IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received October 12, 2020, accepted October 22, 2020, date of publication October 27, 2020, date of current version November 10, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3034101

Genetic Algorithm Based Optimized Feature Engineering and Hybrid Machine Learning for Effective Energy Consumption Prediction

PRINCE WAQAS KHAN AND YUNG-CHEOL BYUN
Department of Computer Engineering, Jeju National University, Jeju-si 63243, South Korea

Corresponding author: Yung-Cheol Byun (ycb@jejunu.ac.kr)

This work was supported by the Ministry of Education and National Research Foundation of Korea, Leaders in INdustry-university Cooperation + Project.

ABSTRACT Smart grids are developing rapidly, leading to the need for accurate forecasts of power consumption. However, developing a precise time series model for energy forecasting is difficult. It has to be trained using optimal meteorological features such as temperature and time lags to qualify for a beneficial model. We have proposed an approach that uses an ensemble machine learning model based on XGBoost, support vector regressor (SVR), and K-nearest neighbors (KNN) regressor algorithms. We have also used the genetic algorithm (GA) to predict total load consumption from optimal feature selection. Using Jeju island's electricity consumption data as a case study shows that the proposed ensemble model optimized with GA is more accurate than the individual machine learning models. Using only the best-selected weather and time features, the proposed model records all the features of a complicated time series and shows a reduction in the mean absolute percentage error (MAPE) and the root mean square log error for the week ahead forecasts. We got 3.35 % MAPE of the three months test data by applying the proposed model. The smart grids operators can manage resources effectively to provide excellent services to the consumers based on the recommended model outcomes.

INDEX TERMS Energy forecasting, ensemble model, feature engineering, genetic algorithm, K-nearest neighbors, meteorological features, power consumption, smart grids, support vector regressor, XGBoost.

I. INTRODUCTION

Energy is essential for national development from a social, economic, and environmental point of view. It has a significant impact on industry and agricultural products, health and hygiene, population, education, and human life quality. Various energy sectors require time-ahead energy forecasting systems. Energy load forecasting is an integral part of the energy planning industry [1]. Smart grids prove helpful in addressing the power grid challenges such as reliability, economics, safety, greenhouse gas, and carbon emissions [2]. Power systems enable companies to make various decisions related to power systems, delivery planning, product storage, safety measures, demand-side management, and financial planning. Prediction is particularly crucial for the future operation of the smart grids. Weak ahead demand forecasts help the smart grids to manage the future supply better. Traditionally, statistical and engineering methods have been used to predict future demand using tables and maps. These traditional methods mainly take into account the influence of the weather and the calendar. These features are currently being used to develop forecasting models in new ways. Load forecasting techniques can be divided into short, medium, and long-term categories [3]. Estimated short-term forecasting consists of Hourly, daily, or weekly forecasts; The medium-term forecasting consists of months to one year, and long-term forecasting range from one year to ten years. It is worth noting that it is an effective way to estimate the load that will significantly impact any power system's financial performance. Most decisions can be made based on the calculated results. Because the relationship between many parameters is complex and unstable, electric forecasting can be separated depending on weather conditions and forward load structure. Various machine learning techniques have been

The associate editor coordinating the review of this manuscript and approving it for publication was Sanjeevikumar Padmanaban.

proposed using algorithms of varying qualities to predict power loads [4]. Due to the rapid economic growth driven by population growth and the steady increase in electricity demand in large cities, power forecasting plays an essential role in monthly time measurements. Weather is one of the main features for power demand, so it is now common for models to include climate changes, for example, temperature, humidity, wind speed, and cloudiness [5]. It is imperative to decide which weather feature can be conceived. There are five weather stations on Jeju Island, and each weather station takes readings on an hourly or sub-hourly basis.

As a case study, we collected time-series data of energy consumption on Jeju island. Jeju island is the largest island in Korea. Jeju energy corporation (JEC) aims to stop energy conveyed from the Korean mainland and consume only renewable energy to satisfy all electricity needs. Renewable energy sources are cheaper as compared to traditional sources of electricity. Hence they are replacing nonrenewable sources worldwide [6]. JEC plans to include renewable energy technology to replace fossil fuel generators and nonrenewable energy production on the island. The objective of the JEC is to achieve this in three significant steps. The first step is converting the Gapa island of Jeju into a test laboratory, making it the first carbon-free island. The following is to increase the share of renewable energy supply by 50 percent by 2020, and the last is to transform Jeju island into a carbon-free city by 2030 [7]. We also collected weather data for five different weather stations on Jeju island. We use the latest machine learning algorithms to suggest energy forecasting systems. We simulated a forecast model based on actual data on energy use and weather data on the island of Jeju, South Korea. A hybrid machine learning algorithm is proposed. We use three advanced algorithms XGBoost, support vector regressor (SVR), and k-nearest neighbors (KNN) regressor algorithms. These models are taught and trained based on optimized features. We have used a genetic algorithm (GA) to obtain optimal meteorological characteristics such as temperature, wind speed, rain, humidity, and time lags. Their performance is better than these simple models. The contributions of this paper are threefold:

- performed exploratory data analysis of weather and load consumption data;
- proposed an approach that uses an ensemble machine learning model based on XGBoost, SVR, and KNN;
- used the genetic algorithm for optimal feature selection to predict total load consumption;
- compared the proposed model with different prediction algorithms.

The rest of the article is arranged as follows. Section 2 introduces related publications and articles. Section 3 introduces the proposed hybrid model, genetic algorithm, and introduces the data collection process. Chapter 4 presents the process of energy consumption prediction based on machine learning; it also performs preprocessing, feature engineering, and training. Chapter 5 presents the performance results of the proposed model evaluated using Jeju energy consumption data. It also compares the results with the current model. Finally, we conclude in the last section.

II. RELATED WORKS

The research in forecasting and prediction is extensive. Various forecasting models were conceived and used to answer the issues posed by the researchers. Electric load estimation is an integral part of the power network. Many countries are opening electricity markets, increasing participation of different entities, creating competitive environments, and reducing costs. In the electricity market, when calculating the electricity load and the cost of shortening the electricity market. Load predictions are becoming more important today, but traditional algorithms' performance is not robust and not acceptable.

Deng *et al.* [8] presented a multi-scale convolutional neural network (CNN) with load prediction and timing cognition at several stages. In the case of multi-scale convolution, CNN's capacity increases with access to the load channel's complex and essential features. Further, they propose a new framework that uses probabilistic distribution in the data, for example, to find relevant properties for better results. The latest models observed that the proposed model could achieve more accurate results and show excellent stability in multi-step point and probabilistic forecasting. However, the proposed methodology performance is not satisfactory in the first step for direct multi-step prediction, and also the network structure's execution is a bit complicated. Wind energy forecasts are a very effective way of solving wind energy problems due to wind energy fluctuation and volatility. Shi *et al.* [9] have introduced hybrid models to independently use each module to predict wind power using grey relational analysis and wind speed distribution features. Each module is measured based on the different wind speed and the same wind speed frequency. The case study shows various applications with a short term of the hybrid predictive model. However, if the wind power output is lower than ten megawatts, the proposed model's performance is inferior to the individual models. Ilbeigi *et al.* [10] performed research to reduce energy consumption in Iran. They trained and employed numerous artificial neural network (ANN) models, assessed the power required by the office, evaluated the impact of the energy factor, and performed a comprehensive sensitivity analysis to find the most effective model. Besides, the building's energy needs were analyzed by the grasshopper and Energy Plus calculation engine. Later, an artificial neural network was modeled using the Levenberg-Marquardt distribution medium to predict the most efficient building parameters based on conventional solar energy. Besides, two different sensitivity analyses were performed to assess the effect of significant pillars on energy intake. Finally, the energy prediction was improved using genetic algorithm research. The results demonstrated that the average energy consumption decreased by 35% by optimizing energy use in the Genetic Algorithm case study. However, considering other parameters governing energy consumption, new design parameters can be included and extended to the stated objectives.

