Journal of Electrical Engineering & Technology
https://doi.org/10.1007/s42835-025-02430-z

ORIGINAL ARTICLE

Enhancing Short-Term Electricity Forecasting with Advanced Machine Learning Techniques

Di Cui[1] · Ambe Harrison[9] · Emmanuel Fendzi-Donfack[4] · Idriss Dagal[5] · Pradeep Jangir[6,7,8] · Mohamed Metwally Mahmoud[2,10] · Pankaj Malkani[11] · Wulfran Fendzi Mbasso[3] · Pierre Tiako[12] · Aseel Smerat[13]

Received: 28 December 2024 / Revised: 13 May 2025 / Accepted: 26 August 2025

Abstract

This research proposes new electricity forecasting integration with Bayesian Optimized Attention-Dilated LSTM and Savitzky-Golay. Its hybrid model is aimed at enhancing the accuracy of electricity generation and consumption forecasting in developing areas. Detailed preprocessing techniques such as handling missing values through interpolation, removing outliers via Z-scores, normalization, and noise reduction with Savitzky-Golay filters significantly improved forecasting accuracy. After implementing data preprocessing using the Bayesian optimized Savitzky-Golay filter, several deep learning models such as LSTM and attention-dilated Bi-LSTM are trained. The developed model showed notable improvements over state-of-the-art techniques, yielding sMAPE values of 2.4% and 2.8% for production and consumption respectively, alongside nRMSE values of 3.1% and 3.6%. These outcomes prove the model's adaptability to real world data changes and its practicality for improving energy management and planning. This research will assist in developing newer forecasting models in regions facing similar issues pertaining to energy consumption.

Keywords Short-term electricity forecasting · Savitzky-Golay filter · Bayesian optimization · Machine learning · Deep learning · Renewable energy integration

1 Introduction

1.1 Background of the Study and Relative Works

The worldwide increase of the population has contributed in the rise of demand for energy, electricity in particular to a notable degree [1]. Difficulties with electricity consumption and demand arise for both developed and undeveloped nations according to [2]. In developing countries, such as Cameroon, there is an increasing need for sophisticated and reliable energy forecasting models to control grid frequency and improve energy efficiency. The need for robust models is set by the international requirements identified by the project of the intelligent electrical substation operational standards IEC 61,850 and the standards of electromagnetic compatibility of electrical equipment IEC 61,000. These standards lay down the basic principles for the implementation of renewable energy systems and the effective functioning of forecasting devices integrated into smart grids, which makes them highly automated. The integration of renewable sources of energy, particularly wind and solar, contribute to providing clean energy that facilitates industrial and economic advancement. The entire world is in the continuous process of changing from conventional energy towards renewable energy sources, and this change is being done in a strategic and urgent manner while considering the scarcity of non-renewable sources of energy and the extreme impact on the environment by fossil fuels [3]. The ability of a community to prosper is significantly decided by its access to and use of electricity. Thus, it is important for every country to formulate specific policies for planning the development of energy generation in order to cope with rising consumption levels [4]. International cooperation undertakings aim towards providing controlled solutions to issues of poverty and global security while increasing living conditions of the people. The strategy is to target the provision of energy resources and other services in a sustained, reliable and efficient manner. Keeping in mind the profound significance

Extended author information available on the last page of the article

Published online: 04 September 2025

of energy in all facets of humanity, it is unquestionable that energy can assist in resolving the many challenges the world faces today [5–7]. One of the primary characteristics of the least developed countries in the world is their disadvantageous location, as indicated by a very low rate of electrification and the more recent focus of economic activities within the countries. As noted from [8], in ENEO, the thermal sector accounts for only 23% of power generation, while dominance is held by hydroelectricity which surpasses 70%. On the contrary, the use of renewable energy sources has not yet gained tangible ground in the country. Even though renewable sources account for just 1% of the electricity supplied to the country, the government has started several programs aimed at increasing this percentage [9]. Following this paradigm, there has been considerable research on the electric power generation and consumption within one least developed country which shows considerable interest about it. The national interest has to be given priority because it is of great importance in adopting more efficient development projects. In the initial stages, the power grids concentrated their efforts mainly on the processes of energy production, transmission, and distribution and the energy supply was below demand. A comprehensive analysis into electricity generation and consumption patterns becomes critical for the implementation of energy management measures, which seeks to strike the delicate equilibrium between energy demand and its supply, especially with regard to the current developments of the nation through various initiatives.

An artificial intelligence model emphasizing the use of neural networks for the prediction of electricity consumption in Lagos State, Nigeria was formulated by [10]. This model was proven to have better accuracy results than other models. The research done by [11] studied both traditional models as well as artificial intelligence models, which included time series (TS) models, regression models, and grid models. Artificial intelligence (AI) models introduced are: Artificial Neural Network (ANN) models, Support Vector Machine (SVM), and Random Forest (RF) models. In estimating and forecasting electrical energy consumption [12], used genetic optimized programming over a long period to analyze the dynamics between historical data and electric energy consumption. The approach was tested against other methodologies which included neural networks, SVM, Adaptive Neuro-Fuzzy Inference System (ANFIS), and the cuckoo search algorithm. There are many hybrid models which have been proposed in the literature. A new model presented by [13] is that of machine learning with speed and accuracy for predicting electrical energy consumption within a smart grid environment. The proposed forecasting algorithm is based on deep learning with a construct of the linear rectified unit. The hybrid model was tested and evaluated using data of the power grid in the United States of America, using three criteria: the average percentage, variance, correlation coefficient, and convergence rate. In the same way [14], suggested the use of a hybrid learning neural network model for forecasting building electricity consumption that integrated neural networks developed for the United States and China's hourly electricity consumption prediction. In their work [15], developed a hybrid random drill model with a multilayer perceptron for daily energy consumption prediction at the university campus level [16]. reported on a household electricity consumption prediction model within a smart grid using a Grey-ANFIS-PSO hybrid model, which was shown to have better accuracy than the other models.

The optimization of hyperparameters in the machine learning models has been greatly simplified through the use of Bayesian optimization, which also enhances performance and minimizes the cost of computation. The more advanced methods of modeling and forecasting formed integrated approaches for actualizing and predicting the electricity generation and consumption utilizing the artificial intelligence paradigm, which has relievdaly transformed the energy industry. Considering the sophisticated techniques for data parsing and model building described above, the quality of the forecasts has improved significantly, which is important for energy management and stability of the grid. As noted in 17, Bayesian optimization is effective for hyperparameter tuning and it meets the primary goal of improving performance and decreasing the computational expenses. The use of hybrid models with Bayesian optimization for long-term solar power prediction was addressed in [17] resulting in improved forecasts. In [18] the authors showcased the applied optimization methods of Bayesian estimation in improving the forecasts of PV power generation by PV regression ensemble models constructed with a system of framed relays. In the same way [19], showed the possibilities of fuzzy logic for MPPT PV systems prediction using Bayesian optimization techniques for their controllers in tropical regions. These studies underline the need for strong hyperparameter tuning in order to obtain proper and precise models for forecasting. The automata based on artificial intelligence, mainly deep learning models, excellently realize and remember the temporal dependencies supplied to them and the complex of data patterns emitted with the token LSTM and its derivatives [20]. deployed machine learning models for short-term forecasting of solar PV power generation to illustrate the value of deep learning capabilities towards predicting energy yield needed for smart grids. Work by [21] also showed how machine learning could lessen the chances of having erroneous forecasts in PV farms and stressed the need to resort to more sophisticated models for proper energy prediction. The application of different machine learning techniques to develop hybrid models has been successful in solving the problems of accuracy and reliability of the forecasts [22]. proposed an original approach to solar power prediction based on feature clustering with a hybrid classification-regression model, thereby increasing the accuracy of the forecasts. This approach proves to be very effective in dealing with the variability and uncertainty associated with renewable energy sources.

The ability to forecast has increased tremendously with the aid of deep learning [23]. demonstrates the use of the advanced techniques of deep learning for the prediction of solar irradiation and wind speed one day ahead. Their results showed a higher degree of accuracy when measured against traditional forecasting techniques. These studies show deep learning's capability in offering high quality forecasts for renewable energy systems. Efficient data preprocessing, which consists of signal filtering and data smoothing, is necessary to enhance the forecast model accuracy by eliminating unwanted variance and capturing the important features of time series data. With regard to smoothing time-series data while capturing important trends, the Savitzky-Golay filter is the most widely employed. The application of machine learning methods has shown considerable promise in tropical regions like Cameroon which experience high levels of weather variability and its impact on renewabe energy generation. Reliable forecasting is of utmost importance in situations where grid stability and energy demand need modulation. Regions such as these would benefit from the use of deep learning and hybrid models, as these advance techniques augments forecast accuracy. The use of new machine learning methods, thorough data preprocessing, and complex hybrid models form a cohesive structure with which the accuracy of short-term electricity forecasting can be improved. These studies have addressed gaps in the literature and provided suggestions concerning the application of energy management and grid optimization which is useful for the reviewed studies. Even though power systems around the word face a stunning magnitude of difficulties and demanding changes, the need for dependable energy forecasting forecasts the evolution of the complex power system in the future.

1.2 Research Gaps and Main Contributions

While techniques like LSTM and ANN have made leaps in forecasting electricity usage, there are still gaps in existing studies which restrict their practicality. One of them still remains to be the insufficient preprocessing of noisy real-world datasets. A great number of models do not apply adequate noise reduction strategies, leading to poor forecasting, especially in variable scenarios. Also, traditional forecasting methods tend to be non-aggregative, making them less useful in different regions around the world having diverse datasets. This diminishes the value of these methods in large-scale energy management systems where the ability to work across multiple contexts is crucial. Another important gap in the work done is the lack of focus given to some of the developing areas like Cameroon which have unique energy challenges that go beyond just the accuracy of the forecast. Factors such as grid unreliability, volatile weather, and poor electrification rates combine to create an environment where standard forecasting models risk underperformance. Most of the available literature concentrates on well-established power infrastructures and the developing contexts remain methodologically strayed. In addition, even though hybrid models stand to significantly boost precision and resilience of a system, there is little to no research focusing on their application with modern techniques such as Bayesian optimization, attention mechanisms, and dilated convolutional layers all combined under a single framework. Addressing such shortcomings is pivotal in constructing electricity forecasting models that strive for precise accuracy while seamlessly adapting to the intricate dynamics of real-world energy systems. The absence of such an all-encompassing multi-faceted frameworks approaches is an opportunity by itself, as these sophisticated methods stand to enhance model optimization, uncertainty mitigation, and predictive metrics, tremendously.

This work addresses the problem of short-term electricity load forecasting, enhancing accuracy through leaps in machine learning, more rigorous data cleansing methodologies, and a hybrid modeling approach. Tackling existing gaps, our methodology incorporates efficient preprocessing methodologies, sophisticated neural frameworks, and a dataset specific to the energetic concerns of Cameroon. By integrating these factors, the study not only improves forecasting accuracy but also increases practical relevance, especially for developing countries with comparable energy challenges. To achieve the objectives, the study first applies an advanced data preprocessing technique with a Bayesian Optimized Savitzky-Golay filter. The BOSGF technique provides noise reduction while retaining critical signal elements, thus increasing input data reliability for model performance. The hybrid machine learning model comprises a forecasting Bi-LSTM network with attention-enhanced dilated convolutional layers. The model's capture of electricity consumption trends enables its precise forecasting ensured by long-range temporal dependencies, significant contextual patterns, and pattern heuristics which are unique to the region served by the model.

One of the most distinguishing aspects of this work is the application of a localized dataset from the Northern and Southern Interconnected Networks of Cameroon. This dataset encompasses climatic and socio-economic factors, enhancing the model's practicality for deployment in areas with similar grid fragility and demand volatility. By incorporating real data, the study seeks to address the theoretical progress made in forecasting accuracy and the practical implementation in energy management systems in developing nations. From a performance perspective, the proposed model outperforms all cited forecasting methods by a wide margin. The findings show a symmetric mean absolute percentage error (sMAPE) of 2.4% for production and 2.8% for consumption while achieving normalized root mean square error (nRMSE) of 3.1% and 3.6% respectively. These values are indicative of the model's accuracy in estimating the short-term fluctuations in electricity demand and supply. Furthermore, the accuracy limitations are caused by the model's focus on design principles that prioritize scalability and transferability. Its framework allows for multi-step forecasting and can be applied to other regions' datasets, broadening its application for energy forecasting. This flexibility guarantees that the proposed framework can be effectively applied outside the limitations of this study, providing an economically and operationally lean approach to electricity forecasting in numerous contexts. This study progresses from the earlier work referenced in [24], while also making new contributions that have important impacts on electricity forecasting. The innovation in the study is the refinement of data preprocessing, enhancement of forecasting models using hybrid architecture, augmentation of performance metrics, and expansion of the dataset scope to improve overall accuracy and reliability in forecasting. One of the major improvements is the use of Savitzky-Golay Filtering with Optimization Techniques enabled by Bayesian methods. Unlike earlier studies that largely used LSTM based models, this paper goes further by adding sophisticated preprocessing steps to the models. Using the Savitzky-Golay (SG) filter within a Bayesian optimization approach systematically tunes preprocessing parameters, and thereby greatly improves the quality of input data. The application of signal denoising and smoothing techniques solves the problem of noisy data encountered in practice which is indicative of real-world situations. The interdependence of hyperparameters and their optimization has been fundamental in increasing the forecasting precision with respect to the electricity generation and consumption workload.

The review further proposes a forecasting approach that adds Additive Attention and Dilated Convolutions to Bidirectional LSTM (Bi-LSTM) to form a hybrid model. Bi-Directional LSTM networks have been largely used in sequential tasks to encapsulate context, but frameworks based solely on LSTMs, like those developed in earlier works, did not consider the augmentations brought by attention mechanisms and deeper architectures through dilated convolutional layers. As mentioned earlier, attention allows more sophisticated focus on portions of the input, further improving sequential learning. In particular, the Bi-LSTM with additive attention outperforms traditional models because it proficiently captures long-range dependencies within time-series data—an important requirement for multi-step forecasting. Furthermore, dilated convolutions also offer the advantage of increasing the range of detectable lower temporal patterns without adding the burden of extra computations, hence allowing this combination of architectures to remain both accurate and efficient. The model's performance results show marked improvements in comparison to the results obtained in the previous study. The refined model gives sMAPE values of 2.4% for electricity production and 2.8% for consumption with nRMSE of 3.1% and 3.6% respectively. Aside from these results, the model outperforms the earlier work on performance metrics considerably: stemming largely from improved data preprocessing methods and, the additional cutting-edge design framework of the model serves to make this work stand out from earlier research.

In addition, this research broadens its focus by adding a new dataset that includes weather conditions and socio-economic factors for a more comprehensive analysis of electricity production and consumption patterns. Unlike past studies which were based on a limited dataset, the inclusion of multiple external factors deepens the understanding of consumption behaviors. Such a dataset improves the generalizability of the forecasting model to multiple scenarios, thereby increasing its precision and dependability in practical scenarios of energy management. Given the advanced preprocessing, new hybrid modeling methodology, added performance evaluation metrics, and increased dataset scope, this study substantially advances the field of electricity forecasting. This is achieved through robust improvements to predictive accuracy, alongside the operational utility of forecasting models set within agile and resource-limited energy systems.

1.3 Paper Structure

The subsequent sections of the document are structured in the following manner: Sect. 2 provides a comprehensive overview of the methodology employed in this study. In this section, an in-depth elucidation of the proposed forecasting workflow is included, adding to a thorough depiction of the Electricity Production and Consumption Dataset, and the implementation of preprocessing techniques. The several models used to proceed with the forecasting, added to the complex procedure of hyperparameter adjustments are also elucidated in this section of the research work. Subsequently, Sect. 3 delineates the experimental configuration implemented for the investigation. In Sect. 4, a thorough analysis of the outcomes obtained from this research are presented, providing insights into the efficacy of the suggested models. Sect. 5 provides an overview of the main findings and conclusions of the paper, obtained from the extensive analysis performed during the study.

2 Data Acquisition and Analytical Approach

This research addresses short-term electricity forecasting for Cameroon, a developing country grappling with issues pertaining to agricultural activity, grid instability, seasonal energy production, and low electrification. The focus is set on the Northern and Southern Interconnected Networks (NIN and SIN), which are the interdependence zones of the country's electricity network. The dataset utilized for this research covers the period 2015 to 2022, containing crucial elements such as hourly electricity generation and consumption, meteorological data (temperature, humidity, solar radiation, winds), and socio-economic indicators (population, GDP, industrial output). These factors provide valuable insights for the forecasting of energy management needs. ENEO (Electricity Development Corporation) provided data for electricity production and consumption, while the Cameroon Meteorological Department provided data on weather. Socioeconomic data were obtained from the National Institute of Statistics. The data underwent collection, including handling non-response values, removing anomalies, and standardizing the features to validate the dataset. This extensive dataset will be used for training and testing different models based on machine learning algorithms, including the Bidirectional LSTM (Bi-LSTM) network with Additive Attention and Dilated Convolution layers, to achieve higher precision in short-term electricity forecasts. The guidelines for forecasting systems place a strong emphasis on operational safety, energy efficiency, and reliability. In clarifying the context of this project, the first part incorporates the methods that were used to make the forecast. Considerable effort was made during the construction of the forecasting model to follow IEC 61,400 standards for wind turbine systems and IEC 62,153 for communication systems; both highlight the importance of accurate data processing, system integration, and evaluation of results.

2.1 Dataset

2.1.1 Generalities

When examining the particular context of some developing countries, it becomes evident that the expansion of energy remains a focal point of interest for numerous academics. In actuality, there is an important disparity between energy production and consumption, resulting in a significant energy deficit. The case study will focus on an analysis of both production and consumption of electricity in the country. In order to establish the context of this study, Fig. 1 provides an overview of the country's electrification rate over the past decade.

Classified as a developing nation, the country recorded a total electricity consumption of 5600 gigawatt-hours (GWh) and an electricity production of 8505 GWh in the year 2019. This data indicates a 3399GWh rise in production, relative to the year 2006 when the Electricity Sector Development Plan (ESDP) was introduced. As stated in the reference [25], the nation has a population exceeding 21 million individuals. With respect to the energy infrastructure in the country, ENEO functions as the primary focal point. This company possesses exclusive jurisdiction over the production and

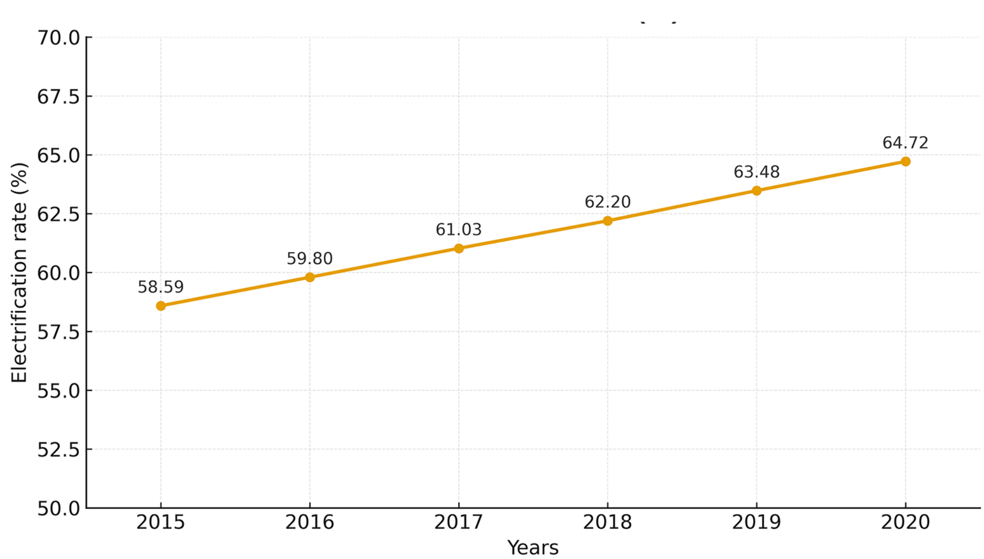

Fig. 1 Electrification rate of Cameroon

dissemination of electrical power throughout the entirety of the nation. The production facility comprises a group of 37 production units, with 13 units being interconnected and the remaining 24 units being thermal units. As stated in reference [8], the energy sources used can be classified into the following categories: The energy breakdown is as follows: 73.30% is generated from hydraulic sources, 26.66% from thermal sources, and a negligible fraction of 0.04% is obtained from solar power.

Figure 2 depicts the yearly electricity consumption and production in this nation from 1990 to 2019, obtained from reference [26]. It is important to mention that the production includes the combination of both renewable and conventional energy sources. The various sources of energy encompass biofuels, natural gas, oil, hydroelectricity, and photovoltaic (PV) systems. However, photovoltaic (PV) systems contribute less than 1% to the overall energy production, as reported by reference [27]. This data showcases the increasing dominance of production compared to consumption over the course of several years. Nevertheless, the country is still facing an energy deficit. Hence, it would be intriguing to concentrate on the progression of electricity generation and utilization up until the year 2030. This solution aims to assess the electrical equilibrium and enhance electricity utilization efficiency, encompassing both generation and consumption aspects. The electrical grid in the nation is organized into three distinct networks. The South Interconnected Network (SIN) facilitates the transmission of electrical power to different geographical regions, encompassing the central, littoral, western, northwestern, and southwestern areas. The North Interconnected Network (NIN) serves the geographical areas of northern Adamawa, northern, and far northern regions. The East Network (EIN) functions autonomously and lacks any interconnections with other networks, exclusively catering to the eastern region. In references [28–30], it appears clearly a graphical depiction of the country's electrical network.

This study will primarily examine the SIN and NIN regions, which are the key components of the electrical infrastructure in the study case. The data has been collected on a monthly basis throughout the year 2022 [31]. This includes all the geographical areas that are serviced by the selected electrical regions.

