DOI 10.24425/aee.2025.153900

Prediction of photovoltaic energy generation using recurrent and transformer neural networks

KRZYSZTOF SIWEK ✉, STANISŁAW ŚWIDER

Institute of Theory of Electrical Engineering, Measurement and Information Systems
Faculty of Electrical Engineering, Warsaw University of Technology
Koszykowa 75, 00-662 Warsaw

e-mail: {✉ krzysztof.siwek/stanislaw.swider.stud}@pw.edu.pl

(Received: 22.09.2024, revised: 17.04.2025)

Abstract: Precise prediction of photovoltaic (PV) energy generation is essential for optimal, profitable and ecological management of electric energy resources all over the world. As a result, attempts are being made to develop more accurate prediction algorithms. This paper compares the application of Long Short-Term Memory (LSTM, a subtype of Recurrent Neural Networks), PatchTST (a type of Transformer Neural Network – TNN) and ensemble models (making use of these two approaches) for estimating PV energy production 24 hours ahead. The results indicate that both analysed single methods have comparable prediction accuracy, though the hybrid approach outperforms them. The experiments were conducted on data from PV sites deployed across campuses at Australian La Trobe University. However, future studies could verify this approach using different datasets. Algorithms and results presented in this study may especially contribute to the development of Recurrent and Transformer Neural Networks as prediction methods of PV energy production.

Key words: LSTM, PatchTST, photovoltaic (PV) energy, prediction, Recurrent Neural Networks, Transformer Neural Networks (TNN)

1. Introduction

The demand for electric energy continues to rise globally, driven by industrialization, urbanization, and an increasing population. Simultaneously, traditional energy production methods face significant challenges, including the depletion of fossil fuel reserves and the adverse environmental effects of carbon emissions. These challenges have necessitated a transition to renewable energy sources, among which solar photovoltaic (PV) technology has emerged as a leading contender due to its renewability and abundance [1].

However, the effectiveness of PV energy systems is intrinsically linked to environmental conditions such as sunlight availability, weather patterns, and geographic factors [2]. These dependencies introduce variability and uncertainty in energy production, which can impede effective energy management and grid stability. To address these challenges, accurate prediction of PV energy generation has become a critical area of research, enabling optimized resource allocation, improved grid integration, and better planning for renewable energy adoption [3].

Machine learning techniques have revolutionized PV forecasting by providing robust models capable of capturing complex temporal patterns in energy production. These techniques range from classical regression methods to advanced neural network architectures like Long Short-Term Memory (LSTM) networks and Transformer-based models. Despite the advancements, challenges remain, such as improving prediction accuracy across diverse conditions and integrating hybrid approaches for enhanced performance.

While LSTM is indeed a more established technique, it remains a widely recognized and validated approach in time series forecasting. Its ability to effectively capture long-term dependencies in sequential data has ensured its continued relevance across various domains, including energy forecasting. Notably, recent studies still leverage LSTM, either independently or in hybrid configurations, to address complex prediction tasks, as it provides a reliable baseline for comparison and practical applications. Its inclusion in this study not only aligns with ongoing research trends but also highlights its complementary strengths when combined with more modern Transformer-based models.

Transformer Neural Networks (TNNs), while relatively newer, have rapidly gained prominence due to their ability to capture complex temporal dependencies and model long-range interactions in time series data. Their flexibility and effectiveness have led to numerous advancements in fields such as energy forecasting, where hybrid models and Transformer variants are increasingly applied to improve prediction accuracy. The growing amount of research on TNN-based solutions underscores their potential to complement traditional approaches like LSTM, particularly in scenarios with dynamic and highly variable datasets.

This paper aims to contribute to this ongoing effort by comparing LSTM and PatchTST (a Transformer-based model) and exploring their ensemble configurations for 24-hour-ahead PV energy predictions.

2. Literature review

Photovoltaic energy has plenty of advantages such as affluence and renewability [1]. Unfortunately, the main drawback of photovoltaics is its dependence on the insolation level and consequently on the time of day and the weather conditions [2]. Thus, precise prediction of photovoltaic (PV) energy generation is crucial for optimal, profitable and ecological management of electric energy resources all over the world. In order to satisfy these requirements, it is necessary to develop the software which can accurately estimate the energy production of photovoltaic cells in different circumstances [3].

Having analysed the review of PV generation forecasting methods in [4], one can notice that in recent years scientists have developed or examined a variety of such algorithms based on a wide range of machine learning techniques. In paper [5] authors describe different PV prediction models which use Lasso Regression, K–Nearest Neighbours Regression, Support Vector Regression (SVR), AdaBoosted Regression Tree, Gradient Boosted Regression Tree, Random Forest Regression and Artificial Neural Network. Research [6] presents the use of Kalman Filter to estimate PV generation. In article [7] there is a study on forecasting solar energy generation under soiling conditions using Linear Regression and Neural Networks. There are also many papers that analyse hybrid PV prediction models such as: [8] combining adaptive k-means and GRU to generate short-term estimations supported by weather data; [9] introducing a parallel architecture method based on the Seasonal Auto-Regressive Integrated Moving Average (SARIMA) model and an Artificial Neural Network (ANN) model with weighing factors computed periodically as a least squares problem, or [10] describing a day-ahead prediction method using a hybrid-classification-regression forecasting engine and a feature selection/clustering based on criteria of relevancy and redundancy.

Despite the fact that LSTM [11] is almost 30-year-old approach, there are still many research works which incorporate it in solar energy generation forecasting models either as single model or as a component of an ensemble model [12], for instance [13] comparing Multi-layer Perceptron (MLP) and simplified Long Short-Term Memory (LSTM) for short-term estimations; [14] proposing a hybrid deep learning framework based on Convolutional Neural Network (CNN) and LSTM; [15] comparing different PV prediction algorithms, including the one combining LSTM, SVR and Bayesian optimization; [16] analyzing the prediction accuracy of five layer CNN-LSTM model in comparison with Lasso and Ridge regression, LSTM and less complex CNN-LSTM or [17] proposing four LSTM based models for multivariate prediction of PV energy generation.

The increasing adoption of Transformer-based architectures [18] has allowed for substantial improvements in forecasting accuracy. For instance, the Transformer variant proposed in [19] integrates numerical weather prediction data with site-specific physical parameters to improve day-ahead forecasting accuracy. Their model, PTFNet, effectively extracts both temporal dependencies and inter-feature relationships, outperforming other state-of-the-art techniques in terms of error metrics such as RMSE and MAPE. Similarly, [20] introduced the Graph Patch Informer (GPI) model, a Transformer-based approach that leverages segment-wise self-attention and graph attention networks to enhance the extraction of temporal dependencies.

Hybrid Transformer models that incorporate recurrent neural networks (RNN) have also been explored to further optimize forecasting performance. Kim et al. proposed the PVTransNet model, which combines Long Short-Term Memory (LSTM) with Transformer encoders to address day-ahead forecasting challenges. By integrating historical PV generation data with weather forecasts and solar geometry information, PVTransNet achieved notable improvements in prediction accuracy, surpassing standalone LSTM and Transformer architectures [21]. Another study [22] compared vanilla Transformer, Informer, and Autoformer models for short-term PV power forecasting in a centralized solar plant. The vanilla Transformer demonstrated superior predictive performance, particularly under scenarios requiring longer forecasting horizons, such as 24-hour predictions.

