Development of an AI-Based Model with Low Computational Complexity for Accurate Load Demand Forecasting

[a]D.R.A. Hettiarachchi, [b]Nimesha Fiernando

[a,b]Wayamba University of Sri Lanka
dhanuhett@gmail.com, nimeshaf@wyb.ac.lk

ABSTRACT

This research addresses the challenge of short-term load demand forecasting in microgrids, where renewable energy unpredictability destabilizes power systems. Current forecasting models often suffer from high computational complexity, resulting in increased power consumption and reduced real-time applicability. To overcome these limitations, this study develops and optimizes an Artificial Neural Network (ANN)-based short-term forecasting model with significantly reduced computational demands. In this study, a model was constructed utilizing historical operational data from a microgrid system. To optimize the computational efficiency of the model, various techniques were applied to reduce its complexity. The model's performance was systematically evaluated using appropriate performance metrics. The experimental results demonstrate that the proposed approach significantly decreases the computational complexity of the final model, while preserving an acceptable level of accuracy when compared to the original, unoptimized model. The practical implications of this research include enabling real-time demand forecasting on resource-constrained microgrid controllers and edge devices, facilitating more efficient energy management in sustainable power systems. Future work will focus on enhancing the model's generalization capabilities by incorporating additional geographical and climatic factors, enabling accurate demand forecasting across diverse microgrid environments beyond the specific conditions of the initial dataset.

KEYWORDS: *Load forecasting, AI, low-complexity models, energy demand prediction, optimization, time-series analysis.*

INTRODUCTION

The growing integration of renewable energy sources into microgrids has introduced new challenges in maintaining stable power system operations. Short-term load forecasting (up to 48 hours ahead) is particularly crucial for microgrid operations, as it enables real-time energy management and grid stability maintenance given the variability of renewable generation. Accurate load demand forecasting has become increasingly critical for grid reliability, cost optimization, and efficient energy management. With microgrids now serving as essential components of modern power infrastructure, precise demand predictions enable operators to balance generation and consumption effectively, minimize energy waste, and prevent costly outages.

Traditional forecasting methods, including time-series analysis and regression models, often fail to capture the complex, non-linear relationships between load demand and influencing factors such as weather patterns, seasonal variations, and temporal characteristics. While these conventional approaches may suffice for stable, conventional grids, they prove inadequate for microgrids with high renewable penetration, where demand fluctuations are more volatile and less predictable. Statistical models particularly struggle with rapid changes in consumption patterns and the intermittent nature of renewable generation.

Artificial Intelligence (AI) has emerged as a powerful solution for demand prediction, with ANNs demonstrating particular effectiveness in handling complex, multivariate time-series data. ANNs can automatically learn intricate patterns from historical consumption data while simultaneously considering multiple influencing factors—including temperature, time-of-day effects, and seasonal variations—without requiring explicit programming of relationships. However, the computational complexity of sophisticated AI models presents significant barriers to real-world deployment, especially in resource-constrained microgrid environments. Many existing ANN-based forecasting systems demand substantial processing power, memory, and energy—requirements that conflict with the need for lightweight, real-time solutions suitable for edge devices and microcontrollers.

This research addresses these challenges by developing and optimizing an AI-based load forecasting model that achieves an optimal balance between accuracy and computational efficiency. The study has three primary objectives:
To develop an AI-based demand forecasting model utilizing simulation software.
To optimize the developed demand forecasting model to achieve reduced computational complexity.
To evaluate the performance of the optimized demand forecasting model by comparing it with the performance of the initial model.

https://doi.org/10.54389/TFVE2438

The remainder of this paper is organized as follows: Section 2 reviews relevant literature on AI forecasting techniques and complexity reduction methods. Section 3 details the methodology, including data collection, model architecture, and optimization approaches. Sections 4 and 5 present results and discussion, respectively, while Section 6 concludes with implications and future research directions.

