulative absolute error of the ANN model in Variant 1 (4 August 2022) and the corresponding predictions in Variant 4 (Fig. 7).
- Example 2: Hourly data forecasts of one of the days with the worst cumulative absolute error of the ANN model in Variant 4 (27 November 2022) and the corresponding predictions in Variant 1 (Fig. 8).



(a) (b)

Fig. 7. Sample data from 4 August 2022 for Variant 1 (a) and Variant 4 (b) of the used features.



(a) (b)

Fig. 8. Sample data from 27 November 2022 for Variant 1 (a) and Variant 4 (b) of the used features.

Both examples show that the higher errors occur mostly on days with varying cloudiness. This behavior is expected, because of the increased errors when estimating the average hourly solar radiation introduced by the period of discretization. Nevertheless, in both situations, the obtained forecasts by the ANN model are slightly better in Variant 4, compared to Variant 1, in which only solar radiation is used as a feature. Other examples are presented in Fig. 9(a) and Fig. 9(b), where the actual and forecasted values from 25 September to 27 September 2022 are shown, representing the models from Variants 1 and Variant 4, respectively. In this case, no significant deviations are observed from the actual values and this refers to both variants of the features used. The maximum absolute difference of the ANN model from the observed values does not surpass 50 Wh/kWp/h for Variant 1 and 36 Wh/kWp/h for Variant 4.

Finally, predicted vs. actual scattered graphs were prepared for all six models with Variant 4 of the selected features, which should provide a clear understanding of their performance. They are presented in Fig. 10, where the models are ordered in the decreasing order of their coefficient of determination.

The following observations could be made:

- One anomaly with all 6 models could be noticed, most likely caused by a maintenance procedure or some technical fault with the PV installation.
- The best performance of the ANN models is also confirmed by the lowest scattering of the predicted vs. actual points [(Fig. 10(a)].
- The kNN [(Fig. 10(b)] and DT [(Fig. 10(c)] models perform almost as well as ANN; however, several points are separated slightly from the main group, which explains their lower score.
- Two of the points of the RF model [(Fig. 10(d)] are significantly separated from the main group. However, if these records are excluded from the testing dataset, the RF model could be a contender for the top spot.
- The performance of the LR [(Fig. 10(e)] and SVM [(Fig. 10(f)] models is significantly worse, and it can be noticed that their predicted vs. actual graph can be better approximated with a polynomial, rather than a straight line.