To address the issue of weather uncertainty, models such as CT-NET [23] employ a hybrid architecture that combines convolutional neural networks (CNN) with multi-head attention mechanisms. This architecture enables the model to extract both local and global features, mitigating the impact of fluctuating weather conditions on PV power generation forecasts. The

CT-NET model achieved lower error rates and reduced computational complexity, making it suitable for real-time applications in smart grids. Similarly, paper [24] utilized the Temporal Fusion Transformer (TFT) for day-ahead PV forecasting. The TFT model integrates interpretable attention-based mechanisms, outperforming traditional methods such as ARIMA, LSTM, and XGBoost in terms of accuracy while offering insights into the temporal dynamics of PV generation.

Despite the improvements brought by Transformer-based models, challenges remain, particularly regarding data availability and computational efficiency. Transfer learning techniques have been proposed as a solution to overcome data scarcity. Paper [25] demonstrated how pre-trained Transformer models could be fine-tuned using limited data from individual households to improve short-term energy demand and PV production forecasts. Their approach resulted in a significant reduction in prediction errors, highlighting the effectiveness of transfer learning for localized forecasting tasks.

Paper [26] describes a CNN-CosAttention-Transformer (CA-Transformer) model based on the Copula function as a method for short-term photovoltaic power generation estimation. In [27] one can find a comparison of models based on several Recurrent Neural Networks (in particular LSTM and GRU) with Transformer models in terms of solar power generation forecasting precision in two types of solar systems (non-transparent and transparent panels). Article [28] proposes the transformer model for predicting ultra-short-term photovoltaic power generation and compares its results with GRU and DNN. However, there are also articles that propose an ensemble of LSTM an Transformer model, such as [29] describing an LSTM-attention-embedding method based on Bayesian optimization for PV power prediction.

As the abovementioned examples show, many algorithms estimating the solar electric energy production have already been implemented. Nevertheless, they can be further adjusted and the new ones can be developed in order to generate even more precise predictions. Therefore, the aim of this paper is to compare Recurrent Neural Networks (RNNs) [30], Transformer Neural Networks [18] and hybrid models (making use of these approaches)in terms of PV energy generation prediction for 24 hours ahead on the basis of the results obtained respectively for Long Short-Term Memory (LSTM) [11,31], PatchTST [32] and ensemble models created based on the averaging technique [33].

3. Methodology

3.1. LSTM (Long Short-Term Memory)

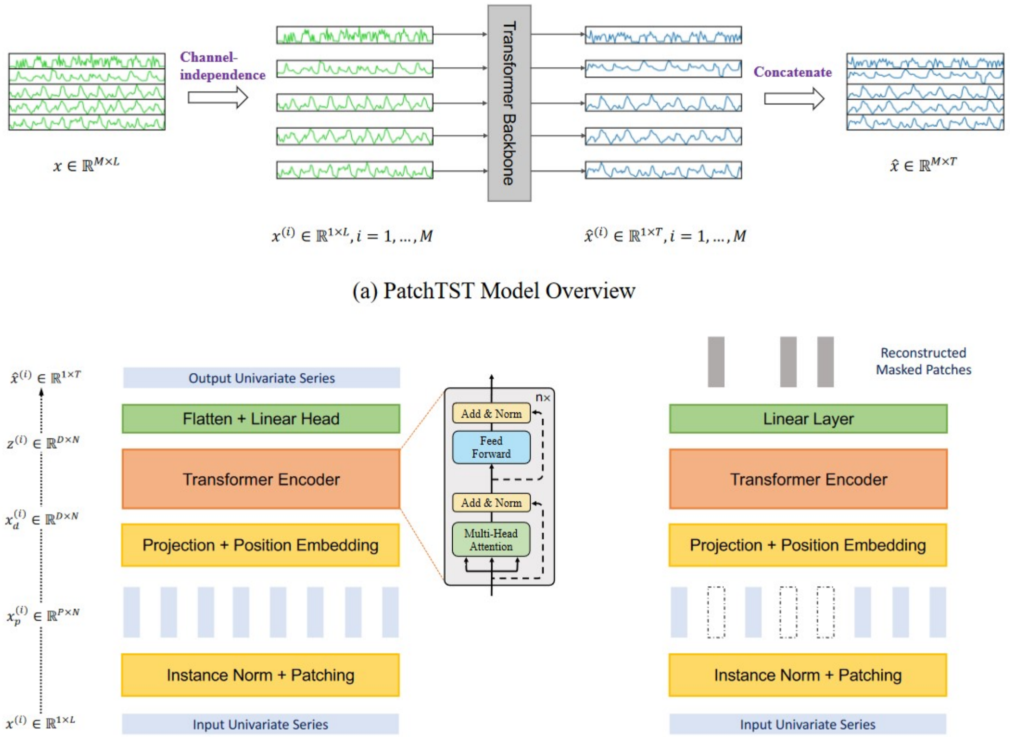
LSTM (described in [11]) is a type of RNN [3] that can handle long-term dependencies and sequential data. Unlike conventional RNNs, which suffer from the vanishing or exploding gradient problem, LSTM networks have a structure that allows them to store and manipulate information over long time intervals [11].

LSTM networks consist of cells that have three gates [31]. The input gate decides which new information to add to the cell state, the output gate determines which part of the cell state to output, and the forget gate decides which part of the cell state to discard [11, 34]. Thanks to them it is possible to regulate the information flow through the network. When using LSTM, it is advisable to normalize the data first [35].

As recent research shows [35, 36], LSTM networks can be applied to various tasks, including PV time series prediction.

3.2. PatchTST

PatchTST (proposed in [32]) belongs to modern Transformer Neural Networks (TNN) and can be applied for both single- and multi- time series modelling tasks. The architecture of PatchTST is shown in Fig. 1.



(a) PatchTST Model Overview

(b) Transformer Backbone (Supervised) (c) Transformer Backbone (Self-supervised)

Fig. 1. PatchTST architecture [32]

In this method, the input time series is first split into separate channels (Fig. 1(a)). After that, the time series in each channel is independently divided into patches inside Transformer Backbone (Fig. 1(b), Fig. 1(c)). That enables PatchTST to place each sample in the context and understand dependencies between samples in different time steps. The aforementioned patches are then used as input tokens in Transformer Encoder. Finally, each series is processed either in a supervised or self-supervised way in order to generate the results. In the former approach the patched set of vectors is used to output the full prediction length, while in the latter randomly selected patches are firstly set to zero and then reconstructed to enable masked time series pre-training.

This approach can significantly enhance the forecasting accuracy for time series problems in comparison with other Transformer Neural Networks [32].

3.3. Ensemble models

Although single models can be accurate predictors, usually it is necessary to increase the precision of generated results and a very common solution to that are ensemble models.

Ensemble model combines the predictions generated by multiple single models trained using the same dataset (or a part of it) in order to calculate the final estimated value [37]. There are many methods that make use of the aforementioned idea. One such technique is averaging in which the predicted value is calculated as a mean of the results generated by each of single predictors the ensemble model comprises of [33].

One of way to average the results is to calculate the value of arithmetic mean of the predictions (1). This type of averaging can improve the estimation results provided that all single predictors the ensemble model uses have a similar accuracy.

$$\hat{x}(t) = \frac{1}{M}\sum_{i=1}^{M} x_i(t), \tag{1}$$

where $\hat{x}(t)$ is the predicted value generated by the ensemble model for the timestep t, M is a number of single predictors and $x_i(t)$ is the predicted value generated by the i–th ($i = 1, 2, \ldots, M$) single predictor for the timestep t.

