BIG DATA MINING AND ANALYTICS
ISSN 2096-0654 01/15 pp247–270
DOI: 10.26599/BDMA.2023.9020030
Volume 7, Number 2, June 2024

Predicting Energy Consumption Using Stacked LSTM Snapshot Ensemble

Mona Ahamd Alghamdi*, Abdullah S. AL-Malaise AL-Ghamdi, and Mahmoud Ragab

Abstract: The ability to make accurate energy predictions while considering all related energy factors allows production plants, regulatory bodies, and governments to meet energy demand and assess the effects of energy-saving initiatives. When energy consumption falls within normal parameters, it will be possible to use the developed model to predict energy consumption and develop improvements and mitigating measures for energy consumption. The objective of this model is to accurately predict energy consumption without data limitations and provide results that are easily interpretable. The proposed model is an implementation of the stacked Long Short-Term Memory (LSTM) snapshot ensemble combined with the Fast Fourier Transform (FFT) and meta-learner. Hebrail and Berard's Individual Household Electric-Power Consumption (IHEPC) dataset incorporated with weather data are used to analyse the model's accuracy with predicting energy consumption. The model is trained, and the results measured using Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and coefficient of determination (R^2) metrics are 0.020, 0.013, 0.017, and 0.999, respectively. The stacked LSTM snapshot ensemble performs better than the compared models based on prediction accuracy and minimized errors. The results of this study show that prediction accuracy is high, and the model's stability is high as well. The model shows that high levels of accuracy prove accurate predictive ability, and together with high levels of stability, the model has good interpretability, which is not typically accounted for in models. However, this study shows that it can be inferred.

Key words: Artificial Intelligence (AI); Deep Learning (DL); energy consumption; snapshot ensemble; prediction

1 Introduction

National development projects, population growth, and urbanization are all driven by energy, making it a highly sought-after resource. Globally, homes and buildings make up a substantial proportion of the energy consumed when they are in operation and contribute to global warming and carbon emissions. Energy performance in many countries is important. But homes and buildings must be designed first to accommodate the occupants' comfort. Where lower energy consumption is preferred, it is a secondary aspect. Therefore, among decision-makers and

• Mona Ahamd Alghamdi is with Information Systems Department, Faculty of Computing and Information Technology (FCIT), King Abdulaziz University (KAU), Jeddah 21589, Kingdom of Saudi Arabia. E-mail: monagh5454@gmail.com.
• Abdullah S. AL-Malaise AL-Ghamdi is with Information Systems Department, FCIT, KAU, Jeddah 21589, Kingdom of Saudi Arabia, and also with School of Engineering, Computing and Design, Dar Al-Hekma University, Jeddah 22246, Kingdom of Saudi Arabia. E-mail: aalmalaise@kau.edu.sa.
• Mahmoud Ragab is with Information Technology Department, FCIT, KAU, Jeddah 21589, Kingdom of Saudi Arabia, and also with Mathematics Department, Faculty of Science, Al-Azhar University, Naser City 11884, Egypt. E-mail: mragab@kau.edu.sa.
* To whom correspondence should be addressed.
Manuscript received: 2023-01-19; revised: 2023-06-04; accepted: 2023-10-16

academic researchers in the energy sector, energy efficiency is a concerning topic that is critical for achieving the low-carbon economy target (green economy)[1].

Many governments appreciate the benefits of efficiently using energy. Efficient energy use affects the capacity of a building to acquire a green building certificate, which is based on the green building rating systems intended to minimize greenhouse effects and carbon emissions. In this regard, predicting energy use is critical for planning, conservation, and management. Also, with increased demand, the call for better energy consumption planning comprises improved consumption measurement and distribution planning. The capacity to optimize and predict energy consumption can aid with energy distribution to support the increased energy demand[1].

Various studies have been conducted to further develop better energy utilization and execution in structures[2–4]. To create applications for smart buildings, buildings must use smart devices, and artificial intelligence and engineering-based methods are typically used to predict energy consumption. Engineering methods and models use principles like thermodynamic equations to forecast energy consumption. For energy performance evaluation, these models and methods frequently necessitate expertise in customizing them to meet specific requirements and programming the thermal parameters. Engineering models and techniques require in-depth information about the building envelope, the thermal properties of the windows and construction layers, and the ventilation, heating, and air-conditioning systems used to predict energy consumption accurately.

Future energy consumption models based on historical data are referred to by Artificial Intelligence (AI) and Machine-Learning (ML) techniques[5]. The ability of models and algorithms to learn about the relationship between future and historical data is one of the benefits of using AI and ML techniques. The prediction models that are developed are fed historical data, in contrast to engineering techniques that use comprehensive building data. Additionally, users are not required to have a comprehensive understanding of the thermodynamic behaviour of a building.

AI models have been developed to predict energy performance, as suggested by some studies. For instance, Song et al.[6] created an evolutionary model to forecast smart building energy consumption. In 2017, Wang and Srinivasan[5] examined ensemble-based AI models to predict building energy use. Other studies have used models from Deep Learning (DL) to conduct research in the creation of a system for managing energy. Jahani et al.[7] incorporated a numerical moment matching technique with a genetic algorithm to create a tool for predicting residential building energy use.

It is essential that national energy efficiency policies be developed and proposed by evaluating trends in electricity consumption and energy structures[8]. The foundation for minimizing energy costs and maximizing energy performance is the capacity to forecast energy consumption beyond buildings[9]. ML and AI are already being used in the building and energy domain[10–14], where models use historical data to forecast energy consumption and generate new insights. For example, based on the temperature of the skin, machine-learning models can predict thermal demand[10]. Using an optimization model and machine learning to identify energy data patterns, Chou et al.[15] evaluated time series energy data in 2016.

Artificial Neural Networks (ANN), linear regression models, and Support Vector Regression (SVR) models are among the most widely used machine-learning models[5]. Ganguly et al.'s research[14] predicted energy consumption in a historical art gallery using an ANN model. In 2019, Seyedzadeh et al.[16] analysed performance with predicting building cooling and heating loads of ML models using SVR, ANN, random forest, Gaussian process, and Gradient-Boosted Regression Trees (GBRT). It is resolved that GBRT shows the best presentation when utilizing the root mean square mistake values. The researchers also concluded that the ANN model evaluates complex datasets best. Additionally, the ANN model computes significantly faster than the other ML models in their study[16].

ML, a subset of artificial intelligence, allows the machine or system to learn without human intervention. Several DL techniques are used in prediction, including Recurrent Neural Networks (RNN), Deep Neural Networks (DNN), and Deep Belief Networks (DBN). These powerful tools help with acquiring robust modelling and prediction performance. RNN is characterized by taking the output of a previous step and feeding it to the current or next step as input. Its most important feature is its hidden state that retains information about the sequence. However, it becomes untrained when the network contains large datasets and layers. Long Short-Term Memory (LSTM) is a DL technique based on RNN that provides the added advantage of successfully training to overcome the problems found with RNN[16].

Although ML models can produce significant and demonstrated prediction accuracy improvements in many cases, research has mainly focused on improving accuracy without dealing with the interpretability of the results. Currently, expert systems, primarily developed using linguistic fuzzy logic systems, give users the ability to have systems modelling capabilities and good interpretability[12]. However, the systems and models often depend on individual expertise and regularly do not produce accurate predictions. Thus, to meet the requirements of interpretability and high accuracy, it is being proposed to combine popular techniques, expert models, and other methods. Despite the application of DL models in energy and the meeting of the accuracy requirement, there is still a need to improve performance in energy consumption and production application.

ML is a rich field with many models that can be and have been implemented in energy consumption prediction, such as ANN, SVR, linear regression models[5], and RNN models[16], to name just a few. With every prediction model proposed and built, there is a problem solved, but each of these machine-learning models has its own problems. The biggest issue is with the training data. ANN is attributed to the lack of energy training data, SVR is not suitable for large datasets, and linear regression is prone to overfitting and noise. Overfitting is where a model will work well with training data but very poorly with new data. Another problem is when a model will work well with the training data, but if there is less training data than new data, the model underperforms. This is a problem with SVR. In linear regression, linearity is quite important, which leaves nonlinear data as a disadvantage. All these data issues greatly impact the accuracy of the prediction model. RNN[16] models also have an issue with training large datasets and layers. Data issues aside, some models may manage to achieve accurate data, but the results achieved cannot be interpreted. Modelling capabilities and good interpretability[12] are essential for any prediction model, and most of these available models focus primarily on result accuracy and fail to consider result interpretability. This created a need to conceptualize a model that would solve the issue of accuracy in an energy prediction model without dataset limitations and produce results that were easily interpretable.

To solve this issue with accuracy and good interpretability while not being data constrained in an energy prediction model, the present study proposes applying stacked LSTM snapshot ensemble. LSTM is a DL technique based on RNN, which can make this one seems like just another LSTM model. However, the proposed model would implement LSTM snapshot, which includes the characteristics of RNN algorithms, LSTM algorithms, and the snapshot characteristics. The advantages that the stacked LSTM snapshot would include inputs as connected time series; solving problems with a disappearing and vanishing gradient; and being able to store snapshots of different data slices, depending on the length of their sequence. The model could work with data with different sequence lengths, as well as train them, which is a problem for most prediction models. Implementation of the Fast Fourier Transform (FFT) would make this possible and allow the proposed model to work with seasonal pattern series. Meta-learning would then be applied to collected base model snapshots and a final estimate of the energy prediction determined. This is where accuracy is achieved because the final estimate is a mean of all the collected snapshots for a given data input. In addition, we improve the viability of the data by adding weather data. The fact that they are climate change and unpredictable weather events implies that the weather component is evolving and thus having new implications on energy consumption and production. The proposed model ensures that every predicted consumption instance is checked for errors and accuracy. The model is found to have the best accuracy relative to the compared models and can be used for accurate energy consumption prediction. With additional performance evaluation, it can be used by energy companies for consumption analysis to improve their service delivery.

This paper discusses works related to various DL models, including LSTM and ensemble learning models in Section 2. The proposed model is described in-depth in Section 3, and in Section 4, the data and related visualizations are presented. Section 5 delves into the developed model, data preparation, settings, feature selection, and result evaluation. Section 6 discusses the model training and related losses. Section 7 compares the stacked snapshot LSTM ensemble to other DL models, and Section 8 concludes this discussion of the study.

2 Related Work

One of the main components of AI is DL, which is defined as a set of layered knowledge-acquiring computer algorithms used by computers and machines to learn without the need for explicit programming[17]. Additionally, DL provides AI with layered algorithms that machines can use to automatically learn and improve actions based on previous experiences. While AI is characterized by airing and applying knowledge, DL is the acquisition of skill and knowledge[17]. When using AI, the goal is to increase the probability of success rather than focusing primarily on accuracy. However, when using DL, the goal is mainly to increase the accuracy of an action, regardless of success. AI can be compared to smart computer software, while DL can be likened to the processes and techniques a machine uses with data in learning. As indicated earlier, DL algorithms are specifically designed to help machines learn. Typically, the DL process involves finding relevant data to identify patterns. After identifying a pattern, the machine can predict outcomes for new data using historical data. There are three ways machines learn: supervised learning, reinforcement, and unsupervised learning[17]. Predicting energy demand frequently makes use of neural networks and DL, particularly LSTM and ensemble learning models.

Through the ANN models, researchers have concluded that building efficiency and rising energy demands are crucial to sustainability[18]. Their exploration aims to decide the general patterns while using ANNs to determine the energy consumption of a structure. They concluded that while zeroing in on the feed-forward brain network, they discovered that there are a few holes, principally in application, because ANN is more fit to time series information, but this is seen in only 14% of the cases they cover. They discovered that 6% of the applications are for general regression and radial bias neural networks. It is determined from their findings that energy management, optimization and conservation forecasting strategies are not as suitable for day-to-day operations as the ANN predictive models[18].

Using 30-minute Short-Term Load Forecasting (STLF) resolutions, the researcher compare the performance of several ANN models with numerous hidden layers and activation functions[19]. The models use 1–10 hidden layers and different activation functions, which comprise the parametric rectified linear unit, rectified linear unit, exponential linear units, leaky rectified linear units, and scaled exponential linear unit[19]. Using electrical consumption data from five specific buildings collected over 2 years and two performance metrics—the Coefficient of Variation of the Root Mean Square Error (CVRMSE) and Mean Absolute Percentage Error (MAPE)—they discovered that the model with five hidden layers has an average superior performance relative to other tested models designed for STLF[19]. Although the researchers produced a standard model for predicting energy consumption, it is possible to create a more precise prediction model by including the input variables, which can show a building's energy consumption characteristics. Additionally, the target's forecasting performance can be anticipated to rise due to hyperparameter tuning in the Scaled Exponential Linear Unit (SELU) prediction models[20].

