Doctoral School of Economics, Business and Informatics

Thesis Booklet

for the Ph.D. Dissertation by

Mutaz AlShafeey

titled

Artificial Intelligence Forecasting Techniques
for Reducing Uncertainties in Renewable Energy Applications

Supervisor: Prof. Csaba Csáki

Budapest, 2022

Corvinus University of Budapest

Department of Business Informatics

Thesis Booklet

for the Ph.D. Dissertation by

Mutaz AlShafeey

titled

Artificial Intelligence Forecasting Techniques for Reducing Uncertainties in Renewable Energy Applications

Supervisor: Prof. Csaba Csáki

Table of Contents

I. BACKGROUND AND RATIONALE

I.1. Introduction

One of the most distinguishing features of this era is the rapid technological development and the increasing demand for natural resources. As a result, a rapid continuous increase in energy demand was observed in the past 50 years. Also, this demand is expected to grow further in the next 50 years.

Traditional energy generation methods which rely on fossil fuels produce air, noise, and environmental pollution (Ashi, Joudeh, Shafeey, Sababha, & Istehkam, 2014). Such pollution has negative impacts on humans and nature. Moreover, the depletion of fossil fuels and inequality in the distribution of energy consumption and reserves causes serious problems for current energy systems (Perea-Moreno, Hernandez-Escobedo, & Perea-Moreno, 2018). Therefore, generating energy from green sustainable resources becomes an imperative necessity on the long timescale to supply the exponentially growing demand and reduce greenhouse gas emissions (Nelson & Starcher, 2015).

One of the promising green resources is renewable energy (Alshafeey & Csáki, 2019). The trend of generating energy from renewable sources, especially the energy generated from solar and wind resources has received wide approval due to its advantages such as the ease of generation and its availability in most geographical locations. Furthermore, renewable resources can complement each other, taken together can contribute to energy security by reducing foreign energy dependency (Bull, 2001).

Irrespective of all the advantages of utilizing renewable technology for energy production, there are some hindrances limiting growth and wider utilization. One of the crucial drawbacks of renewable solutions is variability. Variability problems in some renewable resources come from the reliance of some renewable energy resources on the weather variables for producing energy. Fluctuation and uncertainty in energy production lead to uncertainty in economic benefits. Additionally, uncertainty can affect grid stability in case of grid-connected PV farms (Alshafeey & Csáki, 2019).

Fluctuating energy production has serious consequences. These consequences might be either economic, such as the inability to calculate energy pricing, rate of return, and other economic elements; or technical, such as under or over generation of energy, and more importantly, fluctuation may lead to grid instability.

To achieve renewable energy generation stability there are many solutions that have been used to overcome the above problems. One promising solution for renewable energy generation stability is to enhance renewable energy forecasting (Singh, 2013). If the potential renewable energy can be accurately predicted with lower uncertainty, renewable energy systems can be better designed and optimized helping grid operators in managing power supply and demand (Pazikadin et al., 2020). Accurate forecasts would improve grid stability as well.

Renewable energy forecasting models are software solutions that can be used to forecast the future values of renewable energy generation. Like any system that predicts the future, the forecasted value of energy would have a degree of uncertainty and errors. A good forecasting model can predict future values with minimum errors and uncertainties (Cammarano, Petrioli, & Spenza, 2012).

Many approaches can be used to forecast renewable energy generation such as physical modelling, statistical modelling, artificial intelligence techniques, and their hybrids which have frequently been employed (Wang, Hu, Srinivasan, & Wang, 2018). Each method has its own pros and cons. With the development of computing techniques and hardware, artificial intelligence-based forecasting models can now provide promising forecasting performance compared to other approaches due to their potential abilities in data mining and feature-extracting (Daut et al., 2017).

Renewable energy forecasting is a sophisticated process, many factors affect forecasting accuracy. Yet, forecasting horizon and resolution, forecasting model inputs, and forecasting methods and techniques are the main factors (Ahmed, Sreeram, Mishra, & Arif, 2020).

