

Received July 12, 2020, accepted July 28, 2020, date of publication August 4, 2020, date of current version August 20, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3014086

Machine Learning Based Cost Effective Electricity Load Forecasting Model Using Correlated Meteorological Parameters

MUHAMMAD JAWAD¹, MALIK SAJJAD AHMED NADEEM¹, SEONG-O SHIM²,
ISHTIAQ RASOOL KHAN², ALIYA SHAHEEN³, NAZNEEN HABIB⁴,
LAL HUSSAIN¹, AND WAJID AZIZ^{1,2}

¹Department of Computer Sciences and Information Technology, University of Azad Jammu and Kashmir, Muzaffarabad 13100, Pakistan

²College of Computer Science and Engineering, University of Jeddah, Jeddah 23890, Saudi Arabia

³Department of Mathematics, University of Azad Jammu and Kashmir, Muzaffarabad 13100, Pakistan

⁴Department of Sociology and Rural Development, University of Azad Jammu and Kashmir, Muzaffarabad 13100, Pakistan

Corresponding author: Malik Sajjad Ahmed Nadeem (msajjadnadeem@gmail.com)

ABSTRACT Electricity, a fundamental commodity, must be generated as per required utilization which cannot be stored at large scales. The production cost heavily depends upon the source such as hydroelectric power plants, petroleum products, nuclear and wind energy. Besides overproduction and underproduction, electricity demand is driven by meteorological parameters, economic and industrial activities. Therefore, the region specific accurate electric load forecasting can help to effectively manage, plan, and schedule appropriate low cost electricity generation units to decrease per unit cost and provision of on time energy for maximum financial benefits. Machine learning (ML) offers different supervised learning algorithms including multiple linear regression, support vector regressors with different kernels, k-nearest neighbors, Random Forest and AdaBoost to forecast the time series data, but the performance of these algorithms is data dependent. It is vitally important to consider correlated meteorological parameters of the specific region for accurate prediction of electricity load demand using ML based forecasting models to minimize the price per unit. In this study, an algorithm is proposed to select least cost electric load forecasting model (lcELFM) using correlated meteorological parameters. We developed least cost forecasting models by minimizing root mean squared error, mean absolute error, and mean absolute percentage error. For simulations, the recorded electricity demand data is taken from a substation of water and power development authority Muzaffarabad city from 1st January 2014 to 31st December 2015. The meteorological time series data are obtained from the substation of Pakistan meteorological department for the same period and same region. Empirical results demonstrate the robustness of the proposed algorithm to select lcELFM. Moreover, SVR (Radial) based electric load forecasting model proves to be the robust model when built using correlated features (temperature and dew point) for the said region and in turn can save up to PKR 0.313 million daily.

INDEX TERMS Electricity load demand, electricity load forecasting, least cost forecasting model, machine learning, meteorological parameters.

I. INTRODUCTION

Electricity is used as a major source of energy which is produced from electricity generating units. These units may use water, petroleum products (oil, natural gas etc.), nuclear energy or wind as a fuel to generate electricity [1]. One of the major issues regarding electricity generation is its

The associate editor coordinating the review of this manuscript and approving it for publication was Mauro Tucci¹.

storage at large scales [2]. End users, being one of the main stakeholders, demand minimum per unit cost. The per unit cost of electricity is mainly governed by two factors: type of fuel used; and electricity production as per requirements for a specific region. Both overproduction and under production of electricity cause financial loss to electricity generating and distribution companies (EGDCs). Therefore, EGDCs try to use least cost production units to produce electricity as per demand for a specific time period to maximize economic

benefits [3]. Therefore, an automated and accurate electricity load forecasting (AELF) plays a vital role in devising policies for adequate load management, scheduling the electricity generating units, predicting future trends, contract evaluation, tariff adjustments, and ensuring uninterrupted energy supply to the consumers [4], [5]. More accurate an AELF, more it is financially beneficial for EGDCs. These factors trigger the research for accurate electricity load forecasting so that EGDCs avoid unnecessary power generation to prevent increase in per unit cost and to serve the consumers with low cost and uninterrupted power supply.

During the last few decades, various research studies have been conducted for AELF models suggesting that both under and over forecasts greatly affect profitability of EGDCs [6], [7].

Numerous factors such as air temperature, dew point, wind, relative humidity, holidays, working hours, and seasons are critical and predominantly affect the demand in the electricity consumption [8]. These factors may be collectively called as critical factors affecting electricity load demand. A comprehensive understanding of these factors is necessary for developing accurate load forecasting models. These factors cause deviations from typical load curves and hence, make AELF a challenging task [9].

Irrespective of the approach (short, medium or long term) being used, meteorological parameters are important in AELF. Temperature, relative humidity, visibility, cloud cover, rain fall, precipitation, dew point, wind speed and wind chill are commonly used variables (either individually or in combination) in various techniques [10].

In [11], researchers reported 6% additional electricity consumption per annum due to the global warming in Greece. The authors in [12] reported that the increase in temperature can increase peak load demand up to 15% in Thailand at the end of this century. In [13], the authors showed the strong influence of relative humidity and wind speed on electricity consumption in eight states of America. In [14], different patterns of load demand during different seasons in New South Wales Australia are represented through regression models based on the meteorological parameters. It is found that, the per capita consumption demand can rise up to 6.14% and 11.3% in summer and spring seasons respectively by the end of this century. In [15], the authors showed the sensitivity of weather change to load consumption for different cities of Australia. The temperature may also have an inverse relationship with the load consumption in some countries. For example, the authors in [16] show that in New Zealand, an increase in temperature by 10° C causes a decrease in average yearly load demand by 1.4% and a decrease in temperature by 10°C causes an increase in average yearly load demand by 1.6%. In [13], it was found that the fluctuations in the meteorological parameters especially, the air temperature strongly impact the residential cooling demand in USA.

In short, the consumption of electricity by households and businesses for heating and cooling purposes is primarily driven by weather phenomenon. A change in weather patterns

causes a shift in the electricity load demand, which if not foreseen may result into an imbalance between supply and demand [17]. This imbalance may also trigger an increase in production cost of electricity due to the use of expensive production units (kerosene oil based units for example) and wastage of unutilized produced electricity due to overproduction.

Though, recognizing the significance of this issue, studies have been carried out to analyze the impact of weather on load demand in different regions [18], [19]. As per the best of our knowledge in the existing literature in this dimension, the region of Kashmir is completely missing. In addition, the existing studies for computing energy demand equation all the over the Pakistan, do not consider meteorological parameters as descriptive variables [20], [21]. The authors in [22] studied the relationship with other influencing variables, but didn't consider the impact of meteorological parameters. By the same token, these studies used only monthly load demand data which does not provide information about insights of load-weather relationship on small scales, as the problem requires much disaggregated analysis. A comprehensive investigation of this relationship is imperative because it has consequences on capacity building of electricity generating units. The supply bottlenecks can be avoided by devising strategic plan, which hinges upon accurate estimates of potential power demands. Temperature has been documented as the most influential variable on load demand among others in numerous research studies [8], [23]. However, setting apart temperature, information is completely inconclusive for the region under consideration on the combination of weather variables which provide meaningful input to AELF i.e. the variables with the highest correlation with load demand.

The presented study has a broader objective of estimating the impact of climate variables on the selection of a suitable, least cost, and accurate electricity load forecasting model. For this purpose, in this study, an algorithm is proposed for the choice of robust least cost electric load forecasting model (IcELFM) using meteorological variables and electric load demand. As the first step of proposed algorithm, correlation analysis is performed to determine the correlation among weather parameters and electric load. For this purpose, this study considers all important meteorological variables that may influence the load consumption on hourly, seasonal and annual basis in the region under study (Muzaffarabad city). The meteorological and electricity load demand datasets are acquired from local stations of Muzaffarabad, Azad Kashmir. In the proposed algorithm, the projections of the fluctuations in electricity demand under different weather parameters like temperature, dew point, relative humidity, cloud cover and rainfall are provided on hourly basis using correlation analysis. This helped in identifying only relevant weather variables to load forecasting in Muzaffarabad, because parsimony can be achieved in the prediction models by considering only those weather variables which proved to be relevant to electricity load demand

in the correlation analysis and using extra exogenous variables will create complexity dilemma in forecasting models with negligible or no improvement at all in the results. Literature shows that numerous studies used Pearson's correlation analysis to determine the relationship between electric load and meteorological parameters [24]–[30]. However, we used Pearson's correlation, Spearman correlation analysis (a popular non-linear rank based correlation analysis method) and `nlcor` function of R language to compute nonlinear correlation. The results of Spearman correlation analysis and `nlcor` are in line with the results of Pearson correlation analysis. All three of these correlation analyses show that there is linear relationship between meteorological parameters and electric load variable for available dataset used in this study. The relevant variables are then embedded into forecasting models to determine the accuracy improvement using popular machine learning algorithms. As a next step, the economic significance of developed forecasting models are computed. Finally, the proposed algorithm selects that forecasting model as `lcELFM` which has minimum cost per unit of electricity. Empirical results show the usefulness of proposed `lcELFM` algorithm for least cost electricity production.

The paper is organized as follows. First section includes the introduction and literature review of the study. Data description and preprocessing of data are discussed in the second section. Methodology of the proposed study is detailed in the third section. Proposed algorithm is presented in fourth section. Simulation design to investigate the usefulness of proposed algorithm is discussed in section five. Results of the correlation analysis are presented in section six. Results of electric load forecasting and economic significance are discussed in sections seven and eight, respectively. Finally, discussion and conclusion are detailed in the last section. In this manuscript, electric load and electricity load and load are interchangeably used but refers to the same entity i.e. electricity load demand.

II. DATASETS

Numerous studies from recent past used one to three years' time series data to build energy load forecasting models. For example, in [8], [31]–[34] the authors used datasets of eighteen months to two years duration for developing forecasting models. Data of one year was used in [6], [35]–[37] while three years data was used in [38]. In this study, a time series data of hourly electricity consumption of Muzaffarabad city in kilowatt-hour (KWh), across the time period from 1st January 2014 to 31st December 2015 (two years) is used. Electricity demand data was recorded by water and power development authority (WAPDA) which is responsible for production and supply of electricity across the country. The data was obtained from a local substation in Muzaffarabad city.

The meteorological time series data are obtained from a local meteorological substation of Pakistan meteorological department (PMD) in Muzaffarabad. The meteorological data are collected for the same period as it was for the electricity

consumption data to develop the relationship over the corresponding period i.e. 1st January 2014 to 31st December 2015.

The dataset includes hourly air temperature, dew point temperature, relative humidity, cloud cover, wind speed, and rainfall. Table 1 displays the descriptive statistics of electricity load and meteorological parameters. Between the years, the mean values of load demand, dew point, humidity and cloud cover are higher in 2015 as compared to 2014, contrary to temperature and wind speed which are lower. Overall, similar weather pattern can be observed between the years.

TABLE 1. Descriptive statistics of electric load and meteorological parameters.

Parameter	Year	Mean	Max.	Min.	SD
Electric Load (KWh)	2014	32661.00	50183	19137	6172.00
	2015	34984.00	54497	13353	6641.00
Temperature (°C)	2014	19.54	42	-1	8.98
	2015	18.93	39	1	8.14
Dew Point (°C)	2014	12.39	40	-4.9	7.59
	2015	12.41	50.8	-12.3	6.53
Humidity (%)	2014	65.95	100	15	26.05
	2015	69.19	100	15	18.71
Cloud Cover (0-8)	2014	03.00	8	0	2.90
	2015	03.15	8	0	3.00
Rain Fall (mm)	2014	43.46	102.23	0.12	32.29
	2015	57.69	110.12	9.78	35.24
Wind Speed (km/h)	2014	9.12	10.3	7.3	1.05
	2015	8.1	9.8	5.6	1.12

III. METHODOLOGY

This section discusses the data preprocessing step and methods adopted to determine the correlation between electric load and meteorological parameters. Next, machine learning algorithms used to build forecasting models in present study are presented. Performance evaluation measures and computation of forecasting models' economic significance are detailed at the end of this section.

A. DATA PREPROCESSING

Pakistan has been struggling against severe energy crisis from decades, so as the Muzaffarabad city which caused load shading and sudden power blackouts, resulting into missing values in the load time series. These missing values are interpolated, instead of discarding them in order to obtain real time behavior. Means by nearby points' method is used for this purpose according to the following example criterion:

$$x' = (x_1 + x_2 + x_3 + x_4)/4 \quad (1)$$

where x_1 and x_2 are two previous values and x_3 and x_4 are two values next to missing value (x'). The pre-processing helped in removing irregularities and smoothing the load time

series. Once the data is pre-processed it is divided into different categories like mean daily value, maximum daily value, hour to hour data of each variable, mean of each hour and seasonal data.

