Indonesian Journal of Electrical Engineering and Computer Science
Vol. 38, No. 3, June 2025, pp. 1540~1552
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v38.i3.pp1540-1552

A comparative study on electricity load forecasting using statistical and deep learning approaches

Tehreem Fatima Butt[1], Sana Tameer[1], Muhammad Saleem[1], Jawwad Sami Ur Rehman[2], Sathish Kumar Selvaperumal[3]
[1]Department of Sciences and Humanities, Sir Syed CASE IT (SSCASEIT), Islamabad, Pakistan
[2]Department of Biomedical Engineering, Riphah University, Islamabad, Pakistan
[3]College of Engineering, Asia Pacific University of Innovation and Technology, Kuala Lumpur, Malaysia

Article Info

Article history:

Received May 7, 2024
Revised Nov 30, 2024
Accepted Feb 27, 2025

Keywords:

Autoregressive integrated moving average
Electricity load forecasting
Long short-term memory
Transformer neural network

ABSTRACT

Load forecasting has become reproving aspect of an energy management system (EMS). It gives basic advantage to grid stability, cost effectiveness and battery storage system (BSS). For this purpose, machine learning (ML) is widely adopted to forecast the electricity load. This research paper investigates the performances of various time series estimating models applied to electricity load data for an Irish company. The research mainly adopts the autoregressive integrated moving average (ARIMA) model, long short-term memory (LSTM) networks and transformer neural network (TNN) to forecast the electricity load. A comparison evaluation is conducted encompassing various quantifying measures such as root mean square error (RMSE), mean square error (MSE) and mean absolute error (MAE). The results are then compared to get an understanding whether the TNN using attention-based mechanism is better than the two state of the art models. Hence provides a complete understanding about which of the model needs improvements in its architecture for enhancement of operational efficiency and cost effectiveness in the realm of EMS.

Corresponding Author:

Sathish Kumar Selvaperumal
College of Engineering, Asia Pacific University of Innovation and Technology
57000, Kuala Lumpur, Malaysia
Email: sathish@apu.edu.my

1. INTRODUCTION

Electricity load forecasting is crucial for ecological well-being, steady working of grid with efficient resource management. To foster renewable energy resources one need a robust designed forecasting system that eliminates severance. The load of electricity forecasting is requisite for maintaining grid stability, enhancing resources allocation, incorporating renewable energy, supporting market operations and enhancing resilience to disruption. Machine learning (ML) techniques can address the evolving challenges of the energy sector and build a more sustainable and reliable electricity infrastructure for the future [1].

By applying ML to the acquired data–the challenges in measuring the load can be entertained efficiently. By comparing the three machine-learning algorithms i.e. autoregressive integrated moving average (ARIMA), long short-term memory (LSTM) and transformer neural network (TNN) this study brings forward the best and most reliable solution to the administrated problems. By far TNN are well known for their outstanding results in language translation. In this research TNN is used on time series data and then compared to the most bench marking ML models like ARIMA and LSTM.

In the United States, a research was conducted in 2019 on primary energy consumption that the building zone used up to 32% of prime energy. This usage was increased by 11% which is more than that of the energy consumed in 2010 and it is also considered that global energy usage will be increased by 1.3% per year on regular from 2018 to 2050. Zhang *et al.* [2] used support vector machines (SVM), logistic regression (LR), and different artificial neural network (ANN) models for building load prediction. A case study was conducted by Batlle *et al.* [3], which categorized that four methods- energy audits methods, statistical regression methods, neural networks and Support vectors machines- can be used to analyze energy consumption.

Between 2010 and 2035, global electricity demand is projected to increase by 36%, growing at an average annual rate of 1.2% [4]. Similarly, in Brazil, there has been a sizable rise in electricity consumption, with a 34.1% increase recorded between 2005 and 2016, as reported in the Statistical Yearbook of Electricity 2017. This report indicates a consumption of 461.7 TWh in 2016 alone. Notably, the public sector accounts for around 6.9% of the nation's total consumption, with electricity constituting over 91% of its usage. According to EPE, this sector is expected to experience a yearly growth rate of 4% from 2017 to 2026 [3]. In this current era, where energy is evolving, accurate and precise forecasting is essential for addressing many of the issues resulting from supply and demand of energy. As energy is volatile in nature, a timely and precise methodology is a requisite to mitigate the risks and balance the supply and demand ratio.

As per the above mentioned discussion, this research delves into the importance of load forecasting and developed a methodology aimed to bring accuracy by exploring time series forecasting techniques tailored explicitly for electricity load prediction. By juxtaposing customary statistical methods like ARIMA against contemporary deep learning (DL) structural design such as LSTM networks and attention-based transformers, the study aims to scrutinize their applicability, robustness, and forecasting accuracy.

Wazirali *et al.* [5] provide an in-depth analysis of various forecasting models- ANN, ML, and DL- against traditional statistical models. The paper indicates the strengths and limitations of these models. And shows that forecasting results of models; like ANN, ML and DL; has alleviated against traditional models.

Modeling energy consumption and its forecasting, stands as a focal exertion within both developed as well as developing nations, offering instrumental intuitions to concerning organization and policymakers. Precisely evaluating the needs of energy consumption holds noteworthy inferences; underestimating the risks of dire outages, striking severe impressions on economy and life. Thus, it becomes imperious to model consumption of energy with exactness to forestall inflated errors and to certify optimum allocation of resources [6].

The latent of artificial intelligence (AI) procedures in augmenting load predicting efficiency, mainly with respect to the application of smart meters in the residential sector is of key importance. The survey of these aforementioned advanced AI prototypes wishes to shed light on their competences and influences to the recognition of forecasting systems that is intelligent, and is moving parallel with the sprouting landscape of smart technological models in management of residential energy.

Hayes [7] provides a detailed role of ARIMA model in time-series forecasting. The study highlighted the key components of ARIMA and explains how these components are collectively used to predict next instant using previous instances. Later in methodology section of this paper mathematical representation of ARIMA is also mentioned.

ARIMA denotes a statistical probe model influencing time series data for a twofold purpose: acquisition of insights of dataset and forecasting forthcoming trends [8]. In the demesne of statistical modeling, an autoregressive (AR) model makes prophesies about imminent/forthcoming values by depending on past instances [9]. The origins of ARIMA is tracked back to the drudgery of Box [10] who presented the ARIMA model in the 1970s. Since then, ARIMA has become a foundation stone in the turf of data exploration and foretelling, that hinge on time interludes.

