# A Data Filtration and Deep Learning Approach for Electrical Load Forecasting

**Deepak Yadav[1], Prof. Devkinandan Nagar[2]**
Research Scholar[1], Asst. Professor[2]
PCST, Indore[1,2]

*Abstract* — The advent of AI and Machine Learning has optimized electrical load forecasting which is an important aspect in the power sector for proper planning and maintenance of power systems. Accurate load prediction is generally challenging due to the fact that the load used by the end user lies completely at the discretion of the user but still it is possible to get fair estimates of the average load conditions using surveys and prediction mechanisms. Any information related to pattern to be followed by connected Electrical Load will helps any electric utility organization to make important decisions regarding purchasing and generating electric power, unit commitment decisions, load switching, reduce spinning reserve capacity and infrastructure development. Hence load forecasting is viewed as field of research to develop a model so that efficient and reliable operation of power system could be carried out. The paper also presents a short summary of previous techniques pertaining to the adopted methodology. The proposed algorithm uses the discrete wavelet transform for data pre-processing i.e. removing sudden spikes or irregularities in the electrical load data. The regression back propagation algorithm is used for the training purpose. It is found that the proposed system attains a Mean Absolute Percentage Error (MAPE) of around 2.4%.

***Keywords: Artificial Intelligence, Machine Learning, Wavelet Transform artificial neural Regresion Learning Algorithm, Mean Absolute Percentage Error (MAPE).***

## I. INTRODUCTION

For every electric company its very challenging to supply electricity to consumer in the most economic and secure manner due several operational difficulties. Among all challenges scheduling electric supply with load flow analysis comes at top. As electric load is a continuously varying and there is no direct way to actually guess the most accurate value at any coming time. There are many different factors which contribute to the value of electric load some may be random and other geographical. But all of them at last leads to variation of connected consumer load. Hence a method has to be developed to solve this problem [1] [2]. Many statistical methods have been proposed with time which give average results. In those methods, human instincts also play a big part. Every electric utility analysis network behavior using load forecasting results to determine whether or not if the system might be in a vulnerable situation. If yes, then utility should prepare a plan to deal with such situation which may include use of load shedding or purchasing extra required from market etc. [3].
But estimation of future load with the help of past observation has remained to be difficult task for researchers. Since the recent development in the field of artificial neural network and data mining there is a sharp improvement in load forecasting results. Since beginning of artificial intelligence tool it been possible to solve various time series problems with great accuracy [4].

## II. METHODOLOGY:

Following techniques are utilized in present work for thepurpose of load forecasting. Neural computing systems mimic the learning processes of the brain to discover the relations between the variables of a system. They process input data information to learn and obtain knowledge for forecasting or classifying patterns etc. type of work. ANN consists of number of simple processing elements called neurons. All information processing is done within this neuron only. A network of connected artificial neurons can be designed, and a learning algorithm can be applied to train it [8]. Signals (Input data) are passed between neurons over connection links and Each connection link has an associated weight, which in a neural network, multiplies the signal transmitted. The weights represent information being used by the network to solve a problem. Then the weighted sum is operated upon by an activation function (usually nonlinear), and output data are conveyed to other neurons. The weights are continuously altered while training to improve accuracy and generalize abilities [9] [10].

**The Regression Learning Algorithm:**

Reducing error function is the main reason to use this algorithm. Levenberg-Marquardt algorithm [11] [12] is a very efficient technique for minimizing a nonlinear function. The algorithm includes many different variables like in present study we have output data, weight between neurons and error function, that determine efficiency and success rate of model. The ideal values of these variables are very dependent on the test function.
Levenberg-Marquardt algorithm is fast [13] and has stable convergence. This algorithm was designed to approach second-order training speed without computing the Hessian matrix. When the performance function has the form of a sum of squares, then the Hessian matrix can be approximated and the gradient can be computer as

$$H = {J_x}^T J_x \quad (1)$$
$$g = {J_x}^T e \quad (2)$$

Where $J_k$ is the Jacobian matrix for $k^{th}$ input, which contains first order derivatives of the network errors with respect to the weights and biases, $e$ is a vector of network errors. The Jacobian matrix can be computed through a standard backpropagation technique that is much less complex than computing the Hessian matrix. [14]

