

Long Term Load Forecasting using Grey Wolf Optimizer – Artificial Neural Network

Zuhaila Mat Yasin

*Faculty of Electrical Engineering**Universiti Teknologi MARA*

Shah Alam, Malaysia

zuhai730@uitm.edu.my

Nur Ashida Salim

*Faculty of Electrical Engineering**Universiti Teknologi MARA*

Shah Alam, Malaysia

nurashida606@uitm.edu.my

Nur Fadilah Ab Aziz

*College of Engineering**Universiti Tenaga Nasional*

Bangi, Malaysia

NFadilah@uniten.edu.my

Abstract—This paper presents a new technique namely Grey Wolf Optimizer – Artificial Neural Network (GWO-ANN) as a technique to forecast electrical load. GWO is a meta heuristic technique inspired by the hierarchy of leadership of the grey wolf hunting mechanism in nature. Four types of grey wolves such as alpha, beta, delta, and omega are employed for simulating the leadership hierarchy. In addition, the three main steps of hunting, searching for prey, encircling are also imitated in the algorithm. GWO is utilized to determine the optimal momentum rate and learning rate of ANN for accurate prediction. In the ANN configuration, the temperature, humidity, wind speed, maximum power, and average power were used as the input data. While total power was used as the output data. ANN is trained by adjusting the parameters of momentum rate and learning rate until the output data matches the actual data. The performance of GWO-ANN was compared to the performance of ANN and Particle Swarm Optimization - Artificial Neural Network (PSO-ANN). The results showed GWO-ANN provide better result in terms of the Mean Absolute Percentage Error (MAPE) and coefficients of determination (R^2) as compared to other methods.

Keywords— *Artificial Neural Network (ANN); Grey Wolf Optimizer (GWO), Load Forecasting*

I. INTRODUCTION

Load forecasting is important for planning and operation of a power system. An accurate prediction must be made to ensure the efficiency, reliability and security of power system. An efficient power planning will reduce the operational cost for power generation, transmission and delivery system [1]. The major concern for every electrical utility is the ability to provide reliable and uninterrupted service to their customers. Unlike other commodities, electricity, by nature, must be produced at the same rate as it is consumed [2]. In the past, cost is not a concern for energy firms due to the model of competition is monopoly which means only one generating company is in charge in supplying electric power within an area [3]. This type of competition creates problems such as inefficient energy production and unreasonable electric prices. As time passes, technology keeps on evolving, the model of competition changes to perfect competition as more generating company exist. Electric power is then sold through open energy market which functions to supervise electric price and maintains it. Thus, making it important for generating companies to maintain their competitiveness in the energy market by grasping two major factors; power planning and unit commitment. Overestimation leads to over-investment and therefore, it increase the electricity price. While demand

under estimation may lead to under-investment resulting in unreliable and insecure electricity. Load forecasting helps generating companies in deciding whether to purchase or generate electric power as this decision will affect their financial institutions. Overestimation leads to an over-investment and an increase in the electricity price. While demand underestimation may lead to under-investment resulting in unreliable and insecure electricity [4].

There are three types of load forecasting:

a) Short-Term Load Forecast (STLF): From one hour to a week duration.

b) Medium-Term Load Forecast (MTLF): From one week up to a year duration.

c) Long-Term Load Forecast (LTLF): Duration of more than one year.

STLF predicts the load of substations, feeders, and individual customers [4]. It is also needed for controlling and scheduling power system as well as estimating load flow to prevent overloading. While MTLF and LTLF are used to regulate the limit of generation and transmission, power system expansion, transmission development planning and yearly hydrothermal maintenance scheduling [5].

Load forecasting problem is difficult to solve because the power system is a huge complex graphically, widely distributed and influenced by many factors. Hamedmoghadam et al. stated that industrialization and rural electrification play major role in increasing load demand. Other than that, the usage of electrical appliances and the number of consumers also affect electric power demand [4]. Reference [5] classified the time of the day and weather conditions as two main factors affecting the load demand. The weather is then affected by temperature, humidity and wind speed. Reference [6] considered demographic factors such as weather, climate, and variation of load demands as the input of their research.

