

Received August 12, 2021, accepted September 27, 2021, date of publication October 4, 2021, date of current version October 19, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3117951

Systematic Development of Short-Term Load Forecasting Models for the Electric Power Utilities: The Case of Pakistan

ANEQUE A. MIR^{1,2}, ZAFAR A. KHAN^{3,4}, ABDULLAH ALTMIMI⁵, MARIA BADAR¹,
KAFAIT ULLAH¹, MUHAMMAD IMRAN⁴, AND SYED ALI ABBAS KAZMI¹

¹U.S.-Pakistan Centers for Advanced Studies in Energy (USPCAS-E), National University of Sciences and Technology (NUST), Islamabad 44000, Pakistan

²Lahore Electric Supply Company, Lahore 54000, Pakistan

³Electrical Engineering Department, Mirpur University of Science and Technology, Mirpur 10250, Pakistan

⁴School of Engineering and Applied Science, Mechanical Engineering and Design, Aston University, Birmingham B4 7ET, U.K.

⁵Department of Electrical Engineering, College of Engineering, Majmaah University, Al-Majmaah 11952, Saudi Arabia

Corresponding authors: Aneque A. Mir (anequemireep18.ces@student.nust.edu.pk) and Abdullah Altmimi (a.altmimi@mu.edu.sa)

ABSTRACT Load forecasts are fundamental inputs for the reliable and resilient operation of a power system. Globally, researchers endeavor to improve the accuracy of their forecast models. However, lack of studies detailing standardized model development procedures remains a major issue. In this regard, this study advances the knowledge of the systematic development of short-term load forecast (STLF) models for electric power utilities. The proposed model has been developed by using hourly load (time series) of five years of an electric power utility in Pakistan. Following the investigation of previously developed load forecast models, this study addresses the challenges of STLF by utilizing multiple linear regression, bootstrap aggregated decision trees, and artificial neural networks (ANNs) as mutually competitive forecasting techniques. The study also highlights both rudimentary and advanced elements of data extraction, synthetic weather station development, and the use of elastic nets for feature space development to upscale its reproducibility at global level. Simulations showed the superior forecasting prowess of ANNs over other techniques in terms of mean absolute percentage error (MAPE), root mean squared error (RMSE) and R^2 score. Furthermore, an empirical approach has been taken to underline the effects of data recency, climatic events, power cuts, human activities, and public holidays on the model's overall performance. Further analysis of the results showed how climatic variations, causing floods and heavy rainfalls, could prove detrimental for a utility's ability to forecast its load demand in future.

INDEX TERMS Load forecast, artificial neural networks, multiple linear regression.

I. INTRODUCTION

Load forecasts are fundamental inputs for a reliable, smooth, and resilient operation of any power system. With the advent of smart grid technologies, forecast models are required to consider new elements such as demand response, demand-side management, and distributed energy resources to maintain desired reliability. Forecasting electricity demand is pivotal for maintaining the balance between electricity supply and demand. Therefore, the process of load forecasting introduces several factors which are responsible for influencing consumers' electricity consumption patterns. A thorough consideration of these factors into a forecast model results in reliable forecasts. However, the forecast model may also end

The associate editor coordinating the review of this manuscript and approving it for publication was Bin Zhou^{1b}.

up producing either an over forecast or an under forecast due to multiple reasons. These may include but are not limited to, use of less significant demand determinants (explanatory variables/features) during model development, an unsuitable forecasting technique, or bad quality of data. Since both an over forecast and an under forecast are principally undesirable, a systematic approach to the development of a forecast model hence becomes necessary for producing credible predictions.

Ensuring a reliable operation for a utility complements its business needs. However, all such business needs associated with load forecasts precisely depend on the forecast lead times which are often referred to as forecast horizons. Depending on its forecast horizon, a forecast model inhibits different explanatory variables for the desired application. For example, long term load forecasts (LTLF), with lead

times ranging between three years to decades ahead, find their business needs in energy trading, energy policy, system planning and system maintenance [1], [2]. Medium-term load forecasts (MTLF; lead times of 2 weeks to 3 years) are carried out for unit commitment and power system planning and operations. Whereas short term load forecasts (STLF; lead times of 1 minute to 2 weeks) are primarily carried out for energy purchasing, effective demand-side management, peak load shifting, generation capacity planning and power purchasing [3]–[5]. It must be noted that these conventions may vary among different researchers as there are no universally set criteria for such categorization.

Besides explanatory variables, forecasting techniques also vary as the forecast lead times change. Therefore, load forecasts on all horizons i.e. long, medium, and short term are conventionally carried out using different methods and techniques. Of a variety of methods, for example, short term load forecasting techniques can be broadly categorized based on either the statistical modelling approach or machine learning methods [6]. From the category of statistical methods, regression analysis, especially multiple linear regression (MLR), is considered to be one of the most widely used methods in short term load forecasting literature due to its modelling flexibility [6]–[9]. It is a statistical technique and requires less simulation time as compared to artificial intelligence (AI) techniques like ANNs. Since it develops the model by determining the relationships between dependent and independent variables, MLR methods carry the advantage of higher interpretability and hence are one of the preferred methods adopted by the electric utilities [10], [11]. The availability of detailed load data at the consumer level has enabled the development of improved load forecasts using a variety of forecasting techniques. As a result, regression-based methods are ramping up in the performance hierarchy of short-term load forecasting techniques. This is due to the increased computational power resulting in reduced execution times with extensive training of data. However, these methods rely heavily on explanatory variables and large historical data as detailed data helps in generalizing such relationships [11].

Another forecast technique permeating in the list of short-term load forecasting methods is an AI-based technique called artificial neural networks. ANNs have been effectively in use for load forecasting problems since the 1990s [1]. Their data-driven approach enables forecasters to achieve desired results without pre-postulating the model for parameters estimation [12]. The accuracy of neural networks is subjected to the spread of input data [1]. However, ANNs lack in determining the prominence of the features they are given as an input for determining the forecasted variable; hence additional feature selection techniques have to be used. Consequently, the designed architecture of the ANNs, when given the significant feature space, potentially improves the forecasting accuracy. With their ability to capture the non-linearities of input data, numerous architectures and variants of ANNs have appeared in literature with prominent success [4], [13]–[19]. Although their simulation time can be higher, they have

the potential to approximate the non-linearities in the data and thus are often referred to as global approximators. For example, one of the applications of ANN-based forecasters known for their wide applicability in numerous utilities across the US and Canada has been presented in [20]. In their model, the authors used two ANN forecasters. While one predicted the base-load, the other predicted the change in load. In the end, the final forecast was produced by the adaptive combination of the previously produced two forecasts. Even beyond the scale of an electric utility, ANNs have been reported to outperform other methods when forecasts are made for service territories over wider geographical regions. For example, the short-term electricity demand for Indonesia has been forecasted best when neural networks were used in combination with other forecasting methods. An ensemble method SSA-LRF-NN predicted the load time series with the least MAPE and RMSE values [21].

With advanced computational resources and easy to access graphical user interfaces (GUIs), forecasting load using ANNs has become relatively less challenging today. However, a thorough review of literature on load forecasting in Pakistan shows that most of the literature is missing key elements of a load forecasting process. Authors often acquire the data and, without elaborating the relationship between the variables and the load, produce load forecasts for multiple time leads and geographical regions. Forecasting techniques with few selected variables are used without any justification or validation for their impact on the final forecast. For example, in [22], particle swarm optimization (PSO) based ANNs have been used to develop an STLF model. However, this model incorporates only the lagged load variables without considering any weather information or calendar effects. In another study [23], ANNs were used in comparison with the bagged regression trees (BRT) to forecast day-ahead electricity demand for an electric utility. In this study, authors claimed to have used weather data, calendar effects and lagged load variables in their model. However, no details of the model pertaining to the selection of its explanatory variables and model development were provided. Similarly, another one-hour ahead load forecast model was developed for an educational institution by using historical load data, weather information, and calendar effects [24]. The proposed model used the XGBoost algorithm as a forecasting engine. Although the authors of the study presented satisfactory results in terms of performance indices, they did not provide any insights on the model selection process for the achieved results; hence restricted the reproducibility of the study. In [25], the authors used a non-linear autoregressive (NARX) technique to develop an STLF model for a city's residential load using weather information, calendar effects and lagged load data. In a comparison with multiple other forecasting techniques, the authors demonstrated the highest forecast accuracy by using the proposed technique. Yet again, no information was provided on the feature selection process for the development of the proposed model. While investigating the economic impact of over/under forecasted

load for city-wide electricity consumption, authors in [26] used two different models to test multiple machine learning-based forecasting techniques. These models were: (a) model with input features including lagged load variables plus calendar effects and (b) model with input features including lagged load variables, calendar effects and meteorological parameters. The focus of this study remained on the economic significance of load forecasts on a city-wide scale and did not address the dynamics of a complete electric utility serving multiple cities with different weather profiles in its service territory.

Following the literature produced on load forecasting in Pakistan, it can be established that a standardized and systematic approach to model development remained missing in the published literature. Moreover, most of these studies lacked explaining the major developmental stages of a typical load forecast model. Whereas none of the published work addressed the feature selection process, development of synthetic weather station when required, and the effect of recency on the model's forecast accuracy. Besides that, insufficient information on sources of the acquired data for their respective forecast models yet again renders their reproducibility compromised. Thus, the need for a comprehensive study with reproducible results is imminent to help improve the future forecasting scenarios where the power system is facing tougher operational constraints with increasing penetration of renewable technologies, trigeneration systems and other such technologies.

In this regard, this paper presents a comprehensive and reproducible process of model development, model's training, and testing for a power utility, Islamabad Electric Supply Company (IESCO) in Pakistan which can be potentially reproduced for other similar distribution companies as well. Data used in this study consists of energy readings observed at one-hour resolution for five years (2015-2019) of IESCO. To evaluate the proposed model, both statistical and machine learning-based techniques have been applied and evaluated. Unlike previous studies, a systematic approach of features selection and model development based on a variety of input variables is presented by systematically eliminating the non-significant variables. The feature space includes variables related to meteorological parameters, calendar effects, and lagged time series of load and weather data exhibiting the recency effect as well as multiple seasonalities. Overall, the contribution of this study lies in connecting the previously unrelated facts about short term load forecasting in Pakistan by using synthetic weather station development and feature selection process. Furthermore, the study also discusses multiple prospects of different explanatory variables in a local context and highlights their significance for global applications as well. Therefore, this study promises to advance the knowledge on load forecasting research and benefit researchers and professional engineers to evaluate the performance of their respective forecast models.

Furthermore, the paper is organized as follows. In section 2, an overview of the energy system in Pakistan

is given. Section 3 discusses the methodological framework including data pre-processing, development of synthetic weather station, feature selection process and forecasting techniques. In section 4, results and analysis are detailed. Section 5 includes conclusive remarks and future works.