The Levenberg –Marquardt algorithm is actually a blend of the steepest descent method and the Gauss–Newton algorithm. The following is the relation for LM algorithm computation,

$$W_{k+1} = W_k - [{J_K}^T J_k + \mu I]^{-1} {J_K}^T e_k \quad (3)$$

| Algorithm | Rules | Convergence |
|---|---|---|
| Gradient Newton algorithm | $W_{k+1} = W_k - \alpha g_x,\ \alpha = \frac{1}{\mu}$ | Stable, slow |
| Gauss–Newton algorithm | $W_{k+1} = W_k - [J_K{}^T J_k]^{-1} J_K{}^T e_k$ | Unstable, fast |
| Levenberg–Marquardt (LM) algorithms | $W_{k+1} = W_k - [J^T{}_k J_k + \mu I]^{-1} J_k e_k$ | Stable, fast |

where $I$ is the identity matrix, $W_k$ is the current weight, $W_{k+1}$ is the next weight, $e_{k+1}$ is the current total error, and $e_k$ is the last total error, μ is combination coefficient. [14] [15]
The network tries to combine the advantages of both the methods hence it inherits the speed of the Gauss–Newton algorithm and the stability of the steepest descent method.
The combination coefficient μ is multiplied by some factor (β) whenever a step would result in an increased $e_{k+1}$ and when a step reduces $e_{k+1}$, μ is divided by β. In this study, we used β=l0. When μ is large the algorithm becomes steepest descent while for small μ the algorithm becomes Gauss-Newton. [14]

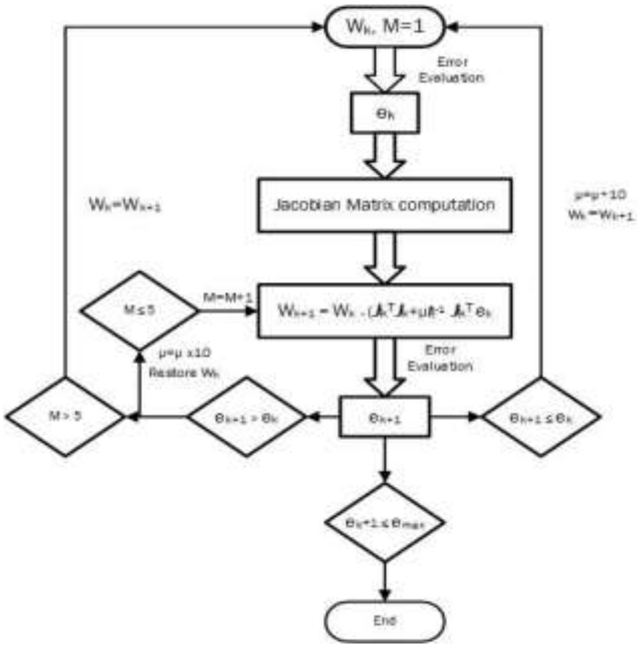


*Fig. 1: Block diagram for training using Levenberg–Marquardt algorithm*

In present study, the Levenberg-Marquardt (LM) learning algorithm was applied with in the input Neurons have no transfer function. The logistic sigmoid transfer (logsig) and linear transfer (purelin) functions were used in the hidden and output layers of the network as an activation function, respectively.

The Neural Network in present study consist of three layers. The first one is Input layer (consist of 11 Neutrons for 11 input element) from where inputs are feed to model for further computation. Then comes second layer called Hidden Layer (consist of 20 Neurons), this is where activation function is used to limit value of output Neuron. At last we have output layer (1 Neuron) form which we take output result for comparison with actual result to calculate error and the feeding it back to model to vary weight accordingly for improving performance.

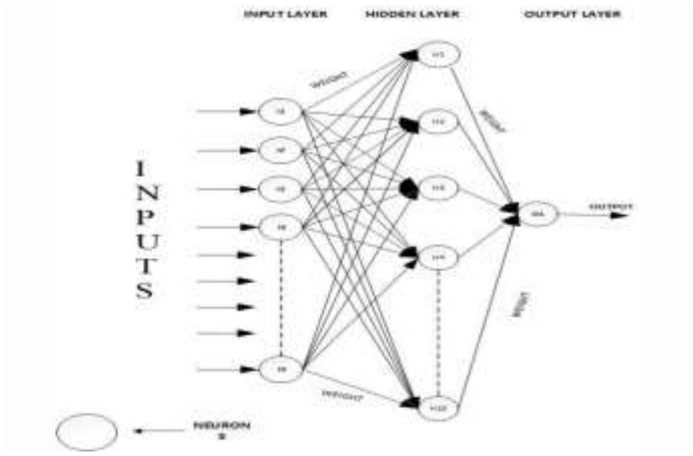


*Fig. 2: Working model*

**Wavelet Transform for Time series problem:**

In the present proposed system, a wavelet transformed signal processing is utilized and applied on each signal in order to obtain its vector of features. In recent times WT has been in limelight in the field of signal processing because of its ability to acquire information from any large signal. [5]

