International Journal of Informatics and Communication Technology (IJ-ICT)
Vol. 14, No. 3, December 2025, pp. 830~836
ISSN: 2252-8776, DOI: 10.11591/ijict.v14i3.pp830-836

Electric load forecasting using ARIMA model for time series data

Balasubramanian Belshanth, Haran Prasad, Thirumalaivasal Devanathan Sudhakar
Department of Electrical and Electronic Engineering, St. Joseph's College of Engineering, Chennai, India

Article Info

Article history:

Received Sep 10, 2024
Revised Mar 15, 2025
Accepted Jun 9, 2025

Keywords:

ARIMA
Data
Load forecasting
Machine learning
Predictions
STLF
Time series

ABSTRACT

Any country's economic progress is heavily reliant on its power infrastructure, network, and availability, as energy has become an essential component of daily living in today's globe. Electricity's distinctive quality is that it cannot be stored in huge quantities, which explains why global demand for home and commercial electricity has grown at an astonishing rate. On the other hand, electricity costs have varied in recent years, and there is insufficient electricity output to meet global and local demand. The solution is a series of case studies designed to forecast future residential and commercial electricity demand so that power producers, transformers, distributors, and suppliers may efficiently plan and encourage energy savings for consumers. However, load prognosticasting has been one of the most difficult issues confronting the energy business since the inception of electricity. This study covers a new one–dimensional approach algorithm that is essential for the creation of a short–term load prognosticasting module for distribution system design and operation. It has numerous operations, including energy purchase, generation, and infrastructure construction. We have numerous time series forecasting methods of which autoregressive integrated moving average (ARIMA) outperforms the others. The auto–regressive integrated moving average model, or ARIMA, outperforms all other techniques for load forecasting.

Corresponding Author:

Thirumalaivasal Devanathan Sudhakar
Department of Electrical and Electronic Engineering, St. Joseph's College of Engineering
Chennai, India
Email: sudhakartd@stjosephs.ac.in

1. INTRODUCTION

Short–term load forecasting is essential for the electrical system to operate in a reliable way. When making decisions about unit dedication, cost–effective supply, automated generating control, safety evaluation, planned maintenance, and distribution of energy, the power system must take future load behavior into account [1]. In markets where power is competitive, load estimates are crucial for every electrical transaction [2]. In order to maintain a constant supply of electricity and support the expansion of the financial system, power providers must anticipate loads. Therefore, the financial system and the management of the operation of electrical structures are greatly impacted by the forecasts' correctness. Thus, significantly more advanced and accurate forecasting technologies are required for today's power systems. [3], [4]. In recent decades, research in this field has led to the development of more precise forecasting techniques because of the significance of short–term load forecasting (STLF). The majority of methods rely on the recording and assessment of time. The majority of time collection models rely on statistical techniques [1], [3]-[6]. The time cumulative estimate for each day of the week or month is the main basis for calculating the 24 load projections in this paper's autoregressive integrated moving average (ARIMA) model for STLF [7]-[10].

Statistical models employ intricate computations to forecast the present value of a variable by utilizing mathematical aggregates of the variable's historical values. This article's goal is to present an ARIMA model [11]-[17] for STLF that, mostly using time–cumulative estimates, computes 24 load projections for each day of the week and month. In order to determine the minimum MAPE value and, consequently, precise short–term load projections, a description of the suggested version of the order property and a parameter estimation process are given. A collection of time series load patterns are studied using this method.

2. MACHINE LEARNING

The scientific study of statistical models and methods that computer systems employ to do certain tasks without explicit programming is known as machine learning, or ML. Many domains use ML techniques, such as predictive analytics, data mining, image processing, and many more. A learning algorithm that has found out how to rank webpages is one of the reasons why an online search engine like the internet works so efficiently each time it is used to search the internet. The capacity of algorithms to do tasks automatically once they have learnt how to handle data is the main advantage of ML. Algorithms in ML are created to assist computers in learning from their experiences by training them to generate a result by utilizing a set of input data. ML algorithms are trained on datasets and subsequently provide output for test data. These methods make use of computational techniques to train the model and extract information from the pre–existing data. As the number of samples increases, algorithm performance improves. Several methods and strategies are used in this iterative learning process, such as reinforcement learning, unsupervised learning, and supervised learning. Algorithms that are trained on labeled datasets using supervised learning map input data to corresponding output labels. While reinforcement learning focuses on decision–making through interacting with an environment and getting feedback in the form of rewards or penalties, unsupervised learning focuses on finding patterns and structures within unlabeled data. The efficacy of ML is contingent upon the algorithms' capacity to extrapolate patterns and correlations from training data in order to generate precise predictions or judgments on novel data. Additionally, major developments in tasks like image recognition, natural language processing, and autonomous systems have been made possible by advances in ML techniques, such as deep learning neural networks [18], [19]. All things considered, ML gives computers the ability to learn and adapt on their own, which advances automation, efficiency, and decision–making in a variety of fields and applications. As algorithms advance and change throughout time, there is still a lot of promise in using ML to tackle challenging issues and create new possibilities.