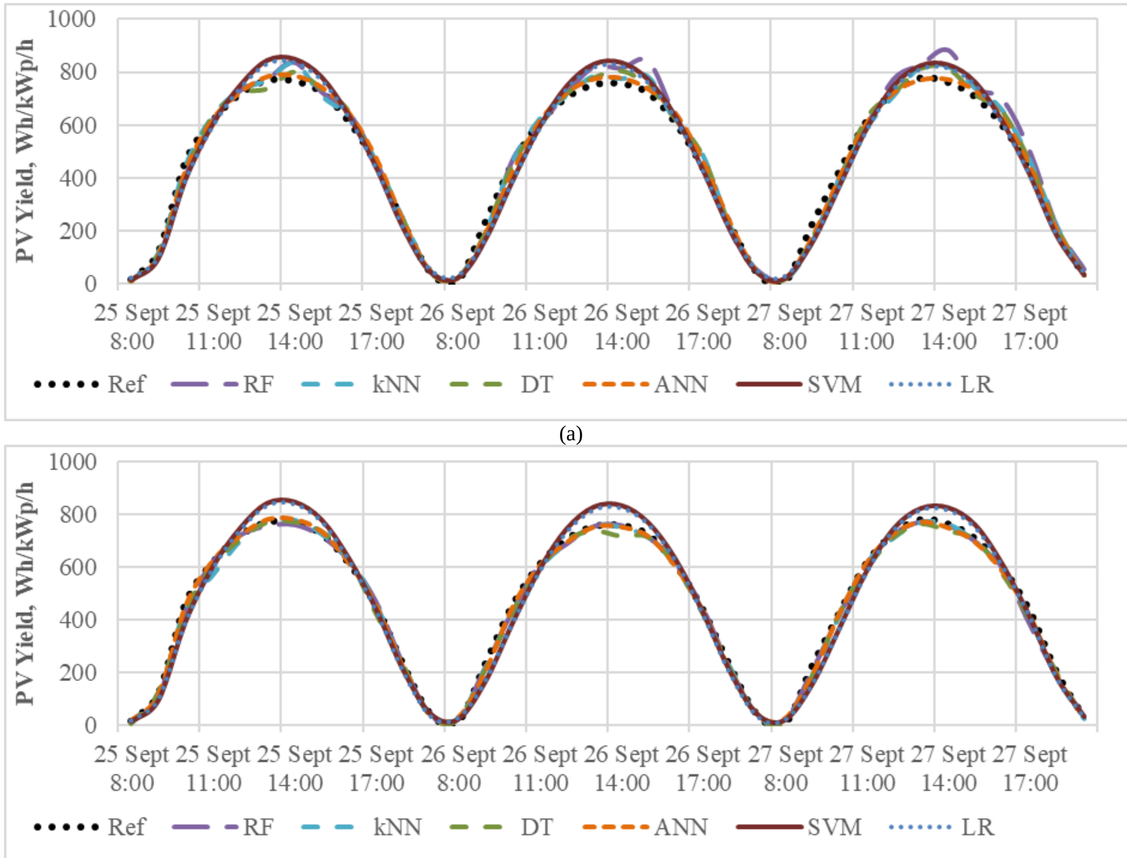


Fig. 9. Sample data from the daylight hours of 25-27 September 2022 for variant 1 (a) and variant 4 (b).

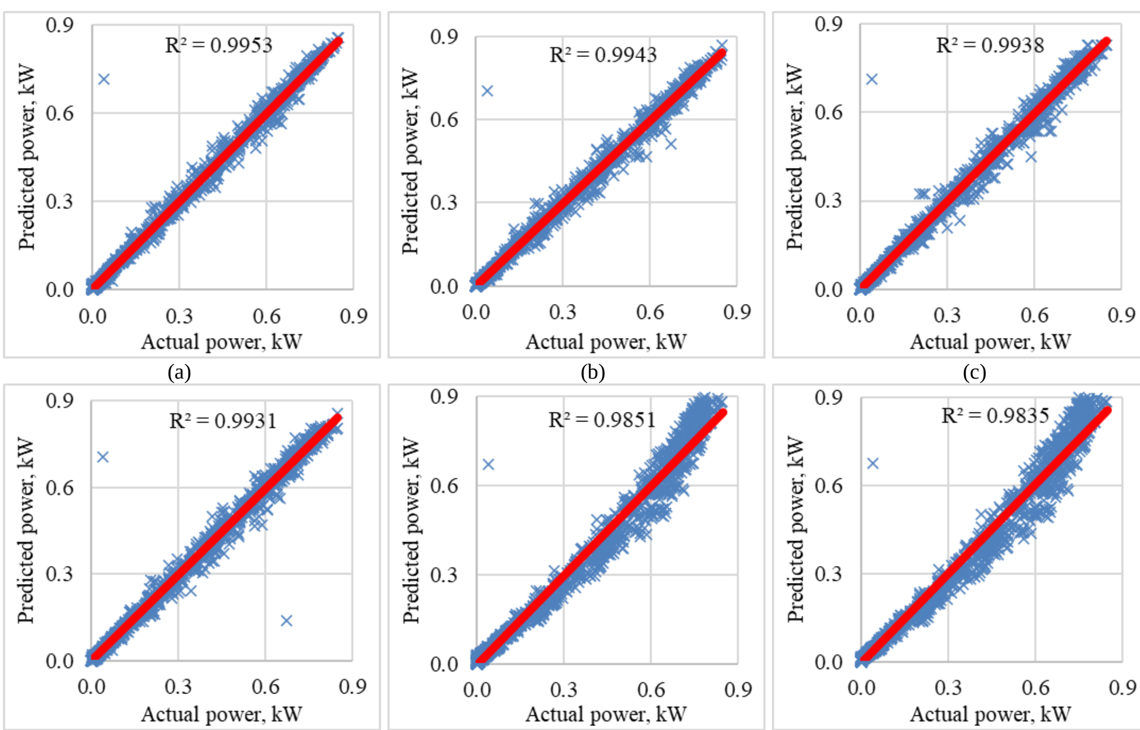


Fig. 10. Comparison between actual and predicted specific yields for the six models with variant 4 of the features: a) ANN; b) kNN; c) DT; d) RF; e) LR; f) SVM.