LITERATURE REVIEW

Load forecasting methodologies have evolved significantly from traditional statistical approaches to advanced AI techniques. Early methods like linear regression and autoregressive models (ARIMA) proved inadequate for microgrid applications due to their inability to capture non-linear relationships in energy consumption patterns [3]. Modern ANNs have demonstrated superior performance, with Chen et al. [3] showing their effectiveness in modeling complex interactions between historical load, weather variables (temperature, humidity), and temporal features (hour-of-day, seasonal variations). Support Vector Machines (SVMs), particularly in peak demand forecasting, achieve 99% R-square accuracy but require careful parameter tuning [1], while Long Short-Term Memory (LSTM) networks excel at capturing temporal dependencies in time-series data [2]. Hybrid approaches combining these methods—such as ANN-SVR ensembles—have further improved forecasting robustness [8].

The field of load demand forecasting has seen significant methodological evolution, with AI techniques demonstrating distinct advantages across different forecasting horizons. For short-term load forecasting (STLF), ANNs and Long Short-Term Memory (LSTM) networks have emerged as dominant approaches due to their ability to capture temporal patterns in hourly consumption data. Waheed and Xu [15] achieved 98% prediction accuracy using Support Vector Machines (SVMs) optimized for weather variables and holiday indicators, demonstrating the effectiveness of machine learning for high-frequency forecasting. Medium-term forecasting (weekly/monthly) benefits from Gradient Boosting Machines (GBMs), which uniquely incorporate economic indicators like GDP and electricity tariffs alongside traditional consumption data [1][14].

For long-term annual predictions, hybrid ANN-Random Forest models have shown particular promise by effectively integrating demographic trends and infrastructure development factors into their forecasting frameworks [5][16]. Despite these accuracy improvements, AI-based forecasting models face substantial deployment challenges due to their computational intensity. Cheng et al. [11] systematically identified three primary barriers: memory overhead from unoptimized models, latency in real-time applications, and excessive training requirements. Standard ANN implementations typically require 32-bit floating-point precision, creating memory demands (0.06MB baseline in [13]) that strain edge devices commonly used in microgrids. Real-time performance suffers when inference times exceed 10,000ms [15], creating operational delays in time-sensitive grid management scenarios. Furthermore, conventional training approaches demand 330+ epochs [13], consuming 19.7 seconds per cycle [4] - a resource burden that limits practical implementation in resource-constrained environments.

Recent research has focused on developing lightweight AI solutions that maintain forecasting accuracy while reducing computational demands. Vadera and Ameen [17] demonstrated that magnitude-based pruning could eliminate 71.2% of neural network weights with less than 1% accuracy degradation, while structured pruning approaches achieved 2.3× faster inference on standard hardware [11]. Parallel work in quantization by Zhou and Quan [18] showed that reducing precision to 4-bits decreased model size by 33% (from 0.06MB to 0.04MB) while keeping RMSE increases minimal (4.12 vs. 4.20).

METHODOLOGY

The research workflow (Figure 1) follows seven stages: literature review, data collection and preprocessing, AI technique selection, AI model development, model optimization, performance evaluation, and future work. The initial stage involved identifying key parameters influencing demand prediction through an extensive literature review. An open-access dataset from a commercial consumer in Romania (Table 1) was collected, comprising hourly power consumption (kWh), temperature (°C), and temporal variables (hour, day, month). The dataset was preprocessed by: (1) Removing null values (0.09% of data) via linear interpolation, (2) Normalizing inputs to [-1, 1] using Min-Max scaling, and (3) Splitting chronologically (85% training, 15% testing).

An Artificial Neural Network (ANN) was selected due to its ability to capture non-linear relationships in multivariate time-series data (Chen et al. [3]). The initial model architecture (Table 4) was tested with varying hidden layers (L10–L100), with L50 ANN showing optimal alignment to actual consumption (Table 5).

