2022 4th International Conference on Sustainable Technologies for Industry 4.0 (STI), 17-18 December, Dhaka

Electricity Demand Forecasting Using Machine Learning Models

Author I*, Author II†, Author III*, and Author IV‡
Department of Computer Science & Engineering
Chittagong University of Engineering and Technology
Chittagong-4349, Bangladesh
Email: authorI@gmail.com, authorII@ieee.org, authorIII@gmail.com, authorIV@universityz.edu

***Abstract*—Electricity demand forecasting is one of the essential ways to manage the power system of a country efficiently and effectively. By facilitating the decision-making process for electricity generation and consumption, electricity demand forecasting serves a useful purpose in decision-making for power system companies. In this work, we develop an electricity demand forecasting model that predicts electricity demand using our own dataset. Our dataset includes numerous parameters relevant to Chittagong's electricity covering more than ten years of data (January 2012 to June 2022). The demand data was taken from the Bangladesh Power Development Board website (BPDB) using web scapping and other features via manual insertion. Several models like Linear Regression(LR), Decision Tree(DT), Random Forest(RF), Support Vector Regression(SVR), and K Nearest Neighbors(KNN) based on machine learning (ML) methods have been applied to observe how these algorithms perform in forecasting electricity. The performance of the discussed ML methods has been analyzed based on various evaluation metrics such as R square, Mean Average Error (MAE), Mean Squared Error (MSE) and Root Mean Squared Error (RMSE). The outcomes obtained from the analysis show that the Random Forest ML approach outperforms the other conventional ML approaches in terms of R square, MAE, MSE, and RMSE since it has the highest accuracy according to R square (86.24%) and the lowest MAE, MSE, and RMSE score (0.0450, 0.0042, and 0.0652, respectively) compared to the other discussed conventional ML approaches. This Random Forest method has the capability to forecast the electricity demand more precisely and can plan ahead for maintenance and load distribution, which can aid electricity generation and distribution companies.**

***Keywords*—Electricity Demand Forecast; Machine Learning; Prediction; Linear Regression; Decision Tree; Random Forest; Support Vector Regression; K Nearest Neighbours.**

I. Introduction

Since electricity has become such an integral element of everyday life in the modern world, each country's economic development is intimately correlated with the infrastructure, network, and accessibility of its power supply. As a result, there is now a greater need for power on a global scale for home and business uses. However, over time, electricity rates have fluctuated, concealing the truth that global demand for electricity exceeds supply. As a result, a number of studies have been conducted to forecast how much electrical energy will be used for industrial, commercial, and residential applications in the future. This information enables power producers, distributors, and suppliers to make plans and encourage customers to conserve energy.

In the entire operation and planning of power systems, electricity demand forecasting is critical. To operate the power system effectively and efficiently, precise electricity demand forecasting is important. Electricity demand forecasting can be beneficial in areas such as power outages, demand-supply coordination, maintenance, operating costs, and structural construction. As a result, power demand forecasting has been a prominent research topic in recent decades [1]. The natures of these projections are also distinct. Forecasts for the short term range from one hour to a week. The term "hourly load forecast" refers to a prediction of short-term power consumption. The medium term predictions span a few weeks to a few months or even a few years. A term for a medium term forecast is the monthly load forecast. Energy trading, cost control, and distribution system planning all require long term electricity demand forecasts. A long term projection must be accurate for a minimum of 5 years and a maximum of 25 years. The system's production and distribution expansion plans are based on this forecast. An annual peak load is a term used to describe a long-term forecast [2].

Despite the fact that there has been a lot of new research in the subject of electricity demand forecasting, more robust and accurate electricity demand forecast models are still needed. Because it is used in decision making, an accurate prediction of variance in future electricity demand is critical for both electric customers and utilities. The main objectives of our work are as follows:

1) To build our own dataset combining energy data, weather data and population density data.
2) To predict the electricity demand accurately using well defined machine learning methods.
3) To compare the performance of the applied algorithms for achieving the best approach for our system.

We will propose a system that will predict the future electricity demand using machine learning approach.