II. OVERVIEW OF ELECTRIC POWER SYSTEM IN PAKISTAN

Before its vertical unbundling, the water and power development authority (WAPDA) of Pakistan used to control power generation, power transmission, and distribution as the only government-controlled body in the country. However, in 1992, all the constituent elements of Pakistan's power infrastructure gained their operational autonomy under the regulation of the national electric power regulatory authority (NEPRA). The only power purchaser in the country is a state-owned central power purchasing agency (CPPA) whereas no utility is allowed to purchase power from the power generators, independently. Most of the power in the country is generated by state-owned generation plants using a variety of energy sources such as hydroelectric, furnace oil, diesel, gas, coal, and nuclear energy sources. In addition to state-owned power generation plants, a large number of independent power producers (IPPs) also add to the country's overall power generation. For the transmission of bulk power, the state directly runs national transmission and dispatch company (NTDC). Similarly, for the distribution of power, there are 10 distribution companies commonly referred to as DISCOs. All these DISCOs operate under a government agency i.e. Pakistan electric power company (PEPCO) [27].

A. OVERVIEW OF DISCOS

The distribution utilities of Pakistan, independent in their regional autonomy and operation, buy their required energy from the central power purchasing agency (CPPA) and collect revenue at the consumer end. CPPA, being a government entity, is exclusively responsible for managing financial aspects of power sale and purchase between generation (both state-owned and IPPs) and distributions companies in the country [28]. Since this study considers specifically a single DISCO/utility for its case development i.e., Islamabad electric supply company (IESCO), the overview of IESCO is given below as well.

B. IESCO

Since its inception in 1998, IESCO has primarily remained responsible for supplying and distributing power to the country's capital (Islamabad) and its immediate surroundings. These include administrative districts of Rawalpindi, Chakwal, Attock, Jhelum, and federal capital Islamabad itself, along with some areas of Azad Jammu and Kashmir region as well. This is shown in **FIGURE 1**. IESCO operates with 108 distribution grid stations and 951 feeders with a total power capacity of 5224 MVAs [29]. With its service territory spreading over 23,160 sq. km, IESCO serves over 2.42 million

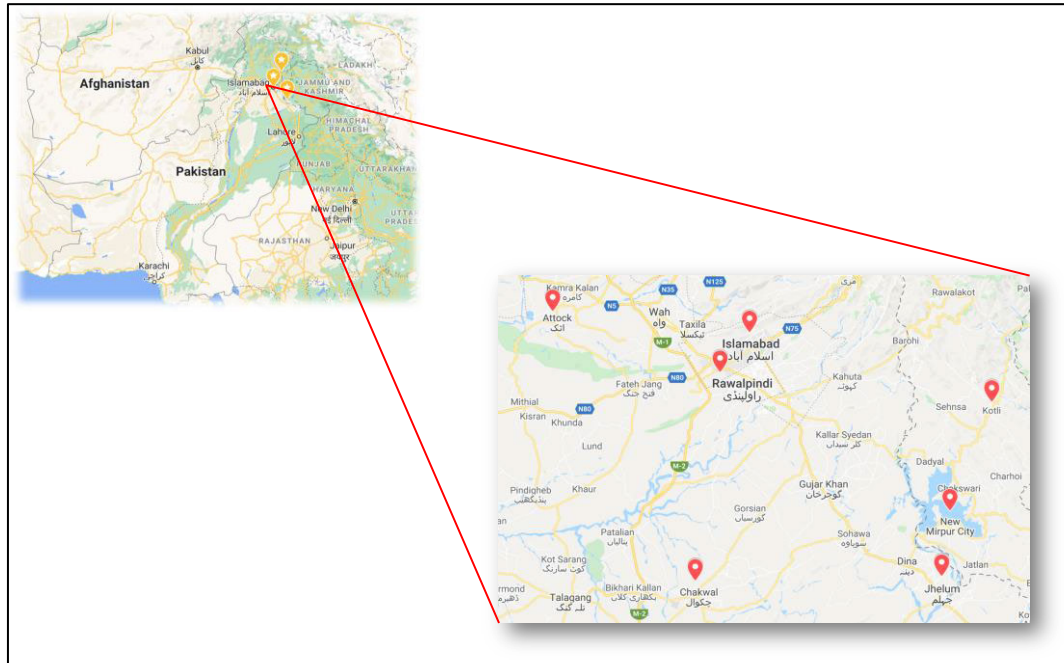


FIGURE 1. Administrative regions of IESCO service territory.

consumers including domestic, commercial, industrial, and agricultural customers [28].

As a large-scale electric utility, accurate load forecasts for IESCO serve as elementary inputs to its utility planning procedures and business needs. Amidst continuously changing fuel prices, ageing infrastructure, and generation restrictions due to environmental regulations, the chance of rising system constraints becomes even higher. Under such circumstances, a reliable forecast for any electric utility becomes crucial for efficient resource planning, tariff restructuring, and power dispatch.

III. METHODOLOGY

A. DATA EXTRACTION, DESCRIPTION, AND PREPARATION

Development of a reliable forecast model while using data-driven forecasting techniques requires quality input data. Since this study aims at developing a robust STLF model, a high quality-high resolution data becomes a prerequisite for producing a reliable forecast. Therefore, the authors acquired the hourly load data for IESCO from NTDC [30] to develop a forecasting model as well as validate its forecast accuracy. As the service territory of IESCO covers a wide region having diverse terrain and weather, data at the same time resolution as the load from multiple weather stations becomes another important requirement for model development.

However, the data available from the meteorological department of Pakistan was not detailed i.e. an average of 24 hours and only minimum and maximum readings of the data was available. Moreover, there were no weather stations installed in many areas of the utility’s territory either. Consequently, the meteorological variables i.e., 2m dry bulb

TABLE 1. Data statistics for load time series.

No. of Observations	Minimum Load	Maximum Load	Mean Load	Standard Deviation
43, 297	601.2 MW	2389 MW	1308MW	372.9

temperature and 2m dew point temperature utilized in this study were acquired from an open-source online climate datastore [31]. This hourly data for both the meteorological parameters were extracted from European Centre for Medium Range Weather Forecasts Re-Analysis-5 (ERA5) land hourly database. ERA5-land hourly data is a reanalysis dataset providing a consistent view of the evolution of land variables over several decades at an enhanced resolution compared to its versions. It is pertinent to mention that the selection of the meteorological parameters i.e. dry bulb temperature and wet bulb temperature was primarily because of the public data availability. Other load determining parameters such as humidity, dewpoint, air pressure, and wind velocity etc. were not publicly available for the service territory of the IESCO and hence were not used in the model development process.

The acquired load data is a time series of IESCO recorded at one-hour time stamps from 00:00 hours on Jan 01, 2015, to 00:00 hours on Dec 10, 2019. As shown in TABLE 1, this accounts for 43, 297 distinct load data points with a minimum load of 601.2 MW and a maximum of 2389 MW. This load time series is also shown in FIGURE 2.

It can be noticed that there are multiple seasonalities present in this time series. A strong seasonal trend can be observed over the years which exists due to two reasons. First,

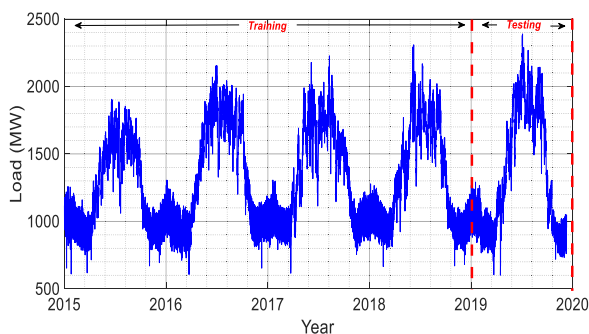
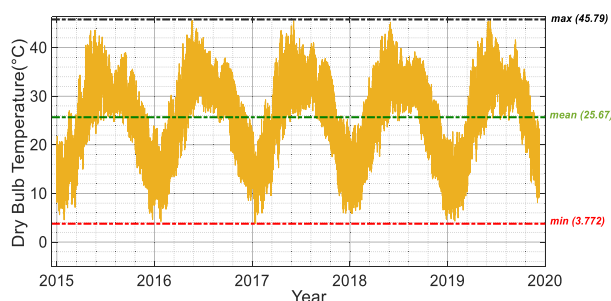
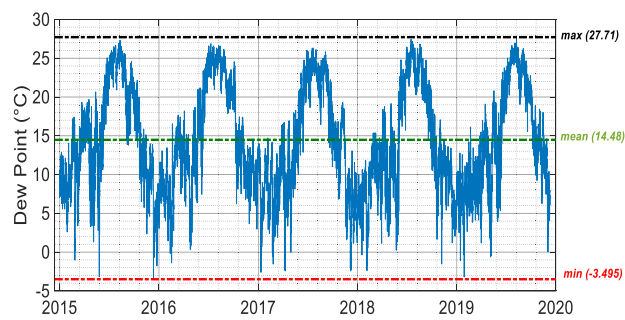


FIGURE 2. Hourly load time series between 2015 to 2019.



(a)



(b)

FIGURE 3. (a) Correlation of load with dry bulb temperature (b) Correlation of load with dewpoint temperature.

as shown in FIGURE 3. (a, b), there exists a strong dependence of load on weather patterns of summers (when the load is maximum) and winters (when the load is minimum). It can be observed that in Pakistan, there is a positive correlation between load and temperature (dry bulb and wet bulb) in summers. This means that when the temperatures rise, the electricity consumption increases primarily due to increased air conditioning. On the contrary, as temperatures start to drop down in winters, electricity consumption develops a negative correlation with temperature. However, the increase in load demand in summers is larger as compared to the winters and hence exhibits a much larger load demand on utility. Secondly, the electricity consumption patterns also depend on human activities which get reflected in holidays and calendar effects.

Such effects of human activities on electricity consumption can be observed through weekly and daily load patterns.

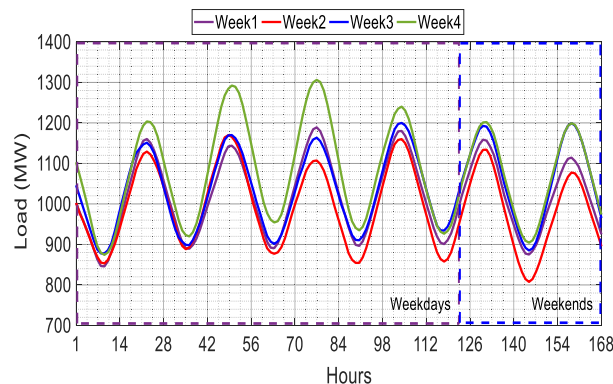


FIGURE 4. Weekly and hourly seasonality in electricity consumption of IESCO consumers.

TABLE 2. Weather stations and their coordinates.

Weather stations	Locations
Islamabad	33.6844 °N, 73.0479 °E
Rawalpindi	33.5651 °N, 73.0169 °E
Attock	33.7660 °N, 72.3609 °E
Murree	33.9070 °N, 73.3943 °E
Jhelum	32.9425 °N, 73.7257 °E
Chakwal	32.9328 °N, 72.8630 °E
Mirpur	33.1480 °N, 73.7537 °E
Kotli	33.5008 °N, 73.9007 °E

As shown in FIGURE 4., it can be observed that there are cyclic patterns in load across consumption hours of day and night within 24 hours' time period. Moreover, on weekends, load generally tends to decrease as compared to weekdays. Later in this study, the empirical effects of these seasonalities and their impact on determination of response variable shall be discussed. Similarly, this study incorporates hourly weather data for eight different weather stations between the same time and dates as it did for the electric load. These weather stations spread across eight cities as presented in TABLE 2. To account for their accumulative impact, data from every individual weather station were averaged and a synthetic weather station was developed.