Similarly, as Fourier Transform WT divides a signal in to small pieces. But in Fourier this partition is done on the basis of standard sine wave which is assumed to be of infinite length of different frequencies whereas WT uses scaled & fixed duration, irregular and asymmetric signal pieces which are known as mother wavelets. Hence by using WT we can accomplish localization of both time and frequency framework as seen from the figure below:

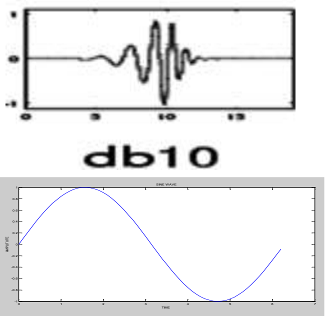


*Fig. 3: (a) db-10 Wavelet (b) Sinewave*

What mother wavelet do is that for high frequency components the time intervals would be short whereas for low frequency components time intervals would be longer. In present problem of load forecasting, Wavelet transform helps in sensing seasonality with time varying period and intensity which eventually improves the quality of forecasting. [6]

The article aims to verify this by comparing the power of classical and wavelet based techniques on the basis of five-time series, each of them having individual characteristics. Depending on the data's characteristics and on the forecasting horizon we either favor a denoising step plus an ANN forecast or a multiscale wavelet decomposition plus an ANN forecast for each of the frequency components.

Just looking at pictures of wavelets and sine waves, we can see intuitively that signals with sharp changes might be better analyzed with an irregular wavelet than with a smooth sinusoid, just as some foods are better handled with a fork than a spoon. It also makes sense that local features can be described better with wavelets, which have local extent.

Discrete Wavelet Transform (DWT) The disadvantage of the continuous wavelet transform lies in its computational complexity and redundancy. In order to solve these problems, the discrete wavelet transform is introduced. Unlike CWT, the DWT decomposes the signal into mutually orthogonal set of wavelets. The discrete wavelet is defined as:

$$\psi_{j,k}(t) = \frac{1}{\sqrt{s_0^j}}\psi\left(\frac{t - k\,\tau_0\, s_0^j}{s_0^j}\right)$$

where j and k are integers, $s_0 > 1$ is a fixed dilation step and the translation factor $\tau_0$ depends on the dilation step. The scaling function and the wavelet function of DWT are defined as:

$$\phi(2^j t) = \sum_{i=1}^{k} h_{j+1}(k)\ \phi(2^{j+1}t - k)$$

$$\psi(2^j t) = \sum_{i=1}^{k} g_{j+1}(k)\ \phi(2^{j+1}t - k)$$

And then, a signal f(t) can be written as:

$$f(t) = \sum_{i=1}^{k} \lambda_{j-1}(k)\ \phi(2^{j-1}t - k) + \sum_{i=1}^{k} v_{j-1}(k)\ \phi(2^{j-1}t - k)$$

The discrete wavelet transform can be done by using the filter bank scheme developed. Figure 4 shows a two-channel filter bank scheme for DWT.

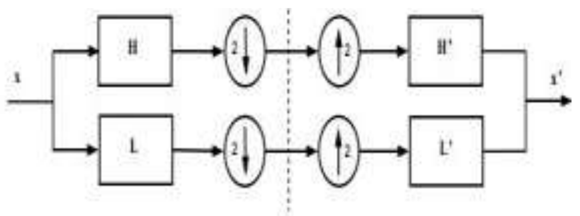


*Fig. 4: Filter Bank Scheme for DWT*

In the figure, H, L, and H', L' are the high-pass and low-pass filters for wavelet decomposition and reconstruction respectively. In the decomposition phase, the low-pass filter removes the higher frequency components of the signal and high pass filter picks up the remaining parts. Then, the filtered signals are down sampled by two and the results are called approximation coefficients and detail coefficients. The reconstruction is just a reversed process of the decomposition and for perfect reconstruction filter banks, we have x = x'. A signal can be further decomposed by cascade algorithm as shown in Equation 5):

$x(t) \quad = A1(t) + D1(t)$

$= A2(t) + D2(t) + D1(t)$

$\quad = A3(t) + D3(t) + D2(t) + D1(t)$

$= An(t) + Dn(t) + Dn_{-1}(t) + \ldots\ldots + D1(t)$ (4)

where Dn(t) and An(t) are the detail and the approximation coefficients at level n respectively. Fig. 5 illustrates the corresponding wavelet decomposition tree.

