Smart Meter Data-Driven Voltage Forecasting Model for a Real Distribution Network Based on SCO-MLP

Emrah Dokur
Department of Electrical Electronics Eng.
Bilecik Seyh Edebali University
Bilecik, Türkiye
emrah.dokur@bilecik.edu.tr

Ibrahim Sengor
MaREI the SFI Research Centre
University College Cork
Cork, Ireland
isengor@ucc.ie

Nuh Erdogan
School of Science and Technology
Nottingham Trent University
Nottingham, UK
nuh.erdogan@ntu.ac.uk

Ugur Yuzgec
Department of Computer Engineering
Bilecik Seyh Edebali University
Bilecik, Türkiye
ugur.yuzgec@bilecik.edu.tr

Barry P. Hayes
MaREI the SFI Research Centre
University College Cork
Cork, Ireland
barry.hayes@ucc.ie

***Abstract*—Advanced metering infrastructure like smart meter technology has enabled the collection of high-resolution data on voltage, active, and reactive power consumption from end-users in real-time. This paper introduces a new machine learning model, named Single Candidate Optimizer (SCO) – Multi-layer perceptron (MLP), for accurate node voltage forecasting in low voltage (LV) distribution networks with high penetrations of low-carbon technologies. The proposed model utilizes historical active and reactive power measurements in one-minute resolution from smart meters to predict node voltage time series values without requiring the network's electrical model topology and parameters. The computational performance of the MLP framework is improved with the SCO algorithm, which reduces the number of required iterations while maintaining accuracy. The model's performance is evaluated with numerical metrics and compared against Particle Swarm Optimization (PSO) and Differential Evolution (DE)-based models, revealing that the proposed model outperforms both, exhibiting a promising voltage forecasting capability with an average deviation of 1.296 volts relative to the measured values. Overall, this study demonstrates the potential of machine learning and smart meter data for enhancing the stability and efficiency of LV distribution networks.**

***Index Terms*—low carbon loads, low distribution network, smart meter, meta-heuristic, single candidate optimizer, voltage regulation,**

I. Introduction

The modern electric power grid is undergoing a transformation from a centralized generation and distribution system to a more decentralized and dynamic network [1]. This transition towards a more sustainable and resilient grid brings new challenges for utilities and grid operators to maintain a reliable and efficient system operation. Ever-increasing low-carbon technologies in the low voltage (LV) network, such as rooftop solar panels, energy storage systems, electric vehicles (EVs), and heat pumps, require the distribution network operators to take on an active management role. This is due to the emergence of power quality issues, such as over- or undervoltage, spikes, etc., which are becoming more common and challenging to control [2]. Several factors, including load demand size, weather conditions, and fluctuations in renewable energy generation, can affect voltage regulation [3]. Due to the dynamic nature of these reasons, active distribution management is increasingly focusing on fast voltage regulation [4]. This requires an accurate prediction of near-future node voltages.

Voltage predictions can be used for both operational purposes to control node voltages within the statutory limits and planning phases, such as the calculation of the hosting capacity of low-carbon loads [5]. The necessity of real-time decisions for fast voltage fluctuations makes the conventional power flow approach impractical [4]. With the deployment of advanced metering infrastructure (AMI) like smart meter technology, machine learning (ML) based voltage regulation is becoming an emerging research focus [2]. Smart meter technology enables the collection of high-resolution data on voltage and real and reactive power consumption from the end-user in real-time. This data can provide valuable insights into the current state of the distribution system and help improve the maintenance of voltage stability with forecasting methods. The work in [6] examined the forecasting capability of short-term load forecasting approaches for local demand, highlighting their limitations and suggesting simple load models from smart meter data may provide similar prediction accuracies. Hayes et al. in [7] developed three services for distribution network

The work of I. Sengor was supported by Science Foundation Ireland (SFI) under grant no. 12/RC/2302_P2.

energy management systems using AMI data. The authors in [8] proposed a new data-driven method to determine the network topology and load phase connectivity of low-voltage distribution networks using smart meter measurements. The proposed method utilized principal component analysis and graph theory to infer the steady-state network topology. Wang et al. in [9] developed a method using historical smart meter data from the head of the feeder to assess the impact of distributed energy resources on voltage behavior. Similarly, in [10], a method was propounded to estimate network topology, line parameters, and phasing connections in LV distribution systems using smart meter measurements. In mining the smart meter data in these studies, the use of ML techniques such as neural networks (NNs) has shown promise in improving voltage forecasting accuracy using smart meter data [11].