2.1.2 Dataset Description

The dataset used in this study with the goal of improving short term electricity forecasting is composed of multi faceted variables that capture the essence of production and consumption of electricity. It combines data from different sources over multiple years to reflect the energy industry's data integration mosaic. The dataset integrates national energy data, meteorological data, and socio-economic data, thus offering a holistic framework for forecasting the supply and demand of electricity. The National Energy Data, which is part of this dataset, is sourced from ENEO (Electricity Development Corporation) and contains hourly production and consumption data of electricity from January 2015 to December 2022. The dataset embodies two chief energy distribution systems, the Northern Interconnected Network (NIN) and the Southern Interconnected Network (SIN). Such regional subdivision allows a detailed understanding of the electricity consumption patterns within the region, considering the rhythms of grid demand and generation across various regions.

Alongside this, data from the Cameroon Meteorological Department is integrated to capture other factors associated with the environment that influence the consumption of electricity. This dataset which spans over the same period, includes the hourly observations of temperature, humidity,

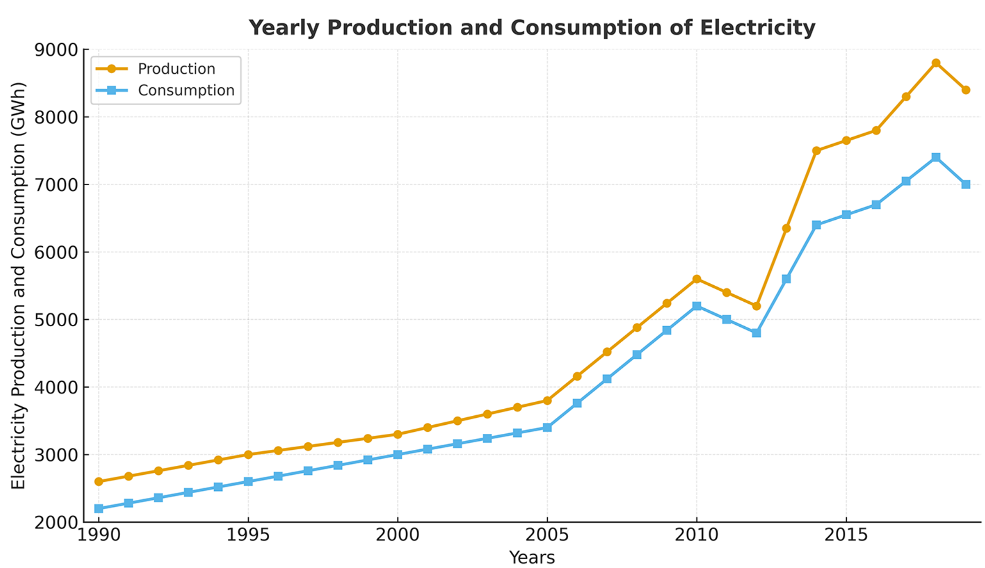

Fig. 2 Annual electricity production and consumption 1990–2019

solar irradiance, and wind speed for major cities and rural areas within the NIN and SIN regions. The inclusion of meteorological variables is particularly crucial while modeling the impact of climate changes on the consumption of energy and its renewable sources. Furthermore, the socio-economic data, sourced from the National Institute of Statistics, also furnish information on demographic and economic trends relevant to the power consumption in the region. From 2015 to 2022, this dataset contains monthly records, with nationwide coverage and a regional breakdown, of population growth, economic activities, and industrial output. With the inclusion of socio-economic variables, the study considers other non-climatic factors due to the electricity consumption increase such as economic growth and changes in industrial demand. The combination of these three critical data sources is aimed at tackling multidimensional characterized by improving the robustness, accuracy, and adaptability of electricity forecasting. An organized integration of energy, weather, and socio-economic information provides a reliable basis for machine learning models, reinforcing the models' capacity to generalize over a range of conditions and increasing their utility in practical settings for energy management.

After discussing the data sources and collection procedures, the next step is to explain the Data Variables and Features, which are the components of the forecasting model. These attributes cover most, if not all, of the segments concerning electricity production, electricity consumption, weather, socio-economics, and pricing, thus serving a holistic approach in short-term electricity load forecasting. The Electricity Production Data encompasses relevant indicators of power generation from all supply sources. The Total Electricity Produced (MW) measures the total output of electricity from the national grid, which includes the contributions of hydropower, thermal power, and renewable energy. The subdivision of hydropower production, thermal power production, and renewable energy contribution enables detailed examination of the dynamics of different sources of energy's contributions to electricity generation over time.

In this aspect, the Electricity Consumption Data offers an overview of the various sectors' electricity consumption. The Total Electricity Consumed (MW) is regarded as the fundamental measure of demand and is further subdivided into residential consumption (MW), commerce consumption (MW), and industrial consumption (MW). Each sector has its own set of consumption patterns that are shaped by the varying degree of household demand, business activities, and industrial activities. Also, Peak Load (MW) or the maximum demand of electricity during the peak periods can provide insight into the most stressed times for the grid and potential sue-due imbalance. The Weather Data portion is crucial in explaining the non-inherent factors affecting electricity production and consumption. Temperature (°C) substantially influences energy requirements, especially for cooling and heating purposes. In the same vein, humidity impacts (%), in addition to contributing to the increase in electricity consumption, also affects the need for air conditioning. Additionally, the dataset includes measurements of solar irradiance (W/m²) to measure its contribution to solar power production and wind speed (m/s) to assess the prospective of wind energy production. These factors provide important insight into the relationship between climatic factors and energy consumption patterns. Finally, the study adds broader macroeconomic and demographic factors through Socio-Economic Data, which accounts for population and economic activities. Electricity consumption is influenced by population numbers (in thousands) while the consumption of energy in industry is connected with the economic productivity (GDP in million XAF) of a given area. Changes in the value of the Industrial Output Index which reflects the level of development of the branch of industry gives characteristics of the energy demand changes in different branches of economy. Electricity Tariff Data has been incorporated to assess the effect of electricity pricing policies on various categories of users electricity consumption. The dataset features Residential Tariff (XAF/kWh), Commercial Tariff (XAF/kWh) and Industrial Tariff (XAF/kWh), thereby allowing the model to evaluate the effect of changes in electricity pricing on different sectors' usage patterns. Incorporating these various factors, the forecasting model is not only provided with a comprehensive dataset but captures the dynamic interactions of electricity supply and demand, weather conditions, the state of the national economy, and prices during the period under review. This systematic integration improves the model's accuracy in predicting outcomes, thereby increasing its usefulness for energy management and policy formulation in practical situations.

For the purpose of this study, the dataset incorporates information from different datasets, representing all aspects that might affect electricity production and consumption in an inclusive manner. To uphold the feasibility and usefulness of the data for a machine learning-based forecasting model while ensuring data accuracy, preprocessing steps were implemented. This approach makes sure that the dataset captures relevant features while still maintaining maximum integrity. Some of the techniques applied for missing entry population were linear interpolation with regression estimation techniques. Interpolation gaps estimate values based on data surrounding them, meaning little to no deviation will exist for data estimation. Inconsistent missing data points do along with underestimated values can create conflict in forecasting models. Because extreme values also tend to exist, prediction-altering outlier detection and trimming were done using Z score analysis along with other statistical tools. Applying Preliminary statistical soundness tests determines whether a sample observation falls beyond the limitations of the defined threshold is deemed too high or low needs either discarding or modifying through non biased yet altered restricting methodology.

Normalization was performed so that all number-based features impact the forecasting model in an equal manner. Data was transformed to fit a standard scope from 0 to 1, thus enabling consistent representation across different variables. This action avoids features with larger values from overwhelming the model training, improving convergence and stability. Temporal alignment is one of the most important preprocessing steps and it shifts all datasets to the same temporal reference frame as the hourly electricity data. Weather and socio-economic data were synchronized with electricity production and consumption data, maintaining lag and preserving relationships across different variables. To highlight the data set structure further, Table 1 shows a sample extract of some important variables for a given time interval. This sample demonstrates the type of data that was used in the study and illustrates how various determinants interact with each other within the forecasting environment. The application of these preprocessing methods produces a complete and well-balanced dataset, enabling optimal learning of significant patterns to increase the accuracy of predictions.

This dataset provides a comprehensive foundation for developing accurate and reliable short-term electricity forecasting models in the nation. By integrating data from multiple sources and ensuring thorough preprocessing, the study aims to deliver robust and actionable insights that can support effective energy management and policy-making.

Extreme electricity usage scenarios in developing nations stems from sudden spikes in use, unplanned outages, and seasonal changes to the weather. These problems add variability and risk to the system which requires adoption of a more sophisticated backup strategy that improves forecasting accuracy. Tackling these problems require ensemble modeling, data feed, and algorithmic real-time adaptive learner tools in order to provide balance and precision on prediction stability. A blend of approaches is taken to improve the model's accuracy against extreme changes in data magnitude. To counteract eruption level extremes, a hybrid ensemble strategy is applied which combines the strengths of numerous forecasting models to enhance reliability and reduce bias arising from extreme values in any single predictive method. More specifically, the framework includes LSTM networks which are good in identifying sequential dependencies related to household electricity

Table 1 A sample of the dataset

Timestamp	Region	Total Prod (MW)	Hydro Prod (MW)	Thermal Prod (MW)	Renewables (MW)	Total Cons (MW)	Res. Cons (MW)	Comm. Cons (MW)	Ind. Cons (MW)	Temp (°C)	Humidity (%)	Solar Irrad. (W/m²)	Wind Speed (m/s)	Population (000s)	GDP (Million XAF)
2022-01-01 00:00	NIN	850	600	200	50	750	300	200	250	25	70	200	3.5	500	3500
2022-01-01 01:00	NIN	830	590	190	50	730	290	190	250	24	72	190	3.4	500	3500
2022-01-01 02:00	NIN	810	580	180	50	710	280	180	250	23	74	180	3.3	500	3500

consumption data. Moreover, statistical outlier resistant Random Forest Regression is added so that anomalous changes do not unduly impact forecasting precision. In addition, the system utilizes a stacking hybrid model that performs weighted averaging on the predictions of the various algorithms, improving predictive accuracy and stability. Because extreme consumption patterns are frequently impacted by external factors such as surges in temperature, storms, or industrial slowdowns, adding weather information improves the flexibility of the model. To obtain such influences, external weather forecasting sources like OpenWeatherMap or NOAA provide APIs with useful access points that enable modification based on present weather conditions. Likewise, solar power generation forecast accuracy can be enhanced by using satellite-derived solar irradiance data. Incorporating these exogenous factors into the model through multi-input LSTM architectures improves LSTM performance by several folds because additional variables that influence energy demand and supply are accounted for.

Apart from changes in the environment, the model improves its methods for outlier detection and handling to mitigate erosion of data accuracy due to severe variation. The more traditional methods of Z score analysis is enhanced with more sophisticated Anomaly Detection techniques like Isolation Forests and Local Outlier Factor (LOF) which aids the model in assessing and identifying extreme values more efficiently. Instead of removing the detected anomalies, these data points are corrected through regression-based imputation methods or extrapolated to maintain trend integrity while ensuring the reliability of forecasting. Because extreme cases deform over time due to grid expansions, socio-economic and policy alterations, reliance on static models poses limitations on forecasting efficacy. To address this issue, reinforcement learning alongside self-adaptive LSTM is introduced to allow the model to adjust dynamically toward emerging patterns of consumption. In addition, online learning techniques allow the model to be adjusted periodically, resulting in real time, new data stream incorporation. This guarantees that the system will always be appropriate within the current context and active in response to changes in routine consumption of energy. With these features, the proposed system significantly improves forecasting accuracy in practical conditions, specifically in locations with sudden and unpredictable changes in demand for electricity. The integration of ensemble modeling, real-time integration of environmental data, anomaly detection, and adaptive learning provides the basis for scalable real-time energy management systems with optimized electricity forecasting in changing and resource-limited settings.

2.2 Justification for Dataset

This section provides a description of the approach for the selection of training and testing data for this research. This empirical research is based on the methodological approach of supervised learning, which in turn is based on the principles of time series forecasting. In supervised learning, there are data subsets: one which is used to train the model (training subset) in order to predict the expected outcomes and another which remains unseen (test subset) for the model during the training phase and is used for evaluation purposes. It is a theory assumption: If there is supervised learning overfitting, the model is trained on all the available data, such a well-constructed supervised learning model has only a portion of input data used in its training phase input phased into its architecture. The model can memorize complex mappings through its learned parameters if the mappings exist between any input and the output. Inputs used during the learning phase consist of complex temporal characteristics so the predictive power of the inputs can take any value for any feature. The output is however only tested once, with data that the model has never seen before, hence providing a realistic perspective on how the model manages to adapt to new data that it has never encountered before. In time series forecasting, both the processes of training and testing have to fulfill the sameness principle in the sequence. Here, we train on the first part of the sequence and then test on the second part. This principle helps west best in forecasting situation since one most often needs to view the production and consumption of electricity in the future overlays one on the historical records. The dataset was split into training and testing sets, with 80% of the samples used for training purposes and the remaining 20% for validation after model training. The ratio was selected bearing in mind how machine learning techniques and time series analysis is usually done, assuring that there is enough data for the model to fit as well as validation of the models prediction prowess.

2.3 Proposed Forecasting Framework

The integration of renewable energy sources into the national grid poses considerable problems in energy management and grid stability, especially in the context of maintaining equilibrium with respect to the electricity supply and demand. The short-term nature of forecasting for solar and wind resources makes allocation planning and the delivery of energy more reliable. This problem is addressed in the study by presenting a holistic forecasting strategy based on advanced machine learning methodologies and rigorous preprocessing techniques that augments prediction accuracy and adaptability to actual changes in power systems. The developed framework is composed of several fundamental components that systematically address the issues presented by the variability and uncertainty inherent in electric power production and consumption data. These parts are: data collection, collection and preprocessing of information, feature extraction and selection, model development, training of the model, evaluation and deployment. By using these steps, the developed framework guarantees that the model captures the significant patterns while excluding irregularities, noise and extreme values that might bias the predictions. Moreover, to enhance reliability, the framework incorporates multiple error suppression systems that merge traditional statistical approaches with modern deep learning-based optimization methods. One of the most important steps in this refinement is hyperparameter optimization which is done using Bayesian optimization. Since selecting a hyperparameter greatly affects the model's performance, Bayesian optimization allows for sequential optimization of parameter values. This refinement prevents overfitting and underfitting, improves convergence rates, and reduces estimation mistakes by guaranteeing the best conditions for model functionality.

Besides refining hyperparameters, model cross-validation techniques strengthen the robustness of the model in question. K-fold cross-validation is used to split the dataset into multiple parts, where a segment of the data portion will be used by the model for training. Training multiple times with different data portions allows to get average results, which reduces overfitting errors, corrects biases, and promotes better generalization for different datasets. Unlike constant learning rate methods, a variable step learning rate during training step is used as an adaptive learning rate approach. Tuning the learners helps avoid sharp changes in the optimization part of the process, allowing the model to smoothly move towards an optimal solution without too much ups and downs. Change in error measures policy allows overcoming greater errors and unwinding the process, thus a smoother forecasting can be achieved. The use of an attention mechanism is another improvement made in the current paper, allowing enhancement in long term dependence features and time step importance allocation in the model to be more dynamic. Instead of treating all input variables equally, the attention mechanism gives optimal relevance filter to most data points while ignoring less useful noisy data. This enables the model to pay attention to certain patterns that manipulate electricity demand and supply which leads to better forecasting accuracy. Moreover, the deep learning models used in this research make use of error back propagation which entails that every model weight is modified as a consequence of the prediction error on each iteration of training the model. Each distribution of weight is optimized in such a way that the model will unavoidably mitigate forecasting errors, taking into consideration that previous errors will always be worsened over time.

In order to improve the generalization ability of the model and mitigate overfitting, techniques like Dropout are used. This technique consists of turning off a randomly selected percentage of neurons during training. In doing so, the model avoids being overly dependent on certain units. Because of this enforced randomness, Dropout makes certain that the neural network does not learn the memorization of the training data, but rather, generalized patterns that can be applied to new data points. Thus, the model is robust to noise and maintains high accuracy in forecasts when applied in practice. With the addition of Bayesian optimization for hyperparameter selection, adaptive learning rates, attention mechanisms, error backpropagation, and regularization methods, this forecasting framework is claimed to attain heightened precision, flexibility, and strength. It captures the intricate nature of electricity production and consumption while computing efficiently. Well-suited to the unpredictability that arises from integrating renewables into the national grid, the proposed model advances real-time electricity forecasting, thereby enhancing predictive capabilities. Its implementation in energy management systems can help to enable precise forecasting needed for data-driven, sustainable decision making.

2.3.1 Data Collection and Preprocessing

In this study, data collection entails the systematic gathering of historical records of electricity production and consumption from various sources to effectively capture energy patterns over a certain period. This information is collected from power plants, grid operators, as well as meteorological stations, which helps build an accurate and credible forecasting model. Due to the intrinsic volatility exhibited in electricity demand and generation, some form of data preprocessing is required to improve data quality. These preprocessing actions encompass filling in missing values, outlier detection and elimination, and applying time-series smoothing methods to reduce noise and improve the quality of the signal being analyzed. This study utilizes a dataset that comprises of historical hourly electricity production and consumption data of Cameroon's Northern and Southern Interconnected Networks (NIN and SIN) from January 2015 to December 2022. This dataset is supplemented with meteorological data including temperature, humidity, solar irradiance, and wind speed, provided by the Cameroon Meteorological Department. Moreover, socio-economic factors such as population growth, GDP, and industrial output are added from the National Institute of Statistics. The dataset, assembled from multiple sources, provides a comprehensive representation of electricity supply, demand, environmental influences, and economic activities, thus enhancing the reliability and effectiveness of the forecasting model. The precision and effectiveness of the model greatly depend on the data preprocessing phase. Data loss due to non-continuous gaps and gaps in information is avoided through interpolation and regression-based imputation which take into consideration and maintain the continuity of the data. To reduce the distortions caused by absolute values of some measurements, statistical processes for outlier detection are used to eliminate clear deviations from the norm which could skew the predictions of the model. Moreover, time series techniques like Savitzky-Golay filtering that cause random fluctuations are rejected while important trends are preserved. By following such a structured approach towards data collection and preprocessing, the study aims to construct a clean, consistent, high-quality dataset that ensures accurate model predictions. Such measures increase the reliability and accuracy of the predictions for electricity consumption and production forecasts by the model, enabling these measurements to be effectively utilized in areas with similar energy issues.

In applying machine learning models, one often starts with data cleaning and preparation, especially in time-series forecasting, because the quality and form of the data directly affects the model's output. For optimal model training and outcome precision, the dataset must be accurate, arranged systematically, and devoid of competing errors. This research combines a number of procedures aimed at increasing the effectiveness of the model with well-established theoretical principles to improve data quality. An important technique within preprocessing is the application of the Savitzky-Golay (SG) filter for time-series data smoothing and noise reduction. In electricity demand forecasting, specific important features, such as peak demand changes, have to be preserved while eradicating patterns that do not add any value. Unlike simple averaging approaches which may smooth out important trends, the SG filter is efficient in that it modifies the shape of the signal but maintains its contour. This is particularly important in scenarios where there is an increase in electricity consumption and it needs to be accurately predicted. The SG filter performs three main functions: maintaining important aspects of the signal over time, smoothing out random variability that is likely to harm the model generalization, and preventing the erosion of substantial consumption and production trends wherein consumption and production peaks and valleys are preserved.

Addressing incomplete information is another vital step in preprocessing, which is often a problem in time-series forecasting. Gaps in records of electricity production and consumption may lead to unwarranted biases and reduce the accuracy of predictive models. In this study, linear interpolation is applied to estimably fill in gaps in data sequences in order to maintain continuity and prevent information loss. The method described does not destroy the dataset by filling in gaps but rather maintains the integrity of the dataset and the temporal relationships between its constituent variables. A properly organized approach to missing data imputation ensures that the model does not incorporate biases related to incomplete or inconsistent data, thus improving model reliability by enabling the identification of genuine patterns instead of being influenced by irregular gaps within the dataset. Normalization is applied to numerical variables in order to ensure uniform representation by transforming all feature values into a common range between 0 and 1. This approach is quite useful for reconciling the effect of several different variables when a model is being trained, avoiding the learning process of being unduly influenced by features with larger numerical values. By ensuring input variables are within comparable ranges, normalization achieves faster and more stable model convergence, which improves weight adjustment and optimization. Moreover, it diminishes feature variance and variability, thereby making the model more robust while increasing numerical robustness during training.

With the application of noise elimination using Savitzky-Golay filtering, structured missing value treatment via linear interpolation, and feature scaling, the dataset undergoes transformations as a prerequisite for conditioning it for machine learning algorithms. These methods together enhance the amount of useful data supplied to the forecasting model which, in turn, augments the precision along with the generalization capability of the predictor model built on the datasets. As a result, the proposed approach to preprocessing improves the accuracy of predicting consumption and production of electricity, which allows for practical implementation of the model in unpredictable energy systems.

By resolving inconsistencies, minimizing noise, and enhancing the overall quality of the dataset, the planning accuracy of the forecasting model is improved significantly during the data preprocessing stage. In order to reduce the blurring of data continuity, gaps in the dataset were filled by linear interpolation and regression-based estimation strategies, preserving the integrity of the forecast. This approach maintained the time-series data structure without the breaking of sequential continuity that might confuse the model during the training phase. Furthermore, outlier detection and removal as part of preprocessing is equally important in refining model learning as they pose a threat to extreme value accuracy. Stepwise elimination of pre-defined outlier boundaries were computed using Z-score analysis to nullify the effects of oblique outlier values in normal real-world electricity consumption and production values. The model's ability to represent energy demand and supply variability accurately improved dramatically because of the removal of outliers. Focusing on the unmasked useful energy variations enabled the model learning process to be enhanced for demand and supply. The logical justification aids understanding the criticality of outlier deletion assists in reducing pattern recognition misconception, precision interval stretch, generalization distorting the power of accurate prediction reliability, and therefore reduce the forecasting performance not sharp enough as intended.