The SVR, LSTM, and predictive model combining SVR and LSTM contain 240 samples with 24-hour load profiles[21]. The goal is to perform short-term microgrid load forecasting. Each hour's load quantity is chosen as the output variable, and the input variables are used as an input sample. The majority of the data (70%) in each network is used for training the model, while the remainder (30%) is used in testing. The short-term load prediction in the microgrid is tested without considering climate data. Instead, it focuses on the application conditions that electricity generators and consumers of the microgrid would encounter at any given time. The researcher's take on the outcomes of the various DL methods is presented in Table 1. Short-term load prediction in the microgrid is more accurate and efficient using the model[21].

The long transient memory LSTM[22] is used to improve the planning capacity of utility companies by improving their ability to predict energy load consumption, which can help in deciding whether a new energy plant, transmission lines, or choosing between different fuel sources during production are needed. The researchers could show that the model is determined to be highly accurate, the MAPE obtained

Table 1 Limitations of the related work.

Model	Measure	Technique employed	Limitation	Source
ANN model	Forecasting building energy use and demand	ANN	• Does not perform well outside the model's training range, • Overfitting, • Inadequate selection of hyperparameters, • Internals of the models are not known.	Runge and Zmeureanu[18]
ANN-based STFL model	Forecasting electrical energy consumption of a building or building clusters	ANN and STLF	Different ANN models are constructed, and the input variables of the data used do not reflect the characteristics of the target building (missing characteristics); therefore, the forecast is not the most accurate.	Moon et al.[19]
SVR-LSTM hybrid model	Forecasting consumption load in microgrid	SVR and LSTF	Microgrid is examined without the presence of Renewable Energy Sources (RES) in the data, and it only includes loads for households and commercial consumption.	Moradzadeh et al.[21]
LSTM-RNN model	Forecasting electricity load demand	LSTM and RNN	There is room for performance improvement by incorporating weather parameters into the training data.	Agrawal et al.[22]
LSTM novel model	Predict time series with periodicity	LSTM	Missing measurement equipment.	Wang et al.[23]
Ensemble model	Forecasting big data time series	Decision tree, gradient-boosted trees, and random forest	–	Galicia et al.[24]
Deep ensemble learning	Probabilistic load forecasting in smart grids	LASSO-based quantile combination strategy and end-to-end ensemble	Difficulty in obtaining a narrow Predicting Interval (PI), which interferes with the model accuracy.	Yang et al.[25]
CNN-DNN model	Mapping landslide susceptibility	CNN for feature extraction and DL neural network for classification by sorting pixels and grouping them into high-susceptibility and low-susceptibility groups	The model only uses the conditions at a point to test for susceptibility to landslides. To achieve more accuracy, a model is recommended that uses a conjunction of several conditions over an area. The researchers leave this as a point for further research.	Azarafza et al.[26]
RNN algorithm model and CNN algorithm model	National-scale landslide susceptibility	Two novel DL algorithms, RNN and CNN	The study lacks some effective data, including soil depth, soil texture, and distance from water table. These factors can help enhance the predictive power of the algorithms.	Ngo et al.[27]
Hybrid model using GeoDetector and ML cluster model	Landslide susceptibility mapping	GeoDetector and ML cluster (ANNs, Bayesian network, logistic regression, and SVM)	The model selects SVM as the most effective output ML, which works well with binary data, but no mention of other presentations of data.	Xie et al.[28]
Spatially explicit DNN model	Landslide susceptibility	DNN	Exclusive for spatial parameters.	Achu et al.[29]

is 6.54 within a confidence interval of 2.25%. Model training takes 30 minutes. For a 5-year forecast, annual offline training is required, making the computational time a benefit. The LSTM–RNN model is suitable for predicting future locational marginal electricity processes[22].

LSTM techniques are used in an attempt to provide credible advice for energy resource allocation, energy saving, and improving power systems[23]. Over five months, experimental data were collected at a minute resolution between March 2018 and July 2018. The experiment demonstrates that time as a variable accurately reflects the periodicity. It is found that LSTM shows better performance than forecast methods, such as Back Propagation Neural Network (BPNN), AutoRegressive Moving Average model (ARMA), and AutoRegressive Fractional Integrated Moving Average model (ARFIMA). For long-term time series predictions, the LSTM's Root Mean Square Error (RMSE) is 19.7% lower than the BPNN value, 54.85% lower than the ARMA value, and 69.59% lower than ARFIMA[23], and shows excellent energy forecasting potential[23].

Ensemble learning[24] is used to determine multistep forecasting for time series data. Three techniques—gradient-boosted trees, decision trees algorithm, and weighted least squares, compute the weight of the ensembles. Through this, it is possible to produce the dynamic or static ensemble model using a two-weight updating strategy technique. The prediction problem is then decomposed into prediction subproblems, where each subproblem value is used in the forecasting horizon to obtain the ensemble member predictions. The researchers determined that their approach is scalable because DL algorithms based on Apache Spark, a big data engine, can solve the subproblems. The data fed to the ensemble models is 10-year electrical data measured at 10-minute intervals. The researchers showed that the static and dynamic ensembles perform better than the individual members. The dynamic model Mean Relative Error (MRE) value is 2%, the highest accuracy level obtained. It is also a promising result for forecasting large time series[24].

The deep ensemble learning study on smart electric grids is based on probabilistic load prediction. It is postulated that accurate load predictions are important in decisions involving benefits and costs for electrical grids[25]. The Least Absolute Shrinkage and Selection Operator (LASSO) model evaluates energy consumption data from 400 small and medium businesses, and 800 consumers. The individual residential data consumption features show higher volatility and diversity than the small and medium businesses data, not withstanding the seasonality and regularity of the aggregated load profiles. When conducting the probabilistic load forecasting on the 800 consumers, the data are classified using one hour and one day intervals. The DNNs used in the ensemble models are randomly chosen, with a total of 7 DNNs with 512 hidden layer nodes, and the randomized numbers between 1 and 4 in the hidden layers. The ensemble forecasts are refined using the LASSO-based quantile combination model.

3 Proposed Model

3.1 Dataset

The electrical energy consumption prediction models are validated with the help of Hebrail and Berard’s IHEPC dataset from the UCI Machine Learning Repository. Between December 2006 and November 2010, 2 075 259 measurements from households were included in the dataset. About 1.25% of the rows have missing measurement values[30], but aside from that, the dataset contains the calendar timestamps. The 12 attributes are date, time, global active power, global reactive power, voltage, global intensity, and sub-metering 1 through 3, which are illustrated in Fig. 1[30]. To select the data to use for the model training, different sequence lengths are identified. This process cannot be handled randomly because random sampling would eliminate the possibility of catching the seasonality in the data. The study therefore uses FFT to extract the right sequence lengths from any given time series. Applying FFT ensures that the sample sequence lengths capture the different seasonalities, patterns, and other time-dependent effects in the entire time series.

3.2 Applying the proposed stacked LSTM snapshot ensemble

Improving energy management services necessitates accurate energy consumption predictions in residential and commercial buildings. However, it is challenging to make accurate predictions about energy consumption due to the unpredictability of noisy data[31]. Complex variables cannot be correlated or evolved using conventional prediction methods. The two-layer ensemble is fed with energy consumption data from the IHEPC along with weather data, allowing for multiple sequence lengths in the proposed model, which addresses these issues based on the photographs. After that, the model is trained, and a base estimate is made. The base estimate has a lot of output, and although the patterns are similar, the different models learn differently. The meta-learner makes it possible to select the appropriate sequences from weighted snapshots, effectively preventing random distribution[32].

The stacked ensemble LSTM DL algorithm, an advanced RNN that takes the place of the original cell neurons, is the tool used for regression. The RNN algorithm’s unique characteristics are passed down to the DL algorithm, allowing the inputs to be considered connected time series. Also, the LSTM cells’ intricate structure can solve problems with disappearing and vanishing gradient limitations[33]. Input, cell status, forget, and output gates are the four essential

Index	Variable	Description
1	Day	A value from 1 to 31
2	Month	A value fron 1 to 12
3	Year	A value from 2006 to 2010
4	Hour	A value from 0 to 23
5	Minute	A vaue from 1 to 60
6	Global active power	Household global minute-averaged active power (In kilowatt)
7	Global reactive power	Household global minute-averaged reactive power (In kilowatt)
8	Voltage	Minute-averaged voltage (in volt)
9	Global intensity	Household global minute-averaged current intensity (in ampere)
10	Sub-metering 1	An oven and a microwave, hot plates being not electric but gas powered (in watt-hour of active energy)
11	Sub-metering 2	This variable corresponds to the laundry room, containing a washing machine, a tumble-drier, a refrigerator, and a light (In watt-hour of active energy)
12	Sub-metering 3	This variable corresponds to an electric water heater and an air conditioner (in watt-hour of active energy)
13	Temperature	The measured temperature in degrees Celsius or Fahrenheit
14	Humidity	Percentage of water vapor in the air
15	Wind speed	Speed of wind, typically measured in kilometers per hour or miles per hour
16	Wind direction	Direction from which the wind is blowing, typically measured in degrees
17	Precipitation	Amount of rain, snow, or other forms of precipitation that fell during a specific time period

Fig. 1 IHEPC dataset features[30].

components of the utilized LSTM algorithm. The forget, input, and output gates are used to, respectively, keep, update, and delete the data in the cell status[34].

The forget gate is responsible for deciding which data should be deleted or kept from the previous step, due to the presence of the sigmoid layer. The data that needed to be saved in the new cell state are then identified. A sigmoid function is used in the input gate layer to determine the update values. New vector values are produced by the tanh layer and injected into the state. The states are then merged to produce a new status update, and the LSTM memory is the cell's state. In this case, the algorithm performs better than standard RNN when processing longer input sequences. Past cell states are coupled to the forget gate in each time step to determine the broadcastable data. The values are later combined in the input gate to create new cell memory. Finally, the LSTM cells produce and distribute energy. The cell state passes beyond the tanh hyperbolic function, filtering the value of the cell state between −1 and 1, as shown in Fig. 2[34, 35].

Where

• i_t represents the input gate,

• $\widetilde{C_t}$ represents the memory gate,

• C_t represents the cell state,

• f_t represents the forget gate,

• o_t represents the output gate,

• h_t represents the hidden state.

And:

• σ is the sigmoid function,

• U_i, W_i, U_c, W_c, U_o, W_o, and W_o are the weight matrices, b_i, b_c, b_f, and b_0 are the biases,

• x_t is the input data.

LSTM has a hidden state, which is the short-term memory. It also has a cell state that is the long-term memory. In Fig. 2[36], they are shown as h_{t-1}, representing the hidden state of the previous timestamp, h_t is the current timestamp hidden state, C_{t-1} is the previous cell state, and C_t is the current cell state. x_t is the input signal into the cell. The equation to solve it represents the input gate. It receives previous predictions and new information as input. It holds the information and manipulates it, hence updating it, with

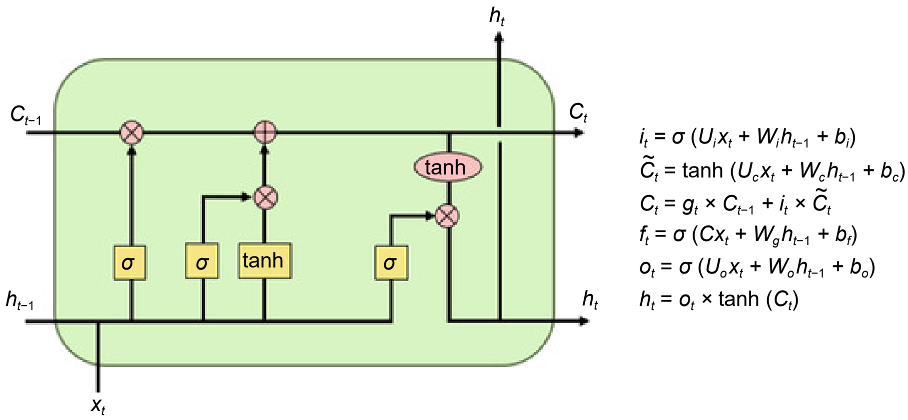


Fig. 2 LSTM structure[36].

the assistance of the sigmoid function. Equation f_t represents the forget gate, which is responsible for removing information no longer needed by a cell. The equation that solves for o_t is the output gate and is responsible for establishing the results of the cell. One LSTM layer can have multiple timestamps. So, a timestamp receives data from a previous timestamp and new information. The new information goes directly to the input gate, while the previous timestamp information passes through the forget gate to select only the needed information for the cell passed to the input gate. The input gate computes the new information, the selected information from the previous gate, and uses the sigmoid function. The results are passed to the output gate, which decides what to output.

Because of their design, LSTMs can only handle sequences of equal length for each epoch. This is following the optimization process's requirements for the matrix operations. In some data, however, a sequence of varying lengths cannot be avoided, so padding is used. This makes it possible to train models with different sequence lengths. However, the patterns they learn are related but different. An FFT must be used to select the data's sequence lengths to accomplish this[36, 37]. The FFT makes it possible to select sequences that distinguish between distinct periods in a given time series. When using FFT to select energy consumption sequences from time series data, it will be possible to capture the series' seasonal patterns and other time-dependent effects, making the chosen sequences work better[36, 37].