I.2. Objectives

The main aim of this research is to employ artificial intelligence technologies in renewable energy forecasting. This will help renewable energy farms and operators in providing better forecasting accuracy based on the available data, which will reflect on grid stability and enhance renewable energy integration with grids. Another aim is to provide researchers, grid operators, and decision-makers with a comprehensive guide for forecasting methods based on the available data. Hence, one of the secondary targets is comparing the widely utilized methods of forecasting (such as MR and ML). Part of creating a comprehensive guide is to study the state of the art and research status by performing tech mining analysis. Therefore, the objectives of this research consist of four main parts.

First, to collect, study and analyze the documents published in the field of renewable energy forecasting using artificial intelligence. Second, to study and analyze the meteorological and past generation data variables to enhance the selection of input data that will be used for designing, training, and building renewable energy forecasting models. Third, to further analyze forecasting horizons and resolutions. Fourth, to study and analyze different algorithms and techniques utilizing different input data.

I.3. Research questions

Since many dimensions are interconnected and must be considered while designing a forecasting system, this research has four main research questions, each question deals with one (or more) aspects of the design problem and the objectives of this thesis.

The first research question deals with data and data availability problems:

RQ1: Which variables should be used to design, train, and build renewable energy forecasting models to improve forecasting accuracy while reducing costs and computational complexity?

The second research question deals with forecasting models and techniques:

RQ2: What are the algorithms and techniques to design, train, and build renewable energy forecasting models that can improve forecasting accuracy based on the available data?

The third research question addresses the forecasting horizon and resolution:

RQ3: What are the resolutions that can be utilized to design, train, and build renewable energy forecasting systems to assure the highest forecasting accuracy?

Most grid operators in the EU require a 15 minutes resolution forecast:

RQ4: Does the regulatory 15-minute forecasting resolution provide similar accuracy when forecasting wind and solar?

II. METHODS

II.1. Approaches

The method consists of four main stages. Each stage deals with one or more research questions. The first stage is where geographical, meteorological, and past power data are collected. Therefore, in this stage, all the collected variables will be used to build a variable selection model. The function of this model is to select the most suitable variables to perform the forecast while reducing costs and computational complexities. The variable selection model consists of sub-models, where statistical methods like correlation and regression will be used. The input of this model is all the collected variables while the output is the variables that should be used.

In the second stage, the output variables of the first stage will be utilized to build ANN and multiple regression models. Based on the input data method, i.e. structural, time-series, or hybrid, six models will be built and tested to forecast solar energy. The input of this model is the output variables of stage one while the output is the forecasted solar energy. This stage contributes to serving one of the aims of this study by providing a comprehensive guide on the effect of different input data methods on forecasting accuracy.

In the third stage, different machine learning techniques will be designed, trained, and tested to forecast wind power using only past-generation time-series data. Three main techniques will be tested including ANN, KNN, SVM. Ultimately, a hybrid machine learning model will be designed and tested. Note that wind energy data sets include only past generation data, and hence, only past generation time series data will only be used to build the mentioned models.

In the fourth stage, different input data resolutions will be used to test their effect on output accuracy. In this stage, solar and wind forecasting models will be built, trained, and tested using the ANN forecasting technique. Note that wind power data sets include only past generation data, and hence, time series past generation data of solar and wind will be used in this stage.

II.2. Tech Mining Analysis

One of the aims of this study is to provide a comprehensive guide on renewable energy forecasting using artificial intelligence. Part of creating this guide involves studying the state of the art. To that end, documents from Scopus database were extracted after applying appropriate search methods. Full tech mining analysis was performed. The most active authors and countries were found. Collaboration between countries and the citations were also analyzed. Advanced analyses like thematic maps were performed as well. The results presented help researchers assess their planned projects in light of relevant literature trends or when looking for publication outlets. The results show huge growth in the number of published documents in recent years. Asian countries like China and Iran have the highest number of documents and citations. The results also show that Chinese institutions are very active in publishing and funding research. Also, by analyzing thematic maps it was found that merging research topics like renewable energy forecasting and artificial intelligence, is highly needed. Finally, trend topics and subtopics in the field were identified. The trendy topics are changing over time. For instance, support vector regression was among the trendy topics back in 2015, while in 2021 topics like climate change, long short-term memory, and covid-19 are among the trendy topics list.