The aforementioned limitation can be though overcome by using weighted averaging, which is defined as (2). In this method the weight of an ensemble model (commitment of each single model to prediction results) is calculated using (3) based on its prediction accuracy obtained for the training dataset [33].

$$\hat{x}(t) = \sum_{i=1}^{M} w_i x_i(t), \tag{2}$$

where: w_i is the weight of the prediction generated by the i–th ($i = 1, 2, \ldots, M$) single predictor and can be calculated using (3)

$$w_i = \frac{\eta_i^m}{\sum\limits_{n=1}^{M} \eta_n^m}, \tag{3}$$

where: η_i^m is the accuracy of the i–th model on the training data and m is a coefficient that enables to increase the influence of the best single predictors on the results of the ensemble model and to decrease the impact of the less accurate ones.

However, there are also more sophisticated approaches to create predictors based on a group of models such as [37]:

– bagging, which every single predictor in the hybrid of models is based on the same machine learning algorithm, but trained using different randomly selected subsets of the training dataset,
– stacking, which introduces an additional model (known as blending predictor) to aggregate the predictions generated by single models,
– boosting, which is based on the technique of sequential usage of different models so that the output of the previous model is the input for the subsequent.

Advanced aggregation methods require additional training (e.g., blending layers in stacking) or iterative optimization (e.g., boosting), significantly increasing computational costs. Averaging and weighted averaging are fast and efficient, even with large sets of predictors. Advanced methods provide marginal benefits at the cost of increased complexity in our case.

3.4. Dataset

All the experiments were conducted using the UNISOLAR dataset [38]. It comprises solar energy generation, weather and location site data from 42 PV sites at five campuses at Australian La Trobe University from approximately two years. The data comes from five separate photovoltaic farm installations, and there is indeed not much variability between stations. Spring and autumn are quite cloudy, while summer is sunny and hot. However, this dataset is well-suited for initial investigations as it includes multiple PV sites and weather data over a substantial period.

The PV data are recorded at 15-minute time intervals and measured using kWh. Prior to using the data in the research, they were preprocessed, statistically analysed and then normalized using min-max normalization, defined as (4)

$$x_k^n = \frac{x_k - \min(\boldsymbol{X})}{\max(\boldsymbol{X}) - \min(\boldsymbol{X})}, \tag{4}$$

where: x_k^n is the value of a normalized sample, $\boldsymbol{X}$ are numerical data from learning and testing sets, and x_k is the value of a $k-$th sample of the set $\boldsymbol{X}$. The minimum and maximum values are calculated only for the learning data and are saved to calculate the actual values for the testing data, which are unknown at the training stage. The data were split into training, validation, and testing in a ratio of 70/15/15 percent of the entire available database.

3.5. Research details

The research included implementation of LSTM and PatchTST algorithms and testing them on their estimation capabilities. The latter included assessing the estimation accuracy using both different network input data configurations and different complexity level of network architectures. The best single models were then used to create ensemble models (based on averaging using the arithmetic and weighted mean) and verify whether such hybrid models can improve the precision of PV generation prediction.

Originally the testing part was to include three main input data configurations for both LSTM and PatchTST:

- only solar generation delayed by 24–72 hours (in steps of 24 hours),
- solar generation and weather data (air temperature, apparent temperature, relative humidity, dew point temperature, wind speed, wind direction) delayed by 24–72 hours (in steps of 24 hours),
- only solar generation delayed by 24–120 hours (in steps of 24 hours).

Due to the fact that Transformer Neural Network requires a context window for all the data, PatchTST data preprocessing process automatically generates it using the step of a given dataset [32]. As a result, is was necessary to modify the aforementioned data configurations for PatchTST to the ones with the step equal to the original step of the dataset used in the experiments.

Recurrent and Transformer Neural Networks make use of different approaches to learning and prediction process. Thus, their architectures cannot be directly compared. However, the comparison of architectures was made in this paper to examine how the complexity of both analysed networks impacts the PV generation prediction results. Architecture tests for LSTM included changes in the number of LSTM layers and the number of neurons in each layer, whereas for PatchTST the results for different number of attention heads were analysed (see section 3). The aforementioned additional configurations were introduced only for the models with solar generation data delayed by 24–72 hours.

The LSTM models presented in this paper were built using Keras package from TensorFlow library (version 2.15.0) [39]. Their architecture did not change during the research (excluding the architecture tests) and was as follows:

– model: sequential,
– LSTM layers: two, with 8 and 4 neurons and hyperbolic tangent as an activation function,
– loss calculation: Mean Absolute Error (MAE),
– optimizer: Adam.

MAE is commonly used alongside MSE to evaluate model performance and as the loss function. In our analysis, we included both metrics during evaluation but opted for MAE as the primary loss function due to its alignment with the goals of this study. Thanks to this, the model learns primarily to fit most of the data, which can be especially useful in the case of a rather dynamic and variable phenomenon of photovoltaic energy generation, for which it is difficult to determine which sample is an outlier and which is not.

The PatchTST models presented in this paper were created using Transformers library, version 4.39.0 [40]. The values of their parameters (apart from the number of attention heads during architecture complexity tests) were as recommended by authors of this method [32].

Prediction results for each created single model were compared and assessed using Root Mean Square Error (RMSE) and Mean Absolute Error (MAE). Unlike MSE, which amplifies the impact of large errors due to its quadratic nature, MAE treats all errors linearly. This characteristic aligns well with our dataset, which occasionally exhibits extreme variations in PV energy generation due to sudden changes in weather conditions. MAE provides a straightforward interpretation of the average prediction error in the same units as the target variable (e.g., kWh in our case). This makes it easier to contextualize the errors for practitioners and stakeholders, especially in energy applications.

In the second part of the research, some of single models were used to create a few hybrid predictors. This included both ensemble models based on calculating the arithmetic mean of predictions and the ones making use of the weighted averaging method (weights were calculated using (3) for $m = 1$ and the values of MAE foreach training dataset as the accuracy of models). The abovementioned hybrid models were tested on their prediction precision on the basis of RMSE and MAE (the same criteria as for the single models).

4. Results and discussion

4.1. Single models

In the first step of this part of the research the impact of different data configurations on the precision of the PV generation prediction using LSTM and PatchTST was analysed. The obtained numerical accuracy results and short description of each model are presented in Table 1, while sample plots of PV generation predictions and prediction errors for LSTM and PatchTST are shown respectively in Fig. 2 and Fig. 3 (for clarity purposes plots show only randomly selected 10-days periods from one PV site).

Table 1. Prediction accuracy of LSTM and PatchTST models for different input data configurations

Data configurationin model	RMSE [kWh]		MAE [kWh]	
	LSTM	**PatchTST**	**LSTM**	**PatchTST**
only solar generation delayed by 24–72 hours	1.9671	1.8248	0.8621	0.9726
solar generation and weather delayed by 24–72 hours	**1.9020**	1.8338	0.8599	1.0247
only solar generation delayed by 24–120 hours	1.9175	**1.7847**	**0.8533**	**0.8552**

Having analysed the data from Table 1, one can notice that the value of RMSE is lower for PatchTST than for LSTM in all the analysed cases. However, the obtained MAE values reveal the opposite tendency. Hence, based on the MAE and RMSE metrics it is impossible to decide which method predicts PV energy generation more precisely. Nonetheless, the presented results indicate a similar and high accuracy of both methods.