The proposed model will feed the LSTM through various data slices, and diversity will rise. As a result, there will be n snapshots stored for any given LSTM, provided that a set $S = \{s_1, s_2, \ldots, s_n\}$ contains sequences of various lengths. The process is repeated with a different data slice following the first data slice's training of the LSTM, and the cycle continues. Snapshots of the various sequences are saved for each data slice. The mean is derived as the base forecast, and the collected snapshot estimates are combined from this point. In this case, meta-learning is used to acquire the mean function. The weight matrices from the first snapshot are used as the second sequence length for the current data slice because there are two LSTMs. Meta-learning is used to combine all of the base model snapshots, resulting in the identification of the final estimate forecast. If 20 sequence lengths are used, for instance, 20 snapshots are stored for the first LSTM and used as the second LSTM's sequence length. The base estimate is compiled from the snapshots obtained from the second LSTM and given to the meta-learner for final estimation[10, 36, 37]. Figure 3 shows the stacked LSTM snapshot ensemble used in the present study.

Figure 3 is a diagram of the proposed model, which received data from the UCI Machine Learning Repository. It is taken as data slices and passed through the FFT, which selects the best sample for the model, that is sequences of different lengths, which makes it to include data with different time-dependent effects including seasonalities and patterns. The varying length sequences length are then recorded as the input data ready to pass through the LSTM layers. The output of the first layer is used as the input of the second layer. The output of the second layer is then passed through the LSTM snapshots. The number of different length sequences that are input determines the number of snapshots taken. So, if the input data has 10 varying

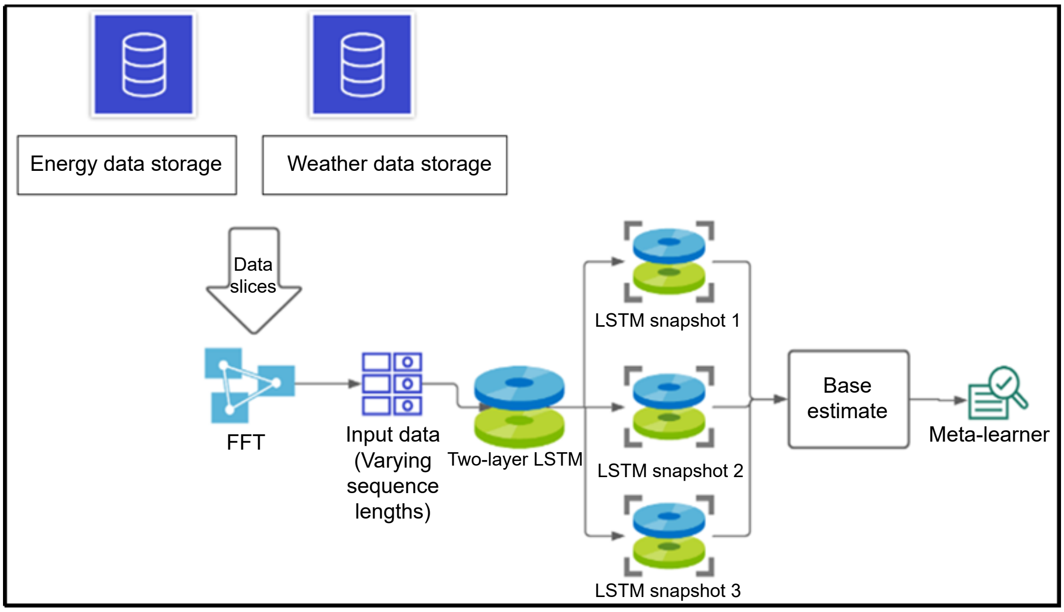


Fig. 3 Stacked LSTM snapshot ensemble.

sequence lengths, there will be 10 snapshots passed to the base estimate. The base estimate stores the snapshots gotten and passes them to the meta-learner, which combines the snapshots to get a final forecast estimate for the model.

4 Data Visualization and Analysis

4.1 Data sampling and resampling

The data contain variables sampled relative to time, month, and date. Resampling is also conducted because it shows much interaction of the data since the periodicity of the systems is changed. Since processing the data is expensive, and even more so with larger datasets, resampling is allowed for better decision-making in a suitable time frame. Figures 4–8 show the data resampled every 30 minutes, hour, day, week, and month, all of which are selected to obtain the best insights.

During resampling, the month, date, and time metrics are important because they show great interaction, as desired. Figures 4–8 show resampling based on 30 minutes, 1-hour, daily, weekly, and monthly resolutions. Sampling is important for obtaining effective features used in the algorithm's performance. Based on Figs. 4–8, sampling the data based on hourly resolutions gives a balanced output of the predictions, which allows the data to be transformed into a dataset with 34 598 observations, from which seasonality could be derived.

4.2 Seasonality

Seasonality refers to the periodic trends in data. Seasonality suggests predictable patterns that repeat at known frequencies within a specified period, such as hourly, weekly, or monthly. From the energy consumption data, the variables evaluated are global active power, global reactive power, and global intensity power. It is observed that the global active power (Fig. 9) and global intensity power (Fig. 10) show similar highs in December and January, and then show low consumption in August. The global reactive power (Fig. 11) shows an opposite trend from the other two variables. Low consumption is detected in the December and January periods, while highs are detected in August.

The trends can be explained by increased consumption in the winter months and reduced consumption in the hotter months of the year. From the trends observed, it can be deduced that energy consumption is affected by weather conditions. Typically, the demand for energy and its consumption is higher during winter than it is in summer, and the peak demand during summer is lower than the peak demand during winter. Similarly, low demand during the summer can be compared to low demand during winter. Electricity demand and consumption tend to vary daily based on human activity.

Energy consumption is typically lower at night, due to reduced domestic consumption. It is reasonable to

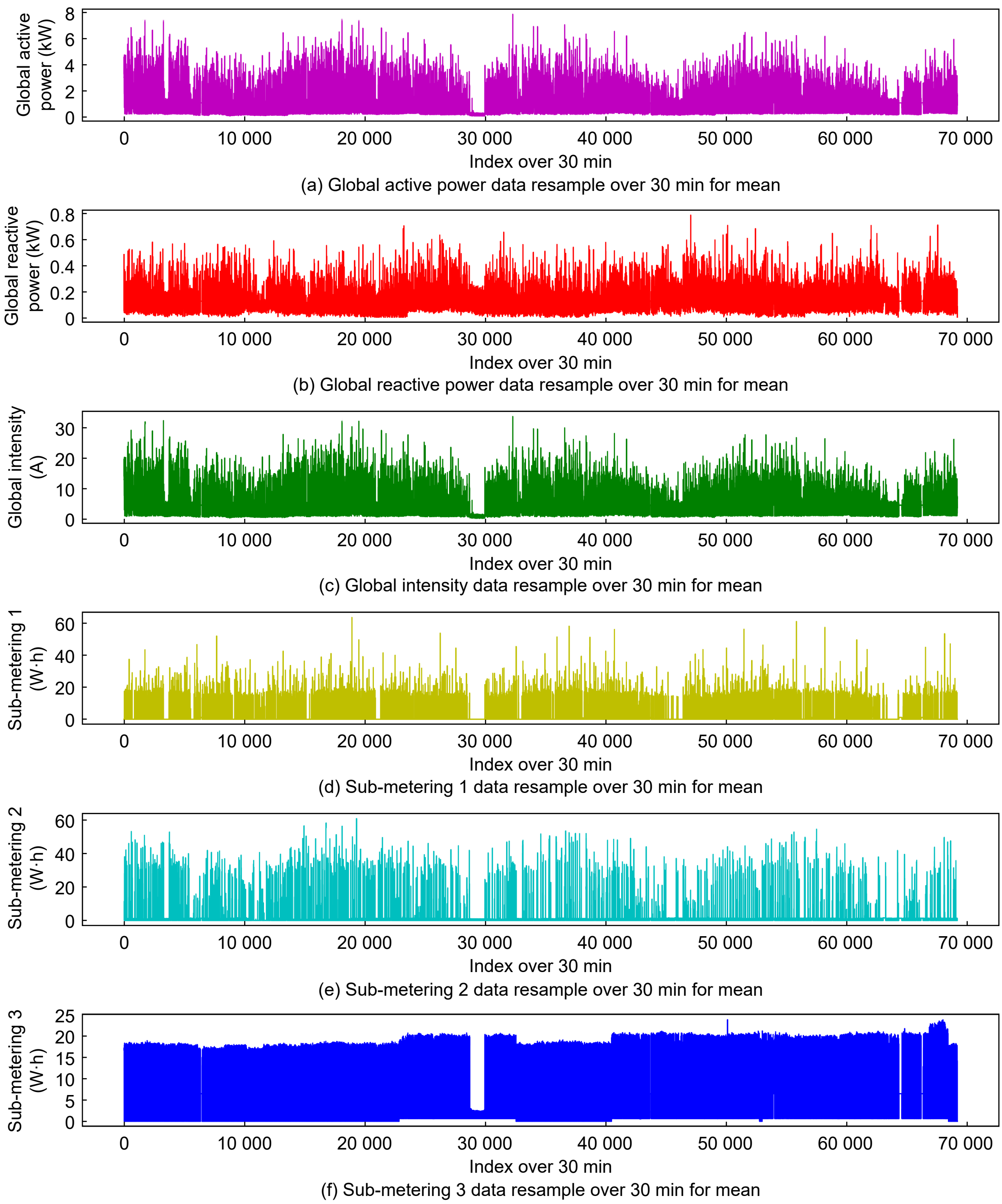


Fig. 4 Data resampling every 30 minutes.

state that there are demands in energy consumption in the morning hours when people wake up and start using electrical appliances, such as showers, toasters, coffee makers, and kettles. However, the surge in energy consumption increases faster over shorter durations during winter. The increase in energy consumption increases and starts stabilizing at a certain time of the day as people leave their homes. It is during these periods that peaks are noted in the sub-metering data. During winter, increased demand is denoted starting at 15:00.

This trend can be attributed to children returning from school and adults returning from work. As people return home, electrical appliances, such as televisions,

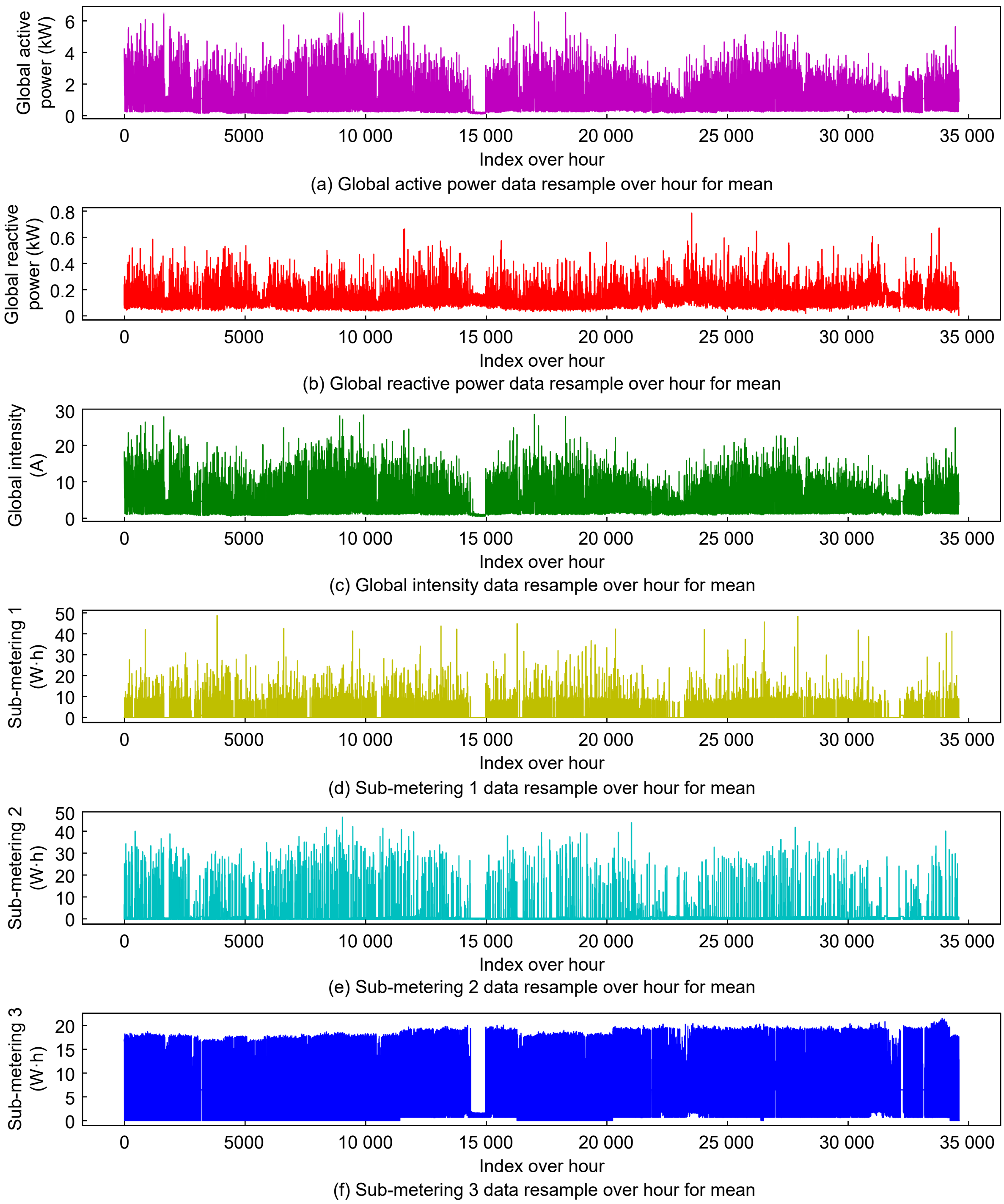


Fig. 5 Data resampling per hour.

dishwashers, microwaves, and air conditioners, are turned on as people warm and light their houses and prepare dinner. Consumption falls and drops as people start going to bed. During summer, the surge is not as evident as in winter because when people return home, it is still light outside, and their houses are warmer. There is an increased use of refrigerators and colder beverages when the weather is warmer, but energy consumption in the evenings is lower in the summer. There are notable peaks in Sub-meterings 2 and 3, such as the air conditioners, refrigerators, and washing machines, are used more often, increasing consumption.