II. Literature Review

Many academic publications on this topic have been published in recent years. This might be because the study of energy economy is seeing increased interest in forecasting electricity demand. Forecasting electricity demand is critical to the management and planning of electrical power and energy systems. Researchers use a variety of approaches to anticipate power consumption. So, it is a very demanding research area.

Camurdan et al. [3], developed a method to predict electricity demand using three different machine learning models. They use different evaluation matrices to test the performance of the results. Here the highest accuracy of 98% was obtained from the random forest model.

Shabbir et al. [4], applied three different ML algorithms: linear regression, tree-based regression, and Support Vector Machine (SVM)-based regression on a large dataset of an Estonian household.The results show that SVM is the most effective method since it provides the lowest Root Mean Square Error(RMSE).

Anik et al. [5], a simultaneous equation framework was applied to an annual database over a 47-year period (1972-2018) to forecast future energy demand.

Liu et al. [6], compared single (ANN and SVM) and hybrid machine learning methods which have different faetures and they are applicable in various situations and they also have their own strengths and weaknesses.

Gajowniczek et al. [7], developed an algorithmic technique for load modeling through peak detection and used the Polish Power System based on data between 1 January 2008 and 31 December to feed the forecasting system for this purpose. SVM gives the most promising results for peak classification and ANN for forecasting.

Jain et al. [8], proposed a method to forecast the electricity consumption using the ARIMA model. For the years 2004-2008, the seasonal ARIMA model was determined to be the best model for predicting energy usage in IIT(ISM). The limitation of this paper is that ARIMA model can be limited in forecasting extreme values.

Behm et al. [9], presented a method for forecasting long-term climate hourly power consumption using ann. Their main contribution is to present a new method to forecast electricity loads as necessary input data for energy system modeling. The limitation is that they didn't use the correct scaling method for loads.

Islam et al. [10], presented a smart grid infrastructure in Chattogram, Bangladesh, using an LSTM-based electrical load forecasting approach, and the results show that the LSTM performs better than the SVM.

Eseye et al. [11], suggested a ml-based hybrid feature selection technique for extracting the most relevant and nonredundant information for improved short-term forecasting of power demand in distributed energy systems.

Patel et al. [12], introduced a time series model for predicting the future load demand. With the aid of projected power system or future consumption, an electric utility or energy management system may plan effectively.

Rabbi et al. [13], proposed a model to forecast Bangladesh's annual electricity demand using a multivariate time series model. They anticipate Bangladesh's power consumption using a Multivariate Time Series since it provides more data.

In [14], authors presented a new Grey-based prediction method for forecasting very-short-term electric power consumption in order to manage electricity demand. Electric consumption waste can be prevented using this approach.

III. Methodology

After conducting extensive research, we developed the technique for our work in a series of steps. The overall design of the framework is depicted in Figure 1.Initially, the dataset is read from the CSV file. The dataset consists of three types of data, such as energy data, weather data, and population density data. Data preprocessing is the combination of following steps: data cleaning, data normalization. After going through the preprocessing stage, we get the final dataset for further implementation. The final dataset is split into two portions: the training set and the testing set. Machine learning algorithms are applied to the training sets. The prediction models are built and they make predictions on the test set, yielding the evaluation metrics and comparing their performances.

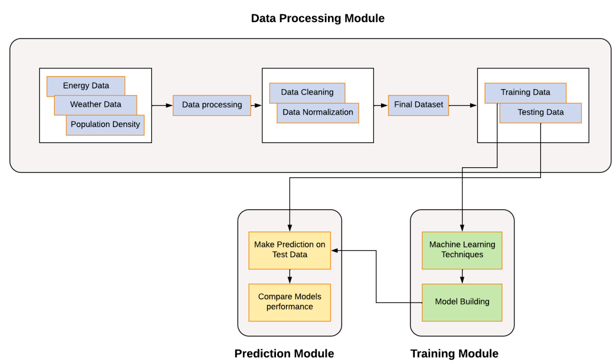

Fig. 1: Overview of framework

A. Data Collection

Initially, the data was collected from various sources for the Chittagong district to build the dataset.The dataset contains data for a period of more than 10 years (01-01-2012 to 31-05-2022). We have considered some features as the main features for our research, and these are: Date, Demand (MW), Temp_max, Temp_min, Humidity, Pressure, Wind_speed, and Population Density. Studying some research papers and articles, we have found that these features have a high correlation with electricity demand.