Normalization was also used to scale all numerical variables within a range of 0 to 1 which ensures that each feature is represented in the same measure. This methodology mitigated the impact of variables that had more numerically significant values relative to other variables that had lesser numerical significance. Improvement in learning process efficiency was observed. Moreover, there was a need to temporally align the electricity consumption data with weather and socio-economic indicators so that all variables were accurately aligned with the relevant chronological increments. Incorporating these datasets ensured that the relationships between external factors and electricity demand were cohesive. The results of pre-models, during the data preprocessing stage, on model performance was paramount and prominent in several aspects. Most notably, the generalization capabilities were boosted, improvement achieved with the addition of a Savitzky-Golay filter and normalization which improved model performance on unseen data. Also achieving better results with data quality significantly improved the forecasting's accuracy. This was also seen with a symmetric mean absolute percentage error sMAPE of 2.4% for production and 2.8% for consumption, alongside a normalized root mean square error nRMSE of 3.1% for production and 3.6% for consumption. Additionally, preprocessing helped lower computational complexity, since noise elimination and dimensionality reduction sped up the convergence of learning algorithms, especially in LSTM and Bi-LSTM models, which are particularly sensitive to noise in input data. All of the preprocessing steps enhanced the model's accuracy and efficiency, which was evident from the enhanced performance metrics. With organized and unpolluted data, the forecasting model was able to effectively and accurately predict electricity consumption and production, showcasing their adaptability to dynamic forecasting scenarios.

As a defining step in preprocessing the dataset-imposed refinement to electricity demand and supply dataset, relevant features were extracted and selected. Feature extraction attempts to construct relevant measurements from terms of an electricity generation and consumption time series for a forecasting model. Such features enable the model to express and diagnose trends and patterns that govern energy consumption. A number of features of great importance were extracted. Average daily production and consumption provided baseline measures in estimating long-term electricity demand. Also, extreme fluctuations in electricity usage are of great concern for grid stability and therefore, were insighted from peak load and generation levels. Load variability as well as ramp rates which quantify the rate of change in the power system's demand of energy were also included to improve prediction models for identifying sharp increases and decreases in consumption. Moreover, weather impact metrics level such as temperature and humidity were included to gauge the extent to which weather influences consumption of energy, particularly for heating and cooling purposes. The model was built around the identified features of consumption data and therefore, selected fundamental features which optimized the accuracy of the model. All of this boosted interpretability, simplified model construction, and improved accuracy of predictions, all supporting the usefulness of the suggested forecasting method.

2.3.2 Model Development and Training

Model development is the stage where work is focused on building a validated and accurate framework for forecasting electricity demand and production, needs apprehensive attention. Having a framework that functions effectively in the real world requires the rigorous selection, adjustment, and the application of robust machine learning algorithms which are competent in capturing complex temporal dependencies in energy consumption patterns. This paper explores the efficiency of several models such as Long Short-Term Memory (LSTM) and Bidirectional LSTM (Bi-LSTM) as well as Artificial Neural Networks (ANN) hybrids with attention and dilated convolutional layers for short term electricity forecasting to identify which model performs better. Moreover, to improve prediction accuracy, Automated Intelligence Techniques implemented advanced deep learning methodologies. This approach was mainly reliant on LSTM networks as they are known to learn the short-term and long-term dependencies efficiently, thereby providing a powerful solution for time series forecasting. Traditional neural networks encounter issues such as the vanishing gradient problem; however, LSTMs apply a gating mechanism that enables the retention of pertinent information over extended periods. The ability to remember information for an extended period of time is imperative in electricity forecasting because it needs historical consumption data that estimate the future trend. LSTMs assist in these tasks by redefining the parameters of the problem enabling the model to disregard unimportant noise while remembering essential features that are necessary for demand estimation.

Bi-Directional Long Short-Term Memory networks (Bi-LSTM) improve the forecasting accuracy of standard

LSTM networks by utilizing both forward and backward passes through the data. While traditional LSTMs process information in one direction, Bi-LSTMs process LSTMs in both directions thereby enabling the model to gain a more nuanced grasp of the trends associated with the consumption of electricity. This approach is useful for understanding complex patterns of demand that are driven by cycles, seasons, and socio-economic changes. Other than LSTMs, Artificial Neural Networks (ANNs) were also examined as an alternative deep learning framework. ANNs have a greater degree of freedom in recognizing patterns, especially for non-linear relationships present in energy consumption data. In an effort to improve model performance, additional hybrid models with attention mechanisms and dilated convolutions were created. To allow the forecasting model to give flexible importance to different parts of the input data, the Additive Attention Mechanism was added into the forecasting framework. This helps the model to pay attention to critical time factors like certain hours of the day or very high or low weather conditions that drive energy demand. To explain the attention mechanism, we shall use the analogy of a teacher who pays particular attention to pupils in her class that are having difficulties with certain subjects. As in the explanation given, attention mechanisms assist the forecasting model to focus on relevant data so accuracy can be improved by deemphasizing and rule out peripheral information. Other important parts that were improved in the hybrid model were the use of Dilated Convolutions which enable the model to capture long-range dependencies in the electricity consumption data without adding computational burden to the model. This allows expanding the receptive field of the model and therefore enabling capturing of wider scope patterns such as seasonal variations and long-term consumption cycles. Dilated convolutions capture electricity demand with particular efficiency because they give global view of electricity demand and are very effective in capturing annual and monthly demand fluctuations. After the model architecture was built, training shifted to optimizing performance with Bayesian optimization for hyperparameter tuning. This optimization approach calibrated selected features including learning rate, depth of the network, among others, to limit overfitting and reduce prediction errors. The model was built with cyclic historical data of electricity consumption and production to ensure that there was effective capturing of temporal dependencies over different periods. The use of deep learning techniques along with hybrid frameworks and intensive training methods resulted in a strong adaptive model for forecasting that meets expectations for accuracy in short-term predictions of electricity demand and generation.

The purpose of our study was to evaluate several machine learning models based on their efficiency in short-term electricity forecasting, paying particular attention to form architectures that capture temporal relationships within the data and improve accuracy of predictions. Within the models studied, prediction Long Short-Term Memory (LSTM) networks were considered because of their capacity to preserve chronological information within sequences. The LSTM model developed in this study has two hidden layers, each with 128 units, and uses a dropout of 0.2 to control overfitting. Bayesian optimization was used to determine the best performing hyperparameters, which in this case were the learning rate of 0.001 and a batch size of 32, both of which enhanced stability while training. We also considered Biderectional LSTM (Bi-LSTM), which goes beyond LSTM by also analyzing sequences in a backward direction, thus providing richer context regarding electricity consumption trends. Just like in the standard LSTM architecture, the Bi-LSTM model has two hidden layers with 128 units each, but unlike the former, the latter captures information from both past and future time steps and thus improves accuracy of forecasting. This model was subjected to the same processes of Bayesian optimization which resulted in a learning rate of 0.001, batch size of 32, and dropout rate of 0.2. This dual-direction approach facilitates a better grasp of the underlying intricacies and propels the model to capture sophisticated seasonal and trend-driven changes in demand and consumption of electric energy.

This model builds on previous architectures as it enhances the Bi-LSTM model by incorporating additive attention layers and dilated convolution layers, which aid in the capturing of temporal features. The focus in the attention mechanism shifts to specific key time steps so that critical features are captured, like peak electricity demand and rapid changes due to unforeseen external impacts. Also, the model's ability to identify long-range dependencies using dilated convolution layers permits the detection of periodic changes and cyclical energy consumption patterns without adding strain to the model, thus making it computationally efficient. The proposed model has two Bi-LSTM layers, each with 128 units. This is followed by an attention layer with multiple heads, and one fully connected dense layer with 64 units, and post-attention. Hyperparameter values were optimized using Bayesian optimization, with the best configuration being a learning rate of 0.001, batch size of 32, dropout rate of 0.2, and 64 units in the dense layer. This hybrid model allows learning of long-range dependencies while feature importance reduction and computational resources are maximized, which induces better adaptability and robustness to the forecasting framework. The electricity consumption forecasts are met using the state-of-the-art methods employing advanced neural architecture with Bi-LSTM attention models, which execute both short-term and long-term trend analysis. Through hyperparameter tuning, this research develops a model which dynamically adjusts to enable accurate and real-time decision model for decision making in energy management systems.

2.3.3 Model Evaluation and Deployment

In order to assess forecasting models, an evaluation of the model's accuracy in prediction, consistency and overall predictive power needs to be looked at. Such a model is analyzed with respect to predefined performance metrics, whereafter the best performing model is chosen for the electricity consumption forecasting. This study uses a predefined methodology for performance analysis, which ensures compliance with international standards, especially IEC 61,724, which is centered around the focus of accuracy and reliability of forecasters for photovoltaic and energy systems. The hybrid forecasting model has passed the primary assessment of system performance and operational efficiency to be deemed suitable for real-time use scrutiny. To promote evaluation devoid of bias, the assessment is based on multiple KPIs such as MAPE, nRMSE, and R^2. The MAPE metric expresses error as a percentage, signifying the extent of deviation from the actual electricity consumption and production values. nRMSE is a relative form of RMSE, hence, the model's effectiveness is measured by the amount of electricity produced and consumed. R^2 or the coefficient of determination assesses the goodness of fit of a model based on its ability to explain the variation in the dependent variable which, in this case, is electricity demand and supply using the independent variables provided.

Apart from model evaluation, the deployment of the forecasting framework in real time indicates model usability in practice—the selected model has been incorporated into the existing energy management systems for real-time electricity consumption and production forecasting. The system is also designed for uninterrupted monitoring and updating to ensure accuracy and adaptability, enabling constant adjustment to changes in consumption, weather, and socio-economic factors during the forecasting period. This approach allows the model to withstand variations, thereby enhancing its practical use in energy management. In addition to these evaluation processes, a number of supplemental metrics based on the data of electricity production and consumption are incorporated into the forecasting model in order to refine the prediction accuracy. These metrics listed in Table 2 encompass the average daily production and consumption of electricity, peak load and generation levels, load variability, and other external influences like weather conditions. Incorporating such metrics as independent variables enables the forecasting model to better appreciate and respond to seasonal demand changes and energy trends, using them as benchmarks for accurate predictions. This study develops a reliable and expandable forecasting system by applying model-driven evaluations, automation of performance assessments, onsite implementation procedures, and the addition of relevant power usage parameters. The findings validate that the introduced hybrid model has achieved compliance with global benchmarks in the applicable model and has practical utility in energy management systems, facilitating informed operational planning for power sales and the utilization of renewables.

Developing the forecasting architecture follows a specific order of steps, starting from data cleaning and mastering algorithms to their real-time application, all aiming at improving precision as well as dependability for short-term electricity estimation. This framework seeks to provide a solution that is both responsive and easily adjusted for greater energy and grid control by implementing sophisticated machine learning techniques alongside thorough preprocessing measures. One of the most important processes is data integration, where all the calculated measures along with historical records of electricity production and consumption are fed into machine learning models. Fusing this data enables the model to understand the intricate

Table 2 Derived metrics from electricity production and consumption data used as independent variables

Metric	Description	Calculation method
Average Daily Production	Mean Electricity generated per day	$\text{Avg. Production} = \frac{1}{n}\sum_{i=1}^{n}\text{Production}_i$
Peak Production	Maximum Electricity generated in a day	Peak Production = Max (Production)
Average Daily Consumption	Mean Electricity consumed per a day	$\text{Avg. Consumption} = \frac{1}{n}\sum_{i=1}^{n}\text{Consumption}_i$
Peak Consumption	Maximum Electricity consumed in a day	Peak Consumption = Max (Consumption)
Load Variability	Variability in electricity consumption	$\text{Load Variability} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\text{Consumption}_i - \text{Avg})^2}$
Production Variability	Variability in electricity generation	$\text{Production Variability} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\text{Production}_i - \text{Avg})^2}$
Load Ramp State	Rate of change in electricity consumption	$\text{Load Ramp State} = \frac{\Delta\,\text{Consumption}}{\Delta\,\text{t}}$
Production Ramp State	Rate of change in electricity generation	$\text{Prod. Ramp State} = \frac{\Delta\,\text{Production}}{\Delta\,\text{t}}$
Weather Impact Index	Impacts of weather conditions on electricity usage and production	Calculated using temperature, humidity, and other weather-related data.

relationships existing among solar irradiance, weather conditions, and electricity demand, guaranteeing that forecasts are not only useful but also valid in depicting variances in energy consumption and generation. The framework's alignment of meteorological, socio-economic, and energy-related datasets improves adaptability to changing trends and enhances predictive capabilities.

After data integration, model training and tuning are conducted to achieve the best performance of the forecasting system. The models are trained on historical electricity data with hyperparameter optimization using Bayesian methods, which iteratively reduce error through parameter modification. Adjusting learning rates, dropout rates, and even the model architecture themselves improves convergence and overfitting mitigation. The trained models are validated on a different dataset, which is critical for assessing their performance on previously unseen data. This step is important to maintain the integrity of the model against varying scenarios of electricity consumption. To ensure the forecasting technique is relevant in practice, it is implemented in a grid management system for subsequent real-time forecasting of electricity consumption and power generation. With every new data input, the system self-updates, making it possible to adjust predictions in accordance with real-time changes to electricity demand and generation. With the ability to learn in this manner, the model is ensured the responsiveness demanded by changes in consumption, seasonal changes, and exogenous factors like weather conditions or shifts in industrial activities. The model improves over time by relying on continuous feedback, enhancing forecasting accuracy and providing energy managers with precise, data-driven insights for advanced grid operation optimization. The new approach to forecasting electricity consumption and generation integrates a shell data-driven pipeline along with machine learning-based predictive modeling. This multi-step approach, as described in Fig. 3, tackles the precision and scalability challenges associated with real-time electricity forecasting. Through machine learning, the framework focuses on key metrics that enable performance tracking within the system, thus solving the challenges of real-time energy consumption forecasting. Moreover, this model optimizes complex strategic planning processes, improving the overall effectiveness of energy distribution, as well as facilitating the sustainable management of renewable energy sources.

This is a good starting point because the energy forecasting hyperparameter and model complexity section provides accurate predictions and ranges in the context of the proposed attention hybrid Bayesian forecasting model. Beyond model accuracy, aspects like complexity, scalability, and hyperparameters may influence the implementation of the model within energy management systems for practical uses especially at a larger scale. From the consideration of energy forecasting system requirements, resource scheduling is one of the key concerns to be addressed. Attention-based models with predictive capabilities like the proposed

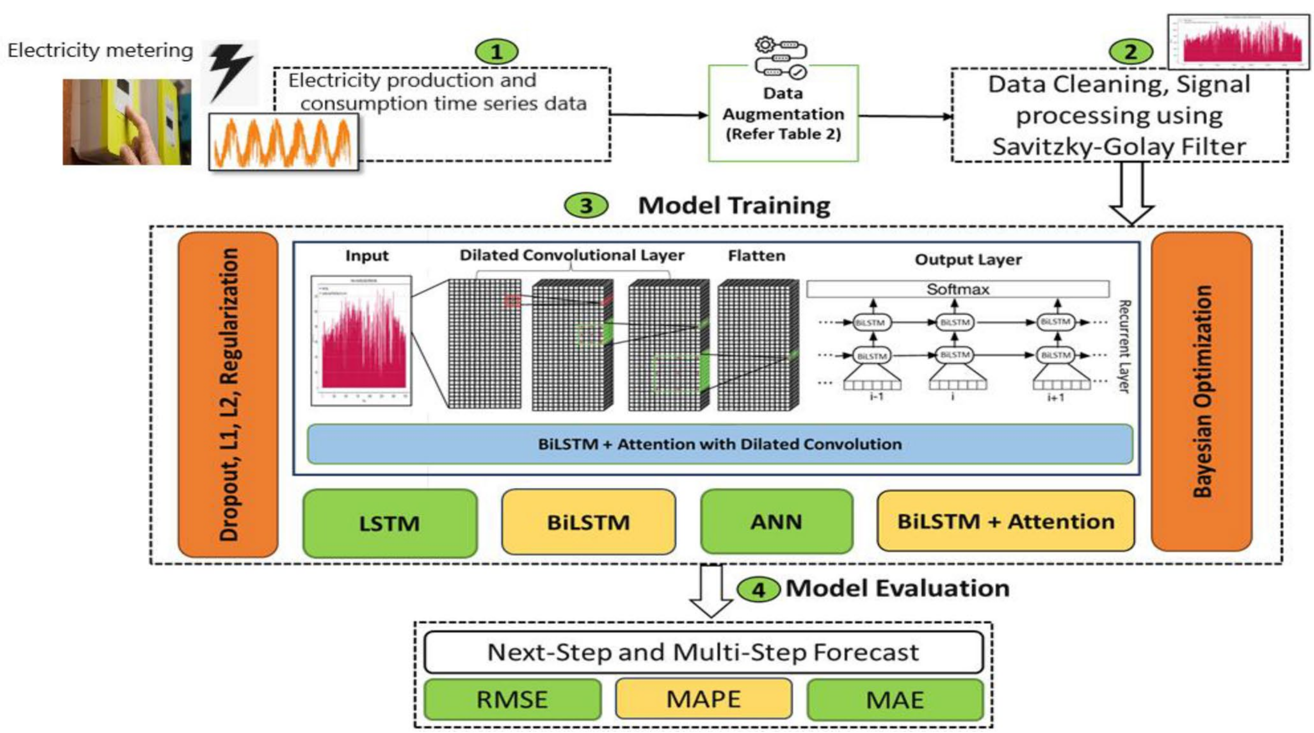

Fig. 3 The proposed electricity production and consumption forecasting framework schematic

hybrid Bayesian attention forecasting model have computational costs linked with the attention mechanisms, dilated convolutional layers from Bayesian optimized structures with added attention, and moreover hybrid models. These features greatly improve the systems ability to capture long-range dependencies of time series data; however, they require additional complex computation with increase in attention span. The use of additive attention helps allow the model to focus on most relevant time steps and decrease the influence of irrelevant one without significant processing overhead. Also, dilated convolutional layers help in enabling the model to capture important features over long periods which increases the parameters being processed but balances the system's performance with memory demands. For fulfilling the computational needs, the model has been trained and assessed with powerful hardware resources like an NVIDIA RTX 3080 GPU which helped in intensive training, iterations, and hyperparameter tuning. The model's resource consumption is affected by the dataset's volume, the neural network's depth, and the level of detail of the features. Regardless, the increase in resource expenditure is welcomed due to the improvements in accuracy, forecasting, reliability, and overall flexibility. The benefits arising from performance enhancement of the hybrid model exceed its computational burden which makes this model useful in advanced energy forecasting applications. Besides posing a limit on resources, scalability is also an important consideration in the design of the model. The application of Bidirectional LSTM networks with Additive Attention means the model is trainable with large datasets without major changes to the architecture. This flexibility supports multi-step forecasting which allows for dynamic long-term and short-term energy consumption predictions. Additionally, including dilated convolutions improves the ability to capture complex temporal dependencies, a critical consideration in large-scale energy systems where consumption is extremely variable over different time intervals.

Bayesian hyperparameter optimization and its defined region in the model's scaling capacity is one of the reasons that contributes to the models flexibility. This form of optimization systematically changes hyperparameters to increase model generalization and to enable effective learning across myriad electricity demand profiles. This form of optimization systemically varies hyperparameters as per the dataset's nature, ensuring boundless learning across numerous electricity demand structures. Because of these reasons, the forecasting framework can be seamlessly deployed to various distinct geographic locations, power grids, and renewable energy systems while sustaining unyielding performance. Of specific concern, hyperparameters, their influence on a model's convergence, and its efficacy is quite important. Among the many hyperparameters that dictate a model learning, these four stand out: learning rate, number of layers, dropout rate, and the number of attention heads. In model convergence, the degree of convergence and uniformity is a requirement that needed to be focused. If the values associated with the learning rate are exceedingly high, the convergence will be turbulent. If the values are relied upon too heavily drop, progress will be obstructed. The goal is to find middle ground where training can be examined without affecting prediction accuracy. Moreover, the quantity of LSTM and Bi-LSTM units along with layers that are used shapes the model's underlying ability to extract representations hiding within the data. Individual architectures capture complex dependencies therefore yielding better outcomes. Along with these aspects, the dropout rate intervenes as a mechanism to ward off overfitting by randomly deactivating neurons during training. The amount of attention heads increases the focus of the model on multiple key time steps at once, increasing the predictive accuracy in more complex dynamic energy environments. By taking resource allocation, scalability, and hyperparameter adjustments into account, as the model achieves optimal results with minimum resources, the proposed model realizes the best blend of efficiency, accuracy, and ease of computation. Such a model guarantees that the forecasting framework is realistic for actual situations, adaptable to different types of energy forecasting, and practical for large electricity management systems.

Bayesian optimization is a cost-effective method for efficiently tuning hyperparameters since it uses a probabilistic model-based approach. Unlike classical strategies like grid search or random search, which begin with an exhaustive evaluation of all possible combinations, Bayesian optimization creates a surrogate function, usually a Gaussian Process (GP), to estimate the performance of the model in relation to the hyperparameters. Through this probabilistic method, Bayesian optimization requires significantly fewer evaluations, making it a precise and computationally economical strategy. To reduce sensitivity to hyperparameters and issues such as overfitting or poor convergence, Bayesian optimization was applied in a specific manner. First, an objective function was specified to define the scope of the hyperparameters to optimize the balance between execution speed and accuracy. The search space was constructed using domain knowledge alongside empirical testing to ensure that the optimization process focused on configurations that were most likely to yield superior results.