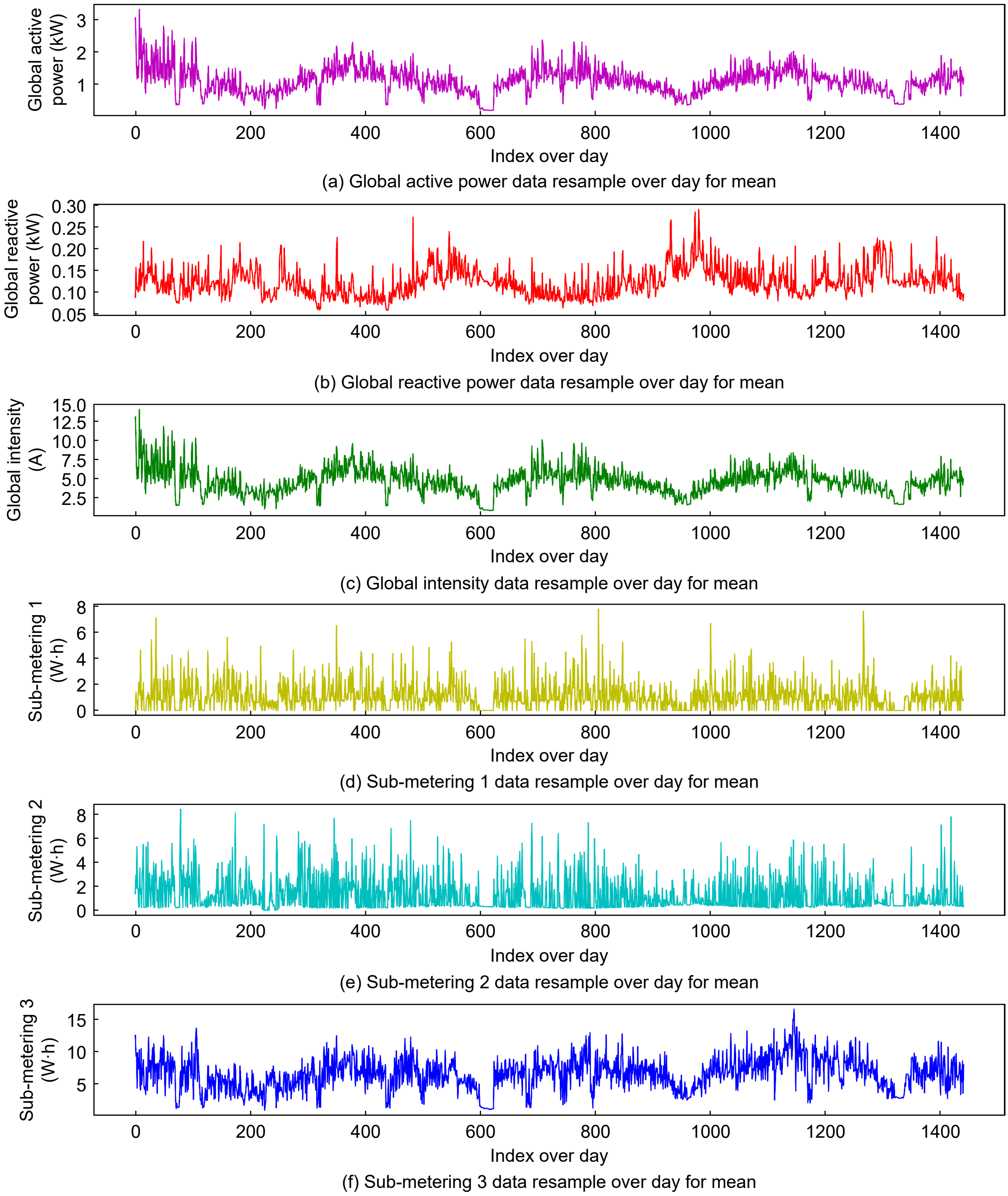
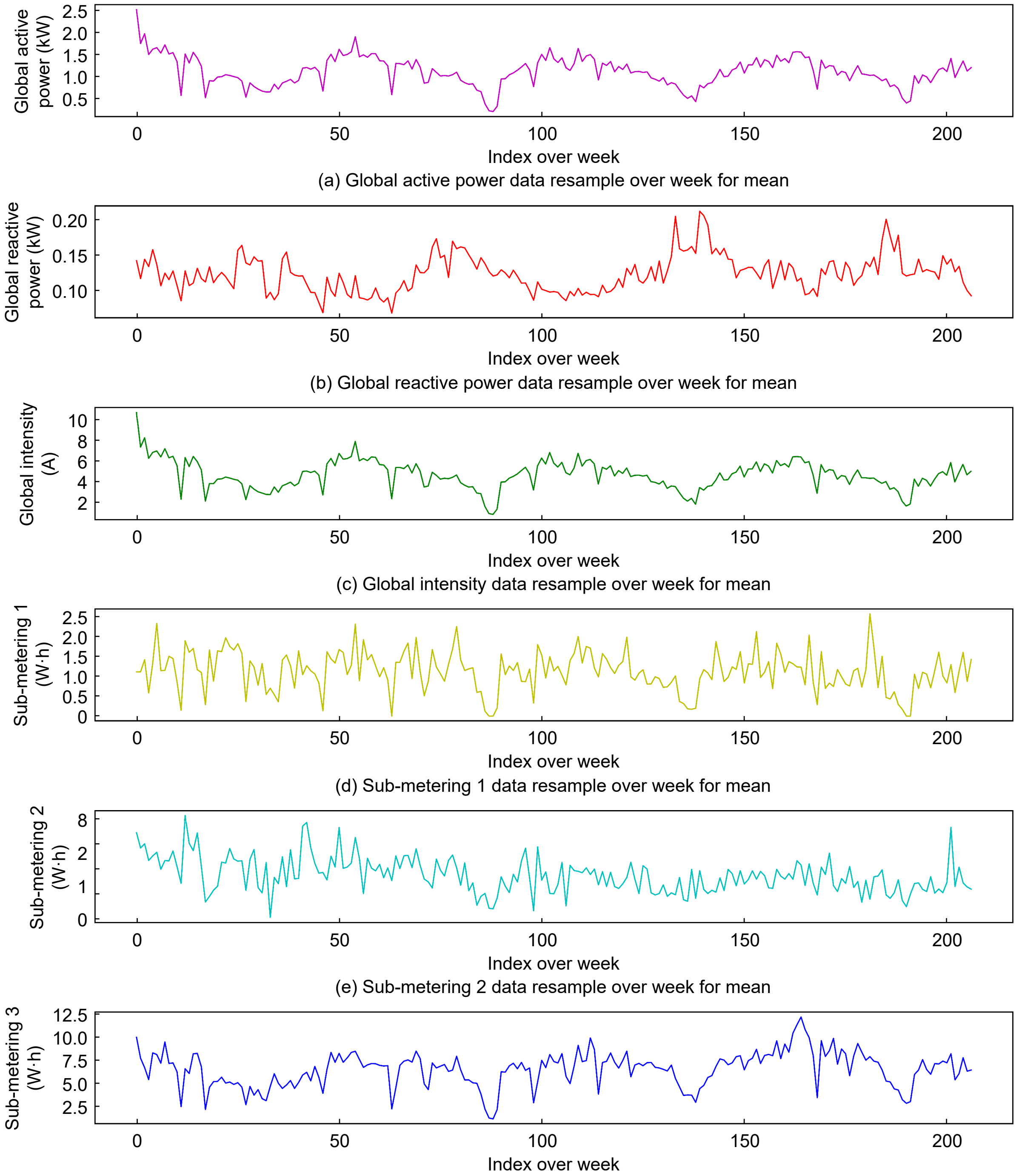


Fig. 6 Resampling per day.

4.3 Rolling average

The rolling average is used to determine a trend's direction. It adds up the data points of the energy consumption over the defined period and divides the total by the data points provided to determine the average. From the rolling curves obtained, it is possible to observe trends in the data where the global active power and global intensity show highs in the months starting in March 2008 that peak in July 2008, after which there is a decreasing trend. This trend can be interpreted as increased energy consumption in the months leading up to July 2008, and there is reduced consumption after July 2008, as observed in the data. The global reactive rolling mean shown in Fig. 12



(a) Global active power data resample over week for mean

(b) Global reactive power data resample over week for mean

(c) Global intensity data resample over week for mean

(d) Sub-metering 1 data resample over week for mean

(e) Sub-metering 2 data resample over week for mean

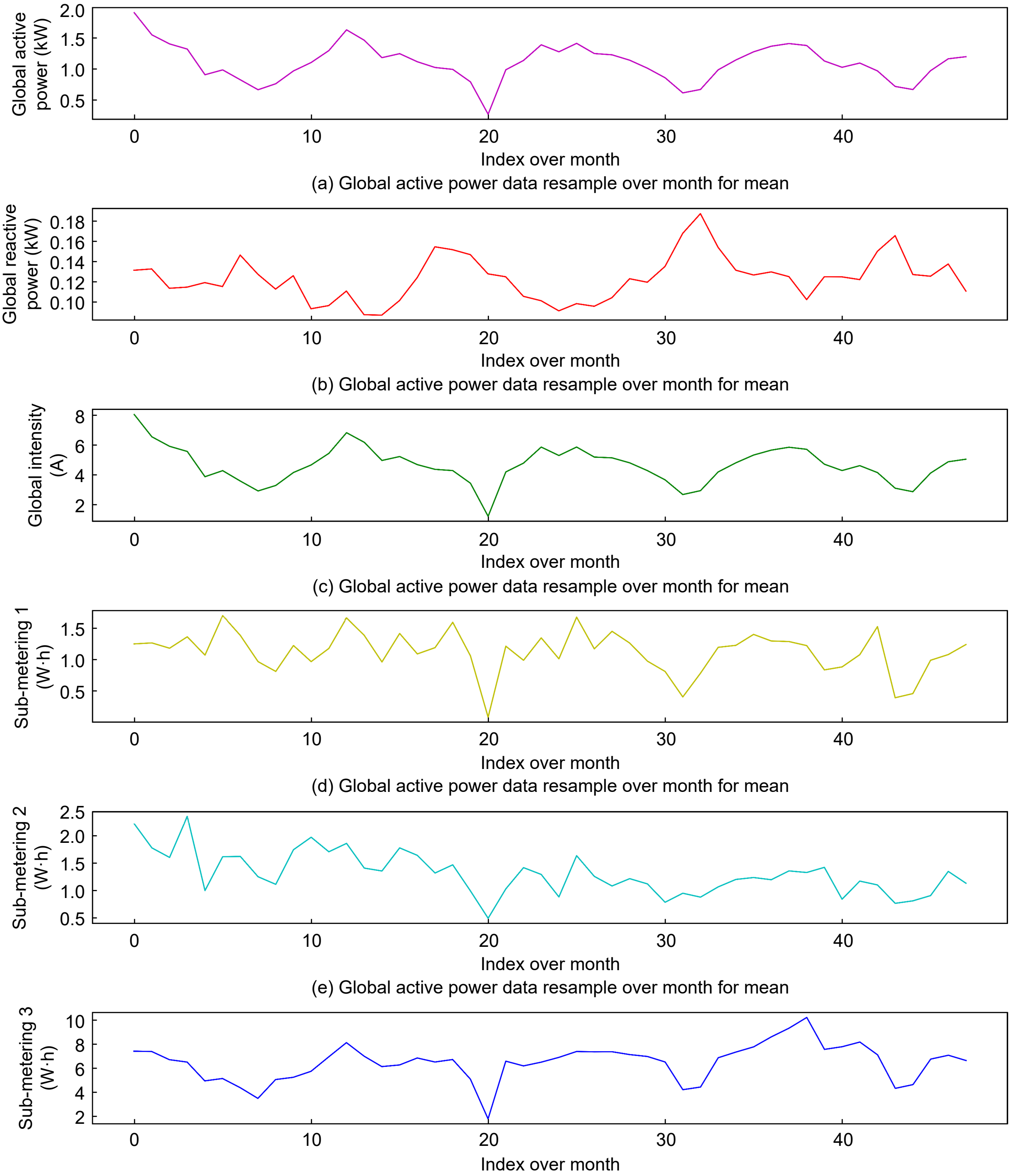
(f) Sub-metering 3 data resample over week for mean

Fig. 7 Data resampling per week.

indicates that the trend is opposite to what is observed and shown in Figs. 13 and 14.