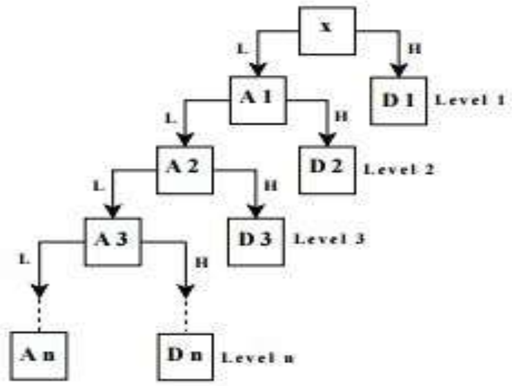


*Fig. 5: Wavelet Decomposition Tree*

A significant potential problem with the DWT is that it is a shift variant transform. Shift-variance is a phenomenon of not necessarily matching the shift of the one-level DWT with the one-level DT of the same shift of a data sequence. Due to shift-variance, small shifts in the input waveform. Wavelet Transform cause large changes in the wavelet coefficients and this makes the DWT suitable for this study because we relate the information at a given time point at the different scales.

## III. LOAD FORECASTING

Load forecasting is the closeness between the actual and predicted future load values. The electrical short-term load forecasting has been emerged as one of the most essential field of research for efficient and reliable operation of power system in last few decades. Load forecasting helps in energy management, load flow analysis, planning and maintenance of power system and contingency analysis.

A Precise forecasting is expected from proposed model since under-forecasted value will lead in negative consequences on demand response and finally on power installation. This under forecasted value will also result in difficulty to manage the overload situations. In case of over-forecasted value the negative effect can be seen in installation and hence the efficiency of the system.

Broadly, the load forecasting techniques can be divided into two categories such as parametric or non-parametric techniques. The linear regression, auto regressive moving average (ARMA), general exponential technique and stochastic time series techniques are some examples of parametric (statistical) technique. The main drawback of this technique is its capability in abrupt change of any types of environment or social changes. However, this shortcoming is overcome by applying non-parametric (artificial intelligence) based technique because of its potentiality to global search. Among these artificial intelligences based methodology, artificial neural network has emerged as one of the most prominent technique that receive much more attention of researchers. The ability to solve the complex relationships, adaptive control, image denoising, decision making under uncertainty and prediction patterns make ANN a powerful performer than previously implemented techniques.

This project proposes a study of short-term hourly load forecasting using Artificial Neural Networks (ANNs). Load forecasts can be divided into three categories: short-term forecasts which are usually from one hour to one week, medium forecasts which are usually from a week to a year, and long-term forecasts which are longer than a year. The forecasts for different time horizons are important for different operations within a utility company. The natures of these forecasts are different as well.

## IV. STAGES FOLLOWE IN LOAD FORECASTING

Following stages are followed in same order as mentioned for performing load forecasting. Stage 3 is skipped in case of performing load forecasting without wavelet transform.

**Stage 1**: Data Extraction Stage

To demonstrate the effectiveness of the proposed approach, publicly available data of load from "The Electric Reliability Council of Texas (ERCOT)" web site [7] has been taken to forecast the hourly load for the Texas state of united states. Also, other than load all weather data of Texas state is obtained from the "Midwestern Regional Climate Center" [8] website where data is openly available for all.
The data used in this study is dated between 1$^{st}$ Jan 2015 to 28$^{th}$ Feb 2017. All input variables have per hour readings, so a total of 18960 samples of each variable are used in study.

**Stage 2**: Data pre-processing stage

In this stage, all the data extracted in previous stage are organized and pre-processed for further stages. In this stage first all the data is copied in single excel sheet, with each column representing value of each input parameter. Now after that insert two more column for weekday and working day values. This two parameters are introduced because load is dependent on the day of the week of observation and weather the day is working day or it's a holiday. Hence, we have represented each day by a respected number like Sunday by 1 and Tuesday by 2. Similarly, we have marked working day by 0 and holiday by 1.