Kaur and Ahuja [11] use the Autoregressive integrated moving average (ARIMA) algorithm to estimate healthcare organizations' electricity consumption. The historical energy consumption dataset from 2005 to 2016 was obtained using Apollo Hospital in India. When choosing the correct estimate, many accuracy tools measure forecasts and statistics, such as root mean squared error and mean percentage error, are used. They train the model for monthly, two monthly, and quarterly electricity consumption forecasting. The results show that ARIMA is the most appropriate model for predicting month range. ANN, particle swarm optimization (PSO), and ARIMA were used by Atienza *et al.* [12] to forecast and determine electricity consumption in the Philippines. They use historical data on electricity consumption from 2008 to 2016 for the prediction of 2017 to 2020. The result revealed that the use of PSO-ANN and ARIMA models yielded the highest accuracy rate than of BP-ANN and ARIMA being tested. However, a more efficient method than grid search for obtaining the hyperparameters, and a different ANN architecture for high consistent prediction accuracy can be used. Short-term Load Forecasting for complex loads amidst hierarchical layers has been analyzed by Fan *et al.* [13]. Some regional forecasting systems are in place for segment analysis to find the region's optimal partitioning according to weather and electricity load values. Therefore, the prediction accuracy is improved. However, the authors have not mentioned the feature selection procedure. Besides, algorithms for collecting information generated by the smart meter in various categories from traditional bottom-up or top-down approaches methods have been proposed [14]. Decomposition based on a load value is also in practice for predicting the day-ahead intermediate and peak loads. They cluster the energy data, applying the k-means approach to ascertain the daily base, medium, and peak loads. Then, a neural network is utilized as a load forecasting system [15]. In the study of Clements *et al.* [16], an algorithm that proposes a method according to each prediction time using various filtering coefficients was implemented. However, the linear prediction model cannot consider the uncertainty of energy shortage and other factors [17]. Therefore, several nonlinear models have been studied over the past decade, such as artificial neural networks [18], Gaussian process regression [19], recurrent neural networks (RNNs) [20], and LSTM [21] have been studied over the past decade to accommodate the nonlinearity of data in a hybrid forecasting model.

Hong *et al.* [22] examined the spatial relationship between different types of appliances to predict individual occupants' short-term strength requirements. An effective short-term Residential prediction framework provides data collection model preprocessing data models, training modules, and load prediction models. Within this framework, several time-series have been run and have accounted for the motion of electricity and certain spatiotemporal relationships. For predicting user usage patterns, a learning method based on deep learning and ResBlocks is proposed. The grid search method is used as the hyperparameter correction network. The proposed method and current prediction techniques are evaluated based on real data sets. The results show that the proposed criteria are better than other comparison methods. However, the correlation described in communication networks can be used to further enhanced the results. The document by Haq and Ni [3] proposed a newer hybrid model for short-term prediction. They decomposed load demand time series by improved empirical mode decomposition. Then they improved the accuracy of the load times during peak periods by increasing the correlation analysis between the system load and any input variables. Energy load data from the Australian and Texas energy markets are used to influence unusual patterns. Thus, the proposed model can produce more robust and more stable, accurate predictions.

Liao *et al.* [23] reported in the paper a method of estimating a similar day short load by XGBoost. They identify the most critical factors affecting transportation and prepare the necessary feature map to determine the exact days. Big data and XGBoost process training techniques were used to estimate the load forecasting. This process is accurate and can improve the accuracy of predictions. However, authors have just focused and targeted similar days for short-term load forecasting. Ceperic *et al.* [24] proposed estimating the load forecasting by support vector regression. To ensure the proposed algorithm's effectiveness, the model has been trained and tested on publicly available datasets. Improvements in the planning strategy can be considered that the efficacy of SVR-based load forecasting depends on the number of available models. Ashfaq and Javaid [25] proposed the use of Support Vector Machine (SVM) and K-Nearest Neighbor (KNN) for load and price forecasting. However, the authors have completely ignored the other factors, such as temperature, holidays, and weekends. Advanced algorithms can be used to choose proper input features for energy load prediction. We have proposed to use the genetic algorithm for optimal feature selection. To obtain better accuracy, we have proposed to use a hybrid algorithm.

III. METHODOLOGY

Machine learning has been extensively used in the energy industry to estimate energy consumption. Machine learning algorithms choose the historical energy consumption data for training of the model [26]. It develops the necessary procedures in the network and uses a particular training system. The accurate forecast also depends on the choice of factors. Figure 1 shows the flow diagram of the designed model. The data on hourly load consumption and several metrological factors are collected from different power companies. Weather readings are also taken on an hourly basis from five different weather stations. Metrological factors such as temperature, wind speed, humidity have a remarkable effect on accurate forecasts. Exploratory data analysis is performed on this data. We analyzed the consumption patterns and their variation with respect to different weather features.

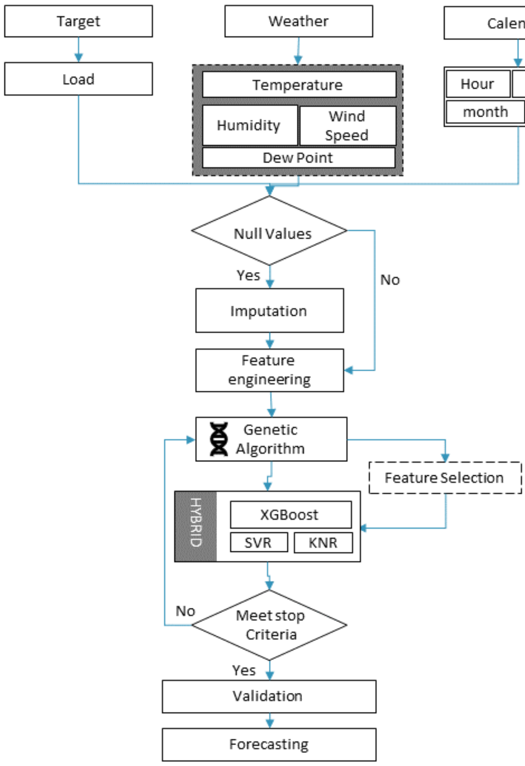


FIGURE 1. Flow diagram of the proposed forecasting model.

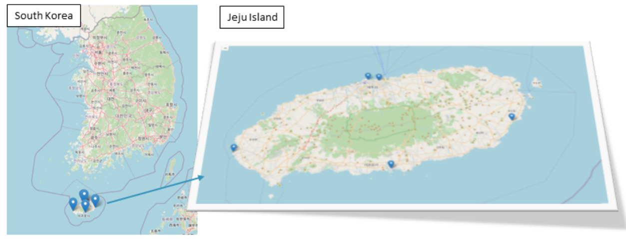


FIGURE 2. Weather stations location on Jeju island, South Korea.

TABLE 1. Weather station information.

#	Station ID	Latitude	Longitude
1	182	33.5167	126.5
2	184	33.5141	126.5297
3	185	33.2938	126.1628
4	188	33.3868	126.8802
5	189	33.2461	126.5653

Null values are imputed, and then a genetic algorithm is applied for the selection of optimal features. The proposed hybrid model consists of three state of the art machine learning algorithm, XGBoost, SVR, and KNR. After training the model, we tested their performance using different measurement methods. We also analyzed the output on different forecasting parameters, such as 48 hours, week-ahead, and two months forecasting.

A. DATA ACQUISITION

As a case study, we have collected Jeju island's energy and weather data. Jeju Energy Corporation is the energy distributor on Jeju Island. Electricity is sourced from two primary sources. The first one is Korea Electric Power Corporation, and the other is the Korea Power Exchange. These organizations are responsible for power market processes, power systems, and real-time shipping to support government planning and policy-making efforts.

Korea Electric Power Corporation and Korea Power Exchange provide renewable energy sources. It has three primary sources. The first one is small-sized solar energy generation without any contract called behind the meter. The second source is the photovoltaic generated solar energy, and the third renewable energy source is wind power energy. The Korea Power Exchange provides energy sources from non-renewable sources, such as fossil fuel-based power sources. Jeju island has five weather stations to cover all sides of the island, as shown in Figure 2. The weather data covers a variety of metrological factors. These factors include temperature, wind direction, wind speed, humidity, precipitation, snowfall, insolation, ground temperature, sea level pressure, local air pressure, dew point temp, and steam pressure. Table 1 shows the station ID, Latitude, and Longitude information of these five weather stations, and Table 2 summarizes the properties of the dataset.

B. ARCHITECTURE OF PROPOSED MODEL

Each machine learning model has its pros and cons. The primary purpose of machine learning is to train a stable model that performs well in all respects. [27]. We propose an ensemble model consisting of XGBoost, SVR, and KNR. Figure 3 describes the structure of the proposed hybrid model. This model requires several input parameters, including the energy source combined with the weather parameters. After clearing the data, the feature configuration is done. Following preprocessing, the function is transferred to the genetic algorithm. GA has performed several tasks, including fitness, selection, reproduction, and mutation. The optimal features are then passed to the hybrid model. The model is validated using test data. Different evaluation metrics are also used to measure the accuracy of predictions.

C. XGBoost

XGBoost algorithm is a supervised machine learning algorithm. XGBoost uses a collection of predictors which come together to answer. It uses boosting rather than bagging technique. In boosting, the predictors are made sequentially, not independently. XGBoost is a gradient boosting algorithm. It is also known as the 'regularized boosting 'technique. It allows cross-validation at every repetition of the boosting process, and thus it is easy to get the exact optimum number of boosting iterations in a single run [28]. XGBoost is a high-performance implementation of the gradient boosting framework. In our case, if we have k regression trees, where each

TABLE 2. Energy Load with weather data summary.