Short term forecasting of electricity load was performed in 2023 using ARIMA and ANN. The study was established on real-time electricity load data from around 709 household utilities and comprises of over 18-months of data. The findings of the research revealed that ARIMA performed slightly better than ANN [11]. The conjecturing of electricity at a very short term time series forecasting (VST-TSF) on a residential customer level is of key importance in planning the management of energy and power. These predictions will help the authorities to make trade-off policies for better consumption. Diverse models, both statistical and recurrent, have been tailored to probe VST electricity load forecasting. But these models have low convergence rate, and are not able to detect underlying complex patterns in data and so forth. To overcome these problems LSTM was developed [12].

LSTM-NN was developed to address some of the shortcomings of obsolete recurrent neural networks (RNNs), a model unable to capture long-term data patterns in chronological sequence. LSTM addressed the challenges of gradient issues faced by RNNs where the gradients of loss function diminished exponentially over time during training. This issue hindered the aptitude of RNNs that effectively acquire and remember data and facts from distant time steps in sequential data [13]. The consumption of electricity from the renewable sources is increasing speedily due to the low maintenance and environmental friendly nature. Apropos, a study was conducted to predict the solar power by using different ML models. It was interpreted that among all chosen models RNN and SVM outperformed and bought highest accuracy in terms of percentage mean absolute percentage error (MAPE) [14].

In particular, the term very short term load forecasting can span from few minutes to one week. This division can be of key importance in accurately predicting the load of electricity on the grid. Studies have shown that if load forecasting errors are decreased by only 1% can save up to worth 10 million pounds for electricity utility operators [15]. The multiplicative gate units within LSTM dynamically regulate access to the constant error flow, demonstrating its adaptability. Importantly, LSTM exhibits local spatial and temporal characteristics, resulting in computational efficiency [16].

Another technique tailored specifically for the acquired data is attention-based TNN. The prevailing models for sequence transduction rely on intricate RNN or convolutional neural network (CNN), incorporating both with an encoder and with a decoder. Most of successful among them is integrating both with an encoder and a decoder using attention mechanism. In contrast, an original and forthright network design called the transformer, which relies exclusively on attention mechanisms, completely discarding the requirement for RNN and CNNs. Tentative outputs from two ML translation chores establish the dominance of these transformer models in terms of parallelizability, efficiency in training and quality [17].

Wang *et al.* [18] presented a transformer-based prototype for forecasting energy system. It was stated in the paper that transformer based model has outperformed the traditional methods. Forecasting energy loads (like electricity, heating, and cooling) in the system. In the same year, Zhang [19] also worked on time augmented transformer architecture. During the research it is stated that their model efficiently address complex time series problems. Moreover, it was already discussed by Hochreiter *et al.* [20] that a cipher-decipher arrangement is included in the majority of viable neural sequence models of transduction.

An AR model, like ARIMA, is the one in which a value at specific instant depends on the erstwhile one. But an attention based model is not an AR classical model hence it can even overawed the curbs of RNN such as, slower computation speed and memory constrictions, and can be of great prominence in triumphing better results in a more speedy way [21]. Through an exhaustive proportional analysis, this research work objects to subsidize to the burgeoning field of energy projecting by enlightening the capabilities of different models and its suitability for real-world applications.

2. PROPOSED METHOD

The proposed methodology encompasses several phases to develop and evaluate predictive models for VST electricity load forecasting.

Stage 1: in this stage, the dataset goes through a nitpicking process of examination and analysis. During this process it was made sure that the acquired data is certainly a time-series data, as time-series data is requisite in domains like electricity load forecasting. Predicting electricity load in time series data is crucial because the sequence of data-points and the time intervals between them are of vital importance for understanding patterns and trends (if any) within the data-points. Particularly in case of forecasting electricity load, each model is required to do two basic tasks; analyzing magnitude of electricity load at different points in time and to identify sequential and any seasonal trends as well as anomalies or sudden changes in electricity usage. The temporal dependencies; like relationship between past, present and future data-points; allows models to predict electricity load by making sure to produce precise predictions.

After the affirmation that the data is time-series, the next stage of preprocessing of data is called for.

Stage 2: after the detailed analysis and nitpicking examination of data for, the next stage is pre-processing of data. This stage is very crucial as it makes sure to prepare data in accordance with the models, which are under examination for comparison. This stage of preprocessing of data is essential, as each model has different requisites and structures. Without proper preprocessing, models may struggle to give optimal performance. Another very tricky thing to do in pre-processing stage is normalization of data. This involves mounting numerical features in a range between 0 and 1. Normalization is also addressed as the dataset has mean of 0 and standard deviation of 1. Normalization is done to make sure that each instance in dataset is having equal importance.

Stage 3: Intermittently, in the third phase, the data is splitted into two sets training and testing sets. This segmentation into training and testing subsets is done to establish model’s ability not only to train efficiently but also to enhance their ability to speculate unforeseen data. Training set, comprises of 80% of total data-points, is used to enlighten the models- like ARIMA, LSTM, attention-based TNN- allowing them to grasp relationships, dependencies and patterns within the data. Once the training process is over the remaining 20% of the data-points are passed from the trained model. The performance of data is evaluated by using several using metrics like mean absolute error (MAE), mean squared error (MSE), and root mean squared error (RMSE).

By going through the above mentioned phases of the proposed methodology, strong prediction models can be obtained for forecasting load of electricity on smart meters. Moreover, this methodology allows a comparative study of the three models- ARIMA, LSTM, and attention-based transformer neural network. The data of smart-meters is acquired from Irish Social Science Data Archive (ISSDA). The evaluation of this household smart meters data, allows researchers to obtain insightful information into the strength, weakness and suitability of models in different scenarios. This proportional analysis is invaluable for understanding the procedures on how to take decisions effectively within the provided resources. This will also help understand the stakeholders how to allocate resources in best optimum way by recognizing the most accurate and efficient model for electricity load. Stakeholders can also enhance the overall efficacy of their forecasting strategies and prioritize model deployment. Consequently, this detailed and ample valuation nurtures knowledgeable decision-making and planning of resources within the energy sector, this will eventually contribute to the progression, improvement and advancement of ecological and dependable management of energy practices. The flowchart of the proposed methodology is shown in Figure 1.