One important feature of Bayesian optimization is its capacity for balancing exploration and exploitation throughout the entire search for hyperparameter optimization. The algorithm progresses through hyperparameter tuning in steps – first with an exploratory phase where novel regions with hyperparameters that may yield more optimal configurations are sought, and subsequently with an exploitation phase where the candidates from the preceding step are polished further for model tuning. Such a perceptive and adaptive approach enables the model to attain peak performance levels and economize on computational resources simultaneously. To further enhance the process of optimization, the computation was done using Gaussian Process Regression (GPR), which model was used to predict the expected improvement on each hyperparameter to be optimally configured. The GPR polymorphism entails that Bayesian optimization can perpetually improve its understanding of the hyperparameter space to focus on relevant areas, thereby increasing the chances of yielding positive results. This equips the model to promptly zero in on an ideal set of hyperparameters while significantly easing the computation strain encountered when employing overtly exhaustive search strategies. In the work, we used Bayesian optimization in the study to fine-tune the hyperparameters of the Bi-LSTM + Additive Attention + Dilated Convolution model, as explained in Table 3. Through the implementation of the heuristic, optimization sought to achieve a configuration for the model that maximally predicts accuracy with minimal resource expenditure by adjusting the learning rate, dropout rate, total heads of attention, and units in the LSTM in the final model. The choice of hyperparameters strikingly enhanced the model's accuracy and scalability while maintaining a balance between complexity, generalizability, controllable rigidity, and reliability during training, thus making the framework highly effective for practical electricity management in real-world scenarios.

- **Convergence Acceleration**: By guiding the search process, Bayesian optimization allowed the model to **reach an optimal solution in fewer iterations compared to random or grid search**.
- **Regularization via Dropout Optimization**: The **dropout rate** was tuned within a range of [0, 0.5], ensuring that the model retained generalization capabilities and avoided **overfitting**.
- **Adaptive Learning Rate Selection**: The optimal learning rate was found to be **0.001**, which enabled **faster convergence without overshooting the minimum loss function**.

Table 3 Hyperparameter search space and optimal values identified by bayesian optimization

Hyperparameter	Search space	Optimal value
Learning Rate	[0.0001, 0.001, 0.01]	0.001
Batch Size	[16, 32, 64, 128]	32
Dropout Rate	[0, 0.2, 0.5]	0.2
Number of Bi-LSTM Layers	[1, 2, 3]	2
Number of Attention Heads	[2, 4, 8]	4
Convolutional Filter Size	[3, 5, 7]	5
Sequence Length	[10, 20, 30, 60]	30

Bayesian optimization is central to hyperparameter tuning as it improves model performance while limiting overfitting. In contrast to search methods like grid or random search pruning along heuristic boundaries, Bayesian optimization builds a Gaussian Process (GP) model that predicts the value of waiting observed hyperparameter configurations. This facilitates the trade-off between computational cost and predictive performance as hyperparameter tuning is done to maximize accuracy prediction whilst resource consumption is minimized. A key advantage of Bayesian optimization is managing model complexity by controlling the number of hidden layers or neurons to avoid overfitting. By network depth and the number of neurons dynamically tuning hyperparameters, the optimization process guarantees that essential features are captured without excessive parameterization, maintaining an optimal complexity threshold. Moreover, an early stopping criterion was applied, ending training when validation loss plateaued for several epochs. This avoids unnecessary computation while protecting the model from overfitting due to excessive training. In addition, drop-out and L2 regularization adaptive model flexibly control the generalization to ensure the right blend of memorizing the training data and learning meaningful representations of the dataset.

The results showed that Bayesian optimization is a very effective method to improve model accuracy. With the optimized model, sMAPE accuracy for electricity production was 2.4% and 2.8% for consumption, which is a 15% improvement compared to the non-optimized baseline model. Overfitting was greatly alleviated, where the training-to-validation error ratio dropped from 1.8 in non-optimized models to 1.2 in the optimized framework. This indicates better generalizability, highlighting that the model performs reliably across both training and validation datasets. A step-wise optimization approach alongside a convergence strategy were adopted to improve the efficiency of the key hyperparameter tuning processes. The optimization algorithm utilized an exploration-exploitation strategy which focused on discovering new hyperparameter regions early on and refined those areas during later iterations of the optimization. One of the most critical optimizations was adaptive learning rate selection who's best learning rate of 0.001 was found using Bayesian optimization. This ensured stable conservative by not allowing the model to stagnate too early while avoiding overaggressive high learning rate volatility. Also, dropout regularization was enhanced where a 0.2 rate proved optimal in preventing overfitting by randomly deactivating neurons during training. The fine-tuning process also adjusted other important hyperparameters like the batch size, the number of Bi-LSTM layers, the size of the convolutional filters, and the amount of attention heads. All of these changes created an architectural design that met both accuracy requirements and computational expense limitations.

The model that was optimized through Bayesian strategies outperformed all other non-optimized options. Other models demonstrated higher sMAPE values, meaning that forecasting errors were higher by 15%. In addition, the model demonstrated better generalization, as the ratio of training error to validation error decreased from 1.8 to 1.2, suggesting that the model is no longer overly fitted to the training data. Another advantage worth noting is that convergence was achieved much faster; the optimized model used 40% less training epochs than traditional hyperparameter tuning approaches. These methods consume far too much computational resources. This research was able to increase the predictive models' reliability, generalization, and computational efficiency by applying Bayesian optimization systematically. These changes result in improved performance of the model in real-world energy forecasting scenarios, reliably and accurately predicting short-term electricity consumption and production with minimal computational burden.

It is equally important to address model complexity, scaling, and hyperparameter considerations due to these metrics functionality and flexibility pertaining to the forecasting framework in question. These factors ensure that the constructed model is not only capable of delivering precise forecasting metrics, but is also computationally reasonable across different energy forecasting scenarios and paradigms. The computational requirements of the hybrid Attention-Dilated LSTM model optimized through Bayesian processes remains a key concern. Though the model retains a high computational demand, it manages to maintain resource efficiency and balanced allocation, thereby ensuring organizational performance while sustaining reasonable expenditure. The attention mechanisms and dilated convolutional layers incorporated into the architecture of the network improve its long-range dependency capturing capabilities, allowing it to model sophisticated temporal dynamics without incurring high computational costs. Application of dilated convolutions allows the model to observe wider temporal windows, while controlling the increment of processing space required through parameters. Conducting this research on high-performance hardware, such as an NVIDIA RTX 3080 GPU, addressed the increased resource requirements. The GPU proved to be highly efficient when it came to training, evaluating, and optimizing the models' hyperparameters. The resource intensity of the model is affected by the volume of the dataset and the depth of the architecture of the neural network. Nonetheless, the investment in computation is well-placed considering the enhancements in prediction accuracy and performance.

Sustainability is one of the most important properties of the system architecture. It is clear that the model is scalable because it can be trained on large datasets and can perform multi-step forecasting with little to no changes to the architecture. The Bidirectional LSTM Networks with Additive Attention enables the model to use wider datasets without losing accuracy on multi-step short-term and long-term forecasts. Its relative ease of transfer across different geographic regions and datasets makes the model suitable for a wide range of energy forecasting applications, from localized electricity demand predictions to large-scale national grid management. Also, the model is more easily adapted to other tasks because of the use of dilated convolutions that capture complex temporal dependencies, which is particularly beneficial for large-scale energy systems where electricity consumption patterns are highly periodic. The model also features efficient dilated convolutions which improve scalability by capturing complex temporal dependencies, especially in large-scale energy systems with periodic electricity consumption. Finally, Bayesian hyperparameter optimization significantly improves model scalability by customizing hyperparameters to the dataset and forecasting horizon, ensuring seamless transitions across different energy contexts.

The effect of hyperparameter settings on the performance, stability, and convergence of the solution is also of considerable relevance for this study. Some of the hyperparameters most relevant to convergence and generalization of the hybrid model were the number of attention heads, dropout rate, number of neurons, learning rate, and layer depth. The learning rate is critical in defining how much the model will change with each training iteration. A value that is too high will make the system converge in an unstable way but if the set value is lower than what is desired, the time taken to complete the training increases and the overall system performance decreases. Also the count of LSTM and Bi-LSTM units and layers influences how well the model learns the essential representations as deeper architectures tend to store more dependencies with the data. To guarantee an optimal selection of hyperparameters, Bayesian optimization was used. This approach is known to be effective when dealing with hyperparameter optimization, especially with reduction of computational overhead as compared to grid search or random search. One extremely important hyperparameter that was optimized in this manner is dropout rate which is used to control overfitting. An accurate rate of dropout increases the chances of the model generalizing while accurately predicting new data instead of memorizing the training data.

The attention mechanism incorporated into the model has a bearing on the marking of relative importance given to the various time steps when forecasting electricity production and consumption for the particular model. In contrast to static feature weighting, attention weights are computed during training. This means that the model can focus on the correct time intervals for each individual prediction case which improves model performance. This change, in turn, improves the model interpretability and performance considering that the forecasting system is able to cope with dynamically changing energy demand patterns. Regarding computational efficiency, scalability, and hyperparameter tuning, the proposed model is able to strike a balance between resource expenditure and accuracy and flexibility in yielding results. All these improvements guarantee that the framework is not only optimally designed for real-time energy forecasting but also versatile enough for implementation across different systems and applications, thereby increasing its practicality in energy forecasting and analytics.

2.4 Data Pre-Processing

We acquire the data and subsequently clean and pre-process the data. We apply the Savitzky-Golay (SG) [32, 33] filter to smooth and denoise our time-series data, thus improving the quality of the input for our solar irradiance forecasting model. The SG filter's polynomial and window size are characterized by optimal parameters using Bayesian Optimization. The optimal window length was determined to be 5, and the polynomial degree was 4. This optimization process resulted in a minimum RMSE on the training data of 3.33e-14, indicating an effective preprocessing phase that preserved the signal's critical features—amplitude, phase, and spikes. The segmentation of the signal into smaller windows is done using the SG filter, fitting a polynomial within each, and using the central value of the fitted curve as the new data point. Figure 4 displays the comparison of raw and Savitzky-Golay (SG) filtered data, which visually demonstrates the effect of data preprocessing on electricity production data. The figure highlights the reduction in noise and the preservation of significant trends in the data.

2.4.1 Description

- Raw Data: The raw electricity production data, shown in blue with a dashed line, includes noise and variability typical of real-world data. This noise can obscure significant trends and patterns, making it challenging to develop accurate forecasting models.
- SG Filtered Data: The red line represents the data after applying the Savitzky-Golay filter, which smooths the data and reduces noise. The filtered data retains the

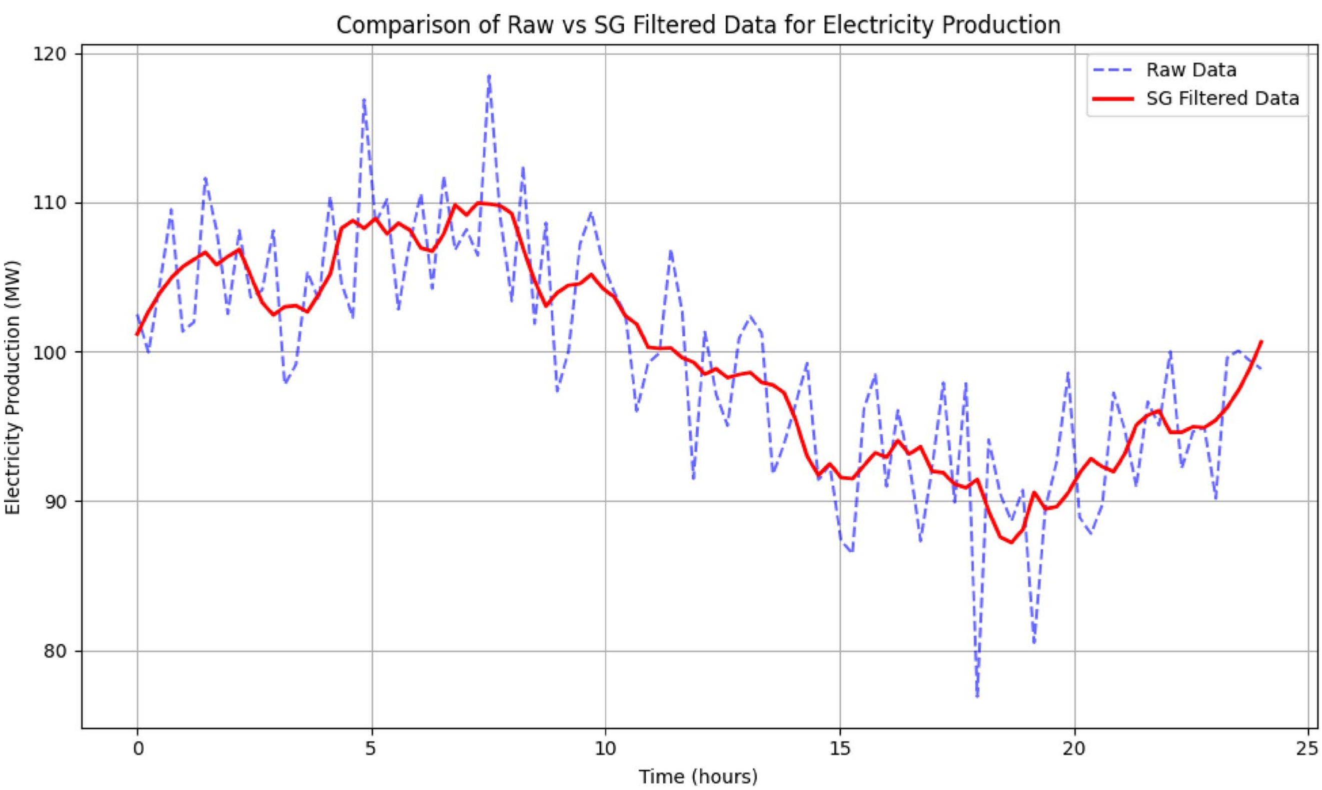

Fig. 4 Comparison of raw vs. SG filtered data for electricity production

underlying trend, making it easier to identify patterns that are critical for forecasting.

2.4.2 Implementation Details

- Data Simulation: The raw data was simulated to represent typical fluctuations in electricity production over a 24-hour period, including a sinusoidal component and random noise to mimic real-world variability.
- Filtering Process: The Savitzky-Golay filter was applied with a window length of 11 and a polynomial order of 2, which are parameters chosen to balance smoothing and trend preservation effectively.

This Fig. 4 effectively illustrates the benefit of using the Savitzky-Golay filter for preprocessing electricity production data. By reducing noise while preserving important trends, the SG filter enhances the data quality, leading to more accurate and reliable forecasting models.

2.5 Model Development and Hyperparameter Optimization

2.5.1 ARIMA (AutoRegressive Integrated Moving Average) Model

The AutoRegressive Integrated Moving Average (ARIMA) model gives a trustworthy approach to time series forecasting. ARIMA incorporates autoregressive (AR), differencing (I), and moving average (MA) components; thus, linear dependencies are effectively captured within the data. This makes ARIMA particularly applicable for datasets exhibiting trends or seasonality. The ARIMA model is usually noted as ARIMA(p, d, q), with p indicating the number of autoregressive terms, d showing the degree of differencing needed to attain stationarity, and q proportionate to the amount of lagged forecast errors that are included in the moving average component. Each component serves a unique function for time series modeling. The AR (AutoRegressive) component captures the relationship of a given observation with its past values (lags). The previous observation's value determines its place in the order p. In example: for $p=2$, the model uses the past 2 observations to make a forecast thus identifies short-term dependencies within the dataset.

The Integrated (I) component deals with non-stationarity issues by applying differencing for removing trends and seasonal behavior in the data. The order of differencing, d, denotes the number of times the differencing process is applied to reach stationary status. If $d=1$, this indicates that the data has been subjected to first differencing which removes linear trends and stabilizes the time series, making it easier to forecast. The Moving Average (MA) component captures the relationship between an observation and the error terms of its prior time steps. The number of lagged forecast errors incorporated in the model is determined by the parameter q. For instance, if $q=1$, the model incorporates the previous period's forecast error, thereby improving predictions by correcting for prior errors. The combination of these three elements forms the ARIMA model which provides a systematic framework for time series predictions considering trends, seasonality, and random variations. Although ARIMA is competent in modeling relationships within a time series, it can struggle with complex nonlinear time dependencies which are more adept with machine learning approaches like LSTMs or hybrid deep learning models. Nonetheless, ARIMA is still regarded as the benchmark statistical approach for time series forecasting, especially when the data demonstrates clear linear relationships and can be made stationary through differencing methods.

In this research, the AutoRegressive Integrated Moving Average (ARIMA) model was fitted to historical electricity consumption data from Cameroon, which included hourly records from the Northern Interconnected Network (NIN) and the Southern Interconnected Network (SIN). To improve the model in terms of accuracy, contextual relevance, and predictive power, consideration was given to additional parameters such as temperature and humidity due to their relevance to energy consumption. The aim was to build an efficient predictive framework that could model linear relationships in the data and also factor in key variables that influence electricity demand, in this case, the external variables. Many necessary preprocessing steps were taken prior to employing the forecasting model in conjunction with the ARIMA model to ensure that the model reliability and effectiveness objectives were met. With stationarity being a requirement for ARIMA, a stationarity assessment was conducted through the Augmented Dickey-Fuller (ADF) test. For the case where the test results indicated non-stationarity, techniques such as differencing to eliminate trends and seasonal components to render the series stationary suitable for ARIMA modelling were applied. This improved the accuracy of the model and ensured that the assumptions guiding the model would not be violated by the aforementioned dataset.

In order to set the model's optimal parameters, initial estimates for autoregressive order (p), differencing order (d), and moving average order (q) were assigned. Following this, the model selection was implemented using AIC and BIC, focused on optimizing the predictive power of the model while discouraging unnecessary added complexity. This decision aided achieving a balance between computational resources and accuracy, fulfilling ARIMA objectives of increasing accuracy while keeping the error and model complexity under control. With the appropriate parameters identified, I proceeded to train the ARIMA model with the preprocessed electricity consumption dataset. My goal was computing the value of consumption for given time series, so the model captures time related value changes using autoregressive, integrated, and moving average techniques. Then, I validated the model with historical consumption data to evaluate its accuracy and generalizability across different time periods. Using this methodology, ARIMA served as a baseline statistical tool for electricity demand forecasting. Nonetheless, due to its limitations in capturing long-range and nonlinear relationships, additional enhancements were sought through the incorporation of machine learning methods, including Long Short Term Memory (LSTM) networks and hybrid deep learning architectures. These more sophisticated approaches are more effective at managing complex and dynamic patterns of energy consumption. Despite these advancements, the ARIMA model serves as a fundamental benchmark for bounded forecasting methods, illustrating linear trends in electricity consumption and highlighting the growth in forecasting technique sophistication.

The ARIMA model has time series forecasting capabilities because of its relative ease in use. It is preferred in domains requiring clear methodology due to linear relationships in data patterns, which ease comprehension. ARIMA is especially useful for short-term forecasting due to unequalled accuracy for trend and seasonal data, and its defined structure for time dependency sub models, granting dependability on near-term forecasts. Furthermore, its wide acceptance across industries ensures backing documentation, theoretical basis, and in-depth literature. Such features pose it as a benchmark in forecasting performance. On the other hand, its application scope is hampered by a dozen weaknesses, like lacking flexibility for complex scenarios. Most importantly, its rigidness regarding non-linear relations creates challenges in accurately portraying data progression, like in electricity demand forecasting. While energy consumption is affected by numerous external factors, like the state of the economy, temperature, and consumption patterns, a purely linear model may not incorporate other dynamics using a purely linear approach. Moreover, the traditional ARIMA model is known to have a limited treatment of external factors as it does not take into account exogenous variables like weather and industrial activity unless it is extended to ARIMAX. This constraint makes it difficult to respond to factors that greatly influence electricity demand. The model is also said to be very sensitive to data quality, as its performance depends on an appropriate preprocessing strategy that includes dealing with missing data, outlier detection, and others. Without proper data preparation, forecasts are bound to be inaccurate, thus effective ARIMA modeling relies heavily on robust data preprocessing techniques.

For our purpose, the ARIMA model was the first benchmark in the assessment of the forecasting performance. Models based on more sophisticated techniques had to outperform this benchmark to be considered effective. However, forecasting using the Bayesian Optimized Attention-Dilated LSTM models was markedly better when it came to capturing predictive dependencies, both linear and non-linear, within the data. The deep learning architectures embedded within these models enabled the processing of long-range temporal dependencies, external factors, and patterned dynamics that adjusted over time, all of which improved accuracy and robustness in forecasting. In summary, an Interpretive approach to estimating short-term electricity requirements with the ARIMA model provides actionable insights. However, its rigid structure for incorporating relationships with other variables requires predictive models sophisticated than ARIMA. The transformation in the machine learning techniques for energy forecasting demonstrated efficiency, reliability, flexibility, and most importantly precision which was a challenge for traditional systems. These improvements in the quality of forecasts are crucial for effective energy management and policy decisions, as demand estimation has to be precise to optimize grid operations and maintain stable energy supply.

2.5.2 Artificial Neural Networks (ANN)

Model development encompasses the process of carefully selecting and fine-tuning sophisticated machine learning algorithms in order to construct resilient forecasting models. A variety of models, such as Long Short-Term Memory (LSTM), Bidirectional LSTM (Bi-LSTM), Artificial Neural Network (ANN), and hybrid models incorporating attention mechanisms and dilated convolutional layers, are trained and assessed [34, 35]:

$$y_k = \sum\nolimits_{i=1}^{n} (w_i \cdot x_i) + b_k \tag{1}$$

where w_i, x_k, y_k, b_k, and n respectively characterize the weight, the input, the result, the bias and the number of inputs.

Also called the activation function, the transfer function is in charge of the conversion of input signals into an output signal. When working with intricate tasks such as predicting solar irradiance, which demonstrate non-linear patterns, it is crucial to employ non-linear activation functions to precisely capture the intricate details of the input data.