In load forecasting, this is a highly recommended and practiced technique for developing weather-based load forecast models [1], [11], [32], [33]. Selection of the weather stations to create a synthetic weather station requires selection of significant weather stations, as an insignificant weather station can skew the synthetic weather station and results in a poor forecasting model. Following the aggregation of weather stations, resulting meteorological observations of the synthetic weather station are shown in FIGURE 5 (a, b).

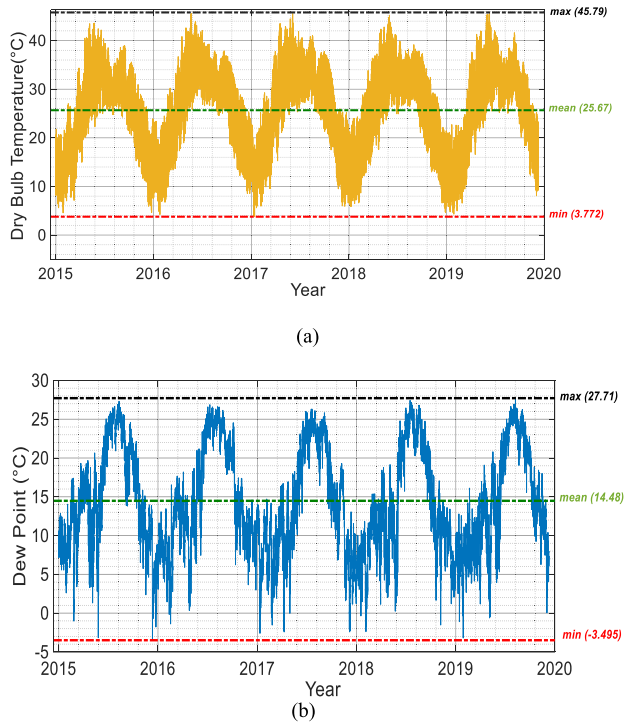


FIGURE 5. (a) Average dry bulb temperature of eight weather stations (b) Average dew point temperature of eight weather stations.

During the model development process, the quality of the input data was assured. This was achieved through the necessary pre-processing of the input data. These data pre-processing steps which were carried out in this study are shown in **FIGURE 6**.

Finally, 80% of the data was used to develop a training/fitting set i.e., from Jan 01, 2015, to Dec 31, 2018, and the remaining 20% for testing/predicting set i.e. from Jan 01, 2019, to Dec 10, 2019.

B. FEATURE SELECTION

In this study, a feature space of 22 different explanatory variables, which have frequently appeared in literature, has been developed [4], [5], [25], [32], [23], [34]–[37]. These explanatory variables are different factors that usually affect the electricity consumption patterns of any utility. However, the complexity of the resulting model is reduced without significantly compromising its accuracy. To actualize an accurate and relatively simple model, authors took advantage of a feature selection method during the model development process. For this purpose, the elastic-nets feature selection technique was used on the selected feature space. Since, for any regression problem where lasso and ridge regression can be used, elastic nets have empirically proven to have performed better [38]. They are a hybrid of both lasso and ridge regression and are used to solve regularization problems. They also have the ability to degenerate a model to its reduced form by generating zero-valued coefficients. Elastic nets come with the ability of grouped/pool variable selection; a feature that a simple lasso regression lacks [39].

TABLE 3. Feature space for model development.

Lagged variables (L)	Meteorological Variables (M)	Calendar days & Seasonal Effects (C)
Previous 24 hour average load (L1)	Temperature (M1)	Year (C1)
Previous 24 hour average temperature (L2)	Dewpoint (M2)	Month-X (C2)
Previous 24 hour average dew point (L3)	-	Month-Y (C3)
Prior day same hour load (L4)	-	Day of the week-X (C4)
Prior day same hour temp (L5)	-	Day of the Week-Y (C5)
Prior day same hour dew point (L6)	-	Hour-X (C6)
Prior week same hour load (L7)	-	Hour-Y (C7)
Prior week same hour temp (L8)	-	Isweekend (C8)
Prior week same hour dew point (L9)	-	Holidays (C9)
-	-	Binary zones (C10)
-	-	Seasons (C11)

Having decided on the feature selection technique, the base model with all the 22 features was first used to produce a load forecast. Following that, the applied feature selection technique gradually refined the model to reach an optimized set of features to yield both accuracy as well as simplicity. This trade-off between the model’s accuracy and parsimony was done by excluding features from the base model one after another while recording their impact on forecast accuracy. Since using all the distinct combinations of 22 different features (~22! i.e. 11×10^{20} combinations) is computationally highly challenging and, the authors made the best use of the relevance-to-load hierarchy of these 22 features produced by the elastic nets algorithm. In this way, a hierarchy of the explanatory variables in terms of their relevancy in determining the response variable (load) was developed. TABLE 3 lists all the explanatory variables that were used in the base model. These variables were further categorized into lagged variables (L), meteorological parameters (M) and calendar days/seasonal effects (C).

In the category of lagged variables, the average of previous 24 hours observations of load, temperature, and dewpoint as L1, L2, and L3 were used, respectively. In addition to that, observations of load, temperature, and dewpoint for the same hour from the previous day as L4, L5, L6 were also considered.

To incorporate the effect of weekly cyclic patterns, observations from the same hour on the same day from the prior week were also used for load, temperature, and dewpoint

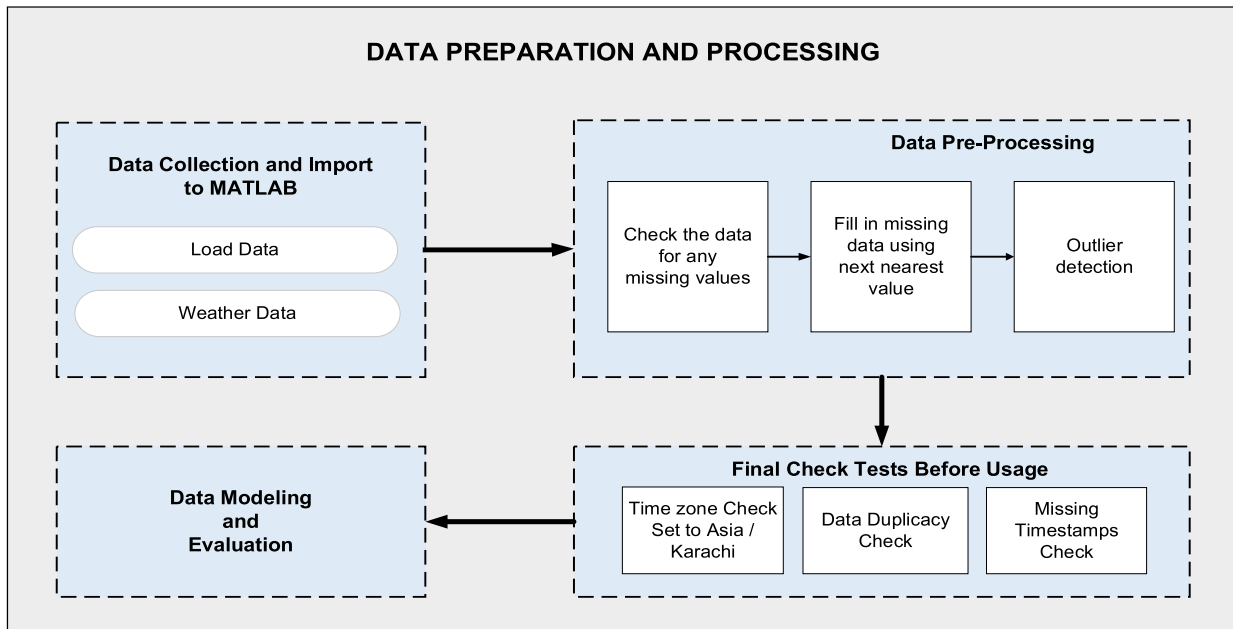


FIGURE 6. Preparation and pre-processing of data.

as L7, L8, and L9, respectively. Similarly, hourly observations of meteorological parameters such as temperature and dewpoint were used as M1 and M2, respectively. In addition to these weather-based parameters, other significant features such as precipitation, wind speed, wind chill index, cloud cover, light intensity, frequency of occurrence of natural disasters like mass flooding etc. could not be incorporated in the proposed method due to the limited access to that data. However, of all the features that were utilized, the majority of features belonged to calendar days and seasonal effects. Under calendar days/seasonal effects, year (C1), month of the year (C2, C3), day of the week (C4, C5), hour of the day (C6, C7), weekends (C8), calendar holidays (C9), binary zones (C10), and seasons (C11) were included. Because month of the year, day of the week, and hour of the day shows cyclic patterns and incorporate long term periodicities, their rectangular coordinates as C2, C3, C4, C5, C6, C7, respectively, were used instead of considering them otherwise. This is shown in (1) and (2).

$$X - coordinate = \frac{\sin(2\pi \times t)}{n} \tag{1}$$

$$Y - coordinate = \frac{\cos(2\pi \times t)}{n} \tag{2}$$

Here, t is the month of the year, day of the week, and hour of the day for the respective values of n as total number of months in a year (12), days in a week (7), and hours in a day (24). Owing to their periodic nature, the resulting features improved the network’s training process for higher testing accuracies.

For capturing the effects of weekends, Monday-Friday was considered as normal weekdays and Saturday and Sunday as typical weekends. To include holidays effect, several national

and religious holidays which are observed every year in Pakistan were considered. These are listed in TABLE 4.

Moreover, based on periodic increase and decrease in load during certain hours of the day, a new binary variable was tested and introduced in the model development process. It was named as ‘binary zones’ and was placed under the category of calendar days and seasonal effects as C10.

The reason behind including C10 was the effect of business hours on daily load patterns. For example, electricity demand tends to rise in the morning hours when all the commercial activities start and decreases as the night falls. Similarly, another binary variable for weather-based bi-yearly seasonality was introduced which stretches across summers (May to September) as well as winters (October to April) as C11.

To forecast a response variable, it is usually a group of certain variables, and not an individual variable, that produces minimum forecast error. It is also not necessary for the complete feature space to be utilized to produce the most accurate results [33]. Therefore, the most suitable combination of explanatory variables must be sought after. Starting with the base model, all the 22 chosen features were grouped as P1.

Following that, four more groups of explanatory variables were developed as P2 (top 20 features), P3 (top 15 features), P4 (top 10 features) and P5 (top 5 features). This is shown in FIGURE 7. Every group contains a combination of explanatory variables from all three categories i.e. lagged variables, meteorological variables, and calendar days/seasonal effects. This grouping was based on the relevance-to-load hierarchy that was produced as a result of feature selection. The significance of each group of variables manifests itself into a varying degree of performance accuracy and model’s simplicity. For example, the variable group P3 best forecasts the

TABLE 4. List of holidays.