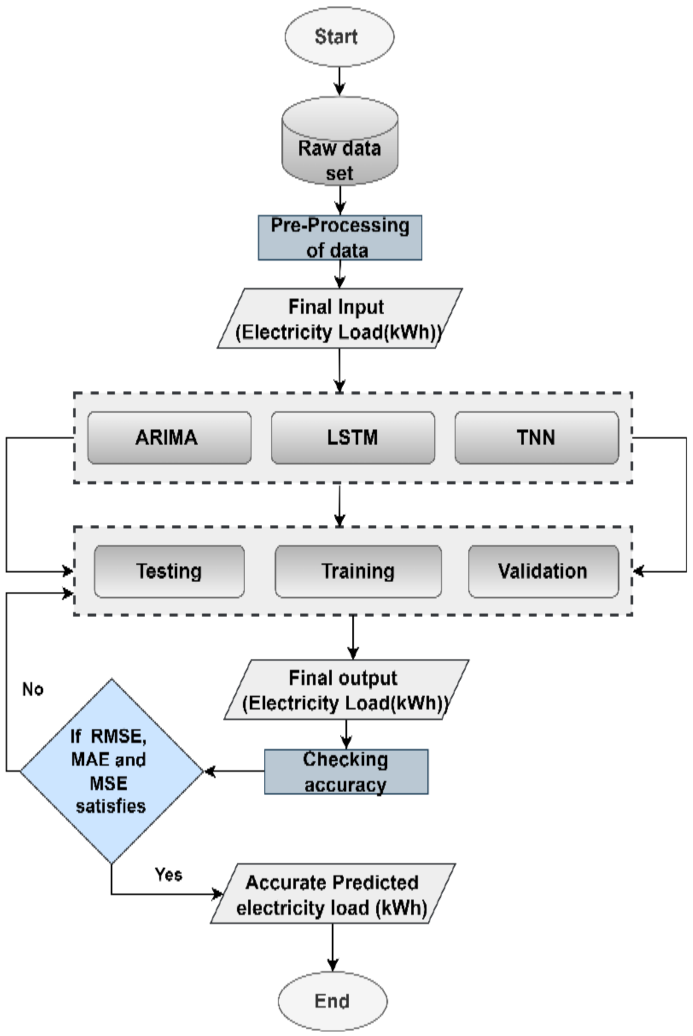

Figure 1. Proposed method

The data obtained from ISSDA comprises of three categories of smart-meters; that is residential, SME and others. This paper compares the three aforementioned models only on residential smart meters. The data provided by each meter consists of two columns. First column shows date and time and second column shows electricity consumed in kWh during 30 minutes interval. Table 1 shows a first 20 of one of the residential meter 1002. The table portrays electricity load data in kilowatt-hours. It displays consumption of electricity throughout the day at specific time. Following are the detailed explanation of mathematical representation of each model and working of each model on the residential smart meters like meter 1002.

Table 1. Electricity load data meter 1002

Date	Load (kWh)
14/07/2009 00:30:00	0.362
14/07/2009 01:00:00	0.064
14/07/2009 01:30:00	0.119
14/07/2009 02:00:00	0.023
14/07/2009 02:30:00	0.14
14/07/2009 03:00:00	0.036
14/07/2009 03:30:00	0.108
14/07/2009 04:00:00	0.083
14/07/2009 04:30:00	0.056
14/07/2009 05:00:00	0.129
14/07/2009 05:30:00	0.015
14/07/2009 06:00:00	0.132
14/07/2009 06:30:00	0.054
14/07/2009 07:00:00	0.082
14/07/2009 07:30:00	0.103
14/07/2009 08:00:00	0.028
14/07/2009 08:30:00	0.136
14/07/2009 09:00:00	0.051
14/07/2009 09:30:00	0.333
14/07/2009 10:00:00	0.384

2.1. Autoregressive integrated moving average

The ARIMA methodology is a prevalent time series forecasting approach that syndicates auto regression, differencing, and moving averages (MA). The AR module comprises of prophesying future values grounded on past values. This models an association amongst an instance and several lagged values [22]. ARIMA consists of three key components that is AR, integrated (I), and MA. The results of ARIMA is obtained by combining the effects of all the three components.

The AR module comprises of prophesying future values grounded on past values. This models an association amongst an instance and several lagged values. Mathematically:

$$y_t = l + a_1 y_{t-1} + a_2 y_{t-2} + \cdots + a_n y_{t-n} + e \tag{1}$$

Here, y_t is the current observation, l is a constant, $a_1, a_2, \ldots, a_n$ are the AR coefficients, $y_{t-1}, y_{t-2}, \ldots, y_{t-n}$ are past observations, and e is the miscalculated term.

The crucial function of the I model is to accomplish stationarity in the data. Stationarity is a key postulation for several time series models, as well as ARIMA. Differencing involves computing difference between consecutive observations.

The differenced series, denoted as y_t is obtained by subtracting the (t−1)-th observation from the t-th observation.

$$y'_t = y_t - y_{t-1} \tag{2}$$

The MA model focuses on modeling the relationship between an observed time series and a linear combination of past error terms (residuals). It is denoted as MA(q), where "q" represents the order of the model.

$$y_t = A_t + \varepsilon_t + \varepsilon_{t-1}\beta_1 + \varepsilon_{t-2}\beta_{2+} \ldots + \varepsilon_{t-n}\beta_n \tag{3}$$

Where, y_t is observed value at instance t, A_t is the mean, ε_t is the white noise term at t, and $\beta_1, \beta_2, \ldots, \beta_n$ are the parameters to be estimated.

The first and the foremost condition for application of ARIMA that data should be stationary. The collected data was tested for the stationarity using augmented dickey-fuller (ADF) test. To determine if the differencing was required to make the data stationary. With all this discussion preprocessing phase began. Continuing this process the data set was resampled to 30 minutes intervals to ensure uniformity across the time series data. These preprocessing steps are crucial for understanding the patterns in data, as it very helpful in fine tuning the model's parameter. The selection of parameter is done with the help of ACF and PACF plots. After training the model the tested set is passed through the trained model and different metrics like- MAE, MSE, and RMSE- was used to evaluate models performance. Later in comparison section the results are shown and compared.

2.2. Long short-term memory

The methodology of LSTM networks involves a deep understanding of the architecture and training process of RNNs. LSTM networks are composed of memory cells, which are responsible for maintaining and updating a cell state. The memory cell is the fundamental building block of LSTM [23], [24].

LSTMs have three natures of gates - input; forget, and output. These gates regulate the movement of data in and out of the memory cell, letting network to reminisce or overlook the selected information [25]. Similar to other RNN, LSTM is coached to use the back-propagation algorithm. However, in the case of LSTM, the BPTT is extended through time to capture long-term dependencies [26]. The activation function used in LSTM is sigmoid function.