Feature	Abbreviation	Count	Null	Mean	Min	Max
Energy Load	PVALUE	29088	0	644.7302	415.015	969.0557
Temp	TA	29088	0	15.76799	-2.35	33.475
Wind direction	WD	29088	0	198.8888	-99.9	360
Wind Speed	WS	29088	0	3.490086	-99.9	15.375
Humidity	HM	29088	0	73.19078	14	100
Precipitation	RN	29088	0	-4.54265	-99.9	54.3
Snow fall	DSNW	10326	18762	0.155263	0	19
Insolation	ICSR	17918	11170	0.560861	0	4
Ground temperature	TS	29088	0	18.23386	-0.75	56.5
Sea level pressure	PS	29088	0	1016.164	-99.9	1036.75
Local Air Pressure	PA	29088	0	1011.752	-99.9	1032.7
Dew point temp	TD	29088	0	10.66538	-10	28.25
Steam pressure	PV	29088	0	16.36473	3	38.25

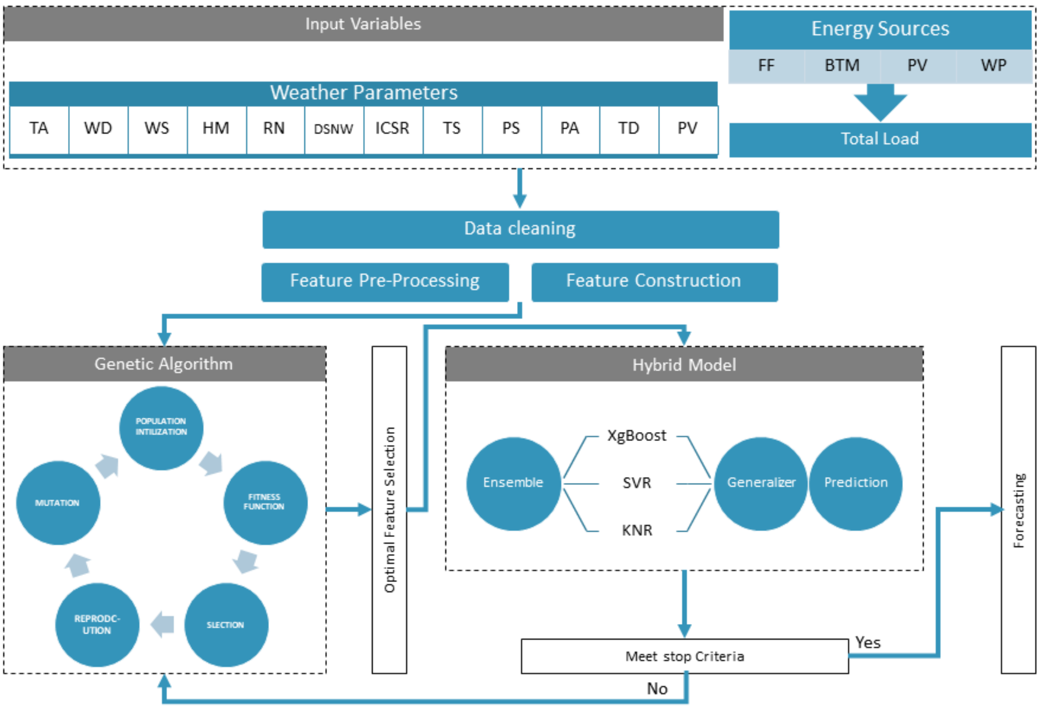


FIGURE 3. Structure of the proposed hybrid model.

tree optimize the previous one, we can define the predicted target $\hat{y}$ as explained in Equation 1

$$\hat{y}_i = \sum_{k=1}^{K} f_k(x_i), f_k \epsilon F \tag{1}$$

We have to minimize the objective function explained in Equation 2. This function contains loss and regularization. Where l explains the loss function, and Ω represents the regularization function.

$$obj(\theta) = \sum_{i}^{n} l(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k) \tag{2}$$

D. SUPPORT VECTOR REGRESSOR

Support vector regression (SVR) belongs to the class of support vector machines(SVM) [29]. SVM has gained much consideration in pattern recognition and regression. The most crucial feature of SVM is that it increases the learning machine's overall capacity according to the principle of structural risk mitigation. SVM's training is about solving integrated programming, so more and more solutions need to be found. Using a slope support vector machine, we often refer to support vector regression, and some researchers have proposed SVR for the problem of electric charge prediction. SVR is a regression algorithm, so using SVR, researchers can work with static values instead of classifying. The support vector is a data point near the border. The distance between the two points is the smallest or minimum. Simple regression tries to reduce the error rate while SVR tries to fit the error within a particular threshold [30].

E. K-NEAREST NEIGHBORS REGRESSOR

K-nearest neighbors (KNN) regressor is one of the most delicate classification systems that uses adjacent "K" to describe predicting data's value [31]. The k nearest neighbors are an incomprehensible and unobtrusive algorithm that stores all known training data information and uses it to predict outcomes based on equations or distance functions. The KNN and the system can use each precautionary measure as input to develop analytical methods that describe future conditions. When KNN is used as a shift, predictions are made by calculating the output as one closest neighbor to the input. K shows the number of neighbors used for prediction. If k is greater than one output, it can be computed or similar to the average of all k nearest neighbors' results with the weight of one.

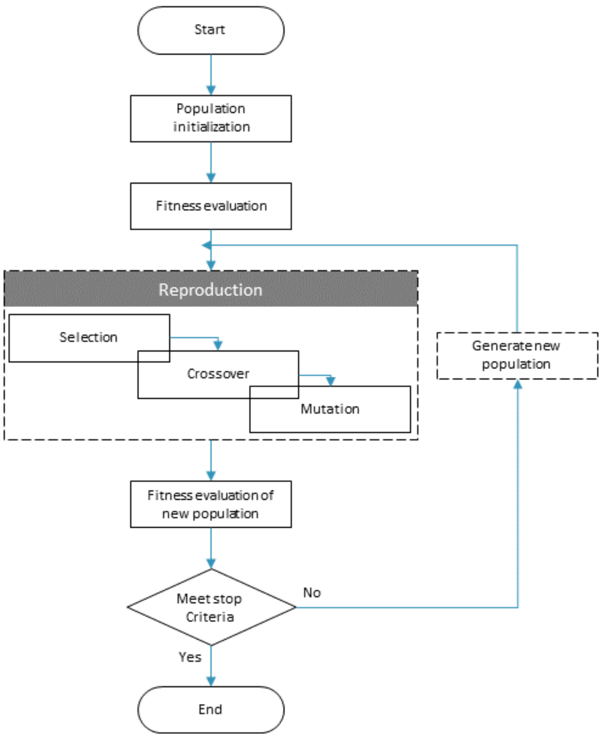


FIGURE 4. **Flow diagram for genetic algorithm.**

F. GENETIC ALGORITHM

A genetic algorithm is used for the optimal feature selection. Figure 4 shows the flow diagram for the genetic algorithm. It starts with the initialization of the population, and then it evaluates the fitness function [32]. The reproduction stage consists of three steps, selection, crossover, and mutation. In every repetition of the GA, the most suitable chromosomes generate new individuals. Those individuals or recent parameters give the foundation for the succeeding generation [33]. Following reproduction, GA evaluates the new population. If the new population does not meet the stopping criteria, it goes back to the reproduction stage and generates a new population; otherwise, it stops and gives the optimal chromosome combination.

IV. FORECASTING

This section covers the three main parts of the proposed load estimation method. The first is exploratory data analysis, the second is feature engineering, and the last is the ensembled model's training. The proposed hybrid method uses a genetic algorithm to determine the best parameters. The data used for estimation are time-series data, and the estimate is based on Equation 3.

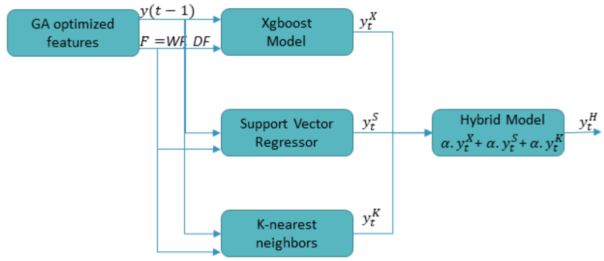


FIGURE 5. **Schematic diagram of hybrid regressor.**

$$y(t+1) = f(y(t), \ldots, y(t-m+1) : F) \tag{3}$$

$$F = WF, DF \tag{4}$$

where $y(t)$ is the vector representation of daily electricity load at time t. m is the order of a dynamical system, which is predetermined constant [34]. F is the set of weather and time features, as explained in Equation 4. These features are further explained in section IV-B.

$$y_t^{hybrid} = \alpha.E_t^{xgboost} + \alpha.E_t^{svr} + \alpha.E_t^{knn} \tag{5}$$

$$\alpha = min \sum_{t=1}^{M} (y_t - \hat{y}_t), 0 \leq \alpha \leq 1 \tag{6}$$

where y_t^{hybrid} is the final output of hybrid model and $alpha.E_t^{xgboost}$, $alpha.E_t^{svr}$, $alpha.E_t^{knn}$ represents the outputs of XGBoost, SVR and KNN models respectively. α is the weight coefficient of each model. Determining the weight coefficients for each model is the crucial step in constructing a hybrid prediction model. This can be achieved by solving optimization problems that minimize the absolute number of errors using Equation 6. Figure 5 depicts the schematic diagram of the hybrid regressor. It takes the GA-based optimized feature and forwards them to XGBoost, SVR, and KNN regressors. We have employed the ensemble technique to get the advantages of different algorithms. Ensemble machine learning enhances machine learning results through consolidating various models. The ensemble approach is a descriptive algorithm that merges multiple algorithms into a predictive model to reduce bias, deviation, and enhances prediction results. The output of these models is combined based on weight α and gives the final outcome y_t^h.