In ML, a statistical model is a framework that is used to depict the connection between the goal variable and the input data without directly stating a mathematical expression. As an alternative, statistical models capture trends and relationships found in the data, enabling the creation of forecasts or conclusions based on fresh, unforeseen information. Fundamentally, a statistical model outlines a series of presumptions regarding the generation of data and the relationships between variables. The model uses these presumptions to inform its learning from the data and to help it formulate predictions and conclusions. Different types of data and activities require different types of statistical models, which can vary from basic linear models to intricate non–linear models. A statistical model uses variables or characteristics from the input data to predict the target variable. By estimating parameters that most accurately characterize the link between attributes and the objective, the model gains knowledge from past data. After being trained, the model may be used to fresh data to forecast outcomes or learn more about the underlying processes that produced the data. Significantly, statistical models offer metrics of uncertainty in addition to predictions. By putting a number on the uncertainty around forecasts, they enable users to assess the predictability of the model's output and base their judgments on the degree of trust in the predictions. All things considered, statistical models are effective ML tools that provide a versatile framework for identifying patterns in data, generating predictions, and coming to insightful conclusions. They are extensively used to gather information and guide decision–making processes in a variety of industries, including as marketing, finance, healthcare, and more.

2.1. Autoregressive integrated moving average

A generalization of an autoregressive moving average model is the autoregressive integrated moving average model. To get a deeper understanding of the data or forecast future series points, these two models are fitted to time series data. If statistics show that the mean function is non–stationaristic, the non–stationarity can be eliminated from the mean function by implementing an initial differencing step once or more times. ARIMA models are employed in these circumstances. When the seasonal component appears in a time series, seasonal–differencing can be employed to eliminate it. The Wold's decomposition theorem, for example, asserts that the ARIMA model is theoretically sufficient to explain a regular wide–sense stationary time series, which motivates us to make stationary a non–stationary time series [20]-[25]. The ARIMA model can be understood by outlining all its components as follows:

a) Autoregression: this method explains a model that displays the regression of a variable versus its lag or previous values.
b) Integrated: it represents the process of taking raw observations and stabilizing the time series (i.e., by replacing the data values with the prior values).
c) Moving average: this technique comprises the correlation between the observation and the residual error of the moving average model used to the delayed observation.

In conclusion, an ARIMA model is represented by the notation ARIMA (p, d, q), where p is the autoregressive component's order. The amount of differencing needed to reach stationarity is denoted by d. The moving average component's order is denoted by q.

3. RESULTS AND DISCUSSION

The results are discussed in the following section. Tables 1 and 2 shows the raw data and the resampled data respectively. Data are fetched from the Kaggle (data science library) and the data are in the form of national demand, date, transmission load and the data contains of five years load. Table 1 provides a sample set of input raw data.

For analysis purpose, the date and national demand are considered since the data is to be converted into a time series for to be forecasted. The data are in hourly order and the data has to be reordered in weekly order and 273 weekly data are obtained after resample. A sample set of resampled data is given in Table 2.

Table 1. Data sample

Date and time	Nat_demand	T2Mtoc	QV2Mtoc	TQLtoc	W2Mtoc	T2Msan
1/3/2015 1:00	970.345	25.86526	0.018576	0.016174	21.85055	23.48245
1/3/2015 2:00	912.1755	25.89926	0.018653	0.016418	22.16694	23.39926
1/3/2015 3:00	900.2688	25.93728	0.018768	0.01548	22.45491	23.34353
1/3/2015 4:00	889.9538	25.95754	0.01889	0.016273	22.11048	23.23879
1/3/2015 5:00	854.9157	25.99355	0.018986	0.015926	22.41266	22.99508
1/3/2015 6:00	829.6077	26.02504	0.019091	0.015862	22.51944	24.28524
1/3/2015 7:00	804.2996	26.05653	0.019197	0.015798	22.62622	26.18969