The sigmoid input in the interval [0, 1]. It is frequently used for binary grouping of problems as it produces an output that can be interpreted as a probability [27]. The mathematical illustration is shown in (2).

$$\sigma(x) = \frac{1}{1+e^{x}} \tag{4}$$

Whereas, the tanh function presses the input data sets in the range (−1, 1). It is widely used in neural networks because it introduces non-linearity and is zero-centered, helping in the optimization process [28]. The formula of tanh is expressed in (3).

$$\tanh(x) = \frac{e^{2x}-1}{e^{2x}+1} \tag{5}$$

These activation functions are crucial to neural networks' operation since they allow them to comprehend the complex and complicated correlations and combinations found within the data. The task's exact requirements and the data's properties determine which activation function should be used. LSTMs are designed to work with sequences of data. The totality of the input sequences and the number of time steps considered during training are important considerations [29].

The preprocessing of data for LSTM is initially same as it was done for ARIMA model. The data is loaded and is resamples to 30 minutes interval to maintain uniformity of data. Additionally, for LSTM input data and its corresponding output data are converted into sequences, which sliced the time-series data into overlapping sequences. These overlapping sequences are the key to LSTM for capturing patterns. And it is because of these formation of overlapping sequences, it is named is LSTM. After preparing these sequences, the data which comprises of sequences now is splitted into training and testing sequences with 80% of data is reserved for training and the remaining 20% is for testing to evaluate the performance of model. Specifically, for the data under study a single layer LSTM model with 50 units, which helped capture any dependencies present in the given temporal data. A one unit dense layer is added and model is compiled by using Adam optimizer and the MSE as the loss function. The model was then trained and tested using the same three metrics used in ARIMA.

The specific details associated with the number of heads, activation functions, and loss functions may diverge, reliant on the architecture and obscurity of model. In this case, a single-layer LSTM with a linear function as an activation function is used, and MSE is engaged as the loss function. Adjustments of these hyper parameters might be made based on the problem and data characteristics.

2.3. Transformer neural network

The attention-based TNN (transformer) has gained prominence for its effectiveness in seizing long range enslavements and patterns in time data. Self-attention: this mechanism is core of the TNN, which calculates and weigh out the consequence of diverse time steps in model when prophesies are made. It computes attention notches for each time step based on its relationship with all other time steps.

It mainly employs multiple attention heads to capture different aspects and patterns in the time series. It includes feed-forward neural networks within each attention head to capture temporal hidden and non-linear affiliations within data. Mathematically, a transformer model is represented as,

$$Attention\ (Q, K, V) = Softmax\left(\frac{Q.K^{T}}{\sqrt{d_k}}\right)V \tag{6}$$

whereas Q is the query matrix that is the information to be calculated, K represents key matrix, a criteria for predictions, and V is the value of actual data. Q, K, V are also known as different projections of same input data. $\sqrt{d_k}$ acts as scaling factor, it is preventing the dot product from becoming too large.

In a nutshell this equation amplifies the model ability to determine how much weight each element in the sequence should be given allowing it to concentrate on the most pertinent data. This methodology adapts the TNN architecture for time series forecasting, that emphasize on the self-attention mechanism's aptitude to seize temporal dependencies [30]. It combines essential components such as positional encoding, multi-head attention, and feed-forward networks to enhance the model's aptitude to learn intricate patterns in time dependent data series. Adjustments to hyper parameters and model architecture may be necessary grounded on the precise values for the dataset [31].

The sliding window method is used to preprocess the time series data set for TNN. This results in input arrays and matching target sequences each of which has a length of 3 to capture short term relationships. Then after all these three processing the data is played into two sections 20% of the data is set aside for evaluating the model performance and the remaining 80% is used for training the model. The Adam optimizer, known for its adeptness in managing extensive data sets, and the means squared error as loss function are utilized in the definition and compilation of transformer models. Means squared error is the optimal choice for continuous data prediction task such as electricity load forecasting.

The model is trained on the 80% of training data over 200 epochs with the batch size of 32. A technique known as early stopping is employed to prevent the model from over fitting. This ensures that the model stops at optimal weights for subsequent predictions by tracking validation performance and stops the training if there is no improvement for each 10 consecutive epochs. This procedure is intended to improve load predictions accuracy. The accuracy and error generated by the transformer model is then calculated by the metrics are discussed in the results and discussion section.

3. RESULTS AND DISCUSSION

This research is meticulously designed to conduct a comparative diagnostic evaluation of the three aforementioned models. While ARIMA and LSTM models have been widely and extensively applied in various time series forecasting studies, the application of attention based transformer networks in forecasting electricity is relatively novel. This is specifically notable as attention mechanism is well known for its ability to capture long-range dependencies and contextual relationship with in sequences. However, their use in forecasting tasks, peculiarly electricity load forecasting remains an unexplored area.

The most important objective of this study is to setup an environment for widespread evaluation of the performance of ARIMA, LSTM, and Attention-Based transformers in apprehending the temporal reliance and sophisticated configurations embedded within the series. By commissioning these cutting-edge models, we wish to assess their effectiveness in forecasting accurateness, adaptableness to composite and compound sequences, and flexibility in handling varied temporal diminuendos.

By combining the results of these models a comparative analysis is conducted which aims to broaden the scope of forecasting approaches beyond traditional methods and RNN resulting in a more comprehensive and flexible approach to time series forecasting and power load prediction. And check whether a TNN, which is originally established for natural language processing, can outperform the well-known ARIMA and LSTM or not.

This section unfolds with an overview of the research results and interpretations carried by each of the architecture, emphasizing the distinct characteristics and parameters involved. Subsequently, the results obtained from the experimentation with each model are presented, followed by a meticulous analysis of the observed patterns, strengths, and limitations of each approach. The graph of actual data with respect to time is shown in Figure 2. The graph is showing electricity load measurement at half hour intervals over the period of time.