II.3. Data Collection

The data was collected from three main sources. Solar energy production data is provided by 3Comm-Hungary. 3Comm provided the authors with full access to three PV grid-connected solar farms. The three solar farms are located in Hungary, more specifically, in Szeged, Mindszent, and Kiskunhalas cities, with a peak power of 546 KWp, 547.8 KWp, and 546.15 kWp, respectively. The access allows getting the amount of produced energy, so a programming code was written to get the produced energy's data every 15 minutes. For the site in Szeged, the data was collected since it was established on April 13th, 2017, and the live data is still collected (for future research beyond this thesis), yet the data used in this work is till April 18, 2020. The second source of data is E.ON. E.ON provided the authors with energy generation data from a 2 MW wind turbine located in Csetény, Hungary. The third source of data is Solcast. Solcast provided the authors with historical data records for the three sites mentioned earlier. The provided datasets include air temperature, cloud opacity, dewpoint temperature, Diffuse Horizontal Irradiance (DHI), Direct Normal Irradiance (DNI), Direct Beam Horizontal Irradiance (EBH), Global Horizontal Irradiance (GHI), precipitable water, relative humidity, snow depth, and wind speed. All the parameters were collected every 15 minutes.

III. RESULTS

III.1. Variable Selection

Structure of the investigation in this stage:

· Some meteorological factors have higher significance than others, so the correlation between PV output power and meteorological variables was calculated;

· Variables that have low correlations (less than 0.1) with the PV power were excluded.

Key findings:

· Solar irradiance components have the highest significant factors, especially GTI fixed-tilt;

· Relative humidity and cloud opacity has a significant inverse relation with PV power output;

· Some variables such as wind speed, snow depth, and precipitable water are excluded.

III.2. Evaluating Neural Network and Linear Regression Photovoltaic Power Forecasting Models Based on Different Input Methods

Structure of the investigation in this stage:

Based on the output selection, two different techniques of PV energy prediction modeling, namely ANN and MR were built and compared;

·Depending on the input variables utilized, forecasting models were built using three different approaches: structural, time-series, and hybrid. The six models were built to predict the PV solar power for a 546 KWp grid-connected solar farm located in Hungary;

· Analysed the PV power forecasting performance of two techniques (ANN and MR) over three different methods of input data: structural, time-series, and hybrid;

· To ensure realistic results, weather and production data were collected over 3 years for a 546 KWp grid-connected PV farm;

· The six models were built and tested under regulatory conditions of 24-hour PV power prediction with 15 mins resolution.

Key findings:

· ANN forecasting models perform better regardless of input method, while the hybrid input method is better in accuracy for both MR and ANN as can be seen in Figure 1 and Figure 2;

· Poor data quality does impact forecasting accuracy negatively, especially for the structural approach;

· Results help grid-connected PV farm operators to achieve better forecasting accuracy depending on the data available to them in an industrial situation.