Furthermore, the data from Table 1 show that the best results for LSTM were obtained for models with the following inputs:

– solar generation and weather data delayed by 24–72 hours,
– only solar generation data delayed by 24–120 hours.

However, for the PatchTST the lowest prediction error was achieved while using the solar generation data delayed by 24–120 hours as a network input.

Observations described in the previous section can also be confirmed by analysis of Figs. 2 and 3. In addition, these plots present more detailed information on daily PV energy generation estimation than error values from Table 1, allowing assessment of the prediction quality in different conditions. Based on that, one can for instance notice the dominance of LSTM over PatchTST in periods with slightly changing generation trend, while the latter surpasses the former in estimating unexpected changes in PV generation. In addition, PatchTST model with solar generation and weather data delayed by 24–72 hours seems to be better than other analysed PatchTSTs in predicting higher amounts of PV energy, but suffers from prediction fluctuations for smaller energy portions. Thus, it is advisable to test hybrid models, for instance different PatchTST configurations or PatchTST together with LSTM.

The second step of this part of the research was analysis of LSTM and PatchTST architectures complexity in terms of prediction accuracy. The results for both tested methods are shown in Table 2 and Table 3 respectively.

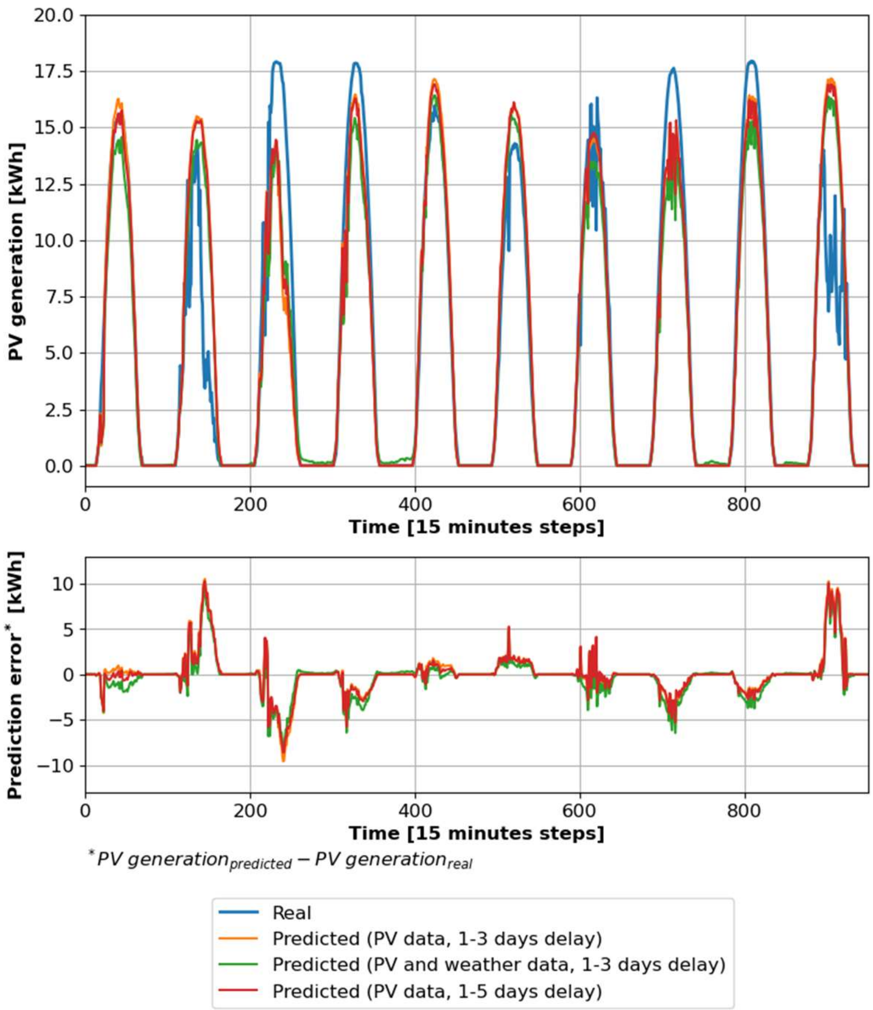


Fig. 2. Sample PV generation predictions (upper) and prediction errors (lower) for different data configurations using LSTM

Prediction accuracy summary for LSTM in Table 2 shows that models with two LSTM layers having less neurons are a little more precise in PV generation prediction than the one having a single LSTM layer with a greater number of neurons. Furthermore, a decrease in the number of neurons in LSTM layers for models with two such layers usually improves both the generalization ability and PV generation prediction results, and reduces the computational complexity of the model.

Having analysed the data presented in Table 3 one can notice that the most appropriate number of attention heads for precise predicting PV energy generation is 8. Moreover, the results show that PatchTST produces the least accurate predictions when the value of the aforementioned parameter is set to 4.

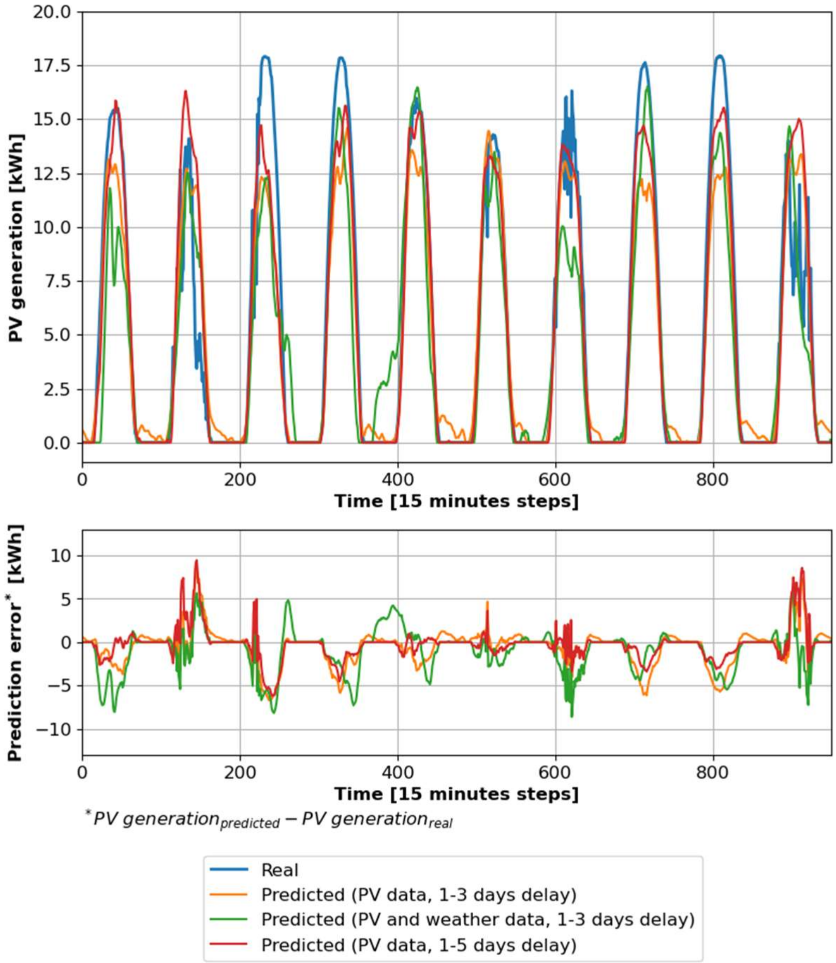


Fig. 3. Sample PV generation predictions (upper) and prediction errors (lower) for different data configurations using PatchTST

Table 2. Prediction accuracy of LSTM models for different number of LSTM layers and their number of neurons

LSTM layers configuration	**RMSE [kWh]**	**MAE [kWh]**
2 LSTM layers, 8 and 4 neurons	**1.9671**	**0.8621**
2 LSTM layers, 32 and 16 neurons	1.9801	0.8656
2 LSTM layers, 64 and 32 neurons	1.9852	0.8667
1 LSTM layer, 100 neurons	1.9878	0.8675

Table 3. Prediction accuracy of PatchTST models for different number of attention heads

Number attention heads	RMSE [kWh]	MAE [kWh]
4	1.8397	0.9860
8	**1.8070**	**0.8672**
16	1.8248	0.9726
32	1.8225	0.9525

Nonetheless, the described LSTM and PatchTST architecture tests were conducted only for one data configuration and their results should be verified for a wider number of cases.