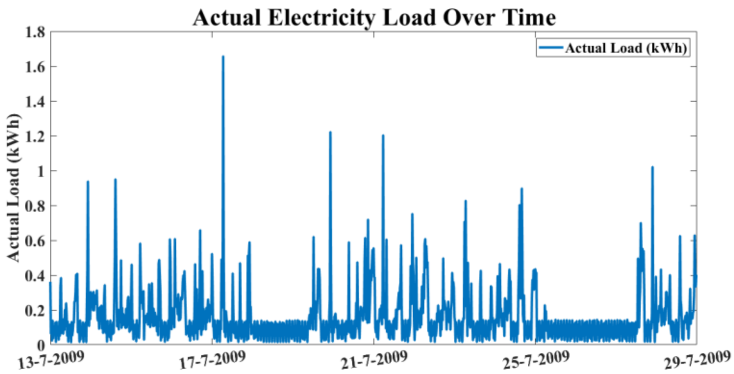

Figure 2. Actual load (kWh) against half hour intervals

Table 2 is showing the comparison of all the techniques under study along with the actual data. It displays the first 50 values of actual data and its comparison with the predicted value. The graph of the actual load along with predicted load of all the aforementioned models is shown in Figure 3. As it can be visualized clearly that the best performing model in all the three of them is LSTM as expected LSTM as outperformed both ARIMA and attention based TNN. While the attention-based model shows promise, the temporal losses observed addressing the intricacy of the electricity loads forecasting chore and the need for careful model design. LSTM, with its ability to capture sequential dependencies, proves to be a robust performer in this context.

Table 2. Targeted vs. predicted values

Time stamp	Target	ARIMA	LSTM	TNN
8/2/2009 5:00	0.02	0.117437	0.10112982	0.1603786
8/2/2009 5:30	0.119	0.123631	0.11837502	0.1603786
8/2/2009 6:00	0.088	0.128043	0.08587683	0.1603786
8/2/2009 6:30	0.046	0.115114	0.11354533	0.1603786
8/2/2009 7:00	0.141	0.13659	0.10868737	0.1603786
8/2/2009 7:30	0.021	0.113346	0.09833882	0.1603786
8/2/2009 8:00	0.113	0.126277	0.11683015	0.1603786
8/2/2009 8:30	0.095	0.124652	0.08403575	0.1603786
8/2/2009 9:00	0.035	0.117337	0.11160741	0.1603786
8/2/2009 9:30	0.142	0.133411	0.10835022	0.1603786
8/2/2009 10:00	0.048	0.113069	0.09145024	0.1603786
8/2/2009 10:30	0.083	0.133713	0.12272968	0.1603786
8/2/2009 11:00	0.118	0.12087	0.09680368	0.1603786
8/2/2009 11:30	0.018	0.119873	0.10649377	0.1603786
8/2/2009 12:00	0.128	0.127397	0.1109148	0.1603786
8/2/2009 12:30	0.062	0.11404	0.07920262	0.1603786
8/2/2009 13:00	0.067	0.125852	0.11130858	0.1603786
8/2/2009 13:30	0.137	0.126907	0.09876965	0.1603786
8/2/2009 14:00	0.019	0.119677	0.10495377	0.1603786
8/2/2009 14:30	0.117	0.12904	0.1195061	0.1603786
8/2/2009 15:00	0.084	0.116495	0.08244327	0.1603786
8/2/2009 15:30	0.046	0.123674	0.11041664	0.1603786
8/2/2009 16:00	0.14	0.129078	0.1053035	0.1603786
8/2/2009 16:30	0.027	0.116168	0.09647609	0.1603786
8/2/2009 17:00	0.103	0.128269	0.11815775	0.1603786
8/2/2009 17:30	0.1	0.120225	0.0855408	0.1603786
8/2/2009 18:00	0.027	0.120262	0.10801794	0.1603786
8/2/2009 18:30	0.141	0.128219	0.10665749	0.1603786
8/2/2009 19:00	0.052	0.115119	0.08561418	0.1603786
8/2/2009 19:30	0.08	0.131547	0.12057918	0.1603786
8/2/2009 20:00	0.122	0.123285	0.09872255	0.1603786
8/2/2009 20:30	0.019	0.119311	0.10694417	0.1603786
8/2/2009 21:00	0.131	0.128196	0.11316586	0.1603786
8/2/2009 21:30	0.07	0.116532	0.08257367	0.1603786
8/2/2009 22:00	0.062	0.127151	0.11630706	0.1603786
8/2/2009 22:30	0.137	0.127946	0.10392039	0.1603786
8/2/2009 23:00	0.018	0.118084	0.10393091	0.1603786
8/2/2009 23:30	0.113	0.128124	0.117836	0.1603786
8/3/2009 0:00	0.092	0.116552	0.0800785	0.1603786
8/3/2009 0:30	0.042	0.124182	0.10940304	0.1603786
8/3/2009 1:00	0.142	0.129943	0.10871811	0.1603786
8/3/2009 1:30	0.035	0.116457	0.09663182	0.1603786
8/3/2009 2:00	0.094	0.129709	0.12082349	0.1603786
8/3/2009 2:30	0.113	0.120496	0.08985633	0.1603786
8/3/2009 3:00	0.019	0.121745	0.1083415	0.1603786
8/3/2009 3:30	0.139	0.127202	0.11061327	0.1603786
8/3/2009 4:00	0.068	0.115609	0.08220327	0.1603786
8/3/2009 4:30	0.06	0.132634	0.1212355	0.1603786
8/3/2009 5:00	0.142	0.123022	0.10352463	0.1603786

Our findings indicate that increasing model complexity is not related with lower performance in the time series forecasting. The proposed methodology may benefit from the strength of LSTM in capturing non-linear correlations while maintaining ARIMA's accuracy in handling stationery data. However, attention based TNN regularly over estimates target values indicating that the attention mechanism requires additional refining to avoid over fitting and improve predictive performance.

The accuracy and error of each model is calculated and compared in Table 3. This table is facilitating a comparative analysis of the performance of each model. This evaluation and comparison of error percentage and accuracy percentage provides a clear understanding of which models demonstrate higher accuracy.