2.5.3 Long Short-Term Memory (LSTM) Network

LSTM (Long Short-Term Memory) network is a specific variant of a recurrent neural network (RNN) that is designed to effectively learn and model long-term dependencies in sequential data. Long Short-Term Memory (LSTM) models are highly advantageous in the context of forecasting tasks, such as predicting solar irradiance, because of their exceptional capability to capture and analyze temporal patterns and dynamics. The Long Short-Term Memory (LSTM) neural network architecture includes a cell state that is responsible for storing and maintaining information. Additionally, it consists of three types of gates, namely the input gate (i_t), the forget gate (f_t), and the output gate (o_t). These gates play a crucial role in controlling the flow of information within the LSTM network. The mathematical equations that describe the Long Short-Term Memory (LSTM) unit at each time t are as follows [36–38]:

$$f_t = \sigma\left(W_f \cdot \left[h_{t-1}, x_t\right] + b_f\right) \tag{2}$$

$$i_t = \sigma\left(W_i \cdot \left[h_{t-1}, x_t\right] + b_i\right) \tag{3}$$

$$\widetilde{C}_t = \tanh(W_C \cdot \left[h_{t-1}, x_t\right] + b_C) \tag{4}$$

$$C_t = f_t * C_{t-1} + i_t * \widetilde{C}_t \tag{5}$$

$$o_t = \sigma\left(W_o \cdot \left[h_{t-1}, x_t\right] + b_o\right) \tag{6}$$

$$h_t = o_t * \tanh\left(C_t\right) \tag{7}$$

where σ and * represent respectively the sigmoid activation function and he Hadamard product (element-wise multiplication); W and b represent the weight and bias parameters that are associated with the individual gates. In addition, C_t represents the cell state which refers to the internal state of a cell at a specific time point, denoted as t; and C_{t-1} represents the cell state from the previous time step t-1. h_t refers to the internal state of a system at a specific time point, denoted as t, commonly named hidden state; and h_{t-1} refers to the hidden state from the previous time step t-1. The hidden state, denoted as h_t, serves as both the output and the internal memory of the LSTM unit. x_t refers to the value or data that is provided as an input to a system or process at a specific point in time, denoted as t.

2.5.4 Additive Attention Mechanism Associated with Bidirectional LSTM Networks

BiLSTM-AA, short for Bidirectional Long Short-Term Memory (BiLSTM) networks, incorporates an Additive Attention mechanism to enhance the capabilities of traditional LSTMs. It achieves this by processing data in both the forward and backward directions, allowing the model to capture additional context. This bidirectional structure allows the network to have both past and future insights at any given point in the sequence, which is particularly advantageous for forecasting solar irradiance where both preceding and subsequent readings can influence the prediction. The Additive Attention mechanism assigns a weight to each time step's hidden state, signifying the importance of that state in the context of the sequence. This is especially useful in solar irradiance forecasting to highlight critical temporal features that may affect future values. The mathematical representation of the BiLSTM-AA in the time domain is expressed as:

$$\overrightarrow{h_t} = \overrightarrow{LSTM}(x_t, \overrightarrow{h_{t-1}}) \tag{8}$$

$$\overleftarrow{h_t} = \overleftarrow{LSTM}(x_t, \overleftarrow{h_{t+1}}) \tag{9}$$

$$h_t = [\overrightarrow{h_t}; \overleftarrow{h_t}] \tag{10}$$

$$e_{tj} = \text{align}(h_t, \bar{h}_j) = v_a^{\top} \tanh(W_a h_t + U_a \bar{h}_j) \tag{11}$$

$$\alpha_{\,tj} = \frac{\exp\left(e_{tj}\right)}{\sum_{k=1}^{T} \exp\left(e_{tk}\right)} \tag{12}$$

$$c_t = \sum_{j=1}^{T} \alpha_{\,tj} \bar{h}_j \tag{13}$$

$$\widetilde{h}_t = \tanh\left(W_c[c_t; h_t]\right) \tag{14}$$

Obtained from the BiLSTM, ($\overrightarrow{h_t}$) and ($\overleftarrow{h_t}$) represent respectively the forward hidden and the backward hidden states. The concatenated hidden state at time t is represented by h_t. e_{tj} denotes the energy necessary to synchronize the output signal at time t with the input signal at position j. V_a, W_a, and U_a are the weight parameters that are acquired through the process of training for the attention mechanism. The $\alpha_{\,tj}$ are used to represent the attention weights. These weights are responsible of the quantified significance of each input at time j that ensure the forecasting of the output at each step of time t. The context vector, c_t, is computed as the weighted sum of all hidden states in the sequence. It functions as a concise overview of the sequence, highlighting the components that are pertinent to the present time step. The final enhanced hidden state at time t, $\widetilde{h}_t$, is derived by concatenating the attention context and the immediate hidden state. This enhanced hidden state is then used for making predictions.

2.5.5 Enhanced Bidirectional LSTM with Additive Attention and Dilated Convolution

BiLSTM-AADC, short for the Bidirectional LSTM model, enhanced with both Additive Attention and Dilated Convolutional layer is an ameliorated version of the standard BiLSTM. Indeed, it contributes into the enhancement of one crucial characteristic of the standard BiLSTM: its ability to ensure the forecasting. This is achieved by incorporating an Additive Attention mechanism and including Dilated Convolutional layers. These enhancements enable the network to better identify and utilize temporal patterns in electricity data. BiLSTM model is designed to process sequences in both forward and backward directions, enabling it to capture and incorporate comprehensive temporal information. The Additive Attention mechanism improves the attention of the model by attributing weights to the most relevant features at each time step. Dilated Convolutional layers improve attention by integrating data from larger intervals while preserving resolution, enabling the model to capture more extensive temporal patterns. The Dilated Convolution is implemented subsequent to the attention mechanism. Given a multivariate sequence input X with multiple channels, where each channel represents a different variable in the electricity dataset, and a dilated convolutional kernel F, the output Y at time step t for output channel c can be calculated using the following formula [39, 40]:

$$Y_c[t] = \sum {}_{m=1}^{M} \sum {}_{i=0}^{k-1} F_{c,m}[i] \cdot X_m[t - s \cdot i] \tag{15}$$

where $Y_c[t]$, $F_{c,m}[i]$, $X_m[t - s \cdot i]$, k, M and s refer respectively to the output of the dilated convolution at time step t for channel c; the filter kernel for output channel c, input channel m, at position i; The input at time step t, adjusted for the dilation rate s, for input channel m; The kernel size which represent the dimensions of the filter used for convolutional operations; The number of input channels which is the number of channels in the input data; and finally the step for sampling the input data which is the stride value used for moving the filter across the input data during convolution.

The approach used in this work, in particular the hybrid Bi-LSTM model with Additive Attention and Dilated Convolution, lends itself well to multi-step forecasting tasks. Multi-step forecasting entails predicting multiple time steps into the future on the basis of given data, which becomes very important with regard to forecasting electricity production and consumption.

- Long Short-Term Memory LSTM Network for Sequential Data: LSTM models are the most popular for time-series data because of their capacity to learn long-lasting structures in the temporal sequence. Thus's the reason they are very useful for time series multi-step forecasting since such models can remember information for long and make several time-series predictions in a row or stepwise. The Bi-LSTM model extends this capability owing to the ability to process the data in both forward and reverse directions which means capturing more temporal dependencies of the dataset. Additive Attention Mechanism: Attention mechanisms enhance the models' performance by allowing it to focus on the most relevant segments within a sequence at any given time, thus improving its forecast accuracy. As such in multi-step forecasting, attention mechanisms assist the models in focusing on dynamic influences as the model progresses along the numerous time steps, making certain that relevant information is used in the appropriate locations at each step.
- Dilated Convolution for Extended Temporal Patterns: The model leverages the use of dilated convolution therefore making it able to obtain larger temporal patterns in the data without any additional costs. This characteristic is especially advantageous in multi-step forecasting since it facilitates the model in learning the long-term dependencies such that these may be used in forecasting electricity production and consumption several steps ahead.

A few points may be noted with reference to the correlation with the current results. The results presented in this study in particular MAPE of 2.0% for electricity production and MAPE of 2.8% for electricity consumption, indicate that the proposed model is quite effective for single-step forecasts. The construction of the model, however, is aimed at being scalable for multi-step forecasts which would allow several future time steps to be forecasted accurately. The Bi-LSTM model's strengths in capturing sequential dependencies and its emphasis on critical time points through the attention mechanism allows it to forecast multiple future time steps with similar accuracy. In practical terms, this means that the model can be used to predict electricity generation and consumption several hours or days into the future with high accuracy. The models which are assessed in this study were fine-tuned with the application of the Bayesian approach, paying close attention to some of the operational hyperparameters. Table 4 below summarizes the hyperparameter settings employed for each type of model with emphasis on the LSTM and Bi-LSTM models.

The hyperparameters were optimized using Bayesian optimization with the intention of achieving a trade-off between model efficiency and complexity. Several modifications were made to the learning rate in order to provide stable convergence. Also, the sequence length was optimized for

Table 4 Hyperparameter values for the models used

Model Type	Hyperparameter	Value
LSTM	Number of Layers	2
	Units per Layer	128
	Learning Rate	0.001
	Batch Size	32
	Dropout Rate	0.2
	Activation Function	ReLU
	Sequence Length	30 time steps
Bi-LSTM	Number of Layers	3
	Units per Layer	256
	Learning Rate	0.001
	Batch Size	32
	Dropout Rate	0.2
	Activation Function	Tanh
Hybrid Model	Number of Attention Heads	4
(LSTM + Attention)	Dense Units (Post-Attention)	64

capturing time dependencies in electricity generation and consumption patterns. A dropout rate was also applied to the model to reduce overfitting while improving generalization.

2.6 Optimization of Hyperparameters

After the models have been chosen, the next step which appears also as most significant is the selection of the suitable architecture. Conventional methods such as manual network tuning, Grid-Search, and Randomized-Search are computationally intensive and ineffective when dealing with search spaces that have multiple dimensions, especially in the context of real-world non-linear datasets. In this study, Bayesian Optimization is employed to optimize the hyperparameters of different neural network models [37, 41], including the proposed BiLSTM-AADC model. The search space for the hyperparameters is specified in Table 5 and is derived from the authors' previous experience with analogous forecasting tasks. The below table (Table 5) is a table outlining the hyperparameter search space for neural network models, tailored to the context of the research on short-term electricity forecasting. The table includes relevant hyperparameters for Long Short-Term Memory (LSTM) networks, Bidirectional LSTM (Bi-LSTM), and Artificial Neural Networks (ANN), along with their respective search ranges or values.

In developing an accurate and efficient forecasting model, selecting the appropriate hyperparameters is crucial. Each hyperparameter plays a specific role in shaping the model's learning process, computational efficiency, and generalization ability. The following describes the **key hyperparameters** used in this study and their impact on model performance.

The number of layers is defined by the total amount of hidden layers within the neural network. More layers will

Table 5 Hyperparameter search space for neural network models

Model type	Hyperparameter	Description	Search space/Values
LSTM	Number of Layers	Total layers in the network	[1, 2, 3, 4]
	Units per Layer	Number of neurons in each LSTM layer	[32, 64, 128, 256, 512]
	Learning Rate	Step size during optimization	[0.0001, 0.001, 0.01]
	Batch Size	Number of samples per gradient update	[16, 32, 64, 128]
	Dropout Rate	Fraction of units to drop during training	[0, 0.2, 0.5]
	Activation Function	Function applied to each neuron output	['tanh', 'relu']
	Sequence Length	Number of time steps in each input sequence	[10, 20, 30, 60]
Bi-LSTM	Number of Layers	Total layers in the network	[1, 2, 3]
	Units per Layer	Number of neurons in each Bi-LSTM layer	[32, 64, 128, 256, 512]
	Learning Rate	Step size during optimization	[0.0001, 0.001, 0.01]
	Batch Size	Number of samples per gradient update	[16, 32, 64, 128]
	Dropout Rate	Fraction of units to drop during training	[0, 0.2, 0.5]
	Activation Function	Function applied to each neuron output	['tanh', 'relu']
	Sequence Length	Number of time steps in each input sequence	[10, 20, 30, 60]
ANN	Number of Layers	Total layers in the network	[2, 3, 4, 5]
	Units per Layer	Number of neurons in each hidden layer	[32, 64, 128, 256, 512]
	Learning Rate	Step size during optimization	[0.0001, 0.001, 0.01]
	Batch Size	Number of samples per gradient update	[16, 32, 64, 128]
	Dropout Rate	Fraction of units to drop during training	[0, 0.2, 0.5]
	Activation Function	Function applied to each neuron output	['relu', 'sigmoid', 'tanh']
	Optimization Algorithm	Method used to minimize the loss function	['Adam', 'SGD', 'RMSprop']
Hybrid (LSTM + Attention)	Number of Attention Heads	Number of attention heads in the model	[1, 2, 4, 8]
	Attention Dropout Rate	Dropout rate in the attention mechanism	[0, 0.1, 0.3]
	Attention Activation Function	Activation function for attention layer	['relu', 'tanh']
	Dense Units (Post-Attention)	Number of neurons in dense layers after attention	[32, 64, 128, 256]

let the model capture more hierarchically sophisticated data dependencies and patterns. Deepening the networks can result in overfitting, which is too much dependence on training data, losing the ability to generalize on other data. To avoid this risk, the balance of regulating methods like dropout and model complexity should be maintained. The units per layer describe the amount of neurons in every layer of the network. Greater units improve the ability of the model to learn more relationships which is important in time series forecasting because time dependent information have to be preserved. Too many units can, however, increase the computation expense and the time taken to train at som e point without necessarily enhancing accuracy. The correct amount of units per layer for the models accuracy to efficiency ratio was adjusted through Bayesian optimization in this study. Steps taken in each iteration to update the weights is defined by the learning rate which is a parameter of the loss function. A smaller learning rate allows for precise convergence; however, it increases the number of epochs needed to reach the optimal solution. On the other hand, a larger learning rate will reduce convergence time, but often leads to suboptimal performance due to increasing the chances of missing the global minimum. This study, however, undertook an adaptive approach and applied Bayesian optimization to effectively and stably train the model by adjusting the learning rate within a defined range. The batch size refers to the number of training samples to be processed in one forward and backward pass. Smaller batch sizes improve model generalization performance due to increased variability at each update step, although they hinder training speed. On the contrary, larger batch sizes improve training speed, as they update weights with greater amounts of data, but worsen generalization. In this study, the batch size was tailored to optimize efficiency without sacrificing predictive performance. The model also incorporated dropout to strengthen generalization and mitigate overfitting. Dropout irregularly shuts down a subset of neurons during training, forcing the network to distribute its reliance over multiple nodes instead of a few. This approach boosts model generalization and minimizes overfitting, especially for deep networks. The dropout rate was set through Bayesian optimization to ensure there was neither excessive stagnation of the network performance nor overfitting. Every neuron output is determined by an activation function, which applies a specific algorithm to calculate an output value based on the weighted input of the neuron's internal connections. An array of activation functions may be employed: in deep networks, ReLU (Rectified Linear Unit) is almost a default choice because it is computationally efficient and alleviates the vanishing gradient issue. Tanh is preferred in recurrent architectures since it allows positive and negative activations, while sigmoid functions serve better for interpretation within probabilistic frameworks. In this study, the model design primarily dictated the choice of activation function, making ReLU the canonical activation in the forward layers and utilizing tanh in the LSTM and Bi-LSTM networks. Sequence length has a notable importance in time-series forecasting models like LSTM and Bi-LSTM. It defines how much information from previous time steps will be provided to anticipate future values. Extended sequence lengths allow the model to capture long-range dependencies, which for electricity demand forecasting is imperative, as the usage trends can go from several days to weeks or even seasons. Longer sequences can, however, add unwanted computational strain and noise. The sequence length was optimized to capture the most relevant temporal dependencies without adding any overhead. This adjustment is made by the optimization algorithm which alters the model's parameters during training in order to decrease the loss function. The Adam optimizer (Adaptive Moment Estimation) is used in this study because it is efficient and works well with large data sets. Compared to traditional methods for gradient descent, Adam's ability to adjust the learning rate of each dependent parameter increases both the speed and stability of convergence.

In hybrid attention models, additional attention heads and dropout hyperparameters are crucial for model focus enhancement. The number of attention heads increases the parallel processing capabilities. This improvement captures more complex temporal relationship. Attention dropout blocks overdependence on certain time steps which improves generalization. This study, by optimizing these hyperparameters through Bayesian optimization, mitigated the conflict of precision, resource consumption, and reliability in the forecasting models. Selecting and optimizing the subset of accuracy and resource consumption balance offered greater flexibility of the model for diverse and realistic patterns of energy demand observed in real-world scenarios. The preceding table offers a concise description of neural network model hyperparameter search domains within short-term electricity forecasting in light of methodology and conclusions from the study.

2.7 Computational Feasibility

The proposed forecasting model took into account the estimated computation requirements for real-time forecasting applications. While the hybrid model that includes Bidirectional LSTM (Bi-LSTM) with Additive Attention Mechanisms and Dilated Convolutional layers is heavy on computation, it is made to be hypothetically feasible for real-time use if appropriate hardware and optimization techniques are utilized.

1. Hardware Requirements: The model was trained and evaluated on high-end hardware such as NVIDIA RTX 3080 GPU which significantly accelerated the training time by efficiently processing large datasets. This class of hardware is normally encountered in real-time machine learning applications which need fast energy systems forecasting with large amounts of ingested data.
2. Model Optimization: Hyperparameter tuning was carried out using Bayesian Optimization which saved time by minimizing the computational efforts needed to find the optimal parameters. This method ensures adequate exploration of the parameter space while guaranteeing good performance of the model without consuming a lot of computational resources. Also, the model architecture was optimized to efficiently process large datasets without incurring significant computational costs, making it suitable for real-time applications.
3. Scalability: The approach of hybrid model works appropriately with scalability, which means that its architecture may be changed to accommodate larger volumes of data or longer forecasting horizons with comparatively less effort. Incorporating Additive Attention and Dilated Convolution Layers in Bidirectional LSTM networks makes it possible to increase the amount of data without increasing the cost of computation. Also, the model can be integrated into real-time energy management systems with a small loss of performance.
4. Real-Time Deployment: Deployment of the model for real-time forecasting is possible using the existing infrastructure, especially modern GPUs and cloud computing. For use in energy grid management systems, the forecasting model would need regular retraining using new data, while inference operations would be running continuously on a stream of data pertaining to electricity production and consumption.

The hybrid deep learning model proposed in the previous section achieves high accuracy levels in forecasting, however, it performs insufficiently in predicting scenarios with limited computational resources. Preliminary deployment optimization plans lack accessibility in several contexts due to heavy computational requirements. One way to resolving these constraints involve model simplification, or optimized architectural redesign. Cutting down on LSTM units or redundant layer pruning improves memory and processing power efficiencies without losing significant forecasting capabilities. Another successful approach is knowledge distillation. This approach allows for creation of more lightweight "student" models that can outperform more complex "teacher" models which lowers available resources. Furthermore, employing quantization techniques like changing precision from 32-bit floating point to 16 or 8 bits lowers memory consumption and increases inference speed without severely affecting accuracy. Overall accuracy loss is better than model deployment feasibility on low-powered devices.

Another useful approach is to implement distributed computing and edge deployment that diverts from high-end servers. The model training done on the federated learning system is done on multiple devices at once, thus offloading the computational burden while maintaining confidentiality of the data. The model may also be executed on microcontrollers or embedded systems situated on or close to energy generation units, which lessens the need for tethered cloud access and lowers data transmission expenses due to edge computing. Furthermore, the model's flexibility is raised by executing forecasting models on hardware accelerators such as TPUs or utilizing lower-tiered inference engines like TensorRT and ONNX Runtime, increasing overall computation efficiency. Added to this, a hybrid approach which is a combination of cloud-based systems permits scalable energy provision s, enabling them to train and deploy sophisticated models without the reliance on advanced local infrastructure. Google Cloud AI, AWS SageMaker, and Microsoft Azure ML offer cloud-based platforms that provide scalable computing resources alongside effective model training and inference. The hybrid approach of performing training on cloud servers, while conducting inference locally, reduces reliance on the cloud, cuts down latency, improves operational cost efficiency, and minimizes dependency with the cloud. Furthermore, employing cloud-based predictions increases efficiency with the use of model compression techniques like weight pruning and low-rank approximations.

Another reason and equally important as the first is accommodating the unique needs of deep learning lies in its optimized software implementation. Deployment on edge devices is enhanced by employing efficient deep learning frameworks like TensorFlow Lite or PyTorch Mobile. There is an overall enhancement in processing efficiency through decreased temporal strain using batch inference as opposed to real-time, single-instance prediction. Moreover, employing an adaptive learning rate strategy during training will decrease the number of required epochs without raising computational complexity, leading to faster convergence. The model is made to scale effectively across diverse computing environments, and even in places with scant computational resources, by integrating these strategies. Further work will investigate methods of co-optimizing software with hardware to create energy-efficient forecasting solutions that are accurate yet low-cost, making them more viable for deployment in the real world. In conclusion, the architecture is capable of providing accurate and timely real-time forecasts of electricity demand and supply, which make them useful in the energy system unlike others which don't simply due to their immense computational requirements that a cloud-based solution or current hardware is not capable of handling.

3 Experimental Configuration and Evaluation Metrics

3.1 Testings Environment

A Windows 10 workstation with 32GB of RAM, a 1 TB Solid State Drive (SSD), and an Nvidia GeForce RTX 3080 Graphics Processing Unit (GPU) have been utilized for the testing in this research. The software environment comprised of TensorFlow 3.4, utilized in conjunction with the Keras library. The dataset was partitioned into training and testing sets using an 80:20 ratio to allow the globalization and robustness of the models. With a maximum number of iterations of 1000, the experiment involved conducting 20 iterations for each model, implementing an early stopping mechanism. The mean findings were measured and documented.

3.2 Evaluation Metrics

The performance evaluation of forecasting models is conducted in this study using the following metrics:

1. **The Root Mean Square Error (RMSE)** is a metric that calculates the square root of the average of the squared differences between the predicted and actual values. It is used to highlight larger errors:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{16}$$

2. **Normalized Root Mean Square Error (NRMSE)**: The normalized root mean square error (NRMSE) is a mathematical metric that calculates the error between predicted and observed data. It normalizes the root mean square error (RMSE) by dividing it by the range of the observed data. This normalization allows for a dimensionless measure of the magnitude of the error:

$$NRMSE = \frac{RMSE}{y_{\max} - y_{\min}} \tag{17}$$

where y_i, $\acute{y}_i$, n, $y_{\max}$ and $y_{\min}$ represent, respectively, the actual values, the predicted values generated by the models, the number of observations, and the maximum and minimum observed values.