4.2. Ensemble models

The aim of this this part of the study was the analysis of estimation accuracy of several hybrid predictors created using models from the first part of the research and averaging technique based on either arithmetic or weighted mean. Table 4 shows single models used in all ensemble models analysed during this part of the research and Table 5 shows the details of the created ensemble models (using names of single models form Table 4).

Table 4. Single models used in ensemble models

Model name	Model description
LSTM 1	Two layers 8–4 neurons LSTM with only solar generation delayed by 24–72 hours (best from Table 2, from input data configurations tests)
LSTM 2	Two layers 8–4 neurons LSTM with solar generation and weather delayed by 24–72 hours
LSTM 3	Two layers 8–4 neurons LSTM with only solar generation delayed by 24–120 hours
PatchTST 1	PatchTST with only solar generation delayed by 24–120 hours (with number of attention heads = 16)
PatchTST 2	PatchTST with only solar generation delayed by 24–72 hours (best from Table 3, with number of attention heads = 8 (from architectures complexity tests))
PatchTST 3	PatchTST with solar generation and weather delayed by24–72 hours (with number of attention heads = 16)

Table 5. Details of created ensemble models (names of single models based on data from Table 4)

Ensemble model	Single models used in a hybrid model			
	Model 1	**Model 2**	**Model 3**	**Model 4**
Hybrid of LSTM models	LSTM 1	LSTM 2	LSTM 3	–
Hybrid of PatchTST models	PatchTST 1	PatchTST 2	PatchTST 3	–
Hybrid of LSTM and PatchTST models 1	LSTM 2	LSTM 3	PatchTST 1	PatchTST 2
Hybrid of LSTM and PatchTST models 2	LSTM 2	LSTM 3	PatchTST 1	PatchTST 3
Hybrid of LSTM and PatchTST models 3	PatchTST 1	PatchTST 2	PatchTST 3	LSTM 2

The obtained numerical accuracy results for each of the aforementioned hybrid models are presented in Table 6.

Having analysed the data from Table 6, it can be noticed that ensemble models based on both presented averaging methods generate almost equally precise PV energy production predictions. Therefore, it might be advisable to use the technique of averaging the results of single predictors using their arithmetic mean, as it requires less computations.

Table 6. Prediction accuracy of LSTM and PatchTST ensemble models

Ensemble model	RMSE [kWh]		MAE [kWh]	
	Averaging (arithmetic mean)	**Weighted averaging**	**Averaging (arithmetic mean)**	**Weighted averaging**
Hybrid of LSTM models	1.9250	1.9249	0.8566	0.8566
Hybrid of PatchTST models	**1.6672**	**1.6693**	0.8536	0.8523
Hybrid of LSTM and PatchTST models 1	1.7959	1.7914	0.8400	0.8400
Hybrid of LSTM and PatchTST models 2	1.6985	1.6936	**0.8360**	**0.8358**
Hybrid of LSTM and PatchTST models 3	1.6788	1.6789	0.8401	0.8402

Moreover, the data from Table 6 indicate differences in prediction accuracy of ensemble models. Hybrid of only LSTM models usually generates more precise estimations than the majority of analysed single predictors using this type of neural network. However, it is the least accurate ensemble model. Unlike the previous hybrid predictor, the one based only on PatchTST method generates the most accurate results among all single models based on transformers. It is also the best PV generation prediction model examined (regarding the value of RMSE), but suffers from one of higher values of MAE (considering other hybrid approaches). The best prediction results are though generated by hybrids of both LSTM and PatchTST, especially by models 2 and 3 from Table 6. The RMSE and MAE values obtained for this two methods are the lowest considering all analysed PV energy production prediction methods.

In order to compare the obtained prediction results to other simple and frequently used prediction methods, tests were conducted using a naive model and linear regression. The obtained results are presented in Table 7.

Conclusions described in the previous paragraphs can be confirmed by the sample plots of PV generation predictions and prediction errors for analysed ensemble models presented in Figs. 4–7.

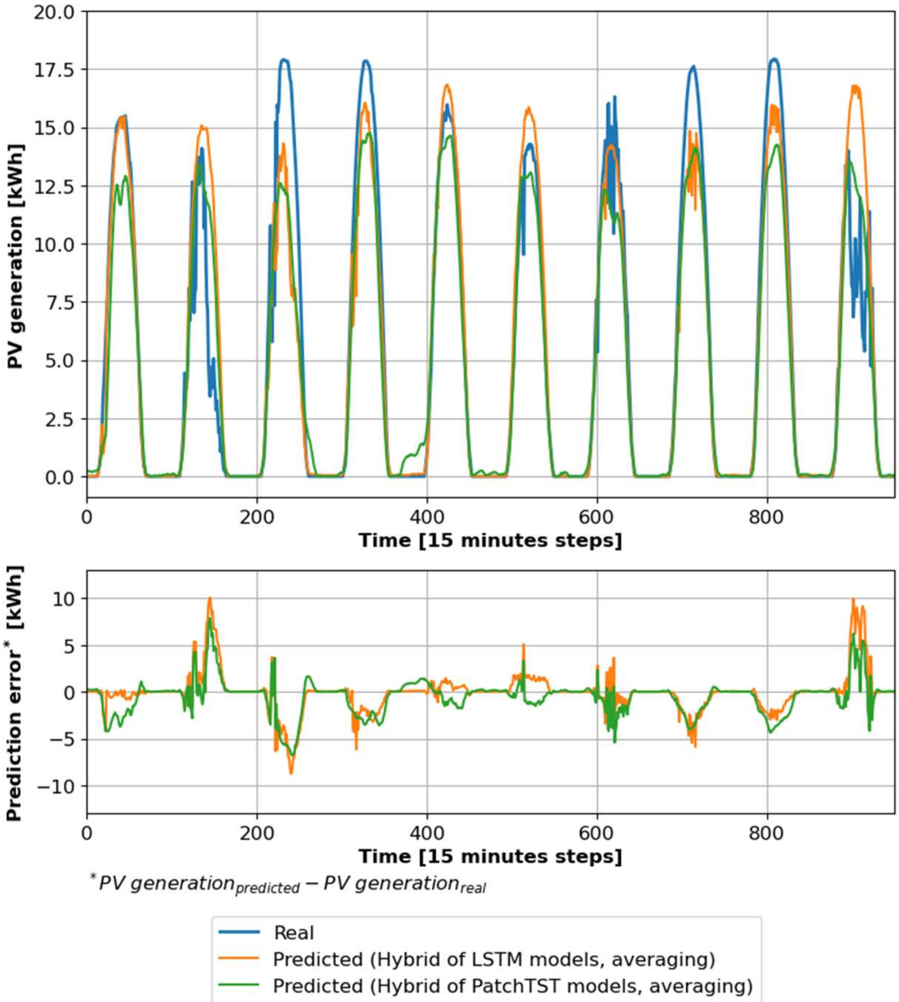


Fig. 4. Sample PV generation predictions (upper) and prediction errors (lower) using hybrid models based on either LSTM or PatchTST using averaging