There are several load forecasting techniques that can be used such as deep neural network [4], [7] system type neural network architecture [5], Artificial Neural Network (ANN), Adaptive Neuro-Fuzzy Inference System (ANFIS) [6], and Long Short-Term Memory (LSTM) Recurrent Neural Network (RNN) to predict electrical load [8]. Ertugrul forecasted electricity load using a novel recurrent extreme learning machines approach that is a training method for Single Hidden Layer Feed Forward Neural Network (SHLFN). SHLFN uses a formula with x as input, y as desired output, and m and n as the number of neurons in hidden and input layers [1]. Reference [4] utilized deep neural network which is a particular type of ANN to predict electricity

consumption rates for 1-24 months ahead in Australia. The inputs used includes socio-economical inputs such as Gross Domestic Product (GDP) and population. Other than that, environmental inputs such as monthly precipitation, Carbon Dioxide (CO₂) emissions, average temperature, average minimum temperature, and average maximum temperature. Hobbs *et al.* proposed a long-term load forecasting using system type neural network architecture [5]. Ammar *et al.* predicted long-term load using ANN and ANFIS [6]. Guo *et al.* used deep learning model to forecast short-term power load and probability density. Deep neural network uses more hidden layers and flexible structures making it more complex and difficult to train. However, when compared to single hidden layer neural network, deep neural network can manage numerous sequences and large image datasets [7]. Kong *et al.* used LSTM RNN to predict short-term demand [8].

Lee and Hong predicted load demand up to four-months using a hybrid dynamic and fuzzy time series model for mid-term load demand forecasting using two typical stochastic models, the Koyck model and the Autoregressive Integrated Moving Average (ARIMA) [9]. Liu *et al.* proposed probabilistic load forecasting by performing quantile regression averaging on a set of sister point forecasts. This method can leverage the development in the point load forecasting literature and does not rely too much on high quality expert forecast. The technique of forecasting is separated into two steps, generating a set of sister load forecasts and applying Quantile Regression Averaging (QRA) [10]. Research in ANN has gain popularity due to its robustness. ANN is also used to predict the failure risk for lightning surge protection of underground medium voltage cables [11], nodal congestion price estimation in spot power market [12], and to identify the primary fault section after contingencies in bulk power systems [13]. In this paper, Grey Wolf Optimizer – Artificial Neural Network (GWO-ANN) is proposed for long term load forecasting. The suggested method is developed to explore the most ideal training parameters such as learning rate and momentum rate. The objective function for the optimization operation is to minimize the error between the expected and targeted output.

II. METHODOLOGY

The required data such as temperature (°F), humidity (%), and wind speed (mph) are obtained from Philadelphia International Station by Wunderground Website [14]. While maximum, average, and total power are retrieved from PJM Website [15] of the load demand in Philadelphia from the 1st January of 2016 until 31st October 2018.

A. Development of ANN Model

In this paper, the ANN model was formed as one input layer, two hidden layers, and one output layer. The measured data such as power, temperature, humidity, and wind speed were utilized as input data. The total power were utilized as output data. In addition, 841 input data was employed for training, and 180 data was employed for testing. The ANN has two hidden layers. Levenberg- Marquardt (LM) algorithm were chosen due to the reliable, simple and rapid properties. Furthermore, Mean Absolute Percentage Error (MAPE) performance function, and Logsig transfer function were chosen in this ANN simulation. The architecture of the ANN model is shown in Fig. 1.

The training process of ANN starts with loading the input data and normalizing it from -1 to 1. The ANN configuration

is then designed with respect to the parameters. The parameters include momentum rate, learning rate, number of hidden layers, and number of neurons in the hidden layers. Momentum rate is utilized to quicken the training process of supervised ANN. Learning rate is utilized to quicken the convergence of ANN training process. The value of momentum rate and learning rate are set within the range of 0 and 1. The selection of these parameters are conducted using a heuristic method. Both momentum rate and learning rate are increased from 0.1 to 1.0 to obtain the least MAPE. The number of nodes in both hidden layers are set in the range of 1 to 20.

The ANN was carried out with the following steps:

1. Perform training and testing data.
2. The training and testing data are normalized.
3. The parameters such as number of hidden layers, learning rate, momentum rate, number of iterations and training goal are set.
4. The program is run until the network converges and produces the expected output.
5. The expected output is de-normalized.
6. The expected output is compared with the actual output.
7. Normalized testing data.
8. Testing data is used to simulate the network.
9. De-normalized testing data.
10. The expected output is compared with the actual output of testing data.
11. The MAPE is used to determine the accuracy of the prediction. MAPE was resolved using (1).

$$MAPE = \left[\frac{1}{N} \sum_{i=1}^N \frac{|y_{p,i} - y_{t,i}|}{|y_{p,i}|} \right] \times 100 \quad (1)$$

Where N is the value of data used, $y_{p,i}$ is the expected data and $y_{t,i}$ is the targeted value of ANN. Calculation of the correlation of determination (R^2) can be determined using (2).