The ML-based prediction relies solely on smart meter data without knowledge of the network topology or parameters, in which a learning model (e.g., a multi-layer perceptron model, MLP) is developed to relate active and reactive power measurements to voltage counterparts for each node in a distribution network. The choice of a suitable ML method is critical from several perspectives, including accuracy, scalability, and fast computation. As such, the established relationship for a studied network can be extrapolated to other LV networks. In [4], a deep reinforcement learning model is proposed to control fast voltage fluctuations in real-time in PV-dominated distribution networks. In [5], a deep NN model is proposed to calculate node voltages in PV-rich LV networks. The authors improved their work in [5] with a new methodology to account for upstream MV network effects and tested its effectiveness on multiple LV feeders simultaneously [12]. These services included demand forecasting, constraint management, and voltage profile forecasting. The study utilized recorded supervisory control and data acquisition (SCADA) and smart meter data from an existing medium voltage distribution network to demonstrate the applicability of their methodology. To achieve fast response, the voltage regulation is formulated as a convex optimization problem and solved by designing a convex NN model in [13]. In these approaches, the learning model aims to minimize the forecasting voltage error by optimizing the model parameters. To tune the parameters, various meta-heuristic algorithms, such as particle swarm optimization (PSO), gray-wolf optimization algorithms [14], or stochastic optimization algorithms, like ADAM [15] have been predominantly used in various ML-based forecasting applications. Among those, Single Candidate Optimizer (SCO) has recently gained considerable attention due to its innovative approach and promising results with significantly reduced computation cost and memory requirements [16]. It is shown that SCO can converge to the optimal solution faster compared to the other algorithms in [17]. However, the performance of SCO depends on the problem type and still needs to be explored, in particular for quasi-real-time applications. This algorithm has also shown the potential to be hybridized with other meta-heuristic algorithms in [17].

The aim of this study is to propose a new ML model to forecast node voltages in LV distribution networks with high penetration of low-carbon loads such as PVs, heat pumps, and EVs, leading to real-time voltage control. The model builds upon an MLP framework that utilizes the SCO algorithm to enhance convergence efficiency by reducing the number of iterations required while maintaining accuracy. As such, the computational performance is improved. The model solely uses smart meter active and reactive powers and voltage measurements in one-minute resolution collected for a real distribution network and does not need to know its electrical model topology and parameters. The model accuracy performance is shown through several numerical metrics and compared with that of well-known optimizer-based models, including PSO and differential evolution (DE).

II. Methodology

A. Single Candidate Optimizer and Multi-Layer Perceptron (SCO-MLP)

The SCO is a new metaheuristic optimization algorithm that belongs to the class of single-solution-based methods proposed by Shami et al. in [16]. It is designed to search for the optimal solution to a given problem by iteratively improving a single-candidate solution. The main advantage of SCO is its simplicity and ease of implementation. The algorithm does not require a population of candidate solutions or complex operators such as selection, crossover, or mutation. Instead, it focuses on improving a single solution, which can be particularly useful for problems where the search space is small, or the evaluation of each candidate solution is computationally expensive.

The algorithm divides the optimization process into two phases, with different ways of updating the position of the candidate solution in each phase. While single-solution-based algorithms and two-phase approaches are established optimization methods, they have been used separately. However, the proposed approach combines the single-candidate approach with the two-phase strategy to create a strong algorithm. The crucial feature of this algorithm is that it uses a unique set of equations to update the candidate solution's position based solely on its current location. The aim of the two-phase strategy is to achieve diversity and balance between exploration and exploitation. During the first phase of SCO, the candidate solution modifies its position according to the following process [16]:

$$x_j = \begin{Bmatrix} gbest_j + (w\left|gbest_j\right|) & if \quad r_1 < 0.5 \\ gbest_j - (w\left|gbest_j\right|), & else \end{Bmatrix}, \tag{1}$$

where r_1 is a random number in the interval [0, 1]. The mathematical definition of w is described as:

$$w(t) = e^{-\left(\frac{bt}{T}\right)^b}, \tag{2}$$

where the variables b, t, and T represent a constant value, the current iteration or function evaluation number, and the maximum number of function evaluations, respectively. As the second phase of SCO, the candidate solution changes its position as seen below:

$$x_j = \begin{cases} gbest_j + ((r_2 w\,(ub_j - lb_j)) & \text{if } r_2 < 0.5 \\ gbest_j - ((r_2 w\,(ub_j - lb_j)) & \text{else,} \end{cases} \tag{3}$$

where the variable r_2 represents a random value between 0 and 1. Additionally, ub_j and lb_j represent the upper and lower bounds of a boundary, respectively. The parameter w is critical in SCO, as it controls the balance between exploring new solutions and exploiting the current best solution. SCO generates a single candidate solution, x, randomly at the outset of the optimization process. This solution is then updated through a series of iterations to improve its performance. The following is the generation of the initial potential solution:

$$x_j = lb_j + r_3(ub_j - lb_j). \tag{4}$$

Detailed information about the SCO process including the candidate solution switch from exploitation to exploration can be obtained from [16].

