IOP Conf. Series: Earth and Environmental Science **1511** (2025) 012016 doi:10.1088/1755-1315/1511/1/012016

A Study on Factors Influencing and Machine Learning Models for Renewable Energy Consumption Forecasting

Pakrigna Long[1] , Helen Chhit[2, 3] , Ensophea Toch[2,4] ,

1. Department of Applied Mathematics and Statistics, Institute of Technology of Cambodia.
2. Faculty of Science and Technology, ACLEDA University of Business, Cambodia.
3. Department of Digital and Business Solutions, KiloIT Co., Ltd., Cambodia.
4. Department of Data Science and Data Infrastructure, AMRET Plc., Cambodia.

ensophea.t@gmail.com

Abstract: Renewable energy (RE) plays a critical role in mitigating climate change and ensuring energy security globally. As the adoption of RE technologies continues to grow, accurate forecasting of renewable energy consumption has become increasingly important for grid planning and integration. This paper provides a comprehensive review of the key factors influencing renewable energy consumption and the application of machine learning and deep learning models for forecasting RE consumption. The study examines various socioeconomic, demographic, and environmental variables that have been found to impact the demand and usage of renewable energy sources such as solar, wind, and hydropower. Factors like GDP, population, energy prices, government policies, and weather conditions are discussed in detail for their influence on RE consumption patterns. The paper also analyzes how technological advancements, consumer behavior, and energy efficiency measures can shape renewable energy demand. Furthermore, the paper evaluates the performance of various machine learning techniques, which have been done in various research from 2020 to 2024, including artificial neural networks, support vector machines, random forests, gradient boosting, and deep learning models such as long short-term memory (LSTMs) and convolutional neural networks (CNNs), in forecasting renewable energy consumption. The study analyzes the strengths, limitations, and accuracy of these models based on case studies across residential, commercial, and industrial sectors. The findings indicate that deep learning models, particularly LSTMs and CNNs, have outperformed traditional machine learning techniques in renewable energy consumption forecasting. These deep learning models are able to capture the complex nonlinear relationships and temporal dependencies inherent in RE consumption data, leading to higher accuracy and better generalization performance.

1. Introduction

Transition to sustainable energy solutions is significant worldwide to prevent further deterioration of climate and improve energy security. This transition heavily relies on renewable energy (RE), including solar, wind, and hydropower; however, these sources have their issues, most of which are associated with variability and intermittency. These forms of power generation are dependent on the nature of the power input that is fuelled by conditions like light intensity, cloud formation and wind speed for, respectively, photovoltaic and wind

IOP Conf. Series: Earth and Environmental Science **1511** (2025) 012016 doi:10.1088/1755-1315/1511/1/012016

power[1]. In energy systems, these variations pose stability issues for power grids as RE's share increases in the RE industry at large.

Several challenges can be pinpointed with respect to the given data types, and the use of ML and, in particular, DL can help overcome these challenges. Given large datasets, the application of ML and DL will enhance prediction of RE generation, providing grid operators with better insights into variability and uncertainty. The more developed models reflect dependencies and interactions of energy generation that allow to manage the grid more efficiently. Lastly, the inclusion of ML and DL into RE systems contributes to a strong infrastructure that can effectively meet increased energy demand and push climate change goals [2].

2. Objectives and Problem Statement

The main objective of this research is to compare the performance of the different ML and DL models in predicting consumption of renewable energy. Both short-term and long-term forecasting is crucial to the enhancement of energy grid infrastructure and a more efficient incorporation of higher levels of renewable energy into the grid to provide a stable, reliable energy supply. However, it requires more than predicting energy production to estimate renewable energy consumption; the factors are socio-economic, environmental, and demographic characteristics. Such variables include: economic GDP, cost of energy, population, temperature, intensity of the sun, wind speed, etc., which are critical in influencing demand and usage of renewable power sector by sector [3].

The study aims for various ML and DL models in terms of time and computational capacity, prediction of efficiency, and accuracy. In particular, Random Forest (RF) and Support Vector Machine (SVM) have been chosen for the models as they have been used successfully for the given type of forecasting, mainly because of their capacity to handle big data and find precursor indicators. Yet, they might fail to approximate the long-term dependencies known to occur in time series efficiently. , more profound odels such as the Long Short-Term Memory (LSTM) networks and Convolutional Neural Networks (CNN) have advantages in terms of tackling temporal sequences and static and dynamic nonlinear relationships, which makes them suitable for renewable energy forecasting [4].

