Future Energy
Open Access Journal
ISSN 2832-0328

August 2024| Volume 03 | Issue 03 | Pages 24-32

Journal homepage: https://fupubco.com/fuen

https://doi.org/10.55670/fpll.fuen.3.3.3

Future Publishing LLC

Article

Predicting power and solar energy using neural networks and PCA with meteorological parameters from Diass and Taïba Ndiaye

Sambalaye Diop[1*], Papa Silly Traore[1], Mamadou Lamine Ndiaye[1], Issa Zerbo[2], Vincent Sambou[1]

[1]Electrical Engineering Department of Polytechnic School of Dakar (ESP), Water, Energy, Environment, Industrial Process Laboratory (LE3PI), Dakar, Senegal

[2]Departmemt of Physics of the Joseph KI-ZERBO UNIVERSITY), Laboratory of Renewable Thermal Energies (LETRE), Ouagadougou, Burkina Faso

ARTICLE INFO

Article history:
Received 03 February 2024
Received in revised form
16 March 2024
Accepted 29 March 2024

Keywords:
Renewable energies, Prediction, Artificial neural networks, Vulnerability, Weather data – PCA

*Corresponding author
Email address:
sambalayediop7@gmail.com

DOI: 10.55670/fpll.fuen.3.3.3

ABSTRACT

The excessive reliance on conventional fossil fuel-based resources poses a significant threat to our environment. To mitigate this impact, it has become increasingly crucial to increase the integration of intermittent and non-polluting energy sources into our electrical grids. However, while this higher penetration rate brings benefits such as improved producer satisfaction and reduced fossil fuel consumption, it also presents challenges for traditional non-smart electrical networks. To promote intermittent energy sources effectively and maintain a balance between consumption and production, accurate forecasting of these energy outputs plays a vital role. This research paper focuses on studying the application of artificial neural networks for predicting the power and energy output of the Diass solar power plant in the short and medium term. The proposed approach utilizes not only the meteorological data from the city where the power plant is located but also data from a nearby city with a data acquisition station. Principal component analysis (PCA) is employed to select the relevant variables for the prediction model. Furthermore, the results obtained from our approach are compared to existing literature that solely uses meteorological data from the power plant's location. The comparison shows that our method achieves more satisfactory results, with mean absolute errors and root mean square errors of 0.0223 KWh and 0.003 KWh, respectively, and a prediction accuracy of 94.57% in terms of energy and power. It is worth noting that the computational resource requirements for our approach are higher, with simulation times ranging between 1788 seconds and 2201 seconds. By utilizing a broader range of data sources and employing advanced techniques like artificial neural networks, this research contributes to improving the accuracy of solar power generation forecasts. The findings highlight the potential of incorporating additional data inputs and advanced modeling techniques to enhance the performance of renewable energy systems, paving the way for a more sustainable and efficient energy future.

1. Introduction

The global energy sector is undergoing a significant transformation, driven by the urgent need to reduce greenhouse gas emissions and mitigate the environmental impact of conventional fossil fuel-based resources. Integrating intermittent and non-polluting energy sources, such as solar and wind power, has become a priority for many countries aiming to achieve a sustainable and low-carbon energy future [1- 2]. However, the increased penetration of these renewable energy sources poses challenges for the existing electrical grid infrastructure, primarily designed for centralized and predictable power generation [3]. To effectively harness the potential of intermittent energy sources and maintain a reliable balance between energy consumption and production, accurate forecasting of their power and energy output is crucial. Accurate predictions enable grid operators to optimize energy dispatch, plan for storage requirements, and ensure grid stability [4]. Over the years, various forecasting models and techniques have been employed to improve the accuracy of renewable energy generation forecasts, including statistical models, time-series analysis, and machine learning algorithms. In recent years, artificial neural networks (ANNs) have emerged as a powerful tool for renewable energy forecasting due to their ability to capture complex nonlinear relationships and adapt to changing conditions [5]. ANNs have demonstrated superior prediction capabilities compared to traditional statistical models, and their performance can be further enhanced by incorporating additional relevant variables [6].

In this context, this research paper focuses on predicting the power and energy output of the Diass solar power plant in Senegal using an ANN-based forecasting model. The proposed approach not only leverages meteorological data from the power plant's location but also integrates data from a nearby city with a data acquisition station. This incorporation of additional data sources aims to enhance the accuracy of the prediction model and address the limitations of existing approaches that rely solely on local meteorological data [7]. To select the most relevant variables for the ANN model, we employ principal component analysis (PCA), a widely used technique for dimensionality reduction and feature selection [8]. By reducing the input variables to a smaller set of principal components, the model can capture the essential information while minimizing computational complexity. The results obtained from our approach are compared with existing literature that utilizes only local meteorological data for solar power generation forecasting. The comparison showcases the superior performance of our method, with reduced mean absolute errors and root mean square errors, indicating higher accuracy in predicting the power and energy output. Additionally, we evaluate the computational resource requirements of our approach to provide insights into its feasibility and scalability [9]. By integrating a broader range of data inputs and leveraging advanced modeling techniques like ANNs, this research contributes to the ongoing efforts to improve the accuracy of solar power generation forecasts [10]. The findings highlight the potential of incorporating additional data sources and advanced algorithms to enhance the performance and reliability of renewable energy systems. Ultimately, these advancements pave the way for a more sustainable and efficient energy future, aligning with the goals of transitioning towards a low-carbon society [11].

The article is organized as follows:

- Section II gives an overview of the study sites and the data that were used for the research.
- Section III describes the methods that were employed, including the neural network model used for prediction.
- The results obtained from the study are presented and discussed in this section.
- Finally, the article concludes with a summary of the findings and their implications.