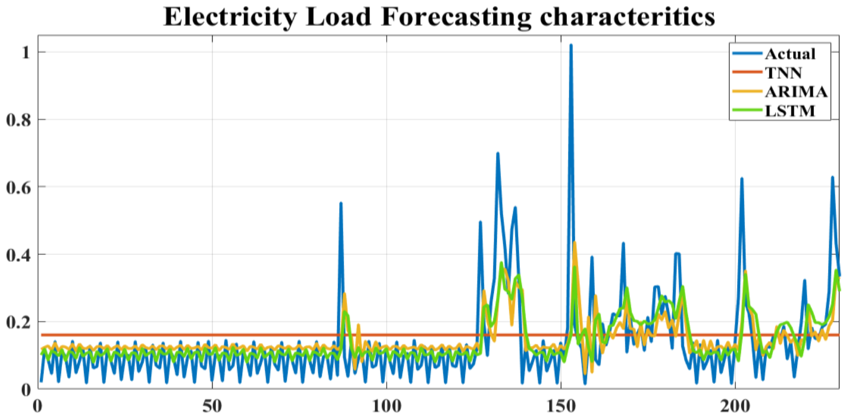

Figure 3. Comparison of all techniques

Table 3. Metrics comparison

	Method	Accuracy	Error
ARIMA	MAE	91.69825234052496	0.08301747659475035
	MSE	98.1431971264334	0.018568028735665914
	RMSE	86.37354457840708	0.13626455421592923
LSTM	MAE	93.23186143188373	0.0676813856811627
	MSE	98.73545225536353	0.012645477446364769
	RMSE	88.75478882085145	0.11245211179148558
Attention-based TNN	MAE	90.90309335762176	0.09096906642378239
	MSE	82.37958122287310	0.17620418777126976
	RMSE	86.72580745313411	0.13274192546865884

Also, for clearer understanding bar plots of errors and accuracy shown in Figures 4 and 5 unveils that LSTM [32], [33] demonstrate higher accuracy as paralleled to both ARIMA and attention models. The conjectures generated and estimated by LSTM aligned more narrowly with the genuine electricity load data, signifying its aptitude and capability to apprehend core dependencies and patterns meritoriously [34]. Such higher accuracy advocates that LSTM outpaces both ARIMA and TNN in forecasting of load on electricity smart meters, making it more trustworthy choice for extrapolative modeling.

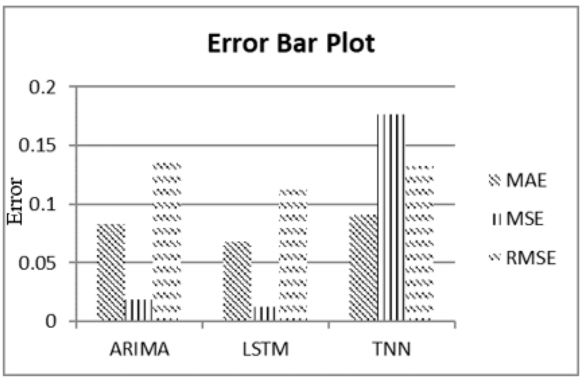

Figure 4. Error comparison

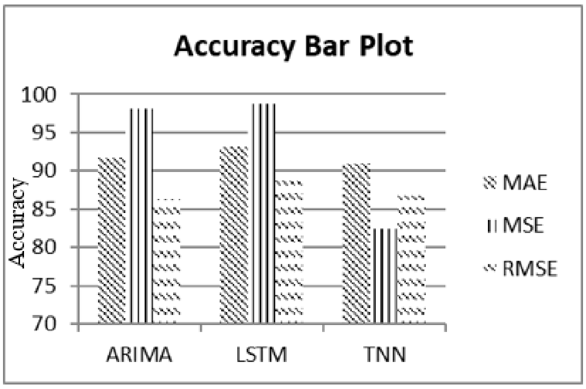

Figure 5. Accuracy comparison

Our study suggests that the attention mechanism require further tweaking or modification in its architecture to boost its overall performance on time series data. Overall, although the non-complexity of ARIMA is beneficial, this study point out the distinct benefits and advantages of more refined and erudite models like LSTM [35]-[37] and highlights the current need for exploration and refinement in the architectures of neural network tasked especially for forecasting time series. Future research in forecasting of load could explore innovative TNN mechanisms and alterations to architectures of TNN, trying to improve the model for chronological dependencies and augment and boost the accuracy of forecasting. Continual of research and efforts in these directions have the latent to harvest noteworthy developments and improvements in the realm of electricity load forecasting, facilitating more and more precise and dependable forecasts to maintain well-organized energy management.

4. CONCLUSION

The purpose of this work was to apply attention mechanism to electricity load data, evaluate and compare its performance against the two states of the art models, ARIMA, and LSTM. The attention mechanisms performance in natural language processing and its capacity to handle vast sequences of chronological data led to the initial hypothesis that it would outperform the cutting-edge models. However, the result shows that, for the dataset utilized, the attention based TNN model did not outperform ARIMA and LSTM. Moreover, despite the introduction of an early stopping mechanism, it is observed that the encoder decoder structure of attention based TNN has demonstrated over fitting. Our future research will concentrate on discovering the novel attention processes and making architectural changes to the TNN model to improve its performance in time series forecasting.

ACKNOWLEDGEMENTS

Since the ISSDA provided the electricity load data utilized in this study, I would like to sincerely thank them. The research and advancement of this study have been made possible by their invaluable donation of high-quality data. I am really grateful and appreciate what they have done to enhance the subject of electricity load forecast by making this data available for academic and research purposes.

FUNDING INFORMATION

Authors state no funding involved.

AUTHOR CONTRIBUTIONS STATEMENT

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Tehreem Fatima Butt	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓		✓	
Sana Tameer		✓		✓	✓		✓			✓	✓			
Muhammad Saleem	✓			✓	✓			✓	✓			✓		
Jawwad Sami Ur Rahman				✓		✓	✓			✓			✓	✓
Sathish Kumar Selvaperumal				✓		✓	✓			✓	✓		✓	✓

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

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

DATA AVAILABILITY

The data that support the findings of the study are openly available in the Irish Social Science Data archive (ISSDA).

REFERENCES

[1] I. K. Nti, M. Teimeh, O. Nyarko-Boateng, and A. F. Adekoya, "Electricity load forecasting: a systematic review," *Journal of Electrical Systems and Inf Technol*, vol. 7, no. 1, p. 13, Dec. 2020, doi: 10.1186/s43067-020-00021-8.

[2] L. Zhang *et al.*, "A review of machine learning in building load prediction," *Applied Energy*, vol. 285, p. 116452, Mar. 2021, doi: 10.1016/j.apenergy.2021.116452.

[3] E. A. O. Batlle, J. C. E. Palacio, E. E. S. Lora, A. M. M. Reyes, M. M. Moreno, and M. B. Morejón, "A methodology to estimate baseline energy use and quantify savings in electrical energy consumption in higher education institution buildings: Case study, Federal University of Itajubá (UNIFEI)," *Journal of Cleaner Production*, vol. 244, p. 118551, Jan. 2020, doi: 10.1016/j.jclepro.2019.118551.