1) The data for demand (mw) was obtained from the Bangladesh Power and Development Board (BPDB), which hosts the area-wise electricity demand. The data was extracted using the web scrapper chrome extension. By installing the web scrapper extension from the chrome web store to extract the data.
2) The weather information (temp_max, temp_min, humidity, pressure, and wind_speed) was collected from the timeanddate.com website.
3) The population density data was collected from the worldpopulationreview.com website.

B. Feature Description

We have collected the dataset for our analysis from several websites. There are 3800 data in the dataset, and there are 8 feature values. The dataset sample is shown in Figure. The description of the features are given below:

Time factor play a major role in forecasting electricity demand. Different weekdays exhibit different patterns in the electricity distribution. Noticeable variances exist between the load on weekdays and weekends, and these differences are significant for load forecasting. During the weekend demand declines, especially on Saturday, the second day of the week-end, and then sharply increases on Sunday. According to the monthly historical analysis of load demand, the power consumption rate behaves in a certain way where it is prominently greater in summer than in winter. [15]

- **Demand data:** Daily electrical demand.
- **Weather Data:** One of the most crucial independent variables for forecasting power is the weather. [16]
  Temperature: During the summer, the electricity demand is about twice as high as the winter electricity demand. In Chittagong, the coldest part of the year begins in early November. Then the temperature drops, and mild weather conditions lower the amount of electricity on the power network.
  Humidity: Humidity is another weather element that affects the total load curve.
  Pressure: The temperature drops when air pressure drops. Additionally, it explains why air becomes colder at higher elevations due to decreased pressure. A low-pressure area with stormy, rainy weather is headed your way, according to falling pressure. Then the temperature drops, also decreasing the electricity demand.
  Wind Speed: A windless day with an air temperature of 15 degrees has no wind effect. 15 degrees is how it feels. A 20-mph wind, on the other hand, makes a 15-degree air temperature feel like only 2! The procedure grows more effective the faster the wind blows, making it appear colder and colder.Then the demand for electricity decreased.
- **Population Density:** Population growth has been named as the main cause of the temperature rise and subsequent increase in power use. [16]

C. Preprocessing of Data

Data preprocessing is the process of preparing the raw data into an understandable format. It is an important step in machine learning models that improves the quality of the data to encourage the extraction of valuable insights from the data. Data preprocessing is also an data mining approach that can't work with raw data.

1) Data cleaning: The dataset ultimately contained 3800 rows as a result of this. One of the important steps in machine learning is data cleaning. There are numerous methods for cleaning data. The handling of missing values is one of them. In our dataset, there are some missing values when collecting the data from the website. Therefore, each row containing a date was checked for any null or zero values before applying ML models. The mean value of all the other values from earlier observations was used to replace any null or zero values that were found.

2) Data normalization: Every dataset does not require normalization for machine learning. It is applied whenever a dataset's features have different ranges.It helps to improve a machine learning model's performance and reliability. Even though there are numerous feature normalization methods in machine learning, we have used the Min-Max Scaling method in our dataset because our dataset's features have a different range of values. Min-Max Scaling: Normalization, is a scaling technique where values are shifted and rescaled to conserve their ranges between 0 and 1.

D. Data Analysis

An illustration of a correlation matrix that shows the relationship between various features is called a correlation heatmap. Correlation can have any value between -1 and 1. A weaker linear relationship between the two variables is indicated by values that are closer to zero. When the values are close to 1 the variables are said to be more positively correlated whereas values close to -1 are said to be more negatively correlated.