The MLP model is a type of artificial NN that consists of multiple layers of interconnected nodes, or neurons. The neurons are organized into input, hidden, and output layers. The input layer receives the input data, and the output layer produces the output values. Each neuron in MLP is associated with a weight vector, which determines the strength of the connection between the neuron and its inputs. Each input v_k to a neuron, m, is multiplied by an adaptive coefficient, w_{mk}, called weight, and then the weighted sum of the inputs is calculated using a nonlinear activation function (φ) such as a sigmoid, hyperbolic tangent, etc. as follows:

$$y_m = \varphi\left(\sum_{k=1}^{n} w_{mk}.v_k + b_m\right), \tag{5}$$

where n and y_m represent the number of inputs and the symbolic function of the predicted results, respectively.

During the training process, the weights are adjusted to minimize the error between the predicted output and the actual output. The SCO is adapted to tune the weight and bias values of the MLP parameters in this paper. This hybrid model is called SCO-MLP.

In this study, the SCO-MLP relates active power (P) and reactive power (Q) measurements from smart meter data in the input layer to voltage measurements in the output layer. Our focus is to forecast voltage deviations for each node in a real LV network. Active and reactive powers and voltage measurements in one-minute resolution for each node collected for a rural LV network with heavily low-carbon loads are used. The choice of input is of crucial importance for the accuracy of the model. By utilizing historical time series measurements of active power $P_i = [P_{i,1}, P_{i,2}, ...P_{i,T}]$, reactive power $Q_i = [Q_{i,1}, Q_{i,2}, ...Q_{i,T}]$, and voltage, $V_i = [V_{i,1}, V_{i,2}, ...V_{i,T}]$), the SCO-MLP model is first trained to capture the nonlinear relationships between the inputs (P_i and Q_i) and the corresponding outputs (V_i) for the i^{th} node, given by 6. Both w_{mk} and b_m are tuned to enhance the performance of the model. The voltage forecasting for each node is represented as follow:

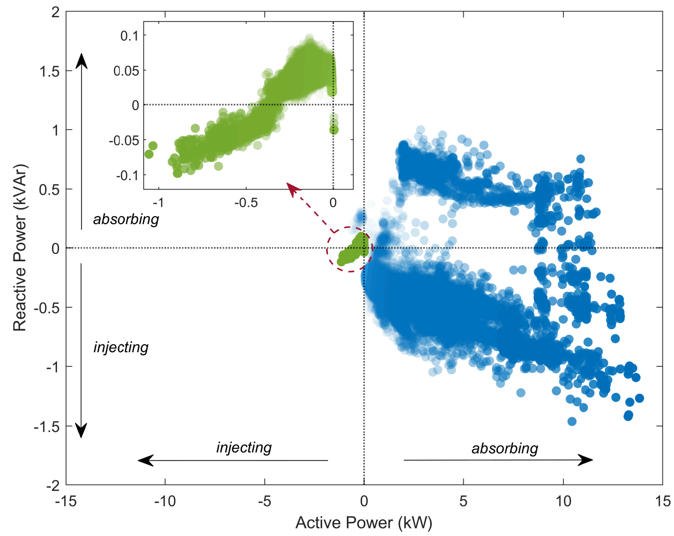

Fig. 1. Reactive Power vs. Active Power measurements for a single customer for one month.

$$V_i = f_{SCO-MLP}\left(P_i, Q_i, w_{mk,i}, b_{m,i}\right) \tag{6}$$

B. Overview of Smart Meter Data

Smart meters have revolutionized the energy sector by enabling the collection of detailed energy consumption data at a high temporal resolution. This data can be used to provide valuable insights into energy consumption patterns and inform decisions related to energy infrastructure planning and management. The historical smart meter data belonging to a distribution network where a number of customers have low-carbon technologies in Ireland has been used in this study. The customers in the network may have one or all of the low-carbon technologies, such as heat pumps, electric vehicles, rooftop PVs, and battery energy storage units. It is worth noting that the used smart meter data has a one-minute time resolution.