Also, a search is to be performed to check for empty data cells in sheet for better performance.

Following inputs elements are taken for forecasting load

- Date
- Time (hour)
- Weekday
- Working Day
- Temperature
- Pressure
- Humidity
- Dew Point
- Wind speed
- load one hours ago
- load two hours ago

and

- Load after 24 hours will be taken as the target.

**Stage 3**: Feature extraction stage

In this stage, a Wavelet Transform is performed on all parameters under consideration including load values. Wavelet transform is done up to 3$^{rd}$ level, so a total of 3 detail coefficient and 1 approximate coefficient are obtained for each signal.

We have saved this values from workspace for further stages. Matlab have a toolbox totally dedicated for various wavelet transform operations. Now after this stage we have 4 coefficients which will form the original signal on performing reverse function. We have total 4 set of data for training. Figure below shows a 3$^{rd}$ level wavelet decomposition of a signal (wind speed).

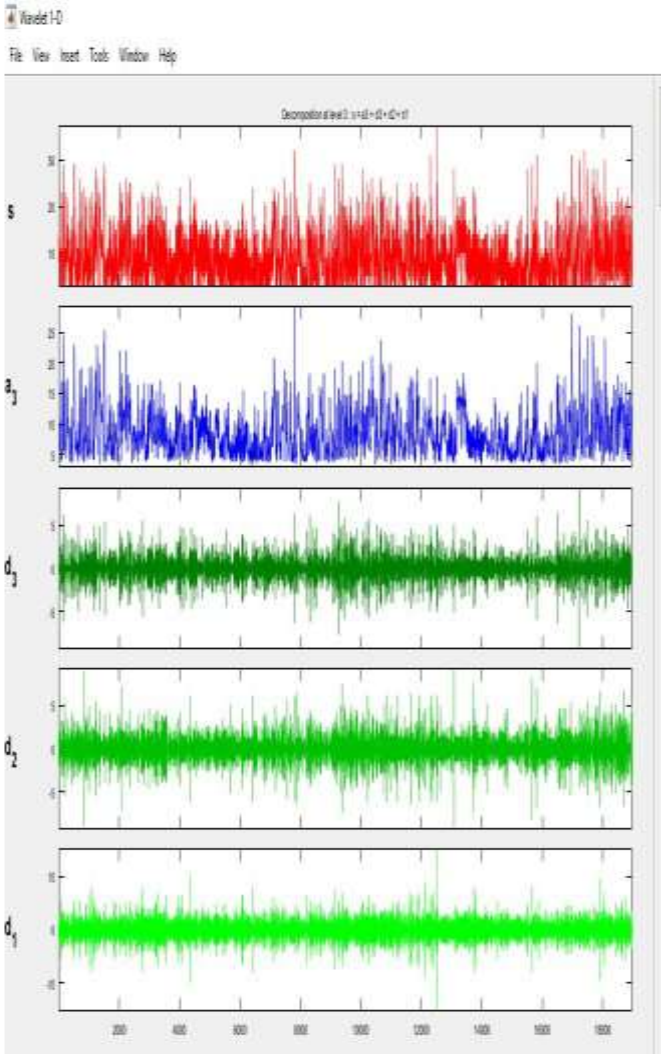


*Fig. 6: Wavelet decomposition up to level 3 of wind speed.*

**Stage 4**: Neural Network Training Stage

Before further operation the data is divide in two parts, one for training and other for testing. So, in this stage we will use training data. Since we have 4 training datasets so we have to train 4 neural networks. We have used LM training algorithm for training.