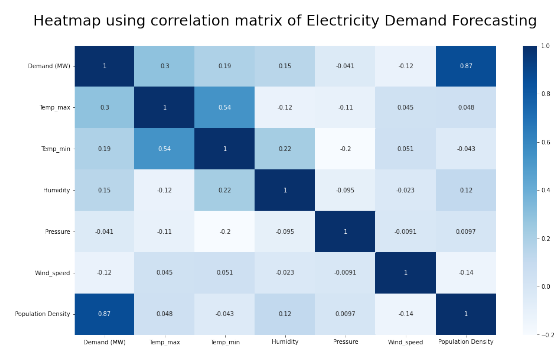

Fig. 2: Correlation Heatmap

E. Training and Testing

In this step, data are split into train(80%) and test(20%) set. we used the cross validation method to assess how well our machine learning models perform on unseen data. In the 5-fold cross-validation (CV) procedure, typically involves dividing the data into k folds at random; in our case, k = 5. After that, k-1 folds are used to train the model, and only one fold is used to test the model. This process is repeated k times.However, all data in this work are first divided into training and testing datasets, with a training dataset being used for cross-validation. We will allow our model to observe the target values Demand (MW), along with other features, during the training session. By doing so, we can make sure that our model is developing the knowledge necessary to predict the target value (Demand (MW)) from features other than the target value. The model gradually discovers any correlations between the features and the target value during the training phase. For it comes time to evaluate the model, it no longer has access to the target values and instead relies entirely on the characteristics of the test set when making predictions. The performance of the model is then calculated by comparing the predictions to the actual values of the target from the test set.

IV. Result Analysis

The findings from the models are determined together with common evaluation metrics like R square, MAE, MSE, and RMSE in this section on evaluation. The best model is picked based on the results received.

A. Evaluation of Linear Regression model and results

Figure 3 shows a comparison between the actual and expected electricity usage based on our model. We can see that an LR virtually predicts data points that match the pattern of electricity usage. In addition, as Table I illustrates, it produces lower MAE, MSE, and RMSE scores. It was observed that linear regression models give 82.53% accuracy according to R square. The Evaluation results of Linear Regression model are as follows:

TABLE I: Prediction Errors and accuracy score of LR

Prediction Errors and accuracy score of LR	
Mean Absolute Error	0.0582
Mean Square Error	0.0057
Root Mean Square Error	0.0761
R square(%)	82.53%

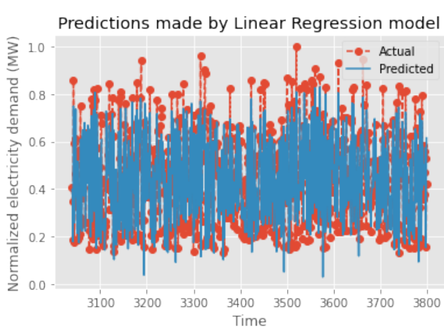

Fig. 3: Graphical representation of Actual vs predicted results

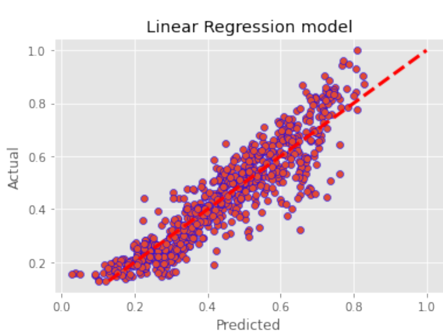

Fig. 4: Actual Vs Predicted observations for Linear Regresion

B. Evaluation of Decision Tree Regression model and results

Figure 5 shows a comparison between the actual and expected electricity usage based on our model. We can see that an DT virtually predicts data points that match the pattern of electricity usage. In addition, as Table II illustrates, it produces lower MAE, MSE, and RMSE scores. It was observed that Decision Tree regression models give 74.16% accuracy according to R square. The Evaluation results of Decision Tree Regression model are as follows:

TABLE II: Prediction Errors and accuracy score of DT

Prediction Errors and accuracy score of DT	
Mean Absolute Error	0.0615
Mean Square Error	0.0081
Root Mean Square Error	0.0903
R square(%)	74.16%