2. Data and site overview

To gain insights into the challenges faced by energy producers and electrical network managers, a field study was conducted to collect data from the Diass and Taïba Ndiaye power plants. The geographical locations of the cities studied, Diass and Taïba Ndiaye, provide valuable insights into the weather and environmental factors influencing power generation in Senegal. Diass is situated at approximately 14°63'92'' North latitude and -17°08'78'' West longitude, while Taïba Ndiaye is located at around 15° 2' 22.1" North latitude and -16° 52' 43" West longitude (Figure 1). Both Diass and Taïba Ndiaye experience a Sahelo-Sudanese climate characterized by a distinct rainy season. In Diass, the rainy season typically spans from June to October, while in Taïba Ndiaye, it extends from July to October [12, 13]. The annual average rainfall in Diass is approximately 440 mm, with an average temperature of 27 °C [12]. On the other hand, Taïba Ndiaye experiences a range of temperatures, with the highest reaching 35°C and the lowest dropping to 16°C. The annual average rainfall in Taïba Ndiaye ranges between 400 and 600 mm [13]. The choice of these specific locations for our study is driven by the significant contribution of the Diass and Taïba Ndiaye power plants to Senegal's renewable energy production capacity. The Diass solar power plant is located in the city of Diass, while the Taïba Ndiaye power plant generates wind power and is situated in Taïba Ndiaye. These sites offer valuable data for analyzing and predicting power generation from renewable sources in Senegal. Understanding the geographical context of the study sites is crucial for comprehending the local weather patterns, solar irradiance levels, and other environmental factors that influence power generation. By considering the specific characteristics of these locations, we can better analyze the data collected and develop accurate prediction models to optimize renewable energy production and management.

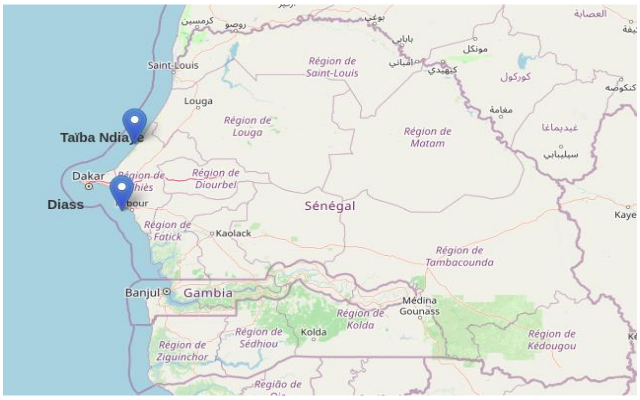

Figure 1. Geographical location of the sites studied

Diass benefits from ample sunlight, which is essential for solar photovoltaic power generation, while Taïba Ndiaye benefits from strong wind resources and solar, making it suitable for generation. By obtaining data from these specific sites, we aim to capture the unique characteristics and challenges associated with renewable energy generation in Senegal. This information will be crucial for developing accurate prediction models and addressing the complexities of integrating intermittent energy sources into the existing electrical grid infrastructure. During the data analysis process, outliers were identified in the recorded panel temperature values in the city of Diass. These outliers showed temperatures exceeding 55°C, while the highest ambient temperature measured was only 38°C (Figure 2). It is important to note that these extreme values are likely the result of measurement errors caused by sensor malfunction or calibration issues. Correcting data errors is crucial for ensuring the accuracy and reliability of the prediction models. Numerous studies emphasize the significance of data quality and the impact of outliers on model performance. For instance, in a study [14], the authors highlight the importance of identifying and handling outliers in data preprocessing to improve the performance of prediction models. To address the outlier issue, a rigorous data cleansing process will be implemented. Techniques such as Winsorization, which replaces extreme values with more reasonable ones, and outlier removal based on statistical analysis can be employed [15]. By correcting these data errors, we can ensure that the prediction models are trained and tested on reliable and consistent data, leading to more accurate and meaningful results.

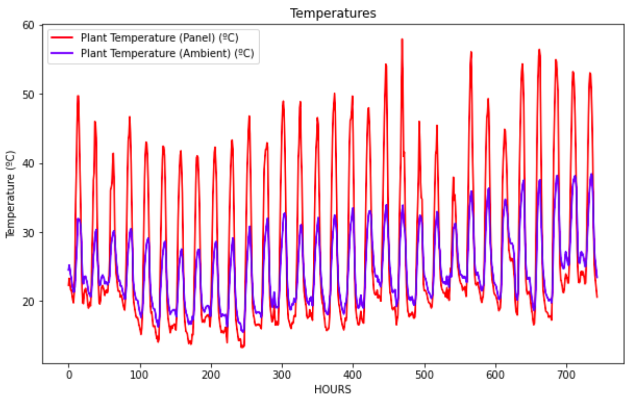

Figure 2. Variation of the temperature of the panels according to the ambient temperature case of Diass

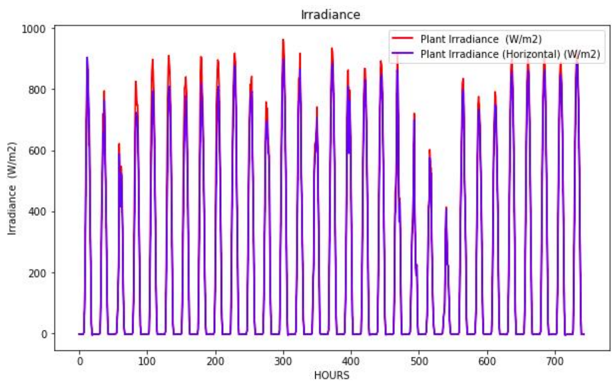

Figure 3. Variation of the irradiance on the panels as a function of the irradiance of the horizontal plane case of Diass