The performance of the trained models could be compared with that achieved in previous studies. In [32] different machine learning algorithms for forecasting the power of a PV installation were evaluated using eight features. The best-performing model was RF, which achieved an R^2 of 0.95 and an MAE of 68.7 W. Similarly, in study [30] different machine-learning models using ambient temperature and solar irradiance as features were compared. The Fine tree model achieved the highest R^2 and RMSE, 0.959 and 5.83 W, respectively.

If compared with studies relying on deep learning, our results are also ranked very well. In study [35] several meteorological parameters, the day, and time were used as features to predict the PV yield. The multiple LSTM neural network achieved an RMSE of 37.1 Wh and an error rate of 13.2%; however, no MAE and R^2 were reported. Similarly, in [40] solar irradiance, windspeed, ambient temperature, and the Sun height were used as input data to predict the PV power. The optimal model was Facebook Prophet, which achieved an R^2 of 0.93, an MAE of 8.77 W, and an RMSE of 3.28 W.

A significantly different approach was used in study [23], where the previous PV yield was used as input data for ANN models to predict the 1-hour-ahead yield. The optimal model achieved R^2, MAE, and RMSE of 0.89, 13.4 Wh, and 27.5 Wh, respectively. A similar approach in [43], where the 3 days ahead solar radiation was used, led to an MAE of 0.00 kW and a RMSE of 35.4 kW with a MLP ANN; though no info was provided about the coefficient of determination.

In study [50] was used a hybrid machine learning model, combining variational mode decomposition (VMD), whale optimization algorithm (WOA), and long short-term memory neural network (LSTM) to forecast power. The study relied on the ambient temperature, relative humidity, global and diffuse horizontal radiation to achieve an R^2 of 0.997.

The above-mentioned is summarized in Table VIII and allows us to conclude that our results position themselves very well. Out of the papers that provided a coefficient of determination, we achieved the second-best results with an R^2 of more than 99%, and were outperformed only by the hybrid model, proposed in study [50]. Similarly, the MAE we achieved is the lowest, compared to the previous studies; however, in terms of RMSE, our optimal models are ranked 3rd. The last information indicates that the models trained in this study returned several wrongly forecasted values, which differ significantly from the actual ones.

TABLE VIII. COMPARISON OF THE ACHIEVED RESULTS WITH THOSE OF PREVIOUS STUDIES

Article	Regression model	Features	Target	R^2	MAE	RMSE
Alhmoud et al [32]	RF	Irradiation, air temperature, module temperature, day of the week, month number, day type, week number, hour of the day, and year type	PV power in W	0.95	68.7 W	N/A
Zulkifly et al. [30]	Fine tree	Ambient temperature, solar irradiance	PV power in W	0.96	34.9 W	5.83 W
Park et al [35]	Multiple LSTM ANN	Ambient temperature, humidity, direct solar radiation, diffuse solar radiation, wind speed, day, and time	PV yield in Wh	N/A	N/A	37.1 Wh
Allam et al [40]	Facebook Prophet	Solar irradiance, wind speed, ambient temperature, sun height	PV power in W	0.93	8.77 W	3.28 W
Cantillo-Luna et al [23]	ConvLSTM1D ANN	Lagged PV yield	1 h ahead PV yield in Wh	0.89	13.4 Wh	27.5 Wh
Stoyanov and Draganovska [43]	MLP ANN	3 days ahead solar radiation	PV power in kW	N/A	0.00 kW	35.4 kW
Hou et al [50]	A hybrid VMD, WOA and LSTM model	Ambient temperature, relative humidity, global and diffuse horizontal radiation	PV power in kW	0.997	15.247	19.753 kW
Ours	ANN kNN DT RF	Solar radiation, wind speed, ambient temperature	Specific PV yield in Wh	0.995 0.994 0.994 0.993	5.71 Wh 6.65 Wh 6.93 Wh 6.56 Wh	17.3 Wh 18.9 Wh 18.9 Wh 19.3 Wh

IV. CONCLUSIONS

The performance of different machine learning algorithms (ANN, kNN, DT, RF, LR, and SVM) for forecasting the yield of a rural photovoltaic installation was evaluated in this study. An almost complete hourly dataset from 2020 and 2021 was used and divided into training/validation and testing datasets. Five combinations of the input features (solar radiation, ambient temperature, wind speed, hour of the day, and month of the year) were evaluated.