B. WEATHER LOAD RELATIONSHIP

The effects of weather variables on load demand are determined by scatter plots and correlation analysis using linear regression trend line. Statistically, the interdependence of two variables is given by the equation:

$$y = ax + b \quad (2)$$

where y is the dependent variable, x denotes the independent variable, a and b represent the regression coefficients.

In statistics, correlation analysis is used to measure the strength of relationship between two variables. The most widespread and frequently employed method for this purpose is the Pearson's correlation analysis. Numerous studies used Pearson's correlation analysis to determine the relationship between electric load and meteorological parameters [24]–[30]. In this study, Pearson's correlation, Spearman correlation analysis (a popular non-linear rank based correlation analysis method) and `nlcor` function of R language are used to investigate the relationship among meteorological parameters and electric load demand. Higher levels of the correlation estimates of nonlinear correlation analysis (values near to 1) indicate that the variables involved in the analysis are nonlinearly correlated and lower values (near to 0) show linear relationship. The results of all three of these correlation analyses show that there is linear relationship between meteorological parameters and electric load variable for available dataset used in this study. Pearson's correlation coefficient r is an index that quantifies; to what extent the given variables are related linearly. Its value can range from +1 to -1 and is given by the following equation:

$$r = \frac{n \sum xy - \sum x \sum y}{\sqrt{\{n \sum x^2 - (\sum x)^2\} \{n \sum y^2 - (\sum y)^2\}}} \quad (3)$$

where

n is the number of pairs

$\sum xy$ is the sum of products of paired scores

$\sum x$ is the sum of x scores

$\sum x^2$ is the sum of squared x scores

$\sum y^2$ is the sum of squared y scores

The strength of the relation is inferred from the value of r according to the following criteria:

$r = +1$ shows linear and perfect positive correlation.

$0.0 < r < 0.09$ shows no correlation.

$0.1 < r < 0.25$ shows small linear correlation.

$0.26 < r < 0.40$ shows medium linear correlation.

$0.41 < r < 1.0$ shows strong linear correlation.

$r = 0$ shows no correlation exists between two variables

$-1 < r < 0$ shows negative linear correlation.

$r = -1$ shows linear and perfect negative correlation.

C. MACHINE LEARNING METHODS FOR LOAD FORECASTING

Many important mechanisms are involved in the energy generation like policy decisions, management, planning, reliable power operations, meeting future demands, contract evaluation and maintenance [10]. These mechanisms require advance information of future electricity load demands. Several machine learning methods have been used to assist EGDCs in estimating these mechanisms through AELF. These methods provide useful information about future load demand trends by analyzing past data and influencing factors. The EGDCs can make smart decisions regarding power generation and distribution based on the information provided by machine learning methods. As evidenced by the scientific literature, the meteorological variables have significant effects on electricity load demand in many regions and in turn, these may have a role in the accuracy of load forecasting. In order to assess the improvement in the load forecasting due to inclusion of highly correlated weather variables in Muzaffarabad region, a predictive analysis is conducted by employing various well known machine learning methods. These are Multiple Linear Regression (MLR) [39], k -Nearest Neighbors (k NN) [40], Support Vector Regressor (SVR) [41] which is based on the principles of Support Vector Machines (SVMs) [42], Random Forest (RF) [43] and AdaBoost [44]. These supervised learning algorithms have applications in decision support systems [45]–[48], sentiment analysis [49], [50] and time series analysis [51]–[53]. To avoid over-fitting and for tuning predictive parameters, 10 fold cross validation method is used.

MLR based models have been used in load forecasting for decades, which are advanced forms of simple linear regression used to predict an output variable (y) based on multiple predictor variables (x) [39]. For example, with four predictor variables (x), the following equation is used to express the output variable (Y):

$$Y = b_0 + b_1 * x_1 + b_2 * x_2 + b_3 * x_3 + b_4 * x_4 \quad (4)$$

where “ b ” values are called beta coefficients (regression weights). They are used to measure the relationship between the predictor (x) and the output (Y).

k -NN has widely been used as non-parametric method in statistical estimation and classification for decades. It predicts any new numerical target point on the basis of a similarity measure (distance functions) i.e. the predicted point is calculated by averaging the k NN nearest neighbors of the numerical target in the training set.

SVR [41] is a popular machine learning method used for regression which is basically a modified version of SVMs [42]. SVMs aim to separate each data point into an n -dimensional feature space after which a hyper-plane is estimated to separate those data points in their corresponding classes. A SVR works on the same principles as support vector machine (SVM) does. SVM has many kernel functions. In this study, three popular kernels, SVR with linear kernel (SVRL), SVM with radial kernel (SVRR) and SVR with

polynomial kernel (SVRP) are employed for electricity load forecasting. The SVR models are developed by employing the SVM function of the ‘e1071’ package of R Language. This package has SVM implementation for regression problems namely SVR. Important parameters in this function are kernel, Vapnik’s insensitive loss function (ε), constrain violation (cost), tolerance of termination criterion, and γ . Linear, Polynomial and Radial are used as values for kernel parameter. ε is set to “0.1”, cost value is “1”, tolerance value of “0.001” and $\gamma = 1/(\text{data dimension})$. These are the typical values used to develop electric load forecasting models. The R Language implementation of SVR function automatically optimizes the other hyper-parameters for the given problem.

RF [43] is an ensemble or decision trees based learning algorithm for classification and regression tasks. It consists of several individual decision trees operating as an ensemble. These trees are trained individually and split out a class guess by searching randomly selected subset of given input variables. For classification task, the class label possessing the highest vote count among individual decision trees, becomes the model’s prediction. In regression, the mean of the individual decision trees becomes the model’s prediction.

Adaptive Boosting [44], in short AdaBoost, is a meta-algorithm which is used to improve the performance of model by conjunction of different weak classifiers. The predictions of the weak classifiers are joined into a weighted sum, representing the final outcome of the boosted learner. It is called adaptive because consequent weak learners are pulled towards those instances that were misclassified by preceding classifiers. In this article, Adaboost.RT algorithm [54] is used, which follows the same boosting procedure as that of [44] but in context of regression problems.

D. PERFORMANCE EVALUATION MEASURES

The performances of forecasting models can be evaluated using the mean absolute error (MAE), the root mean square error (RMSE) and the mean absolute percentage error (MAPE). These are widely used statistical measures for time series based models to measure their performances [51], [52]. These are defined as follows:

$$MAE = \frac{1}{N} \sum_{l=1}^N |\bar{y}_l - y_l| \quad (5)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{l=1}^N (\bar{y}_l - y_l)^2} \quad (6)$$

$$MAPE = \frac{1}{N} \sum_{l=1}^N \frac{|\bar{y}_l - y_l|}{y_l} \quad (7)$$

where N is the size of test samples, and \bar{y}_l and y_l are the estimated and actual electric load values, respectively.

E. COMPUTING ECONOMIC SIGNIFICANCE OF FORECASTING MODELS

MAPE is a commonly used statistical measure to determine the accuracy of ELF models. Improvement in the MAPE can be converted into a financial benefit (e.g. rupees or dollar amount). It was found in [3] that 1% improvement in the MAPE can decrease variable generation cost by around 0.1%-0.3% and with a conservative estimate, 1% decrease in the forecasting error can save approximately \$1.6 million annually for a 10,000 MWh utility. The economic significance can be estimated from MAPE using the following equations:

$$MAPE_{redu} = MAPE_{ELnMP} - MAPE_{EL} \quad (8)$$

$$AvgED_{MW/d} = AvgED_{MW/h} \times H_d \times MAPE_{redu} \quad (9)$$

$$AvgDS = AvgED_{MW/d} \times GC_{MWh} \quad (10)$$

where, $MAPE_{EL}$ is the MAPE of ELF model built using electric load data only, $MAPE_{ELnMP}$ is the MAPE of ELF model built using electric load data coupled with meteorological parameters (as described before), $MAPE_{redu}$ is the reduction in of forecasting error due to inclusion of meteorological parameters, $AvgED_{MW/d}$ is the average daily demand in megawatts $AvgED_{MW/h}$ is average hourly electricity demand in megawatts, H_d is the hours per day. In equation (10), $AvgDS$ is average daily saving and GC_{MWh} is the generation cost of per MWh of electricity. Once $AvgDS$ for each of the developed electricity load forecasting models using different machine learning algorithms are computed, the following proposed equation can be used to obtain the least cost ELF model.

$$lcELFM = \arg \min_i g(AvgDS_i | (TMLA_i, \min(MAPE_{ELnMP}))) \quad (11)$$

where $lcELFM$ represents the least cost electric load forecasting model, and $TMLA_i$ is the list of traditional machine learning algorithms. Equation (11) can be helpful for exploring least cost ELF model.

IV. PROPOSED ALGORITHM

In this section, proposed algorithm to select lcELFM is described as follows.

According to the best of our knowledge, before this study, no algorithm has yet been proposed to select the lcELFM among the available ELF models. In this section, the proposed algorithm to select lcELFM is described, which is one of the main contributions of this study. The algorithm starts by taking lists of original EL , $MPrams$, $TMLAs$, and GC_{MWh} . Here EL is the hourly electricity load demand, $MPrams$ are the meteorological parameters, $TMLAs$ is the list of traditional machine learning algorithms to build forecasting models, and GC_{MWh} is the generation cost per MWh of electricity, respectively. The algorithm returns $lcELFM$ which shows the least cost ELF model having minimum $MAPE_{ELnMP}$ value among the built ELF models as per provided $TMLAs$ using the details listed in the algorithm.

Algorithm 1: An algorithm to select *lcELFM* using correlated meteorological parameters

Input: (EL , $MPrams$, $TMLAs$, GC_{MWh}) where

- EL ; a time series of hourly electric load demand.
- $MPrams$; a list of meteorological parameters (hourly air temperature, dew point temperature, relative humidity, cloud cover, wind speed, pressure, rainfall and participation).
- $TMLAs$; a list containing traditional machine learning algorithms.
- GC_{MWh} ; generation cost of electricity in MWh

Output: Selected least cost forecasting model (*lcELFM*) built using $TMLAs$, and Eq. (11).

1. Repeat steps 2 to 3 for each value a in EL , where $a := 1$ to m
2. if ($EL[a] == NA$) ▷ Value is missing
3. $EL[a] = \frac{EL[a-2]+EL[a-1]+EL[a+1]+EL[a+2]}{4}$ ▷ Fill missing value using Eq. (1)
4. $AvgED_{MWh} := average(EL_{MWh})$ ▷ Compute average EL demand per 24hrs
5. $CorrMPrams = findCorrParams(MPrams)$ ▷ Find Correlated Meteorological parameters as per criteria defined in section "WEATHER LOAD RELATIONSHIP"
6. $\chi = combine(EL, CorrMPrams)$ ▷ Combine EL and $MPrams$ time series
7. Split data EL into k folds i.e. $ELD_k = EL_{k=1}^n$ ▷ Electric load demand (EL) data split for k -Fold CV
8. Split data χ into k folds i.e. $ELnMPD_k = \chi_{k=1}^n$ ▷ EL and $MPrams$ time series data split for k -Fold CV
9. Repeat steps 10 to 13 for each fold k , where $k := 1$ to n
10. $ELDtr_k := ELD_{n-1}$ ▷ train EL set
11. $ELDts_k := ELD_n$ ▷ test EL set
12. $ELnMPDtr_k := ELnMPD_{n-1}$ ▷ train $ELnMP$ set
13. $ELnMPDts_k := ELnMPD_n$ ▷ test $ELnMP$ set
14. Repeat steps 15 to 18 for each $TMLA_c$ where $c := 1, 2, 3, \dots, j$
15. $ELFModel_{k,TMLA_c} := buildFModel(ELDtr_k, TMLA_c)$
16. $AllAlgosELFModelTestResults[k, TMLA_c] := testFModel(ELFModel_{k,TMLA_c}, ELDts_k)$
17. $ELnMPFModel_{k,TMLA_c} := buildFModel(ELnMPDtr_k, TMLA_c)$
18. $AllAlgosELnMPFModelTestResults[k, TMLA_c] := testFModel(ELnMPFModel_{k,TMLA_c}, ELnMPDts_k)$
19. Repeat steps 20 to 24 for each $TMLA_c$ where $c := (1, 2, 3, \dots, j)$
20. $MAPEsELFMsAllAlgos[c] := MAPE(AllAlgosELFModelTestResults[1 : n, TMLA_c], EL)$
▷ Computation of MAPE based on original (EL data) and predicted values from ELF models as per Eq. (7)
21. $MAPEsELnMPFMsAllAlgos[c] := MAPE(AllAlgosELnMPFModelTestResults[1 : n, TMLA_c], \chi[1 : n, EL])$
▷ Computation of MAPE based on original (EL data) and predicted values from $ELnMPF$ models as per Eq. (7)
22. $MAPEs_{redu}[c] := MAPEsELnMPFMsAllAlgos[c] - MAPEsELFMsAllAlgos[c]$ ▷ Computation as per Eq. (8)
23. $AvgED_{MWh/d}[c] := AvgED_{MWh} \times 24 \times MAPEs_{redu}[c]$ ▷ Computation as per Eq. (9)
24. $AvgDS[c] := AvgED_{MWh/d}[c] \times GC_{MWh}$ ▷ Computation as per Eq. (10)
25. $lcELFM := ELnMPFModel_{TMLA[1]}$
26. Repeat steps 27 to 28 for each $TMLA_c$ where $c := (2, 3, \dots, j)$ ▷ As per criteria set in Eq. (11)
27. if ($(AvgDS[c] < AvgDS[c-1]) \& (MAPEsELnMPFMsAllAlgos[c] == \min(MAPEsELnMPFMsAllAlgos))$)
28. $lcELFM := ELnMPFModel_{TMLA[c]}$
29. Return $lcELFM$