Figure 2 displays the irradiance on the horizontal plane compared to the irradiance received in the plane of the photovoltaic panels. It is evident that there are a few values that do not align with the irradiance received in the panel plane, indicating potential measurement errors. To address these outliers, they can be either removed from the dataset or replaced with the average value. This step ensures the accuracy and consistency of the data used for prediction. In Figure 3, the variation of irradiance in the two cities under study is depicted. It is noteworthy that there is a correlation between the irradiances of the two cities, suggesting a potential positive impact on our prediction. The small difference observed between the irradiances of the two cities implies that incorporating data from both locations can provide valuable insights for improving the accuracy of our forecasting model. These observations align with previous studies that highlight the significance of considering multiple data sources and correlations for accurate solar power generation predictions [16, 17]. By leveraging the correlated irradiance data from different cities, our prediction model can benefit from a broader and more diverse dataset, leading to improved forecasting capabilities.

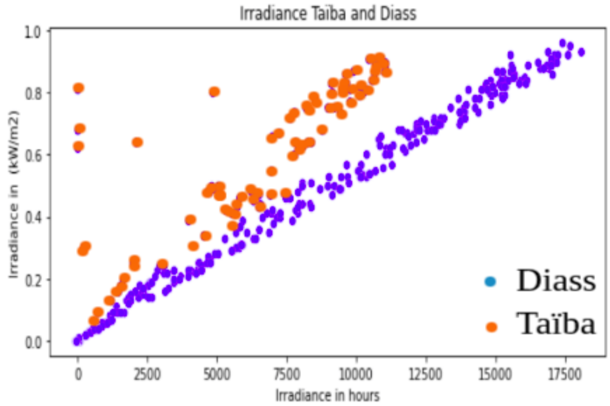

Figure 4. Variation of the irradiance of the city of Taïba Ndiaye and Diass

In Figure 4, we can observe the irradiance data specifically for the Taïba area, which we have associated with the data collected from the sensors at the Diass power plant. The objective of this analysis is to determine whether incorporating environmental parameters from a neighboring town can enhance the accuracy of our prediction model. Upon examining the plot, it is evident that there are periods of both high and low sunshine potentials in the Taïba area. This variability in irradiance levels is similarly observed in the Diass area, as shown in Figure 3. These findings indicate that the solar energy potential in both locations is subject to fluctuations due to meteorological factors such as cloud cover, atmospheric conditions, and seasonal variations. By integrating the irradiance data from the Taïba area into our prediction model, we can benefit from additional insights and a more comprehensive understanding of the environmental conditions that influence solar power generation. This approach aligns with previous research that emphasizes the importance of incorporating diverse and geographically distributed data sources to improve the accuracy of solar power prediction models [18, 19]. The inclusion of data from the Taïba area allows us to capture localized variations in solar irradiance (Figure 5), which may not be fully captured by the data collected solely at the Diass power plant. This broader perspective enhances the robustness of our prediction model and provides valuable information for grid operators and energy managers to optimize energy production and distribution. Our objective is to develop a prediction model for the power and energy output of the Diass power plant using meteorological parameters and the irradiance data from Taïba Ndiaye. The target data to be predicted are represented in Figure 6. A proportional correlation between power and energy can be observed. This means that when power increases, the consumed or generated energy also increases, and vice versa. This proportional relationship is consistent with the fundamental principles of electricity, where power is the amount of energy consumed or produced per unit of time. By visualizing these data in Figure 6, we can better understand this relationship and use it to develop more accurate prediction models. Analyzing this correlation between power and energy can also contribute to optimizing energy management and making more informed decisions in the field of renewable energy.

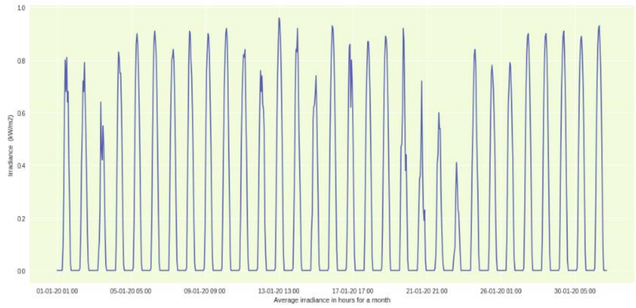

Figure 5. Variation of the irradiance of the city of Taïba Ndiaye

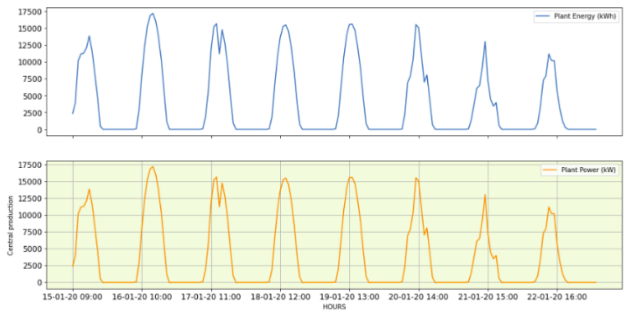

Figure 6. Variation in power and energy produced by the Diass power plant as a function of time

To accomplish this task, we will employ a comprehensive methodology that takes into account the complexity of the study. The dataset will be divided into two parts: 80% of the data will be allocated for training the prediction model, while the remaining 20% will be used for testing the model's performance. This division allows us to effectively evaluate the accuracy and reliability of the model in predicting the power and energy output. By dedicating a significant portion of the dataset for training, we ensure that the model captures the underlying patterns and relationships between the meteorological parameters, irradiance, and power generation. This enables us to create a robust and accurate prediction model that can be applied to real-time scenarios. The testing phase with the remaining data is crucial for assessing the model's performance and verifying its predictive capabilities. By evaluating the model on unseen data, we can gauge its generalization ability and ensure that it can provide accurate predictions beyond the training data. This methodological approach ensures that our prediction model is built on a solid foundation of training and testing, allowing us to effectively forecast the power and energy output of the Diass power plant based on the meteorological parameters and Taïba Ndiaye's irradiance data. Overall, by following this approach, we aim to develop a reliable and accurate prediction model that can support decision-making processes and optimize the integration of renewable energy sources into the electrical grid.