During the 5-fold cross-validation step the ANN achieved the highest R^2 (0.995), closely followed by RF, kNN, and DT (0.994). LR and SVM returned a lower coefficient of determination (0.985 and 9.984, respectively), though it is not significantly lower. During the testing stage, the worst results were achieved with solar radiation as the only feature, and the best results with solar radiation, ambient temperature, wind speed, and hour of the day. In all cases, the ANN model had the highest performance in terms of R^2, MAE, RMSE, and NZMAE, though once again it was very closely followed by kNN, DT, and RF.

The obtained results allow us to conclude that when a PV installation is located in a rural or ruruban area, which is characterized by a lack of significant shadings influencing its operation:

- the optimal combination of features for forecasting the output power is solar radiation, ambient temperature, wind speed, and hour of the day;
- the optimal models are ANN, kNN, DT, and RF;
- in case of limited availability of meteorological data, it is acceptable (in terms of forecasting errors) to use solar

radiation and ambient temperature or only solar radiation data as features.

The results obtained in this study could be useful to energy experts and farm owners, who are trying to maximize their profit and added value. However, it should not be forgotten that with such an approach the models also need reliable input data, such as solar radiation, ambient temperature, and wind speed. Therefore, it is also important to investigate the influence of the forecasted feature errors on the precision of the trained models, i.e. if a certain error is added to the meteorological data, what absolute and relative difference will it create. Moreover, in the present paper, we accepted that the PV installation produces only active power, which is not always the case. The presence of reactive consumers in the industry might be a significant problem when PV installations produce only active power and require a thorough investigation. The abovementioned problems were not addressed in this study and are an object for future research.

ACKNOWLEDGMENT

This research is financed by the Bulgarian National Science Fund under Project КП-06-Н77/2 “Research and optimization of hybrid system with renewable energy sources for power supply of livestock farm”.

This research is supported by the European Union-NextGenerationEU, through the National Recovery and Resilience Plan of the Republic of Bulgaria, project № BG-RRP-2.013-0001-C01.

REFERENCES

[1] K. Manohar, R. Ramkissoon, A. Adeyanju, “Cost benefit analysis of implementing a solar photovoltaic system,” International Journal of Innovative Research in Science, Engineering and Technology, vol. 4, no. 12, pp. 1-8, 2015, doi: https://doi.org/10.15680/IJIRSET.2015.0412006

[2] Y. Wang, R. Das, G. Putrus, R. Kotter, “Economic evaluation of photovoltaic and energy storage technologies for future domestic energy systems – A case study of the UK,” Energy, vol. 203, 2020, doi: https://doi.org/10.1016/j.energy.2020.117826

[3] K. Natesan, C.K. Nagaraj, N.K. Chandran, “Studies on improvement of solar PV panel performance,” Journal of Chemical Technology and Metallurgy, vol. 58, no. 6, pp. 1065-1070, 2023, doi: http://dx.doi.org/10.59957/jctm.v58i6.145

[4] L.P. Panagoda, R.A. Sandeepa, W.A. Perera, D.M. Sandunika, S.M. Siriwardhana, M.K. Alwis, S.H. Dilka, “Advancements in Photovoltaic (PV) Technology for Solar Energy Generation,” Journal of Research Technology & Engineering, vol. 4, no. 30, pp. 30-72, 2023, https://www.jrte.org/wp-content/uploads/2023/07/Advancements-In-Photovoltaic-Pv-Technology-for-Solar-Energy-Generation.pdf

[5] D. Rekioua, “Energy Storage Systems for Photovoltaic and Wind Systems: A Review,” Energies, vol. 16, no. 9, 2023, doi: https://doi.org/10.3390/en16093893

[6] A.S. Hassan, L. Cipcigan, N. Jenkins, “Optimal battery storage operation for PV systems with tariff incentives,” Appl. Energy, vol. 203, pp. 422–441, 2017, doi: https://doi.org/10.1016/j.apenergy.2017.06.043

[7] G.G. Zanvettor, M. Casini, A. Vicino, “Optimal Operation of Energy Storage Facilities in Incentive-Based Energy Communities,” Energies, vol. 17, no. 11, 2024, doi: https://doi.org/10.3390/en17112589

[8] H. Beltran, P. Ayuso, E. Pérez, “Lifetime Expectancy of Li-Ion Batteries used for Residential Solar Storage,” Energies, vol. 13, no. 3, 2020, doi: https://doi.org/10.3390/en13030568

[9] K.J. Iheanetu, “Solar Photovoltaic Power Forecasting: A Review,” Sustainability, vol. 14, no. 24, 2022, doi: https://doi.org/10.3390/su142417005