The scatter plot displayed in Fig. 1 illustrates the relationship between the active power and reactive power of a customer. The plot indicates whether power is absorbed or imported from the grid (positive values) or injected or exported into the grid (negative values). It should be noted that this customer has a rooftop PV generation unit together with its residential loads. In this figure, blue dots represent the net active-reactive power profile, and green dots stand for PV's active-reactive power profile. While the maximum active power consumption of the customer is 13.83 kW, the maximum power generated by its PV unit is 1.13 kW. Figure 2 shows the time-series voltage values of a single customer for a given period with a one-minute resolution. As seen in the figure, the customer's voltage level varies in a voltage range

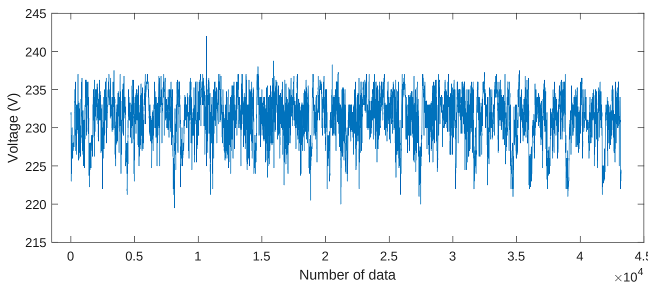

Fig. 2. Time-series voltage for a customer.

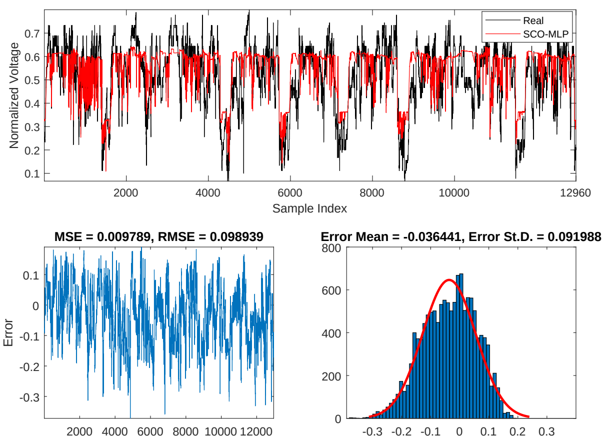

Fig. 3. Forecasting node voltage test results of the proposed model.

of 207 V to 253 V, which conforms to European Standard EN50160.

III. Results and Discussion

In this section, forecasting results of SCO-MLP is discussed in detail. Moreover, the SCO-MLP results are compared with well-known metaheuristic algoritms-MLP such as Particle Swarm Optimization (PSO-MLP) and Differential Evolution (DE-MLP). Herein, the purpose of the first experiment is to demonstrate the performance of the SCO-MLP data algorithm over in detail its fit with the target data. Secondly, the voltage forecasting results are to be compared with time series graphs, and performance metrics.

The collected time series data for one month, comprising 43,200 points, is divided into 70 and 30 percents to train and test the model, respectively. The data is first normalized according to their maximum and minimum values. Fig. 3 displays the forecasting normalized voltage results for a single node. As can be seen in Fig. 3, the error values are also gathered on the zero axis. It is an acceptable indicator of the reliability of the model. It is seen that the SCO-MLP model mostly captured the trend of voltage behaviors, whereas undervoltage values are better captured than overvoltage values.

The SCO-MLP algorithm was run for 3000 iterations, and Fig. 4 shows the variation of the root mean square error (RMSE) with respect to the number of iterations. It is shown

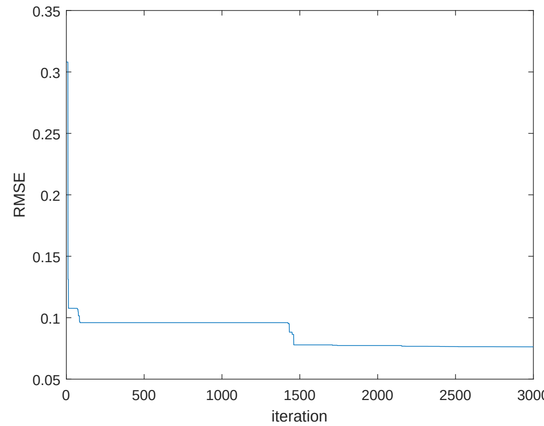

Fig. 4. RMSE deviation with respect to the number of iterations.