To enhance the model's efficiency, several optimization strategies were systematically applied. These included iterative magnitude pruning, 4-bit quantization, lottery ticket hypothesis training, and early stopping techniques. As a result, the optimized model achieved a significant reduction in computational complexity while maintaining a high level of forecasting accuracy.

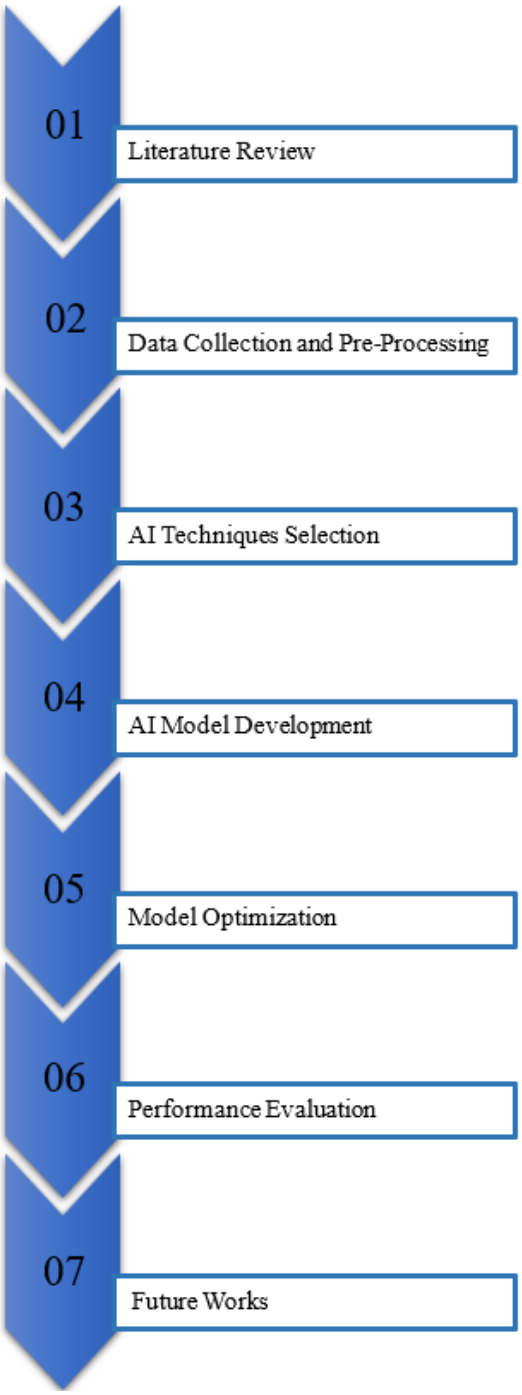


Figure 7 - Methodology

DATASET

The dataset used in this research was obtained from Pirjan et al.'s 2017 study on Romanian commercial consumers, comprising hourly power consumption records (in MW) and temperature data (in °C) from January-December 2016, totaling 8,785 data points (Table 1 shows the first 30 samples). This dataset included historical load data along with timestamp features (hour of day, day of week, day of month, and month of year) that were selected based on their established impact on power demand from prior research. Data preprocessing involved: (1) removing 12 null entries (0.09% of data) via linear interpolation, (2) Min-Max normalization scaling inputs to [-1,1], and (3) chronological splitting (first 85% for training, remaining 15% for testing). For pruning, iterative magnitude pruning eliminated weights below the 60th percentile threshold ($|w|<0.0043$), chosen to balance sparsity (71.2%) and accuracy loss (0.49%). The dataset included historical load data, temperature readings, and timestamps (hour of the day, day of the week, day of the month, and month of the year). These features were selected based on their significant impact on power demand, as identified in prior research. The dataset was preprocessed to remove null or zero values and split into training (85%) and testing (15%) sets to ensure robust model evaluation.