3. Methodology

The methodology for this study consists of several steps to ensure an effective prediction model for the power and energy output of the Diass power plant:

- Dimension Reduction using Normalized PCA: The first step involves applying a normalized Principal Component Analysis (PCA) method. This technique helps reduce the dimensionality of the model training data by identifying highly correlated variables in the dataset. By reducing redundancy and eliminating irrelevant variables, PCA improves the efficiency and accuracy of the prediction model. This step ensures that only the most informative and relevant features are considered, leading to better predictions.
- Data Subdivision: The dataset is then divided into two subsets: a testing set and a training set. The testing set accounts for 20% of the data, while the remaining 80% is used to train the model. This division allows us to evaluate the model's performance on unseen data and assess its generalization ability. The training set is utilized to optimize and fine-tune the neural network model for accurate power and energy prediction.
- Neural Network Model Application: With the training data, the neural network model is applied to learn the underlying patterns and relationships between the meteorological parameters, irradiance data, and power generation. The model is designed to capture complex nonlinear dependencies and adaptively adjust its internal parameters to make accurate predictions.
- Performance Evaluation: After training the model, its performance is evaluated using the testing set. Various metrics and indicators, such as mean absolute error and root mean square error, are used to quantify the model's performance and assess its predictive capabilities. This evaluation provides insights into the accuracy and reliability of the model in predicting the power and energy output of the Diass power plant.
- Power and Energy Prediction: Finally, the trained and evaluated model is utilized to predict the power and energy output of the Diass region. By incorporating the relevant meteorological parameters, irradiance data, and the insights gained from the previous steps, the model can provide reliable and accurate predictions for short and medium-term time horizons.

By following this methodology, we can ensure an efficient and reliable prediction model that leverages dimension reduction, data subdivision, neural network modeling, and performance evaluation. This approach enhances the model's accuracy and applicability for predicting the power and energy output of the Diass power plant, contributing to the effective management of renewable energy resources.

3.1 Dimension reduction using the PCA method for variable selection

The field of big data has witnessed the rise of principal component analysis (PCA) as an effective method for unsupervised variable selection [20]. PCA is commonly employed for numerical value prediction, as it helps eliminate redundant variables and identify the underlying population structure for classification purposes. Essentially, PCA aims to find the optimal linear subspace that minimizes the information loss when projecting the data. Consequently, the selected variables will capture the essence of the entire dataset. PCA offers a powerful tool to streamline and enhance the predictive modeling process by reducing dimensionality and improving the interpretability of the selected variables. The essence of PCA lies in finding the best linear subspace that retains the maximum amount of information from the original data. By projecting the data onto this subspace, the variables selected by PCA capture the essential characteristics of the entire dataset. In other words, they provide a condensed representation of the data while preserving its key properties. This not only simplifies the modeling process but also enhances the interpretability of the selected variables, as they collectively reflect the overall image of the original variables [21].

3.2 Sample

When dealing with samples of data from different sites, each characterized by multiple random variables (X1, X2, ..., XN), applying Principal Component Analysis (PCA) becomes a valuable approach. By examining the correlation matrix of the data, as depicted in Figure 7, we can further justify the relevance of employing PCA in this study.

Figure 7. Correlation matrix of study data

The correlation matrix provides a comprehensive view of the relationships between the variables, allowing us to assess their interdependencies. By analyzing the correlation matrix, we can identify variables that exhibit strong correlations, indicating a high degree of linear association. Conversely, variables with weak correlations suggest a lower level of linear dependence [22]. This information is crucial in understanding the underlying structure and patterns present in the dataset. PCA leverages this correlation information to transform the original variables into a new set of uncorrelated variables, known as principal components. These principal components are linear combinations of the original variables, and they are ordered in terms of their ability to explain the variance in the data. The first principal component accounts for the largest variance, followed by the second principal component, and so on [21]. By selecting a subset of the principal components, we can effectively capture the essential information contained in the original variables while reducing dimensionality. By observing the correlation matrix in Figure 7, we can gain insights into the strength and nature of the relationships between the variables. Variables with high positive or negative correlations indicate a significant linear association, suggesting they may convey similar information and exhibit redundancy [23]. In such cases, PCA can help identify the dominant underlying factors driving the data, facilitating variable selection and dimensionality reduction. Furthermore, the correlation matrix allows us to detect any potential multicollinearity issues where variables are highly correlated with each other. Multicollinearity can lead to instability and unreliable estimates in regression models, making it necessary to address this problem. PCA can effectively mitigate multicollinearity by identifying the principal components that capture the most significant sources of variation and combining variables with high correlations into a reduced set of uncorrelated components.