V. SIMULATION DESIGN

To investigate the usefulness of proposed algorithm, the obtained electric load and meteorological datasets are first checked for missing values. Missing values are filled by using equation (1). As a second step of simulation design, the correlation between available meteorological parameters and electric load is computed by Pearson's correlation analysis (Refer to equation (3)). Correlated dataset is then used in two settings (Setting-I and Setting-II). Setting-I corresponds to the electric load dataset while Setting-II contains dataset having electric load and correlated meteorological parameters. For both of the settings, the datasets are divided into

10 equal folds for cross validation purposes. The divided data is alternatively provided to machine learning algorithms (MLR, k NN, SVML, SVMR, SVMP, RF and AdaBoost) to build predictive models as per the notion of well know k -fold cross validation technique. The built models are then tested using test dataset to have predicted values of electric load. The obtained electric load values are used to compute error measures (MAE, RMSE and MAPE). Using equations (8), (9) and (10), the economic significance of the predictive models is computed. Equation (11) is used to select the least cost electric load forecasting model. The whole process is illustrated in the Fig. 1.

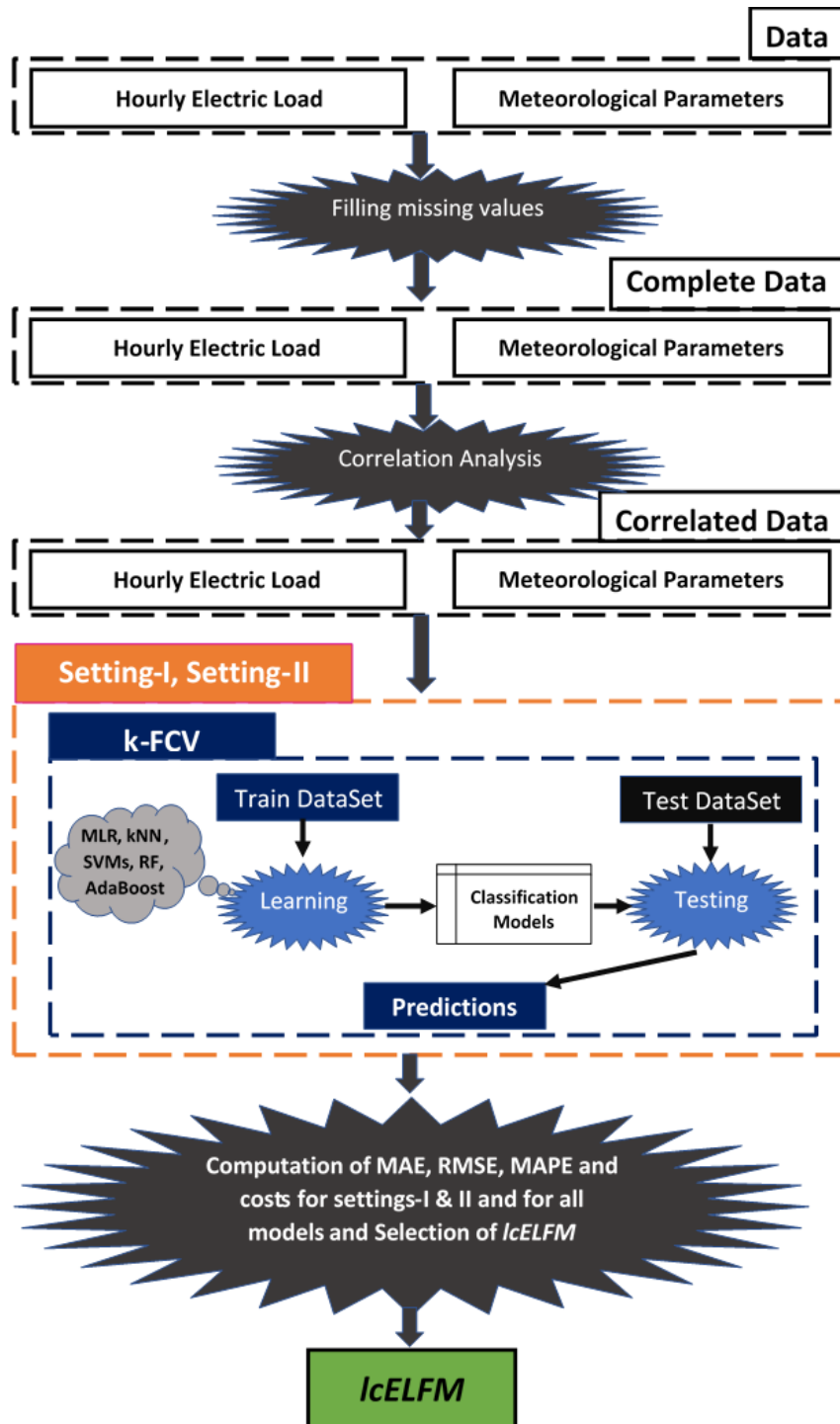


FIGURE 1. Simulation design for the selection of ML based cost effective electricity load forecasting model using correlated meteorological parameters.

VI. CORRELATION ANALYSIS RESULTS

As discussed before, according to scientific literature, electricity load demand relies on the human reactions to weather conditions. Therefore, in order to quantify the relationship between meteorological variables (temperature, relative humidity, dew point, cloud cover, rainfall and wind speed) and electric load demand, scatter plots are drawn on annual

and seasonal bases and corresponding correlation estimates are computed.

A. HOUR TO HOUR LOAD TEMPERATURE RELASHIONSHIP

In order to analyze the hour to hour correlation between load and temperature on yearly and seasonal basis, the scatter plots

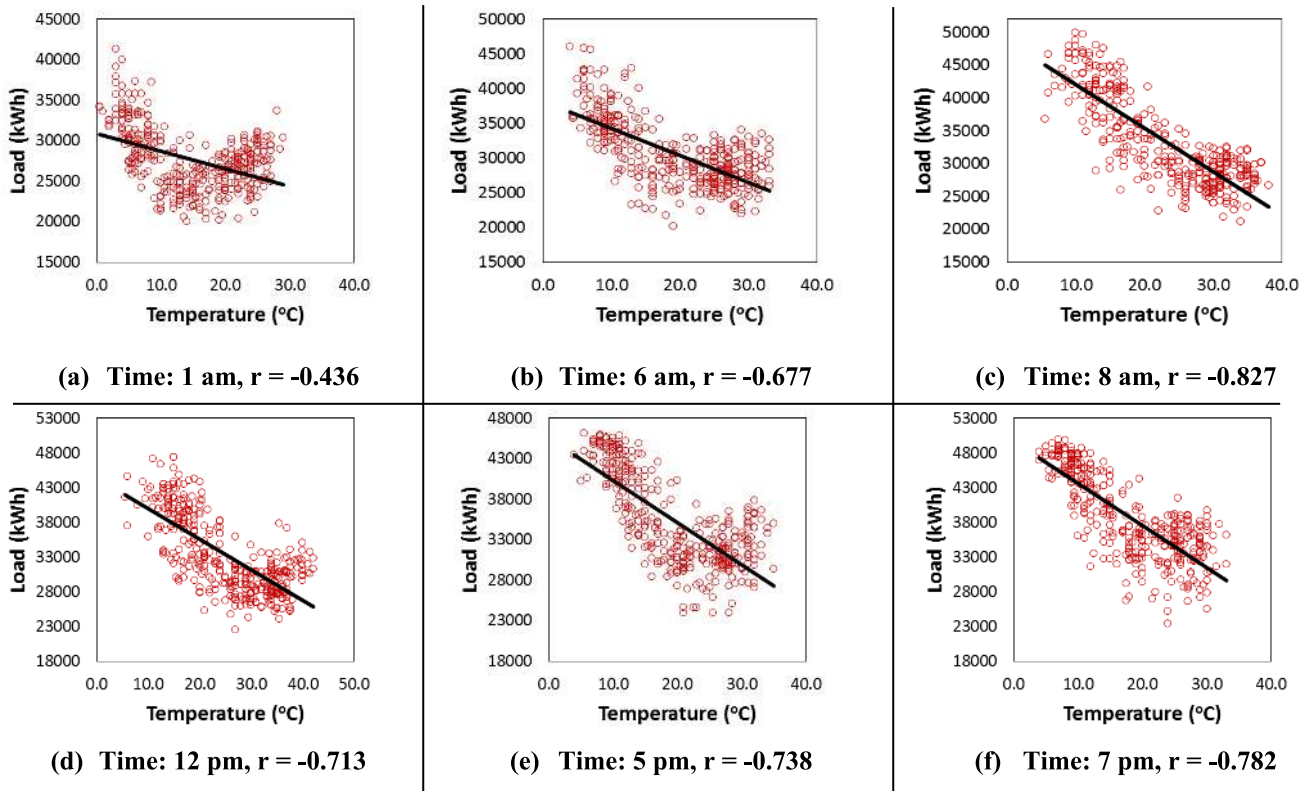


FIGURE 2. Load vs temperature correlation of the most crucial times of the day (hourly basis).

with linear trend lines are generated for the peak hours of the day. Fig. 2 shows the hour to hour correlation on yearly basis for 1am, 6am, 8am, 12pm, 5pm and 7pm. The subgraphs in Fig. 2, other than subgraph for 1am, show the readings at five different peak hours.

Power demand tends to increase in the morning, when people usually start their day and similarly secondary peaks are observed in the evening, when most of the people return home. Annual hour to hour scatter plots of the most significant hours (Fig. 2) of Muzaffarabad during the study period resembled this trend. Higher correlation exists between load and temperature in the morning at 6am and 8am (Fig. 2(b), 2(c)) and in the evening at 5pm and 7pm (Fig. 2(e), 2(f)) due to maximum activities being taken place during these hours. Simply, all the times of the day (Fig. 2 (a-f)) clearly indicated a strong inverse linear relationship between load and temperature, which shows notable response of the residents of Muzaffarabad to temperature changes during prominent hours.

B. SEASONAL HOUR TO HOUR LOAD TEMPERATURE RELATIONSHIP

The relationship between electric load and temperature can be defined in a better way on seasonal time scales. In the Figs. 3-6, hour to hour relationship between electric load and temperature is shown (only for important time slots of

the day) on seasonal basis. Important time slots include before sun rise, morning, noon, afternoon and evening. Hours with maximum correlation are plotted for each important time slots of the day which can vary for each season. For instance, plots at 1am, 5am, 9am, 12pm, 5pm and 7pm are shown for summer season whereas plots at 1am, 6am, 8am, 12pm, 5pm and 7pm are shown for spring season because corresponding hours have maximum correlation. (In order to preserve maximum possible details of the plot, the x-axis and y-axis scales of the scatter plots are altered).

Fig. 3 (a-f) shows correlation of temperature against electric load at 1am, 6am, 8am, 12pm, 5pm and 7pm for spring season. Inverse linear correlation is observed for all time slots. The correlation is small during predawn hours i.e. at 1am and 6am (Fig. 3 (a-b)) but as majority of the citizens wakeup and begin a combination of routine activities, the correlation become stronger at 8am (Fig. 3 (c)). This is due to the gradual rise of temperature from uncomfortable levels of late winter temperatures to pleasant early spring temperatures causing a decline in electric load demand.