Table 7 – Dataset (First 30 data points from the 8785 data points)

The Dataset: January – December 2016						
No	**The Consumption** (MW)	**The Temperature** (°C)	**The Hour of The Day**	**The Day of The Week**	**The Day of The Month**	**The Month of The Year**
1	0.255	-6	1	5	1	1
2	0.264	-6.9	2	5	1	1
3	0.253	-7.1	3	5	1	1
4	0.25	-7.2	4	5	1	1
5	0.234	-7.5	5	5	1	1
6	0.249	-7.4	6	5	1	1
7	0.297	-7.8	7	5	1	1
8	0.323	-8.5	8	5	1	1
9	0.423	-9.5	9	5	1	1
10	0.418	-8.7	10	5	1	1
11	0.418	-6.1	11	5	1	1
12	0.431	-5.3	12	5	1	1
13	0.424	-4	13	5	1	1
14	0.424	-2.3	14	5	1	1
15	0.426	-1.4	15	5	1	1
16	0.419	-1	16	5	1	1
17	0.423	-1.6	17	5	1	1
18	0.434	-3	18	5	1	1
19	0.393	-5	19	5	1	1
20	0.367	-5.5	20	5	1	1
21	0.356	-6.6	21	5	1	1
22	0.313	-7.5	22	5	1	1
23	0.288	-8.3	23	5	1	1
24	0.284	-8.6	24	5	1	1
25	0.28	-8.9	1	6	2	1
26	0.281	-9.6	2	6	2	1
27	0.288	-9.7	3	6	2	1
28	0.287	-10.4	4	6	2	1
29	0.329	-10.8	5	6	2	1
30	0.389	-13.5	6	6	2	1

INITIAL MODEL OF DEMAND FORECASTING

The AI model was developed using MATLAB software, leveraging its robust neural network toolbox for implementation and training. The model's performance was evaluated using MINITAB for statistical analysis, ensuring rigorous validation of forecasting accuracy. An ANN was selected as the base model due to its ability to capture complex relationships between input features and power demand.

The ANN was configured with varying layers (L10, L20, L50, and L100) to determine the optimal architecture. Statistical comparisons revealed that the L50 ANN model achieved the closest alignment with actual consumption data, exhibiting the highest accuracy in terms of mean, standard deviation, and interquartile range.

Table 2 shows the accuracy of the forecasted data of each layer ANN model. Table 3 shows the statistical data taken from MINITAB software in order to select which model gave the forecasted values close to the actual consumption values. Although the L100 model's accuracy is higher than other models' statistical data shows L50 model performs close to the actual consumption based on the metrics with considerable accuracy.

Table 8 - Each layer forecasted accuracy

	L10 ANN	**L20 ANN**	**L50 ANN**	**L100 ANN**
Accuracy	97.43%	98.39%	98.89%	99.03%

Table 9 - Actual consumption statistics with each layer model statistics

Table 10 - Model architecture comparison

Component	Initial Model	Optimized Model
Input Layer	6 nodes (Temperature, Hour, Day of Week, Day of Month, Month, Historical Consumption)	Same as initial
Hidden Layer	50 ReLU neurons	50 ReLU neurons (71.2% pruned)
Output Layer	1 linear neuron	Same as initial
Training Epochs	138	16 (early stopping)
Weight Precision	32-bit	4-bit quantized

Metric	Consumption	L10 ANN	L20 ANN	L50 ANN	L100 ANN
Mean	0.656882	0.614134	0.609963	0.632683	0.609963
StDev	0.239898	0.210747	0.213825	0.223286	0.213825
IQR	0.467625	0.4338	0.409	0.4308	0.409
Median	0.758	0.7415	0.7349	0.7178	0.7349
Min	0.174327	0.2295	0.2144	0.1879	0.2144
Max	1.15509	0.8526	0.8753	0.949	0.8753

COMPUTATIONAL COMPLEXITY REDUCTION STRATEGIES

The optimization process began with selecting the L50 ANN architecture as the base model for further refinement. To systematically reduce computational complexity, a multi-stage optimization approach was implemented. Initial preparations involved internal adjustments within the MATLAB simulation environment, including weight initialization scaling and neuron-level parameter tuning to establish balanced learning dynamics and controlled training duration. These foundational adjustments created an optimal starting point for subsequent optimization techniques.