Holidays					
National Holidays (Gregorian calendar)	Kashmir Day	Pakistan Day	Labour Day	Independence Day	Iqbal Day
Yearly date of occurrence	5th Feb	23rd Mar	1st May	4th Aug	9th Nov
Religious Holidays (Lunar calendar)	Eid Milad	Eid-ul-Fitr	Eid-ul-Adha	Ashura	-
2015	4th Jan	17th - 20th Jul	24th - 27th Sep	23rd - 24th Oct	-
2016	12th Dec	10th - 12th Jul	14th - 15th Sep	11th - 12th Oct	-
2017	1st Dec	29th - 30th Jun	2nd - 5th Sep	30th Sep - 1st Oct	-
2018	21st Nov	16th - 18th Jun	22nd - 24th Aug	21st - 22nd Sep	-
2019	10th Nov	5th - 7th Jun	12th - 15th Aug	9th - 10th Sep	-

load with the least MAPE. However, it does not make the simplest of the possible models. Whereas the simplest model with the least number of explanatory variables (P5) incurs the highest error. This qualifies the basic axiom of prioritizing a combination of variables instead of using only highly correlated or all the available variables in the feature space. Across the spectrum of accuracy and simplicity, a trade-off must therefore be made to maintain both the properties in the final forecast model.

To make a better sense of this trade-off, every group of variables was assigned an arbitrary value as its parsimony score or P-score. From a range of values between 0.2 to 1.0, the variable group making the simplest model (P5) was assigned

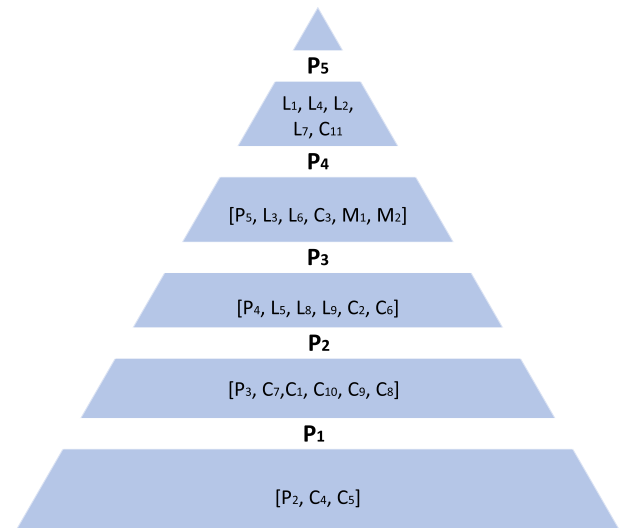


FIGURE 7. Variable groups containing different sets of features.

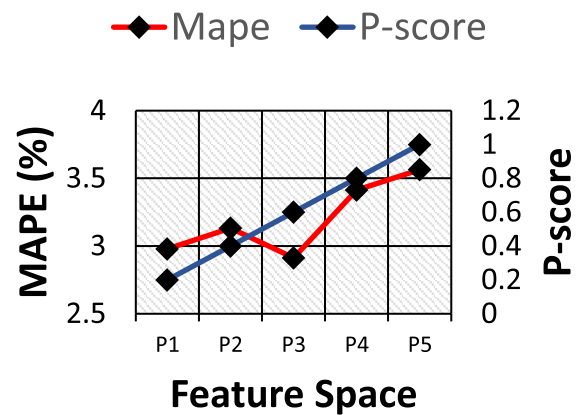


FIGURE 8. Trade-off between model's accuracy and simplicity.

the maximum score of 1.0. Similarly, the variable group that made the least simple model was given the minimum P-score of 0.2. This quantification was then plotted along with the relevant accuracies in terms of MAPE for all five models. This is shown in FIGURE 8. It can be seen that P3 results in the minimum MAPE; hence giving the most accurate forecast.

It can also be noticed that as additional features are introduced to this model, it loses both the relative accuracy as well as simplicity as compared to P3. Similarly, if some features are excluded from it, it becomes simpler but relatively less accurate.

C. LOAD FORECASTING TECHNIQUES

Short term load forecasts can be produced using a variety of statistical and nonstatistical methods. Since different techniques in both domains have their advantages and disadvantages, their suitability keeps changing from one model to another. In this study, three different techniques to forecast electric load for an electric utility have been used. These include multiple linear regression, decision trees, and

artificial neural networks. All three techniques were applied to the final model and a comparison has been made according to each model’s performance.

In the absence of the best available forecasting technique in literature [1], the selection of any such method for a load forecasting application becomes highly subjective to that particular model. This means that while one technique performs better on a particular data set, it may perform poorly on the other, and vice versa. Consequently, forecast performance, instead of the forecast technique itself, becomes desirous in the majority of the load forecasting applications. Similarly, in this study, all forecasting techniques have been chosen for their comparative performance with each other. All three techniques have been finalized for their global applicability and widely reported success in load forecasting applications in industry and academia. Moreover, the absence of their usage in Pakistan’s power distribution sector for high-resolution data (hourly observations for five years), its granularity, as well as for a systematically developed model for an hour ahead load forecasting resulted in their selection for this research.

1) MULTIPLE LINEAR REGRESSION

While using multiple linear regression (MLR), a load forecast model is conceived by developing a relationship between the target variable (dependent variable i.e. load) and multiple explanatory variables (load determinants such as meteorological parameters etc.). The mathematical representation of an MLR based forecast model is given below.

$$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_n X_n + \epsilon \quad (3)$$

In the above expression, y corresponds to the target variable (i.e. load), $X_1, X_2, X_3, \dots, X_n$ are the explanatory variables, $\beta_0, \beta_1, \beta_2, \beta_3, \dots, \beta_n$ are the regression coefficients whereas ϵ is the error value; a numerical difference between the observed and predicted values. For multiple values of y , the following representation can be used.

$$y_k = \beta_0 + \beta_1 X_{k1} + \beta_2 X_{k2} + \beta_3 X_{k3} + \dots + \beta_n X_{kn} + \epsilon_k \quad (4)$$

where $k = 1, 2, 3, \dots, k$. In a matrix form, this model can be represented as:

$$y = X\beta + \epsilon \quad (5)$$

where:

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_k \end{bmatrix}, \quad X = \begin{bmatrix} 1 & X_{11} & X_{12} & \dots & X_{1k} \\ 1 & X_{21} & X_{22} & \dots & X_{2k} \\ & \vdots & & \ddots & \vdots \\ 1 & X_{n1} & X_{n2} & \dots & X_{nk} \end{bmatrix}$$

$$\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} \quad \text{and} \quad \epsilon = \begin{bmatrix} \epsilon_0 \\ \epsilon_1 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

Since the regression coefficients betas (β_s) are still unknown, Eq. 2 can be used to find out these coefficients.

$$\beta = (X'X)^{-1} X'y \quad (6)$$

After finding out the values of regression coefficients, the above model can be used to predict future values of y following the below mathematical representation.

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \hat{\beta}_3 X_3 + \dots + \hat{\beta}_p X_p \quad (7)$$

Here, \hat{y} (read as y-hat) corresponds to the predicted value. Similarly, $\hat{\beta}$ (read as beta-hat) are the values of estimated regression coefficients.

2) BOOTSTRAP AGGREGATED DECISION TREES FOR REGRESSION

When used for a regression problem, bootstrap aggregated decision trees are commonly known as bagged regression trees or BRT. Since the data set contained multiple predictive features, regression trees was applied as one of the forecasting techniques to test the proposed forecast model. Moreover, regression trees have a history of wide-scale usage in load forecasting across the world and, in some studies, have performed better than other forecasting techniques as well [23], [10], [40]–[44]. Regression trees leverage bagging (which is a smoothing operation) and a bootstrap methodology which results in variance reduction [45]. In an attempt to reduce the overfitting in training data, 30 decision trees were used whose individual results had to be combined to produce the final forecast model [46].

3) ARTIFICIAL NEURAL NETWORKS

Given the input and output sample data, ANN is a promising technique to extrapolate the observed past values of the data into the future. These networks learn the relationship between the samples of input and output values by the supervised learning method. This learning method processes patterns of input values and generates the output closest to the value in the dataset by updating the weights. A variety of ANN models have been used in the literature. However, a feed-forward multi-layered perceptron (MLP) ANN model was used in this study. These models are primarily used for supervised learning as the data is neither sequential nor time-dependent. A simple MLP ANN model is shown in **FIGURE 9(a)**. The MLP feedforward networks have hidden layers that help them to learn the non-linear relationship between inputs and outputs, hence dealing effectively with the system complexity.

This is the reason that these networks are widely used in solving large datasets problems like load forecasting [47], [15]. The architecture of the feed-forward network is governed by the connection of neurons that are systematized into layers. Each neuron consists of one or more inputs x_i that are linearly combined. These inputs are attuned by specific weights w_i . In addition to this, a bias term θ is also used to fit the data in a better way by shifting the activation function to left or right. A simple neuron model is shown in **FIGURE 9(b)**.

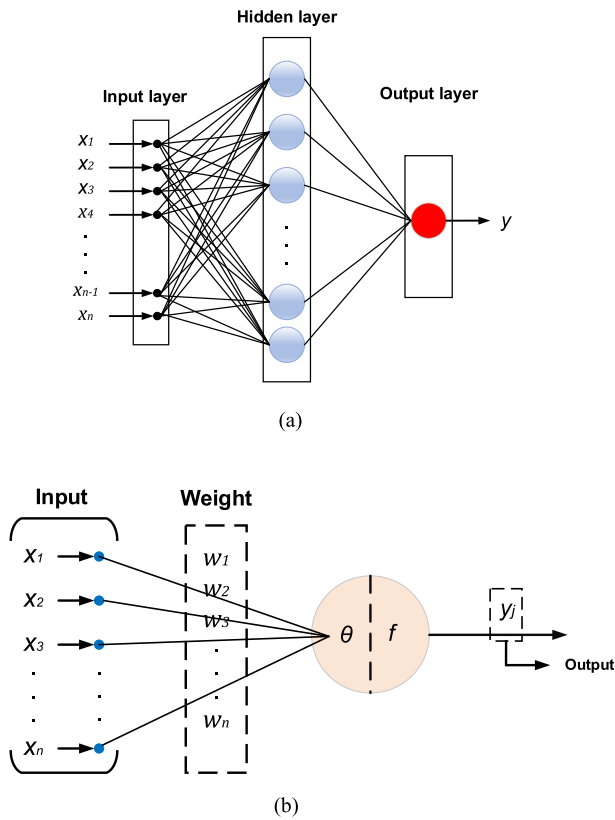


FIGURE 9. (a). Feed forward multi-layered perceptron (MLP) ANN model (b) A simple neuron model.

a: CREATING A FITTING NETWORK

The neural network developed in this study has five hidden layers in addition to one input and one output layer. Having decided the number of hidden layers, neurons in each layer form a connection with all the neurons of the former layer through specific weights. These weights act as network encoders. As the network is feedforwarded, the information is transferred layer by layer in the forward direction.

To predict the output in a feed-forward network, a two-step procedure is followed as shown in FIGURE 10. First, the input parameters of each neuron are multiplied by its weight and are then linearly combined. It is then passed to the activation function that simulates the behavior of neurons to get the output from them. The activation function can either be an identity function or a sigmoid function [48]. These functions must be differentiable and non-decreasing so that the non-linearity can be mapped.