Fig. 4 (a-f) shows correlations of temperature and electric load at 1am, 5am, 9am, 12pm, 5pm and 7pm for summer season. This season exhibited different and a nonlinear correlation for corresponding hours as compared to other seasons. A small positive correlation is found for all hours in this setting. The reason is that the increase in temperature during summer and hence the use of electric equipment for

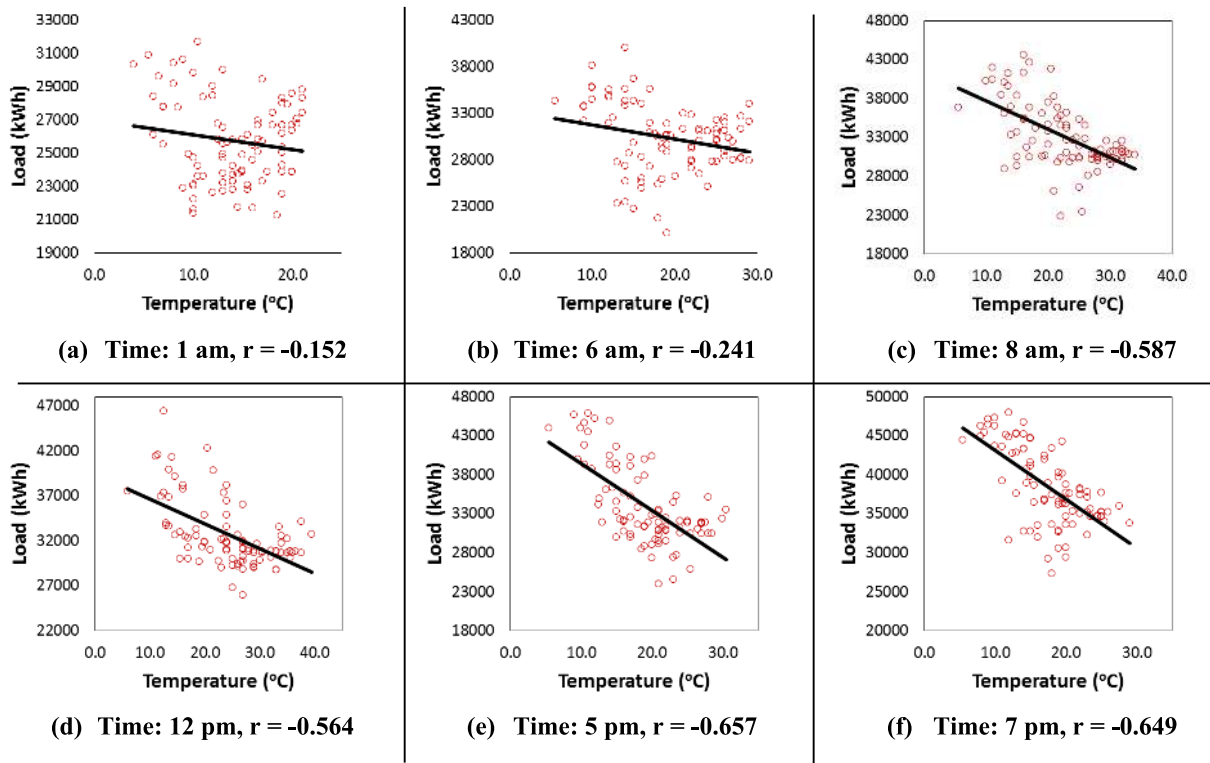


FIGURE 3. Load vs temperature correlation of the most peak times of the day (spring season).

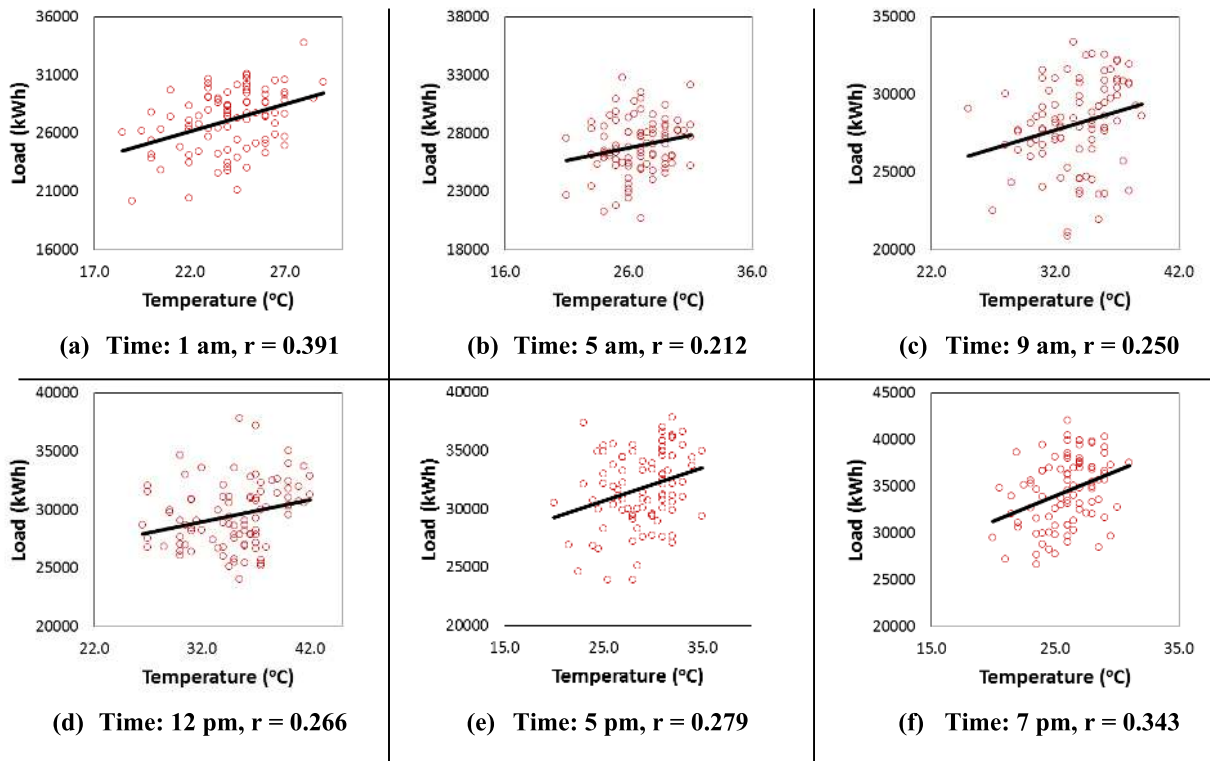


FIGURE 4. Load vs temperature correlation of the most crucial times of the day (summer season).

cooling purposes triggered the increase in electric load demand. However, the correlation for this season was not much impressive because the temperature of Muzaffarabad is

not as higher as it is in other parts of the country and pleasant climate conditions restrict citizens not to show abnormal trends in the summer.

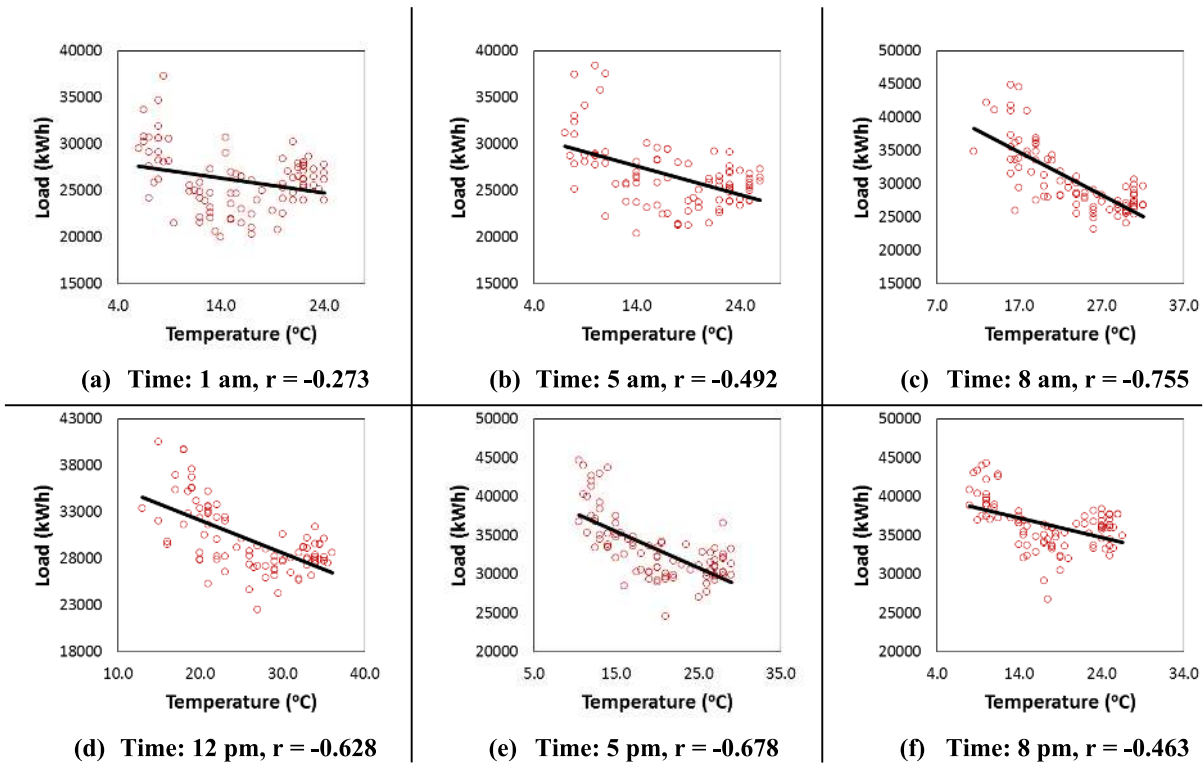


FIGURE 5. Load vs temperature correlation of the most crucial times of the day (fall season).

Fig. 5 (a-f) shows the correlation between load and temperature at 1am, 5am, 8am, 12pm, 5pm and 8pm for fall season. In this figure, the last peak hour in the evening changes from 7pm to 8pm. The possible reasons are: (1) during fall season, the temperature starts decreasing in the evening as compared to summer; (2) the sun sets later in the evenings as compared to winter and the citizens engaged later in their indoor activities. An inverse strong relationship is developed during these hours. The correlation is well-defined for 5am (Fig. 5 (b)) in the morning (when most of the residents wake up for their day) to 8pm at night (when most of the residents are involved in the variety of post evening activities after returning home). The highest correlation is found at 12pm (Fig. 5 (d)) because, the temperature remains comfortable during midday hours in the fall season which results into a decreased load demand, however the fall values of other prominent hours in the scatter plot mimicked that of 12pm. The correlation between pleasant fall temperatures and corresponding loads of the prominent hours (Fig. 5 (b-f)) is comparable to those of spring season (Fig. 3 (c-f)).

Fig. 6 (a-f) shows the correlation between load and temperature at 1am, 6am, 8am, 12pm, 5pm and 8pm for winter season. Muzaffarabad is known for its frigid temperatures in the winter so, elevated levels of load consumption are seen during this season. An inverse relationship between load and temperature is observed for all hours, which is most noticeable at 8pm. This is because of the combined effect of dropping temperature and rush oriented post evening routines of the residents. People struggle against cold temperatures of

the winter throughout the day, so handsome correlation exists between load and temperature. Usage of heating equipment is a major part of the customer’s adjustment to uncomfortable temperatures and hence the load demand increases. This adjustment is clearly reflected in the scatter plots for all hours in Fig. 6 (a-f).

C. LOAD DEW POINT RELATIONSHIP

Dew point, relative humidity and air temperature are reported together because these variables are related with each other i.e. when dew point, temperature and air temperature are equivalent, the relative humidity would be 100% and fog, clouds or frost are created depending upon the season. However, dew point is more accurate measure of “how humid the air is”, because it is an absolute measure whereas the relative humidity is a relative measure and can sometime give misleading results [8], [9].

Dew normally appears during the clear sky, when the surface emits heat and become cooler as there are no clouds to trap the heat. The emission of radiations from earth’s surface cools down the air more rapidly overnight, elevating its chances to drop under the dew point. Heavy dew becomes more common in the mornings of fall season. It is considered as the peak dew season among all because the air is normally cool enough to drop under the dew point temperature, but not enough to make frost. Fall season of Muzaffarabad is known for its clearer and calmer nights that are important for cooling and surface radiation which causes surface temperature to drop under the dew point temperature.