Table 7. Prediction accuracy of naive model and linear regression model

Model	**RMSE [kWh]**	**MAE [kWh]**
Naive model	2.1895	0.9256
Linear regression	2.0710	0.9129

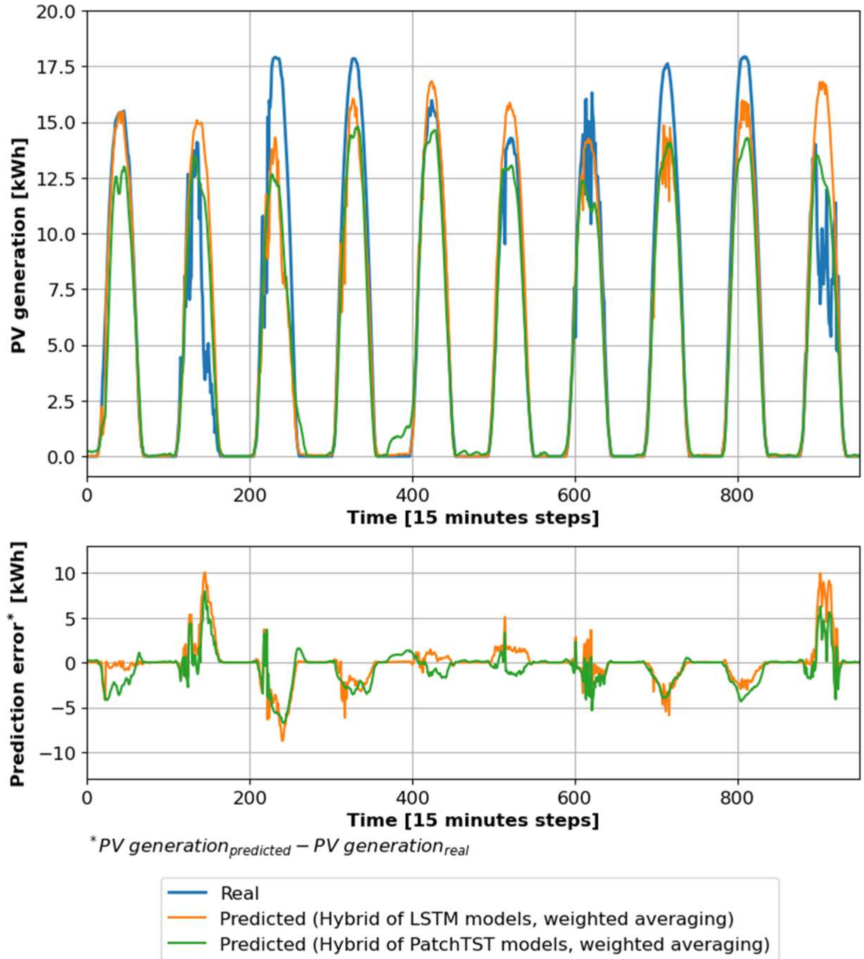


Fig. 5. Sample PV generation predictions (upper) and prediction errors (lower) using hybrid models based on either LSTM or PatchTST using weighted averaging

Based on the aforementioned plots, one can notice that making use of the ensemble models resulted in the reduction of the adverse effect of predicted value fluctuations observed for PatchTST models for smaller energy portions (Fig. 4 and Fig. 5). Furthermore, it was also possible

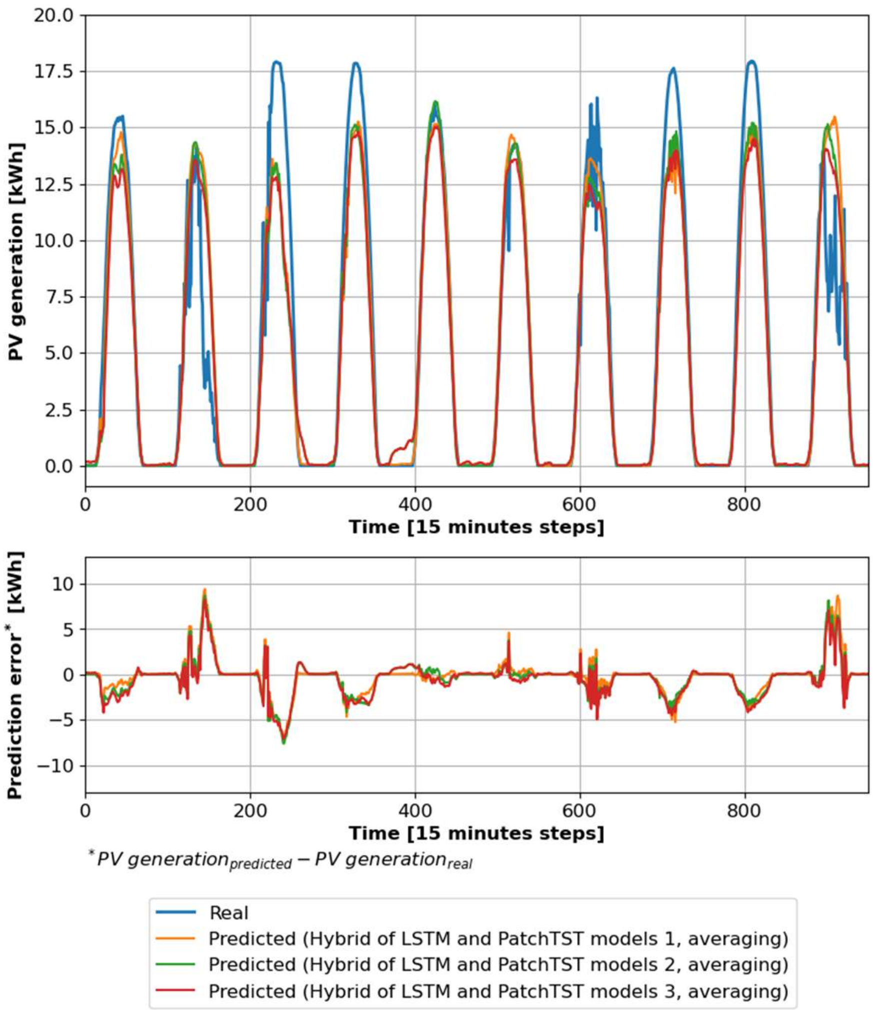


Fig. 6. Sample PV generation predictions (upper) and prediction errors (lower) using hybrid models based on LSTM and PatchTST using averaging

to diminish the phenomenon of under- and overestimating the predicted value of PV energy production. Unfortunately, ensemble models did not manage to address the problem of less precise predictions for unexpected and rapid changes in PV generation. Therefore, it would be advisable to further test other hybrid approaches to PV energy production prediction, for instance the ones considering the season or/and the part of a day.

The studied prediction methods calculate energy generation values for night hours, when it is known that PV installations are not working. The obtained forecast quality can be easily improved by not taking into account forecasts for hours after sunset and before sunrise.