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_{p,i} - y_{t,i})^2}{\sum_{i=1}^N (y_{t,i} - y_{avg})^2} \quad (2)$$

Where y_{avg} is the average value of the targeted output of ANN. The value of R^2 that is close to 1 shows that the network execution is accurate and dependable.

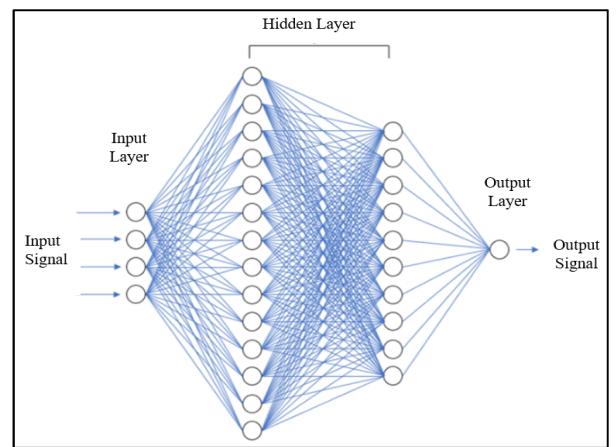


Fig. 1. ANN Architecture

B. Grey Wolf Optimizer – Artificial Neural Network

Grey Wolf Optimizer (GWO) was first introduced in [16]. The effectiveness of GWO algorithm for solving various problems has been demonstrated in [17], [18], [19]. The GWO algorithm imitates the ranking and hunting technique of grey wolves.

Four types of grey wolves are:

- Alpha (α), the leader of the pack
- Beta (β), the subordinate that assists the α
- Delta (δ), submissive towards α and β
- Omega (ω), the scapegoat

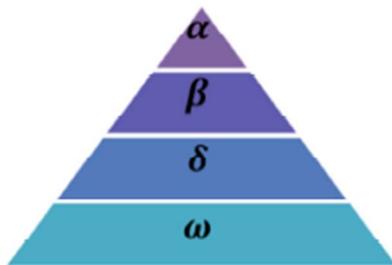


Fig. 2. Hierarchy of grey wolves

The heuristic selection of the training parameters of ANN might be bothersome. Therefore, GWO is introduced to deduce the most ideal momentum rate and learning rate of ANN so that the prediction execution could be put on a spurt and optimized. The hierarchy of grey wolves is shown in Fig. 2. The proposed GWO-ANN model is presented in Fig. 3.

At the start, the grey wolf population X_a , A , and C is initialized. In this paper, four parameters are considered which are hidden node 1, hidden node 2, learning rate (lr) and momentum rate (mc). MAPE is set as the fitness function by running the ANN model. Then, the iteration starts as the fitness equation is optimized. X_α , X_β , and X_δ are determined to update the position of the search agent. As a , A , and C are updated, iteration increases until the convergence goal is met and X_α is taken as the best solution as shown in (3). Grey wolves encircle their prey during hunting.

$$\begin{aligned} \vec{D} &= |\vec{C} \cdot \vec{X}_p - \vec{X}(t)| \\ \vec{X}(t+1) &= \vec{X}_p(t) - \vec{A} \cdot \vec{D} \\ \vec{A} &= 2\vec{a} \cdot \vec{r}_1 - \vec{a} \\ \vec{C} &= 2\vec{r}_2 \end{aligned} \quad (3)$$

Where t is the ongoing iteration

\vec{A} and \vec{C} are coefficient vectors

\vec{X}_p is the position vector of the prey

\vec{X} is the position of the grey wolf

Grey wolves can recognize the location of prey and encircle them.

$$\vec{D}_\alpha = |\vec{C}_1 \cdot \vec{X}_\alpha - \vec{X}|$$

$$\vec{D}_\beta = |\vec{C}_2 \cdot \vec{X}_\beta - \vec{X}|$$

$$\vec{D}_\delta = |\vec{C}_3 \cdot \vec{X}_\delta - \vec{X}|$$

$$\begin{aligned} \vec{X}_1 &= \vec{X}_\alpha - \vec{A}_1 \cdot (\vec{D}_\alpha) \\ \vec{X}_2 &= \vec{X}_\beta - \vec{A}_2 \cdot (\vec{D}_\beta) \\ \vec{X}_3 &= \vec{X}_\delta - \vec{A}_3 \cdot (\vec{D}_\delta) \\ \vec{X}(t+1) &= \frac{\vec{X}_1 + \vec{X}_2 + \vec{X}_3}{3} \end{aligned} \quad (4)$$

Equation (4) lets search agents to updates its position according to α , β , and δ in n-dimensional search space. The process is repeated until the solution converged. The convergence criterion is set as the different between maximum and minimum fitness.