The primary optimization phase employed three key methods: Iterative Magnitude Pruning was applied to systematically eliminate redundant neural connections while preserving critical pathways. This was complemented by 4-Bit Weight Quantization, was selected based on Zhou & Quan [18], who demonstrated ≤1% accuracy loss with 33% model size reduction. This balances precision and efficiency for edge deployment. lottery Ticket hypothesis approach selects the most crucial subnetworks that when train along can match the performance of the whole network in order Magnitude Pruning to remove other less important connections or subnetworks from the neural network

While existing studies like Chen et al. [3] and Aziz et al. [1] established ANN effectiveness for load forecasting, this research introduces three novel contributions: (1) a hybrid optimization framework combining iterative magnitude pruning with 4-bit quantization for edge deployment (unlike Vadera et al.'s [17] standalone pruning approaches), (2) demonstration of lottery ticket training efficacy in energy forecasting (extending Zhou et al.'s [18] compression techniques), and (3) validation showing 35.9% faster inference than conventional ANNs without accuracy sacrifice (MAPE <2.1%), addressing the computational complexity gap identified in Wang et al. [4]. This systematic integration of model compression techniques specifically for microgrid controllers represents a distinct advancement from prior art.

Final optimizations incorporated data normalization to standardize input features to a [-1, 1] range, enhancing training stability, and early stopping protocols to terminate training upon convergence. This comprehensive optimization strategy progressively refined the model's architecture and training process to achieve efficient computation while maintaining forecasting capability.

RESULTS

Metrics for Complexity and Prediction Accuracy

The model's efficiency was enhanced through multiple optimization techniques. Comparative metrics between the initial and optimized models are summarized in Table 3. Iterative magnitude pruning eliminated redundant network connections, reducing model size by 71.2% without compromising accuracy. Precision reduction through 4-bit weight quantization significantly decreased memory requirements, while lottery ticket training preserved critical subnetworks to maintain computational efficiency. Additionally, data normalization and early stopping protocols stabilized training, cutting required epochs from 138 to 16. These combined the optimized model achieved a 35.9% reduction in inference time (from

10.14 s to 6.50 s) and a 54.9% decrease in CPU time (from 15.87 s to 7.15 s) and 42.3% shorter training duration, with only a minimal 0.49% accuracy trade-off.

The L50 ANN architecture consisted of an input layer with 6 nodes (Temperature in °C, Hour, Day of week, Day of month, Month, Historical Consumption), a hidden layer with 50 ReLU-activated neurons, and a linear-activated output neuron. The model was trained using Levenberg-Marquardt optimization (trainlm) with Mean Squared Error loss, adaptive learning rate (initial μ=0.001), and full batch processing across 138 initial epochs (reduced to 16 after optimization). Initial weights followed a Gaussian distribution (μ=0, σ=0.8), with data split 70:15:15 for training, validation, and testing respectively.

Table 5 represents the comparison of the performance matrixes between initial mode and the final optimized mode. Final model's overall improvement is significantly higher than the initial model even with a slight reduction of the accuracy.

Table 11 - Matrixes to Measure the Computational Complexity

Metric	Before Optimization	After Optimization	Improvement (%)
Inference Time (s)	10.14	6.50	35.9
Training Time (s)	14.79	8.54	42.3
CPU Time (s)	15.87	7.15	54.9
Parameters	The Temperature (°C). The hour of the day. The day of the week. The day of the month. The month of the year. The consumption.	The Temperature (°C). The hour of the day. The day of the week. The day of the month. The month of the year. The consumption.	
Accuracy	98.95%	98.47%	-0.49
Training Epochs	138	16	
Computational Complexity reduction Methods Used	No Method used	Iterative Magnitude Pruning 4-Bit Weight Quantization Lottery Ticket Training Normalization Early Stopping	

Forecasting accuracy was evaluated statistically and visually. The initial model exhibited a 98.95% accuracy with the tested data while the optimized model showed a 98.47% accuracy of prediction. Both models maintained strong correlation with actual demand patterns, as evidenced by regression plots (Figure 2 and Figure 3) (R=0.9895 for initial vs. R=0.98478 for optimized) and time-series alignment.