A. EXPLORATORY DATA ANALYSIS

For experimental purposes, the latest updated energy consumption and metrological data from Jeju Island have been collected from January 2017 until April 2020. Table 2 summarizes the properties of the dataset. It shows the function name, abbreviation, the count of each data feature, null values, mean, minimum, and maximum value. The data consists of hourly energy consumption and different weather observations such as temperature, wind direction, wind speed, humidity, precipitation, snowfall, insolation, ground temperature, sea level pressure, local air pressure, dew point temperature, and steam pressure. The data contains 29088 rows. This table shows that snowfall and insolation have a large number of null values. The average energy consumption throughout the observed dates is 644.73 MW. This energy load comes from both renewable and nonrenewable energy sources. The alliance of renewable power sources into the grid is advantageous for the atmosphere furthermore holds financial benefits [35]. Figure 6 shows the combined load of renewable and nonrenewable energy sources. This graph shows the overall trend of energy consumption on Jeju island without the impact of other weather features. We can observe a symmetrical pattern in energy consumption.

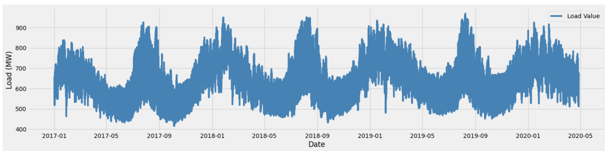


FIGURE 6. Combined load of renewable and nonrenewable energy sources.

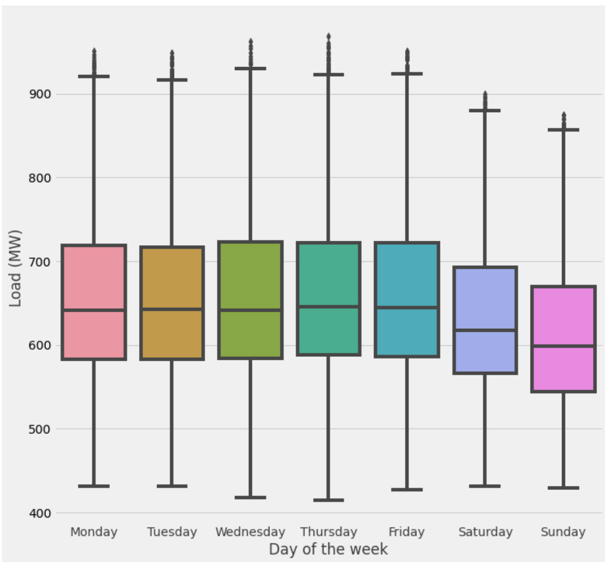


FIGURE 7. Box-plot for daily energy consumption.

Figure 7 shows the box-plot for total daily consumption. The X-axis shows the name of the weekday, and Y-axis represents the daily energy consumption. It can be observed clearly that days are essential when it comes to consumption. The lowest use is usually marked on weekends, as most commercial and industrial areas remain closed, leading to reduced overall consumption in the region.

To better understand the power consumed per hour, we made Figure 8. This graph shows the mean energy load distribution on an hourly basis. The X-axis indicates the number of hours; Y-axis represents the mean hourly

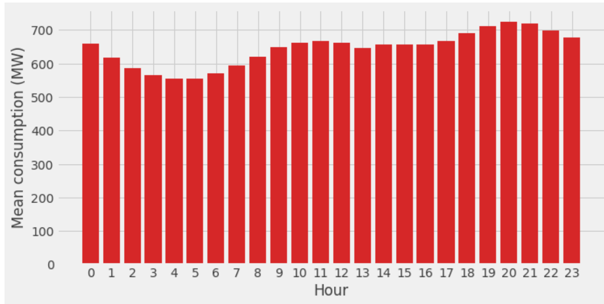


FIGURE 8. Mean hourly consumption.

energy consumption. On the hour graph, we can see that energy consumption is high between 19:00 and 22:00 and relatively low between 2:00 and 07:00.

Figure 9 shows the rolling mean of energy consumption compared to meteorological factors. The X-axis shows the recording date of the load and meteorological factor. Left Y-axis with the red label indicates the energy, and the right y-axis shows the meteorological feature. Subfigure 9(a) shows the rolling mean of energy compared to temperature. It is observed that a decrease in temperature leads to an increase in energy consumption. Subfigure 9(b) shows the rolling mean of energy compared to humidity. Subfigure 9(c) shows the rolling mean of energy compared to precipitation. According to a report by Korea's meteorological administration, most parts of Korea had below-normal rainfall in 2017 [36]. Hence, we can see a smooth trend in the year 2017. Subfigure 9(d) shows the rolling mean of energy compared to wind speed.

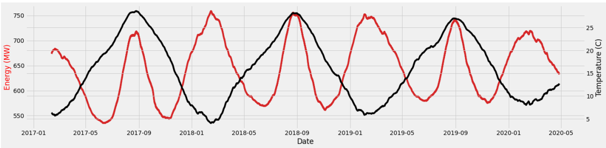
B. FEATURE ENGINEERING

The second step is the feature engineering. In this step, abstractions and critical features are selected and extracted from the prepared data, redundancy and irrelevant elements are removed. Feature selection is the process of choosing a subset of relevant and informative features for use in model building [37]. There are many advantages to creating predictive models using feature selection techniques. These techniques can improve the model's predictive accuracy and generalization ability by reducing the problem's size and preventing overfitting. It also provides a smaller feature set as input to the model, useful for building simpler models with shorter training times.

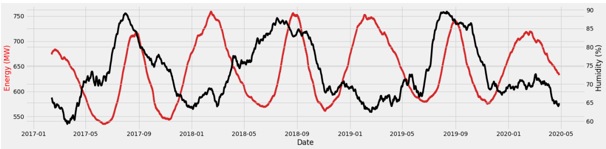
There are four different sources of energy for Jeju island. These sources consist of fossil fuel-based energy sources, behind the meter, photovoltaic, and wind power. These sources are combined using Equation 7. Where E_t^{total} is the total energy consumption with respect to time. E_t^{ff}, E_t^{btm}, E_t^{pv} and E_t^{wp} represents the fossil fuel-based energy sources, behind the meter, photovoltaic and wind power energy consumption with respect to time t.

$$E_t^{total} = E_t^{ff} + E_t^{btm} + E_t^{pv} + E_t^{wp} \tag{7}$$

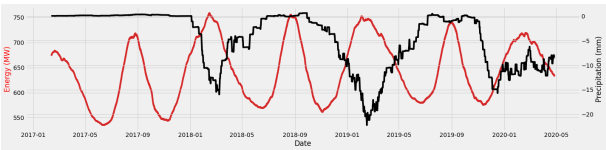
Equation 8 describes the weather features and Equation 9 expresses the date features with respect to time t.