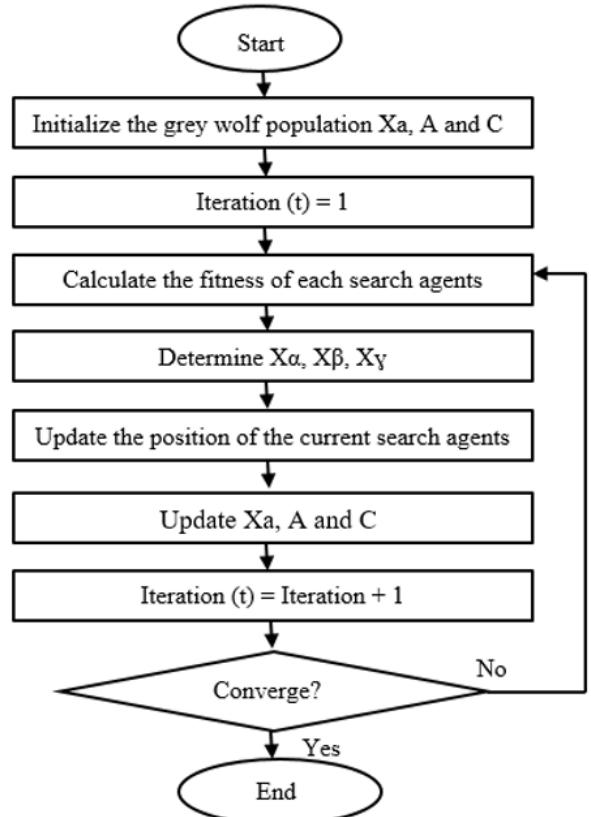


Fig. 3. Flowchart of GWO-ANN

C. Particle Swarm Optimization – Artificial Neural Network

Particle Swarm Optimization (PSO) is based on the movement of a bird flock or fish school and developed by Eberhart and Kennedy. PSO is a meta heuristic method which can search very large spaces of solutions according to the search agents [20], [21]. PSO has been applied to numerous areas in optimization. This method performs the search of the optimal solution through agents, referred to as particles. In PSO-ANN, PSO is utilized to optimize the ANN parameters such as momentum rate and learning rate. The algorithm began with initialization process. In the initialization process, the position of particle swarm is randomly generated using (5). Momentum rate and learning rate are set between 0 and 1. While hidden node 1 and hidden node 2 are set between 1 to 20.

$$\mu_0^j = \text{rand}(\mu_{\max}^j - \mu_{\min}^j) \quad (5)$$

Where μ_0^j is the position of j dimension.

μ_{\max}^j is the upper bound

μ_{\min}^j is the lower bound

The initial velocity of each swarm particle is calculated based on (6).

$$v_0^j = [\mu_{\min}^j + r(\mu_{\max}^j - \mu_{\min}^j)/\Delta t] \quad (6)$$

Where Δt = time step value (generation value).

The fitness of each particle swarm is evaluated by running the ANN program. In this paper, MAPE is set as the fitness function. Then, the local and global best particle are determined by searching for the local best optimum position P_k^j at the current iteration k and finds the global best optimum particle position P_g^k within the swarm population. Then, the position of each particle are updated according to (7).

$$\mu_{k+1}^j = \mu_k^j + v_{k+1}^j \Delta t \quad (7)$$

Where,

μ_{k+1}^j is the position of particle j at iteration $k + 1$

v_{k+1}^j is the velocity vector

Δt is the small-time step value.

In order to find the optimal solution in the further generations, the position of the particle need to be modified with respect to the velocity. The velocity of each particle is updated using (8).

$$v_{k+1}^j = w_{v_k}^j + c_1 r_1 \left(\frac{p_k^j - \mu_k^j}{\Delta t} \right) + c_2 r_2 \left(\frac{p_g^k - \mu_k^j}{\Delta t} \right) \quad (8)$$

where,

$r1$ and $r2$ are the random numbers ranging 0 to 1

P_k^j is the best position found by particle j

P_g^k is the best position in the swarm by at time k

W is the Inertia of the particle which controls the exploration properties of the algorithm.