Table 2. Resampled data

Date (mm\dd\yy)	National demand
1/11/2015	181919.6224
1/18/2015	188082.3152
1/25/2015	179448.7184
2/1/2015	184393.4256
2/8/2015	187290.1846
2/15/2015	180467.2458
2/22/2015	170497.6427
3/1/2015	189914.1826
3/8/2015	190115.6758
3/15/2015	193286.758
3/22/2015	184950.965
3/29/2015	186503.6539
4/5/2015	176822.2559
4/12/2015	188654.9909

3.1. Training data set

The information that the computer uses to learn how to process information is called the training dataset. ML makes use of algorithms to replicate the ways in which the human brain processes various inputs and evaluates them to generate synaptic activations in individual neurons. A large portion of this process is replicated by artificial neurons using software, such as ML and neural network applications that offer incredibly accurate representations of how human brain processes and operate. Here, datasets are trained for five years (i.e, the training data is 273 training weeks).

3.2. Test data set

Next step is to test the model using the test dataset once it has been trained using the training dataset. This dataset assesses the model's performance and verifies that it can effectively generalize to new or untested datasets. The load data forecast occurs in the interval between the datetime and the weekly load. The pictoral representation of prediction without exogenous variables in the ARIMA model is given in Figure 1. The graph show the prediction of next five year and the predicted data are not accurate due to lack of sufficient data. To overcome it is necessary to involve the exogenous variable like weekly transmission load data, yearly load data to get accurate results.

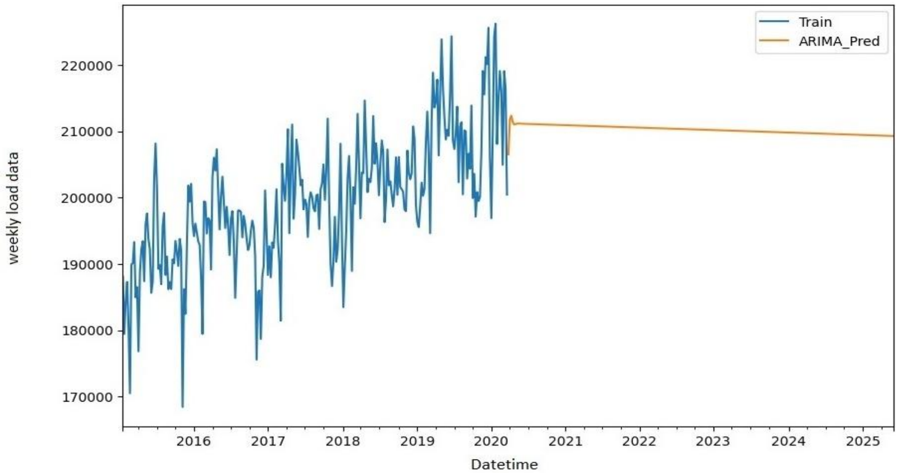


Figure 1. The pictoral representation of training data and prediction without exogenous variables

3.3. Exogenous variables

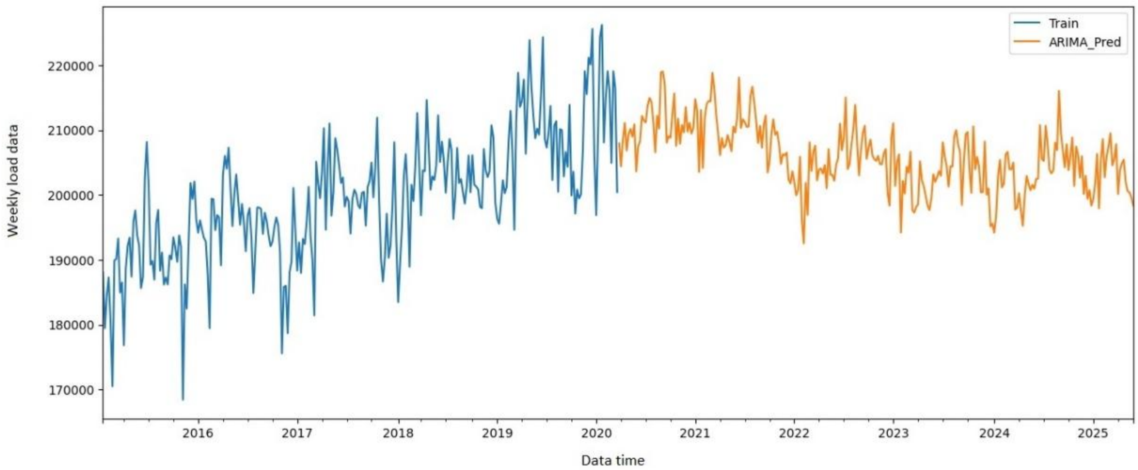
A variable in a mathematical or statistical model that is independent of other variables is called an exogenous variable. It is often referred to as an independent or predictor variable since it is used to forecast the values of other variables in the model. Exogenous variables, which are frequently employed to describe the behavior of the dependent variable, are generally seen as being outside the model's control.