In this study, a sigmoid function has been used as it can handle nonlinearities in the data. The sigmoid function is given as:

$$f = \frac{1}{1 + e^{-z_j}} \tag{8}$$

If a logistic activation function is used for hidden layer neurons i.e. a sigmoid and a linear function is used for output layer neurons, then the ANN model can be expressed

mathematically as [48]:

$$y_k = \sum_{j=1}^n w_{jk} \times \frac{1}{1 + \exp(-z_j)} + \theta_k \tag{9}$$

where y_k is the predicted output, w_{jk} represent the weight of link joining the output of neuron j and k , z_j is the output from the hidden layer node represented by $\sum_{i=1}^m x_i w_{ij} + \theta_j$. The values of weights w and biases θ can be determined by a training algorithm. These values are important and need to be tuned to minimize the error.

b: THE NEURAL NETWORK TRAINING PROCESS

The training algorithm that has been used to train neural network is a back-propagation algorithm, famously known as Levenberg Marquardt (LM) [49]. This algorithm is an advanced version of the gauss newton method. LM works by back-propagating the error and is suitable for function fitting problems like load forecasting and nonlinear regression problems [47]. It eliminates the need to find a Hessian matrix by computing a Jacobian matrix which speeds up the process [50]. This process is shown in FIGURE 11.

For the LM algorithm [50], let us first consider Newton’s method to minimize the performance function with respect to x .

$$w_{i+1} = w_i - H^{-1} \nabla V \tag{10}$$

where H is the hessian matrix, V is a function of a sum of squares of errors that is $V = \sum_{i=1}^N e_i^2$, ∇V is the gradient of V and w_{i+1} and w_i are the updated and current weights respectively. When performance index is approximated as V , then:

$$\nabla V = J^T . e \tag{11}$$

where J is the Jacobian matrix containing first derivatives of errors w.r.t to weights w and biases θ and e is the matrix of network errors. From here, a Hessian matrix can be represented as:

$$H = J^T J + S \tag{12}$$

Here S is a product of combination coefficient μ and identity matrix I . The value of S is very small as compared to the product of the Jacobian matrix, so it can be assumed zero. Therefore, the equation becomes,

$$H \approx J^T J \tag{13}$$

Using this approximation, the Gauss-Newton algorithm can be represented as follows:

$$w_{i+1} = w_i - [J^T J]^{-1} J^T . e \tag{14}$$

where w_{i+1} is the updated value of weight and w_i is the current value of weight. The Hessian matrix (H) in (13) is noninvertible. This is the reason why a modification is made in the matrix to make it invertible. A modified version of

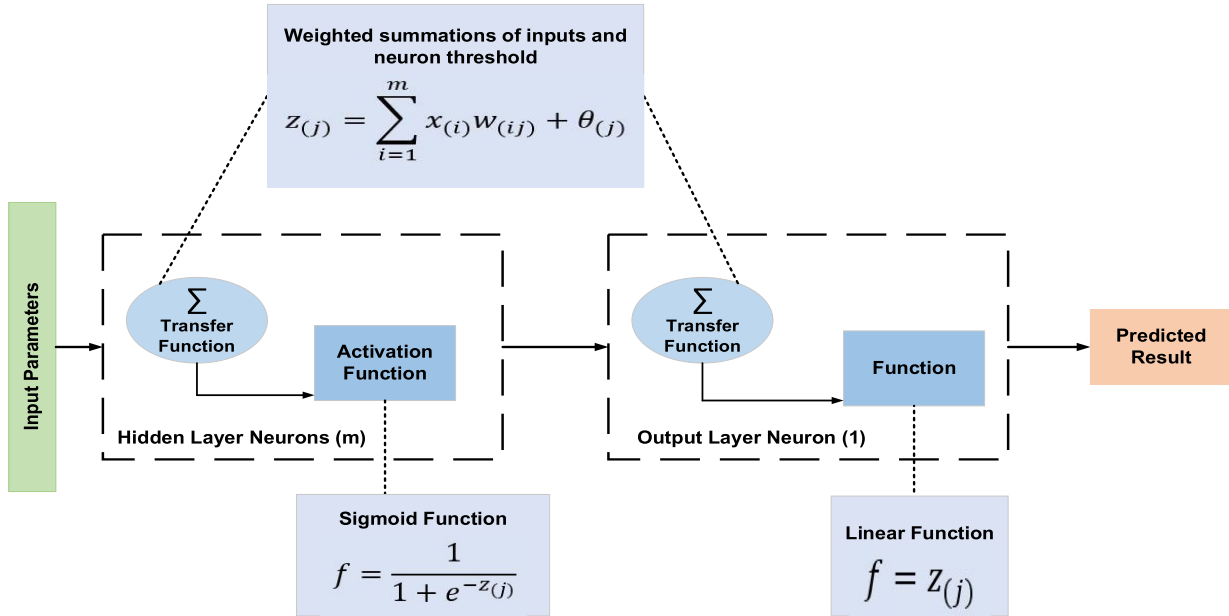


FIGURE 10. Two step procedure for developing a feed forward network.

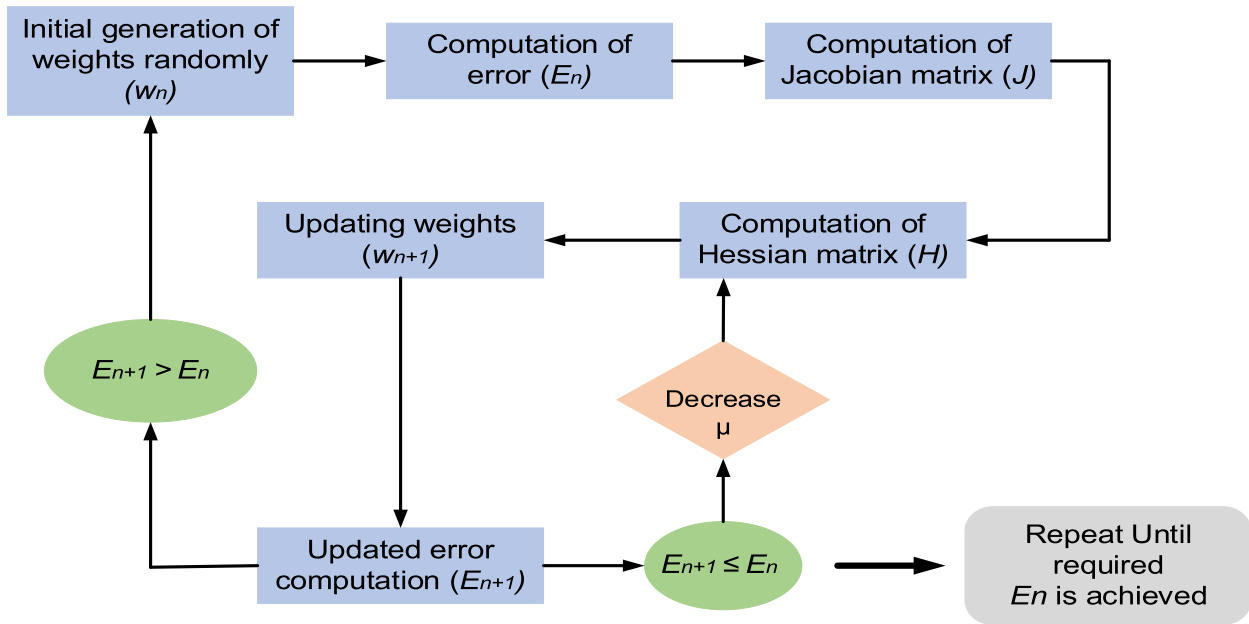


FIGURE 11. Levenberg Marquardt back propagation algorithm.

Hessian matrix with identity matrix I and variable μ is given as follows:

$$H \approx J^T J + \mu I \quad (15)$$

This modification leads to the LM algorithm as follows:

$$w_{i+1} = w_i - [J^T J + \mu I]^{-1} J^T .e \quad (16)$$

where the value of μ can be varied. Levenberg-Marquardt is a combination of two algorithms i.e. gradient descent and

Gauss-Newton. Its values shift between the two based on the value of μ . For a large value of μ , the algorithm becomes gradient descent and for a small value of μ , it shifts to Gauss-Newton. As the Gauss-Newton method is fast in terms of processing, it was shifted towards its corresponding value by decreasing μ .

This implies that the performance function will always decrease in this iterative process. While training the network, maximum number of iterations was set to 500. This selection is again arbitrary and depends on the network's performance.

TABLE 5. Variations in performance metrics across different feature groups.

Feature Space	MAPE						RMSE					
	MLR		BRT		ANNs		MLR		BRT		ANNs	
	Train	Test	Train	Test	Test	Train	Train	Test	Train	Test	Train	Test
P1	3.397	3.572	1.005	3.222	2.506	2.978	57.961	64.169	18.807	62.942	46.471	54.733
P2	3.411	3.607	1.039	3.263	2.551	3.132	58.255	64.816	19.506	63.112	47.352	56.533
P3	3.700	3.592	1.162	3.3294	2.883	2.913	59.998	65.050	20.184	64.822	49.396	56.792
P4	4.267	4.327	1.205	3.751	3.180	3.413	67.921	70.656	22.208	68.535	56.514	61.700
P5	4.741	4.720	1.498	4.011	3.491	3.564	76.014	78.192	27.951	70.433	62.945	64.526

It must also be noted that the number of iterations is model specific and hence, therefore, are subject to change as the model varies.

IV. RESULTS AND DISCUSSIONS

In pursuit of achieving an accurate as well as efficient model, model's most optimized performance on scales of both accuracy and simplicity was considered paramount.

A. BETWEEN PARSIMONY AND ACCURACY

Of all the five pools of explanatory variables, it was observed that P3 showed the most optimized performance in terms of forecasting accuracy as well as model's simplicity. As shown in TABLE 5, there is a significant percentage increase in MAPE as model selection traverses from P3 to P4 and P5; hence more parsimonious but less accurate. This remained consistent for MLR, BRT, and ANNs. However, while looking back from P3 to P2 and P1, the percentage decrease in MAPE for both MLR and BRT is relatively less significant as the model becomes more accurate as well as more complex. Whereas, while applying ANN, the model showed higher MAPE as it moved from P3 to P2. This too was undesirable and facilitated in deciding P3 as the most optimized choice of all the other pools of explanatory variables. Since deciding between accuracy and complexity is a choice based on the business needs, a tradeoff between the model's complexity and efficiency needs to be considered for making a final decision.

B. THE RECENCY EFFECT

The term 'recency effect' refers to the effect of variables of preceding time stamps on the variable being predicted. Variables with such properties of recency are known as lagged variables [51]. In load forecasting, literature reports ample evidence where numerous studies empirically suggest that the consideration of recency effect in load forecast models significantly improves the forecasting results [34], [52]–[54].

To elaborate on the significance of the recency effect, this study takes advantage of significantly relevant lagged variables to demonstrate their forecasting prowess. The selected pool of features (P3) was further into P3_{RE} (with only lagged

TABLE 6. Division of P3 into P3_{RE} lagged and P3_{NRE} without lagged variables.