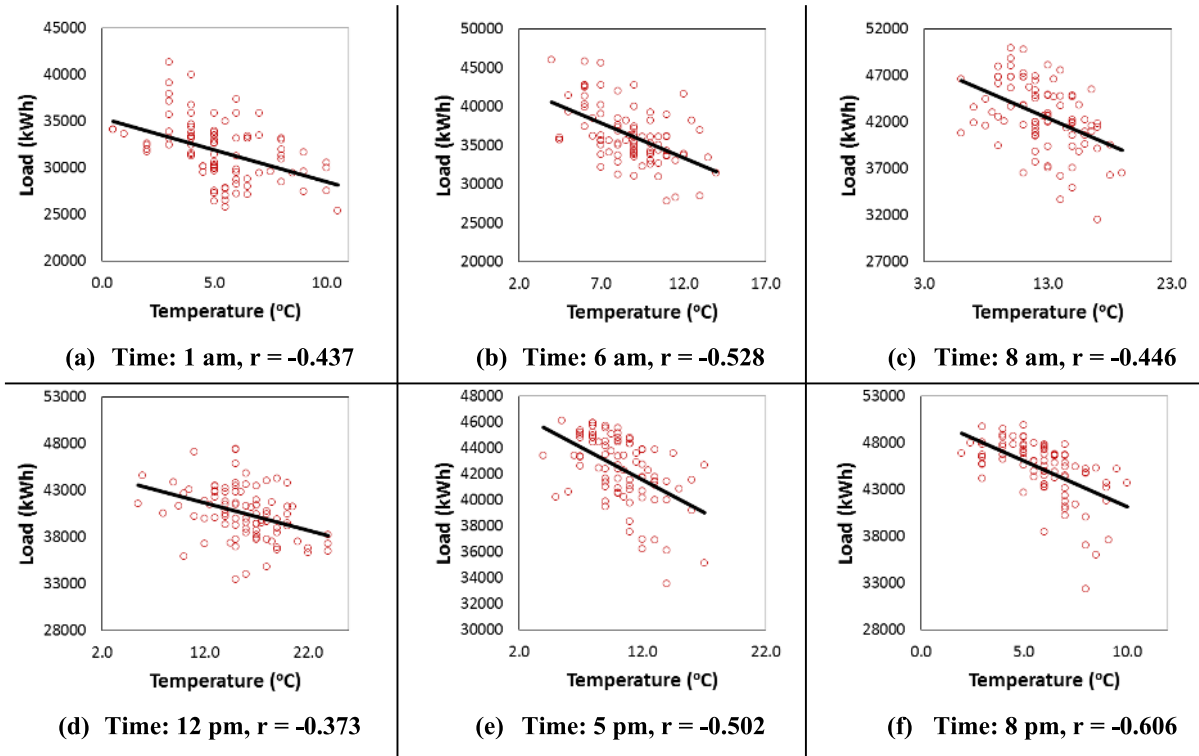


FIGURE 6. Load vs temperature correlation of the most crucial times of the day (winter season).

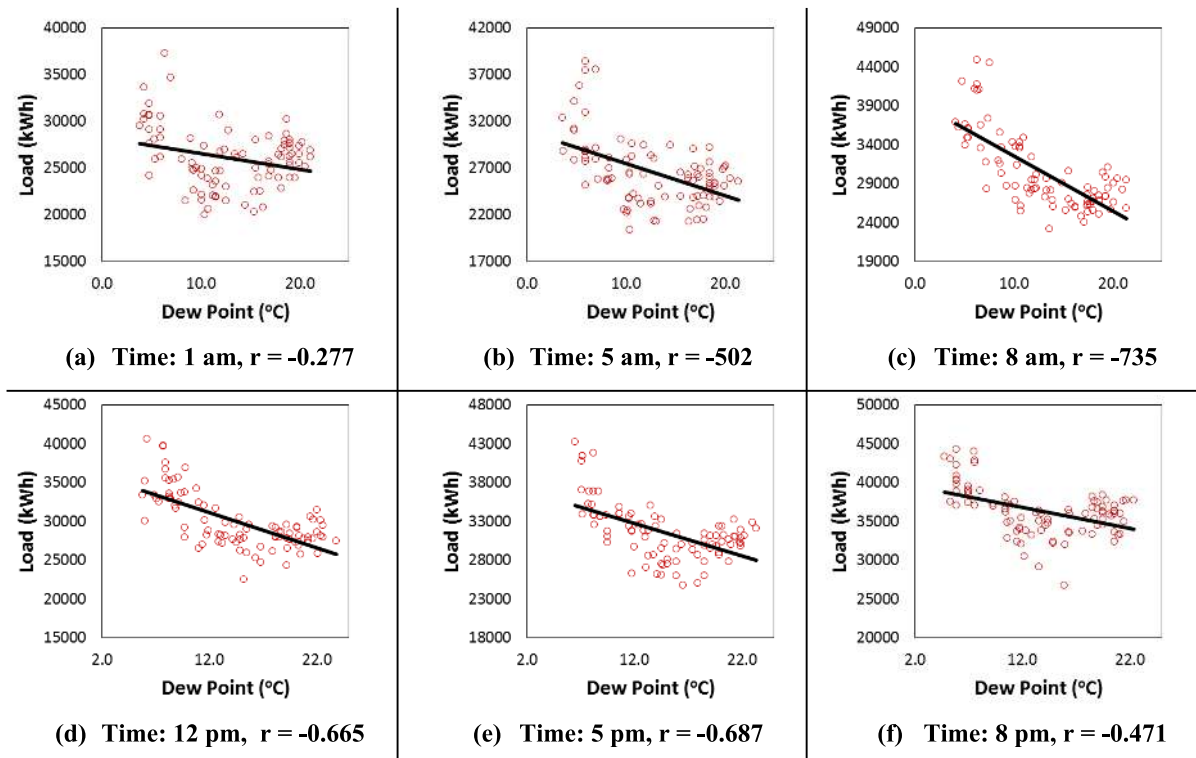


FIGURE 7. Load vs dew point correlation of the most crucial times of the day (fall season).

Fig. 7 (a-f) shows the correlation between load and dew point at 1am, 5am, 8am, 12pm, 5pm and 8pm for fall season. The air temperature normally reaches down to the lowest

value from the midnight at 12am to the morning around sunrise, so during this time, it is more likely that the dew point temperature is reached. As a result of this phenomenon,

correlation between load and dew point began to develop at 1am (Fig. 7 (a)) but this is not good enough because most of the residents are sleeping at midnight and load consumption prone activities are rear during this time. As the sunrise approaches (Fig. 7 (b)), the correlation becomes stronger and stronger because the overnight decreased temperature is nearly equivalent to the dew point temperature at this time and the highest relative humidity is there. The relationship reaches its highest value at 8am (Fig. 7 (c)) because majority of the residents have awoken and begin their day through a combination of activities and people respond to dew point temperature by increasing the load demand at this time. The inverse relationship is the combined effect of customer's post dawn activities and dew point temperature, which results into an increase in the load demand. As the day proceeds after the sun rise, the dew point temperature moves away from the air temperature and the correlation between dew point temperature and load demand decreases (Fig. 7 (d-f)). For example, the highest customer's activities are also seen at 8pm as these are seen at 8am but the correlation is not that high as it is at 8am.

D. LOAD RELATIVE HUMIDITY RELATIONSHIP

The correlation between electric load and other meteorological variables is also explored. Relative Humidity is one of them, having significant relationship with electric load demand. It is defined as "*the amount of water vapors present in the air expressed as a percentage of the amount needed for saturation at the same temperature*". Higher humidity during the hotter months results into hotter feelings as compared to the actual temperature. Fig. 8 (a) shows the scatter plot of relative humidity verses electric load demand for the whole year of 2014. The figure revealed a small positive correlation between relative humidity and load demand. Summer is the most relevant season for relative humidity because of higher temperatures. So, scatter plots of relative humidity and temperature are also shown in Fig. 8 (b) along with scatter plots of relative humidity and load demand Fig. 8 (c) for summer season. The following inferences can be drawn from the figures.

During the summer season, the electric load decreased as the relative humidity increased (Fig. 8 (c)). Because, an inverse relationship exists between temperature and humidity (Fig. 8 (b)) during the summer season, so when temperature increased, humidity decreased and in turn the load demand increased. Whereas, the electric load demand seemed independent of relative humidity during the winter season (Fig. 8 (d)). This is so because in winter season, wind speed and wind chill are more relevant to load vagaries as compared to relative humidity. Spring and fall seasons do not show any relevancy to load demand and resembled the similar winter trend.

E. LOAD CLOUD COVER RELATIONSHIP

Cloud cover is another important variable of meteorological data which is related to solar radiation. It is defined

by three aggregated cloud layers i.e. low clouds, medium clouds and high clouds depending upon the height from the earth's surface. Measurement of cloud cover is a difficult task. Satellite cameras, ground sensors and visual observation are commonly used methods for this purpose each has its own limitations. Muzaffarabad meteorological department uses a scale from 0 to 8 for cloud determination where 0/8 means no clouds at all and 8/8 means complete sky cover.

So widespread scatters are found for transitional seasons (not shown) that the load demand seemed nearly independent of cloud cover however, small to medium level correlation existed between load and cloud cover for some hours of different seasons which are narrated here.

Cloud cover plays an important role for blocking unpleasant solar radiation striking the earth in the summer season. Solar radiation reaches to maximum at solar noon. In summer season, from 1pm to 4pm are the most relevant hours for load-cloud cover relationship because solar radiation is at its maximum during these hours if clouds are not there. Therefore, small inverse correlation between load and cloud cover is found from 2pm to 5pm ($r = -0.12312$, $r = -0.29609$, $r = -0.22755$ and $r = -0.12553$ respectively). The reason is obvious. As the clouds started disappearing, uncomfortable effect due to solar radiation increased, and in turn the load demand increased.

The highest correlation is found at 1pm during springtime ($r = 0.33834$) and prior two hours showed medium correlation i.e. $r = 0.32204$ at 12pm and $r = 0.29616$ at 11am whereas from 1pm onwards the correlation ranged 0.1 to 0.3. Perhaps, solar radiation was not strong enough to create the inverse correlation from pleasant early to 2/3 of spring season even in the absence of clouds, therefore positive correlation is reported.

Clear sky during the fall season results in dew points at nights which create cold conditions as compared to the sky covered with clouds. Therefore strong inverse load-cloud cover relationship is found from 8pm to 5am (r values between -0.2 to -0.5) with the highest value occurred at 10pm ($r = -0.4679$) when residents are involved in closing activities of the day. The same damp conditions occurred for winter season in the absence of clouds at night, therefore inverse relationship existed during the night hours of the winter but it was not well pronounced as it is for fall season and ranged from -0.1 to -0.2 .