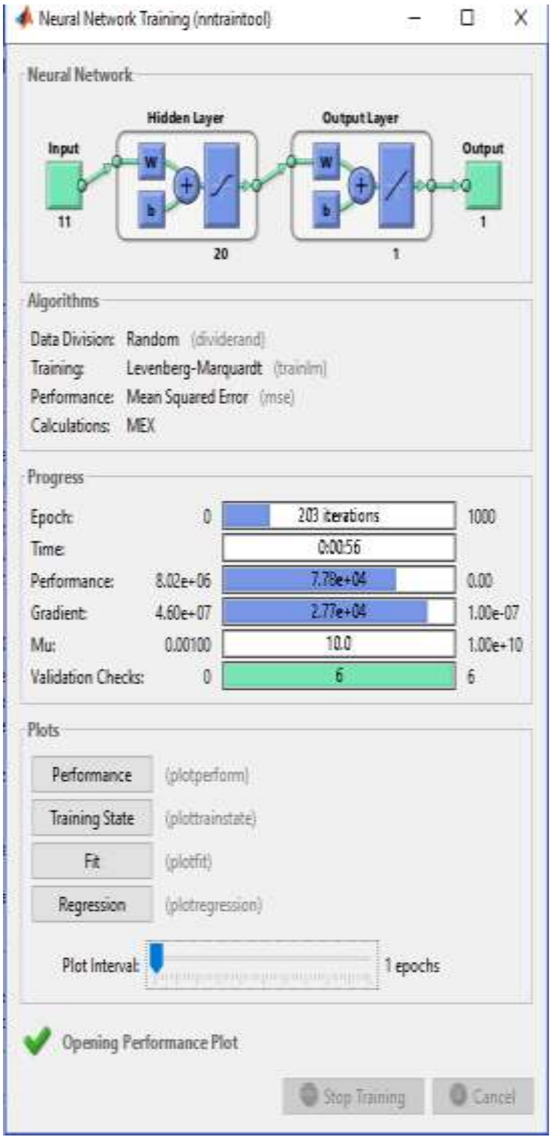


*Fig. 7: GUI of ANN during training.*

The above figure shows a training GUI of neural network, which will give all details of related to training. We have taken 20 hidden layer neurons 11 input layer neurons and 1 output layer neuron.

**Stage 5**: Neural Network Testing Stage

At this stage, the second part of dataset is used. Although only inputs are provided to already trained neural network and output is calculated from all four neural networks. These four outputs will then be added to form output signal. This signal is then compared to original target load to observe the closeness between the two.

## V. RESULTS

The results of the proposed system have been simulated in Matlab with the Neural Networks and Deep Learning Libraries.

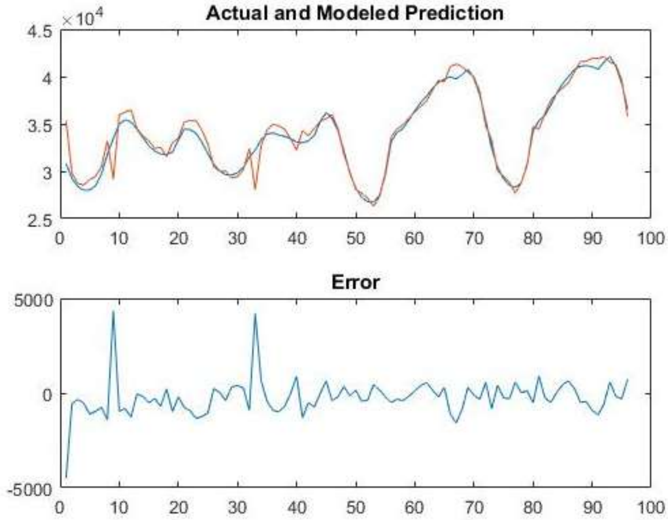


*Fig. 8: Comparison between the predicted and actual load employing the proposed model using LM training with WT.*

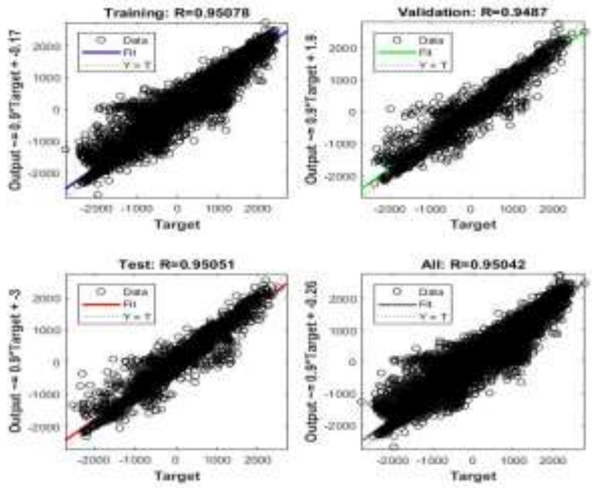


*Fig. 9: Regression plot during training, testing & validation for Proposed for Proposed Levenberg-Marquardt (LM) training algorithm with WT.*