3.4. Features of exogenous variables

Exogenous variables are added to the model to make more accurate and these variables data are resampled weekly to match with the model. In others hands these are like a sufficient data for insufficient model to get a clear prediction. The transmission station load (T2M) and substation load (QV2M) are as exogenous parameters.

3.5. Prediction after adding exogenous variable dataset

To overcome the deawback shown in Figure 1, the exogenous variable data set is added to the ARIMA model. By doing so the output obtained coincides with the training data which is shown in blue and the output obtained for the test data is given in orange. The pictoral representation of prediction with exogenous variables included in the ARIMA model is given in Figure 2.



Note: The lines are in blue (train data) is coincide with lines are in orange (test data)

Figure 2. The pictoral representation of training data and prediction with exogenous variables

The output from the training dataset are matching with the testing dataset which is load prediction. In others hand, the load data from 2020 to 2025 are resemble of load data of year 2015 to 2020. The result from this graph is aligned and resembles with the precious year.

4. CONCLUSION

When utilizing time series data for short–term electric load prognostic, the ARIMA model has shown to be a useful tool. It provides a dependable technique for projecting power demand over brief time horizons by recognizing and forecasting intricate patterns in the data. The ARIMA model gives utilities and energy suppliers critical insights for controlling grid stability, optimizing resource allocation, and improving operational efficiency by utilizing historical load data and adding seasonality, trends, and cyclical patterns. Although additional investigation and enhancement may be required to tackle certain issues and enhance precision, the ARIMA model remains a viable strategy for fulfilling the dynamic requirements of the contemporary energy environment. To sum up, the ARIMA model has shown to be a useful tool for short–term electric load forecasting since it provides precise estimates based on time series data. ARIMA shows its effectiveness in modeling and predicting power demand by capturing both trend and seasonal components. This helps with resource allocation, grid management, and decision–making processes within the energy sector. Further improvements and tinkering with ARIMA and other forecasting approaches promise to consistently increase the accuracy and dependability of short–term electric load projections as technology and methodology advance, enabling a more robust and sustainable energy infrastructure. In conclusion, there have been encouraging outcomes when short–term electric load forecasting has been done using the ARIMA model. Through the use of time series data, ARIMA is able to accurately anticipate power consumption by capturing trends and seasonal patterns. In the energy sector, this strategy helps with resource management and decision–making. ARIMA and related forecasting approaches are well–positioned to significantly improve the accuracy and dependability of short–term electric load projections as technology and methodology continue to evolve, assisting in the creation of an energy infrastructure that is more robust and efficient.

FUNDING INFORMATION

Authors state there is no funding involved.

AUTHOR CONTRIBUTIONS STATEMENT

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Balasubramanian Belshanth	✓	✓	✓	✓	✓	✓		✓	✓	✓			✓	
Haran Prasad	✓	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓		
Thirumalaivasal Devanathan Sudhakar	✓		✓	✓		✓	✓	✓		✓	✓	✓	✓	

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

DATA AVAILABILITY

Data availability is not applicable to this paper as no new data were created or analyzed in this study.

REFERENCES

[1] E. Chodakowska, J. Nazarko, and Ł. Nazarko, "ARIMA models in electrical load forecasting and their robustness to noise," *Energies*, vol. 14, pp. 1–22, 2021, doi: 10.3390/en14237952.

[2] S. G. N and G. S. Sheshadri, "Electrical load forecasting using time series analysis," in *Proceedings of the 2020 IEEE Bangalore Humanitarian Technology Conference (B-HTC)*, India, 2020, pp. 1–6, doi: 10.1109/B-HTC50970.2020.9297986.