4.4 Autocorrection

It should be noted that after two lags seen in Figs. 15–17, the lines get inside the confidence interval (light blue area). The lag is caused by the 12–13 months used in defining a season in the data. Once the algorithm detects the extended season, there is an autocorrection in the lag, where the lines fall within the confidence level. The benefit of having the extended 12–13 months and resultant lags is to show the ability of the algorithm to adapt to unconventional data and still produce accurate results.



(a) Global active power data resample over month for mean

(b) Global active power data resample over month for mean

(c) Global active power data resample over month for mean

(d) Global active power data resample over month for mean

(e) Global active power data resample over month for mean

(f) Global active power data resample over month for mean

Fig. 8 Data resampling per month.

5 Experimental Evaluation

5.1 LSTM settings

The stacked LSTM snapshot ensemble is implemented using Python, and the front and back ends are developed using Keras and TensorFlow, respectively. As a high-level DL API, Keras is used to develop and train DL models with TensorFlow as the backend. The IHEPC data are cleaned, and noisy data are removed and integrated into the model before training, testing, and validation.

During training and testing, a MacBook Pro M2 clocked at 3.2 GHz with 32 Gb of RAM is used. Additionally, the M2 chip has a 10-core GPU, an 8-

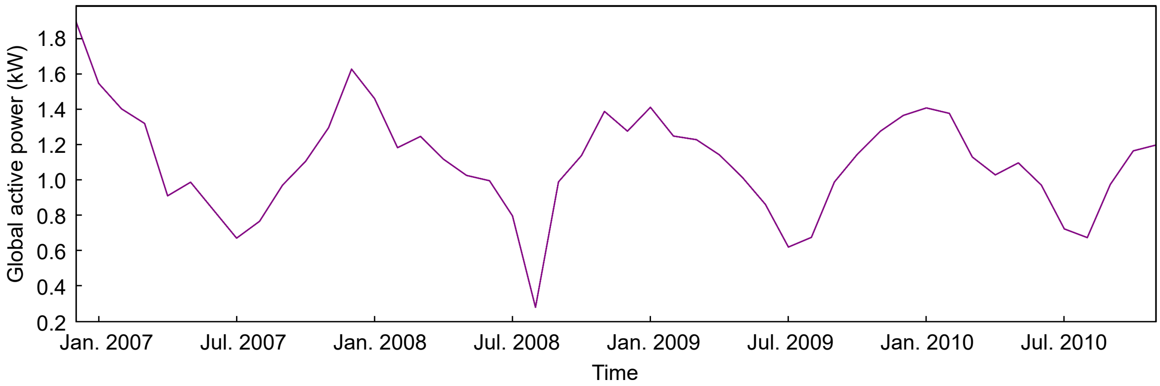


Fig. 9 Global active power.

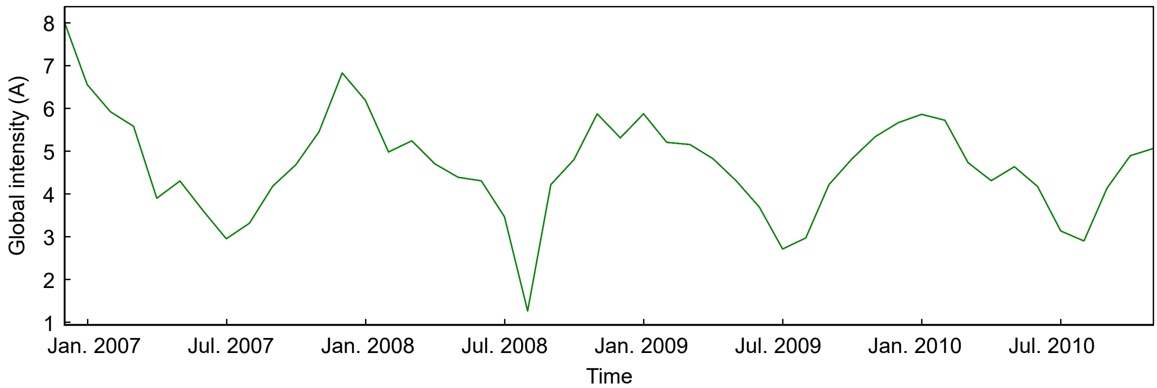


Fig. 10 Global intensity power.

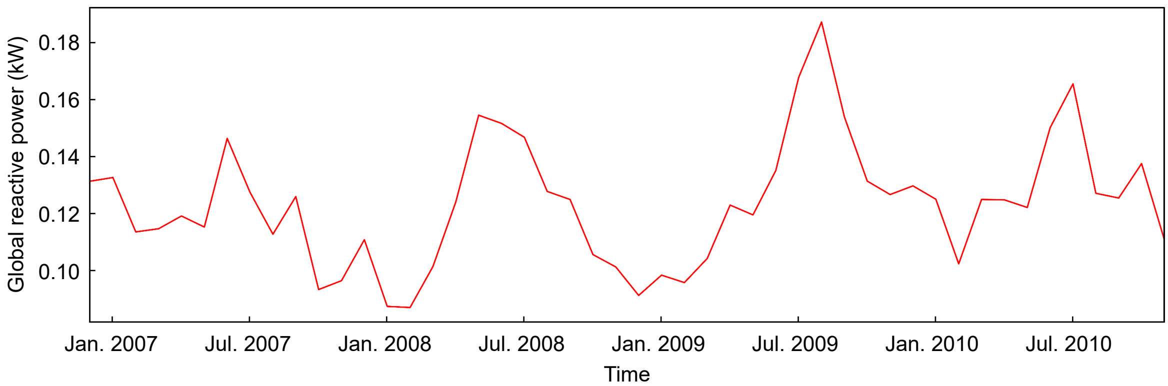


Fig. 11 Global reactive power.

core CPU with four efficiency and four performance cores, and a 16-core neural engine. From the dataset, 70% of the data are used in training, and 30% are used for testing and validation. The technique used for model validation is a train/test split. Validation occurs between the training and test stages. Based on the new dataset obtained, the LSTM settings used are normalization set to between 0 and 1, the batch size is 50, the epoch number is 100, and there are 4 LSTM layers used.

5.2 Data preparation

The dataset contained 2 075 259 rows and 7 columns. The data and date fields are parsed to the date/time column and converted to the index column during importation. The outliers in the data, noisy data, are cleaned by filling the null values and noise with the mean values in their respective fields. The data are successfully integrated after cleaning. From the dataset containing 2 075 259 rows, 25 979 rows contain null values, which are filled with the mean value. This is

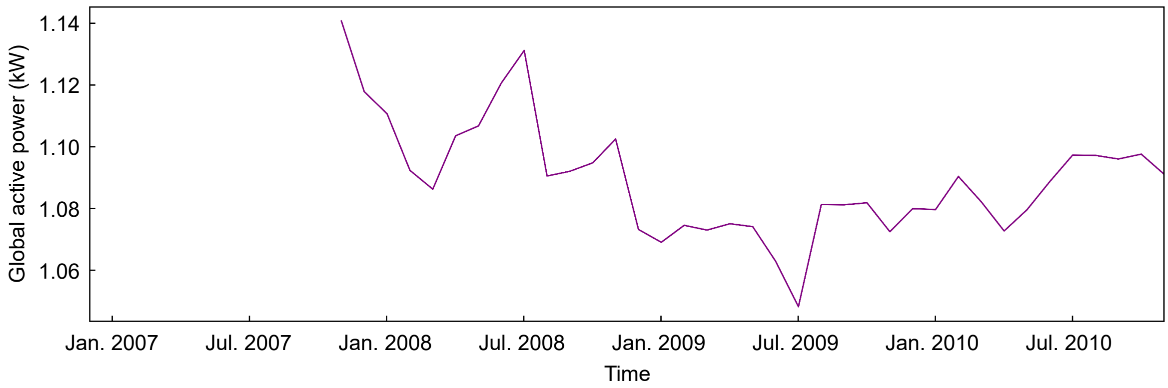


Fig. 12 Rolling mean for global active power over 12 month period.

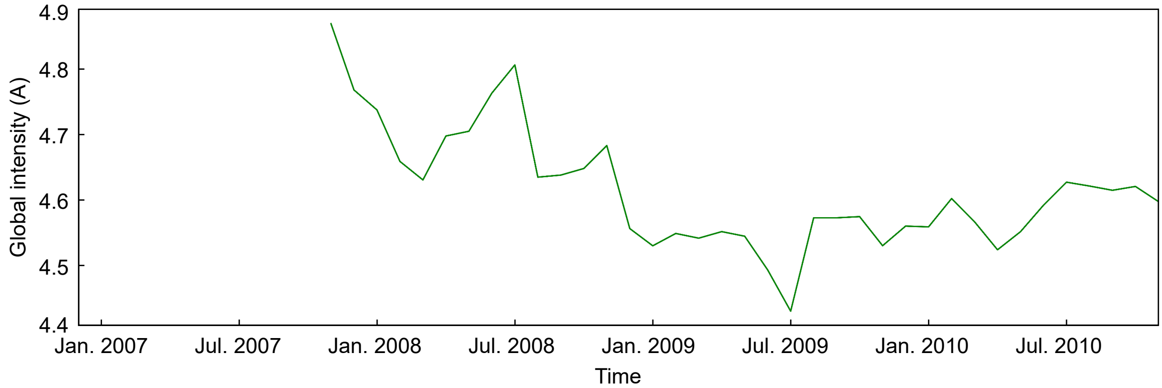


Fig. 13 Rolling mean for global intensity over 12 month period.

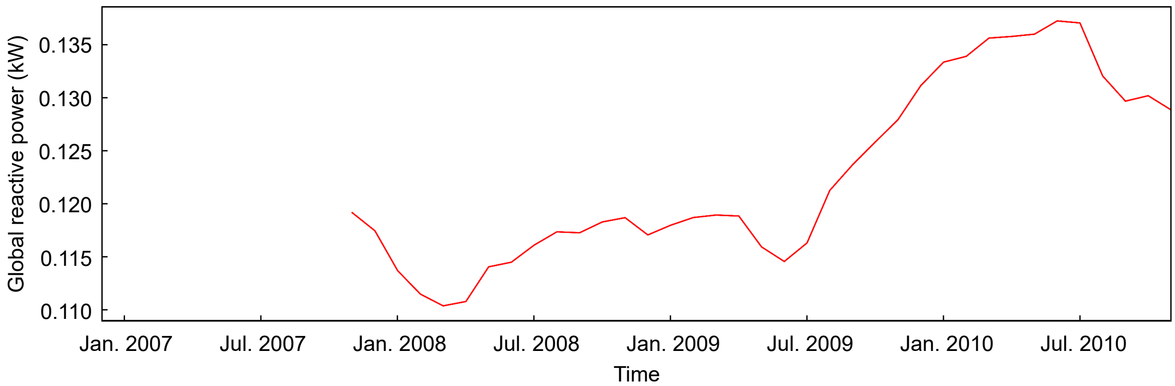


Fig. 14 Rolling mean for global reactive power over 12 month period.

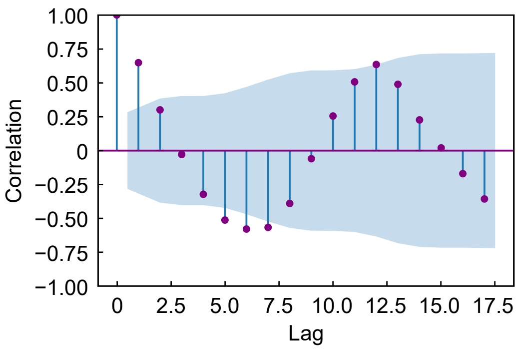


Fig. 15 Autocorrection of global active power.

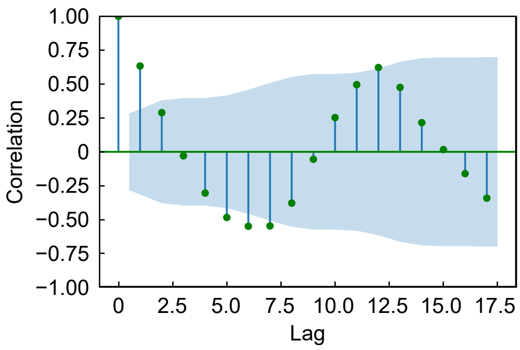


Fig. 16 Autocorrelation of global intensity.