F. LOAD WIND SPEED: RAIN FALL RELATIONSHIP

The relationship between electric load and wind speed, electric load and rain fall is not significant during any season. Inconsistent and non-apparent correlation (not shown here) was found between the variables. However, correlation from 0.1 to 0.3 exist for some random hours of the day for all seasons, for example a correlation of 0.312773 is noticed in load-wind speed relationship at 5pm during summertime. So, the independent consideration of wind speed and rain fall variables for electric load forecasting seems to be of

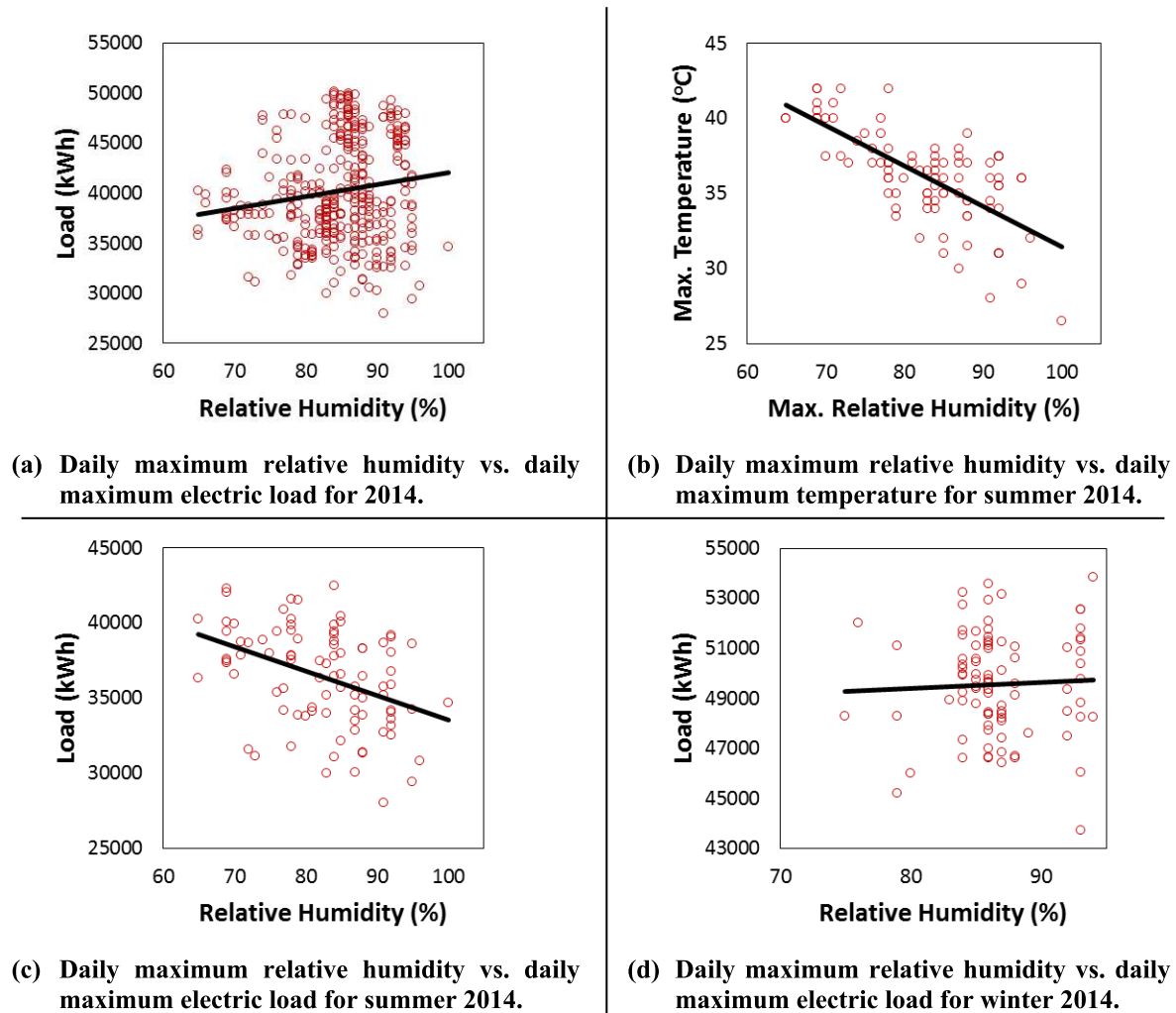


FIGURE 8. (a) Correlation between daily maximum load and maximum relative humidity during the whole year 2014. (b): Daily maximum relative humidity vs daily maximum temperature for summer season 2014. (c): Daily maximum relative humidity vs daily maximum load for summer season 2014, and (d): Daily maximum relative humidity vs daily maximum load for winter season 2014.

no use. However, these may be useful when used with the combination of other variables like wind chill.

VII. ELECTRIC LOAD FORECASTING RESULTS

Two model simulations, one with electric load variable only and other with meteorological parameters coupled with electric load variable are performed in order to estimate the null hypothesis i.e. “*relationship between weather variables and electricity load can improve the electricity load forecasting for Muzaffarabad city*”. Meteorological parameters that showed strong relationship with the electric load demand in the correlation analysis are integrated into a forecasting model to access the accuracy improvement. Hourly observations of the corresponding variables are used to create two ELF models by considering following two settings:

Setting I: Use only electric load demand variable to build electric load forecasting model (ELFModel).

Setting II: Electric load demand variable coupled with meteorological parameters to build forecasting model termed as ELnMPFModel.

For both of the settings ML algorithms MLR, k NN, SVRL, SVRR, SVRP, RF and AdaBoost are used to build and test both types of ELF models. The results of the modeling studies are used to determine which combination of variables i.e. setting-I or setting-II gives accurate forecasting.

The performances of both types of models are evaluated using the RMSE, MAE, and MAPE.

Short term load forecasting (24 hours ahead) is performed with both types of models. The outputs from each model are compared with the original observations to estimate MAPE, RMSE and MSE over the specified forecasting period. The model with the lowest evaluation values is considered better. Previous research studies showed that many exogenous variables like day of the week, month, holidays, year etc. are essential to load forecasting [8]. So, these are used in both settings (I and II) during the training phase to develop forecasting models. Other model specifications are listed in Table 2. This table shows common variables of the models (which are used in both models) and the list of variables used to build models for both of the settings (I and II).

TABLE 2. Exogenous variables used in settings (I & II).

List of exogenous variables
Years (2014, 2015)
Months (January to December)
Day of the week (Monday to Sunday)
Holidays (0 indicates holiday and 1 indicates working day)
PREDICTOR VARIABLES USED IN SETTING I
Specifications list of predictor variables for model 1 (only load variables)
Years (2014, 2015)
Months (January to December)
Day of the week (Monday to Sunday)
Holidays (0 indicates holiday and 1 indicates working day)
Days (1 to 31)
Previous Load (2014, 2015)
PREDICTOR VARIABLES USED IN SETTING II
Specifications list of predictor variables for model 2 (only weather variables)
Years (2014, 2015)
Months (January to December)
Day of the week (Monday to Sunday)
Holidays (0 indicates holiday and 1 indicates working day)
Days (1 to 31)
Hourly load
Hourly temperature
Hourly dew point
Hourly relative humidity
Hourly cloud cover

Table 3 shows the actual values of electric load and the predicted values by ELFModel (referred as M1 in Table 3) and ELnMPFModel (referred as M2) used in this study. Tabulated values (for twenty-four hours ahead forecasting) show that ELnMPFModel based values are in close agreement with actual electric load values than those produced by ELFModels.

Tables 4 shows the performance indexes of both types of models built using Setting-I and Setting-II. It can be seen from the Table 4 that MAPE of the ELFModel is high (MAPE = 14.6) as compared to ELnMPFModel (MAPE = 9.2) when MLR is used as learner. High MAPE of 7.6 is reported in ELFModel as compared to 6.1 of ELnMPFModel based on *k*NN learning algorithm. Moreover, the MAPE of ELFModel is 17.5, whereas that of ELnMPFModel is 10.3 for SVR Linear. For SVR Radial, the ELnMPFModel has better MAPE of 5.6 versus 7.8 of ELFModel. Likewise, SVRP, RF and AdaBoost based ELnMPFModels have MAPEs of 8.39, 5.71 and 8.43 respectively, whereas the ELFModels achieve 8.99, 7.56 and 10.51 MAPEs values respectively. Similarly, for other evaluation measures i.e. RMSE and MSE, the ELnMPFModels significantly reduced the forecasting error for all machine learning algorithms used.

MAPE can also be used as the accuracy measure. Considering this, the improvement in the prediction accuracy given by the ELnMPFModels is 5.34%, 1.48%, 7.21 %, 2.22%, 0.6%, 1.85% and 2.08% for MLR, *k*NN, SVRL, SVRR, SVRP, RF and Adaboost respectively.

Table 4 shows the reduction of error for one day ahead load demand for MLR model. Compared with

ELFModel, the forecasting framework with the ELnMPFModel decreases the prediction error significantly by 36.64% of MAPE, 27.33% of RMSE and 47.19% of MSE respectively. Table 4 also shows the forecasting results of *k*NN based models. On the same data set, the ELnMPFModels achieve the error reductions of 19.63%, 12.41% and 23.28% for MAPE, RMSE, and MSE respectively. The results of ELFModel and ELnMPFModel using SVRL are presented in the Table 4. Clearly, the ELnMPFModel decreases the forecasting error by 41.26% of MAPE, 31.71% of RMSE and 53.36% of MSE, respectively. The results based on SVRR show the significant decrease in the forecasting error by the ELnMPFModels as compared to ELFModels by 28.33%, 22.08%, and 39.29% for MAPE, RMSE and MSE respectively. Table 4 indicates error reductions of the ELnMPFModel by 6.67% of MAPE, 7.85% of RMSE and 15.13% of MSE, when SVRP is used. The results of RF are also shown in Table 4, where the ELnMPFModel reduces the errors by 24.47%, 14.84% and 27.48% for MAPE, RMSE and MSE respectively. Moreover, the results of AdaBoost based ELFModel and ELnMPFModel also show significant reductions of the ELnMPFModel as compared to ELFModel, namely 19.79%, 20.32% and 36.50% for MAPE, RMSE and MSE respectively.

It can be seen from the Table 4 that the model errors are reduced, when ELnMPModels are used (as compared to using load data alone) in all cases. Big differences in the error measure are reported by the results. In case of load variables, the errors are inconsistent and possess higher values for simulated forecasting period. However, two exogenous variables, weekday and IsWIsH (is working day or holiday) did not resemble a statistical significance from other variables including exogenous, load and weather variables i.e. the load is independent of these two variables.

The simulated results revealed that the electricity use at Muzaffarabad responded to temperature and dew point temperature more than other two metrological parameters because these are greatly related to human comfort. The strongest relationship is shown between load and these variables as compared to others in the correlation analysis. The model learned the load-temperature and load-dew point temperature relationship more effectively compared with the other weather and exogenous variables. The results uncovered the benefit of using temperature and dew point in the simulated models, however relative humidity and cloud cover also looked beneficial in some studies and the model can still capable of creating an acceptable forecast when equipped with them but not for the region under consideration in this study.

VIII. RESULTS OF ECONOMIC SIGNIFICANCE

The economic impact of ELnMPFModel is quantified by computing a rupee amount saved by WAPDA if the MAPE could be reduced in the load forecasts. These economic benefits should be portable across the Azad Kashmir.

TABLE 3. Actual and forecasted values of both models: ELFModel (M1) and ELnMPFModel (M2) using different machine learning methods.

Actual	MLR		kNN		SVRL		SVRR		SVRP		RF		AdaBoost	
	M1	M2	M1	M2	M1	M2	M1	M2	M1	M2	M1	M2	M1	M2
32112	31477	32919	31763	30485	29621	32827	31985	30430	31812	26602	31773	30040	31833	31904
33491	31807	33862	32211	31093	29986	33473	30647	29857	32615	29417	32228	31143	31833	31904
34526	32137	34215	31573	31125	30351	33860	30631	30161	33469	31764	31549	31141	31833	31904
39310	32467	34568	33280	33222	30717	34247	32172	31900	34360	34065	33268	35287	31833	31904
36780	32796	35869	33837	34120	31082	34178	35023	34449	35273	35307	33755	35709	31833	31904
41509	33126	38652	38868	40868	31448	35436	38506	38921	36193	39072	38844	35816	40881	40721
43524	33456	38715	44014	43090	31813	35775	41761	41192	37108	39282	43987	39675	43902	43730
44513	33786	38777	45705	41325	32179	36115	44051	41671	38001	39009	45693	38555	42924	42555
41358	34116	38594	44518	41567	32544	36399	44999	40223	38859	38228	44534	43993	40822	39433
41638	34446	38929	45069	41456	32909	37879	44658	39651	39667	39771	45094	41691	35402	36241
40625	34775	39531	43311	42420	33275	37920	43426	40046	40411	39442	43318	41643	35524	36241
39235	35105	40897	42682	43480	33640	39212	41888	41010	41077	41876	42675	42539	40891	39433
38858	35435	41500	41665	41010	34006	39678	40653	40595	41649	42737	41619	41153	42009	40823
39528	35765	41529	41244	40309	34371	39208	40227	39636	42115	41938	41228	38965	42993	42555
39332	36095	41371	40645	39158	34736	38848	40893	39044	42459	41413	40673	39247	43847	43730
39720	36425	40736	40475	38630	35102	38543	42598	39311	42666	40955	40485	39708	43134	43322
40388	36755	42374	42787	39597	35467	39300	44897	40988	42723	41434	42760	40149	43832	43730
45862	37084	42255	47487	41048	35833	39341	47025	42198	42615	40686	47486	41781	43573	43498
44828	37414	43120	48333	41046	36198	40043	48125	43839	42328	40480	48348	41759	43873	43730
46994	37744	42655	48456	42683	36564	39793	47582	43380	41847	39236	48452	42703	42668	42895
43826	38074	42536	45684	38705	36929	39835	45299	41099	41159	37619	45678	41720	42668	42895
34451	38404	42614	41198	38149	37294	40246	41775	38535	40247	36296	41200	35490	42439	42754
32942	38734	42179	37879	33269	37660	40288	37917	33564	39099	33825	37875	32073	35999	36058
28103	39063	42567	35695	31153	38025	40674	34691	29143	37700	31017	35694	29164	35999	36058

Azad Kashmir region has the load demand of around 400 MWh. Considering, mean demand of Muzaffarabad reaches 50 MWh and assume the mean cost of electricity is \$70/MWh (considering the cost of hydropower projects as Muzaffarabad receives electricity from these resources). Table 4 shows that the best prediction is given by SVRR with 2.2% reduction in MAPE for 1-day-ahead forecasting. Total daily savings of WAPDA in Muzaffarabad using ELnMPF-Models and one day-ahead load forecasts are calculated using equation (5) and 6 as below:

$$\begin{aligned}
 50\text{MWh/h} \times 24\text{h/d} &= 1200\text{MW/d} \times 2.22\%(\text{MAPE}) \\
 &= 2664\text{MWh/d} \times \$70/\text{MWh} = \$1865/\text{d} \\
 &= \text{PKR}0.313\text{M/d}.
 \end{aligned}$$

Reductions in MAPE with corresponding reductions in cost due to ELnMPFModels using other ML algorithms are shown in Table 5 (calculated by equation (5) and (6)).