$c1$ is the coefficient of the self-recognition component

$c2$ is the coefficient of the social component.

III. RESULTS AND DISCUSSIONS

The initial parameters of ANN are tabulated in Table 1. Number of hidden nodes, learning rate and momentum rate are tested heuristically in order to obtain the best MAPE and R^2 . Number of hidden nodes are varied between 1 to 20. Meanwhile, learning rate and momentum rate are varied between 0 to 1. The comparison between actual and predicted training results are illustrated in Fig. 4. Fig. 5 presents the details performance of ANN.

TABLE 1. INITIAL PARAMETERS OF ANN

Parameters	ANN
Number of training data	841
Number of testing data	180
Number of input nodes	1
Number of nodes in hidden layer 1	Range from 1 to 20
Number of nodes in hidden layer 2	Range from 1 to 20
Number of nodes in the output layer	1
Learning Rate	Range from 0 to 1
Momentum Rate	Range from 0 to 1
Training algorithm	trainlm
Type of transfer function	Logsig, logsig, purelin

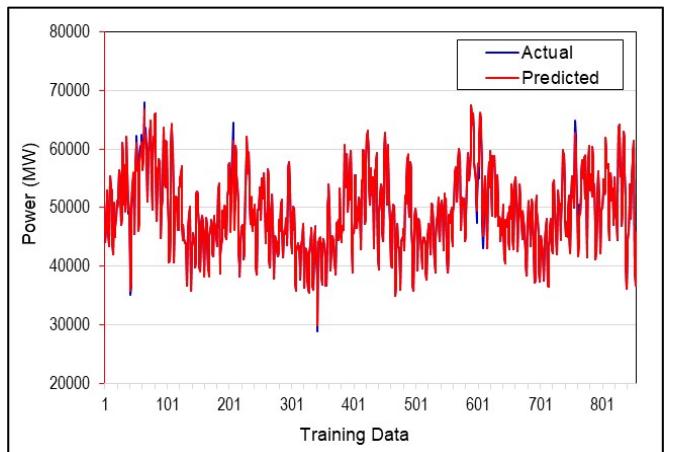


Fig. 4: Results of training data of ANN

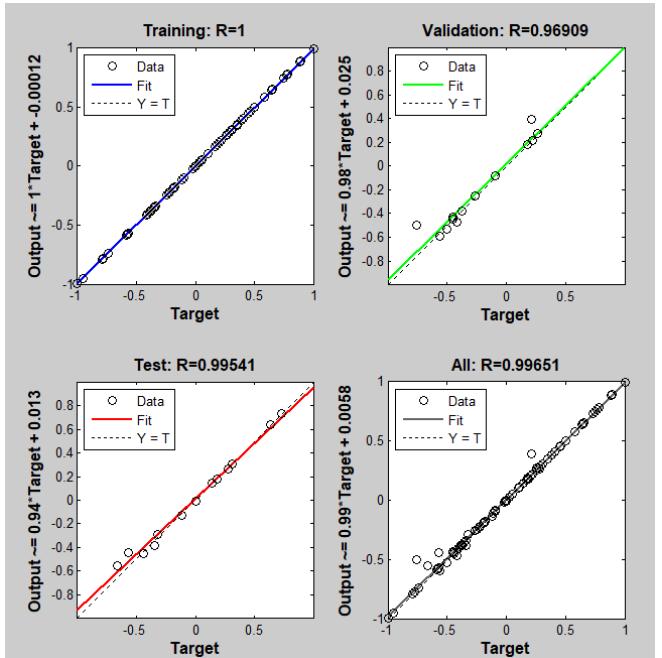


Fig. 5. Regression of ANN

From Fig. 4, it can be observed that the predicted data are similar with the actual data except for certain point where there is slightly different between the predicted and actual data. MAPE value of 0.5249% was obtained when the value of momentum rate is 0.65 and learning rate equals to 0.55. The

coefficient of determination (R^2) value is 0.99651. The R^2 value near to unity indicate that the prediction is good.

In order to improve the accuracy of the prediction, GWO-ANN is introduced. Four variables are optimized which are hidden node 1, hidden node 2, learning rate and momentum rate. The parameters of GWO-ANN are set according to Table 2. The lower and upper bound value for hidden nodes is determined to be in the range of 1 to 20. On the other hand, the lower and upper bound value of both momentum and learning rate are set to 0 and 1 respectively. The results of GWO-ANN and ANN are shown in Table 3.