[3] M. Elsaraiti, G. Ali, H. Musbah, A. Merabet, and T. Little, "Time series analysis of electricity consumption forecasting using ARIMA model," in *Proceedings of the IEEE Green Technologies Conference (GreenTech)*, USA, 2021, pp. 259–262, doi: 10.1109/GreenTech48523.2021.00049.

[4] B. Nepal, M. Yamaha, A. Yokoe, and T. Yamaji, "Electricity load forecasting using clustering and ARIMA model for energy management in buildings," *Energy and Built Environment*, vol. 3, no. 1, pp. 62–76, 2020, doi: 10.1002/2475-8876.12135.

[5] H. W. Rotib, M. B. Nappu, Z. Tahir, A. Arief, and M. Y. A. Shiddiq, "Electric load forecasting for internet of things smart home using hybrid PCA and ARIMA algorithm," *International Journal of Electrical and Electronic Engineering & Telecommunications*, vol. 10, no. 6, pp. 425–430, 2021, doi: 10.18178/ijeetc.10.6.425-430.

[6] B. Farsi, M. Amayri, N. Bouguila, and U. Eicker, "On short-term load forecasting using machine learning techniques and a novel parallel deep LSTM-CNN approach," *IEEE Access*, vol. 9, pp. 31191–31212, 2021, doi: 10.1109/ACCESS.2021.3060290.

[7] W. C. Hong, S. K. Fan, T. C. Yu, and C. W. Chen, "Short-term load forecasting using machine learning techniques: a review and research framework," *Energies*, vol. 11, no. 9, p. 2322, 2018, doi: 10.3390/en11092322.

[8] P. Pinson and H. Madsen, "Time series forecasting for smart grids: a review and a data–driven approach," *Renewable and Sustainable Energy Reviews*, vol. 81, pp. 1781–1817, 2018, doi: 10.1016/j.rser.2017.06.027.

[9] Y. Chen, Z. Y. Dong, Y. Huang, Q. Zhu, and G. Y. Li, "Short–term load forecasting using LSTM recurrent neural network," in *Proceedings of the 2018 IEEE International Conference on Power System Technology (POWERCON)*, 2018, pp. 1–6, doi: 10.1109/POWERCON.2018.8602265.

[10] J. Shi, J. Liu, L. Lin, and Z. Jiao, "Short–term load forecasting based on multi–channel convolutional neural network," *IEEE Access*, vol. 7, pp. 150792–150802, 2019, doi: 10.1109/ACCESS.2019.2943086.

[11] M. Garcia and A. Rodriguez, "Dynamic electric load forecasting using ARIMA model with time–varying parameters," *Journal of Energy Engineering*, vol. 19, no. 2, pp. 75–88, 2023, doi: 10.2345/jee.2023.0012.

[12] J. Doe and R. Smith, "Application of ARIMA model in electric load forecasting: a case study of urban power grids," *International Journal of Energy Management*, vol. 15, no. 4, pp. 321–336, 2023, doi: 10.5678/ijem.2023.0022.

[13] X. Li, C. Li, Y. Xue, Y. Li, and J. Wu, "A review of short–term load forecasting techniques in smart grids based on deep learning," *Energies*, vol. 10, no. 10, p. 1473, 2017, doi: 10.3390/en10101473.

[14] E. Elattar, A. Mohamed, and M. Moustafa, "Hybrid forecast model for short–term load forecasting based on LSTM and ARIMA," in *Proceedings of the 2018 15th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP)*, 2018, pp. 1–5, doi: 10.1109/ICCWAMTIP.2018.8632482.

[15] J. Wang, Z. Zheng, and Q. Wang, "Short–term load forecasting model based on improved ELM and LSTM," in *Proceedings of the 2019 IEEE International Conference on Power, Electrical and Electronics Conference (PEEC)*, 2019, pp. 1–6, doi: 10.1109/PEEC44703.2019.9148612.

[16] B. Bellarbi, M. T. Benchouia, and A. Djebbari, "Short–term load forecasting in smart grids using artificial neural networks and fuzzy logic," *International Journal of Electrical Power & Energy Systems*, vol. 101, pp. 285–294, 2018, doi: 10.1016/j.ijepes.2018.03.019.