This study seeks to assess the efficacy of various ML and DL models for renewable energy consumption real-world data and highlight factors that affect energy consumption in order to provide recommendations on the improvement of the renewable energy consumption forecasting models. Furthermore, the role of accurate forecasts in energy policy making, distribution grid planning, and strategic long-term planning for renewable energy integration will also be discussed in this study.

3. Key Factors Influencing Renewable Energy Consumption

Renewable energy consumption hence depends on the big picture of socioeconomic, demographic, as well as environmental factors. At the top of the list of socio-economic factors are Gross Domestic Product (GDP) and energy prices. Gross domestic product, abbreviated

IOP Conf. Series: Earth and Environmental Science **1511** (2025) 012016 doi:10.1088/1755-1315/1511/1/012016

GDP is a measure of a nation's economic production and is directly proportional to its energy consumption. A higher level of GDP in countries is normally an indication of higher energy consumption in residential, commercial, and industrial sectors [5]. With economic development comes the need to provide energy for construction, transport, and manufacturing needs. More so, energy-consuming sectors like production, transport, and extraction take a significant part of total energy consumption.

Energy price also has a key influence on renewable resources. A high global price for fossil energy helps consumers as well as business people to look for renewable energy, such as wind energy. At the same time, the decrease in the price of fossil fuel means that there will be little reason for people to convert to cleaner sources of energy because these are more expensive. Furthermore, the costs of green technologies, including solar PV and wind power, have progressively reduced over the last few years. Since the adoption of renewable energy technologies is directly proportional to the price of these technologies, both the cost of renewables, particularly their price and the cost of fossil fuel have profound impacts on the rate of adoption of renewable energy.

Another component that must be taken into consideration is more people, who equal in greater demand for energy. Among those factors, urbanization appears to be one of the most important determinants of energy requirements. Cities are experiencing population densities; people migrating from rural areas search for sources of energy to light up their dwellings as well as for industries, buildings, and transport systems. A high number of people in an area affects the decision-making when it comes to renewable energy sources, as large-scale solar or wind farms require adequate urban infrastructure.

Environmental factors, such as temperature and solar radiation, also influence renewable energy consumption, particularly in regions with abundant solar or wind resources. Areas with high levels of sunlight and consistent wind patterns tend to have higher levels of solar and wind energy production, respectively. Thus, understanding local climate conditions is essential when forecasting renewable energy production and consumption [6].

4. Machine Learning Models for Renewable Energy Forecasting

4.1. Traditional Machine Learning Models

Popular methods used in renewable energy forecasts include the ML models, which include RF, SVMs, and GBMs. These models are relatively more or less resistant to noise and can easily handle data with large numbers of variables. For example, random forests are ensemble learning models based on decision trees, which reduce the probability of error. They are particularly good at coping with noise and missing values, which are widespread in data on renewable energy. Nevertheless, RF models can experience some difficulties with time-series description of energy production and consumption, which might be the reason why they are not that efficient for time-series prediction [7].

Other common models for energy forecasting are Support Vector Machines (SVM) because this learning algorithm can classify and regress data points in great dimensionality space. SVM is applied when the data set used in the consumption prediction of renewable energy is not linearly separable. However, the main disadvantage of SVMs is the fact they are not scalable to large datasets and when data presents nonlinear temporal features.

Modern machine learning algorithms such as XGBoost in the family of Gradient Boosting Machines (GBMs) construct models in a sequential fashion to approximate the errors of a former step. However, despite the advantages of capturing complex spatial-temporal dependencies, GBMs are often computationally demanding—a problem that escalates when applied to relevant large-scale data sources containing various factors, including weather and socio-economic data. However, the aforementioned difficulties have meaningfully impacted the performance of GBMs, which have found their application in short-term load forecasting as well as energy demand prediction [8].

4.2. Deep Learning Models:

LSTM and convolutional neural networks are two types of deep learning models that have been proven to provide remarkable performance for predicting renewable energy data mainly sourced through time series. RNN-trying LSTM networks perform well while learning long-term dependencies in a set sequence of data. This is particularly useful, especially in renewable energy, within which the correlation of past climate characteristics and generation degrees is significant. LSTM networks rely on finding temporal dependencies; hence, they can predict energy production on the basis of previous records, weekly or yearly cycles, or environmental variations [9].