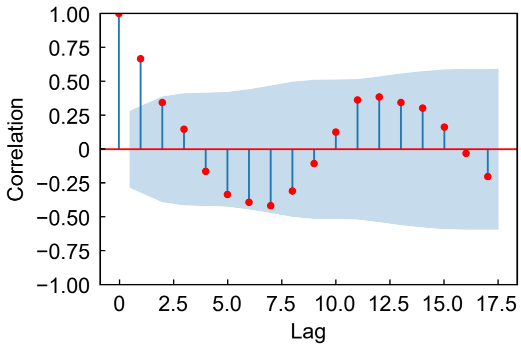


Fig. 17 Autocorrection of global reactive power.

done while cleaning the data to deal with biasing the results of the DL model used, and minimizing the model's inaccuracy.

5.3 Data transformation

The Dickey-Fuller test for the presence of a unit root is used in the analysis to test our time series dataset, A p-value is obtained from the test, which is used to make inferences about the dataset. The null hypothesis assumes the presence of a unit root; therefore, the p-value should be less than 0.05 for the null hypothesis to be true. The p-value obtained from the data here is 0, which implies a unit root is existed and leads to make transformation on the data to become stationary by taking lag differences of the series.

5.4 Normalization

The goal of normalization is that since the variables in the datasets are measured on different scales, that do not all contribute equally to the model learning and fitting and could cause bias. Therefore, to handle the probable problem, normalization of the data is completed using the MinMaxScalar function. The function normalizes the data used in the model between a minimum value of 0 and a maximum value of 1. This is done so that all feature values are within the 0–1 range.

5.5 Feature selection

The Pearson Correlation Coefficient (PCC) is used to select the most relevant features from the dataset, with the value ranging from −1 to 1. The technique applies covariance and two other factors to determine the strength of relationship between the features and how strongly the features correlate. After applying PCC, it is found that the voltage feature has a negative correlation. Figure 18 shows correlations of the features in the dataset.

5.6 Evaluation metrics and results

The performance of the models is evaluated using the metrics of Mean Absolute Error (MAE), coefficient of determination (R^2), RMSE, and MAPE as follows:

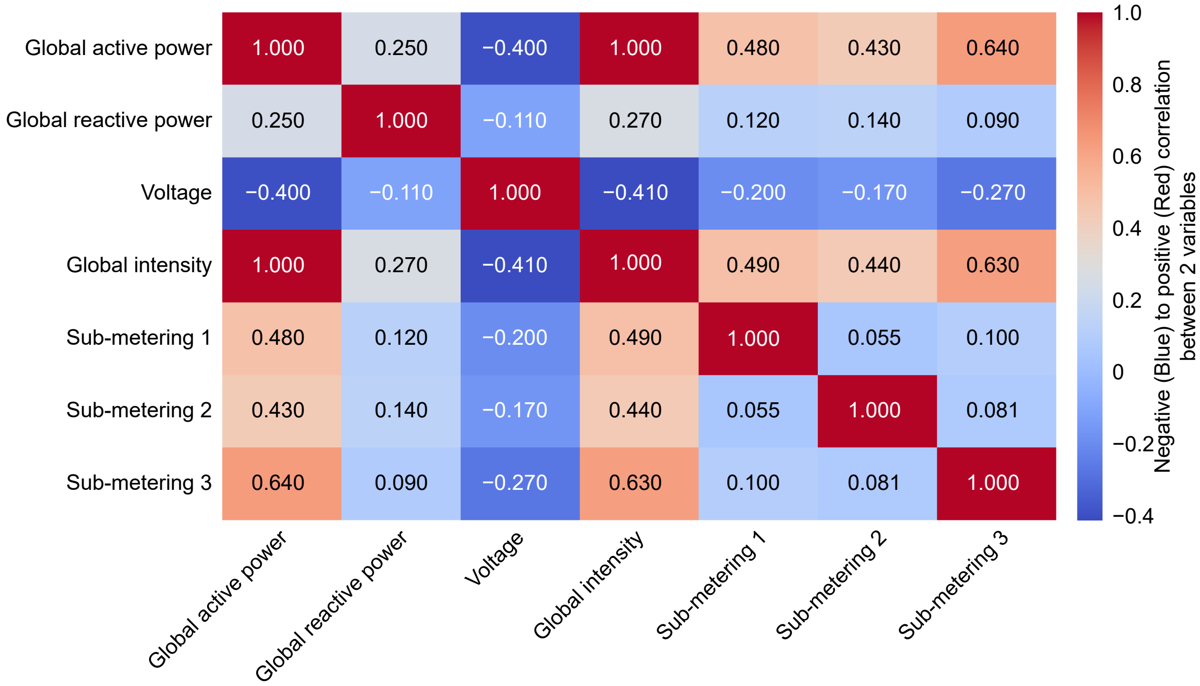


Fig. 18 Feature selection-based PCC.

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \widehat{y_i})^2} \tag{1}$$

$$\text{MAE} = \frac{\sum_{i=1}^{n}|y_i - \widehat{y_i}|}{n} \tag{2}$$

$$\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\frac{|y_i - \widehat{y_i}|}{|y_i|} \times 100 \tag{3}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \widehat{y_i})^2}{\sum_{i=1}^{n}(y_i - \overline{y})^2} \tag{4}$$

where n is the sample size, y_i is the true value, $\widehat{y_i}$ is the predicted value, and $\overline{y}$ is the mean of the sample. The prediction error standard deviation, or RMSE, is a measure of how far apart the data points are from the regression line. MAE is the average of the absolute differences between the actual and predicted values for every instance in the test dataset, considering that all individual differences have the same weight. It is used to quantify the average magnitude of the prediction errors and ignores the direction. MAPE is a relative metric that represents the actual data percentage as the average value of the relative error. MAPE is used to evaluate the model's accuracy through the ratio reflection of the actual value to absolute error values of all samples. When the index is closest to 0 is when it is most accurate. The R^2 score is used to assess the performance of the linear regression model leading to variations in the output dependent variables that are predictable from the input independent variables. At the same time, the coefficient of determination (R^2) shows how well the model fits in the prediction, with values close to 1, indicating the best prediction performance. RMSE is best to use in describing the degree of deviation between the true value and the predicted result. Lower RMSE values indicate a more stable model.

After testing, the RMSE, MAE, MAPE, and R^2 are found to be 0.020, 0.013, 0.017, and 0.999, respectively. Lower values of MAE, RMSE, and MAPE indicate the good accuracy of the developed model. Also, the closer the coefficient of determination (R^2) is to 1, the better the model's performance. For this model, the coefficient of determination (R^2) is 0.999. From the values of the evaluation metrics, it can be deduced that the model fits the datasets and produces highly accurate results. Accuracy contributes to the trustworthiness of the model. Trust attests to the predictivity of the model. The RMSE value, at close to 0, is on the lower side, which means it is a more stable model. In this case, we can deduce that the model is highly stable. The predictivity and stability of the model will contribute immensely to its interpretability, which is how people understand the workings of the model. However, interpretability is only needed when the model gives results that are out of the estimated result bounds. It references the need to understand why the results are out of the expected range.

6 Model Training and Loss

The train/test validation model trains, validates, and tests the splits of the data, with the percentage used for training, validating, and testing being 70%, 10%, and 20%, respectively. The first step is training the data with the training set. Next is usually the validation process, in which the results from the training process are validated, and the hyperparameters are tuned with the validation set. The results are assessed using RMSE, MAPE, MAE, and R^2. The purpose of this is to achieve satisfactory performance metrics. Once this stage is completed, the process moves on to testing the data.

Because this is a supervised learning model in which the DL algorithm looks at many examples to find a good model that reduces loss, model training requires finding the best values for each weight and bias. Losses are penalties for poor predictions, where loss is the number indicating how bad the data are upon which the prediction is made by the developed model. Losses also indicate insufficient data preprocessing practices. When the model's prediction is perfect, the loss is zero, and the objective of training is to select bias and weight sets with low average data loss. The loss values are evaluated using RMSE.

In Fig. 19, we can see that the model performed optimally, with a value close to zero. From the training and test graphs obtained, convergence has occurred. There is no overfitting or underfitting of the data, based on these test and train graphs, and the learning rate is based on a decaying function. Figure 19 thus implies that validation is successful.

If validation is not effective, underfitting or overfitting can be identified from the loss graph. Underfitting implies that the developed model cannot

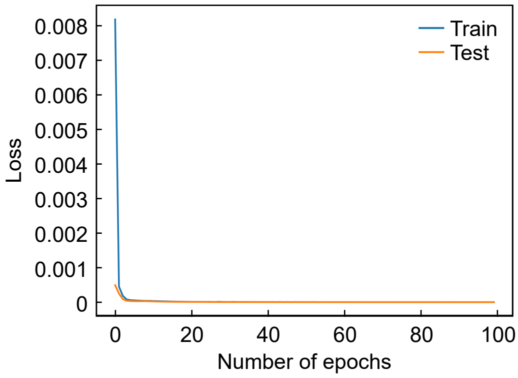


Fig. 19 Model loss.

learn from the training dataset. Underfit models are identified in the training loss learning curve. A typical underfit learning curve has noisy or flat values of comparatively high losses, indicating the model could not completely learn the training dataset. An overfitting curve describes a model that learns too well from the training dataset and includes the random fluctuations or statistical noise found in that data. The issue with overfitting is that when the model is specialized to use training data, it loses its capacity to generalize new data, which results in an increased chance of encountering generalization errors. Overfitting usually occurs when a model contains more capacity than is needed for the problem, which translates into too much flexibility.

This can also happen when training goes on for too long. Overfitting plots exhibit a continuing loss as the experience (epoch) decreases. A good fit curve, like Fig. 19, is the goal of the DL model and can be found between the underfit and overfit models. The good fit curve can be recognized by the test and training loss that reduce to the point of stability and with a minimal gap between the final loss values. The model's loss is almost always lower on the training data than on the test data. The implication is that there should be a gap (generalization gap) between the training and test loss learning graphs. As shown in Fig. 19, the model's training loss curve reduces until stability is obtained. Since a good fit is obtained, the model's prediction capacity is observed to be highly accurate, as shown in Fig. 20, The conclusion that the model is highly accurate is warranted because the actual and predicted values match along the graph from the data points in the crests and troughs. Further, there are no outliers in the data to suggest that there is an anomaly.

7 Comparative Analysis of Models Using the IHEPC Dataset

The comparison of the stacked LSTM ensemble to other DL models is thoroughly explained in this section. The consumption prediction is used to determine whether the stacked LSTM ensemble outperforms other DL algorithms that use the IHEPC dataset. Figure 20 depicts the model and dataset's performance, indicating that the obtained results are satisfactory. Various DL models are compared with the IHEPC dataset to determine performance, as shown in Table 1. The models developed for the datasets include linear regression[38], ANN[39], CNN[40], CNN-LSTM[38], CNN Bidirectional LSTM (CNN-BiLSTM)[41], CNN-LSTM autoencoder[38], CNN Gated Recurrent Unit (CNN-GRU)[42], CNN Echo State

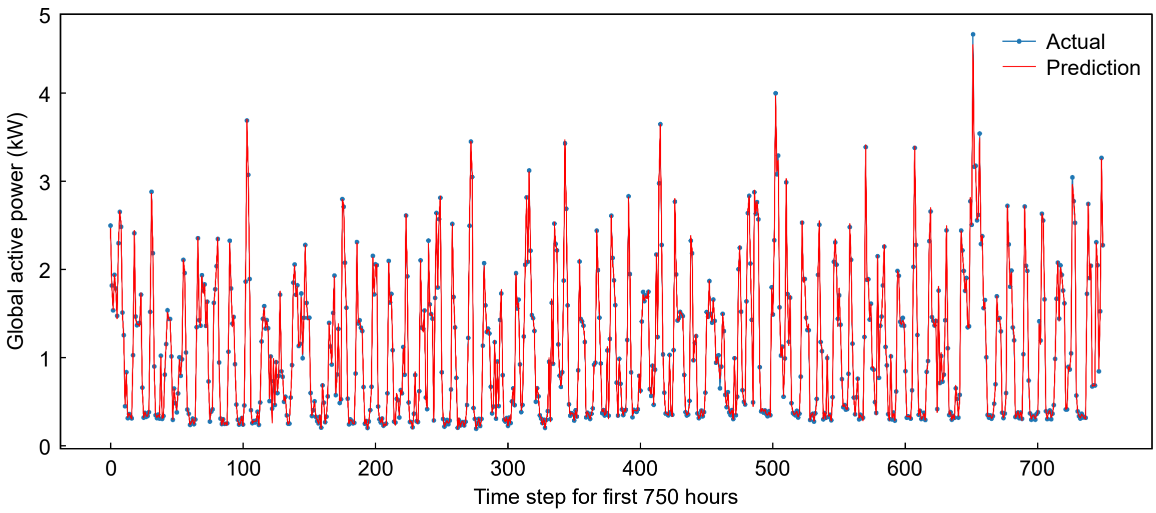


Fig. 20 Model's prediction accuracy.