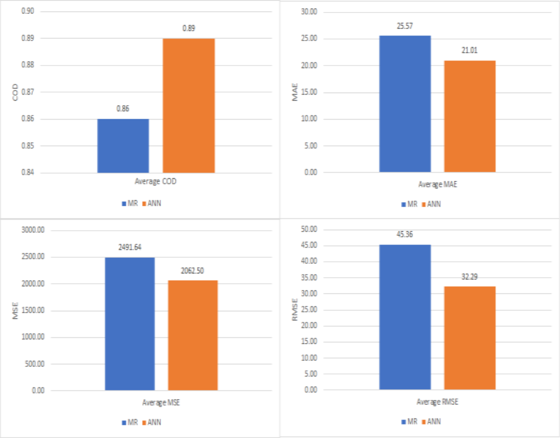

Figure 1 Average performance measures comparison between MR and ANN

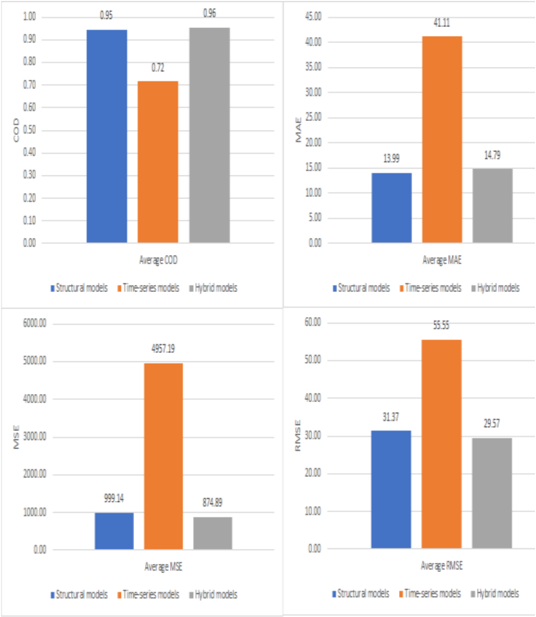

Figure 2 Average performance measures comparison between the methods

III.3. Wind Power Forecasting Discussion and Analysis

Structure of the investigation in this stage:

·After building the models and the evaluation methods based on the mathematical background, the models were utilized to perform a 24 hours wind energy forecasting;

·All models were trained and tested for 406 days, starting on the 5th of May, 2019, till the 13th of June, 2020;

The data used in this chapter was collected from a 2 MW wind turbine located in Hungary;

· First, three machine learning techniques namely ANN, SVR, and KNN were utilized to build three forecasting models to predict wind energy for the next 24 hours.;

· Then, a hybrid model based on the three tested techniques was designed, trained, and tested. Moreover, a performance analysis was provided, comparing and benchmarking the different models.

Key findings:

· The results show that ANN has very poor performance in short-term (24 hours ahead) prediction, while KNN shows a very good performance in predicting wind energy for the next 24 hours. The hybrid model shows a good performance, way better than ANN and SVR, but slightly lower than KNN as can be seen in Table 1;

· On the contrary to ANN's poor performance on short-term forecasting, it shows excellent long-term forecasting abilities as can be seen in Figure 3;

· KNN might be powerful for short-term forecasting, but in long term, some models perform better;

· The hybrid method shows a very good long-term forecasting performance with a 1.28% error. The hybrid model maintains good forecasting abilities in both long and short-term forecasting, thus it can be used in both cases while performing pretty well.

Table 1. Short-term forecasting performance measures

Model	Performance measure		
	MAE	MSE	COD
ANN	117.83	21200	0.061
SVR	68.22	12995.24	0.42
KNN	39.18	3983.48	0.82
Hybrid	63.70	9167.50	0.60

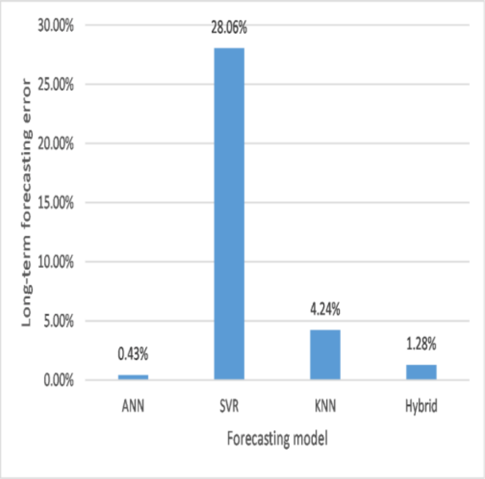

Figure 3. Long-term forecasting error

III.4. The Impact of Input Data Resolution on Neural Network Forecasting Models for Wind and Photovoltaic Energy Generation

Structure of the investigation in this stage:

· While it has been established that different forecasting horizons lead to different accuracies, the impact of input data resolution could bear some clarification;

·The objective here is to investigate how input data resolution affects prediction performance with the goal of helping wind and solar energy producers to improve their scheduling and to inform regulators on potential scheduling policy and planning issues;

· ANN modelling technique was used utilizing past energy values as time-series input data;

· Energy values were collected from a 546 KWh grid-connected solar farm and a 2 MW wind turbine, both located in Hungary. All energy values were collected with three different resolutions of 15, 30, and 60 minutes;

· The forecasting models were trained to predict the energy of both PV and wind farms for a 24-hour ahead horizon, utilizing the above input data resolutions.