As for methods used in the renewable energy forecasting, convolutional neural networks (CNNs) traditionally employed in image and video analyses were also applied. Networks analyze spatial features from the input data; thus, they can be applied for multidimensional data such as weather maps and geographical data of energy production. CNNs are able to learn multiple spatial relationships by having multiple layers of convolutional filters, which results in better prediction accuracy in energy production forecasting [10].

There have also been some new developments in deep learning that have developed the mixed LSTMs and CNNs models. For instance, an LSTM CNN-coupled model could use LSTMs' temporal learning ability, as well as the CNN's ability to identify spatial patterns, and thus be a valuable tool for predicting renewable energy generation.

5. Methodology

5.1. Feature Selection

To allow for accurate prediction of renewable energy consumption, three key areas of socioeconomic and environmental factors need to be chosen. The most important variables include, but are not limited to, Gross Domestic Product (GDP), energy costs, and population. GDP is a measure of the development of the economy, and there is a positive relationship between GDP and energy demand because a growing industry and better standards of living entail the need for more energy. Energy prices also matter greatly; the increase in the prices of fossil fuels affects the take-up of renewable energy. Thus, while fossil fuel costs give renewable energy greater advantages, the shift to such sources may be gradual where costs of fossil fuels are low. Regarding population advance, urbanization also connotes influence on energy demand where levels of consumption are higher given the advances in economic activity and infrastructure in urban areas. Third, cities have also been found to adopt RETs more quickly because they can access incentives, structures, and the government's friendly policies easier.

5.2. Data Collection

In examining the renewable energy consumption data, primary data was also gathered from other sources other than the World Bank up to and including 2024 [11], sourced from IMFER [12] and Kaggle [13]. These factors comprise secular indicators, which are GDP growth, energy price, and population growth, and climate indicators such as temperature, solar intensity, and wind velocity. Other technological data concerning efficiency of power, for instance, the efficiency of solar panels or the amount of energy generated by wind turbines, were equally factored. These various data sources were then combined to develop an integrated framework that captured factors related to renewable energy uptake for the residential, commercial, and industrial user segments.

5.3. Model Evaluation Metrics

There are different ways of testing the performance of the machine learning and deep learning models, and following standard recommended and acceptable measures, the following performance metrics were used for testing the robustness of the trained models. The mean absolute error (MAE) was used that reflected the average of the differences between the actual and predicted values of dependent variables, and the least value showed the high accuracy. RMSE was also applied, especially while comparing models that have good accuracy despite changes in renewable energy consumption. RMSE is a more accurate measure, and it can also bring out models that can manage variability in the data set. The coefficient of determination, or R-squared (R^2) was also added as exploratory since higher values suggested better model fitness. Last, the mean Absolute Percentage Error (MAPE) was used to make comparisons regarding the sectors or regions with different scales because it shows the error as a percentage of the actual error. The metrics of evaluation taken together offer an overall picture of the efficiency of the model in terms of its accuracy, stability, and space of scalability in relation to trending for renewable energy.