3.3 Normalization of variables

Normalization is an essential preprocessing technique employed to simplify the complexity of the Artificial Neural Network (ANN) model used in our study. It aims to ensure that the input data is within an optimal range for the neural network to operate effectively, typically between -1 and 1. This normalization approach has been widely adopted in various studies documented in the literature [24-26]. The normalization process involves scaling the data to a specific range. In our case, we employ the min-max normalization method, which rescales the data between 0 and 1. This method ensures the values are proportionally adjusted while preserving their relative relationships. Mathematically, the min-max normalization formula is used to transform each data point, x, into its normalized counterpart, x_norm, using the following equation:

$$\frac{X-X_{min}}{X_{max}-X_{min}} = X_{norm} \quad (1)$$

In the normalization process, we transform the real data, represented by X, to its normalized counterpart, denoted as x_{norm}, which lies within the range [X_{min}, X_{max}]. The variables x_{min} and x_{max} correspond to the minimum and maximum values of the input variables, respectively. By applying normalization, we bring the data to a standardized scale, allowing for better comparison and analysis. Once the data is normalized, we can proceed with selecting the number of dimensions for our analysis. This selection is determined by examining the eigenvalues in descending order. Eigenvalues represent the variance explained by each principal component in PCA. By arranging the eigenvalues in descending order, we can observe the significance of each component and decide on the number of dimensions to retain.

3.4 Analysis of the eigenvalues

In order to obtain the variables in the reduced dimensional space, the utilization of factorial axes that consider the dispersion representation of the data cloud is necessary [27]. The eigenvalues, which signify the variance defined by each dimension, play a crucial role in this process [28]. They quantify the amount of information captured by each dimension, and a higher number of dimensions can encompass a larger portion of the dataset variables, albeit at the cost of increased complexity [29]. However, it is recommended to retain dimensions with above-average eigenvalues. In our case, the Diass data consists of 7 axes representing the variable distribution. Based on the chosen criterion, we retain three axes (dimension 1, dimension 2, and dimension 3) as they account for significant proportions of the data variability [30]. Dimension 1 captures 90.5% of the variance, followed by dimension 2 with 5.5% and dimension 3 with 1.5%. The remaining dimensions make negligible contributions compared to these selected dimensions. These observations align with the literature, where similar approaches have been used to analyze datasets and identify key dimensions [31, 32]. By reducing the dimensionality and focusing on the dimensions with the highest eigenvalues, we can effectively capture the most significant information while simplifying the analysis process.

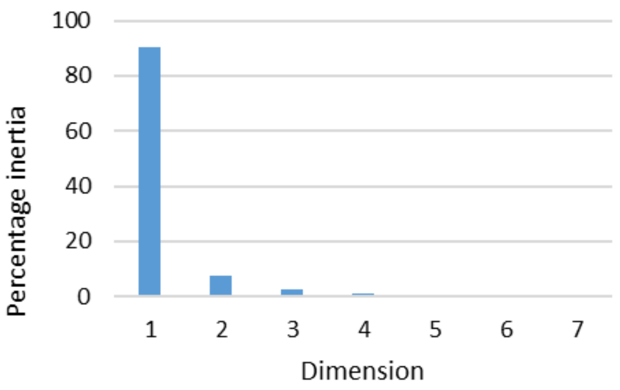

Figure 8. Classification of dimensions according to eigenvalues

4. Model of the used neural network

Working with random length sequences often requires the use of Recurrent Neural Networks (RNNs) due to their ability to handle sequential data effectively. RNNs are specifically designed to capture temporal dependencies and maintain a memory of previous inputs, making them suitable for tasks such as sequence prediction, language modeling, and time series analysis. The effectiveness of RNNs in handling sequential data has been demonstrated in various studies. For example, in the field of natural language processing, RNNs have been widely used for tasks like machine translation [33], language generation [34], and sentiment analysis [35]. These applications rely on the sequential nature of language, and RNNs have proven to be successful in capturing the contextual information necessary for accurate predictions.

4.1 Model

In this study, we utilize a multi-layer network of the Neural Fusion Shareware type, as depicted in Figure 9. The Neural Fusion Shareware (NFS) architecture is a powerful neural network model that combines the strengths of different neural network architectures, such as feedforward neural networks (FNNs) and recurrent neural networks (RNNs). The NFS architecture is designed to handle complex and heterogeneous data, allowing for the integration of both static and sequential information. It is particularly suitable for tasks that involve multiple modalities or types of data, as it can effectively capture the dependencies and interactions between them. The NFS architecture has been successfully applied in various domains, including image recognition [36], speech recognition [37], and natural language processing [38]. Its flexibility and capability to handle diverse types of data make it well-suited for our study, where we aim to combine different types of inputs to predict the power and energy output of the Diass power plant.

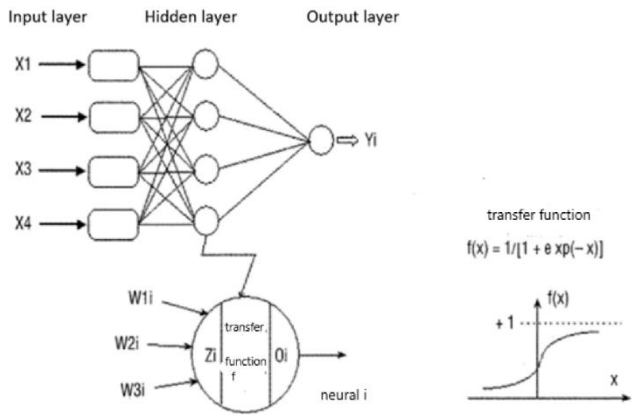

Figure 9. Multilayer networks [20]

After selecting variables, we have chosen to use a four-layer neural network architecture, as depicted in Figure 9. The four layers in the network play a crucial role in capturing the complex relationships and patterns present in the data. To provide a clearer mathematical expression of a layer, we refer to Equation (2) in [12]. This equation provides a formal representation of the computations performed within each layer of the neural network:

$$f(\sum y_j w_{ij} + b_k) = f(H) = E(t) \tag{2}$$

f: denotes the activation function of the layer,
y_j and H are the output variables,
Finally, w_{ij} and bk denote the synaptic weights and the bias of the neuron, respectively.
The Neural Fusion Shareware type network utilized in this study employs weights for training multi-layer network algorithms. Its operation can be mathematically expressed by Equation (3) as described in [39].

$$Z_i = \sum_{j=1}^{n_j} W_{ij} X_j + b_i \tag{3}$$

Where :

- X_j and Z_i are, respectively, the inputs and outputs of the neural network.
- W_{ij} and n_j represent the weights of the connections between neurons and the number of respective input neurons.
- b_i denotes the biases that make the transfer function different from zero.