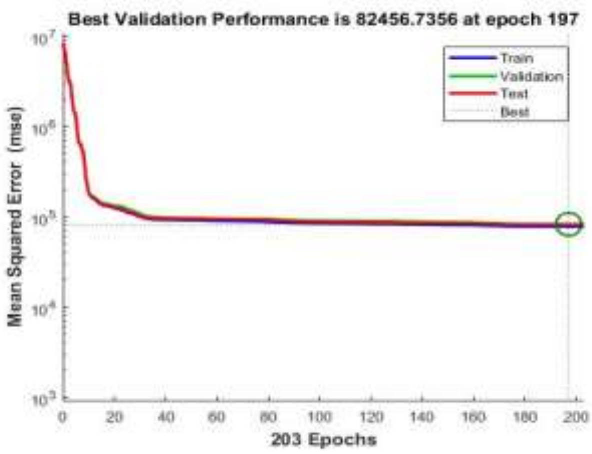


*Fig. 10: Neural network performance during training, testing & validation with WT.*

Mean Absolute Percent Error (MAPE): ~1.7 %
Mean Absolute Error (MAE): ~500 MW

**Without Wavelet transform**

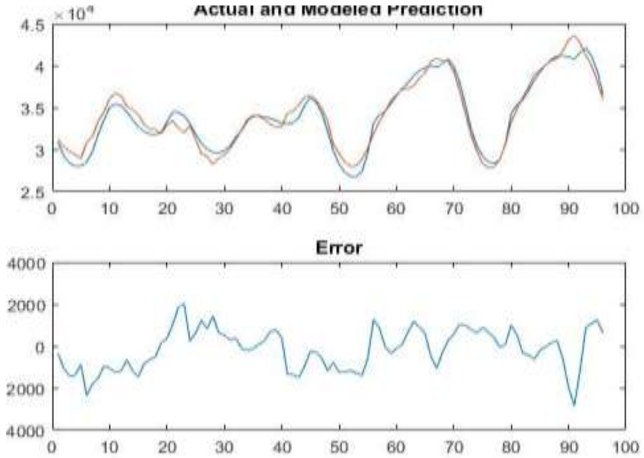


*Fig. 11: Comparison between the predicted and actual load employing the proposed model using LM training without WT.*

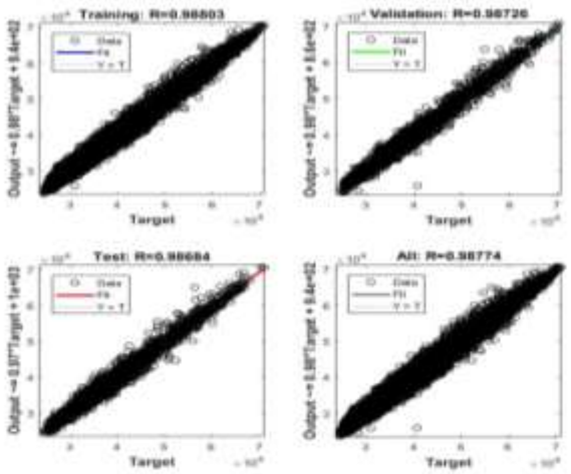


*Fig. 12: Regression plot during training, testing & validation for Proposed for Proposed Levenberg-Marquardt (LM) training algorithm without WT.*

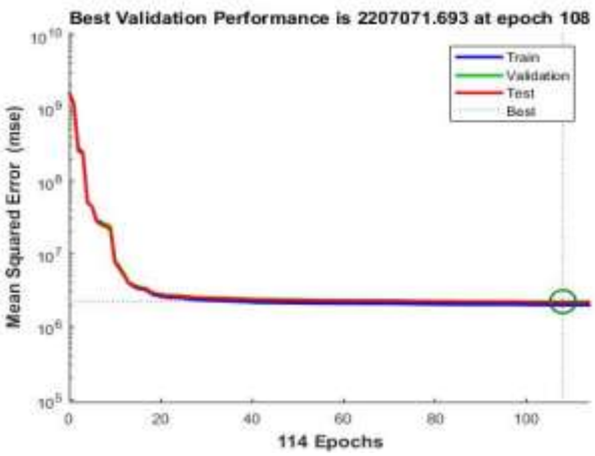


*Fig. 13: Neural network performance during training, testing & validation without WT.*

Mean Absolute Percent Error (MAPE): ~2.4%
Mean Absolute Error (MAE): ~800 MW

## VI. CONCLUSION

The present work can be considered to be an amalgamation of two standard processes i.e. the Discrete Wavelet Transform for data pre-processing and the LM algorithm for implementing Back Propagation. The results obtained are:
The developed ANN model with the configuration of 11-20-1 gives the result with MAPE ~ 2.4 % and corresponding MAE ~ 800 MW. Thus it can be seen that the accuracy in load prediction is pretty high. Out of the two techniques WT-ANN has least error hence it will give the most accurate forecasting.