Forecasting models built using different ML algorithms give different reductions in MAPE, hence different reductions in cost, because of their own performance limitations and advantages but one thing is certain that each model shows

MAPE reduction by including weather parameters in the forecasting models which leads to cost reduction. However, the cost reduction of the method with the lowest MAPE can be considered economical and effective for the utilities. Although, SVRL based ELnMFModel has the maximum reduction in MAPE and maximum reduction in cost of forecasting model but here the proposed algorithm selects that ELnMPFModel which is built using SVRR learning algorithm because Table 5 shows that SVRR based ELnMPF-Model is the one that has least MAPE value. Least MAPE value means that forecasting model has least forecasting error and hence the most AELF model. In this case reduction in cost is 0.313 million PRS per day for the utility providing electricity to the City of Muzaffarabad. Therefore, the consideration of significant meteorological parameters in the electric load forecasting models due to their influence on the electric load demand can save WAPDA around RKR. 0.313M daily in Muzaffarabad. As, WAPDA is responsible for supplying electricity to whole country and larger cities need a huge amount of energy than Muzaffarabad, so, if these findings are applied to the operational decisions of the WAPDA across

TABLE 4. Performance indexes of settings (I & II) for different ML models.

RESULTS OF MLR MODELS				
Error Measures	ELFModel	ELnMPF Model	Differences	Error Reduction
MAPE	14.57	9.23	5.34	36.64%
RMSE	6436.07	4677.23	1758.84	27.33%
MSE	41422978	21876437	19546541	47.19%
RESULTS OF kNN MODELS				
MAPE	7.55	6.07	1.48	19.63%
RMSE	3366.13	2948.49	417.64	12.41%
MSE	11330825	8693570	2637255	23.28%
RESULTS OF SVRL MODELS				
MAPE	17.46	10.26	7.21	41.26%
RMSE	7497.02	5120.05	2376.967	31.71%
MSE	56205279	26214929	29990350	53.36%
RESULTS OF SVRR MODELS				
MAPE	7.84	5.62	2.22	28.33%
RMSE	3524.97	2746.58	778.39	22.08%
MSE	12425438	7543721	4881717	39.29%
RESULTS OF SVRP MODELS				
MAPE	8.99	8.39	0.6	6.67%
RMSE	4115.66	3791.53	324.13	7.85%
MSE	16938655	14375682	2562973	15.13%
RESULTS OF RANDOM FOREST MODELS				
MAPE	7.56	5.71	1.85	24.47%
RMSE	3370.72	2870.55	500.17	14.84%
MSE	11361740	8240061	3121679	27.48%
RESULTS OF ADABOOST MODELS				
MAPE	10.51	8.43	2.08	19.79%
RMSE	4911.04	3913.34	997.7	20.32%
MSE	24118359	15314691	8803668	36.50%

TABLE 5. Results of cost reductions using settings (I & II) for different ML models.

Method	MAPE		Reduction in MAPE	Reduction in Cost (Million RS)
	ELFModel	ELnMPFModel		
MLR	14.57	9.23	5.34	0.754
kNN	7.55	6.07	1.48	0.209
SVRL	17.46	10.26	7.21	1.017
SVRR	7.84	5.62	2.22	0.313
SVRP	8.99	8.39	0.6	0.085
RF	7.56	5.71	1.85	0.261
AdaBoost	10.51	8.43	2.98	0.421

Pakistan, the economic benefits could approach billions of rupees per day.

IX. CONCLUSION

Electric load demand is associated with the uncertainties in load consumption due to human behavior related to weather fluctuations. Moreover, the cost of electricity production mainly depends upon the source such as water, petroleum products (oil, natural gas etc.), nuclear energy or wind and it cannot be stored at large scales. End users (individuals and industries etc.) are interested in cheaper price per unit while EGDCs require maximum profit. Therefore, AELF requires both weather fluctuations and human behavior towards such fluctuations reflected as historical electric load consumption to be studied together for a machine learning based AELF model. At the same time AELF may also be used to reduce the production cost of electricity. This study suggests

an algorithm for the selection of least cost forecasting model and gives an exclusive and first research for correlated meteorological parameters driven electric load demand of the Muzaffarabad city. An extensive analysis on annual and seasonal basis reveals that the meteorological parameters especially air temperature and dew point temperature clearly influenced the electricity consumption in Muzaffarabad, though the effects varied depending upon the factors like time of the day and seasons. Strong correlation was found between load and dew point temperature during prominent hours of the fall season. The relationship is not significant during summertime, however good inverse correlation existed during the night hours of the winter season. The cloud cover does not show enough strong relationship with the electric load demand that it can be considered as major predictor of electric load however, small to medium level correlation is revealed for some hours of different seasons. Inconsistent and non-apparent correlation is found between load and wind speed: load and rain fall during all seasons.

Forecasting results revealed significant reduction in the prediction errors for ELnMPFModels which showed the use of meteorological parameters can increase forecasting accuracy in Muzaffarabad and model learned more effectively from load-weather relationships than using exogenous and load variables alone. The reduction in the (MAPE) is translated into economic value. The computed daily savings are RS 0.313 million for day ahead forecast of Muzaffarabad. If these results are projected to bigger cities that consume a huge amount of electricity, WAPDA can reduce its operating expenses up to billions of rupees per year.

Though this study primarily focused on suggesting an algorithm for the choice of least cost electric load forecasting model and the interrelation between load demand and meteorological parameters in Muzaffarabad city, but the general methodology and findings of this study can be tested and applied to other cities of Pakistan and even every region of the globe. Based on the findings of this study, power utilities in Muzaffarabad can make economic decisions regarding their generation operations and can decrease operating cost, allowing profitable sales and reduction in purchases of power if applicable (they are buying from the market). The findings can be replicated nationwide and globally.

ACKNOWLEDGMENT

Special thanks to WAPDA (Muzaffarabad station) and PMD (Muzaffarabad station) for providing electricity load demand and meteorological data of Muzaffarabad region, respectively.

COMPETING INTERESTS

The authors declare no conflict of interests regarding the publication of this article.

REFERENCES

- [1] R. Turconi, A. Boldrin, and T. Astrup, "Life cycle assessment (LCA) of electricity generation technologies: Overview, comparability and limitations," *Renew. Sustain. Energy Rev.*, vol. 28, pp. 555–565, Dec. 2013.