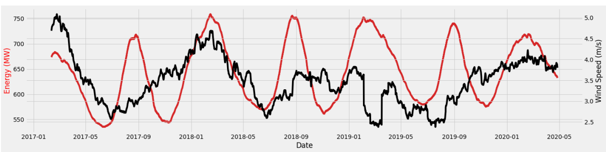
(a) Energy compared to temperature



(b) Energy compared to humidity



(c) Energy compared to precipitation



(d) Energy compared to wind speed

FIGURE 9. **Rolling mean of energy consumption compared to meteorological factors.**

Weather features consist of temperature TA_t, wind direction WD_t, wind speed WS_t, humidity HM_t, precipitation RN_t, snowfall $DSNW_t$, insolation $ICSR_t$, ground temperature TS_t, sea level pressure PS_t, local air pressure PA_t, dew point temperature TD_t, and steam pressure PV_t. Date feature consists of hour H_t, month M_t, year Y_t, quarter Q_t, and day of week DoW_t.

$$WF_t = \begin{Bmatrix} TA_t, WD_t, WS_t, HM_t, RN_t, DSNW_t, \\ ICSR_t, TS_t, PS_t, PA_t, TD_t, PV_t \end{Bmatrix} \tag{8}$$

$$DF_t = \{H_t, M_t, Y_t, Q_t, DoW_t\} \tag{9}$$

We collected weather data from five different weather stations and then calculated the average of the grouped data. The first step in estimating the average of grouped data is to define the midpoint of each interval. Then multiply these midpoints by the frequency of the corresponding class. The total of products divided by the aggregate of values is the mean value. The mean μ for each meteorological element can be obtained by dividing Σmf by the total number of stations N, where m is the midpoint of the category, f is the frequency. As a result, the Equation 10 can be written to summarize the steps used to determine the mean of a weather station's data.

$$\mu = \frac{\Sigma mf}{N} \tag{10}$$

A genetic algorithm is used for the optimal feature selection [38]. The pseudo-code for the genetic algorithm is expressed in Algorithm 1. It starts with the initialization of

Algorithm 1 Pseudo-Code for Genetic Algorithm

```
initialize population p(t)
evaluate p(t)
while Until stopping criteria do
    for each chromosome do
        crossover c(t) from p(t)
        compute fitness()
        select p(t+1) from p(t) and c(t)
        if chromosome available then
            mutate p(t)
        end if
        output best and stop
    end for
end while
```

population $p(t)$ with respect to time t. The population is then evaluated for each feature. For every set of a chromosome, GA performs crossover $c(t)$, compute fitness, and mutate. It continues to do so until it gets the best combination of the chromosome.

Figure 10 displays the chromosome representation of the feature. Where the features with 1 bit are considered for further training purposes, and features with 0 or False bit are not considered.

C. TRAINING AND TESTING PHASE

The preprocessed data is used to define the training model. The preprocessed data set is divided into two parts: the test group and the training group. Training data consist of January 1, 2017, to January 14, 2020, and test data include from January 15, 2020, to April 27, 2020. Typically, 70% of training and 30% of the testing data is used in practice, but we have used 36 months for training and three months for testing instead of 27 months for training and 12 months of testing. The main reason for leaving less data for testing is to get a more practical energy consumption prediction by employing more training data. There are several typical training parameters. When combining hyperparameters, the root mean square error (RMSE) is defined as the predictive metric. RMSE is the square root of the deviation expressed in the Equation 11, where y_t is the real value, and $\hat{y}_t$ is the estimated value. This statistical parameter is also called the standard deviation of the regression system [39]. A low RMSE value indicates that the model is well trained. For training purposes, the loss function is set to RMSE.

$$RMSE = \sqrt{\frac{1}{M}\sum_{t=1}^{M}(y_t - \hat{y}_t)^2} \times 100 \tag{11}$$

V. EXPERIMENTAL RESULTS AND DISCUSSION

This section covers feature importance analysis, forecasting results, and model evaluation indicators. The proposed model is also compared with other existing models.

TA	WD	WS	HM	RN	TS	PS	PA	TD	PV	H	DoW	Q	Y	DoY	DoM	WoY
1	0	0	1	0	1	0	1	1	1	1	0	0	1	1	1	1
TRUE	FALSE	FALSE	TRUE	FALSE	TRUE	FALSE	TRUE	TRUE	TRUE	TRUE	FALSE	TRUE	TRUE	TRUE	TRUE	TRUE

FIGURE 10. Chromosome representation of features.

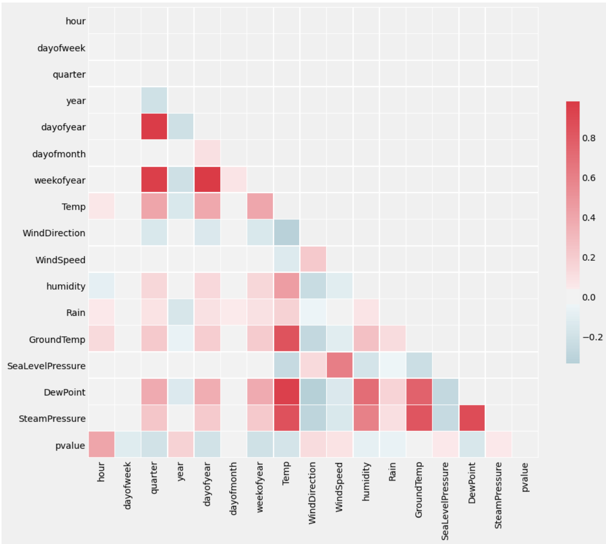


FIGURE 11. Correlation graph.

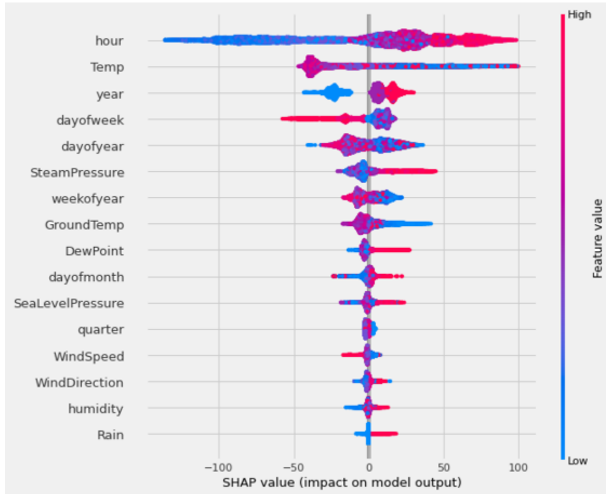


FIGURE 12. SHapley Additive explanations (SHAP) graph.

A. FEATURE IMPORTANCE ANALYSIS

This section covers the correlation diagram, Shapley, feature importance diagram, and feature selection based on genetic algorithms for feature significance analysis. The correlation matrix is used in an important way for the measured values obtained from the interval scale. The linear relationship between two continuous variables can be evaluated through correlation. Correlation can be calculated using the Equation 12.

$$\rho_{x,y} = \frac{Cov(x,y)}{\sigma_x \sigma_y} = \frac{\sum(x-\mu_x)(x-\mu_y)}{\sqrt{\sum(x-\mu_x)^2 \sum(y-\mu_y)^2}} \tag{12}$$

where $rho_{x,y}$ is the correlation coefficient between variable x and variable y. σ_x, σ_y are standard deviation of variable x and variable y resppectivley. μ_x, μ_y represents the mean of variable x and variable y. Figure 11 shows the correlation graph.

Shapley (SHAP) is a different approach to examine the significance of features. The purpose of SHAP is to estimate the participation of individual prediction function and to interpret the predictions [40]. The graphics summarized in Figure 12 shows the features' importance and influence score. Each spot on the graph has a particular value. The rating feature names determined in the y-axis, and while the x-axis depicts the Shapley value. Colors designate independent values from bottom to top. These features are listed as critical. The horizontal position symbolizes that the value is more prominent or less than the expected value. The red color indicates that the variable value is high, and the blue color shows that the importance is low. The ''hour'' feature's performance is very decisive; as in red, it is positive in the x-axis. Humidity and rain have an insignificant effect on training data.

Figure 13 shows the bar graph of feature importance. Feature scaling is imperative for forecasting models. To perform feature scaling, it is necessary to calculate the number of times each function is distributed over the boosting trees. Then display the results as a bar graph and sort the features according to their importance status. XGBoost calculates feature significance based on the effect of component value changes on average prediction difference. If the value is more significant, it means that it will have more influence on changing the expected value.

B. FORECASTING RESULTS

The test data consist of January 01, 2020, to April 04, 2020. The comparison of actual and forecasted load values of week ahead prediction is depicted in Figure.14. Subfigure 14(a) explains the week ahead prediction between 15 to January 22. Subfigure 14(b) displays the 168 hours prediction within 12 to 29th February. Subfigure 14(c) presents the 168 hours forecast among 18 to 25th March.

Figure 15 shows the actual and prediction graph of test data. This graph is for the whole test data. The X-axis represents the date, and Y-axis represents the load value.