## REFRENCES

1. D. Cao et al., "Robust Deep Gaussian Process-based Probabilistic Electrical Load Forecasting against Anomalous Events," in IEEE Transactions on Industrial Informatics, 2025.
2. L. Yin, Z. Sun, F. Gao and H. Liu, "Deep Forest Regression for Short-Term Load Forecasting of Power Systems," in IEEE Access, vol. 8, pp. 49090-49099, 2024.
3. D. Wu, B. Wang, D. Precup and B. Boulet, "Multiple Kernel Learning-Based Transfer Regression for Electric Load Forecasting," in IEEE Transactions on Smart Grid, vol. 11, no. 2, pp. 1183-1192, 2023.
4. N. J. Johannesen, M. Kolhe and M. Goodwin, "Comparison of Regression Tools for Regional Electric Load Forecasting," 3rd International Conference on Smart and Sustainable Technologies (SpliTech), 2022, pp. 1-6.
5. L. Li, K. Ota and M. Dong, "Everything is Image: CNN-based Short-Term Electrical Load Forecasting for Smart Grid,"11th International Conference on Frontier of Computer Science and Technology & 2021, pp. 344-351.
6. K. P. Warrior, M. Shrenik and N. Soni, "Short-term electrical load forecasting using predictive machine learning models," 2020 IEEE Annual India Conference (INDICON), 2020, pp. 1-6.
7. Hao Quan, Dipti Srinivasan, Abbas Khosravi,"Short-Term Load and Wind Power Forecasting Using Neural Network-Based Prediction Intervals" , IEEE Transactions on Neural Networks and Learning Systems, IEEE, Volume: 25, Issue: 2, pp 303 – 315, 2019.
8. Sharad Kumar, Shashank Mishra, Shashank Gupta, "Short Term Load Forecasting Using ANN and Multiple Linear Regression", Second International Conference on Computational Intelligence & Communication Technology, IEEE, 2018
9. Kishan Bhushan Sahay, Suneet Sahu, Pragya Singh, "Short-Term Load Forecasting of

Toronto Canada by Using Different ANN Algorithms", 6th International Conference on Power Systems (ICPS), 2017.

10. Renuka Achanta, "Long Term Electric Load Forecasting using Neural Networks and Support Vector Machines", International Journal of Computer Science and Technology, Vol. 3, Issue 1, pp 266-269, 2016.
11. Anand Mohan, "Mid Term Electrical Load Forecasting for State of Himachal Pradesh Using Different Weather Conditions via ANN", Model International Journal of Research in Management, Science & Technology, Vol. 1, No. 2, pp 60-63, 2015.
12. S.S. Somam, H. Zareipour, O. Malik, and P. Mandal, "A Review of Wind Power and Wind Speed Forecasting Methods with Different Time Horizons". North American Power Symposium, Arlington, pp. 26-28, 2015.
13. Wen-Yeau Chang, "A Literature Review of Wind Forecasting Methods", Journal of Power and Energy Engineering, pp. 161-168, 2014.
14. A.D. Papalexopoulos, T.C. Hesterberg, 'A Regression Based Approach to Short Term Load Forecasting', IEEE Transactions on Power Systems, 40 – 45, 1990.
15. N. Amjady, 'Short-term hourly load forecasting using time-series modeling with peak load estimation capability', IEEE Transactions on Power Systems, 2013, pp.498 – 505
16. J. Yang, H. Cheng, "Application of SVM to power system short-term load forecast", Power System Automation Equipment China, pp 30 – 32, 2012.
17. W. Christianse, "Short Term Load Forecasting Using General Exponential Smoothing", IEEE Transactions on PAS, 2019, 900 – 910.
18. S.A. Villalba, C.A. Bel, "Hybrid demand model for load estimation and short-term load forecasting in distribution electrical systems", IEEE Transactions on Power Delivery, pp 764 – 769, 2020.