- [2] L. L. Grigsby, *Electric Power Generation, Transmission, and Distribution*. Boca Raton, FL, USA: CRC Press, 2016.
- [3] B. F. Hobbs, S. Jitprapaikulsum, S. Konda, V. Chankong, K. A. Loparo, and D. J. Maratukulam, "Analysis of the value for unit commitment of improved load forecasts," *IEEE Trans. Power Syst.*, vol. 14, no. 4, pp. 1342–1348, Nov. 1999.
- [4] A. Srivastava, A. S. Pandey, and D. Singh, "Short-term load forecasting methods: A review," in *Proc. Int. Conf. Emerg. Trends Electr. Electron. Sustain. Energy Syst. (ICETEESSES)*, 2016, pp. 130–138.
- [5] K. Methaprayoon, C. Yingvivatanapong, W.-J. Lee, and J. R. Liao, "An integration of ANN wind power estimation into unit commitment considering the forecasting uncertainty," *IEEE Trans. Ind. Appl.*, vol. 43, no. 6, pp. 1441–1448, Nov./Dec. 2007.
- [6] Y. Kim, H.-G. Son, and S. Kim, "Short term electricity load forecasting for institutional buildings," *Energy Rep.*, vol. 5, pp. 1270–1280, Nov. 2019.
- [7] J. Moral-Carcedo and J. Pérez-García, "Time of day effects of temperature and daylight on short term electricity load," *Energy*, vol. 174, pp. 169–183, May 2019.
- [8] M. Sobhani, A. Campbell, S. Sangamwar, C. Li, and T. Hong, "Combining weather stations for electric load forecasting," *Energies*, vol. 12, no. 8, p. 1510, Apr. 2019.
- [9] T. K. Mideksa and S. Kallbekken, "The impact of climate change on the electricity market: A review," *Energy Policy*, vol. 38, no. 7, pp. 3579–3585, Jul. 2010.
- [10] R. Schaeffer, A. S. Szklo, A. F. P. de Lucena, B. S. M. C. Borba, L. P. P. Nogueira, F. P. Fleming, A. Troccoli, M. Harrison, and M. S. Boulahya, "Energy sector vulnerability to climate change: A review," *Energy*, vol. 38, pp. 1–12, Feb. 2012.
- [11] S. Mirasgedis, Y. Sarafidis, E. Georgopoulou, V. Kotroni, K. Lagouvardos, and D. P. Lalas, "Modeling framework for estimating impacts of climate change on electricity demand at regional level: Case of Greece," *Energy Convers. Manage.*, vol. 48, no. 5, pp. 1737–1750, May 2007.
- [12] S. Parkpoom, G. Harrison, and J. Bialek, "Climate change impacts on electricity demand," in *Proc. 39th Int. Univ. Power Eng. Conf. (UPEC)*, 2004, pp. 1342–1346.
- [13] D. Sailor, "Air conditioning market saturation and long-term response of residential cooling energy demand to climate change," *Energy*, vol. 28, no. 9, pp. 941–951, Jul. 2003.
- [14] T. Ahmed, K. M. Muttaqi, and A. P. Agalgaonkar, "Climate change impacts on electricity demand in the State of New South Wales, Australia," *Appl. Energy*, vol. 98, pp. 376–383, Oct. 2012.
- [15] S. M. Howden and S. Crimp, "Effect of climate and climate change on electricity demand in Australia," in *Proc. Int. Congr. Modelling Simulation (MODSIM)*, Soc. Aust. New Zealand, Dec. 2001, pp. 655–660.
- [16] G. H. Chen and T. T. Lie, "The impact of climate change on New Zealand's electricity demand," in *Proc. IEEE 11th Int. Conf. Probabilistic Methods Appl. Power Syst.*, Jun. 2010, pp. 808–813.
- [17] R. Mahmood, S. Saleemi, and S. Amin, "Impact of climate change on electricity demand: A case study of Pakistan," *Pakistan Develop. Rev.*, vol. 55, no. 1, pp. 29–47, Mar. 2016.
- [18] I. Staffell and S. Pfenninger, "The increasing impact of weather on electricity supply and demand," *Energy*, vol. 145, pp. 65–78, Feb. 2018.
- [19] T. G. Cassarino, E. Sharp, and M. Barrett, "The impact of social and weather drivers on the historical electricity demand in Europe," *Appl. Energy*, vol. 229, pp. 176–185, Nov. 2018.
- [20] A. A. Chaudhry, "A panel data analysis of electricity demand in Pakistan," *Lahore J. Econ.*, vol. 15, pp. 75–106, Sep. 2010.
- [21] M. Nasir, M. S. Tariq, and A. Arif, "Residential demand for electricity in Pakistan," *Pakistan Develop. Rev.*, vol. 47, no. 4II, pp. 457–467, Dec. 2008.
- [22] F. Jamil and E. Ahmad, "Income and price elasticities of electricity demand: Aggregate and sector-wise analyses," *Energy Policy*, vol. 39, no. 9, pp. 5519–5527, Sep. 2011.
- [23] M. S. Owayedh, A. A. Al-Bassam, and Z. R. Khan, "Identification of temperature and social events effects on weekly demand behavior," in *Proc. Power Eng. Soc. Summer Meeting*, 2000, pp. 2397–2402.
- [24] T. Ahmed, D. H. Vu, K. M. Muttaqi, and A. P. Agalgaonkar, "Load forecasting under changing climatic conditions for the city of Sydney, Australia," *Energy*, vol. 142, pp. 911–919, Jan. 2018.
- [25] D. H. Vu, K. M. Muttaqi, and A. P. Agalgaonkar, "Assessing the influence of climatic variables on electricity demand," in *Proc. IEEE PES Gen. Meeting | Conf. Expo.*, Jul. 2014, pp. 1–5.
- [26] A. Falkoni and G. Krajacic, "Linear correlation and regression between the meteorological data and the electricity demand of the dubrovnik region in a short-term scale," *Thermal Sci.*, vol. 20, no. 4, pp. 1073–1889, 2016.
- [27] A. Dedinec and A. Dedinec, "Correlation of variables with electricity consumption data," in *Proc. 6th Int. Conf. Inf. Soc. Technol.*, 2016, pp. 1–6.
- [28] M. Ali, M. J. Iqbal, and M. Sharif, "Relationship between extreme temperature and electricity demand in Pakistan," *Int. J. Energy Environ. Eng.*, vol. 4, no. 1, p. 36, 2013.
- [29] M. Y. A. Khan, L. Saeed, B. Zeb, N. Ahmed, M. Azhar, M. A. Y. Rustam, A. Basit, J. Soomro, and S. U. Khan, "Environmental impacts on energy utilization of smart grid consumers," in *Proc. 2nd Int. Conf. Comput., Math. Eng. Technol. (iCoMET)*, Jan. 2019, pp. 1–6.
- [30] G. R. Duarte, L. G. D. Fonseca, P. V. Z. C. Goliatt, and A. C. D. C. Lemonge, "Uma comparação de técnicas de aprendizado de máquina para a previsão de cargas energéticas em edifícios," *Ambiente Construído*, vol. 17, no. 3, pp. 103–115, Jul. 2017.
- [31] S. Haben, G. Giasemidis, F. Ziel, and S. Arora, "Short term load forecasting and the effect of temperature at the low voltage level," *Int. J. Forecasting*, vol. 35, no. 4, pp. 1469–1484, Oct. 2019.
- [32] C. Tian, J. Ma, C. Zhang, and P. Zhan, "A deep neural network model for short-term load forecast based on long short-term memory network and convolutional neural network," *Energies*, vol. 11, no. 12, p. 3493, Dec. 2018.
- [33] J. Moon, Y. Kim, M. Son, and E. Hwang, "Hybrid short-term load forecasting scheme using random forest and multilayer perceptron," *Energies*, vol. 11, no. 12, p. 3283, Nov. 2018.
- [34] N. Liu, V. Babushkin, and A. Afshari, "Short-term forecasting of temperature driven electricity load using time series and neural network model," *J. Clean Energy Technol.*, vol. 2, no. 4, pp. 327–331, 2014.
- [35] Y.-T. Chen, E. Piedad, and C.-C. Kuo, "Energy consumption load forecasting using a level-based random forest classifier," *Symmetry*, vol. 11, no. 8, p. 956, Jul. 2019.
- [36] P.-H. Kuo and C.-J. Huang, "A high precision artificial neural networks model for short-term energy load forecasting," *Energies*, vol. 11, no. 1, p. 213, Jan. 2018.
- [37] A. Gastli, Y. Charabi, R. A. Alammari, and A. M. Al-Ali, "Correlation between climate data and maximum electricity demand in qatar," in *Proc. 7th IEEE GCC Conf. Exhib. (GCC)*, Nov. 2013, pp. 565–570.
- [38] K. Chapagain and S. Kittipiyakul, "Performance analysis of short-term electricity demand with atmospheric variables," *Energies*, vol. 11, no. 4, p. 818, Apr. 2018.
- [39] D. A. Freedman, *Statistical Models: Theory and Practice*. Cambridge, U.K.: Cambridge Univ. Press, 2009.
- [40] N. S. Altman, "An introduction to kernel and nearest-neighbor non-parametric regression," *Amer. Statistician*, vol. 46, no. 3, pp. 175–185, Aug. 1992.
- [41] H. Drucker, C. J. C. Burges, L. Kaufman, A. Smola, and V. Vapnik, "Support vector regression machines," presented at the 9th Int. Conf. Neural Inf. Process. Syst., Denver, CO, USA, 1996.
- [42] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, 1995.
- [43] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.
- [44] R. E. Schapire, "Explaining AdaBoost," in *Empirical Inference*. Berlin, Germany: Springer, 2013, pp. 37–52.
- [45] M. H. Waseem, M. S. A. Nadeem, A. Abbas, A. Shaheen, W. Aziz, A. Anjum, U. Manzoor, M. A. Balubaid, and S.-O. Shim, "On the feature selection methods and reject option classifiers for robust cancer prediction," *IEEE Access*, vol. 7, pp. 141072–141082, 2019.
- [46] M. R. Abbas, M. S. A. Nadeem, A. Shaheen, A. A. Alshdadi, R. Alharbey, S.-O. Shim, and W. Aziz, "Accuracy rejection normalized-cost curves (ARNCCs): A novel 3-dimensional framework for robust classification," *IEEE Access*, vol. 7, pp. 160125–160143, 2019.
- [47] M. S. A. Nadeem, J. D. Zucker, and B. Hanczar, "Combining rejection regions for microarray-based classification," in *Proc. Conf. Francophone d'Apprentissage Automatique CAp*, Chambéry, France, 2011, pp. 377–392.
- [48] M. Nadeem, J.-D. Zucker, and B. Hanczar, "Accuracy-rejection curves (ARCs) for comparing classification methods with a reject option," *J. Mach. Learn. Res.*, vol. 8, pp. 65–81, Jan. 2010.
- [49] M. Bibi, W. Aziz, M. Almarashi, I. H. Khan, M. S. A. Nadeem, and N. Habib, "A cooperative binary-clustering framework based on majority voting for Twitter sentiment analysis," *IEEE Access*, vol. 8, pp. 68580–68592, 2020.
- [50] M. Bibi, M. S. A. Nadeem, I. H. Khan, S.-O. Shim, I. R. Khan, U. Naqvi, and W. Aziz, "Class association and attribute relevancy based imputation algorithm to reduce Twitter data for optimal sentiment analysis," *IEEE Access*, vol. 7, pp. 136535–136544, 2019.

- [51] M. Rafique, A. D. K. Tareen, A. A. Mir, M. S. A. Nadeem, K. M. Asim, and K. J. Kearfott, "Delegated regressor, a robust approach for automated anomaly detection in the soil radon time series data," *Sci. Rep.*, vol. 10, no. 1, Dec. 2020, Art. no. 3004.
- [52] A. D. K. Tareen, K. M. Asim, K. J. Kearfott, M. Rafique, M. S. A. Nadeem, T. Iqbal, and S. U. Rahman, "Automated anomalous behaviour detection in soil radon gas prior to earthquakes using computational intelligence techniques," *J. Environ. Radioactivity*, vol. 203, pp. 48–54, Jul. 2019.
- [53] L. Hussain, W. Aziz, S. Saeed, M. Rafique, M. S. A. Nadeem, S.-O. Shim, S. Aftar, and J.-U.-R. Pirzada, "Extracting mass concentration time series features for classification of indoor and outdoor atmospheric particulates," *Acta Geophys.*, vol. 68, no. 3, pp. 945–963, Jun. 2020.
- [54] D. P. Solomatine and D. L. Shrestha, "AdaBoost.RT: A boosting algorithm for regression problems," in *Proc. IEEE Int. Joint Conf. Neural Netw.*, Jul. 2004, pp. 1163–1168.



MUHAMMAD JAWAD received the B.S. and M.Phil. degrees in computer sciences from the University of Azad Jammu and Kashmir, Muzaffarabad, Pakistan, in 2012 and 2016, respectively, where he is currently pursuing the Ph.D. degree. His research interests include machine learning, neural networks, electricity load forecasting, digital signal processing, and HRV analysis.



MALIK SAJJAD AHMED NADEEM received the Ph.D. degree from Paris, France, in 2011. He has been serving as an Assistant Professor with the Department of Computer Sciences and Information Technology, University of Azad Jammu and Kashmir, Muzaffarabad. He has published various journal articles in the area of machine learning, deep learning, bioinformatics, data analytics, biomedical informatics, computational physics, and data mining.



SEONG-O SHIM received the B.S. degree in electronics engineering from Ajou University, Suwon, South Korea, in 1999, and the M.S. degree in mechatronics and the Ph.D. degree in information and mechatronics from the Gwangju Institute of Science and Technology, Gwangju, South Korea, in 2001 and 2011, respectively.

He was with LG Electronics DTV Labs, Seoul, South Korea, working on research and development of digital TV, from 2003 to 2007. He is currently working as an Associate Professor with the College of Computer Science and Engineering, University of Jeddah, Jeddah, Saudi Arabia. His research interests include computer vision, image processing, 3D shape recovery, and medical imaging.



ISHTIAQ RASOOL KHAN received the B.Sc. degree in electrical engineering from the University of Engineering and Technology, Taxila, Pakistan, in 1992, the M.S. degree in systems engineering from Quaid-i-Azam University, Islamabad, Pakistan, in 1994, and the M.S. degree in information engineering and the Ph.D. degree in digital signal processing from Hokkaido University, Japan, in 1998 and 2000, respectively. He was a JSPS Fellow with Hokkaido University, from 2000 to 2002. He has worked with the University of Kitakyushu, Japan, the Kyushu Institute of Technology, Japan, and the Institute for Infocomm Research, A*STAR, Singapore. He is currently a Professor with the College of Computer Science and Engineering, University of Jeddah. His research interests include high-dynamic range imaging, data analytics, and digital signal processing.



ALIYA SHAHEEN received the M.Phil. degree from the University of Azad Jammu and Kashmir, Muzaffarabad, Pakistan. She is currently an Assistant Professor with the Department of Mathematics, University of Azad Jammu and Kashmir. Her research interests include computational mathematics, computational informatics, and mathematical modeling.



NAZNEEN HABIB started her career as a Lecturer with the University of Azad Jammu and Kashmir (UJ&K), in 2002, where she is currently an Assistant Professor in sociology. Her core research expertise is in sociology of health and illness and focused area of research is socio-cultural factors contributing towards anemia in women. Her other research interests include social and network informatics, women health, and public health. Her main research interest includes computational intelligence.



LAL HUSSAIN received the M.S. degree in communication and networks from Iqra University, Islamabad, Pakistan, in 2012, with Gold medal, and the Ph.D. degree from the Department of Computer Science and IT (DCS&IT), University of Azad Jammu and Kashmir (UJ&K), Muzaffarabad, Pakistan, in February 2016. He is currently an Assistant Professor with DCS&IT, UJ&K. He has worked as a visiting Ph.D. Researcher with Lancaster University, U.K., for six months, under HEC International Research Initiative Program and worked under the supervision of Dr. Aneta Stefanovska, Professor of biomedical physics, Physics Department, Lancaster University. He is the author of more than 20 publications of highly reputed peer reviewed and Impact Factor Journals. He presented various talks at Pakistan, U.K., and USA. He has worked as a Visiting Researcher for two months at the AI MRI Research Group, Stonybrook University, NY, USA, in supervision of Prof. Dr. T. Duong. His research interests include biomedical signal processing with concentration in complexity measures; time-frequency representation methods; cross frequency coupling to analyze the dynamics of neurophysiological and physiological signals; neural networks and machine learning classification; regression and prediction techniques for detection of cancer, epilepsy, brain dynamics, and diseases (i.e., autism spectrum disorder (ASD) and attention-deficit/hyperactivity disorder (ADHD), Alzheimer's disease, and brain tumor); and image processing and segmentation.



WAJID AZIZ received the Ph.D. degree from the Pakistan Institute of Engineering and Applied Sciences (PIEAS), in 2006.

He started his career as a Lecturer with the University of Azad Jammu and Kashmir (UJ&K), in 1998, where he has been serving as a Professor. He was a Postdoctoral Researcher with the University of Leicester, U.K., in 2011. His core research interest includes biomedical information systems. His research interests include biomedical signal processing, time series analysis, and intelligent data analytics. He has published three books and more than 45 research articles in the reputed national and International journals and conference proceeding. Based on his academic and research contributions, he was a recipient of the HEC University Best Teacher Award for the year 2012–2013 by HEC Pakistan, in 2014, and the University Best Teacher Award by UJ&K, in 2013.

...