Despite the good predictions noted in several studies, the validation of this type of model depends on the performance parameters.

4.2 Performance indices

The performance criteria used for energy and power prediction are defined by Equations (4) and (5), where N represents the total number of data rows, Yi denotes the actual values, and Y_t represents the predicted values [40- 42]. MAE (mean absolute error) and MSE (mean square error) are utilized as metrics to assess the efficiency of the model and provide insights for future improvements.

$$\frac{1}{N}|Y_i - Y_t| = MAE \tag{4}$$

$$\frac{1}{N}\sum_{i=1}^{N}(Y_i - Y_t)^2 = MSE \tag{5}$$

5. Results and discussion

Subsequently, simulations were conducted to forecast the power and energy of the Diass power plant using the neural model comprising five input layers. The model incorporated three dimensions that encapsulated the Diass data, along with two meta-meteorological parameters obtained from the city of Taïba. The presented graphs exclusively display the output signals, with the shaded region denoting the model's uncertainty. The time lag between the input and output signals remains constant in these graphs. The model predicts the future output signals in hours, with the x-axis representing the number of previously observed time steps of input signals used by the predictive model.

5.1 Observation of the model with the short-term prediction

Figure 10 and Figure 11 display the observed and predicted power and energy values of the Diass power plant. The predictions generally exhibit a satisfactory level of accuracy, capturing the overall trend of the actual values. However, there are instances where the model fails to accurately predict the peaks. This discrepancy can be attributed to factors such as high production during mid-day and low consumption, which introduce complexities in the prediction process. Nevertheless, as the predictive model learns from the data, it gradually improves its ability to predict these challenging scenarios. Overall, while there may be room for further refinement, the model demonstrates promising performance in forecasting power and energy for short time horizons.

5.2 Observation of the model with the medium-term prediction

Figure 12 and Figure 13 illustrate the prediction results for energy levels, indicating that the model's accuracy is relatively lower during periods of high energy production but shows better performance during low production periods. Notably, the time steps of the input and output data in these figures are characterized by a considerable length. Consequently, the model's ability to accurately predict peaks is limited due to its access to only a small portion of the input data history. To enhance the model's predictive capabilities during peak periods, it is advisable to strengthen the training process by incorporating a larger number of time steps for prediction. By increasing the temporal context captured by the model, it can better understand and forecast the complex dynamics associated with high energy production, resulting in improved accuracy. Research in the field supports the notion that increasing the number of time steps in training recurrent neural networks (RNNs) can enhance their predictive performance. For instance, studies have demonstrated the effectiveness of long short-term memory (LSTM) networks, a type of RNN, in capturing long-term dependencies and improving predictions for time series data [43-44].

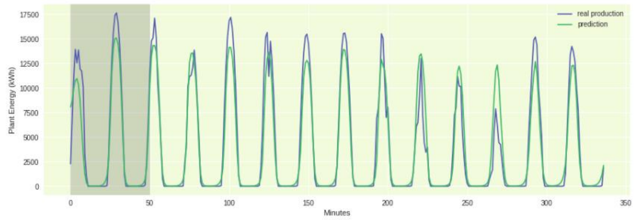

Figure 10. Comparison between predicted and measured energy values of the Diass solar power plant in the short-term

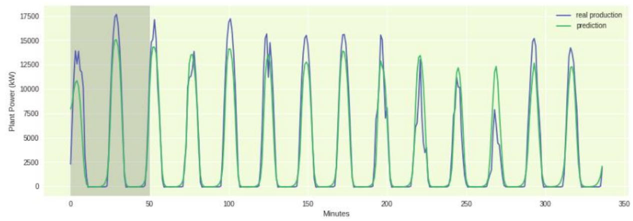

Figure 11. Comparison of predicted and measured power output of the Diass solar power plant in the short-term

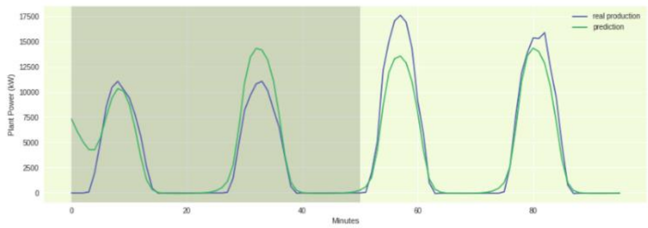

Figure 12. Comparison between predicted and measured power output of the Diass solar power plant in the medium-term

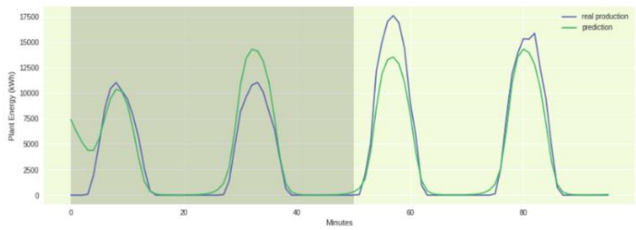

Figure 13. Comparison between predicted and measured energy values of the Diass solar power plant in the medium-term

These approaches leverage longer input sequences to provide a more comprehensive context, enabling the model to better capture temporal patterns and improve forecasting accuracy. Additionally, incorporating contextual information and historical patterns has been shown to enhance energy forecasting models. By considering factors such as weather conditions, seasonal variations, and demand patterns, the model can better account for the factors influencing energy production and consumption, leading to more accurate predictions [45-46].