Network (CNN-ESN)[43], two-stream deep network STLF (namely STLF-Net)[44], residual GRU[45], ESN-CNN[46], Region-based CNN (R-CNN) with Meta-Learning LSTM (ML-LSTM)[20], standard LSTM with LSTM-based Sequence-to-Sequence (S2S) architecture[47], Multiplicative LSTM (M-LSTM)[48], Deep-Broad Network (DB-Net)[49], explainable autoencoder[50], residual GRU-based hybrid model[45], hybrid DL network[51], multi-headed attention model[52], and Conventional LSTM-based hybrid architecture Network (CL-Net)[53]. Compared to all other models, the stacked LSTM ensemble reveals lower error rates of 0.020, 0.013, 0.017, and 0.999 for RMSE, MAE, MAPE, and R^2, respectively. In an examination, based on RMSE, MAE, and MAPE, Kim and Cho[38] determined that the linear regression model's performance has an hourly resolution of 0.6570, 0.5022, and 83.74, respectively. Rajabi and Estebsari[39] determined that the ANN model's performance has RMSE and MAE values of 1.15 and 1.08, respectively. The structure uses recurrence plots to encode time series data into images, and the model performs better than CNN, SVM, and ANN. Khan et al.'s[40] work found that the LSTM autoencoder hybrid CNN model has RMSE and MAE values of 0.67 and 0.47, respectively. The model performs best with daily predictions as opposed to hourly predictions of household electricity consumption. Using the CNN-LSTM DL algorithm, Kim and Cho[38] developed a model to predict residential energy consumption. Their analysis reveals that the RMSE, MAE, and MAPE have values of 0.595, 0.3317, and 32.82, respectively. Ullah et al.[41] created a CNN multilayer bidirectional LSTM network based model for predicting residential energy consumption. They discovered that the RMSE, MAE, and MAPE values are 0.565, 0.346, and 29.10, respectively.

Using the CNN-LSTM autoencoder, Kim and Cho[38] created a model that could anticipate residential energy consumption. According to their research, the RMSE and MAE metrics have model errors of 0.47 and 0.31, respectively. Sajjad et al.'s[42] study utilizing the hybrid model of CNN and GRU to forecast energy consumption shows that the error values are 0.47 (RMSE) and 0.33 (MAE). Khan et al.[43] wanted to use DL algorithms to improve energy harvesting and selected the CNN-ESN model. Their analysis reveals error values of 0.0472 (RMSE) and 0.0266 (MAE)[43]. In 2022, Abdel-Basset et al.[44] used the STLF-Net model for DL analysis and short-term load prediction in residential buildings. The MAPE, RMSE, and MAE are found to be 36.24, 0.4386, and 0.2674, respectively. Khan et al.[45] forecasted energy demand and supply using a residual GRU model. They discovered that the RMSE and MAE are 0.4186 and 0.2635, respectively. The ESN-CNN DL model is utilized by Khan et al. in 2022[46] to enhance energy prediction. The study's error values are 0.2153 (RMSE) and 0.1137 (MAE). Alsharekh et al.[20] improved short-term load prediction using the hybrid model and R-CNN. The RMSE, MAE, MAPE, and R^2 values they discovered are 0.0325, 0.0144, 1.024, and 0.9841, respectively.

Based on its consumptive nature, the S2S model is studied and evaluated using the standard LSTM and an LSTM model based on sequence/no sequence. Findings show that LSTM performs better at hourly resolution but not at a per minute resolution[47]. The RMSE is 0.625. Researchers Kim and Cho[38] developed the CNN-LSTM model, which uses a hybrid connection between the LSTM and CNN networks. The CNN network in the model extracts intricate features from variables that impact consumption. The LSTM algorithm is utilized for modelling temporal information. The RMSE, MAE, and MAPE values are 0.595, 0.3317, and 32.83, respectively. The explainable autoencoder DL model is used to forecast consumption for 15, 30, 45, and 60 minutes in another model with sample data. The specialists utilize a t-SNE calculation to make sense of and imagine the estimated results. The MAE value produced by their model is 0.3953[50]. Khan et al.[49] published works that utilize a hybrid connection of bidirectional LSTM and CNN networks along with the DB-Net algorithm to forecast consumption. The model's error values are 0.1272 (RMSE) and 0.0916 (MAE). Ullah et al.[48] utilized the conventional ML and DL sequential models for energy consumption predictions. Based on error metrics, their investigations reveal that the M-LSTM model has superior prediction ability over the DL and ML algorithms. The M-LSTM model's error values are 0.3296 (RMSE) and 0.3086 (MAE), based on an hourly resolution. Haq et al.[51] predicted energy consumption by residential and commercial users using a novel hybrid DL model. The model acquires RMSE and MAE upsides of 0.324 and 0.311, respectively. The RNN model incorporating multi-headed attention is created by Bu and Cho[52] to forecast energy consumption and selectively determine spatiotemporal characteristics. The MSE value is 0.2662, but the model provides no error metrics. Khan et al.[45] created a hybrid model with Residual GRU (R-GRU) and dilated convolutions. The RMSE and MAE error metric scores are 0.4186 and 0.2635, respectively, when this model is used to predict energy generation and consumption. Khan et al.[53] modelled the CL-Net architecture using the ConvLSTM hybrid to assess the model's accuracy in predicting energy consumption. Their testing of the model results in an RMSE score of 0.122 and an MAE score of 0.088. In this comparison, most of the models perform better than this one. The RMSE, MAE, MAPE, and R^2 of the proposed model are 0.020, 0.013, 0.017, and 0.999, respectively. The developed model has the lowest error scores of any model, indicating that it accurately predicts energy consumption. The performance comparison of different prediction models is summarized in Table 2.

8 Conclusion

Using the developed model, it is possible to accurately predict energy consumption. Compared to other studies that lack dimension reduction algorithms[54] to allow for seasonality observation, the stacked snapshot LSTM ensemble shows that it is possible to investigate seasonality attributed to energy consumption. Another advantage of using the model is that it is easy to train and validate. Furthermore, it supports big data and could dynamically support the model weights used without many adjustments to the dataset. The model is designed to be simple and functional, such that it could provide a relatively inexpensive method of evaluating big energy datasets. Finally, the model includes an algorithm that trains the LSTM model sequentially. This allows the model to learn different patterns. The advantage of this feature is that the estimates provided by the final model are very robust and accurate due to the high levels of generalization. The model's accuracy and stability are measured using RMSE, MAE, MAPE, and R^2 as 0.020, 0.013, 0.017, and 0.999, respectively. 0.020 for the RMSE demonstrates the model's high level of stability, while 0.017 for MAPE is very close to 0, signalling high-level accuracy with the model. The R^2 results of 0.999 is nearly 1, which shows the model's good performance. Accuracy, stability, and performance give the model consistent results, but if the results are out of the ordinary, the model allows for humans to understand the causes of those results. This refers to interpretability, which permits a human to explain the cause-and-effect of such anomalous results.

Table 2 Prediction model comparison.

Model	RMSE	MAE	MAPE	R^2
Linear regression[38]	0.6570	0.5022	83.740	–
ANN[39]	1.1500	1.0800	–	–
CNN[40]	0.6700	0.4700	–	–
CNN-LSTM[38]	0.5950	0.3317	32.830	–
CNN-BDLSTM[41]	0.5650	0.3460	29.100	–
CNN-LSTM autoencoder [38]	0.4700	0.3100	–	–
CNN-GRU[42]	0.4700	0.3300	–	–
CNN-ESN[43]	0.0472	0.0266	–	–
STLF-NET[44]	0.4386	0.2674	36.240	–
ESN-CNN[46]	0.2153	0.1137	–	–
R-CNN with ML-LSTM[20]	0.0325	0.0144	1.024	0.9841
Standard LSTM and LSTM-based S2S architecture[47]	0.6250	–	–	–
M-LSTM[48]	0.3296	0.3086	–	–
DB-NET[49]	0.1272	0.0916	–	–
Explainable autoencoder[50]	–	0.3953	–	–
Residual GRU-based hybrid model[45]	0.4186	0.2635	–	–
Hybrid DL network[51]	0.3240	0.3110	–	–
Multi-headed attention model[52]	0.2662	–	–	–
CL-Net architecture[53]	0.1220	0.0880	–	–
Proposed model	**0.0200**	**0.0130**	**0.0170**	**0.9990**

Despite the model's advantages, there are some limitations encountered, such that the model's performance is not tested with real-time performance to determine its robustness. Therefore, the model should be tested in the future to determine its dynamic performance in a real-time environment. In addition, in the present study, the model's accuracy and stability results are used to infer interpretability. Future studies should determine an independent way to measure a model's interpretability.

Based on these findings, the energy sector could use such a model, as it provides high-value insights, value-addition, and service improvements based on its effective use of big data regarding energy consumption. Because energy data reliably and in real time reflects economic activity trends of populations, businesses, and the community, by virtue of these technological advantages and data resources, this can be regarded as an important data resource for energy companies to develop data platforms with high-level accuracy and performance algorithms that can integrate data from multiple industries, facilitating the transformation and upgrading of governments and organizations.

For energy companies, the application of the stacked LSTM snapshot ensemble and other DL models to energy consumption is still in early stages of development. The data points have numerous compound values that must be discovered and mined by internal and external businesses to obtain additional insights and trends. Energy generation companies need to pay close attention to how big data about energy consumption work and how businesses and multinational corporations use them by working closely with these businesses to learn more from them. Businesses are conducting energy research to speed up the sharing of energy data. Companies that produce energy need to pay close attention to their key dynamics, intensify learning, and use big data and learning applications. Additionally, they should investigate cooperative endeavours and share mechanisms for collective improvement.

Translating the findings from the stacked LSTM snapshot ensemble energy consumption prediction model into the analysis of the energy usage dataset for reporting and determining faults, optimization, and forecasted maintenance in households and businesses is the future direction of this study.

Acknowledgment

The authors gratefully acknowledge the support and invaluable guidance provided by the Faculty of Computing and Information Technology (FCIT), King Abdulaziz University (KAU), Jeddah, Kingdom of Saudi Arabia.

References

[1] M. Cellura, F. Guarino, S. Longo, and G. Tumminia, Climate change and the building sector: Modelling and energy implications to an office building in southern Europe, *Energy Sustain. Dev.*, vol. 45, pp. 46–65, 2018.

[2] T. Zhang, D. Wang, H. Liu, Y. Liu, and H. Wu, Numerical investigation on building envelope optimization for low-energy buildings in low latitudes of China, *Build. Simul.*, vol. 13, no. 2, pp. 257–269, 2020.

[3] A. D. Pham, N. T. Ngo, T. T. Ha Truong, N. T. Huynh, and N. S. Truong, Predicting energy consumption in multiple buildings using machine learning for improving energy efficiency and sustainability, *J. Cleaner Prod.*, vol. 260, p. 121082, 2020.

[4] A. Raza, T. N. Malik, M. F. N. Khan, and S. Ali, Energy management in residential buildings using energy hub approach, *Build. Simul.*, vol. 13, no. 2, pp. 363–386, 2020.

[5] Z. Wang and R. S. Srinivasan, A review of artificial intelligence based building energy use prediction: Contrasting the capabilities of single and ensemble prediction models, *Renew. Sustain. Energy Rev.*, vol. 75, pp. 796–808, 2017.

[6] H. Song, A. K. Qin, and F. D. Salim, Evolutionary model construction for electricity consumption prediction, *Neural Comput. Appl.*, vol. 32, no. 16, pp. 12155–12172, 2020.

[7] E. Jahani, K. Cetin, and I. H. Cho, City-scale single family residential building energy consumption prediction using genetic algorithm-based numerical moment matching technique, *Build. Environ.*, vol. 172, p. 106667, 2020.

[8] J. Xu, W. Gao, and X. Huo, Analysis on energy consumption of rural building based on survey in northern China, *Energy Sustain. Dev.*, vol. 47, pp. 34–38, 2018.

[9] A. Zeng, H. Ho, and Y. Yu, Prediction of building electricity usage using Gaussian Process Regression, *J. Build. Eng.*, vol. 28, p. 101054, 2020.

[10] C. Dai, H. Zhang, E. Arens, and Z. Lian, Machine learning approaches to predict thermal demands using skin temperatures: Steady-state conditions, *Build. Environ.*, vol. 114, pp. 1–10, 2017.

[11] C. Xu, H. Chen, J. Wang, Y. Guo, and Y. Yuan, Improving prediction performance for indoor temperature in public buildings based on a novel deep learning method, *Build. Environ.*, vol. 148, pp. 128–135, 2019.

[12] D. K. Bui, T. N. Nguyen, T. D. Ngo, and H. Nguyen-Xuan, An artificial neural network (ANN) expert system enhanced with the electromagnetism-based firefly algorithm (EFA) for predicting the energy consumption in buildings, *Energy*, vol. 190, p. 116370, 2020.

[13] M. A. Jallal, A. González-Vidal, A. F. Skarmeta, S. Chabaa, and A. Zeroual, A hybrid neuro-fuzzy inference system-based algorithm for time series forecasting applied to energy consumption prediction, *Appl. Energy*, vol. 268, p. 114977, 2020.