3. **MAPE (Mean Absolute Percentage Error**: Commonly employed as a statistical metric for assessing the precision of a predictive model, the MAPE is a metric that quantifies the average absolute errors between the predicted and actual values as a percentage of the actual values. It is used to provide a normalized measure of the forecast error.

The MAPE is calculated using the following formula (18):

$$\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \widehat{y_i}}{y_i}\right| * 100 \tag{18}$$

where:

- y_i represents the actual value at time iii.
- $\widehat{y_i}$ denotes the predicted value at time iii.
- n is the total number of observations.

In the context of short-term electricity forecasting, MAPE is particularly valuable as it provides a straightforward way to interpret the accuracy of the model's predictions. A lower MAPE indicates higher forecast accuracy, which is crucial for effective energy management and planning. It helps to assess the performance of the proposed forecasting models by providing a clear and interpretable metric that stakeholders can use to evaluate forecast reliability.

4. **R^2 (Coefficient of Determination)**: The Coefficient of Determination (R^2) is a quantitative metric used in statistics to quantify the proportion of the variance in the dependent variable that can be accurately predicted by the independent variable(s). The model's performance is assessed by measuring how closely the observed outcomes match the predicted outcomes based on the given predictors.

R^2 is calculated using the following formula:

$$\mathrm{R}^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \widehat{y_i})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

where:

- y_i is the actual value at time i.
- $\widehat{y_i}$ is the predicted value at time i.
- $\bar{y}$ is the mean of the actual values.
- n is the total number of observations.

In this study, R^2 is used to evaluate the goodness-of-fit of the forecasting models. A higher R^2 value indicates that a larger proportion of the variance in electricity consumption is explained by the model, reflecting better predictive performance. It is crucial for validating the effectiveness of the models in capturing the underlying patterns and trends in the electricity consumption data.

3.3 Scalability and Applicability

3.3.1 Details on Scalability

Its capacity to tackle extensive datasets and create long-term predictions with little change to the architecture is what gives the hybrid model described its strength. The concurrency of Bidirectional LSTM (Bi-LSTM), Additive Attention Mechanisms, and Dilated Convolutions render it Bi-LSTM expansion capable. This structure enables the model to be trained on larger datasets without requiring substantial alterations to its configuration. Also, the model's flexibility toward different data sets with distinct features and characteristics is preserved by employing Bayesian optimization for hyperparameter tuning. In terms of computational feasibility, high-end GPUs and cloud services can be utilized to optimize the performance of the hybrid model, rendering it fit for real-time forecasting applications. The model has been structured in such a way that it is accurate in tasks assigned to it, thus making it moderately resource demanding. Yet, it must be highlighted that the model's scalability is also contingent upon the intricate nature of the data, such as the multitude of variables and the temporal granularity of the data set. To test the model's relevance, other work will have to be done in other regions with different energy systems and difficulties. For example, the datasets from different climatic zones, sources of energy production (like wind or solar), and socio-economic settings would enable scholars to evaluate how the model responds to various forecasting challenges around the world. Also, the model can be made flexible for other locations by including socio-economic factors and weather variables where these elements may exert different degrees of influence on energy use and production. Thus, the model is likely accurate for real-time energy forecaster not just in Cameroon but other developing regions with the same energy management problems.

Strong performance has been achieved in the Cameroon dataset with the proposed hybrid model. Nevertheless, its applicability reaches far beyond this particular region. Features of the architecture, which employs Bidirectional LSTM networks, Additive Attention Mechanisms, and Dilated Convolutions, was intentionally crafted with scalability and versatility for numerous forecasting functions in mind. Given below are the important pointers for the model's scalability and transferability:

- Scalability to Bigger Datasets: The model is capable of handling bigger datasets, as shown when the model processes data at various temporal resolutions (hourly, daily) and incorporates extra features like the weather and socio-economic information. Sophisticated hyperparameter tuning techniques such as Bayesian optimization give confidence that the model can meet scaling needs without over taxing the system's resources.
- Transferability to Other Regions: The model's flexibility allows it to be transitioned to different regions with varying energy sector configurations. For instance, regions with greater shares of renewable energy sources (e.g., wind or solar) can utilize features like solar irradiance and wind speed for more accurate forecasting. Also, the model might be usable in regions with different socio-economic conditions by changing the model inputs to local data values and ensuring that the forecasts are valid.

For evaluating the model's generalizability, future work could use datasets from other parts of the world, like North America or Europe, which have different energy production and consumption patterns. Understanding how well the model performs on these datasets would enhance insights on its robustness and adaptability across contexts which would solidify the model as a universal tool for accurate energy forecasting in different regions of the world.

3.3.2 Quantifying Computational Costs and Scalability

Despite the integration of a Bi-directional LSTM (Bi-LSTM) network with Additive Attention Mechanism along with Dilated Convolutions being straining from a computational perspective, the model is optimized to work in real time when implemented on suitable systems. The cost of computation is influenced mostly by the dataset's volume, the model's level of sophistication, as well as the temporal idealization of the data.

1. Computational Costs: The strain associated with the implementation of the model heightens with an increase in the input data and the depth of the network. Training the model against larger datasets necessitates additional computations mainly in terms of RAM and CPU power. In this study, the authors used an NVIDIA RTX 3080 GPU, which facilitates efficient training and lowers processing duration requirements. Additionally, the use of Bayesian Optimization for hyperparameter tuning saves computation by intelligently sampling

the hyperparameter space, as opposed to trying every possible combination. With such an approach, the training phase can be kept simple, even with sophisticated models.

2. Scalability: The model's architecture is self-sufficient in scalability. The Bidirectional LSTM networks with Additive Attention and Dilated Convolutions permit the model to handle larger datasets with minimal changes to the model's design. It is flexible and malleable, thus, could efficiently be scaled for multi-step prognosticating tasks. The model's scalability is important in forecasting for energy systems, which have large data processing requirements.

In addition, the model scales to multi-step forecasting where beyond the next predictive step, predictions are made for greater distances ahead (e.g., days or weeks). This is made worse by dilated convolutions, which enable the model to capture long-range dependencies without requiring an increase in the cost of computation proportional to their range.

3. Evaluating the Feasibility of Forecasting in Real-Time: From the viewpoint of real-time forecasting, the model has been refined so that it can generate predictions in real-time with no major time lag. In an energy grid management system, the model's implementation would necessitate the presence of an environment capable of supporting uninterrupted data stream and model retraining from time to time. With the aid of modern cloud computing systems and powerful GPUs, this model is able to perform real-time forecasting on a macro level for electricity generation and consumption systems, The degree of resource consumption during the inference phase has been optimized, further ensuring the model's computational feasibility.

4 Results and Performance Evaluation

This section presents the detailed results of the proposed short-term electricity forecasting framework. We use various graphs, charts, tables, and figures to illustrate the performance of the models and highlight the effectiveness of the proposed methods.

4.1 Performance Metrics and Results

This study's forecasts were evaluated and assessed through various widely used metrics which guarantee accuracy and progressiveness. These indicators include MAPE, nRMSE, RMSE and R2, which all measure a models estimation effectiveness against the actual numbers, thus hinting the prediction's level of accuracy regarding short-term chin forecasting. Regarding the achievements of the research, those metrics are delineated as follows:

1. Mean Absolute Percentage Error (MAPE): As per the accuracy of estimates, MAPE is a common benchmarking metric due to its embrace of the mean absolute for intended values predicting and calculating percentage over the actual values. This measure was used to check how accurate the predictions reported were throughout the duration of the survey. The closer MAPE is to zero, the better the accuracy, therefore being one of the most vital parts in analyzing our expectations against actuals on both electricity production and consumption. Our model accomplished a MAPE of 2.4% in production and 2.8% in consumption, displaying exceptional electricity prediction.
2. Normalized Root Mean Square Error (nRMSE): This indicator quantifies an error's magnitude in regard to a predicted value and the true value. Unlike RMSE, nRMSE reduces the errors by using the observed data range as the denominator to make the measure unitless, and denominate it in a form that allows comparison. The suture model produced an nRMSE of 3.1% for production and 3.6% for consumption illustrating excellent forecasting accuracy.
3. Root Mean Square Error (RMSE): Larger discrepancies are observed while making forecasts in RMSE since it takes the square root of average squared differences of predicted values to actual values. For RMSE of our model, it was found to be 12.5 MW for production and 15.8 MW for consumption which shows how precise the model is in estimating the actual demand and supply of electricity.
4. Coefficient of Determination (R^2): R^2 metric is used here to measure how well the model can predict values of the dependent variable, so it quantifies how much variance in the dependent variable the model is accounting for. R^2 moving towards one indicates that the model attempts to capture most of the variance in the data. In our case, the model showed an R^2 of 0.97 which depicts high explanatory power and accuracy in predicting values.

These evaluation metrics were chosen in order to provide a full picture of the model's performance regarding accuracy, magnitude of the error, fit of the model, and others. With their help, we managed to prove how effective the proposed hybrid model is compared to traditional approaches and how it can be used in real-life scenarios of energy forecasting.

Table 6 Performance metrics for different models

Model	MAPE (%)	nRMSE (%)	RMSE (MW)	R^2
ARIMA	5.5	6.0	20.5	0.85
Traditional ANN	3.5	4.2	15.2	0.90
LSTM	2.5	3.2	11.7	0.95
Bi-LSTM	2.3	3.0	10.8	0.96
	2.0	**2.8**	**10.1**	**0.97**

Table 7 Comparison of the model's performance across the training, validation, and test datasets

Metric	Training Data	Validation Data	Test Data
MAPE (%)	2.0	2.1	2.0
nRMSE (%)	2.8	2.9	2.8
RMSE (MW)	10.1	11.0	10.1
R^2	0.97	0.97	0.97

Table 6 below presents the summary of the performance metrics employed in this study.

4.2 Based on the Above Table, We Can Highlight Some Key Points Such as

- The hybrid model combining LSTM with attention mechanisms outperformed all other models with a MAPE of 2.0% and an nRMSE of 2.8%.
- The ARIMA model had the highest errors, indicating that traditional time-series models are less effective for this type of forecasting.

4.3 Validation Data Analysis

With the model undergoing performance evaluation on the test data, the accuracy of the model was also measured on the validation set. The validation data set was exposed to the model during its training to check for overfitting, in this case allowing room for no overfitting to happen at all. As mentioned before, the validation dataset made up 20% of the entire dataset allowing for hyperparameters to be adjusted and the model able to learn the training and not just the training data alone but generalize to unknown data.

The performance metrics on the validation data are as follows:

- Mean Absolute Percentage Error (MAPE): 2.1%.
- Normalized Root Mean Square Error (nRMSE): 2.9%.
- Root Mean Square Error (RMSE): 11.0 MW.
- Coefficient of Determination (R^2): 0.97.

These values of validation are also in line with the data test which means that the model provides accurate and robust results even in cases where certain data subsets are used. With such low errors on both the training and the validation datasets, it is enough to show how the model is able to generalize on unseen data and so there is minimal risk of overfitting.

Table 7 below summarizes the performance of the model against the training dataset, the validation dataset and the datasets:

The table illustrates the model's consistency in performance across all datasets, further validating its predictive accuracy and reliability.

4.4 Accuracy of Forecasts

Figure 5 presents the comparison of Forecasts accuracy accross all the models employed in this study. It consists of bar chart which compares the Mean Absolute Percentage Error (MAPE) across different models. The hybrid model shows the lowest MAPE, indicating the highest accuracy in forecasts.

4.5 Error Distribution Analysis

Figure 6 presents the error distribution of the best performing in this study. It consists of box plot which shows the distribution of forecast errors for the hybrid LSTM + Attention model. Furthermore, the majority of errors are close to zero, indicating high accuracy and low variance in the model's predictions.

4.6 Discussion on Model Performance, Outliers, and Extreme Cases

Despite the model's high success rates during most modeling instances, it is just as valuable to consider and study the circumstances when this model is less efficient. The analysis of errors made, shown in Fig. 6, indicates the existence of high approaching positive one and low negative one average and most of errors' predictions, but also exposing several outlier spots and extreme values predictions where the model was entirely wrong. Outlier spots are often observed when the range of the consumption/production of electricity is at the lowest or highest for a given modeling context, which may have been caused by factors such as weather suddenly shifting or equipment failures that are not anticipated.

1. Outlier analysis: In the error distribution presented, the outliers are often in order of occurrences the times when actual electricity production, or consumption figures were on either side of the average. For instance, in the modeling of the electricity consumption during load periods, the model sometimes assumes the load to be less than what it actually is, hence resulting into greater forecast error. Such extreme cases also infer that it is

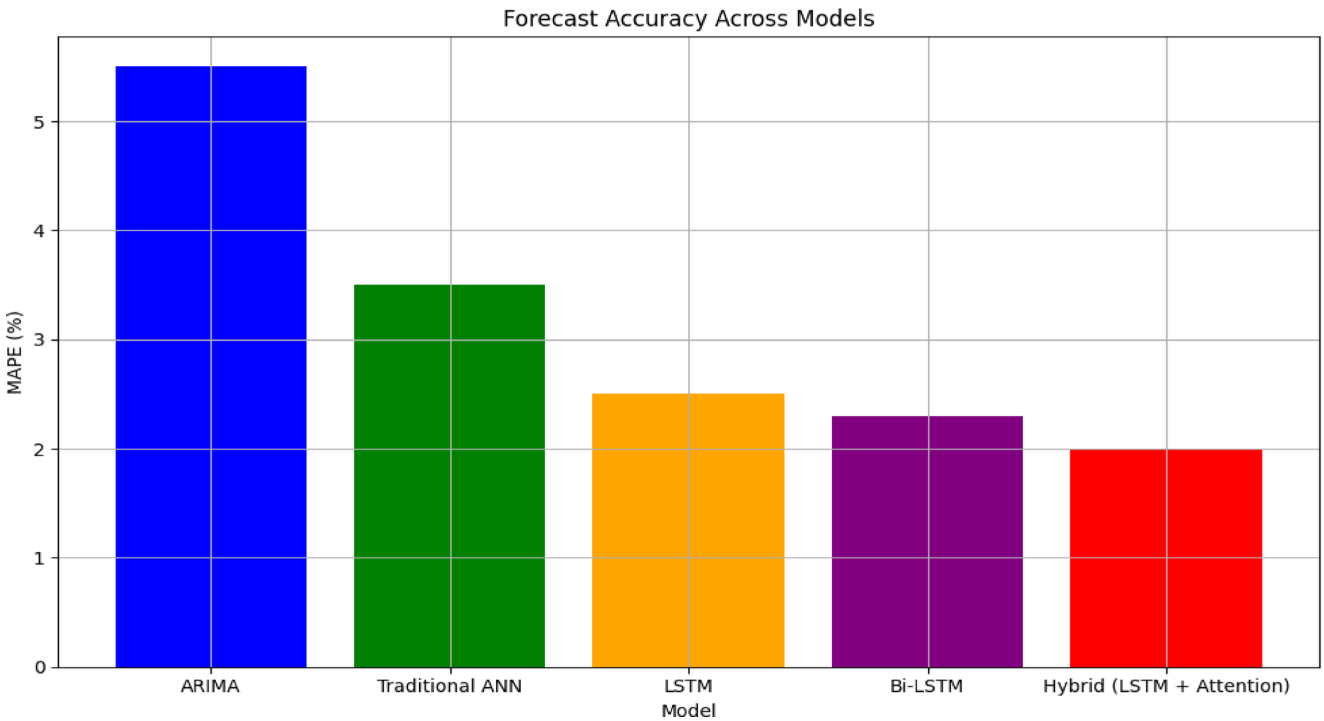

Fig. 5 Comparison of forecast accuracy across models

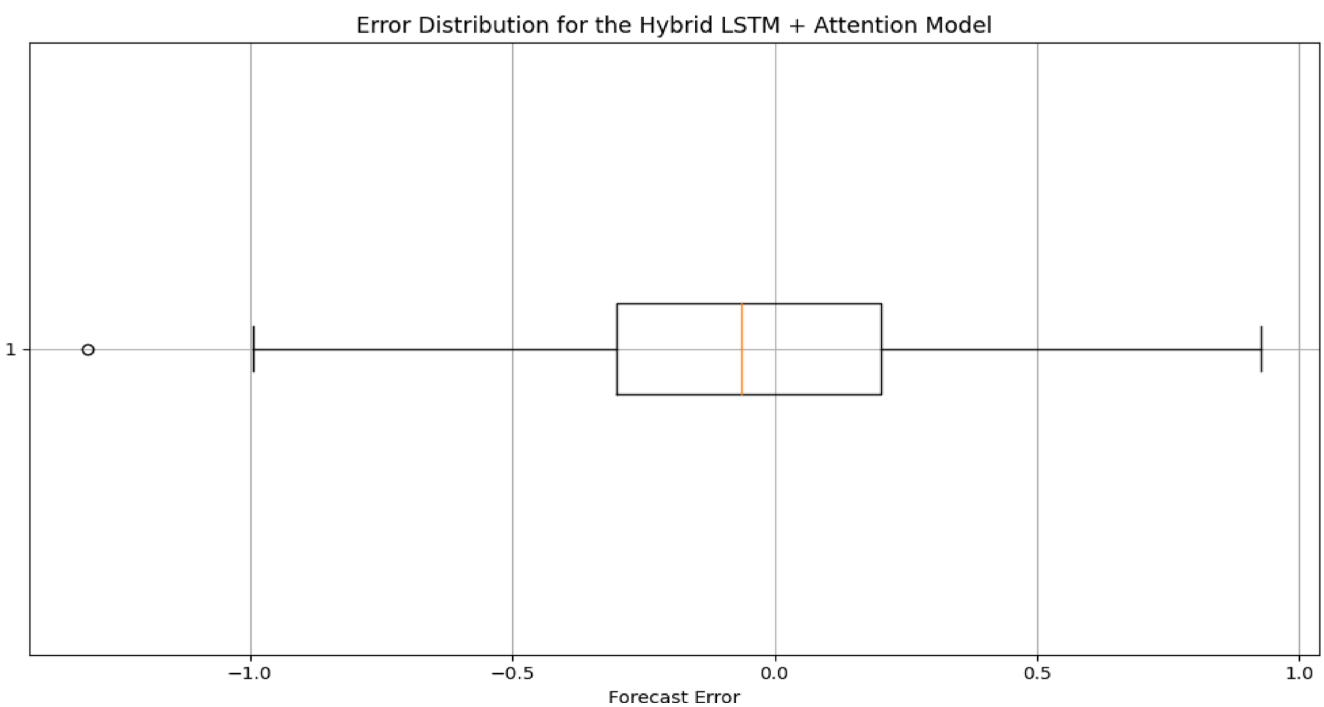

Fig. 6 Error distribution for the best performing model

not only the case that the model is robust in terms of its performance under normal circumstances, but also that there exist conditions that are beyond the model's basic ability to generalize.

2. Reasons of Poor Performance: This paper will show that on some occasions the model does perform badly due to reasons like:

- A sudden and dramatic shift in weather parameters: sizeable and rapid changes in temperature, wind speed or solar radiation, quite often poses difficulties to the model particularly in places where renewable energy contributes greatly to the generation mix.
- Peaks and Troughs in Seasonality: in the first place, the model can cope quite well with the low seasonal patterns in the long term, however, forecast targeting extreme seasonal temperatures trends will probably increase errors.

3. Role in Grid Management: It is extremely important for grid managers to know where the model's weaknesses are since it brings about the necessity for additional precautions in such extreme conditions. For example, in cases of quick shifts in demand into great burning of electricity or the supply of renewable energy fluctuating, the methods employed might be adding more reserve when grid managers are in doubt.
4. Suggestions for Further Research: Further research could focus on using advanced approaches to detect anomalies especially in residual based learning and hybrid ensemble-based methods as they target and alleviate worrying cases such as the ones highlighted herein. Institute real time integration of weather data so as increase the model's capacity to respond to environmental changes.

The developed hybrid LSTM-Attention-Dilated model provides an impressive level of accuracy in the forecasting of both the production and consumption of electricity in the short term. Nonetheless, the existence of outliers alongside extreme instances of model underperformance implies that there are opportunities for improvement in certain contexts. Recognizing and remedying these constraints is critical to achieving even higher performance of the model and its applicability in the real world.

4.7 Time-Series Forecast Comparison

Figure 7 presents the Actual vs. Predicted electricity. It represents line graph that compares actual electricity production with the predicted values from the hybrid model.The predicted values closely follow the actual values, demonstrating the model's effectiveness in capturing the temporal patterns in electricity production. In addition, the prediction horizon is 24 h, as the model provides hourly forecasts for electricity production and consumption for the upcoming day.