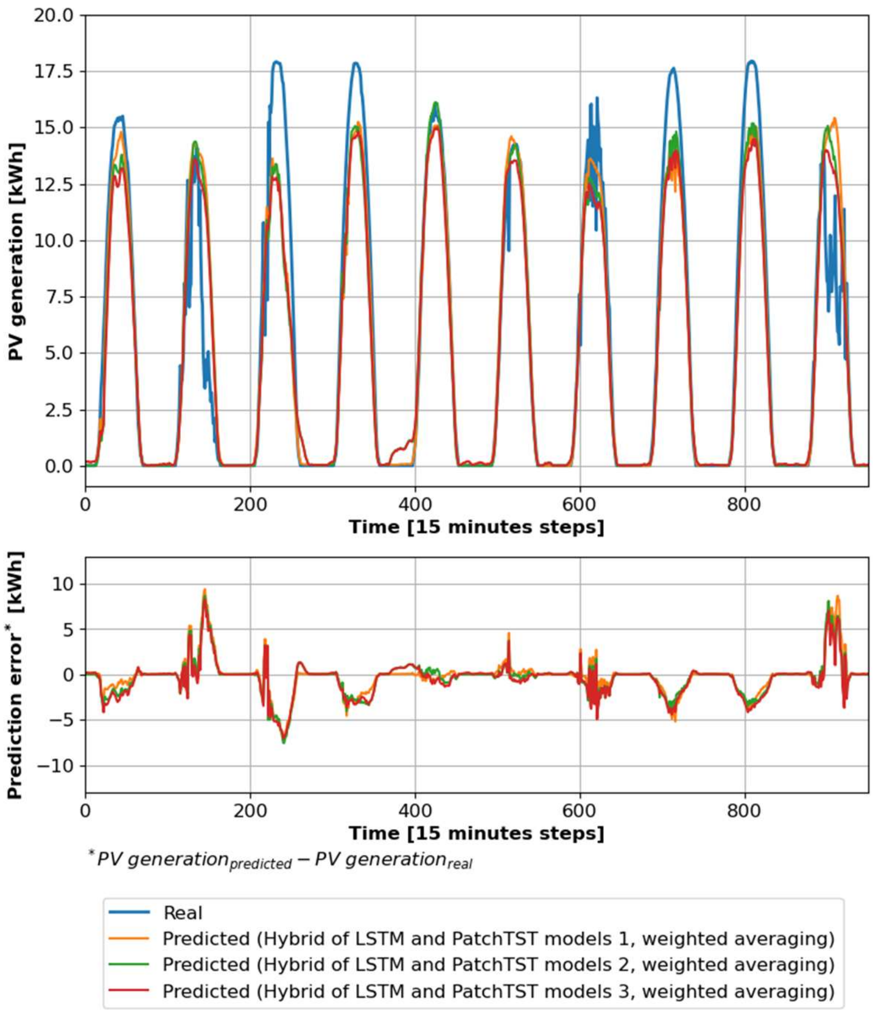


Fig. 7. Sample PV generation predictions (upper) and prediction errors (lower) using hybrid models based on LSTM and PatchTST using weighted averaging

5. Conclusions

This paper compared different approaches to estimating future photovoltaic energy generation: LSTM (a subtype of Recurrent Neural Networks), PatchTST (a type of Transformer Neural Networks) and ensemble models based on the above-mentioned methods. The obtained results suggest that both analysed single approaches present a comparable and high overall accuracy of PV energy generation prediction, though their precision varies from each other depending on the amount of produced photovoltaic energy and changes in its generation trend. Thus, hybrids of the two above-mentioned techniques were also researched. The accuracy assessment tests have shown that the best ensemble models of LSTM and PatchTST outperform all analysed single PV energy generation predictors. However, such approach is not able to address the problem of less precise predictions for unexpected and rapid changes in PV generation.

Therefore, it would be advisable to evaluate the estimation precision of different ensemble models in such application. Furthermore, due to the fact that the described experiments were conducted using the data from one country, future studies could also verify the presented methodology using a larger number of datasets from many localizations with varied weather.

Despite the aforementioned limitations, this study may contribute to the development of Recurrent and Transformer Neural Networks and their hybrid combination as a prediction methods of PV energy production.

References

[1] Singh G.K., *Solar power generation by PV (photovoltaic) technology: A review*, Energy, vol. 53 (2013), DOI: 10.1016/j.energy.2013.02.057.

[2] Jobayer M., Shaikat M.A.H., Naimur R., Rakibul H., *A systematic review on predicting PV system parameters using machine learning*, Heliyon, vol. 9, no. 6, e16815 (2023), DOI: 10.1016/j.heliyon.2023.e16815.

[3] Chandel S.S., Gupta A., Chandel R., Tajjour S., *Review of deep learning techniques for power generation prediction of industrial solar photovoltaic plants*, Solar Compass, vol. 8, 100061 (2023), DOI: 10.1016/j.solcom.2023.100061.

[4] Tsai W.-C., Tu C.-S., Hong C.-M., Lin W.-M., *A Review of State-of-the-Art and Short-Term Forecasting Models for Solar PV Power Generation*, Energies, vol. 16, no. 14, 5436 (2023), DOI: 10.3390/en16145436.

[5] Krechowicz M., Krechowicz A., Licholai L., Pawelec A., Piotrowski J.Z., Stępień A., *Reduction of the Risk of Inaccurate Prediction of Electricity Generation from PV Farms Using Machine Learning*, Energies, vol. 15, no. 11, 4006 (2022), DOI: https://doi.org/10.3390/en15114006.

[6] Suksamosorn S., Hoonchareon N., Songsiri J., *Post-Processing of NWP Forecasts Using Kalman Filtering with Operational Constraints for Day-Ahead Solar Power Forecasting in Thailand*, IEEE Access, vol. 9, pp. 105409–105423 (2021), DOI: 10.1109/ACCESS.2021.3099481.

[7] Shapsough S., Dhaouadi R., Zualkernan I., *Using Linear Regression and Back Propagation Neural Networks to Predict Performance of Soiled PV Modules*, Procedia Computer Science, vol. 155, pp. 463–470 (2019), DOI: 10.1016/j.procs.2019.08.065.

[8] Liu Y., *Short-Term Prediction Method of Solar Photovoltaic Power Generation Based on Machine Learning in Smart Grid*, Mathematical Problems in Engineering, vol. 2022, pp. 1–10 (2022), DOI: 10.1155/2022/8478790.

[9] Vrettos E., Gehbauer C., *A Hybrid Approach for Short-Term PV Power Forecasting in Predictive Control Applications*, in 2019 IEEE Milan PowerTech, Milan, Italy, pp. 1–6 (2019), DOI: 10.1109/PTC.2019.8810672.

[10] Nejati M., Amjady N., *A New Solar Power Prediction Method Based on Feature Clustering and Hybrid-Classification-Regression Forecasting*, IEEE Trans. Sustain. Energy, vol. 13, no. 2, pp. 1188–1198 (2022), DOI: 10.1109/TSTE.2021.3138592.

[11] Hochreiter S., Schmidhuber J., *Long Short-Term Memory*, Neural Computation, vol. 9, no. 8, pp. 1735–1780 (1997), DOI: 10.1162/neco.1997.9.8.1735.

[12] Asghar R., Fulginei F.R., Quercio M., Mahrouch A., *Artificial Neural Networks for Photovoltaic Power Forecasting: A Review of Five Promising Models*, IEEE Access, vol. 12, pp. 90461–90485 (2024), DOI: 10.1109/ACCESS.2024.3420693.