Key findings:

· It was found that ANN time-series model was efficient in predicting the PV energy regardless of the input data resolution. In fact, input data resolutions have only a small effect on the accuracy of the ANN time-series PV forecasting model as forecasting measures are fairly close when utilizing 15 or 30 minutes input data resolution. Yet, the 15 minutes resolution shows better forecasting performance compared to the 60-minute resolution as it improves some performance measures by 1.3% - 4.1%;

· The same model approach shows poor performance in predicting wind energy. ANN time-series wind forecasting model has huge errors in forecasting wind energy regardless of the input data resolution. Yet, the 30 minutes input data resolution shows a slightly better performance. Utilizing the 30-minutes improves some performance measures by 0.4% - 31%;

· These results show that forecasting energy production in a 15-minute resolution might not assure high prediction accuracy for all renewable resources;

· Different renewable energy resources might need different input data resolutions to attain better forecasting accuracy.

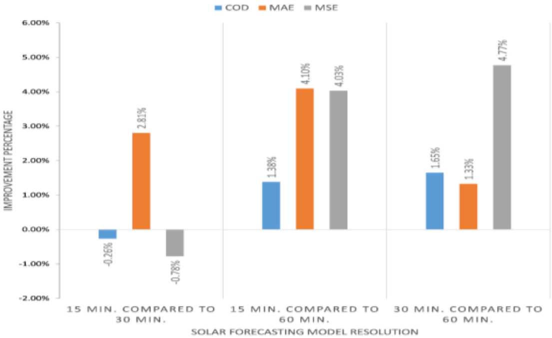

Figure 4. Performance measures comparison of PV energy forecasting

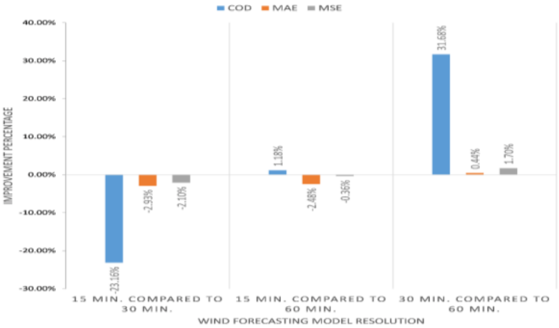

Figure 5. Performance measures comparison of wind energy forecasting

III.5. Addressing the Research Questions

RQ1: Which variables should be used to design, train, and build renewable energy forecasting models to improve forecasting accuracy while reducing costs and computational complexity?

Based on the key findings of Stage 1, it was found that some meteorological variables have a very low correlation with the PV power and hence can be eliminated.

RQ2: What are the algorithms and techniques to design, train, and build renewable energy forecasting models that can improve forecasting accuracy based on the available data?

For PV forecasting, based on the key findings of Stage 2, it was found that the ANN models perform better regardless of input method, while the hybrid input method is better in accuracy for both MR and ANN.

For wind forecasting, based on the key findings of Stage 3, it was found that KNN shows very good day-ahead forecasting abilities, yet its long-term performance can be improved. The opposite case was observed in ANN, where the short-term prediction is poor, while the long-term is very accurate. The hybrid model inherited good prediction abilities from each model, thus the suggested hybrid model shows very good abilities in both long and short-term prediction, as a result, it can be utilized in long and short wind energy forecasting while maintaining good forecasting accuracy.

RQ3: What are the resolutions that can be utilized to design, train, and build renewable energy forecasting systems to assure the highest forecasting accuracy?

Based on the key findings of Stage 4, it was found that input data resolutions have only a small effect on the accuracy of the ANN time-series PV forecasting model. Yet, the 15 minutes resolution shows better forecasting performance compared to the 60-minute resolution as it improves some performance measures by 1.3% - 4.1%. The same model approach shows poor performance in predicting wind energy. ANN time-series wind forecasting model has huge errors in forecasting wind energy regardless of the input data resolution. Yet, the 30 minutes input data resolution shows a slightly better performance. Utilizing the 30-minutes improves some performance measures by 0.4% - 31%.