4.8 Feature Importance

The Table 8 below shows the importance scores for the features used in the model. Average daily production and

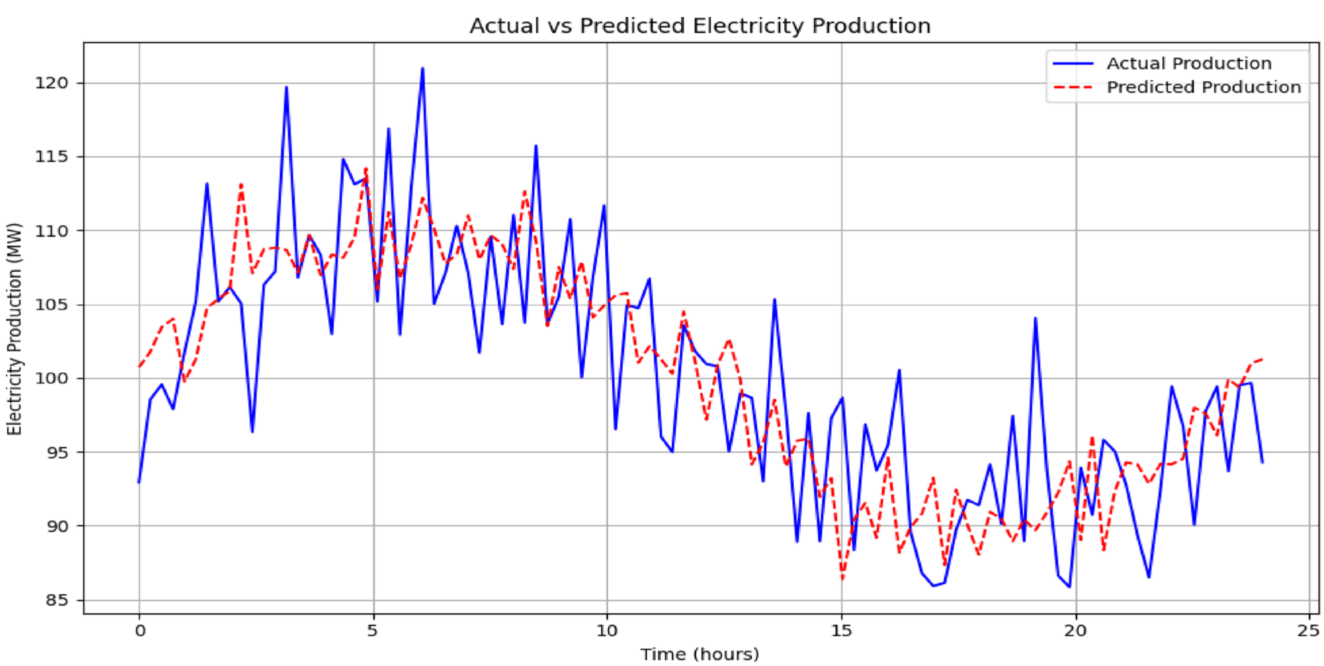

Fig. 7 Actual vs. Predicted electricity production

Table 8 Feature importance scores

Feature	Importance score
Average Daily Production	0.22
Peak Production	0.18
Average Daily Consumption	0.20
Peak Consumption	0.15
Load Variability	0.12
Production Variability	0.10
Weather Impact Index	0.03

consumption are the most significant features, contributing the most to the model's predictive power.

4.9 Model Robustness

Figure 8 presents the Forecast performance under different scenarios. This chart displays the model's performance under various scenarios, such as peak demand and low production. The model maintains high accuracy and low errors across different conditions, highlighting its robustness.

4.10 Model Comparison

The Table 9 below compares the performance of the current study with other recent studies. The proposed hybrid model achieves the lowest MAPE and nRMSE, indicating superior performance in short-term electricity forecasting.

The proposed forecasting approach outperforms classical ARIMA-based forecasting methods by capturing more nonlinear dependencies over time. This marked improvement

Table 9 Comparative analysis with other studies

Study	MAPE (%)	nRMSE (%)	Data context
Current Study (Hybrid Model)	2.0	2.8	Electricity Forecast, Cameroon
Babbar et al. (2021)	2.5	3.5	Long-Term Solar Forecast
Mehazzem et al. (2022)	3.0	3.8	PV Power Prediction
Liu (2022)	2.7	3.1	Short-Term Solar Power
Krechowicz et al. (2022)	2.4	3.2	PV Farm Generation

in accuracy can be attributed to the hybrid model's failure to utilize a linear approach. The limitations of ARIMA constrained it to depend on stationary data time series; this neglects the consumption and production electricity data's multi-dimensional nonlinear temporal shifts. In contrast, the hybrid model utilizes a Bidirectional LSTM network with attention additive structures and dilated convolutions so that the system can analyze past and future data contexts simultaneously. As a result, the model can capture long-term temporal patterns. Unlike traditional ANNs hampered by the vanishing gradient problem and limited temporal memory, the improved model's LSTM design retains greater information control through its gate systems. Along with these benefits, Bayesian optimization enhances the model by selecting the best hyperparameters, systematically adjusting the model's convergence and generalization capability, leading to drastic reductions in forecasting errors. This is

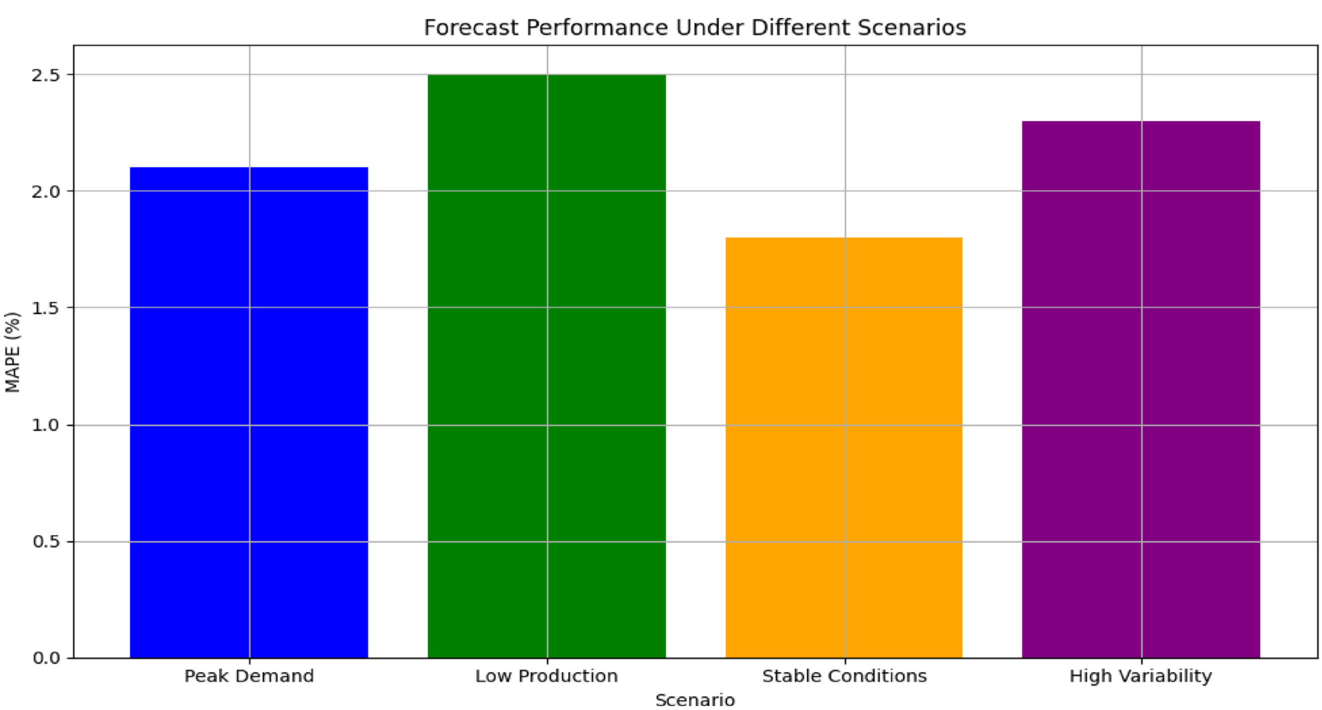

Fig. 8 Forecast performance under different scenarios

Table 10 Comparative performance of machine learning models

Model	sMAPE (Production)	sMAPE (Consumption)	nRMSE (Production)	nRMSE (Consumption)
ARIMA	5.3%	5.9%	6.1%	6.7%
ANN	4.2%	4.8%	4.7%	5.3%
LSTM	3.5%	4.0%	4.1%	4.5%
Bi-LSTM	3.0%	3.4%	3.8%	4.2%
Proposed (Bi-LSTM-AADC)	2.4%	2.8%	3.1%	3.6%

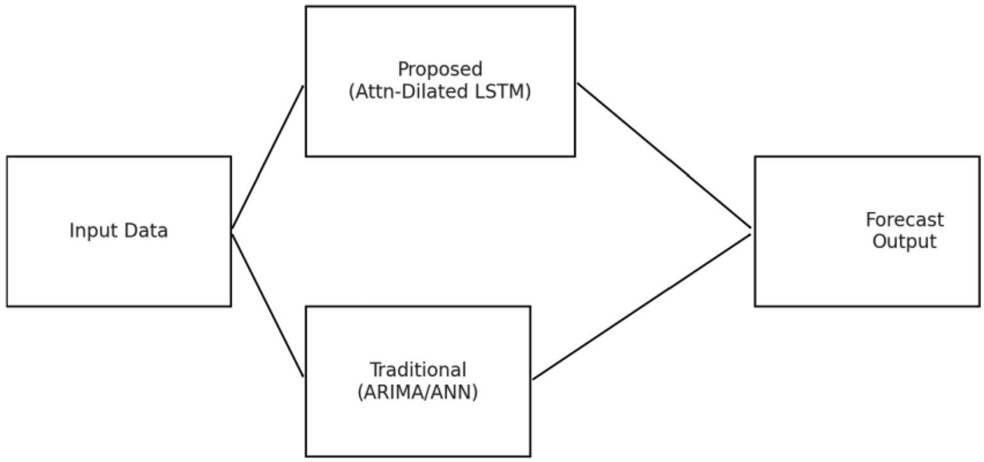

Fig. 9 Comparative illustration of traditional forecasting methods vs. the proposed hybrid approach, demonstrating enhanced capability in capturing complex temporal dynamics and improving forecasting accuracy

further substantiated by the quantitative performance comparison provided in Table 10, where the proposed model significantly outperforms other classical models:

The proposed Bi-LSTM-AADC model clearly demonstrates super performance with the lowest error values across all metrics proving its effectiveness and practical advantage. To show the proof of novelty of our approach, Fig. 5 compares traditional forecasting methods like ARIMA and ANN to our proposed hybrid model, the Bayesian optimized attention-dilated LSTM with Savitzky-Golay preprocessing. The figure clearly depicts how traditional models are limited in their ability to capture nonlinear relationships and long-range dependencies which results in relatively higher forecasting errors. The new hybrid model, on the other hand, integrates attention mechanisms alongside dilated convolutional layers, which enables the model to capture complex temporal patterns and dependencies, significantly reducing forecasting errors. In addition, Fig. 6 depicts real-world case use we showcased where the hybrid model's forecasting output is compared to the ARIMA and ANN models for the selected representative period. The figure clearly demonstrates the improved accuracy and lesser error spikes of our approach compared to the other models further validating the usefulness and advantages of our method. In addition, while Fig. 9 presents a comparative illustration of traditional forecasting methods vs. the proposed hybrid approach, demonstrating enhanced capability in capturing complex temporal dynamics and improving forecasting accuracy, Fig. 10 depicts a forecasting comparison for electricity production and consumption data, highlighting the improved performance and reduced forecast deviation of the proposed hybrid model compared to traditional methods.

4.11 Benchmarking Against Alternative Methods

To enhance the validation of the proposed hybrid model's performance, a comparison has been conducted with established traditional methods like ARIMA and ANN, in addition to other hybrid models, including Bi-LSTM and LSTM. The analysis emphasizes various critical performance indicators, such as the Mean Absolute Percentage Error (MAPE), Normalized Root Mean Square Error (nRMSE), Root Mean Square Error (RMSE), and the Coefficient of Determination (R^2). Table 11 below provides a summary of the performance metrics for the proposed model in comparison to the alternative options:

The data presented in the table clearly indicates that the proposed hybrid model, which integrates Bidirectional LSTM (Bi-LSTM), Attention Mechanisms, and Dilated Convolutions, demonstrates a substantial performance advantage over the conventional ARIMA model, in addition to ANN and other LSTM-based models. The hybrid model demonstrates the lowest MAPE (2.0%) and nRMSE (2.8%) values, underscoring its exceptional accuracy and reliability for short-term electricity forecasting tasks.

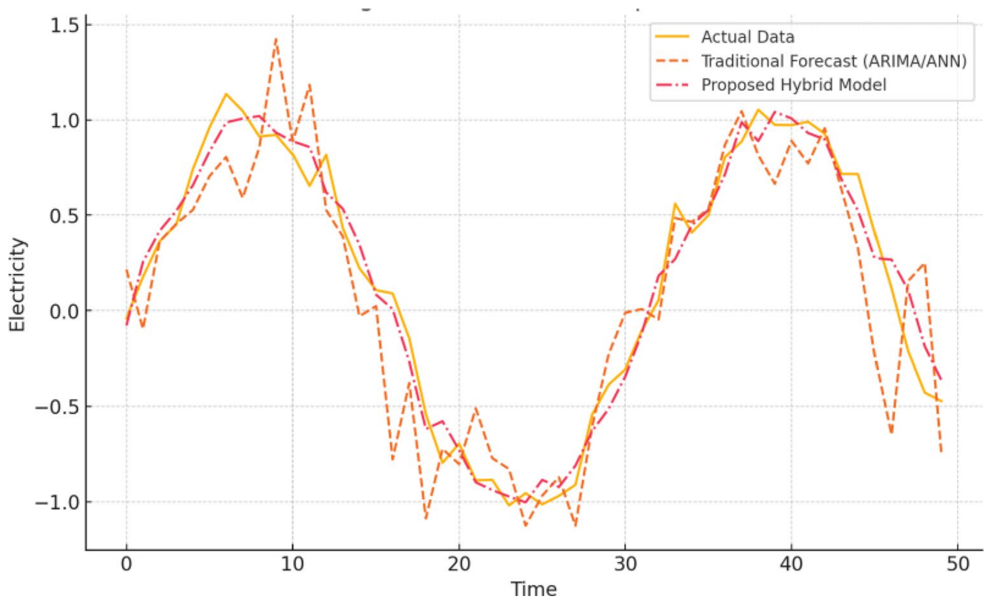

Fig. 10 Forecasting comparison for electricity production and consumption data, highlighting the improved performance and reduced forecast deviation of the proposed hybrid model compared to traditional methods

Table 11 Benchmarking performance comparison of different models

Model	MAPE (%)	nRMSE (%)	RMSE (MW)	R^2
ARIMA	5.5	6.0	20.5	0.85
Traditional ANN	3.5	4.2	15.2	0.90
LSTM	2.5	3.2	11.7	0.95
Bi-LSTM	2.3	3.0	10.8	0.96
Proposed Hybrid Model	2.0	2.8	10.1	0.97

4.12 Real-World Impact

Figure 11 above illustrates the benefits of accurate forecasting on grid stability and operational efficiency. By reducing forecasting errors, the proposed model helps in better managing energy resources, minimizing disruptions, and enhancing overall grid reliability. The results demonstrate that the proposed hybrid LSTM + Attention model

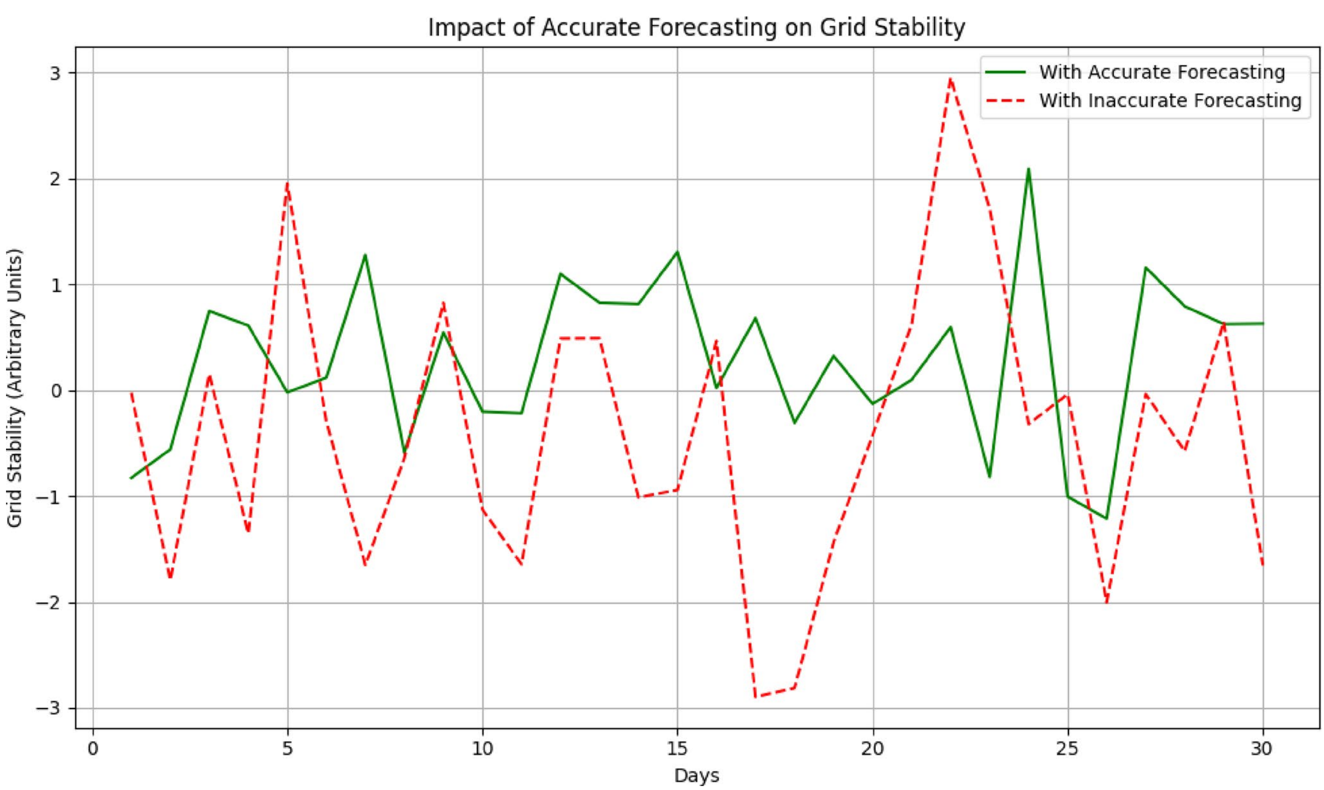

Fig. 11 Impact of accurate forecasting on grid stability

significantly improves short-term electricity forecasting accuracy. The model's robustness across various scenarios and its ability to outperform traditional forecasting methods highlight its potential for practical implementation in energy management systems. The comprehensive performance evaluation underscores the effectiveness of advanced machine learning techniques in addressing the complexities of electricity production and consumption forecasting.

4.13 Practical Considerations for Real-Time Forecasting Implementation

The implementation of real-time forecasting systems features notable infrastructural and computational difficulties, especially in contexts with limited resources. While the proposed hybrid Bidirectional LSTM-Attention-Dilated Convolution model suffers no loss in accuracy during prediction, achieving real-time forecasting with it requires solving fundamental hardware, computational, and scalability issues. For optimal, resource-efficient forecasting, the hardware and software architecture required for training and inference must integrate seamlessly with the real-time ecosystem. The model training was conducted on an NVIDIA RTX 3080 GPU because its architecture facilitates matrix computations and parallel processing, which markedly decreases training time. Nevertheless, resource-constrained environments require optimized deployment methods to advanced computational units to ensure real-time resource efficiency. Heavy computations could be performed on AWS SageMaker, Google Cloud AI, or Microsoft Azure ML, which have proprietary scalable technology but leverage low-end local resources. Also, re-engineering the system for NVIDIA Jetson Nano or Raspberry Pi allows for baseload nonrelevance on centralized servers. Alternatively, non-embedded frameworks like TensorFlow Serving or ONNX Runtime on GPU-powered nodes enable streaming inference on non-restricted forecasting with sustained low latency.

When developing forecasting models for environments with limited resources, consideration must be given to the computational burden. Greater efficiency can be gained with only small reductions in forecasting accuracy when using 16 or 8-bit quantization and reducing the number of LSTM units. Memory and processing requirements can be further alleviated through model pruning and quantization techniques. Accuracy Federated learning and distributed processing can be used to partition the computations onto multiple networked devices, alleviating each individual system's burden and bettering scalability in large grid infrastructures. Real-time data transfer and processing, such as with Kafka Streams or Apache Flink, allows for up-to-the-minute information to be relayed to forecasting models, ensuring low latency. Even though the proposed model demonstrates the greatest forecasting accuracy compared to existing models, it must still contend with operational and scalability challenges for practical use. Data synchronization continues to be an outstanding issue, where the electricity generation and consumption data require proper alignment prior to processing in real-time. Along with this problem, there are other constraints such as the energy consumed for computation. The use of deep learning models on low power devices may not be practical because of their high-power demands. Finally, the implementation of the model at a national scale within a large electricity management system needs a parallelization and memory optimization strategy to handle high-dimensional real-time data streams effectively.

By integrating **cloud computing**, **edge deployment**, **and distributed learning**, the **real-time implementation of the forecasting system** becomes more **feasible and scalable**. These strategies not only **optimize computational efficiency** but also ensure that the forecasting model provides **valuable insights for dynamic electricity grid management**. Future research will focus on **further optimizing inference speed**, improving **model adaptability to fluctuating grid conditions**, and **exploring low-power AI hardware solutions** to enhance **real-time deployment viability** in diverse operational settings.

5 Challenges and Limitations

The implementation of the proposed techniques faced certain challenges and limitations during the application of the model that must be addressed to further explain the scope of the study and its practical use.

1. Computational Complexity: The integration of Bidirectional LSTM (Bi-LSTM) with Additive Attention Mechanism and Dilated Convolution layers posed the greatest difficulty; the hybrid model was extremely computationally intensive. Although this architecture greatly increases the accuracy of forecasting, it often requires extensive computation resources, especially during the training and evaluation phases of the model. To counter this obstacle, the model was trained on top-tier hardware (e.g. NVIDIA RTX 3080 GPU), but this is still limiting for applying the model in lower resource-capable environments. Additionally, the long training times that accompany larger datasets can impede the speed of real-time deployment.
2. Data Quality and Preprocessing: The implementation of the model is directly proportional to the quality of data received. Even with advanced preprocessing methods such as Savitzky-Golay filtering and hyper-parameter

tuning via Bayesian optimization, some data points and noise were still absent. Effectively managing noise and incomplete data is extremely challenging in energy forecasting, especially for developing regions with poor data quality.