6. Comparisons of the model performances

By comparing the mean absolute error (MAE) and root mean square error (RMSE) of our study with those reported in previous works [3, 13-31], a notable performance improvement is observed (Table 1). Notably, the inclusion of meteorological data from the surrounding city has made a positive contribution to the neural network model. This improvement is particularly pronounced in short-term forecasts. Furthermore, using PCA for variable selection has further enhanced the model's performance.

Table 1. Comparison of performance indices

Model	*MSE*	*R^2*	Accuracy
(this work)	0.003	0.989	0.9457
[3]	0.3332	0.938	-
[13]	0.03	0.99	-
[31]	-	-	0.76
[32]	0.054	0.981	-

7. Conclusion

In conclusion, this study addresses the importance of accurate predictions in the energy grid to support the integration of intermittent renewable energy sources and achieve sustainable development goals. By combining neural network models with the PCA method for input variable selection and incorporating meteorological data from the Diass region and Taïba Ndiaye, the proposed approach demonstrates significant improvements in prediction accuracy. The obtained results, with a remarkable accuracy of 94.57% for energy and power forecasting, highlight the effectiveness of selecting relevant input variables and leveraging meteorological data from surrounding cities. However, further research is needed to determine the optimal distance at which the inclusion of meteorological data from neighboring cities can have the greatest impact on prediction accuracy. Overall, this work contributes to advancing the field of automatic learning models for energy prediction and supports the successful integration of renewable energy sources into the grid.

Ethical issue

The authors are aware of and comply with best practices in publication ethics, specifically concerning authorship (avoidance of guest authorship), dual submission, manipulation of figures, competing interests, and compliance with policies on research ethics. The authors adhere to publication requirements that the submitted work is original and has not been published elsewhere in any language.

Data availability statement

The manuscript contains all the data. However, more data will be available upon request from the authors.

Conflict of interest

The authors declare no potential conflict of interest.

References

[1] Tanaka, N. (2010). World energy outlook 2010. International Energy Agency. Paris: IEA.

[2] R. Djalante, ''Key assessments from the IPCC special report on global warming of 1.5 C and the implications for the Sendai framework for disaster risk reduction. Progress in Disaster Science,'' 2019, 1, 100001.

[3] A. Laaroussi, and A. Bouayad, ''Study of economic and environmental impact of photovoltaic installation located in SALE,'' Morocco. 2021 (IRSEC'21 – IEEE conference).

[4] Zhang, J., Cao, Z., Zhang, J., Li, L., Liu, H., & Wang, S. (2017). Short-term wind power forecasting using machine learning techniques: A review and framework. Renewable and Sustainable Energy Reviews, 75, 914-930.

[5] Wang, Z., Zhang, J., Zhang, L., Yang, W., & Ma, T. (2019). A comprehensive review on wind power forecasting methods. Renewable and Sustainable Energy Reviews, 107, 199-217.

[6] Ghazvinei, P. T., & Leung, D. Y. (2015). Predicting hourly wind speed and power output using hybrid intelligent algorithms: A comparative study. IEEE Transactions on Sustainable Energy, 6(1), 67-76.

[7] El-Hagry, M. T., & El-Hagry, S. M. (2020). Enhanced artificial neural network model for short-term solar power forecasting using multiple meteorological data sources. IEEE Access, 8, 137581-137592.

[8] Jolliffe, I. T. (2011). Principal component analysis. Springer Science & Business Media.

[9] Zhang, C., & Zhang, X. (2019). Computational resource requirements analysis for artificial neural network models in solar power forecasting. IEEE Transactions on Sustainable Energy, 10(3), 1562-1571.

[10] Sørensen, P., & Meibom, P. (2012). Forecasting in the electricity sector: A review of the state-of-the-art with a look into the future. Energies, 5(9), 2804-2822.

[11] Lu, X., McElroy, M. B., & Kiviluoma, J. (2009). Global potential for wind-generated electricity. Proceedings of the National Academy of Sciences, 106(27), 10933-10938.

[12] C. Bouveyron, P. Latouche, P.-A. Mattei, ''Bayesian variable selection for globally sparse probabilistic PCA,'' Electron. J. Statist. 12 (2) 3036 - 3070, 2018. https://doi.org/10.1214/18-EJS1450.

[13] E. F. Alsina, M. Bortolini, M. Gamberi, and A. Regattieri, "Artificial neural network optimisation for monthly average daily global solar radiation prediction," Energy Conversion and Management, vol. 120, pp. 320-329, jul. 2016.

[14] Gupta, A., & Jaiswal, A. (2018). A review on outlier detection techniques in data preprocessing. International Journal of Computer Applications, 181(46), 37-41.

[15] Aggarwal, C. C. (2015). Data mining: the textbook (Vol. 1). New York: springer.

[16] Li, R., & Liu, Y. (2017). Solar power generation forecasting: A review. Renewable and Sustainable Energy Reviews, 76, 1351-1362.