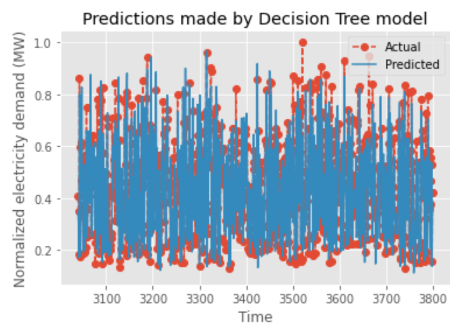

Fig. 5: Graphical representation of Actual vs predicted results

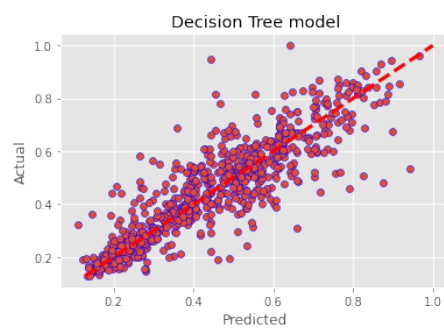

Fig. 6: Actual Vs Predicted observations for Decision Tree Regression

C. Evaluation of Random Forest Regression model and results

Figure 7 shows a comparison between the actual and expected electricity usage based on our model. We can see that an RF virtually predicts data points that match the pattern of electricity usage. In addition, as Table III illustrates, it produces lower MAE, MSE, and RMSE scores. It was observed that Random forest regression models give 86.24% accuracy according to R square. The Evaluation results of Random Forest Regression model are as follows:

TABLE III: Prediction Errors and accuracy score of RF

Prediction Errors and accuracy score of RF	
Mean Absolute Error	0.0450
Mean Square Error	0.0042
Root Mean Square Error	0.0652
R square(%)	86.24%

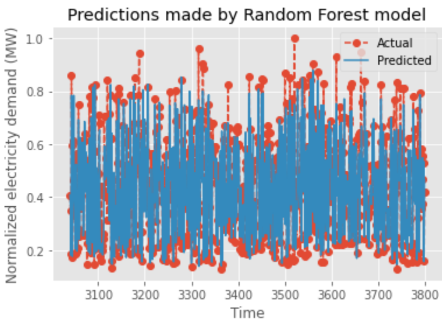

Fig. 7: Graphical representation of Actual vs predicted results

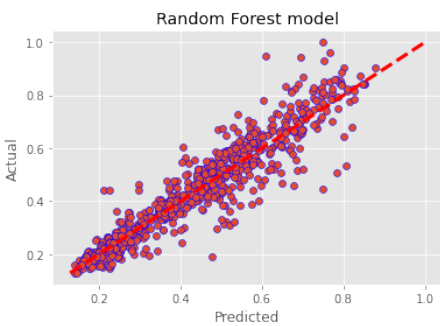

Fig. 8: Actual Vs Predicted observations for Random Forest Regression

D. Evaluation of Support Vector Regression model and results

Figure 9 shows a comparison between the actual and expected electricity usage based on our model. We can see that an SVR virtually predicts data points that match the pattern of electricity usage. In addition, as Table IV illustrates, it produces lower MAE, MSE, and RMSE scores. It was observed that Support vector regression models give 84.52% accuracy according to R square. The Evaluation results of Support Vector Regression model are as follows:

TABLE IV: Prediction Errors and accuracy score of SVR

Prediction Errors and accuracy score of SVR	
Mean Absolute Error	0.0540
Mean Square Error	0.0052
Root Mean Square Error	0.0724
R square(%)	84.52%

E. Evaluation of K Nearest Neighbours model and results

Figure 11 shows a comparison between the actual and expected electricity usage based on our model. We can see that an KNN virtually predicts data points that match the pattern of