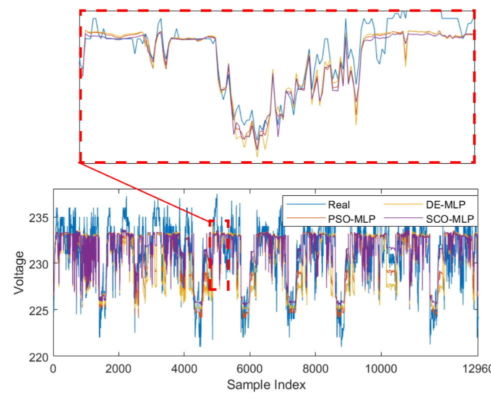

Fig. 5. Voltage forecasting results of a single node.

that the global convergence point has been reached rapidly, as seen from the 100th and 1500th iterations. It is evident that the SCO algorithm can minimize error rapidly due to its simple structure, which does not rely on a population of candidate solutions or complex operators such as selection, crossover, and mutation.

This performance of the proposed SCO-MLP is assessed against well-known meta-heuristic-based forecasting models, namely PSO-MLP and DE-MLP. The time series voltage forecasting results for a single node are shown in Fig. 5 for all implemented models. It is seen that the trend of voltage at the time of forecasting was captured by all the models. To quantify the performance, some metrics, such as RMSE, mean square error (MSE), and mean absolute error (MAE), have been employed. The numerical results are summarized in Table I. The proposed model outperforms for all the metrics considered with RMSE, MSE, and MAE values of 1.7595 V, 3.0958 V^2, and 1.4282 V respectively. Since the analysis is conducted at minute resolution and the results have been denormalized, the voltage error ranges obtained from all three models can be considered acceptable.

Fig. 6 shows the Taylor diagrams for the results of the

TABLE I
COMPARISON OF ERROR PERFORMANCE METRICS VALUES OF THE IMPLEMENTED MODELS.

Models	RMSE	MSE	MAE
PSO-MLP	1.8476	3.4137	1.4782
DE-MLP	2.1689	4.7040	1.7086
SCO-MLP	**1.7595**	**3.0958**	**1.4282**

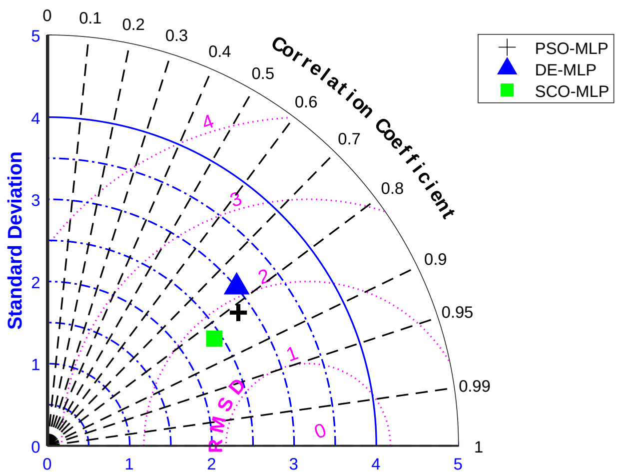

Fig. 6. Comparative results using Taylor diagram.

implemented models. The diagram presents how the correlation coefficient, root mean square deviation (RMSD), and standard deviation are interrelated, enabling a straightforward comparison of forecasting accuracy. The closer the correlation coefficient value to 1, the more linear the relationship between the original and predicted data is. A correlation coefficient value closer to 1 indicates a stronger linear association between the forecasted and original data. The green square represents the results for the proposed SCO-MLP method, which has lower RMSD and standard deviation values and a correlation coefficient closer to 1. Based on these findings, the SCO-MLP model is deemed reliable and effective for voltage forecasting in the studied distribution network.

IV. CONCLUSIONS

In this study, a voltage forecasting model has been proposed for LV distribution networks with low carbon loads where fast voltage deviations occur due to the nature of the loads. The proposed MLP model incorporates SCO to take advantage of its fast convergence rate to global optimum points while minimizing the forecasting error. The proposed model has been tested for the calculation of node voltage time series values for a real LV distribution network. The model can demonstrate promising voltage forecasting with an average voltage deviation of 1.296 volts relative to the measured values. The model's performance has also been shown superior as compared to two well-known meta-heuristic optimization algorithms. Future research will focus on adapting the model to representative LV distribution networks (e.g., urban, suburban, rural, etc.) and improving the model's performance to further predict overvoltage values.