RQ4: Does the regulatory 15-minute forecasting resolution provide similar accuracy when forecasting wind and solar?

Based on the key findings of Stage 4, it was found that the 15-min. ANN time-series forecasting models are suitable for predicting PV output energy, Yet, for wind other resolutions such as 30-min. shows better accuracy.

IV. CONCLUSIONS AND PRACTICAL IMPLICATIONS

This thesis has demonstrated the use of artificial intelligence forecasting techniques in wind and solar energy prediction utilizing different input data methods. To that end, the work was done in four major parts. In the first part, a systematic literature review was conducted. In the second part, neural network and linear regression photovoltaic power forecasting models based on different input methods were evaluated. The third part studies wind energy forecasting. Finally, the fourth part discusses the impact of input data resolution on neural network forecasting models for wind and photovoltaic energy generation.

In the first part, tech mining analysis in the field of renewable energy forecasting using AI technologies was discussed. The results show that there is huge growth in the number of published documents in recent years. Asian countries like China, India, and Iran have the highest number of documents, the highest number of citations, and high number of citations per document.

In the second part, two different techniques of PV energy prediction modelling, namely ANN and MR were analysed. Depending on the input variables utilized, forecasting models were built using three different approaches: structural, time-series, and hybrid.

The results indicate that ANN forecasting models have higher COD and lower MAE, MSE, RMSE values compared to the MR, regardless of the method used for building the forecasting models. It was also found that using the hybrid method to build prediction models results in better prediction accuracy for both MR and ANN while using the time-series method results in the least accurate forecasting models.

After analysing the results of this work using real farm data, it was confirmed that ANN technique performs better than the MR. This is true regardless of the input method used to build the models.

In the third part, four machine learning wind forecasting models were built and tested. First, three machine learning techniques namely ANN, SVR, and KNN were utilized to build three forecasting models to predict wind energy for the next 24 hours. Then, a hybrid model based on the three tested techniques was designed, trained, and tested.

The results show that ANN has very poor performance in short-term (24 hours ahead) prediction, while KNN shows a very good performance in predicting wind energy for the next 24 hours. The hybrid model shows a good performance, way better than ANN and SVR, but slightly lower than KNN.

On the contrary of ANN's poor performance on sort-term forecasting, it shows excellent long-term forecasting abilities. The hybrid method shows a very good long-term forecasting performance with a 1.28% error. It can be observed from the analysis that the hybrid model maintains good forecasting abilities in both long and short-term forecasting, thus it can be used in both cases while performing pretty well.

In the fourth part, different input data resolutions were used to build and test PV and wind energy forecasting. ANN modelling technique was used utilizing past energy values as time-series input data.

It was found that ANN time-series model was efficient in predicting the PV energy regardless of the input data resolution. In fact, input data resolutions have only a small effect on the accuracy of the ANN time-series PV forecasting model as forecasting measures are fairly close when utilizing 15, 30, or 60 minutes input data resolution. Yet, the 15 minutes resolution shows slightly better forecasting performance.

The same model approach shows poor performance in predicting wind energy. ANN time-series wind forecasting model has huge errors in forecasting wind energy regardless of the input data resolution. Yet, the 30 minutes input data resolution shows a slightly better performance.

V. MAIN REFERENCES

Ahmed, R., Sreeram, V., Mishra, Y., & Arif, M. (2020). A review and evaluation of the state-of-the-art in PV solar power forecasting: Techniques and optimization. Renewable and Sustainable Energy Reviews, 124, 109792.

Alshafeey, M., & Csáki, C. (2019). A Case Study of Grid-Connected Solar Farm Control Using Artificial Intelligence Genetic Algorithm to Accommodate Peak Demand. Paper presented at the Journal of Physics: Conference Series.

Ashi, A., Joudeh, A. A., Shafeey, M., Sababha, B. H., & Istehkam, S. N. (2014). A PV solar tracking system: Design, implementation and algorithm evaluation. Paper presented at the Information and Communication Systems (ICICS), 2014 5th International Conference on.