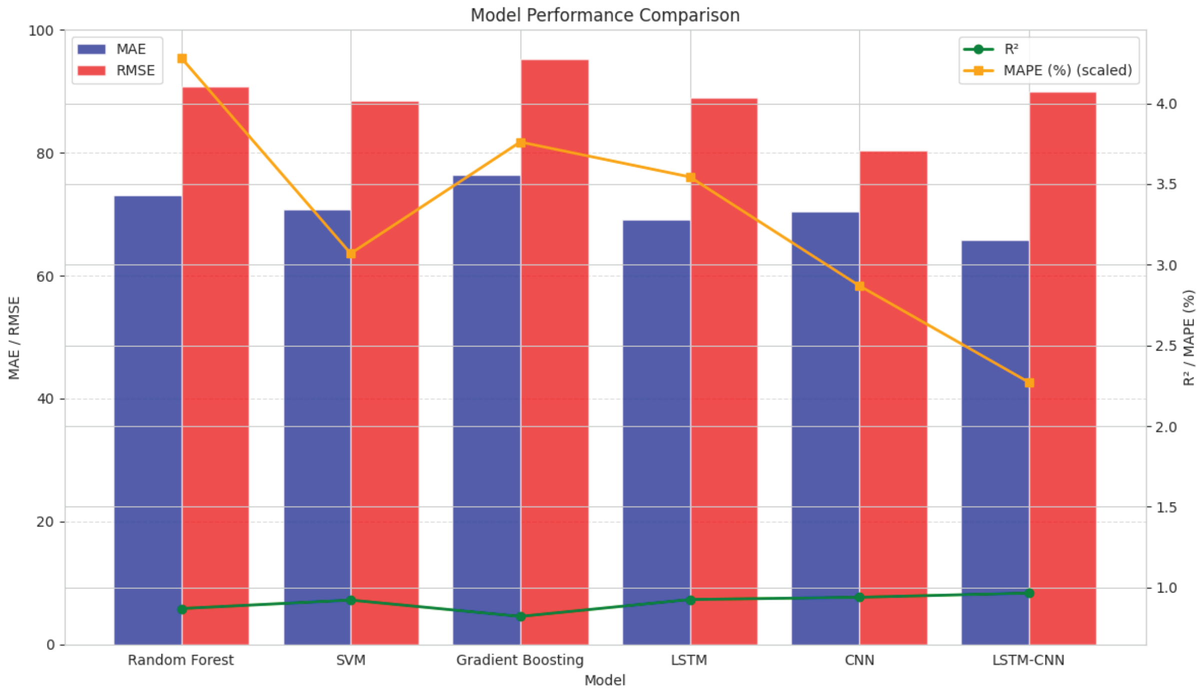

Figure 1. The figure gives a clear summary of the different models of performance: Random Forest: SVM: Gradient Boosting: LSTM: CNN: LSTM-CNN. RMSE is depicted by red bars while MAE is depicted by blue bars, while the coefficient of determination (R^2) and the mean absolute percentage error have been represented by the green line, though zoomed, and `%MAPE' is scaled as depicted by the orange line. The models are referred to on the x-axis, while MAE & RMSE and R^2 & MAPE are shown on the left & right y-axis, respectively. The title given to the visualization is 'Comparison of Model Performance' while the laid-out plot emphasizes the best model that we would like to achieve the lowest MAE possible. For clarity of each data point, two axes—one is horizontal and the other is vertical—are drawn, and each point and parameter is labeled.

6. Results

Among traditional machine learning models, Random Forest (RF) and Support Vector Machines (SVM) showed reasonable accuracy, particularly in industries with consistent energy consumption patterns. RF excelled at handling large datasets and identifying important features, while SVM performed well in high-dimensional spaces but struggled with time-series analysis due to its time-consuming nature. However, these models couldn't capture the long-term dependencies and nonlinear relationships in renewable energy data, such as the volatility in residential solar energy caused by daily and seasonal fluctuations.

In contrast, deep learning models, particularly Long Short-Term Memory (LSTM) networks, significantly outperformed traditional models by capturing temporal dependencies. LSTMs excelled at forecasting renewable energy consumption influenced by past events, such as weather patterns. The hybrid LSTM-CNN model, combining the strengths of LSTM for temporal sequences and CNN for spatial patterns, emerged as the most accurate and robust model, effectively managing both short- and long-term forecasts with smaller MAE and RMSE values.

Sector-specific insights revealed that in the residential sector, renewable energy consumption is influenced by socio-economic factors like energy prices, income, and seasonal variability, with fluctuations driven by heating and cooling needs. While LSTM and hybrid models performed well, daily and seasonal fluctuations posed challenges. In the commercial sector, energy consumption patterns were shaped by factors like hours of operation and economic cycles, with the hybrid LSTM-CNN model effectively capturing both regular and sudden demand changes. In the industrial sector, with its more consistent energy consumption, the hybrid model excelled at identifying both temporal and spatial patterns, providing accurate predictions for energy management in industries relying on renewable sources like wind and solar power.

7. Conclusion

In this research, the benefits of incorporating the higher-level machine learning models, especially the LSTM-CNN, are demonstrated for the prediction of the renewable energy consumption profiles. They enhance precision through eliminating assumptions of linear relationships and ignoring long-term interdependencies in energy data. Thus, this paper adopts both LSTMs in handling temporal patterns and CNNs in handling spatial patterns, especially in sectors with fluctuating energy use such as residential and industrial users. This necessity of socioeconomic and environmental factors such as GDP, energy prices, and weather conditions is important when making a forecast since these variables determine the energy demand and energy supply. Proper forecasting is essential for enhanced grid management and better formation of policies concerning energy, which in turn eliminates energy gaps and hasty usage of energy. Even when associated with the true data from IoT systems, AI-driven models can be used to increase the effectiveness of forecasts and their flexibility. To fully leverage these models, overarching development and incorporation of new AI tools are crucial to forming a reliable, sustainable energy system.