Figure 16 shows the 48 hours forecasting. It consists of actual values, predicted values, and also the difference

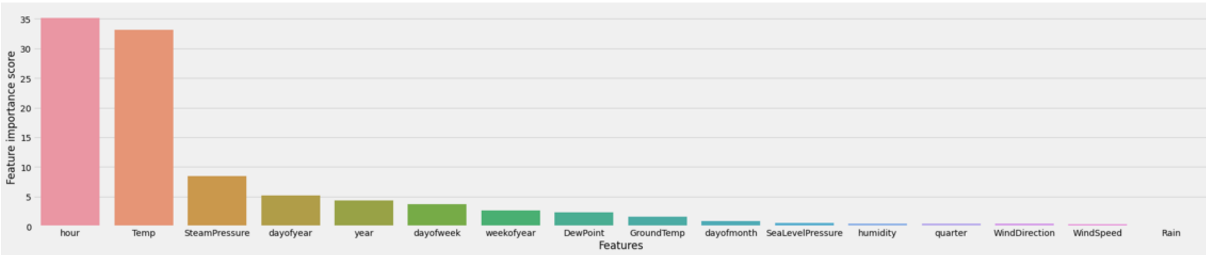
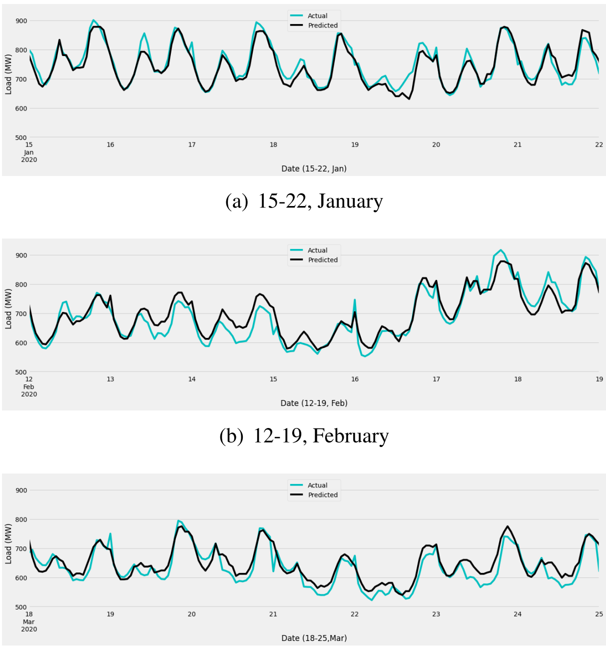


FIGURE 13. Feature importance graph.



(a) 15-22, January

(b) 12-19, February

(c) 18-25, March

FIGURE 14. Week-Ahead predictions using hybrid model.

between those values. Energy companies have different forecasting windows. We have tested the forecasting results based on the week ahead or 168 hours prediction.

C. MODEL EVALUATION INDICATORS

A comparison of the proposed model with the latest model is performed in this section. We compared the model with Lasso, Ridge, Gradient Boosting, XGBoost, Multilayer Perceptron (MLP) Regressor, and SVR.

Figure 17 shows a graphical representation for comparison with these models.Where the name of models are mentioned on X-axis, and the values are specified on Y-Axis. The root of the logarithmic means square error (RMSLE) is obtained by equation 13. Where M is the total number of data points, y_t is the actual value, and $\hat{y}_t$ is the forecasted value. RMSLE error is the logarithmic relationship between the model's actual data value and the forecasted value [41]. By applying the proposed model, we obtained 0.02 RMSLE, which is the lowest value compared to other models. Various evaluation metrics are used to validate the model's pros and cons, such as mean absolute error, mean square error, and root mean square error.

TABLE 3. Evaluation metrics.

Name	MAE	MSE	RMSE
Lasso	61.16	6087.15	78.02
Ridge	61.17	6085.77	78.01
GradientBoost	20.27	772.86	27.80
XGBoost	11.80	268.34	16.38
MLPRegressor	31.94	1709.08	41.34
LSTM	19.84	750.85	27.40
Hybrid [7]	20.21	769.04	27.73
Hybrid [42]	20.79	807.57	28.41
Proposed	10.05	192.32	13.87

$$RMSLE = \sqrt{\frac{1}{M}\sum_{t=1}^{M}\left(\log\left(y_t+1\right)-\log\left(\hat{y}_t+1\right)\right)^2} \tag{13}$$

We also chose the numerous advanced models for comparison with the proposed hybrid model. Besides benchmark models, we additionally evaluate the results with two other hybrid models. The first hybrid model is composed of catboost, Support vector regressor, and Multilayer perceptron [7]. The other hybrid model consists of XGBoost, Random Forst, and CatBoost models [42]. Table 3 shows the evaluation indicators used to examine the different model's performance with the proposed model. The mean square error (MSE) is the difference between the initial and predicted value [43]. It is extracted by squaring the mean squared error of the data set using Equation 14. The observed MSE for the proposed model is 192.32. The mean absolute error (MAE) describes the difference between the initial value and the forecasted value and is extracted as the mean absolute difference in the data set [44]. The MAE for the proposed model is 10.05, calculated using the Equation 15.

$$MSE = \frac{1}{M}\sum_{t=1}^{M}\left(y_t-\hat{y}_t\right)^2 \tag{14}$$

$$MAE = \frac{1}{M}\sum_{t=1}^{M}\left(y_t-\hat{y}_t\right) \tag{15}$$

Figure 18 shows the training graph of the genetic algorithm. It shows the improvement in the accuracy of

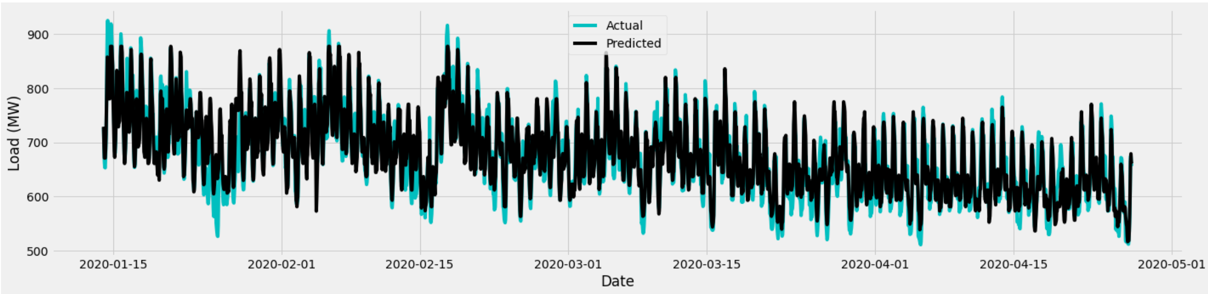


FIGURE 15. Comparison of actual and forecasted load values from test data.

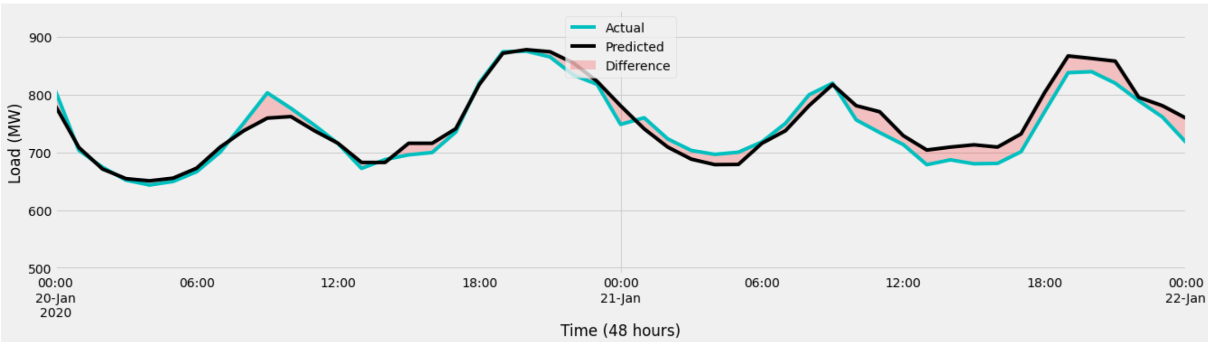


FIGURE 16. Forecasted load values for Operating 48 hours.

TABLE 4. MAPE and RMSLE comparison on week-ahead test data with other models.

Performance Metrics		Load Only Min %	Max %	Avg %	RMSLE	Load With Weather Min %	Max %	Avg %	RMSLE
Jan-20	Lasso	0.0588	27.3549	7.3452	0.1281	2.9222	46.7229	17.539	0.1185
	Ridge	0.2550	23.7445	6.2608	0.125	3.172	47.4651	17.975	0.1185
	Gradient Boost	0.0420	15.8999	4.1385	0.0591	0.0023	15.3365	3.7259	0.0425
	XGBoost	0.0700	11.4716	3.4616	0.0408	0.0501	12.9527	2.5427	0.0252
	MLP Regressor	0.0219	33.2032	13.1385	0.1335	0.0954	22.9120	7.2840	0.0750
	Hybrid	**0.0402**	**12.5415**	**4.6784**	**0.0613**	**0.0280**	**10.4438**	**2.3725**	**0.0194**
Feb-20	Lasso	0.0762	28.1682	9.0824	0.1281	0.0362	38.6889	9.9615	0.1185
	Ridge	0.2263	22.2846	10.3505	0.1255	0.0128	39.5296	10.2178	0.1185
	Gradient Boost	0.0055	25.1776	9.8775	0.0591	0.0195	14.0919	3.8520	0.0425
	XGBoost	0.0522	19.5277	10.4904	0.0408	0.0400	15.1055	5.0981	0.0252
	MLP Regressor	0.0248	36.6111	10.5004	0.1335	0.0410	25.3533	5.5538	0.0750
	Hybrid	**0.0886**	**26.0154**	**13.1052**	**0.0613**	**0.0186**	**12.3889**	**3.6291**	**0.0194**
Mar-20	Lasso	0.2777	24.7307	10.9014	0.1281	0.0465	37.1241	8.1921	0.1185
	Ridge	0.0893	25.1147	11.2858	0.1255	0.0074	37.8016	8.3334	0.1185
	Gradient Boost	0.0081	15.8904	6.5333	0.0591	0.1304	11.1739	4.7604	0.0425
	XGBoost	0.0825	17.9607	7.9463	0.0408	0.0125	15.0319	3.3507	0.0252
	MLP Regressor	0.0083	25.6280	7.4706	0.1335	0.0216	24.0328	4.9875	0.0750
	Hybrid	**0.3399**	**22.0757**	**10.2001**	**0.0613**	**0.0283**	**14.1376**	**3.7774**	**0.0194**

every generation. The blue line represents every generation's best score, and the red line shows every generation's average score.