[10] A. Rhouma, Y. Said, “Solar Energy Forecasting Based on Complex Valued Auto-encoder and Recurrent Neural Network,” International Journal of Advanced Computer Science and Applications(IJACSA), vol. 14, no 4, 2023, doi: https://dx.doi.org/10.14569/IJACSA.2023.0140443

[11] Z.R. Tahir, A. Kanwal, M. Asim, M. Bilal, M. Abdullah, S. Saleem, M.A. Mujtaba, I. Veza, M. Mousa, M.A. Kalam, “Effect of Temperature and Wind Speed on Efficiency of Five Photovoltaic Module Technologies for Different Climatic Zones,” Sustainability, vol. 14, no. 23, 2022, doi: https://doi.org/10.3390/su142315810

[12] M.H. Alomari, O. Younis, S. Hayajneh, “A Predictive Model for Solar Photovoltaic Power using the Levenberg-Marquardt and Bayesian Regularization Algorithms and Real-Time Weather Data,” International Journal of Advanced Computer Science and Applications (IJACSA), vol. 9, no.1, 2018, doi: http://dx.doi.org/10.14569/IJACSA.2018.090148

[13] H.A. Kazem, M.T. Chaichan, I.M. Al-Shezawi, H.S. Al-Saidi, H.S. Al-Rubkhi, K. Alsinani, A.H. Al-Waeli, “Effect of Humidity on the PV Performance in Oman,” Asian Transactions on Engineering, vol. 2, no. 4, pp. 29–32, 2012.

[14] D.S. Hasan, M.S. Farhan, H. Alrikabi, “Impact of Cloud, Rain, Humidity, and Wind Velocity on PV Panel Performance,” Wasit Journal of Engineering Sciences, vol. 10, no. 2, pp. 34-43, 2022, doi: https://doi.org/10.31185/ejuow.Vol10.Iss2.237

[15] B. Stankov, A. Terziev, M. Vassilev, M. Ivanov, “Influence of Wind and Rainfall on the Performance of a Photovoltaic Module in a Dusty Environment,” Energies, vol. 17, no. 14, 2024, doi: https://doi.org/10.3390/en17143394

[16] M.R. Maghami, H. Hizam, C. Gomes, M.A. Radzi. M.I. Rezadad, S. Hajighorbani, “Power loss due to soiling on solar panel: A review,” Renewable and Sustainable Energy Reviews, vol. 59, pp. 1307-1316, 2016, doi: https://doi.org/10.1016/j.rser.2016.01.044

[17] P. dos Santos Vicente, E.M. Vicente, M.G. Simoes, E.R. Ribeiro, “Shading position effects on photovoltaic panel output power,” International Transactions on Electrical Energy Systems, vol. 30, no. 1, e12163, 2020, doi: https://doi.org/10.1002/2050-7038.12163

[18] D. Wang, T. Qi, Y. Liu, Y. Wang, J. Fan, Y. Wang, H. Du, “A method for evaluating both shading and power generation effects of rooftop solar PV panels for different climate zones of China,” Solar Energy, vol. 205, pp. 432-445, 2020, doi: https://doi.org/10.1016/j.solener.2020.05.009

[19] H. Yang, J. Chang, H. Wang, D. Song, “Power degradation caused by snail trails in urban photovoltaic energy systems,” Energy Procedia, vol. 88, pp. 422-428, 2016, doi: https://doi.org/10.1016/j.egypro.2016.06.018

[20] R.A.G. Burbano, G. Petrone, P. Manganiello, “Early Detection of Photovoltaic Panel Degradation through Artificial Neural Network,” Applied Sciences, vol. 11, no. 19, 2021, doi: https://doi.org/10.3390/app11198943

[21] S. Baratsas, F. Iseri, E.N. Pistikopoulos, “A hybrid statistical and machine learning based forecasting framework for the energy sector,” Computers & Chemical Engineering, vol. 188, 2024, doi: https://doi.org/10.1016/j.compchemeng.2024.108740

[22] A. Gholami, M. Ameri, M. Zandi, R.G. Ghoachani, S.J. Gerashi, H.A. Kazem, A.H. Al-Waeli, “Impact of harsh weather conditions on solar photovoltaic cell temperature: Experimental analysis and thermal-optical modeling,” Solar Energy, vol. 252, pp. 176-194, 2023, doi: https://doi.org/10.1016/j.solener.2023.01.039