[4] N. Jollands *et al.*, "The 25 IEA energy efficiency policy recommendations to the G8 gleneagles plan of action," *Energy Policy*, vol. 38, no. 11, pp. 6409–6418, Nov. 2010, doi: 10.1016/j.enpol.2009.11.090.

[5] R. Wazirali, E. Yaghoubi, M. S. S. Abujazar, R. Ahmad, and A. H. Vakili, "State-of-the-art review on energy and load forecasting in microgrids using artificial neural networks, machine learning, and deep learning techniques," *Electric Power Systems Research*, vol. 225, p. 109792, Dec. 2023, doi: 10.1016/j.epsr.2023.109792.

[6] Q. Hassan *et al.*, "The renewable energy role in the global energy transition," *Renewable Energy Focus*, p. 100545, Jan. 2024, doi: 10.1016/j.ref.2024.100545.

[7] A. Hayes, "Autoregressive integrated moving average (ARIMA) prediction model. investopedia," *Investopedia*, 2022.

[8] V. Kotu and B. Deshpande, *Data science: concepts and practice*. Morgan Kaufmann, 2018.

[9] B. L. Smith, B. M. Williams, and R. K. Oswald, "Comparison of parametric and nonparametric models for traffic flow forecasting," *Transportation Research Part C: Emerging Technologies*, vol. 10, no. 4, pp. 303–321, 2002.

[10] G. Box, "Box and jenkins: time series analysis, forecasting and control," in *A Very British Affair*, London: Palgrave Macmillan UK, 2013, pp. 161–215. doi: 10.1057/9781137291264_6.

[11] C. Tarmanini, N. Sarma, C. Gezegin, and O. Ozgonenel, "Short term load forecasting based on ARIMA and ANN approaches," *Energy Reports*, vol. 9, pp. 550–557, 2023.

[12] S. Hochreiter, "Long short-term memory," *Neural Computation MIT-Press*, 1997, Accessed: Oct. 13, 2024. [Online]. Available: https://sophieeunajang.wordpress.com/wp-content/uploads/2020/10/lstm.pdf.

[13] W. Budiharto, V. Andreas, and A. A. S. Gunawan, "Deep learning-based question answering system for intelligent humanoid robot," *Journal Big Data*, vol. 7, no. 1, p. 77, Dec. 2020, doi: 10.1186/s40537-020-00341-6.

[14] S. M. Babbar and L. C. Yong, "Solar power prediction using machine learning algorithms: a comparative study," in *2022 3rd International Conference on Smart Electronics and Communication (ICOSEC)*, IEEE, 2022, pp. 1313–1319. Accessed: Oct. 13, 2024. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9951972/

[15] S. Huang, J. Shen, Q. Lv, Q. Zhou, and B. Yong, "A novel node approach combined with LSTM for short-term electricity load forecasting," *Future Internet*, vol. 15, no. 1, p. 22, 2022.

[16] K. Cho, "Learning phrase representations using RNN encoder-decoder for statistical machine translation," *arXiv preprint arXiv:1406.1078*, 2014.

[17] V. Ashish, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, p. I, 2017.

[18] C. Wang, Y. Wang, Z. Ding, T. Zheng, J. Hu, and K. Zhang, "A transformer-based method of multienergy load forecasting in integrated energy system," *IEEE Transactions on Smart Grid*, vol. 13, no. 4, pp. 2703–2714, 2022, doi: 10.1109/TSG.2022.3166600.

[19] G. Zhang, C. Wei, C. Jing, and Y. Wang, "Short-term electrical load forecasting based on time augmented transformer," *International Journal of Computational Intelligence Systems*, vol. 15, no. 1, p. 67, Aug. 2022, doi: 10.1007/s44196-022-00128-y.

[20] S. Hochreiter, Y. Bengio, P. Frasconi, and J. Schmidhuber, "Gradient Flow in Recurrent Nets: The Difficulty of Learning LongTerm Dependencies," *A Field Guide to Dynamical Recurrent Neural Networks. IEEE press*, 2001, doi: 10.1109/9780470544037.ch14.

[21] N. Mohamed, M. H. Ahmad, and Z. Ismail, "Double seasonal ARIMA model for forecasting load demand," *Matematika*, pp. 217–231, 2010, doi: 10.11113/matematika.v26.n.565.

[22] W. Yu, X. Li, and J. Gonzalez, "Fast training of deep LSTM networks," in *Advances in Neural Networks – ISNN 2019*, vol. 11554, H. Lu, H. Tang, and Z. Wang, Eds., in Lecture Notes in Computer Science, vol. 11554, Cham: Springer International Publishing, 2019, pp. 3–10. doi: 10.1007/978-3-030-22796-8_1.

[23] M. Liu, L. Chen, X. Du, L. Jin, and M. Shang, "Activated gradients for deep neural networks," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 34, no. 4, pp. 2156–2168, 2021.