Variations of P3	MLR		BRT		ANNs	
	Train	Test	Train	Test	Train	Test
P3_{NRE}	9.908	12.295	4.104	9.747	7.646	9.994
P3_{RE}	4.092	4.231	1.186	3.63	3.172	3.315
P3_{NRE} + P3_{RE}	3.700	3.592	1.162	3.294	2.883	2.913

variables i.e. with recency effect) and P3_{NRE} (without lagged variables i.e. no recency effect). This is shown in TABLE 6. Our experiments showed a significant improvement in model accuracy after introducing the recency effect to it. As lagged variables were added to the model, a threefold decrease in MAPE across all the techniques (MLR, BRT, ANNs) was observed.

Another significant observation was the composite effect of both lagged and non-lagged explanatory variables in determining our response variable (load) across all the forecasting techniques. This has been demonstrated using MAPE as a performance metric as shown in TABLE 6. It must be noted for convenience that authors have used P3_{NRE+RE} instead of P3 interchangeably in TABLE 6 solely for explanatory reasons.

C. THE HOLIDAY EFFECT

Having finalized the feature space for explanatory variables, it is important to highlight an explanatory variable i.e. Holiday Effect. This variable was dropped during the feature selection process. Since, in Pakistan, national holidays appear according to the Gregorian calendar and religious holidays appear according to the lunar calendar. Therefore, dates appearing on these two calendars are highly incompatible with each other with a difference of 11 days. This means that any religious holiday that appears on a particular date in the Gregorian calendar in 2015 will appear on a different Gregorian date in 2016 and so on. This resulted in a year-wise incoherence in holiday occurring dates and eventually lead to a low relevance status of holidays as a load determining variable in our forecast model.

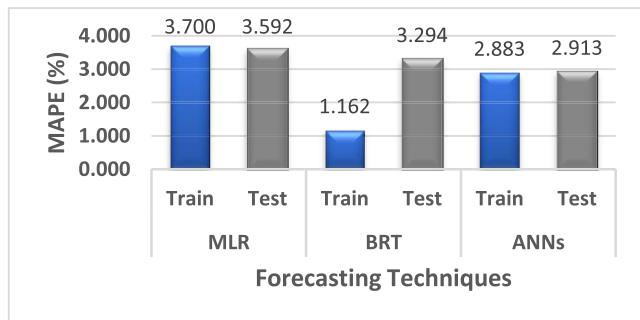


FIGURE 12. Results of proposed forecasting techniques.

D. FORECASTING TECHNIQUES AND THE WINNING MODEL

On the selected pool of features i.e. P3, three different forecasting techniques were applied. As shown in FIGURE 12, our results showed that ANNs outperformed both MLR and BRT with a considerable margin across all the performance metrics. In terms of MAPE, for example, while testing our model, forecasts produced by ANNs showed the least overall MAPE of 2.913% followed by 3.294% and 3.592% for BRT and MLR, respectively. This is a consequence of the extensive training process of neural networks and their ability to capture the non-linear relationships between input and the target variables. It was also noticed that compared to ANNs and MLR, regression trees showed large variance and tended to overfit the model during their training process. Whereas the element of overfitting seemed to be insignificant, and comparatively minimum, in ANN-based training and testing of the model. Such a property is desirable while developing accurate predictive models for load forecast applications.

The comparative performance of forecasted load series has been widely measured in terms of performance metrics such as MAPE, RMSE etc. irrespective of the employed forecast technique [33]. To ensure a fair comparison between different techniques, the same simulation conditions were used for all scenarios. Moreover, different error matrices including MAPE, RMSE and R² have been used to compare the performance of all algorithms. The consistency of the results across these statistical performance measures shows a fair comparison between the proposed and the comparative methods.

E. COMPARISON OF OBSERVED AND FORECASTED LOAD

Since ANNs showed the least forecasting error, it can be observed in FIGURE 13. (a, b) that in comparison to other techniques, ANNs very well followed the observed load trend in the test set of our data. It can also be noticed that, unlike ANNs and MLR, forecasts produced by regression trees performed particularly poor on load peaks and valleys. This is due to the high variance and overfitting of the BRT model. Another observation could be the behavior of MLR on peaks and valleys in the load series as it tended to under forecast the load on these instances more than ANNs and BRTs. If applied, such forecasting discrepancies could have

TABLE 7. Monthly performance of MLR, BRT and ANN.

Months	MAPE			RMSE		
	MLR	BRT	ANN	MLR	BRT	ANN
Jan	4.992%	2.681%	2.660%	62.817	36.074	35.710
Feb	3.987%	3.501%	3.142%	46.425	41.305	35.863
Mar	3.915%	3.248%	3.034%	43.767	40.562	36.334
Apr	3.370%	3.583%	3.165%	69.993	72.082	69.429
May	3.206%	2.999%	2.460%	59.193	56.886	46.775
Jun	3.349%	3.540%	3.319%	76.557	81.258	76.284
Jul	3.173%	3.841%	2.964%	78.255	99.895	72.820
Aug	3.797%	3.300%	3.148%	85.407	75.437	72.600
Sep	4.020%	3.919%	3.150%	93.465	95.268	80.012
Oct	2.655%	3.424%	2.760%	38.309	46.950	38.961
Nov	3.327%	2.649%	2.862%	37.566	29.630	31.901
Dec	3.318%	2.847%	2.290%	36.812	31.877	24.711
Average	3.592%	3.294%	2.913%	60.714	58.935	51.783

TABLE 8. Performance metrics and the computational cost.

Forecasting Techniques:		MLR	BRT	ANN	
MAPE	Monthly	Min	2.651%	2.657%	2.293%
		Max	4.991%	3.919%	3.319%
	Weekly	Min	1.812%	1.847%	1.445%
		Max	6.500%	7.381%	6.246%
RMSE (MW)	Monthly	Min	36.81	29.62	24.71
		Max	93.46	99.89	80.01
	Weekly	Min	21.63	24.28	19.22
		Max	136.27	147.89	123.69
R ² Score		0.975	0.977	0.982	
Runtime (s)		13.54	18.64	29.13	

severe consequences on a utility in terms of peak hour load dispatch and generation scheduling. Therefore, the observations helped us concede the forecasting superiority of ANNs on our proposed set of explanatory variables and rendered it the most suitable forecasting technique for electric utilities of Pakistan.

F. COMPARISON AND ANALYSIS OF MODEL'S PERFORMANCE ON MONTHLY AND WEEKLY BASIS

Furthermore, the performance of all the applied techniques has been evaluated on a monthly and weekly basis. While looking at the monthly performance as shown in TABLE 7, it is observed that ANNs remained consistent in performing better than their counterparts. However, during the winters season, ANNs performed better (2.845% MAPE) than in the summers season (3.008% MAPE). The underlying factors resulting in seasonal performance will be elaborated in the proceeding section. Similarly, TABLE 8 shows the minimum and maximum values of performance metrics on a monthly and weekly basis along with the computational cost of each method.

It is observed that the applied ANN structure yet again performed better than other techniques and showed least of the maximum MAPEs both on monthly and weekly resolutions.

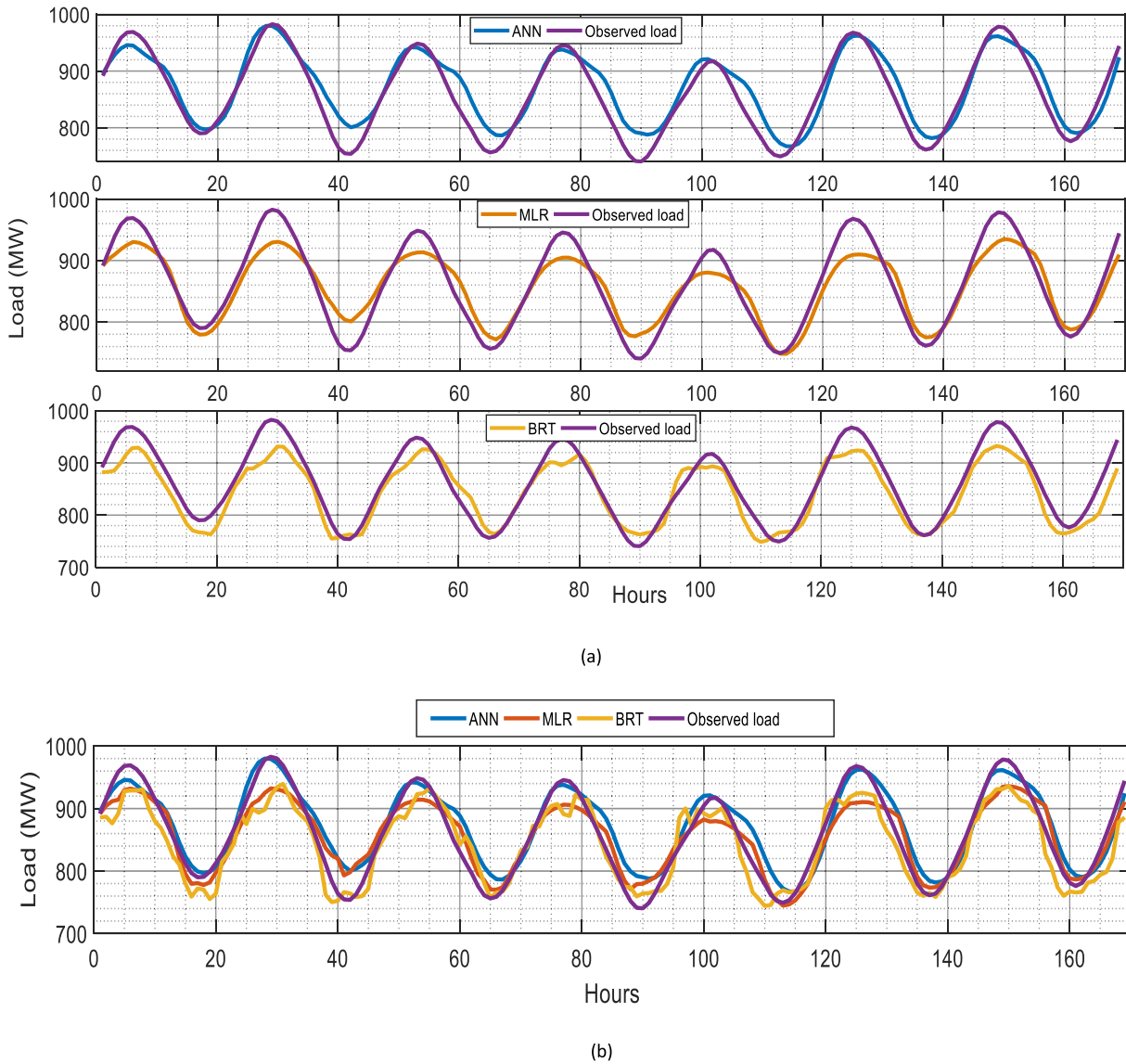


FIGURE 13. (a, b) Comparison of all techniques.