[23] S. Cantillo-Luna, R. Moreno-Chuquen, D. Celeita, G. Anders, “Deep and Machine Learning Models to Forecast Photovoltaic Power Generation,” Energies, vol. 16, no. 10, 2023, doi: https://doi.org/10.3390/en16104097

[24] T. Berghout, M. Benbouzid, T. Bentrcia, X. Ma, S. Djurović, L.-H. Mouss, “Machine Learning-Based Condition Monitoring for PV Systems: State of the Art and Future Prospects,” Energies, vol. 14, no. 19, 2021, doi: https://doi.org/10.3390/en14196316

[25] E. Garoudja, A. Chouder, K. Kara, S. Silvestre, “An enhanced machine learning based approach for failures detection and diagnosis of PV systems,” Energy Convers. Manag., vol. 151, pp. 496–513, 2017, doi: http://doi.org/10.1016/j.enconman.2017.09.019

[26] A. Eskandari, J. Milimonfared, M. Aghaei, “Line-line fault detection and classification for photovoltaic systems using ensemble learning model

based on I-V characteristics," Sol. Energy, vol. 211, pp. 354–365, 2020, doi: http://doi.org/10.1016/j.solener.2020.09.071

[27] M. Dhimish, "Defining the best-fit machine learning classifier to early diagnose photovoltaic solar cells hot-spots," Case Stud. Therm. Eng., vol. 25, 2021, doi: http://doi.org/10.1016/j.csite.2021.100980

[28] K.Y. Yap, C.R. Sarimuthu, J.M. Lim, "Artificial intelligence based MPPT techniques for solar power system: A review," Journal of Modern Power Systems and Clean Energy, vol. 8, no. 6, pp. 1043-1059, 2020, doi: https://doi.org/10.35833/MPCE.2020.000159

[29] M. Gautam, S. Raviteja, R. Mahalakshmi, "Household energy management model to maximize solar power utilization using machine learning," Procedia Computer science, vol. 165, pp. 90-96, 2019, doi: https://doi.org/10.1016/j.procs.2020.01.075

[30] Z.A. Zulkifly, K.A. Baharin, C.K. Gan, "Improved machine learning model selection techniques for solar energy forecasting applications," International Journal of Renewable Energy Research (IJRER), vol. 11, no. 1, pp. 308-319, 2021, doi: https://doi.org/10.20508/ijrer.v11i1.11772.g8135

[31] S. Wendlandt, F. Popescu, "Photovoltaic Energy Yield Prediction Using an Irradiance Forecast Model Based On Machine Learning For Decentralized Energy Systems," The European Photovoltaic Solar Energy Conference and Exhibition (EU PVSEC), France, pp. 1860–1864, 2019, doi: https://doi.org/10.4229/EUPVSEC20192019-6CV.1.6

[32] L. Alhmoud, A.M. Al-Zoubi, I. Aljarah, "Solar PV power forecasting at Yarmouk University using machine learning techniques," Open Engineering, vol. 12, pp. 1078–1088, 2022, doi: https://doi.org/10.1515/eng-2022-0386

[33] M. Tucci, A. Piazzi, D. Thomopulos, "Machine Learning Models for Regional Photovoltaic Power Generation Forecasting with Limited Plant-Specific Data," Energies, vol. 17, no. 10, 2024, doi: https://doi.org/10.3390/en17102346

[34] S.M. Babbar, C.Y. Lau, K.F. Thang, "Long Term Solar Power Generation Prediction using Adaboost as a Hybrid of Linear and Non-linear Machine Learning Model," International Journal of Advanced Computer Science and Applications(IJACSA), vol. 12, no. 11, 2021, doi: https://dx.doi.org/10.14569/IJACSA.2021.0121161

[35] M.K. Park, J. Lee, J.; W.H. Kang, J.M. Choi, K.H. Lee, "Predictive model for PV power generation using RNN (LSTM)," Journal of Mechanical Science and Technology, vol. 35, no. 2, pp. 795-803, 2021, doi: http://doi.org/10.1007/s12206-021-0140-0

[36] K.R. Kumar, M.S. Kalavathi, "Artificial intelligence based forecast models for predicting solar power generation," Materials today: proceedings, vol. 5, no. 1, pp. 796-802, 2018, doi: https://doi.org/10.1016/j.matpr.2017.11.149