TABLE 2. PARAMETERS OF GWO-ANN

Parameters	GWO-ANN
Number of search agents	10
Maximum iterations	100
Lower bound value	0.00001, 0.00001, 1,1
Lower bound value	1,1,20,20

TABLE 3. RESULTS OF GWO-ANN VERSUS ANN

Parameters	GWO-ANN	ANN
Number of neurons	5, 8	12, 10
Learning Rate	0.2550	0.65
Momentum Rate	0.8602	0.55
Mean Absolute Percentage Error (MAPE)	0.1084	0.5249
Coefficient of determination, R^2	0.99906	0.99651

From the results tabulated in Table 3, it can be observed that the MAPE value obtained from GWO-ANN is much lower than the ANN. The optimum momentum rate is 0.8602 with the learning rate of 0.255. A lower MAPE and higher R^2 value indicated that the predicted data is near to the targeted data. The comparison between actual and predicted data during training process is illustrated in Fig. 6. The overall regression results are shown in Fig. 7. The result had shown that GWO-ANN has better performance than ANN model as it produced a lower MAPE and better R^2 value.

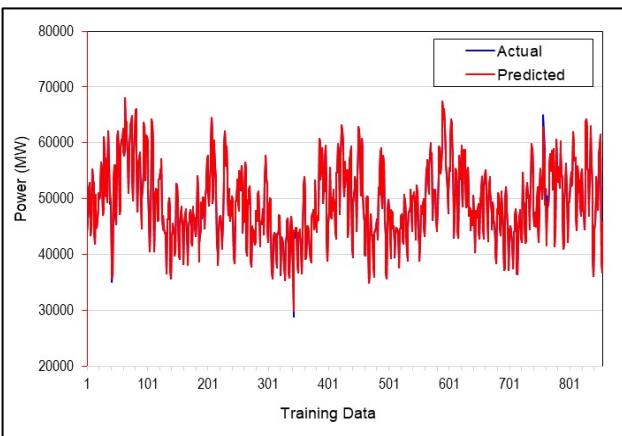


Fig. 6: Results of training data of GWO-ANN

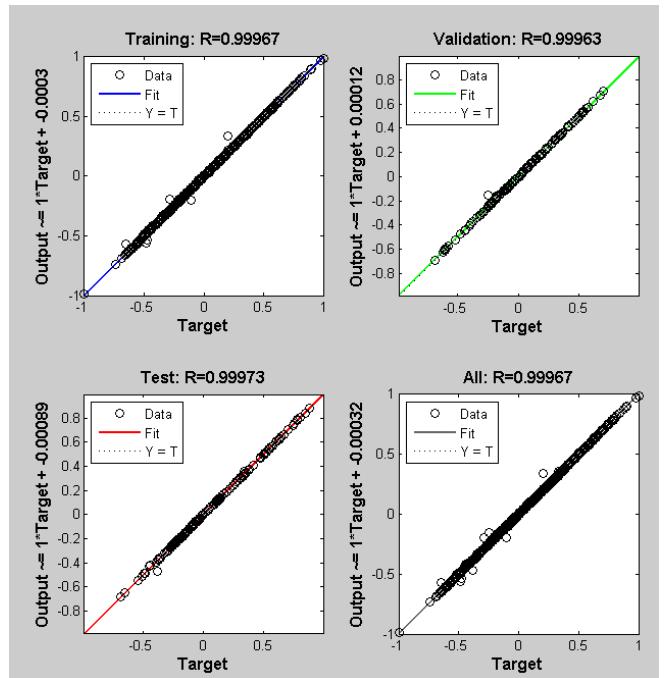


Fig. 7. Regression of GWO-ANN

The performance of GWO-ANN is also compared to PSO-ANN in terms of convergence rate. The convergence graph of GWO-ANN and PSO-ANN is presented in Fig. 8 for 100 iterations. PSO-ANN converges at 70th iteration. The optimum MAPE is 0.3232%. While GWO-ANN converges at 70th iteration. Although GWO-ANN converges later than PSO-ANN, the MAPE value is lower than PSO-ANN. The optimal value of MAPE is 0.1084%. A graph comparing the results of actual and predicted value of the testing data using PSO-ANN and GWO-ANN are represented in Fig. 9 and Fig. 10 respectively.