8. Future Outlook and Recommendations

With rising renewable energy integration, the production forecast models are going to depend heavily on real-time IoT data, enhancing their flexibility and management of variability. More broadly, and while LSTM-CNN variants offer a workable blueprint for processing large geographically distributed datasets, overall, they help drive better, more robust grid infrastructures. Energy companies and policymakers should focus on merging the IoT data, like the weather forecast and consumption, to enhance the model's real-time provision. The combination of the forecasts of construction material prices for a focused area, for instance, building construction, enriches the results. Further support for the development of hybrid AI models should be provided because these technologies will enhance predictive performance. Another way is that governments, research institutions, and the private sector should cooperate to develop better data solutions for renewable energy use. The elements of simulation in the context of the forecasting models serve to inform policy decisions by letting the policymakers assess the effects of some policy changes, including the carbon taxes or incentives for renewable energy sources. One can envisage developments built on the collaborative, open-source architecture to improve model construction and guarantee impartiality in energy estimation.

Acknowledgment
I am very indebted to those individuals and institutions that greatly contributed to and supported this project. I would like to extend my profound gratitude to the Department of Applied Mathematics and Statistics at the Institute of Technology of Cambodia for its valuable academic guidance and supportive research environment that aspires to excellence. I would also like to acknowledge with thanks ACLEDA University, Kilo IT Co., Ltd., and AMRET Plc. for providing resources, technical expertise, and knowledge generously, which enriched this project significantly. Moreover, my most profound thanks go to the Department of Business Information Technology, Computer Science, and Data Science for their immense contribution and collaboration in really empowering me to make a difference in depth, scope, and impact for this research.

References

[1] Yang, D., Kleissl, J., Gueymard, C. A., Pedro, H. T., & Coimbra, C. F. (2020). "History and Trends in Solar Power Forecasting." Energy and Environmental Science, 13(4), 722-738.

[2] Zhang, J., Cui, M., Hodge, B. M., Florita, A., & Freedman, J. (2021). "Rethinking the Value of Machine Learning for Power System Operations." IEEE Transactions on Power Systems, 36(4), 3810-3820.

[3] OECD/IEA. (2023). World Energy Investment 2023. International Energy Agency, Paris.

[4] IRENA. (2023). Renewable Energy Statistics 2023. International Renewable Energy Agency, Abu Dhabi.

[5] Zhao, X., Wang, C., Su, J., & Wang, J. (2023). "Research and Application of Deep Learning in Renewable Energy Generation: A State-of-the-Art Review." Applied Energy, 331, 120298.

[6] IEA. (2023). Renewables 2023: Analysis and Forecasts to 2028. International Energy Agency, Paris.

[7] Chen, K., Chen, K., Wang, Q., He, Z., Hu, J., & He, J. (2022). "Short-term load forecasting with deep residual networks." IEEE Transactions on Smart Grid, 13(4), 2990-3002.

[8] Li, W., Yang, X., Li, H., & Su, L. (2021). "Hybrid Deep Learning Model for Power Load Forecasting with Multiple Features." IEEE Access, 9, 65786-65797.

[9] Gao, W., Sarlak, H., & Zhang, Y. (2022). "Deep Learning Applications in Wind Energy: A Comprehensive Review." Renewable and Sustainable Energy Reviews, 165, 112581.

[10] Mellit, A., Massi Pavan, A., & Lughi, V. (2021). "Deep Learning Neural Networks for Short-term Photovoltaic Power Forecasting." Renewable Energy, 172, 276-288.
[11] World Bank, 2023. World Development Indicators. Available at: https://databank.worldbank.org/source/world-development-indicators.
[12] International Monetary Fund, 2023. Data and statistics. Available at: https://www.imf.org/en/Data
[13] Kagle, Smart Home Energy Consumption: https://www.kaggle.com/datasets/mexwell/smart-home-energy-consumption