[14] S. Ganguly, A. Ahmed, and F. Wang, Optimised building energy and indoor microclimatic predictions using knowledge-based system identification in a historical art gallery, *Neural Comput. Appl.*, vol. 32, no. 8, pp. 3349–3366, 2020.

[15] J. S. Chou, N. T. Ngo, W. K. Chong, and G. E. Jr Gibson, Big data analytics and cloud computing for sustainable building energy efficiency, in *Start-Up Creation*: *The Smart Eco-Efficient Built Environment*, F. Pacheco-Torgal, E. Rasmussen, C. G. Granqvist, V. Ivanov, A. Kaklauskas, and S. Makonin, eds. Cambridge, UK: Woodhead Publishing, 2016, pp. 397–412.

[16] S. Seyedzadeh, F. Pour Rahimian, P. Rastogi, and I. Glesk, Tuning machine learning models for prediction of building energy loads, *Sustain. Cities Soc.*, vol. 47, p. 101484, 2019.

[17] S. Badillo, B. Banfai, F. Birzele, I. I. Davydov, L. Hutchinson, T. Kam-Thong, J. Siebourg-Polster, B. Steiert, and J. D. Zhang, An introduction to machine learning, *Clin. Pharmacol. Ther.*, vol. 107, no. 4, pp.

871–885, 2020.

[18] J. Runge and R. Zmeureanu, Forecasting energy use in buildings using artificial neural networks: A review, *Energies*, vol. 12, no. 17, p. 3254, 2019.

[19] J. Moon, S. Park, S. Rho, and E. Hwang, A comparative analysis of artificial neural network architectures for building energy consumption forecasting, *Int. J. Distrib. Sens. Netw*., vol. 15, no. 9, p. 155014771987761, 2019.

[20] M. F. Alsharekh, S. Habib, D. A. Dewi, W. Albattah, M. Islam, and S. Albahli, Improving the efficiency of multistep short-term electricity load forecasting via R-CNN with ML-LSTM, *Sensors*, vol. 22, no. 18, p. 6913, 2022.

[21] A. Moradzadeh, S. Zakeri, M. Shoaran, B. Mohammadi-Ivatloo, and F. Mohammadi, Short-term load forecasting of microgrid via hybrid support vector regression and long short-term memory algorithms, *Sustainability*, vol. 12, no. 17, p. 7076, 2020.

[22] R. K. Agrawal, F. Muchahary, and M. M. Tripathi, Long term load forecasting with hourly predictions based on long-short-term-memory networks, in *Proc. 2018 IEEE Texas Power and Energy Conf.* (*TPEC*), Station, TX, USA, 2018, pp. 1–6.

[23] J. Q. Wang, Y. Du, and J. Wang, LSTM based long-term energy consumption prediction with periodicity, *Energy*, vol. 197, p. 117197, 2020.

[24] A. Galicia, R. Talavera-Llames, A. Troncoso, I. Koprinska, and F. Martínez-Álvarez, Multi-step forecasting for big data time series based on ensemble learning, *Knowl.-Based Syst.*, vol. 163, pp. 830–841, 2019.

[25] Y. Yang, W. Hong, and S. Li, Deep ensemble learning based probabilistic load forecasting in smart grids, *Energy*, vol. 189, p. 116324, 2019.

[26] M. Azarafza, M. Azarafza, H. Akgün, P. M. Atkinson, and R. Derakhshani, Deep learning-based landslide susceptibility mapping, *Sci. Rep.*, vol. 11, no. 1, p. 24112, 2021.

[27] P. T. T. Ngo, M. Panahi, K. Khosravi, O. Ghorbanzadeh, N. Kariminejad, A. Cerda, and S. Lee, Evaluation of deep learning algorithms for national scale landslide susceptibility mapping of Iran, *Geosci. Front.*, vol. 12, no. 2, pp. 505–519, 2021.

[28] W. Xie, X. Li, W. Jian, Y. Yang, H. Liu, L. F. Robledo, and W. Nie, A novel hybrid method for landslide susceptibility mapping-based GeoDetector and machine learning cluster: A case of Xiaojin County, China, *ISPRS Int. J. Geo-Inf.*, vol. 10, no. 2, p. 93, 2021.

[29] A. L. Achu, G. Gopinath, and U. Surendran, Landslide susceptibility modelling using deep-learning and machine-learning methods—A study from southern Western Ghats, India, in *Proc. 2021 IEEE Int. India Geoscience and Remote Sensing Symp.* (*InGARSS*), Ahmedabad, India, 2021, pp. 360–364.

[30] G. Hebrail and A. Berard, Individual household electric power consumption, https://archive.ics.uci.edu/dataset/235/individual+household+electric+power+consumption, 2012.

[31] J. F. Torres, A. Galicia, A. Troncoso, and F. Martínez-Álvarez, A scalable approach based on deep learning for big data time series forecasting, *Integr. Comput.-Aided Eng.*, vol. 25, no. 4, pp. 335–348, 2018.

[32] J. F. Torres, D. Gutiérrez-Avilés, A. Troncoso, and F. Martínez-Álvarez, Random hyper-parameter search-based deep neural network for power consumption forecasting, in *Proc. 15th Int. Work-Conf. Artificial Neural Networks*, Gran Canaria, Spain, 2019, pp. 259–269.

[33] W. Kong, Z. Y. Dong, Y. Jia, D. J. Hill, Y. Xu, and Y. Zhang, Short-term residential load forecasting based on LSTM recurrent neural network, *IEEE Trans. Smart Grid*, vol. 10, no. 1, pp. 841–851, 2019.

[34] N. M. Ibrahim, A. I. Megahed, and N. H. Abbasy, Short-term individual household load forecasting framework using LSTM deep learning approach, in *Proc. 5th Int. Symp. Multidisciplinary Studies and Innovative Technologies* (*ISMSIT*), Ankara, Turkey, 2021, pp. 257–262.

[35] C. Xu, W. Li, M. Yu, J. Xu, J. Liu, Y. Wang, and L. Zhu, A correlation sorting-LSTM method for high accuracy short-term load forecasting based on smart meter data, in *Proc. 7th Int. Conf. Information, Cybernetics, and Computational Social Systems* (*ICCSS*), Guangzhou, China, 2020, pp. 755–760.

[36] M. Ma, C. Liu, R. Wei, B. Liang, and J. Dai, Predicting machine's performance record using the stacked long short-term memory (LSTM) neural networks, *J. Appl. Clin. Med. Phys.*, vol. 23, no. 3, p. E13558, 2022.

[37] G. Huang, Y. Li, G. Pleiss, Z. Liu, J. E. Hopcroft, and K. Q. Weinberger, Snapshot ensembles: Train 1, get M for free, arXiv preprint arXiv:1704.00109, 2017.

[38] T. Y. Kim and S. B. Cho, Predicting residential energy consumption using CNN-LSTM neural networks, *Energy*, vol. 182, pp. 72–81, 2019.

[39] R. Rajabi and A. Estebsari, Deep learning based forecasting of individual residential loads using recurrence plots, in *Proc. 2019 IEEE Milan PowerTech*, Milan, Italy, 2019, pp. 1–5.

[40] Z. A. Khan, T. Hussain, A. Ullah, S. Rho, M. Lee, and S. W. Baik, Towards efficient electricity forecasting in residential and commercial buildings: A novel hybrid CNN with an LSTM-AE based framework, *Sensors*, vol. 20, no. 5, p. 1399, 2020.

[41] F. U. M. Ullah, A. Ullah, I. U. Haq, S. Rho, and S. W. Baik, Short-term prediction of residential power energy consumption via CNN and multi-layer bi-directional LSTM networks, *IEEE Access*, vol. 8, pp. 123369–123380, 2020.

[42] M. Sajjad, Z. A. Khan, A. Ullah, T. Hussain, W. Ullah, M. Y. Lee, and S. W. Baik, A novel CNN-GRU-based hybrid approach for short-term residential load forecasting, *IEEE Access*, vol. 8, pp. 143759–143768, 2020.

[43] Z. A. Khan, T. Hussain, and S. W. Baik, Boosting energy harvesting via deep learning-based renewable power generation prediction, *J. King Saud Univ.—Sci.*, vol. 34, no. 3, p. 101815, 2022.

[44] M. Abdel-Basset, H. Hawash, K. Sallam, S. S. Askar, and M. Abouhawwash, STLF-Net: Two-stream deep network

for short-term load forecasting in residential buildings, *J. King Saud Univ.—Comput. Inf. Sci*., vol. 34, no. 7, pp. 4296–4311, 2022.

[45] S. U. Khan, I. U. Haq, Z. A. Khan, N. Khan, M. Y. Lee, and S. W. Baik, Atrous convolutions and residual GRU based architecture for matching power demand with supply, *Sensors*, vol. 21, no. 21, p. 7191, 2021.

[46] Z. A. Khan, T. Hussain, I. U. Haq, F. U. M. Ullah, and S. W. Baik, Towards efficient and effective renewable energy prediction via deep learning, *Energy Rep.*, vol. 8, pp. 10230–10243, 2022.

[47] D. L. Marino, K. Amarasinghe, and M. Manic, Building energy load forecasting using deep neural networks, in *Proc. 42nd Annu. Conf. IEEE Industrial Electronics Society*, Florence, Italy, 2016, pp. 7046–7051.

[48] F. U. M. Ullah, N. Khan, T. Hussain, M. Y. Lee, and S. W. Baik, Diving deep into short-term electricity load forecasting: Comparative analysis and a novel framework, *Mathematics*, vol. 9, no. 6, p. 611, 2021.

[49] N. Khan, I. U. Haq, S. U. Khan, S. Rho, M. Y. Lee, and S. W. Baik, DB-Net: A novel dilated CNN based multi-step forecasting model for power consumption in integrated local energy systems, *Int. J. Electr. Power Energy Syst*., vol. 133, p. 107023, 2021.

[50] J. Y. Kim and S. B. Cho, Electric energy consumption prediction by deep learning with state explainable autoencoder, *Energies*, vol. 12, no. 4, p. 739, 2019.

[51] I. U. Haq, A. Ullah, S. U. Khan, N. Khan, M. Y. Lee, S. Rho, and S. W. Baik, Sequential learning-based energy consumption prediction model for residential and commercial sectors, *Mathematics*, vol. 9, no. 6, p. 605, 2021.

[52] S. J. Bu and S. B. Cho, Time series forecasting with multi-headed attention-based deep learning for residential energy consumption, *Energies*, vol. 13, no. 18, p. 4722, 2020.

[53] N. Khan, I. U. Haq, F. U. M. Ullah, S. U. Khan, and M. Y. Lee, CL-Net: ConvLSTM-based hybrid architecture for batteries' state of health and power consumption forecasting, *Mathematics*, vol. 9, no. 24, p. 3326, 2021.

[54] U. Ugurlu, I. Oksuz, and O. Tas, Electricity price forecasting using recurrent neural networks, *Energies*, vol. 11, no. 5, p. 1255, 2018.

Mona Ahamd Alghamdi received the master degree in information systems from King Abdulaziz University (KAU), Jeddah, Kingdom of Saudi Arabia. Currently, she is a PhD candidate at Information Systems Department, KAU. Also, she is working as a lecturer at Faculty of Information Technology, Arab Open University, Jeddah, Saudi Arabia. She has published several articles in reputable peer-reviewed international journals. Her research interests include artificial intelligence, data analytics, and decision support systems.

Mahmoud Ragab obtained the PhD degree from Christian-Albrechts-University at Kiel (CAU), Schleswig-Holstein, Germany. Currently, he is a professor of data science at Information Technology Department, Faculty of Computing and Information Technology (FCIT), KAU, Jeddah, Kingdom of Saudi Arabia, and also working at Mathematics Department, Faculty of Science, Al-Azhar University, Egypt. He workes in different research groups at many universities, such as the Combinatorial Optimization and Graph Algorithms Group (COGA), Faculty II Mathematics and Natural Sciences, Berlin University of Technology, Berlin, Germany, the British University in Egypt BUE, Automation, Integrated Communication Systems Group, Ilmenau University of Technology TU Ilmenau, Thüringen, Germany. His research focuses on AI algorithms, deep learning, sorting, optimization, mathematical modeling, data science, neural networks, time series analysis, and quantum computation.

Abdullah S. AL-Malaise AL-Ghamdi is a professor at FCIT, KAU, Jeddah, King of Saudi Arabia, and also a professor at Information Systems Department, School of Engineering, Computing and Design, Dar Al-Hekma University, Kingdom of Saudi Arabia. He received the PhD degree in computer science from George Washington University, USA in 2003. He is a member of the Scientific Council and holding position as a secretary general of Scientific Council, KAU. In addition, he is working as the head of consultant's Unit at Vice-President for Development Office, as a consultant to Vice-President for Graduate Studies & Scientific Research, KAU. Previously, he has worked as the head of Information Systems Department, Dar A1-Hekma University, Egypt, the vice dean for Graduate Studies and Scientific Research, and the head of Computer Skills Department at FCIT. His main research areas are software engineering and systems, artificial intelligence, data analytics, business intelligence, and decision support system.