[13] Liu C.-H., Gu J.-C., Yang M.-T., *A Simplified LSTM Neural Networks for One Day-Ahead Solar Power Forecasting*, IEEE Access, vol. 9, pp. 17174–17195 (2021), DOI: 10.1109/ACCESS.2021.3053638.

[14] Li G., Xie S., Wang B., Xin J., Li Y., Du S., *Photovoltaic Power Forecasting with a Hybrid Deep Learning Approach*, IEEE Access, vol. 8, pp. 175871–175880 (2020), DOI: 10.1109/ACCESS.2020.3025860.

[15] Wang L., Mao M., Xie J., Liao Z., Zhang H., Li H., *Accurate solar PV power prediction interval method based on frequency-domain decomposition and LSTM model*, Energy, vol. 262, 125592 (2023), DOI: 10.1016/j.energy.2022.125592.

[16] Tovar M., Robles M., Rashid F., *PV Power Prediction, Using CNN-LSTM Hybrid Neural Network Model. Case of Study: Temixco-Morelos, México*, Energies, vol. 13, no. 24, 6512 (2020), DOI: 10.3390/en13246512.

[17] Succetti F., Rosato A., Araneo R., Panella M., *Deep Neural Networks for Multivariate Prediction of Photovoltaic Power Time Series*, IEEE Access, vol. 8, pp. 211490–211505 (2020), DOI: 10.1109/ACCESS.2020.3039733.

[18] Vaswani A. *et al*., *Attention Is All You Need*, arXiv (2017), DOI: 10.48550/arXiv.1706.03762.

[19] Tao K., Zhao J., Tao Y., Qi Q., Tian Y., *Operational day-ahead photovoltaic power forecasting based on transformer variant*, Applied Energy, vol. 373, 123825 (2024), DOI: 10.1016/j.apenergy.2024.123825.

[20] Liu J., Fu Y., *Renewable energy forecasting: A self-supervised learning-based transformer variant*, Energy, vol. 284, 128730 (2023), DOI: 10.1016/j.energy.2023.128730.

[21] Kim J., Obregon J., Park H., Jung J.-Y., *Multi-step photovoltaic power forecasting using transformer and recurrent neural networks*, Renewable and Sustainable Energy Reviews, vol. 200, 114479 (2024), DOI: 10.1016/j.rser.2024.114479.

[22] Wu J., Zhao Y., Zhang R., Li X., Wu Y., *Application of three transformer neural networks for short-term photovoltaic power prediction: A case study*, Solar Compass, vol. 12, 100089 (2024), DOI: 10.1016/j.solcom.2024.100089.

[23] Munsif M., Ullah F.U.M., Khan S.U., Khan N., Baik S.W., *CT-NET: A Novel Convolutional Transformer-Based Network for Short-Term Solar Energy Forecasting Using Climatic Information*, Computer Systems Science and Engineering, vol. 47, no. 2, pp. 1751–1773 (2023), DOI: 10.32604/csse.2023.038514.

[24] López Santos M., García-Santiago X., Echevarría Camarero F., Blázquez Gil G., Carrasco Ortega P., *Application of temporal fusion transformer for day-ahead PV power forecasting*, Energies, vol. 15, no. 14, 5232 (2022), DOI: 10.3390/en15145232.

[25] Gokhale G., Van Gompel J., Claessens B., Develder C., *Transfer learning in transformer-based demand forecasting for home energy management*, arXiv (2023), DOI: 10.48550/arXiv.2310.19159.

[26] Hu K., Fu Z., Lang C., Li W., Tao Q., Wang B., *Short-Term Photovoltaic Power Generation Prediction Based on Copula Function and CNN-CosAttention-Transformer*, Sustainability, vol. 16, no. 14, 5940 (2024), DOI: 10.3390/su16145940.

[27] Sherozbek J., Park J., Akhtar M.S., Yang O.-B., *Transformers-Based Encoder Model for Forecasting Hourly Power Output of Transparent Photovoltaic Module Systems*, Energies, vol. 16, no. 3, 1353 (2023), DOI: 10.3390/en16031353.

[28] Tian F., Fan X., Wang R., Qin H., Fan Y., *A Power Forecasting Method for Ultra-Short-Term Photovoltaic Power Generation Using Transformer Model*, Mathematical Problems in Engineering, vol. 2022, pp. 1–15 (2022), DOI: 10.1155/2022/9421400.

[29] Yang T., Li B., Xun Q., *LSTM-Attention-Embedding Model-Based Day-Ahead Prediction of Photovoltaic Power Output Using Bayesian Optimization*, IEEE Access, vol. 7, pp. 171471–171484 (2019), DOI: 10.1109/ACCESS.2019.2954290.

[30] Marhon S.A., Cameron C.J.F., Kremer S.C., *Recurrent Neural Networks*, Handbook on Neural Information Processing, vol. 49, Bianchini M., Maggini M., Jain L.C., Eds., in Intelligent Systems Reference Library, vol. 49, Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 29–65 (2013), DOI: 10.1007/978-3-642-36657-4_2.

[31] Gers F.A., Schmidhuber J., Cummins F., *Learning to forget: continual prediction with LSTM*, 9th International Conference on Artificial Neural Networks: ICANN '99, Edinburgh, UK: IEE, pp. 850–855 (1999), DOI: 10.1049/cp:19991218.

[32] Nie Y., Nguyen N.H., Sinthong P., Kalagnanam J., *A Time Series is Worth 64 Words: Long-term Forecasting with Transformers*, arXiv (2023), DOI: 10.48550/arXiv.2211.14730.

[33] Osowski S., *Metody i narzędzia eksploracji danych*, BTC (2013).

[34] Yadav H., Thakkar A., *NOA-LSTM: An efficient LSTM cell architecture for time series forecasting*, Expert Systems with Applications, vol. 238, 122333 (2024), DOI: 10.1016/j.eswa.2023.122333.

[35] Lindemann B., Muller T., Vietz H., Jazdi N., Weyrich M., *A survey on long short-term memory networks for time series prediction*, Procedia CIRP, vol. 99, pp. 650–655 (2021), DOI: 10.1016/j.procir.2021.03.088.

[36] Konstantinou M., Peratikou S., Charalambides A.G., *Solar Photovoltaic Forecasting of Power Output Using LSTM Networks*, Atmosphere, vol. 12, no. 1, 124 (2021), DOI: 10.3390/atmos12010124.

[37] Géron A., *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: concepts, tools, and techniques to build intelligent systems*, Second edition, O'Reilly Media, Inc (2019).

[38] Wimalaratne S., Haputhanthri D., Kahawala S., Gamage G., Alahakoon D., Jennings A., *UNISOLAR: An Open Dataset of Photovoltaic Solar Energy Generation in a Large Multi-Campus University Setting*, 15th International Conference on Human System Interaction (HSI), Melbourne, Australia: IEEE (2022), DOI: 10.1109/HSI55341.2022.9869474.

[39] Abadi M. *et al.*, *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*, arXiv (2016), DOI: 10.48550/arXiv.1603.04467.

[40] Wolf T. *et al.*, *Transformers: State-of-the-Art Natural Language Processing*, Conference on Empirical Methods in Natural Language Processing: System Demonstrations, pp. 38–45 (2020), DOI: 10.18653/v1/2020.emnlp-demos.6.