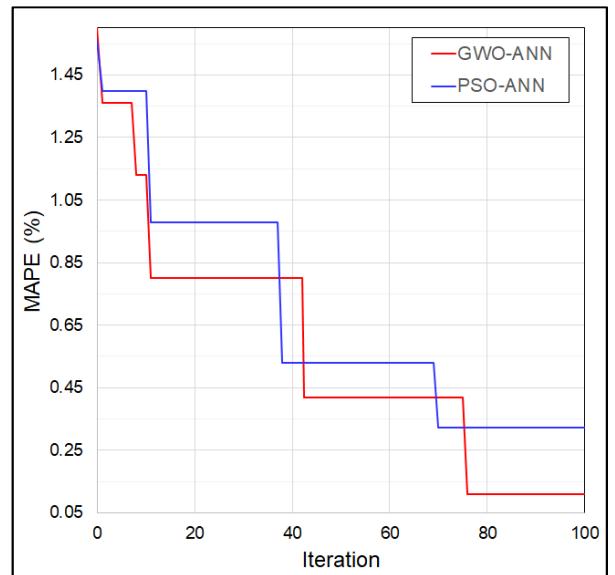


Fig. 8. Convergence graph of GWO-ANN and PSO-ANN

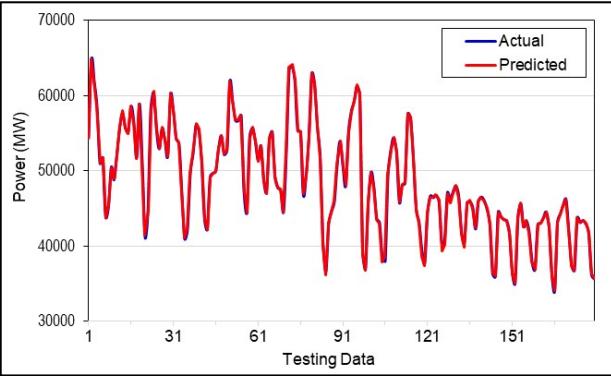


Fig. 9. Results of testing data using PSO-ANN

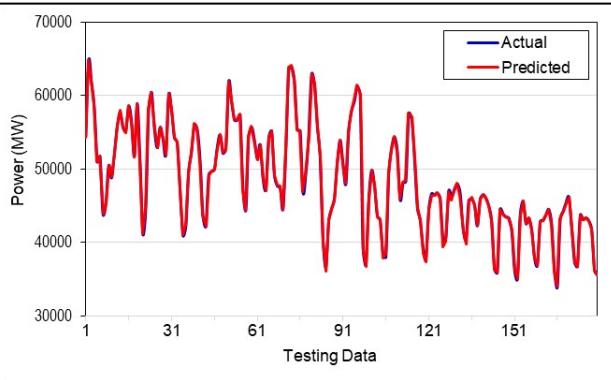


Fig. 10. Results of testing data using GWO-ANN

The red line represents the predicted output while the blue line represents the actual output. The less visible the blue line, the more accurate the load forecast. From the results presented in Fig. 9 and Fig. 10, it can be observed that the predicted values for both techniques are similar to the actual output. However, GWO-ANN has better accuracy as the MAPE is lower.

IV. CONCLUSION

This paper had presented various techniques on load forecasting. In spite of ANN already showing a good prediction performance with a MAPE of 0.5249%, a hybrid ANN was developed to refine the performance of prediction. GWO is used as an optimizer to search for the best momentum rate and learning rate. GWO algorithm mimics the ranking and hunting technique of grey wolves in nature. There are four inputs to the ANN which are power, temperature, humidity and wind speed. The output of the prediction is total power. For comparison, the performance of GWO-ANN was compared to ANN and PSO-ANN. The outcome uncovered that GWO-ANN performed better than other techniques such as ANN and PSO-ANN in terms of MAPE and coefficient of determination. The optimum fitness obtained by GWO-ANN, PSO-ANN and ANN were 0.1084%, 0.3232% and 0.5249% respectively.