REFERENCES

[1] Y. Li, M. Han, Z. Yang, and G. Li, "Coordinating flexible demand response and renewable uncertainties for scheduling of community integrated energy systems with an electric vehicle charging station: A bi-level approach," *IEEE Transactions on Sustainable Energy*, vol. 12, no. 4, pp. 2321–2331, 2021.

[2] A. F. Bastos, S. Santoso, V. Krishnan, and Y. Zhang, "Machine learning-based prediction of distribution network voltage and sensor allocation," in *2020 IEEE Power Energy Society General Meeting (PESGM)*, 2020, pp. 1–5.

[3] R. Smolenski, P. Szczesniak, W. Drozdz, and L. Kasperski, "Advanced metering infrastructure and energy storage for location and mitigation of power quality disturbances in the utility grid with high penetration of renewables," *Renewable and Sustainable Energy Reviews*, vol. 157, p. 111988, 2022.

[4] D. Cao, J. Zhao, W. Hu, F. Ding, N. Yu, Q. Huang, and Z. Chen, "Model-free voltage control of active distribution system with pvs using surrogate model-based deep reinforcement learning," *Applied Energy*, vol. 306, p. 117982, 2022.

[5] V. Bassi, L. Ochoa, and T. Alpcan, "Model-free voltage calculations for pv-rich lv networks: Smart meter data and deep neural networks," in *2021 IEEE Madrid PowerTech*, 2021, pp. 1–6.

[6] B. Hayes, J. Gruber, and M. Prodanovic, "Short-term load forecasting at the local level using smart meter data," in *2015 IEEE Eindhoven PowerTech*, 2015, pp. 1–6.

[7] B. P. Hayes and M. Prodanovic, "State forecasting and operational planning for distribution network energy management systems," *IEEE Transactions on Smart Grid*, vol. 7, no. 2, pp. 1002–1011, 2016.

[8] S. J. Pappu, N. Bhatt, R. Pasumarthy, and A. Rajeswaran, "Identifying topology of low voltage distribution networks based on smart meter data," *IEEE Transactions on Smart Grid*, vol. 9, no. 5, pp. 5113–5122, 2018.

[9] Y. Wang, M. Z. Liu, and L. F. Ochoa, "Assessing the effects of der on voltages using a smart meter-driven three-phase lv feeder model," *Electric Power Systems Research*, vol. 189, p. 106705, 2020.

[10] V. C. Cunha, W. Freitas, F. C. L. Trindade, and S. Santoso, "Automated determination of topology and line parameters in low voltage systems using smart meters measurements," *IEEE Transactions on Smart Grid*, vol. 11, no. 6, pp. 5028–5038, 2020.

[11] P. Rodríguez-Pajarón, A. H. Bayo, and J. V. Milanović, "Forecasting voltage harmonic distortion in residential distribution networks using smart meter data," *International Journal of Electrical Power & Energy Systems*, vol. 136, p. 107653, 2022.

[12] V. Bassi, L. F. Ochoa, T. Alpcan, and C. Leckie, "Electrical model-free voltage calculations using neural networks and smart meter data," *IEEE Transactions on Smart Grid*, pp. 1–1, 2022.

[13] Y. Chen, Y. Shi, and B. Zhang, "Data-driven optimal voltage regulation using input convex neural networks," *Electric Power Systems Research*, vol. 189, p. 106741, 2020.

[14] T. İnaç, E. Dokur, and U. Yüzgeç, "A multi-strategy random weighted gray wolf optimizer-based multi-layer perceptron model for short-term wind speed forecasting," *Neural Computing and Applications*, vol. 34, no. 17, pp. 14 627–14 657, 2022.

[15] E. Dokur, N. Erdogan, and S. Kucuksari, "Ev fleet charging load forecasting based on multiple decomposition with ceemdan and swarm decomposition," *IEEE Access*, vol. 10, pp. 62 330–62 340, 2022.

[16] T. M. Shami, D. Grace, A. Burr, and P. D. Mitchell, "Single candidate optimizer: a novel optimization algorithm," *Evolutionary Intelligence*, pp. 1–25, 2022.

[17] X. Yuan, M. A. Karbasforoushha, R. B. Syah, M. Khajehzadeh, S. Keawsawasvong, and M. L. Nehdi, "An effective metaheuristic approach for building energy optimization problems," *Buildings*, vol. 13, no. 1, p. 80, 2023.