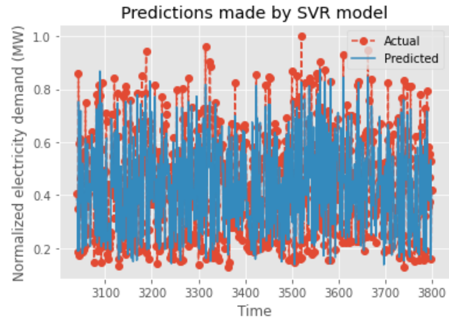

Fig. 9: Graphical representation of Actual vs predicted results

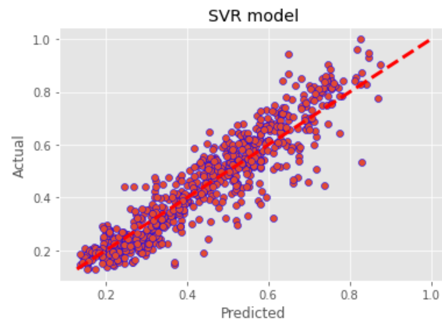

Fig. 10: Actual Vs Predicted observations for Support Vector Regression

electricity usage. In addition, as Table V illustrates, it produces lower MAE, MSE, and RMSE scores. It was observed that K Nearest Neighbours models give 74.16% accuracy according to R square.

The Evaluation results of K Nearest Neighbours model are as follows:

TABLE V: Prediction Errors and accuracy score of KNN

Prediction Errors and accuracy score of KNN	
Mean Absolute Error	0.0534
Mean Square Error	0.00612
Root Mean Square Error	0.0782
R square(%)	81.15%

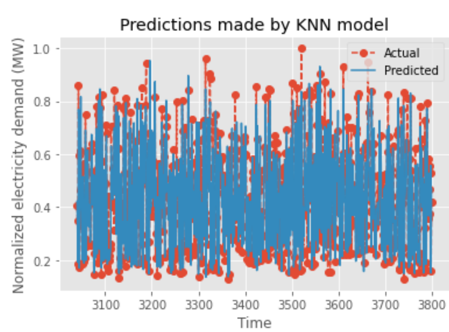

Fig. 11: Graphical representation of Actual vs predicted results

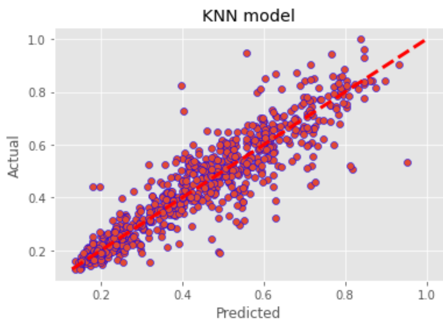

Fig. 12: Actual Vs Predicted observations for K Nearest Neighbours

V. Comparison and Result Summary

A comparison of various machine learning-based demand forecasting models is shown in Table VI. Future electricity demand can be predicted using these models. The findings show that, in comparison to earlier investigations, our approach provides significantly better predictions. In this work, many machine learning methods are employed. Table VI displays a thorough performance comparison of different algorithms. The outcomes obtained from the analysis represent that the Random Forest ML approach outperforms the other conventional ML approaches in terms of R square, MAE, MSE, and RMSE error rates since it has the highest accuracy (86.24%) according to R square and lowest MAE, MSE, and RMSE error rates (0.0450, 0.0042, and 0.0652, respectively) compared to other discussed conventional ML approaches.

The result presented in Table VI demonstrates the obtained accuracy and error values generated in the applied algorithms.

TABLE VI: Comparison among Various Algorithms

Algorithms	R square	MAE	MSE	RMSE
Linear Regression	82.53%	0.0582	0.0057	0.0761
Decision Tree	74.16%	0.0615	0.0081	0.0903
Random Forest	86.24%	0.0450	0.0042	0.0652
Support Vector Regression	84.52%	0.0540	0.0052	0.0724
K Nearest Neighbours	81.84%	0.0534	0.0061	0.0782

VI. Conclusion

Energy shortages could seriously impede Bangladesh's economic development given its rapidly expanding economy. Although electricity demand has been a hot topic for some time, there haven't been many in-depth studies done on the state of our energy production. Artificial intelligence and other contemporary computation techniques can undoubtedly produce useful results that can reduce the gap between supply and demand.