[24] F. Xiao, Y. Honma, and T. Kono, "A simple algebraic interface capturing scheme using hyperbolic tangent function," *Numerical Methods in Fluids*, vol. 48, no. 9, pp. 1023–1040, Jul. 2005, doi: 10.1002/fld.975.
[25] B. Bakker, "Reinforcement learning by backpropagation through an LSTM model/critic," in *2007 IEEE International Symposium on Approximate Dynamic Programming and Reinforcement Learning*, IEEE, 2007, pp. 127–134. Accessed: Oct. 13, 2024. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/4220824/
[26] J. Brownlee, *Long short-term memory networks with python: develop sequence prediction models with deep learning*. Machine Learning Mastery, 2017.
[27] N. Passalis, A. Tefas, J. Kanniainen, M. Gabbouj, and A. Iosifidis, "Deep adaptive input normalization for time series forecasting," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 31, no. 9, pp. 3760–3765, 2019, doi: 10.1109/TNNLS.2019.2944933.
[28] G. Zerveas, S. Jayaraman, D. Patel, A. Bhamidipaty, and C. Eickhoff, "A transformer-based framework for multivariate time series representation learning," in *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, Virtual Event Singapore: ACM, Aug. 2021, pp. 2114–2124, doi: 10.1145/3447548.3467401.
[29] S. Ahmed, I. E. Nielsen, A. Tripathi, S. Siddiqui, R. P. Ramachandran, and G. Rasool, "Transformers in time-series analysis: a tutorial," *Circuits Syst Signal Process*, vol. 42, no. 12, pp. 7433–7466, Dec. 2023, doi: 10.1007/s00034-023-02454-8.
[30] A. Zeng, M. Chen, L. Zhang, and Q. Xu, "Are transformers effective for time series forecasting?," in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2023, pp. 11121–11128, Accessed: Oct. 13, 2024. [Online]. Available: https://ojs.aaai.org/index.php/AAAI/article/view/26317
[31] T. Bashir, C. Haoyong, M. F. Tahir, and Z. Liqiang, "Short term electricity load forecasting using hybrid prophet-LSTM model optimized by BPNN," *Energy Reports*, vol. 8, pp. 1678–1686, 2022.
[32] A. Montañés and A. Sanso, "The Dickey-fuller test family and changes in the seasonal pattern," *Annales d'Economie et de Statistique*, pp. 73–90, 2001.
[33] M. Alagirisamy and C.O. Chow, "An energy based cluster head selection unequal clustering algorithm with dual sink (ECH-DUAL) for continuous monitoring applications in wireless sensor networks," *Cluster Computing*, vol. 21, no. 1, pp. 91-103, 2018.
[34] S. Bhadula, M. Almusawi, A. Badhoutiya, A. Deepak, N. Bhardwaj, and G. Anitha, "Time series analysis for power grid anomaly detection using LSTM networks," *Proceedings of International Conference on Communication, Computer Sciences and Engineering*, IC3SE 2024, 2024.
[35] V. R. Kaneti, M. Venkatesan, H. Sharma, A. Bodhankar, R. Bajaj, R. G. Kumar, H. S. Pokhariya and A. Deepak, "Detecting fake information dissemination using leveraging machine learning and DRIMUX with B-LSTM," *International Journal of Intelligent Systems and Applications in Engineering*, Vol. 12, No. 5s, pp. 91-100, 2024.
[36] A. Y. A. B. Ahmad, P. William, D. Uike, A. Murgai, K. K. Bajaj, A. Deepak, A. Shrivastava, "Framework for sustainable energy management using smart grid panels integrated with machine learning and IoT based approach, "*International Journal of Intelligent Systems and Applications in Engineering*, vol. 12, no. 2s, pp. 581-590, 2024.
[37] R. Priyadarshini, M. Alagirisamy, N. Rajendran, A. K.Marandi, V. V. Patil and Vivek, "Minimization of makespan and energy consumption in task scheduling in heterogeneous cloud environment," *International Journal of Intelligent Systems and Applications in Engineering*, vol. 10, no. 2s, 2022.

BIOGRAPHIES OF AUTHORS

Tehreem Fatima Butt boarded on her journey of mathematical brilliance with a voracious interest in numbers and patterns. In 2013, she graduated with a Bachelor's degree in Mathematics from Islamic University with a CGPA of 3.63. Undeterred by challenges, she pursued further academic excellence, graduating from Sir Syed Case University in 2024 with a Master's degree in mathematics, earning a gold medal and a CGPA of 3.88. Her passion extended to teaching, spending three years as a lecturer at Ghazali College, Phalia, M.B. Din, Pakistan. Her commitment not only inspired students but also contributed significantly to nurturing future mathematical talents. Throughout her journey, she remained steadfast in her pursuit of excellence, leaving an enduring legacy in mathematics. She can be contacted at email: ftehreem@ymail.com.

Sana Tameer received the BS degree in electronics in 2015 from COMSATS University, Islamabad, Pakistan in, the MS degree in electrical engineering in 2019 from the Abasyn University of Islamabad., Pakistan and the Ph.D. degree in engineering in 2023 from Asia Pacific University of Innovation and Technology, Kuala Lumpur, Malaysia. She used to hold several administrative posts with the Department of Sir Syed CASE IT, Islamabad, from 2021 to 2022. She is currently an adjunct faculty with the Department of Humanities and Sciences at Sir Syed CASE IT. Her research interests include soft machine learning, artificial neural network, and deep learning. She can be contacted at email: sana.mohsin11@gmail.com.

Muhammad Saleem completed his Ph.D. from COMSATs University Islamabad, Pakistan in 2013. He Joined Center for Advanced Studies in Engineering (CASE) in 2013 and presently working as professor of mathematics at Siryed CASE Institute of Technology Islamabad Pakistan. He is the head of Department of Sciences and Humanities at Sir Syed CASE IT (SSCASEIT). His main research areas are optimization, data science, and CFD. He can be contacted at email: saleem.cfd@gmail.com.

Dr. Jawwad Sami Ur Rahman has done BS in biomedical engineering in 2011 from Sir Syed University of Engineering and Technology. From Pakistan and M.Sc. in technology management in 2013 and done Ph.D. in engineering from Asia Pacific University in 2023. His research interest revolves around analysis of brain tumour by applying various medical image segmentation techniques in MRI images. Currently, he is working as assistant professor in Riphah International University in Islamabad, Pakistan. Moreover, he is the branch counselor of IEEE for Faculty of Engineering and Applied Sciences (FEAS) and Co-Chiarman of MDT Board in Pakistan for tumour research. He can be contacted at email: jawwad_sami_87@hotmail.com.

Dr Sathish Kumar Selvaperumal completed his Ph.D. program at Sri Chandrasekharendra Saraswathi Viswa Mahavidyalaya University, Chennai, India in the year 2014. He completed his B.E degree in electronics and communication engineering in the year 2001 and M.E applied electronics in the year 2006, at Arulmigu Meenakshi Amman College of Engineering, Kanchipuram, and Chennai, India. He has 20 years of teaching experience and he is currently working as associate professor and program leader for telecommunication engineering in Asia Pacific University (APU) of Technology and Innovation, Technology Park of Malaysia, Kuala Lumpur, Malaysia. He is the Final Year Project Manager in the School of Engineering, APU. He is a charted engineer (CEng,U.K) and member of the Institution of Engineering and Technology (IET), Institute of Electrical and Electronics Engineers (IEEE), Indian Society for Technical Education (ISTE), International Association of Computer and Information Technology (IACIST), Singapore and International Association of Engineers (IAENG), Hong Kong. He has been the reviewer for more than 50 international conferences and journals. He has published and presented more than 50 research papers in national and international conference and reputed journals. He has been the keynote speaker for various international conferences. His research interests are image segmentation in angiogram images, brain and liver images, image enhancement, image compression, image retrieval, watermarking and speech detection and speech processing, optical communication, IoT, antenna design, AI, and robotics. He can be contacted at email: sathish@staffemail.apu.edu.my.