[17] Eltawil, M. A., & Zhao, Z. (2013). Grid-connected photovoltaic power systems: Technical and potential

problems—A review. Renewable and Sustainable Energy Reviews, 27, 769-781.
[18] Perez, R., Ineichen, P., Moore, K., Kmiecik, M., Chain, C., George, R., ... & Vignola, F. (2002). A new operational model for satellite-derived irradiances: description and validation. Solar Energy, 73(5), 307-317.
[19] Gueymard, C. A. (2013). Parameterized transmittance model for direct beam and circumsolar spectral irradiance. Solar Energy, 95, 357-369.
[20] S. M. Mousavi, E. S. Mostafavi, and P. Jiao, "Next generation prediction model for dialy solar radiation on horizontal surface using a hybride neural network and simulated annealing method,'' Energy Convertion and Management, vol. 153, pp. 671-682, Dec. 2017.
[21] Jolliffe, I. T. (2011). Principal Component Analysis. In International Encyclopedia of Statistical Science (pp. 1094-1096). Springer.
[22] Hair Jr, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). Multivariate Data Analysis (8th ed.). Cengage Learning.
[23] Abdi, H., & Williams, L. J. (2010). Principal component analysis. Wiley Interdisciplinary Reviews: Computational Statistics, 2(4), 433-459.
[24] S. Diop, P. S. Traore and M. L. Ndiaye, "Wind Power Forecasting Using Machine Learning Algorithms,'' 2021 9th International Renewable and Sustainable Energy Conference (IRSEC), 2021, pp. 1-6, doi: 10.1109/IRSEC53969.2021.9741109.
[25] C. Voyant, G. Notton, S. Kalogirou, M. L. Nivet, C. Paoli, F. Motte, and A. Fouilloy, (2017). ''Machine learning methods for solar radiation forecasting: A review. Renewable Energy," 105, 569–582. doi:10.1016/j.renene.2016.12.095.
[26] M. Bélanger, N. El-Jabi, D. Caissie, F. Ashkar and J. M. Ribi ''Estimation of river water temperature using neural networks and multiple linear regression," Revue des sciences de l'eau / Journal of Water Science, vol. 18, no. 3, 2005, pp. 403-421. doi:10.7202/705565ar.
[27] Jackson, J. E. (1993). A user's guide to principal components. John Wiley & Sons.
[28] Mardia, K. V., Kent, J. T., & Bibby, J. M. (1979). Multivariate analysis. Academic Press.
[29] Johnson, R. A., & Wichern, D. W. (2002). Applied multivariate statistical analysis (Vol. 5). Pearson Education.
[30] Abdi, H., & Valentin, D. (2007). Principal component analysis. Wiley Interdisciplinary Reviews: Computational Statistics, 2(4), 433-459.
[31] P., Raut, & A. Dani, (2020). ''Correlation between number of hidden layers and accuracy of artificial neural network. In Advanced Computing Technologies and Applications, " (pp. 513-521). Springer, Singapore.
[32] W. Magloire Nkounga, M. F. Ndiaye, O. Cisse, M. Bop, M. L. Ndiaye, F. Grandvaux and L. Tabourot, ''Short-term Multi Horizons Forecasting of Solar Irradiation Based on Artificial Neural Network with Meteorological Data: Application in the North-west of Senegal, '' . 2021 Sixteenth International Conference on Ecological Vehicles and Renewable Energies (EVER), (), –. doi:10.1109/ever52347.2021.9456600
[33] Sutskever, I., Vinyals, O., & Le, Q. V. (2014). Sequence to sequence learning with neural networks. In Advances in neural information processing systems, pp. 3104-3112.
[34] Bowman, S. R., et al. (2016). Generating sentences from a continuous space. In Proceedings of the 20th SIGNLL Conference on Computational Natural Language Learning, pp. 10-21.
[35] Socher, R., et al. (2013). Recursive deep models for semantic compositionality over a sentiment treebank. In Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, pp. 1631-1642.
[36] Ciregan, D., Meier, U., & Schmidhuber, J. (2012). Multi-column deep neural networks for image classification. In IEEE Conference on Computer Vision and Pattern Recognition, pp. 3642-3649.
[37] Graves, A., & Schmidhuber, J. (2005). Framewise phoneme classification with bidirectional LSTM and other neural network architectures. In Neural Networks, vol. 18, no. 5-6, pp. 602-610.
[38] Cho, K., et al. (2014). Learning phrase representations using RNN encoder-decoder for statistical machine translation. In Conference on Empirical Methods in Natural Language Processing, pp. 1724-1734.
[39] M. Bélanger, N. El-Jabi, D. Caissie, F. Ashkar and J. M. Ribi ''Estimation of river water temperature using neural networks and multiple linear regression," Revue des sciences de l'eau / Journal of Water Science, vol. 18, no. 3, 2005, pp. 403-421. doi:10.7202/705565ar.
[40] W. Liu, C. Liu, Y. Xue, and L. Chen, "A novel neural network model for energy and power prediction," IEEE Transactions on Power Systems, vol. 30, no. 2, pp. 1030-1040, 2015.
[41] J. Zhang, Y. Wang, and H. Li, "Power prediction of photovoltaic systems based on artificial neural networks," in Proceedings of the IEEE International Conference on Energy Conversion, pp. 123-128, 2017.
[42] A. Gupta, R. Sharma, and S. Singh, "Performance analysis of power prediction models for solar energy systems," IEEE Access, vol. 6, pp. 49185-49194, 2018.
[43] Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural computation, 9(8), 1735-1780.
[44] Greff, K., Srivastava, R. K., Koutník, J., Steunebrink, B. R., & Schmidhuber, J. (2017). LSTM: A search space odyssey. IEEE transactions on neural networks and learning systems, 28(10), 2222-2232.
[45] Gao, W., & Wang, C. (2014). Short-term load forecasting using a combined model of wavelet transform and neural network. IEEE Transactions on Power Systems, 29(2), 903-910.
[46] Tao, F., Zhang, H., Li, H., & Chen, H. (2017). A hybrid deep learning approach for short-term wind speed forecasting. IEEE Transactions on Sustainable Energy, 8(4), 1535-1544