Overall, outcomes show that the proposed ML approach outperforms the conventional ML approaches in terms of R square, MAE, MSE, and RMSE error rates since it has the highest accuracy (86.24%) according to R square and the lowest MAE, MSE, and RMSE error rates (0.0450, 0.0042, and 0.0652, respectively) compared to other discussed conventional ML approaches. The RF method provides better performance due to the bagging algorithm, which is the foundation of Random Forest, which employs ensemble learning. Therefore, the RF method has the ability to predict the future demand for electricity more accurately and also be able to plan ahead for maintenance and load distribution using the obtained results, which can help electricity generation companies.

References

[1] S. Masum, Y. Liu, and J. Chiverton, "Multi-step time series forecasting of electric load using machine learning models," in *International conference on artificial intelligence and soft computing*. Springer, 2018, pp. 148–159.

[2] N. Phuangpornpitak and W. Prommee, "A study of load demand forecasting models in electric power system operation and planning," *GMSARN International Journal*, vol. 10, no. 2016, pp. 19–24, 2016.

[3] Z. Ãamurdan and M. C. Ganiz, "Machine learning based electricity demand forecasting," in *2017 International Conference on Computer Science and Engineering (UBMK)*, 2017, pp. 412–417.

[4] N. Shabbir, R. Ahmadiahangar, L. Kütt, and A. Rosin, "Comparison of machine learning based methods for residential load forecasting," in *2019 Electric power quality and supply reliability conference (PQ) & 2019 symposium on electrical engineering and mechatronics (SEEM)*. IEEE, 2019, pp. 1–4.

[5] A. R. Anik and S. Rahman, "Commercial energy demand forecasting in bangladesh," *Energies*, vol. 14, no. 19, p. 6394, 2021.

[6] Z. Liu, D. Wu, Y. Liu, Z. Han, L. Lun, J. Gao, G. Jin, and G. Cao, "Accuracy analyses and model comparison of machine learning adopted in building energy consumption prediction," *Energy Exploration & Exploitation*, vol. 37, no. 4, pp. 1426–1451, 2019.

[7] K. Gajowniczek and T. Zabkowski, "Two-stage electricity demand modeling using machine learning algorithms," *Energies*, vol. 10, no. 10, p. 1547, 2017.

[8] P. K. Jain, W. Quamer, and R. Pamula, "Electricity consumption forecasting using time series analysis," in *International Conference on Advances in Computing and Data Sciences*. Springer, 2018, pp. 327–335.

[9] C. Behm, L. Nolting, and A. Praktiknjo, "Forecasting long-term electricity demand time series using artificial neural networks," 2020.

[10] M. R. Islam, A. Al Mamun, M. Sohel, M. L. Hossain, and M. M. Uddin, "Lstm-based electrical load forecasting for chattogram city of bangladesh," in *2020 International Conference on Emerging Smart Computing and Informatics (ESCI)*. IEEE, 2020, pp. 188–192.

[11] A. T. Eseye, M. Lehtonen, T. Tukia, S. Uimonen, and R. J. Millar, "Machine learning based integrated feature selection approach for improved electricity demand forecasting in decentralized energy systems," *IEEE Access*, vol. 7, pp. 91 463–91 475, 2019.

[12] M. R. Patel, R. B. Patel, and N. A. Patel, "Electrical energy demand forecasting using time series approach," 2020.

[13] F. Rabbi, S. U. Tareq, M. M. Islam, M. A. Chowdhury, and M. A. Kashem, "A multivariate time series approach for forecasting of electricity demand in bangladesh using arimax model," in *2020 2nd International Conference on Sustainable Technologies for Industry 4.0 (STI)*. IEEE, 2020, pp. 1–5.

[14] A. W. Yao, S. Chi, and J. Chen, "An improved grey-based approach for electricity demand forecasting," *Electric Power Systems Research*, vol. 67, no. 3, pp. 217–224, 2003.

[15] B. N. Mahmud, Z. Ferdoush, and L. T. Mim, "Modelling and forecasting energy demand of bangladesh using ai based algorithms," Ph.D. dissertation, Brac University, 2019.

[16] A. Istiaque, S. I. Khan *et al.*, "Impact of ambient temperature on electricity demand of dhaka city of bangladesh," *Energy and Power Engineering*, vol. 10, no. 07, p. 319, 2018.