Bull, S. R. (2001). Renewable energy today and tomorrow. Proceedings of the IEEE, 89(8), 1216-1226.

Cammarano, A., Petrioli, C., & Spenza, D. (2012). Pro-Energy: A novel energy prediction model for solar and wind energy-harvesting wireless sensor networks. Paper presented at the 2012 IEEE 9th International Conference on Mobile Ad-Hoc and Sensor Systems (MASS 2012).

Daut, M. A. M., Hassan, M. Y., Abdullah, H., Rahman, H. A., Abdullah, M. P., Hussin, F. J. R., & Reviews, S. E. (2017). Building electrical energy consumption forecasting analysis using conventional and artificial intelligence methods: A review. 70, 1108-1118.

Nelson, V. C., & Starcher, K. L. (2015). Introduction to renewable energy: CRC press.

Pazikadin, A. R., Rifai, D., Ali, K., Malik, M. Z., Abdalla, A. N., & Faraj, M. A. (2020). Solar irradiance measurement instrumentation and power solar generation forecasting based on Artificial Neural Networks (ANN): A review of five years research trend. Science of The Total Environment, 715, 136848.

Perea-Moreno, M.-A., Hernandez-Escobedo, Q., & Perea-Moreno, A.-J. (2018). Renewable energy in urban areas: Worldwide research trends. Energies, 11(3), 577.

Singh, G. K. (2013). Solar power generation by PV (photovoltaic) technology: A review. Energy, 53, 1-13.

Wang, Y., Hu, Q., Srinivasan, D., & Wang, Z. J. I. T. o. S. E. (2018). Wind power curve modeling and wind power forecasting with inconsistent data. 10(1), 16-25.

VI. PUBLICATIONS RELATED TO THIS DISSERTATION

VI.1. Journal papers

AlShafeey, M., & Csáki, C. (2021). Evaluating neural network and linear regression photovoltaic power forecasting models based on different input methods. Energy Reports, 7, 7601-7614.

Alshafeey, M., & Csaba, C. (2019, August). A case study of grid-connected solar farm control using artificial intelligence genetic algorithm to accommodate peak demand. In Journal of Physics: Conference Series (Vol. 1304, No. 1, p. 012017). IOP Publishing.

VI.2. Conference papers, abstracts and presentations

AlShafeey, M., & Csaki, C. (2022, April). Tech Mining Analysis: Renewable Energy Forecasting Using Artificial Intelligence Technologies. In 2022 IEEE Nigeria 4th International Conference on Disruptive Technologies for Sustainable Development (NIGERCON) (pp. 1-5). IEEE.

Al Shafeey, M., & Harb, A. M. (2018, March). Photovoltaic as a promising solution for peak demands and energy cost reduction in Jordan. In 2018 9th International Renewable Energy Congress (IREC) (pp. 1-4). IEEE.

AlShafeey, M., & Csaki, C (2021, November). Disruptive Innovation in Renewable Energy Technologies. In: Raffai, Mária; Kosztyán, Zsolt Tibor (szerk.) OGIK'2021 Országos Gazdaságinformatikai KonferenciaVeszprém, Magyarország : Platina Nyomda és Kiadó Kft. (2021) 64 p. p. 30.

Csáki, Csaba., & **AlShafeey, Mutaz** (2021, April). Artificial neural network forecasting models for wind and photovoltaic energy prediction utilizing time-series input data with different resolutions. In: Kurt, Erol (szerk.) 9. European Conference on Renewable Energy Systems (ECRES 2021) Ankara, Törökország : Gazi University (2021) pp. 103-108. , 6 p.

AlShafeey, M., & Csaki, C (2020, November). Artificial Intelligence prediction models for solar power forecasting. In: Aniko, Kelemen-Erdos; Anett, Popovics; Pal, Feher-Polgar (szerk.) XV. FIKUSZ 2020 International Conference Abstract Book Budapest, Magyarország : Óbudai Egyetem, Keleti Károly Gazdasági Kar (2020) 72 p. pp. 69-70. , 2 p.