REFERENCES

- [21] T. P. Latchoumi, K. Balamurugan, K. Dinesh, and T. P. Ezhilarasi, "Particle Swarm Optimization approach for waterjet cavitation peening," *Measurement: Journal of the International Measurement Confederation*, vol. 141, pp. 184-189, 2019.
- [1] Ö. F. Ertugrul, "Electrical Power and Energy Systems Forecasting electricity load by a novel recurrent extreme learning machines approach," *Int. Journal of Electrical Power and Energy Systems*, vol. 78, p. 429–435, 2016.
- [2] H. K. Alfares and M. Nazeeruddin, "Electric Load Forecasting: Literature Survey and Classification of Methods," , vol. 33, no. 1, 2002," *Int. Journal of Systems Science*, vol. 33, no. 1, pp. 1-7, 2002.
- [3] S. Rasmidatta, "Hourly Weather Forecast Analysis," *Thesis University of Texas*, 2006.
- [4] H. Hamedmoghadam, N. Joorabloo, and M. Jalili, "Australia's Long-Term Electricity Demand Forecasting using Deep Neural Networks," *Neural and Evolutionary Computing*, pp. 1-14, 2018.
- [5] N. J. Hobbs, B. H. Kim, and K. Y. Lee, "Long-Term Load Forecasting using System Type Neural Network Architecture," *Int. Conference on Intelligent Systems Applications to Power Systems*, pp. 1-8, 2017.
- [6] N. Ammar, M. Sulaiman, and A. F. M. Nor, "Long-Term Load Forecasting of Power Systems using Artificial Neural Network and ANFIS," , " *ARPJN Journal of Engineering and Applied Sciences*, vol. 13, no. 3, p. 828–834, 2018.
- [7] Z. Guo, K. Zhou, X. Zhang, and S. Yang, "A Deep Learning Model for Short-Term Power Load and Probability Density Forecasting," *Energy*, vol. 160, pp. 1186-1200, 2018.
- [8] W. Kong, Z. Y. Dong, Y. Jia, D. J. Hill, Y. Xu, and Y. Zhang, "Short-Term Residential Load Forecasting based on LSTM Recurrent Neural Network," *IEEE Transactions on Smart Grid*, vol. 3053, pp. 1-11, 2017.
- [9] W. J. Lee and J. Hong, "A Hybrid Dynamic and Fuzzy Time Series Model for Mid-Term Power Load Forecasting," , " *Int. Journal of Electrical Power and Energy Systems*, vol. 64, p. 1057–1062, 2015.
- [10] B. Liu, J. Nowotarski, T. Hong, and R. Weron, "Probabilistic Load Forecasting via Quantile Regression Averaging on Sister Forecasts," vol. 8, no. 2, pp. 1–3, 2017, " *IEEE Transactions on Smart Grid*, vol. 8, no. 2, pp. 730 - 737, 2017.
- [11] Á. L. Orille-fernández, N. Khalil, and S. B. Rodríguez, "Failure Risk Prediction Using Artificial Neural Networks for Lightning Surge Protection of Underground MV Cables," *IEEE Trans. on Power Delivery*, vol. 21, no. 3, p. 1278–1282, 2006.
- [12] S. N. Pandey, S. Tapaswi, and L. Srivastava, "Nodal congestion price estimation in spot power market using artificial neural network," *IET Generation, Transmission & Distribution*, vol. 2, no. 2, p. 280–290, 2008.
- [13] G. Cardoso, J. G. Rolim, and H. H. Zürn, "Identifying the Primary Fault Section After Contingencies in Bulk Power Systems," , " *IEEE Transactions on Power Delivery* , vol. 23, no. 3, p. 1335–1342, 2008.
- [14] Wunderground, "Philadelphia Daily Forecast," 2019.
- [15] "PJM," *Dayton Load Demand (1/1/2016 - 31/10/2018)*.
- [16] S. Mirjalili, S. M. Mirjalili and A. Lewis, "Grey Wolf Optimizer," *Advances in engineering software*, vol. 69, pp. 46-61, 2014.
- [17] A. Lakum and V. Mahajan, "Optimal placement and sizing of multiple active power filters for radial distribution system using grey wolf optimizer," *7th IEEE Int. Conference on Power Systems*, pp. 1-6, 2017.
- [18] L.I. Wong, M.H. Sulaiman, M.R. Mohamed and M.S. Hong, "Grey Wolf Optimizer for solving economic dispatch problems," *IEEE International Conference on Power and Energy* , pp. 1-6, 2014.
- [19] S. Sukumar, M. Marsadek, A. Ramasamy, H. Mokhlis, "Grey Wolf Optimizer Based Battery Energy Storage System Sizing for Economic Operation of Microgrid," *IEEE Int. Conference on Environment and Electrical Engineering* , pp. 1-6, 2018.
- [20] S. Dahal and H. Salehfar, "Impact of distributed generators in the power loss and voltage profile of three phase unbalanced distribution network," *Int. Journal of Electrical Power and Energy Systems*, vol. 77, pp. 1-8, 2016.