[37] A. Mellit, A.M. Pavan, "A 24-h forecast of solar irradiance using artificial neural network: Application for performance prediction of a grid-connected PV plant at Trieste, Italy," Solar energy, vol. 84, no. 5, pp. 807-821, 2010, doi: https://doi.org/10.1016/j.solener.2010.02.006

[38] M.A. Hassan, N. Bailek, K. Bouchouicha, S.C. Nwokolo, "Ultra-short-term exogenous forecasting of photovoltaic power production using genetically optimized non-linear auto-regressive recurrent neural networks," Renewable Energy, vol. 171, pp. 191-209, 2021, doi: https://doi.org/10.1016/j.renene.2021.02.103

[39] D.V. Tai, "Solar photovoltaic power output forecasting using machine learning technique," Journal of Physics: Conference Series, vol. 1327, 2019, doi: https://doi.org/10.1088/1742-6596/1327/1/012051

[40] G.H. Allam, B.E. Elnaghi, M.N. Abdelwahab, R.H. Mohammed, "Using Machine Learning to forecast Solar Power in Ismailia," International Journal of Scientific and Research Publications (IJSRP), vol. 11, no. 12, pp. 238-44, 2021, doi: http://doi.org/10.29322/IJSRP.11.12.2021.p12033

[41] W. Tercha, S.A. Tadjer, F. Chekired, L. Canale, "Machine Learning-Based Forecasting of Temperature and Solar Irradiance for Photovoltaic Systems," Energies, vol. 17, no. 5, 2024, doi: https://doi.org/10.3390/en17051124

[42] L. Benali, G. Notton, A. Fouilloy, C. Voyant, R. Dizene, "Solar radiation forecasting using artificial neural network and random forest methods: Application to normal beam, horizontal diffuse and global components," Renewable energy, vol. 132, pp. 871-884, 2019, doi: https://doi.org/10.1016/j.renene.2018.08.044

[43] L. Stoyanov, I. Draganovsk, "Application of ANN for forecasting of PV plant output power–Case study Oryahovo," 2021 17th Conference on Electrical Machines, Drives and Power Systems (ELMA), Sofia, Bulgaria, 2021, doi: https://doi.org/10.1109/ELMA52514.2021.9503087

[44] L. Stoyanov, I. Draganovska, "Comparison of hybrid models for pv power output forecasting—application to Oryahovo, Bulgaria," 2023 18th Conference on Electrical Machines, Drives and Power Systems (ELMA), Varna, Bulgaria, 2023, doi: https://doi.org/10.1109/ELMA58392.2023.10202257

[45] A.M. García, R.G. Perea, E.C. Poyato, P.M. Barrios, J.R. Díaz, "Comprehensive sizing methodology of smart photovoltaic irrigation systems," Agricultural Water Management, vol. 229, 105888, 2020, doi: https://doi.org/10.1016/j.agwat.2019.105888

[46] A.S. Maia, E. de Andrade Culhari, V.D. Fonsêca, H.F. Milan, K.G. Gebremedhin, "Photovoltaic panels as shading resources for livestock," Journal of Cleaner Production. vol. 258, 120551, 2020, doi: https://doi.org/10.1016/j.jclepro.2020.120551

[47] A. M. Bartkowiak, "Energy-saving and low-emission livestock buildings in the concept of a smart farming," Journal of water and land development, vol. 51, 2021, pp. 272-278, doi: https://doi.org/10.24425/jwld.2021.139935

[48] J.K. Kaldellis, M. Kapsali, K.A. Kavadias, "Temperature and wind speed impact on the efficiency of PV installations. Experience obtained from outdoor measurements in Greece," Renewable energy, vol. 66, pp. 612-624, 2014, doi: https://doi.org/10.1016/j.renene.2013.12.041

[49] J. Demsar, T. Curk, A. Erjavec, C. Gorup, T. Hocevar, M. Milutinovic, M. Mozina, M. Polajnar, M. Toplak, A. Staric, M. Stajdohar, L. Umek, L. Zagar, J. Zbontar, M. Zitnik, B. Zupan, "Orange: Data Mining Toolbox in Python," Journal of Machine Learning Research, vol. 14, pp. 2349−2353, 2013.

[50] Z. Hou, Y. Zhang, Q. Liu, X. Ye, "A hybrid machine learning forecasting model for photovoltaic power," Energy Reports, vol. 11, 2024, pp. 5125-5138, doi: https://doi.org/10.1016/j.egyr.2024.04.065