This must be noticed that the performance of neural networks remained consistent in terms of all the applied performance metrics (MAPE, RMSE, R^2). Throughout this work, the primary aim has been to develop a fairly accurate short term load forecast model for electric power utilities considering the model’s performance paramount. However, it is not necessary to have the highest performance and lowest runtime in a single model. As shown in Table 8, the ANNs appear to outperform other techniques in terms of forecasting accuracy. But at the same time, the ANN model takes the highest runtime (29.13 s) to execute; reflecting its high computational complexity. This is due to the rigorous training process of neural networks as compared to traditional regression methods, which arguably, results in better accuracy as well. Better in performance than other competing models, the runtime

of the ANN model turned out to be 2.1 times the runtime of the MLR model (13.54 s) and 1.56 times that of the BRT model (18.64 s). Although comparatively the fastest and computationally less challenging model, MLR loses its appeal when it comes to model’s forecast accuracy; hence becoming a secondary choice for utilities to adopt.

All the measured runtimes were calculated by running all the models individually on a 64-bit Intel(R) Core (TM) i7-8550U CPU @ 1.80GHz 1.99 GHz with 8GB RAM.

Overall results suggest that the ANNs outperform all other approaches and tend to perform better in terms of forecast accuracy. Further detailed analysis of the results was carried out to ascertain the reasons where and why the developed model performed relatively poorer. On monthly basis, the maximum MAPE occurred in June. While finding the

possible reasons behind the model's poor performance in a summers' month i.e. June, it was noted that in Pakistan, there is unscheduled load shedding that mostly occurs in the summers season because of the peak load demand in the season. Moreover, the monsoon season that lasts from mid-June to September in Pakistan also brought heavy flooding in the country in 2019 [55]. This resulted in unprecedented power cuts and unaccounted unserved energy for IESCO's consumers. Unfortunately, it was not possible to obtain any data to account for the quantitative effect of this unserved energy on the model's performance. On account of these two events, the major reasons behind the maximum performance error in June were both load shedding and massive flash flooding. Similarly, the maximum weekly error occurred in the second week of April 2019. Detailed analysis of the model's performance during that time of the year again traced back to the events like flash flooding which affected many parts of Pakistan including the southern service territory of IESCO. These floods took away multiple human lives while damaging dozens of houses and other infrastructure [56]. Yet again, with no access to data for the unserved energy during the second and third weeks of April, it was not possible to quantify the impact of this flood in the forecast model.

A detailed analysis of the results suggests that it is possible to develop a good forecasting model for a country like Pakistan where data availability is scarce to date and hence makes the development of research more challenging. However, while comparing different models, it is important to consider the special events including natural disasters, calamities, regional weather dynamics, and other human factors. If the availability of data allows, a comparison should be carried out during periods where no such events occur. Overall, the results suggest that the proposed model shows a higher level of accuracy as compared to other similar models developed in past. In addition to its widescale reproducibility, the proposed model followed a systemic approach to its development and hence becomes the right contender for serving as a base model for future load forecasting studies in Pakistan.

V. CONCLUSION

Load forecasting is considered a key input for the safe operation of the power systems. In this paper, the problem of systematic development of a short-term load forecast model for electric utilities in Pakistan was addressed. Maintaining the global applicability of the conducted research, the case of IESCO was presented for the systematic development of the proposed forecast model. The model was trained and tested on real-time data from the utility's five years hourly energy consumption. To produce desired load forecasts, three forecasting techniques i.e. MLR, BRTs, and ANNs were considered for evaluation by using multiple performance metrics. Significant weather profiles from eight different cities were selected to develop a synthetic weather station. For systematic model development, twenty-two explanatory variables were selected for the variable selection process by using elastic nets based feature selection algorithm. Experiments showed that

compared to MLR and BRTs, neural networks showed higher accuracy for hourly load predictions at hourly as well as weekly and monthly basis. The model was further evaluated to highlight the significance of data recency by incorporating lagged time series of load and weather data. An interesting aspect of the effect of public holidays in Pakistan was also examined which empirically suggested the low relevancy of public holidays for determining the load. Finally, the performance of the proposed model was evaluated on a monthly and weekly basis as well. Analysis showed that unprecedented events like load shedding and power cuts during flash flooding negatively impacted the model's forecasting accuracy. With limited or no access to data regarding climate variability and resulting natural disasters, the proposed model could not capture the effects of such special events which have had prominence in determining utility's final load demands in past.

In future, multiple prospects of this research can be explored for further development. For example, the dynamics of the 'holiday' variable can be modelled to capture its impact on forecasting accuracy. Moreover, the effect of unserved energy and power cuts can be modelled given the availability of the relevant modelling data. Another important development could be the design of an online forecast model for electric utilities using recurrent neural networks through a univariate analysis approach.

REFERENCES

- [1] T. Hong and S. Fan, "Probabilistic electric load forecasting: A tutorial review," *Int. J. Forecasting*, vol. 32, no. 3, pp. 914–938, 2016.
- [2] P. E. McSharry, S. Bouwman, and G. Bloemhof, "Probabilistic forecasts of the magnitude and timing of peak electricity demand," *IEEE Trans. Power Syst.*, vol. 20, no. 2, pp. 1166–1172, May 2005.
- [3] C. Baladron, J. M. Aguiar, B. Carro, and A. J. Sanchez-Esguevillas, "A survey on electric power demand forecasting: Future trends in smart grids, microgrids and smart buildings," *IEEE Commun. Surveys Tuts.*, vol. 16, no. 3, pp. 1460–1495, 3rd Quart., 2014.
- [4] T. Hong, *Short Term Load Forecasting*. Raleigh, NC, USA: North Carolina State Univ., 2010.
- [5] J. W. Taylor, "An evaluation of methods for very short-term load forecasting using minute-by-minute British data," *Int. J. Forecasting*, vol. 24, no. 4, pp. 645–658, Oct. 2008.
- [6] T. Hong, M. Gui, M. E. Baran, and H. L. Willis, "Modeling and forecasting hourly electric load by multiple linear regression with interactions," in *Proc. IEEE PES Gen. Meeting*, Jul. 2010, pp. 1–8.
- [7] N. Amral, C. S. Ozveren, and D. King, "Short term load forecasting using multiple linear regression," in *Proc. 42nd Int. Univ. Power Eng. Conf.*, Sep. 2007, pp. 1192–1198.
- [8] G. Heinemann, D. Nordmian, and E. Plant, "The relationship between summer weather and summer loads—A regression analysis," *IEEE Trans. Power App. Syst.*, vol. PAS-85, no. 11, pp. 1144–1154, Nov. 1966.
- [9] M. E. Baran, L. A. A. Freeman, F. Hanson, and V. Ayers, "Load estimation for load monitoring at distribution substations," *J. Inst. Electr. Eng.*, vol. 6, no. 66, pp. 348–351, 2005.
- [10] B. Yildiz, J. I. Bilbao, and A. B. Sproul, "A review and analysis of regression and machine learning models on commercial building electricity load forecasting," *Renew. Sustain. Energy Rev.*, vol. 73, pp. 1104–1122, Jun. 2017.
- [11] T. Hong and M. Shahidehpour, "Load forecasting case study," U.S. Dept. Energy, Washington, DC, USA, Tech. Rep. NARUC-2014-RFP042-DE0316, 2015, p. 171.
- [12] H. S. Hippert, C. E. Pedreira, and R. C. Souza, "Neural networks for short-term load forecasting: A review and evaluation," *IEEE Trans. Power Syst.*, vol. 16, no. 1, pp. 44–55, Feb. 2001.