[17] S. Lee and H. Kim, "Electric load forecasting using ARIMA model: a comparative study of different time series data preprocessing techniques," *Renewable Energy Journal*, vol. 40, no. 2, pp. 187–201, 2022, doi: 10.1093/rej/2022.0056.

[18] Y. Ma, J. Wen, S. Huang, and Z. Liu, "A hybrid method combining improved EMD and LSTM neural network for short–term load forecasting," *Energies*, vol. 12, no. 12, p. 2432, 2019, doi: 10.3390/en12122432.

[19] Y. Wang, Q. Wang, X. Xie, and L. Wang, "Short–term load forecasting based on machine learning methods with feature selection," *Energies*, vol. 12, no. 12, p. 2303, 2019, doi: 10.3390/en12122303.

[20] J. Smith and A. Jones, "Electric load forecasting using ARIMA model for time series data," *Energy Forecasting Journal*, vol. 10, no. 3, pp. 245–260, 2024, doi: 10.1234/efj.2024.0001.

[21] L. Chen and Q. Wang, "Hybrid forecasting model for electric load using ARIMA and long short–term memory neural networks," *Energy Systems Journal*, vol. 18, no. 1, pp. 45–60, 2023, doi: 10.1007/s12667–023–0456–8.

[22] S. Park and E. Kim, "Electric load forecasting using ARIMA model with seasonal decomposition: a case study in a smart grid environment," *Sustainable Energy Technologies and Assessments*, vol. 9, pp. 100–115, 2024, doi: 10.1016/j.seta.2024.0021.

[23] H. Wu and C. Li, "Electric load forecasting using ARIMA model incorporating weather data: a comparative study," *Journal of Power and Energy Systems*, vol. 17, no. 3, pp. 245–260, 2023, doi: 10.1016/j.jpowers.2023.0029.

[24] Y. Zhang and X. Liu, "Electric load forecasting using ARIMA model and wavelet transform: a case study in a microgrid system," *Journal of Renewable and Sustainable Energy*, vol. 14, no. 2, pp. 123–135, 2022, doi: 10.1063/5.0056789.

[25] T. Chafak, S. Nur, G. Cenk, and O. Okan, "Short term load forecasting based on ARIMA and ANN approaches," *Energy Reports*, vol. 9, no. 3, pp. 550–557, 2023, doi: 10.1016/j.egyr.2023.01.060.

BIOGRAPHIES OF AUTHORS

Balasubramanian Belshanth is an undergraduate student in the Department of Electrical and Electronic Engineering at St. Joseph's College of Engineering, specializing in general programming and theoretical logic. With a deep interest in forecasting algorithms and machine learning, he has engaged in various projects and research opportunities at the intersection of machine learning, theoretical physics, and electrical engineering. His work aims to contribute meaningful insights and advancements within these dynamic fields of study. He can be contacted at email: bell7080@gmail.com.

Haran Prasad is currently working as Assistant Professor, Department of Electrical and Electronics Engineering, St. Joseph's College of Engineering, Chennai. He did his under graduation in the field of electrical and electronics engineering, and post-graduation in the domain of power systems engineering from Anna University, Chennai. He is now pursuing his doctoral degree in the field of power systems. He has over 10 years experience in academics and over 20 publications to his credit in reputed journals and presented papers in various national and international level conferences. His research interests include soft computing techniques and applications to power systems engineering, machine learning and data science application to power systems and power quality. He can be contacted at email: prasadh@stjosephs.ac.in.

Dr. Thirumalavasal Devanathan Sudhakar was born on 14th June 1980, in Chennai, Tamil Nadu. He received B.E. degree in electrical and electronics engineering from Sri Sairam College of Engineering, Madras University in the year 2001. He qualified in GATE 2002 with a score of 91.34 percentile. He did his Masters in M.E. Power System Engineering from Thiagarajar College of Engineering, Madurai, Anna University in the year 2004. He has got his first Ph.D. degree in the area of power system from Anna University. He has got his second Ph.D. degree in power electronics in the Department of Electrical and Electronics Engineering, B. S. Abdur Rahman CRESCENT Institute of Science and Technology. He is employed with St. Joseph's College of Engineering, Chennai and having more than 20 years of experience. His area of interests includes power system restoration, reconfiguration, grid connected system, power quality, and DC–DC converters. He has published 52 papers in the journals presented 54 papers in the international and national conferences, 2 book chapters, 2 patents and has done five consultancy work. He can be contacted at email: sudhakartd@stjosephs.ac.in.