For testing purposes, we have used three weeks ahead of slots. We chose a week from January, February, and March. We made two cases. First, we trained and tested a load of energy consumption data without weather features, and then we carried out training and testing by adding weather features. Mean absolute percent error (MAPE) is a measure of the prediction accuracy of a prediction [45]. It measures the size of the error, calculated using Equation 16 and expressed as a percentage. We got 3.35 % MAPE of the overall model by applying the proposed model.

$$MAPE = \frac{1}{M}\sum_{t=1}^{M}\left|\frac{y_t - \hat{y}_t}{y_t}\right| \times 100\% \tag{16}$$

Table 4 shows the comparison of the minimum, maximum, and average MAPE in the two cases. This table also indicates the RMSLE of each algorithm. It is evident from the

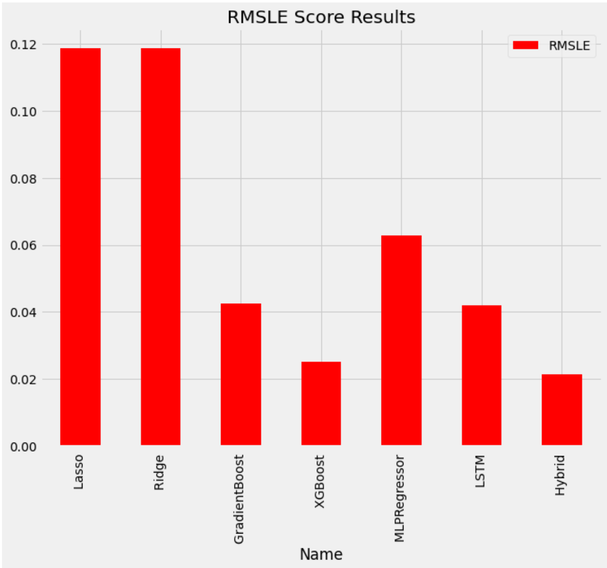


FIGURE 17. Comparison of RMSLE with different models.

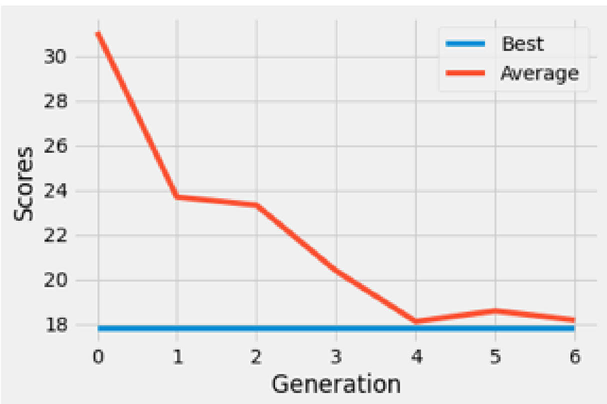


FIGURE 18. Genetic Algorithm.

correspondence that the recommended hybrid model works well contrasted with different existing models.

VI. CONCLUSION

This article presents a concise presentation to the energy consumption forecasting applying a genetic algorithm based optimized feature engineering and machine learning. It concentrates on the comparison of various load forecasting methods with the proposed method. It also focuses on optimal meteorological features such as temperature, wind speed, rain, humidity, and time lags to qualify for a beneficial model. This study has suggested an approach that employs an ensemble of machine learning models, namely XGBoost, support vector regressor, and k-nearest neighbor regressor algorithms. It turns out that individual forecasting models are limited in terms of performances. Therefore, combinations of prediction methods are gaining rising concentration. We have also used the genetic algorithm to predict total load consumption for optimal feature selection. We have obtained and use the Jeju islands' actual energy consumption and weather data for the experimental purpose. We performed exploratory data analysis, preprocessing, and train-test split before the training of the model. Moreover, we used various metrics to test the advantages of the proposed model: absolute mean error, absolute percent error, root mean square error, and log root error. We also selected the latest model for comparison with the proposed hybrid model. We got MAPE of 3.35 % for the three months test data by applying the proposed model. Electricity providers can effectively organize and manage supplies based on the prescribed model results to provide excellent customer services. In the future, this work can be extended by taking into account the other parameters such as the number of residents, electric vehicles, and tourists coming in different seasons of the year.

REFERENCES

[1] S. Fallah, M. Ganjkhani, S. Shamshirband, and K.-W. Chau, "Computational intelligence on short-term load forecasting: A methodological overview," *Energies*, vol. 12, no. 3, p. 393, Jan. 2019.

[2] A. Ghasempour and J. Lou, "Advanced metering infrastructure in smart grid: Requirements, challenges, architectures, technologies, and optimizations," in *Smart Grids: Emerging Technologies, Challenges and Future Directions*. Hauppauge, NY, USA: Nova Science, 2017.

[3] M. R. Haq and Z. Ni, "A new hybrid model for short-term electricity load forecasting," *IEEE Access*, vol. 7, pp. 125413–125423, 2019.

[4] A. A. Mamun, M. Sohel, N. Mohammad, M. S. Haque Sunny, D. R. Dipta, and E. Hossain, "A comprehensive review of the load forecasting techniques using single and hybrid predictive models," *IEEE Access*, vol. 8, pp. 134911–134939, 2020.

[5] I. Staffell and S. Pfenninger, "The increasing impact of weather on electricity supply and demand," *Energy*, vol. 145, pp. 65–78, Feb. 2018.

[6] E. Hossain, F. Shi, R. Bayindir, and J. Hossain, "Feasibility analysis: Evaluating sites for possible renewable energy options and their implications to minimize the cost," *Int. J. Elect. Eng. Educ.*, pp. 1–16, Jun. 2020.

[7] P. W. Khan, Y.-C. Byun, S.-J. Lee, D.-H. Kang, J.-Y. Kang, and H.-S. Park, "Machine learning-based approach to predict energy consumption of renewable and nonrenewable power sources," *Energies*, vol. 13, no. 18, p. 4870, Sep. 2020.

[8] Z. Deng, B. Wang, Y. Xu, T. Xu, C. Liu, and Z. Zhu, "Multi-scale convolutional neural network with time-cognition for multi-step short-term load forecasting," *IEEE Access*, vol. 7, pp. 88058–88071, 2019.

[9] J. Shi, Z. Ding, W.-J. Lee, Y. Yang, Y. Liu, and M. Zhang, "Hybrid forecasting model for very-short term wind power forecasting based on grey relational analysis and wind speed distribution features," *IEEE Trans. Smart Grid*, vol. 5, no. 1, pp. 521–526, Jan. 2014.

[10] M. Ilbeigi, M. Ghomeishi, and A. Dehghanbanadaki, "Prediction and optimization of energy consumption in an office building using artificial neural network and a genetic algorithm," *Sustain. Cities Soc.*, vol. 61, Oct. 2020, Art. no. 102325.

[11] H. Kaur and S. Ahuja, "Time series analysis and prediction of electricity consumption of health care institution using ARIMA model," in *Proc. 6th Int. Conf. Soft Comput. Problem Solving*. Patiala, India: Springer, 2017, pp. 347–358.

[12] N. A. C. Atienza, J. R. A. T. Jao, J. A. D. S. Angeles, E. L. T. Singzon, and D. D. Acula, "Prediction and visualization of electricity consumption in the philippines using artificial neural networks, particle swarm optimization, and autoregressive integrated moving average," in *Proc. 3rd Int. Conf. Comput. Commun. Syst. (ICCCS)*, Apr. 2018, pp. 135–138.

[13] S. Fan, K. Methaprayoon, and W.-J. Lee, "Multiregion load forecasting for system with large geographical area," *IEEE Trans. Ind. Appl.*, vol. 45, no. 4, pp. 1452–1459, Jul. 2009.

[14] S. B. Taieb, J. W. Taylor, and R. J. Hyndman, "Hierarchical probabilistic forecasting of electricity demand with smart meter data," *J. Amer. Stat. Assoc.*, pp. 1–17, Feb. 2020.