- [13] A. S. Alfuhaid, M. A. El-Sayed, and M. S. Mahmoud, "Cascaded artificial neural networks for short-term load forecasting," *IEEE Trans. Power Syst.*, vol. 12, no. 4, pp. 1524–1529, Nov. 1997.
- [14] A. T. A. Loadforecaster, "Very short-term load forecasting using artificial neural network," *IEEE Trans. Power Syst.*, vol. 15, no. 1, pp. 263–269, Feb. 2000.
- [15] E. Banda and K. A. Folly, "Short term load forecasting using artificial neural network," in *Proc. IEEE Lausanne Power Tech.*, Jul. 2007, pp. 159–163.
- [16] M. Hayati and Y. Shirvany, "Artificial neural network approach for short term load forecasting for illam region," *Int. J. Electr. Comput. Eng.*, vol. 1, no. 4, pp. 667–671, 2007.
- [17] N. Kandil, R. Wamkeue, M. Saad, and S. Georges, "An efficient approach for short term load forecasting using artificial neural networks," *Int. J. Electr. Power Energy Syst.*, vol. 28, no. 8, pp. 525–530, 2006.
- [18] F. J. Marin, F. Garcia-Lagos, G. Joya, and F. Sandoval, "Global model for short-term load forecasting using artificial neural networks," *IEE Proc.-Gener., Transmiss. Distrib.*, vol. 149, no. 2, pp. 121–125, Mar. 2002.
- [19] Y. Shimakura, Y. Fujisawa, Y. Maeda, R. Makino, Y. Kishi, M. Ono, J.-Y. Fann, and N. Fukusima, "Short-term load forecasting using an artificial neural network," in *Proc. 2nd Int. Forum Appl. Neural Netw. Power Syst.*, 1993, vol. 7, no. 1, pp. 233–238.
- [20] A. Khotanzad, R. Afkhami-Rohani, and D. Maratukulam, "ANNSTLF-artificial neural network short-term load forecaster-generation three," *IEEE Trans. Power Syst.*, vol. 13, no. 4, pp. 1413–1422, Nov. 1998.
- [21] W. Sulandari, Subanar, M. H. Lee, and P. C. Rodrigues, "Indonesian electricity load forecasting using singular spectrum analysis, fuzzy systems and neural networks," *Energy*, vol. 190, Jan. 2020, Art. no. 116408.
- [22] S. R. U. Hassnain and A. Khan, "Short term load forecasting using particle swarm optimization based ANN approach," in *Proc. Int. Joint Conf. Neural Netw.*, Aug. 2007, pp. 1476–1481.
- [23] A. R. Khan, S. Razaq, T. Alquthami, M. R. Moghal, A. Amin, and A. Mahmood, "Day ahead load forecasting for IESCO using artificial neural network and bagged regression tree," in *Proc. 1st Int. Conf. Power, Energy Smart Grid (ICPESG)*, Apr. 2018, pp. 1–6.
- [24] A. Y. Kharal, A. Mahmood, and K. Ullah, "Load forecasting of an educational institution using machine learning: The case of NUST, Islamabad," *Pakistan J. Sci.*, vol. 71, pp. 252–257, Dec. 2019.
- [25] F. Abbas, D. Feng, S. Habib, U. Rahman, A. Rasool, and Z. Yan, "Short term residential load forecasting: An improved optimal nonlinear autoregressive (NARX) method with exponential weight decay function," *Electronics*, vol. 7, no. 12, p. 432, Dec. 2018.
- [26] M. Jawad, M. S. A. Nadeem, S.-O. Shim, I. R. Khan, A. Shaheen, N. Habib, L. Hussain, and W. Aziz, "Machine learning based cost effective electricity load forecasting model using correlated meteorological parameters," *IEEE Access*, vol. 8, pp. 146847–146864, 2020.
- [27] K. Ullah, "Electricity infrastructure in Pakistan: An overview," *Int. J. Energy, Inf. Commun.*, vol. 4, no. 3, pp. 11–26, 2013.
- [28] U. Qazi and M. Jahanzaib, "An integrated sectoral framework for the development of sustainable power sector in Pakistan," *Energy Rep.*, vol. 4, pp. 376–392, Nov. 2018.
- [29] Islamabad Electr. Supply Company, Islamabad, Pakistan, 2020.
- [30] National Transmission & Despatch Company Limited (NTDC) Pakistan. Accessed: May 17, 2020. [Online]. Available: <https://www.ntdc.com.pk/transmission>
- [31] J. M. Sabater, "ERA5-Land hourly data from 1981 to present, Copernicus Climate Change Service (C3S) Climate Data Store (CDS)," Eur. Centre Medium-Range Weather Forecasts, Eur. Union, Tech. Rep., 2019, doi: 10.24381/cds.e2161bac.
- [32] T. Hong, P. Pinson, and S. Fan, "Global energy forecasting competition 2012," *Int. J. Forecasting*, vol. 30, no. 2, pp. 357–363, Apr./Jun. 2014.
- [33] S. Fan and R. J. Hyndman, "Short-term load forecasting based on a semi-parametric additive model," *IEEE Trans. Power Syst.*, vol. 27, no. 1, pp. 134–141, Feb. 2012.
- [34] P. Wang, B. Liu, and T. Hong, "Electric load forecasting with recency effect: A big data approach," *Int. J. Forecasting*, vol. 32, no. 3, pp. 585–597, Sep. 2015.
- [35] A. A. Mir, M. Alghassab, K. Ullah, Z. A. Khan, Y. Lu, and M. Imran, "A review of electricity demand forecasting in low and middle income countries: The demand determinants and horizons," *Sustainability*, vol. 12, no. 15, p. 35, 2020.
- [36] I. Dimoulkas, P. Mazidi, and L. Herre, "Neural networks for GEFCom2017 probabilistic load forecasting," *Int. J. Forecasting*, vol. 35, no. 4, pp. 1409–1423, Oct. 2019.
- [37] J. Luo, T. Hong, and S.-C. Fang, "Robust regression models for load forecasting," *IEEE Trans. Smart Grid*, vol. 10, no. 5, pp. 5397–5404, Sep. 2019.
- [38] H. Zou and T. Hastie, "Erratum: Regularization and variable selection via the elastic net (journal of the royal statistical society. Series B: Statistical methodology (2005) 67 (301-320))," *J. Royal Stat. Soc., B Stat. Methodol.*, vol. 67, no. 5, p. 768, 2005.
- [39] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," *J. Royal Stat. Soc., B Stat. Methodol.*, vol. 67, no. 2, pp. 301–320, 2005.
- [40] C. Fan, F. Xiao, and S. Wang, "Development of prediction models for next-day building energy consumption and peak power demand using data mining techniques," *Appl. Energy*, vol. 127, pp. 1–10, Aug. 2014.
- [41] V. Dehalwar, A. Kalam, M. L. Kolhe, and A. Zayegh, "Electricity load forecasting for urban area using weather forecast information," in *Proc. IEEE Int. Conf. Power Renew. Energy (ICPRE)*, Oct. 2016, pp. 355–359.
- [42] C. González, J. Mira-McWilliams, and I. Juárez, "Important variable assessment and electricity price forecasting based on regression tree models: Classification and regression trees, bagging and random forests," *IET Gener., Transmiss. Distrib.*, vol. 9, no. 11, pp. 1120–1128, Aug. 2015.
- [43] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An introduction to Statistical Learning chap*, vol. 7, no. 10. New York, NY, USA: Springer, 2014.
- [44] R. Richman and M. V. Wäthrich, "Bagging predictors," *Risks*, vol. 8, no. 3, pp. 1–26, 2020.
- [45] P. Bähmann, "Bagging, boosting and ensemble learning," *Handb. Comput. Stat.*, no. 1, pp. 1–38, 2012.
- [46] Y. L. Pavlov, *Random Forests*. Berlin, Germany: De Gruyter, 2019, pp. 1–122.
- [47] M. Abuella and B. Chowdhury, "Solar power forecasting using support vector regression," 2017, *arXiv:1703.09851*. [Online]. Available: <http://arxiv.org/abs/1703.09851>
- [48] J. Schachter and P. Mancarella, "A short-term load forecasting model for demand response applications," in *Proc. 11th Int. Conf. Eur. Energy Market (EEM14)*, May 2014, pp. 1–5.
- [49] M. T. Hagan and M. B. Menhaj, "Brief papers," *Brain Cogn.*, vol. 32, no. 2, pp. 273–344, 1996.
- [50] H. Liu, "On the Levenberg-Marquardt training method for feed-forward neural networks," in *Proc. 6th Int. Conf. Natural Comput.*, Aug. 2010, pp. 456–460.
- [51] T. Hong, "Short term electric load forecasting," Ph.D. dissertation, North Carolina State Univ., Raleigh, NC, USA, 2010. [Online]. Available: <https://repository.lib.ncsu.edu/handle/1840.16/6457>
- [52] G. A. Darbellay and M. Slama, "Forecasting the short-term demand for electricity Do neural networks stand a better chance?" *Int. J. Forecast.*, vol. 16, no. 1, pp. 71–83, 2000.
- [53] A. D. Papalexopoulos and T. C. Hesterberg, "A regression-based approach to short-term system load forecasting," *IEEE Trans. Power Syst.*, vol. 5, no. 4, pp. 1535–1547, Nov. 1990.
- [54] S. B. Taieb and R. J. Hyndman, "A gradient boosting approach to the Kaggle load forecasting competition," *Int. J. Forecasting*, vol. 30, no. 2, pp. 382–394, Jun. 2014.
- [55] *Extreme Monsoon Rains in Pakistan*, Nat. Aeronaut. Space Admin., Earth Observatory, MD, USA, Sep. 2020.
- [56] *South Asia—Storms and Rain Kill Over 100 in India, Pakistan and Afghanistan—FloodList*. Accessed: Dec. 12, 2020. [Online]. Available: <http://floodlist.com/asia/storms-india-pakistan-afghanistan-april-2019>

ANEQUE A. MIR received the master's (M.S.) degree in electrical power engineering from the U.S.-Pakistan Centers for Advanced Studies in Energy, National University of Sciences and Technology, Islamabad, Pakistan. From 2016 to 2017, he attended Oklahoma State University as a Fulbright Scholar with the College of Engineering, Architecture, and Technology. He is currently working as an Assistant Director (Planning & Design) at Lahore Electric Supply Company, Lahore, Pakistan. He has published numerous journal articles and conference proceedings, and policy briefs in areas, such as energy forecasting and load management, energy policy, energy management systems, and power control engineering.

ZAFAR A. KHAN received the Ph.D. degree from the University of Birmingham, U.K. He is currently completing his postdoctoral training at Aston University, U.K. He is also an Assistant Professor with the Mirpur University of Science and Technology. He has experience in energy industry, research, and teaching. He has published a significant number of research papers in scientific journals, conference proceedings, and as book chapters in the fields of energy management, load forecasting, load profiling, smart grids, power system security, reliability, and active distribution network operation and risks in power systems.

ABDULLAH ALTMIMI received the B.Sc. degree in electrical engineering from the University of Hail, Saudi Arabia, in 2011, the M.Sc. degree in electrical engineering for renewable and sustainable energy from the University of Nottingham, U.K., in 2015, and the Ph.D. degree in electrical power engineering from the University of Birmingham, U.K., in 2020. He joined Majmaah University, Saudi Arabia, as an Assistant Teacher and a Lecturer, between 2012 and 2020. He is currently an Assistant Professor with the Electrical Engineering Department, Majmaah University. He has authored a significant number of research articles in international scientific journals and conference proceedings. His current research interests include renewable energy technologies and integration, smart grids, power system reliability and stability, climate change impacts, and distributed generation. He is also a reviewer and an editor in number of recognized international journals.

MARIA BADAR received the B.Sc. degree in electrical engineering from the University of the Punjab, Lahore, Pakistan, in 2017, and the master's (M.S.) degree in energy systems engineering from the U.S.-Pakistan Centers for Advanced Studies in Energy, National University of Sciences and Technology (NUST), Islamabad, Pakistan, in 2021. As an academic, she has also served as a Power Transmission, Distribution, and Utilization Lab Engineer with the University of Central Punjab, Lahore. Her research interests include control applications, energy management systems, microgrids, neural networks, and fuzzy logic-based controllers.

KAFAIT ULLAH received the M.S. degree in development economics from the University of Life Science, Norway, and the Ph.D. degree in energy economics and policy from the University of Twente, The Netherlands. He is currently working as an Assistant Professor with the Energy Economics, Policy and Governance with the U.S.-Pakistan Centre for Advanced Studies in Energy (USPCAS-E), National University of Science and Technology (NUST), where he is involved in teaching and supervising M.S. and Ph.D. students. He has published various articles in international peer-reviewed journals. He has also worked on different research projects relating to the energy sector in liaison with international organizations, such as the World Bank and USAID. He is also involved with energy-related public and private organizations in advisory and consultancy assignments on energy policy. His research interests include energy sector liberalization, institutional economics of public sector utilities, energy poverty, diffusion of energy innovations, and integrated energy planning and economics of renewable energy resources.

MUHAMMAD IMRAN is currently an Established Researcher in the area of thermal energy systems. His research activities aimed at developing innovative thermal energy systems and improving the energy performance of existing energy systems. He has been working in the area of waste heat to power conversion technologies (organic Rankine cycle), since 2012. His work in the organic Rankine cycle technology can be categorized into design and optimization, working fluid selection, flow boiling and condensation of refrigerants, binary mixtures, volumetric expanders, system integration, and dynamic modeling and control. His research involves both the numerical modeling and experimental approach. He has the opportunity to conduct research in the above-mentioned research areas in the Asia and the Europe. He was part of the team with the Korea Institute of Energy Research that successfully developed and commercialized 5kW, 30kW, and 100kW ORC system. He is particularly interested in the waste and low temperature heat to power conversion technologies (ORC, sCO₂), hybrid energy systems (geothermal, solar, wind, biomass), combined heat and power systems, thermal energy management, district heating and cooling, heat pump, and thermal energy storage. He has successfully completed number of research projects in above mentioned research areas. His research work has been published in various internationally recognized ISI indexed journals and international peer-reviewed conference proceedings. He is also serving as a reviewer for number of international journals in the field of thermal energy.

SYED ALI ABBAS KAZMI received the bachelor's degree in electrical engineering from the University of Engineering and Technology, Taxila, Pakistan, the master's degree in electrical power engineering from the University of Engineering and Technology, Peshawar, Pakistan, and the Ph.D. degree in electrical power engineering from Sungkyunkwan University, South Korea. He is currently an Assistant Professor with the Electrical Engineering Department, U.S.-Pakistan Center for Advanced Studies in Energy (USPCAS-E), National University of Sciences and Technology (NUST). His research interests include voltage stability, distributed generation, smart grids, power system modeling, power system planning, micro grids, multi-microgrids, and virtual power plants. Mr. Author's awards and honors include the Frew Fellowship (Australian Academy of Science), the I. I. Rabi Prize (APS), the European Frequency and Time Forum Award, the Carl Zeiss Research Award, the William F. Meggers Award, and the Adolph Lomb Medal (OSA).

• • •