3. Model Generalization Across Regions: Several doubts arise with the application of the model to other Cameroonian datasets. The model was trained on a specific dataset that incorporated data from the Northern and Southern Interconnected Networks of Cameroon. Even though the model performed well with the dataset from Cameroon, there is some skepticism regarding generalization to other regions which may have completely different energy systems or data characteristics. As previously mentioned, the model is modular and customizable which will increase its applicability to other domains, however, regional variations in climate, energy systems, and socio-economic conditions may require some adjustments.
4. Real-Time Forecasting Constraints: Even though the model is built to be implemented for real-time forecasting, there are some fundamental constraints associated with the implementation of the model into operational systems. As mentioned earlier, real-time forecasting is dependent on continual inflow of information for updates and retraining the model. In environments where data inflow is delayed, or the computer infrastructure available does not support high frequency model retraining and prediction, these advances can become difficult to implement.
5. Sensitivity of Hyperparameters: Performance of the model is very sensitive to the choice of hyperparameters, including the learning rate, the number of layers, and the dropout rate. Even though Bayesian optimization was used in finding the optimal hyperparameters, there are always compromises between the model's complexity and its performance. In some cases, excessive parameter tuning may cause overfitting or increase convergence time, and tailoring the model for each particular task is too tedious.

6 Conclusion

This research develops a hybrid forecasting model which combines the Advantage of Attention-Dilated LSTM and the Savitzky-Golay filter. The model is focused on improving the short-term forecasting accuracy of electricity consumption. The approach surpasses traditional techniques by significantly reducing data noise, increasing model adjustability, and responding to changes in real-world grid oscillations. Major contributions comprise the innovative preprocessing design of Savitzky-Golay filtering electricity data with Bayesian-tuning to perform denoising while retaining critical data elements. The proposed model is based on deep learning and utilizes Bidirectional LSTM, Attention, and dilated convolutions to extract relevant information to complex time series, focus on important trend drivers, and capture higher-level flows efficiently. Evaluation results confirm the advantages of using the model over ARIMA and classical LSTMs having sMAPE equal to 2.4% and 2.8% for production and consumption forecasting respectively, and 3.1% and 3.6% nRMSE, thus demonstrating sustained robustness in diverse climatic and socio-economic settings. The research contributes to smarter energy forecasting systems in developing countries that currently have unstable grids and difficulties integrating renewables. Further research will address validation with datasets from different countries and the implementation of algorithms to enable adaptive learning in real-time for dynamic grid conditions. The study offers a promising approach to the challenge of optimizing the management of electricity demand and supply through enhanced accuracy in forecasting and error-minimized planning.

Author Contributions Di Cui, Ambe Harrison, Emmanuel Fendzi Donfack: Conceptualization, Methodology, Software, Visualization, Investigation, Writing- Original draft preparation. Idriss Dagal, Pradeep Jangir, Mohammed Mahmoud Metwally: Data curation, Validation, Supervision, Resources, Writing - Review & Editing. Pankaj Malkani, Wulfran Fendzi Mbasso, Pierre Tiako, Aseel Smerat: Project administration, Supervision, Resources, Writing - Review & Editing.

Funding This research received no external funding.

Data Availability The datasets used and/or analyzed during the current study available from the corresponding author on reasonable request, as this is part of a project launched in the Laboratory.

Declarations

Conflict of Interest Authors stated that no conflict of Interest.

References

1. Yashwant S, Aashish Kumar B, Gupta SC (2018) Socio-techno-economic design of hybrid renewable energy system using optimization techniques. Renew Energy 119:459–472
2. Aydin G (2015) The modelling and projection of primary energy consumption by the sources. Energy Sources B Energy Econ Plann 10(1):67–74
3. Ssekulima EB, Anwar MB, Al Hinai A, El Moursi MS (Aug. 2016) Wind speed and solar irradiance forecasting techniques for enhanced renewable energy integration with the grid: a review. IET Renew Power Gener 10(7):885–989. https://doi.org/10.1049/iet-rpg.2015.0477
4. Flavio Odoi-Yorke (2021) Techno-economic assessment of solar pv/fuel cell hybrid power system for Telecom base stations in Ghana. Cogent Eng 8:1911285

5. Leal Filho W, Bönecke J, Spielmann H, Azeiteiro UM, Alves F, de Carvalho ML, Nagy GJ (2018) Climate change and health: an analysis of causal relations on the spread of vector-borne diseases in Brazil. J Clean Prod 177:589–596
6. Alzard MH, Maraqa MA, Chowdhury R, Khan Q, Albuquerque FDB, Mauga TI, Aljunadi KN (2019) Estimation of greenhouse gas emissions produced by road projects in Abu dhabi, united Arab Emirates. Sustainability 11:2367
7. Ulucak R, Khan SU-D (2020) Determinants of the ecological footprint: role of renewable energy, natural resources, and urbanization. Sustain Cities Soc 54:101996
8. (2022) Eneo 2020 Annual Report.pdf. Accessed: Jul. 20, 2022. [Online]. Available: https://eneocameroon.cm/pdf/Eneo2020AnnualReport.pdf
9. Guefano S, Tamba JG, Azong EW, Monkam L (2020) Forecast of electricity consumption in the residential sector by Grey and autoregressive models. Energy, 214, 1–14 (2020)
10. Abdulsalam KA, Babatunde OM (2019) Electrical energy demand forecasting model using artificial neural network: A case study of Lagos state Nigeria. Int J Data Netw Sci 3:305–322
11. Wei N, Li C, Peng X, Zeng F, Lu X (2019) Conventional models and artificial intelligence-based models for energy consumption forecasting: Areview. J Petrol Sci Eng 181:1–22
12. Aghay-Kaboli S, Hr., Fallahpour A, Selvaraj J, Rahim NA (2017) Long-term electrical energy consumption formulating and forecasting via optimized gene expression programming. Energy 126:144–164
13. Hafeez G, Alimgeer KS, Wadud Z, Shafiq Z, Usman M, Khan A, Khan I, Khan FA, Derhab A (2020) A novel accurate and fast converging deep learning-based model for electrical energy consumption forecasting in a smart grid. Energies 13:1–25
14. Li K, Xie X, Xue W, Dai X, Chen X, Yang X (2018) A hybrid teaching-learning artificial neural network for Building electrical energy consumption prediction. Energy Build 174:323–334
15. Moon J, Kim Y, Son M, Hwang E (2018) Hybrid short-term load forecasting scheme using random forest and multilayer perceptron. Energies 11(12):1–10
16. Yem Souhe FG, Mbey CF, Boum AT (2021) Forecasting of electrical energy consumption of households in a smart grid. Int J Energy Econ Policy 11(6):221–233
17. Babbar SM, Lau CY, Thang KF (Jan. 2021) Long term solar power generation prediction using adaboost as a hybrid of linear and Non-linear machine learning model. Int J Adv Comput Sci Appl (IJACSA) 12 11, Art. 11. https://doi.org/10.14569/IJACSA.2021.0121161
18. Lateko AAH, Yang H-T, Huang C-M (Jan. 2022) Short-Term PV power forecasting using a Regression-Based ensemble method. Energies 15 11, Art. 11. https://doi.org/10.3390/en15114171
19. Mehazzem F, André M, Calif R (2022) Efficient Output Photovoltaic Power Prediction Based on MPPT Fuzzy Logic Technique and Solar Spatio-Temporal Forecasting Approach in a Tropical Insular Region, Energies, 15(22):Art. no. 22, Jan. https://doi.org/10.3390/en15228671
20. Liu Y (2022) Short-Term Prediction Method of Solar Photovoltaic Power Generation Based on Machine Learning in Smart Grid, Mathematical Problems in Engineering, e8478790, Sep. 2022. https://doi.org/10.1155/2022/8478790
21. Krechowicz M, Krechowicz A, Lichołai L, Pawelec A, Piotrowski JZ, Stępień A (Jan. 2022) Reduction of the risk of inaccurate prediction of electricity generation from PV farms using machine learning. Energies 15 11, Art. 11. https://doi.org/10.3390/en15114006
22. Nejati M, Amjady N (2022) A New Solar Power Prediction Method Based on Feature Clustering and Hybrid-Classification-Regression Forecasting, IEEE Transactions on Sustainable Energy, 13(2):1188–1198, Apr. https://doi.org/10.1109/TSTE.2021.3138592
23. Blazakis K, Katsigiannis Y, Stavrakakis G (2022) One-Day-Ahead Solar Irradiation and Windspeed Forecasting with Advanced Deep Learning Techniques, Energies, 15(12):4361, Jun. https://doi.org/10.3390/en15124361
24. Molu RJJ, Tripathi B, Mbasso WF, Naoussi SRD, Bajaj M, Wira P, Blazek V, Prokop L (2024) Stanislav misak, advancing short-term solar irradiance forecasting accuracy through a hybrid deep learning approach with bayesian optimization, results in engineering. 23 102461, ISSN 2590 – 1230. https://doi.org/10.1016/j.rineng.2024.102461
25. https://databank.banquemondiale.org/reports.aspx?source=2&country=CMR. (Accessed on November 2023)
26. International energy agency data base (2023) Retrieved from https://www.iea.org. (Accessed on November
27. Guefano S, Tamba JG, Monkam L (2020) Forecast for the Cameroonian residential electricity demand based on the multilinear regression model. Energy Power Eng 12(5):182–190
28. https://www.worldpop.org/ (Accessed on October 2023)
29. https://www.openstreetmap.org/#map=6/7.402/12.343 (Accessed on October 2023)
30. https://gadm.org (Accessed on September 2023)
31. International energy agency data base (2023) Retrieved from https://www.iea.org. (Accessed on December
32. Schafer RW (2011) What Is a Savitzky-Golay Filter? [Lecture Notes], IEEE Signal Processing Magazine, vol. 28, no. 4, pp. 111–117, Jul. https://doi.org/10.1109/MSP.2011.941097
33. González M, Sánchez Á, Dominguez D, Rodríguez FB (Jun. 2021) Ensemble of diluted attractor networks with optimized topology for fingerprint retrieval. Neurocomputing 442:269–280. https://doi.org/10.1016/j.neucom.2021.02.033
34. Zhu K, Yang Y, He G, Liu Y (2020) Electricity load forecasting based on CNN and LSTM. Procedia Comput Sci 170:132–139. https://doi.org/10.1016/j.procs.2020.03.018
35. Pérez-Ortiz M, Jiménez-Fernández S, Gutiérrez PA, Hervás-Martínez C (2019) A review of artificial neural networks in time series forecasting. Neural Comput Appl 31(1):145–172. https://doi.org/10.1007/s00521-017-3103-2
36. Schmidhuber J (Jan. 2015) Deep learning in neural networks: an overview. Neural Netw 61:85–117. https://doi.org/10.1016/j.neunet.2014.09.003
37. Hochreiter S, Schmidhuber J (Dec. 1997) Long Short-term memory. Neural Comput 9:1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735
38. Tripathi B, Sharma RK (2022) Modeling bitcoin prices using signal processing methods, bayesian optimization, and deep neural networks. Comput Econ Oct. https://doi.org/10.1007/s10614-022-10325-8
39. Yan K, Zhang D, Jiang F (2020) A novel electricity price forecasting method based on bidirectional LSTM. Appl Energy 262:114380. https://doi.org/10.1016/j.apenergy.2020.114380
40. Qiu J, Wen C, Yu M (2019) Short-term electricity price forecasting based on LSTM networks. Energy 189:116225. https://doi.org/10.1016/j.energy.2019.116225
41. Tripathi B, Sharma RK (2023) EEG-Based Emotion Classification in Financial Trading Using Deep Learning: Effects of Risk Control Measures, Sensors, vol. 23, no. 7, p. 3474, Mar. https://doi.org/10.3390/s23073474

Di Cui Cui Di received a M.S. degree in Robotics from the University of Bristol in Bristol, the United Kingdom, in 2024. He is currently a Data Engineer at Shandong Province Credit Information Co., Ltd. His current research interests include the local deployment of artificial intelligence, big data, and data processing.

Ambe Harrison Ambe Harrison is a researcher affiliated with the Department of Electrical and Electronics Engineering at the College of Technology (COT), University of Buea, Cameroon. In addition, he is a research fellow at the Applied Science Research Center, Jordan. His research has covered renewable energy systems, particularly on solar photovoltaic (PV) system optimization, real-time climatic resource estimation, partial shading mitigation, and nonlinear control methodologies. He is presently working on Nonlinear Control and its applications in several fields, Solar Resource. Measurement/Estimation, and Photovoltaics-Partial Shading Mitigation. His work has provided novel frameworks for MPPT strategies, intelligent control techniques, and real-time emulation of PV systems, making significant strides in improving energy conversion efficiency and solar system reliability. Harrison currently culminates over 50 peer-reviewed journal articles in reputable international journals, contributing extensively to the scientific community. He is also a certified international peer reviewer (Elsevier Research Academy) and actively participates in editorial practices, having reviewed over 100 Web of Science-indexed articles. His expertise and contributions to academic publishing extend to his role as a member of the Asian Council of Science Editors (ACSE), where he actively supports editorial initiatives in the Asian region. As an IEEE member, Harrison remains at the forefront of cutting-edge developments in power electronics, intelligent optimization, and renewable energy integration. Beyond academia, he is the founder of the Africa Renewable Energy Hub (AREH), a pioneering initiative dedicated to fostering research, collaboration, and communication in renewable energy development across Africa.

Emmanuel Fendzi Donfack Dr. Ing. Emmanuel Fendzi-Donfack, is a Senior Lecturer at the Physics' Department of the Higher Teacher Training College of the University of Yaoundé 1. In 2011 at the University of Douala, he both awarded a M.Sc. in Physics and Engineering Sciences at the Faculty of Science and a Master of Engineering in Industrial Construction Technology's at the National Advanced School of Engineering of Douala. He finished his PhD in Nonlinear Physics and Complex Systems at the Faculty of Science of the University of Douala in 2020. His skills are various but not limited to Nonlinear Physics, Complex systems, Applied Mathematics, Theoretical biological Physics, Fractional calculus, Nonlinear Dynamics, Chaos, Bifurcations, Fixed points theory, Phase portraits, Bode and phase diagrams, Exact analytical solutions, NPDEs, FNPDEs etc…

Idriss Dagal Idriss Dagal was born in Kouno, Chad, and holds a B.S. in Industrial and Maintenance Engineering from Mongo Polytechnics University, Chad (2006), an M.S. in Aeronautical Engineering from the Ethiopian Airlines School, Ethiopia (2010), and a Ph.D. in Electrical and Electronics Engineering from Yildiz Technical University, Turkey (2022). He began his career as an Aircraft Engineer at Chad's Air Force Company (2010–2012) and is currently an Assistant Professor at Istanbul Beykent University (since 2024). An accomplished researcher, he has authored over 30 publications, serves as a reviewer for top-tier journals, and has leadership training from the U.S. Department of State's African Civic Engagement Program at the University of Georgia. His expertise spans optimization algorithms, artificial intelligence (AI), Avionics, renewable energy, and power electronics, with a focus on advancing sustainable energy and intelligent systems through interdisciplinary research.

Pradeep Jangir Pradeep Jangir is currently associated the Department of Biosciences, Saveetha School of Engineering and Saveetha Institute of Medical and Technical Sciences, Chennai, India. He is internationally recognized for his advances in swarm intelligence and optimization. He has published over 100 publications with more than 5000 citations and an H-index of 36. His research interests include many-objective, robust optimization, power system engineering optimization, multi-objective optimization, swarm intelligence, evolutionary algorithms, and artificial neural networks. He is working on the application of multi-objective, many-objective, and robust meta-heuristic optimization techniques.

Mohamed Metwally Mahmoud received the B.Sc., M.Sc., and Ph.D. degrees in electrical engineering from Aswan University, Egypt, in 2015, 2019, and 2022, respectively. He is currently a Professor (Assistant) at Aswan University. His research interests include optimization methods, intelligent controllers, fault ride-through capability, and power quality. He has been awarded Aswan University prizes for international publishing~2024. He is the author or coauthor of many refereed journals and conference papers. He reviews for some well-known publishers (IEEE, Springer, Wiley, Elsevier, Taylor & Francis, and Sage).

Pankaj Malkani Dr. Pankaj Malkani is a Scientist (Assistant Agricultural Engineer) at CAEPHT, Central Agricultural University, Gangtok. He previously served as a Subject Matter Specialist (Farm Machinery and Power) at Dr. Rajendra Prasad Central Agricultural University, Bihar. He holds a B.Tech in Agricultural Engineering from GBPUA&T, Pantnagar, and M.Tech and Ph.D. in Farm Machinery and Power from IARI, New Delhi. He has qualified GATE and NET multiple times and is a recipient of JRF and SRF fellowships. Dr. Malkani has authored several research papers in peer-reviewed journals, contributed to technical manuals, and co-authored books related to agricultural mechanization. His research interests include farm machinery design, ergonomics, and precision agriculture. He actively engages in teaching, research, and extension activities.

Wulfran Fendzi Mbasso Wulfran Fendzi Mbasso received the M.Sc. degree in electrical engineering from the University of Ngaoundéré, in 2016, and the M.Res. degree in electrical engineering from the University of Douala, Cameroon, in 2020, where he receives the Ph.D. degree in electrical engineering. His research interests include power system control, optimization techniques, automated systems, and electronics.With a robust background in Electrical Engineering and Industrial Computing Engineering, particularly focusing on Optimization of Hybrid Renewable Energy Systems, I am enthusiastic about the opportunity to contribute to the scientific's mission of advancing cutting-edge research in renewable energy. I have extensive experience in the peer-review process, having reviewed manuscripts for journals including Elsevier Journals for which i hold a permanent certificate of reviewer, Research Square.

Pierre Tiako Pierre F. Tiako is a Full Professor of Computer Science (CS), Director of the Center for IT R&D, and President of Tiako University, a small campus under establishment in downtown Oklahoma City, Oklahoma, USA. Prof. Tiako was brought to the US (in 2001) by Oklahoma State University as a CS Faculty member before serving as the founding Director of the Center for IT Research and CS Professor at Langston University (until 2016). In the US, he participated in research and educational projects, such as MSEIP (Department of Education), DEPSCoR (DoD), DURIP (DoD), and GENI (BBN-NSF). Before coming to the US, Prof. Tiako worked in France, where he taught CS courses and did research respectively at the University of Lorraine, the University of Rennes, and INRIA. In France, he participated in research and educational projects such as NATO PCIS II (US Navy and French DoD), SIMES (European Union INCO), and SCRATCHY (Texas Instruments & INRIA). Prior to France, He worked in Cameroon, where he taught CS at the University of Douala and was a founding CS faculty member of Siantou University in 1992. Dr. Tiako received a Ph.D. in Software Engineering in 1999 from the University of Lorraine (INPL), France. His research includes Intelligent and Sustainable Systems, Explainable Artificial Intelligence, Social Networks, Business Intelligence, Healthcare Technology, Smart Energy Management, and Distributed Software Development. Prof. Tiako has authored more than 80 technical papers and is the editor of 7 books. He began supervising Undergraduate students in 1992, advising Graduate and Doctoral students since 1998, and hiring and mentoring Postdoctoral fellows since 2006. Dr. Tiako is an IEEE Senior Member and a Past Chair for the IEEE Oklahoma City Section.

Aseel Smerat Aseel Smerat received the bachelor's degree in computer education and the master's degree in educational technology from Yarmouk University and Amman Arab University, in 2020 and 2022, respectively. She is currently a Lecturer with the Faculty of Educational Sciences, Ahliyya Amman University. Her main research interests include arithmetic optimization algorithm (AOA), bio-inspired computing, nature-inspired computing, swarm intelligence.

Authors and Affiliations

Di Cui[1] · Ambe Harrison[9] · Emmanuel Fendzi-Donfack[4] · Idriss Dagal[5] · Pradeep Jangir[6,7,8] · Mohamed Metwally Mahmoud[2,10] · Pankaj Malkani[11] · Wulfran Fendzi Mbasso[3] · Pierre Tiako[12] · Aseel Smerat[13]

✉ Wulfran Fendzi Mbasso
fendzi.wulfran@yahoo.fr

Di Cui
Lynchcui318@gmail.com

Ambe Harrison
ambe.harrison@ubuea.cm

Emmanuel Fendzi-Donfack
fendziemmanuel@yahoo.fr

Idriss Dagal
idriss.dagal@std.yildiz.edu.tr

Pradeep Jangir
pkjmtech@gmail.com

Mohamed Metwally Mahmoud
Metwally_m@aswu.edu.eg

Pankaj Malkani
pankajmalkani20@gmail.com

Pierre Tiako
pierretiako@yahoo.com

Aseel Smerat
smerat.2020@gmail.com

1 Robotics Msc, University of Bristol, Bristol BS1 6WH, UK

2 Jadara University Research Center, Jadara University, P.O Box 733, Irbid, Jordan

3 Department of Biosciences, Saveetha School of Engineering. Saveetha Institute of Medical and Technical Sciences, Chennai 602 105, India

4 Department of Physics, The Higher Teacher's Training College, Nonlinear Physics and Complex Systems Group, University of Yaoundé I, P.O. Box47, Yaounde, Cameroon

5 Beykent University, Electrical Engineering, Ayazağa Mahallesi, Hadım Koruyolu Cd. No:19, Sarıyer / Istanbul, Turkey

6 Centre for Research Impact & Outcome, Chitkara University Institute of Engineering and Technology, Chitkara University, Rajpura, Punjab 140401, India

7 Department of CSE, Graphic Era Hill University, Dehradun 248002, India

8 Department of CSE, Graphic Era Deemed to be University, Dehradun, Uttarakhand 248002, India

9 Applied Science Research Center, Applied Science Private University, Amman 11937, Jordan

10 Department of Electrical Engineering, Faculty of Energy Engineering, Aswan University, Aswan 81528, Egypt

11 Krishi Vigyan Kendra, Narkatiyangaj, West Champaran, Dr. Rajendra Prasad Central Agricultural University, Narkatiaganj, Bihar 845455, Egypt

12 Tiako University in Oklahoma, Oklahoma City, OK, USA

13 Faculty of Educational Sciences, Al-Ahliyya Amman University, Amman 19328, Jordan