[15] L. C. P. Velasco, N. R. Estoperez, R. J. R. Jayson, C. J. T. Sabijon, and V. C. Sayles, "Day-ahead base, intermediate, and peak load forecasting using K-means and artificial neural networks," *Int. J. Adv. Comput. Sci. Appl.*, vol. 9, no. 2, pp. 62–67, 2017.

[16] A. E. Clements, A. S. Hurn, and Z. Li, "Forecasting day-ahead electricity load using a multiple equation time series approach," *Eur. J. Oper. Res.*, vol. 251, no. 2, pp. 522–530, Jun. 2016.

[17] J. Zheng, C. Xu, Z. Zhang, and X. Li, "Electric load forecasting in smart grids using Long-Short-Term-Memory based recurrent neural network," in *Proc. 51st Annu. Conf. Inf. Sci. Syst. (CISS)*, Mar. 2017, pp. 1–6.

[18] G. P. Zhang, "Time series forecasting using a hybrid ARIMA and neural network model," *Neurocomputing*, vol. 50, pp. 159–175, Jan. 2003.

[19] J. Song and E. Hwang, "Hybrid day-ahead load forecasting with atypical residue based Gaussian process regression," in *Proc. 9th Int. Conf. Future Energy Syst.*, Jun. 2018, pp. 631–634.

[20] Y. Shui-Ling and Z. Li, "Stock price prediction based on ARIMA-RNN combined model," DEStech Trans. Social Sci., Educ. Hum. Sci., Wuhan, China, Tech. Rep., 2017.

[21] H. Y. Kim and C. H. Won, "Forecasting the volatility of stock price index: A hybrid model integrating LSTM with multiple GARCH-type models," *Expert Syst. Appl.*, vol. 103, pp. 25–37, Aug. 2018.

[22] Y. Hong, Y. Zhou, Q. Li, W. Xu, and X. Zheng, "A deep learning method for short-term residential load forecasting in smart grid," *IEEE Access*, vol. 8, pp. 55785–55797, 2020.

[23] X. Liao, N. Cao, M. Li, and X. Kang, "Research on short-term load forecasting using XGBoost based on similar days," in *Proc. Int. Conf. Intell. Transp., Big Data Smart City (ICITBS)*, Jan. 2019, pp. 675–678.

[24] E. Ceperic, V. Ceperic, and A. Baric, "A strategy for short-term load forecasting by support vector regression machines," *IEEE Trans. Power Syst.*, vol. 28, no. 4, pp. 4356–4364, Nov. 2013.

[25] T. Ashfaq and N. Javaid, "Short-term electricity load and price forecasting using enhanced KNN," in *Proc. Int. Conf. Frontiers Inf. Technol. (FIT)*, Dec. 2019, pp. 266–2665.

[26] S. Walker, W. Khan, K. Katic, W. Maassen, and W. Zeiler, "Accuracy of different machine learning algorithms and added-value of predicting aggregated-level energy performance of commercial buildings," *Energy Buildings*, vol. 209, Feb. 2020, Art. no. 109705.

[27] J. Schmidt, M. R. G. Marques, S. Botti, and M. A. L. Marques, "Recent advances and applications of machine learning in solid-state materials science," *npj Comput. Mater.*, vol. 5, no. 1, pp. 1–36, Dec. 2019.

[28] Y. Jiang, G. Tong, H. Yin, and N. Xiong, "A pedestrian detection method based on genetic algorithm for optimize XGBoost training parameters," *IEEE Access*, vol. 7, pp. 118310–118321, 2019.

[29] L. Barolli and O. Terzo, "Complex, intelligent, and software intensive systems," in *Proc. 11th Int. Conf. Complex, Intell., Softw. Intensive Syst. (CISIS)*, vol. 611. Turin, Italy: Springer, 2017.

[30] T. Parhizkar, E. Rafieipour, and A. Parhizkar, "Evaluation and improvement of energy consumption prediction models using principal component analysis based feature reduction," *J. Cleaner Prod.*, vol. 279, Jan. 2021, Art. no. 123866.

[31] H. Gan, M. H. B. M. Khir, G. Witjaksono Bin Djaswadi, and N. Ramli, "A hybrid model based on constraint OSELM, adaptive weighted SRC and KNN for large-scale indoor localization," *IEEE Access*, vol. 7, pp. 6971–6989, 2019.

[32] I. Maries and E. Scarlat, "Computational intelligence techniques for communities network formation," in *Proc. IEEE Int. Conf. Grey Syst. Intell. Services*, Sep. 2011, pp. 887–893.

[33] S. A. G. Shirazi and M. B. Menhaj, "A new genetic based algorithm for channel assignment problems," in *Computational Intelligence, Theory and Applications*. Dortmund, Germany: Springer, 2006, pp. 85–91.

[34] S. Fan, J. Liao, K. Kaneko, and L. Chen, "An integrated machine learning model for day-ahead electricity price forecasting," in *Proc. IEEE PES Power Syst. Conf. Expo.*, 2006, pp. 1643–1649.

[35] E. Hossain, H. M. R. Faruque, M. S. H. Sunny, N. Mohammad, and N. Nawar, "A comprehensive review on energy storage systems: Types, comparison, current scenario, applications, barriers, and potential solutions, policies, and future prospects," *Energies*, vol. 13, no. 14, p. 3651, Jul. 2020.

[36] K. M. ADMINISTRATION. (2017). *Annual Report*. [Online]. Available: https://www.kma.go.kr

[37] G. Hafeez, K. S. Alimgeer, A. B. Qazi, I. Khan, M. Usman, F. A. Khan, and Z. Wadud, "A hybrid approach for energy consumption forecasting with a new feature engineering and optimization framework in smart grid," *IEEE Access*, vol. 8, pp. 96210–96226, 2020.

[38] Y. Kang and D. Zhang, "A hybrid genetic scheduling algorithm to heterogeneous distributed system," *Appl. Math.*, vol. 3, no. 7, pp. 750–754, Jul. 2012.

[39] A.-L. Schubert, D. Hagemann, A. Voss, and K. Bergmann, "Evaluating the model fit of diffusion models with the root mean square error of approximation," *J. Math. Psychol.*, vol. 77, pp. 29–45, Apr. 2017.

[40] P. Calleja and F. Llerena, "Consistency, weak fairness, and the shapley value," *Math. Social Sci.*, vol. 105, pp. 28–33, May 2020.

[41] Z. Zhang, W. Yang, and S. Wushour, "Traffic accident prediction based on LSTM-GBRT model," *J. Control Sci. Eng.*, vol. 2020, pp. 1–10, Mar. 2020.

[42] P. Waqas Khan, Y.-C. Byun, S.-J. Lee, and N. Park, "Machine learning based hybrid system for imputation and efficient energy demand forecasting," *Energies*, vol. 13, no. 11, p. 2681, May 2020.

[43] R. J. Bessa, V. Miranda, and J. Gama, "Entropy and correntropy against minimum square error in offline and online three-day ahead wind power forecasting," *IEEE Trans. Power Syst.*, vol. 24, no. 4, pp. 1657–1666, Nov. 2009.

[44] M. U. Yousuf, I. Al-Bahadly, and E. Avci, "Current perspective on the accuracy of deterministic wind speed and power forecasting," *IEEE Access*, vol. 7, pp. 159547–159564, 2019.

[45] H. Dong, Y. Gao, X. Meng, and Y. Fang, "A multifactorial short-term load forecasting model combined with periodic and non-periodic features-a case study of Qingdao, China," *IEEE Access*, vol. 8, pp. 67416–67425, 2020.

PRINCE WAQAS KHAN received the master's degree in computer science from the University of Agriculture, Faisalabad, Pakistan, in 2017. He is currently pursuing the Ph.D. degree with the Machine Learning Laboratory, Department of Computer Engineering, Jeju National University, Jeju, South Korea. He was a Lecturer with the Department of Computer Science, University of Agriculture. He also worked as a Researcher with Chongqing Key Laboratory of Cyberspace and Information Security, Chongqing University of Posts and Telecommunications, Chongqing, China. His research interests include machine learning, image processing, blockchain, and the Internet of Things.

YUNG-CHEOL BYUN received the B.S. degree from Jeju National University, in 1993, and the M.S. and Ph.D. degrees from Yonsei University, in 1995 and 2001, respectively. He worked as a Special Lecturer with SAMSUNG Electronics, in 2000 and 2001. From 2001 to 2003, he was a Senior Researcher with the Electronics and Telecommunications Research Institute (ETRI). He was an Assistant Professor with Jeju National University, in 2003. He is currently an Associate Professor with the Computer Engineering Department, Jeju National University. His research interests include AI and machine learning, pattern recognition, blockchain and deep learning-based applications, big data and knowledge discovery, time series data analysis and prediction, image processing and medical applications, and recommendation systems.