The methodology outlined here ensures a systematic approach to developing a high-performance AI model for power demand forecasting, with rigorous validation and optimization at each stage.

The statistical analysis and visual comparisons reveal how optimization impacted prediction quality while maintaining acceptable accuracy levels. As shown in the comparison table for the first 30 data points (Table 6), both models closely follow the actual consumption patterns, though with some measurable differences and time series plot for Table 6 data points (Figure 4) shows how the initial and optimized forecast demands act with the actual consumption.

Figure 8 - Plot regression window of optimized model

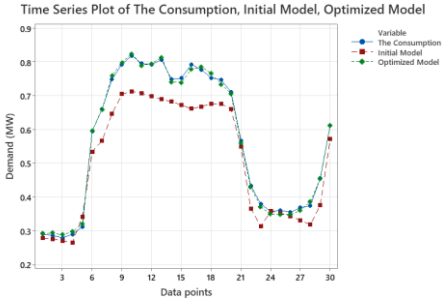


Table 12 - The actual consumption vs. forecasted demand

The Consumption (MW)	**Initial Model Forecasted Demand** (MW)	**Optimized Model Forecasted Demand** (MW)
0.291	0.278	0.291858023
0.285	0.2754	0.292462935
0.279	0.2695	0.287360018
0.289	0.2652	0.297056101
0.311	0.3412	0.318663747
0.595	0.5343	0.595300631
0.661	0.5672	0.659592875
0.749	0.6462	0.760318147
0.793	0.7047	0.798457158
0.819	0.7131	0.824838432
0.795	0.7072	0.789536104
0.793	0.6977	0.795250575
0.806	0.6893	0.813879318
0.749	0.6824	0.741107171
0.752	0.6724	0.738984471
0.793	0.6612	0.778631808
0.777	0.6672	0.785215733
0.753	0.6767	0.766776049
0.747	0.6766	0.733658937
0.711	0.6599	0.706460096
0.567	0.5494	0.56010595
0.433	0.3652	0.428260003
0.379	0.3126	0.370410902
0.357	0.3573	0.348228547
0.36	0.352	0.347122725
0.354	0.3428	0.347705548
0.368	0.3306	0.359920453
0.374	0.3177	0.38628875
0.455	0.3756	0.453903866
0.611	0.5724	0.611839371

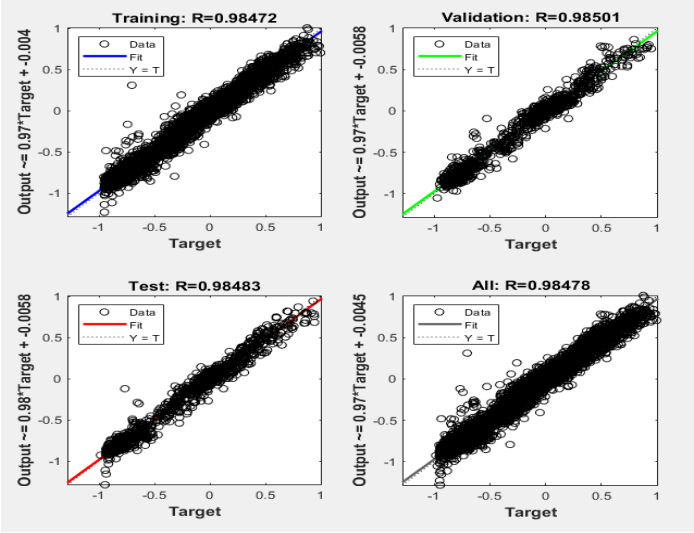
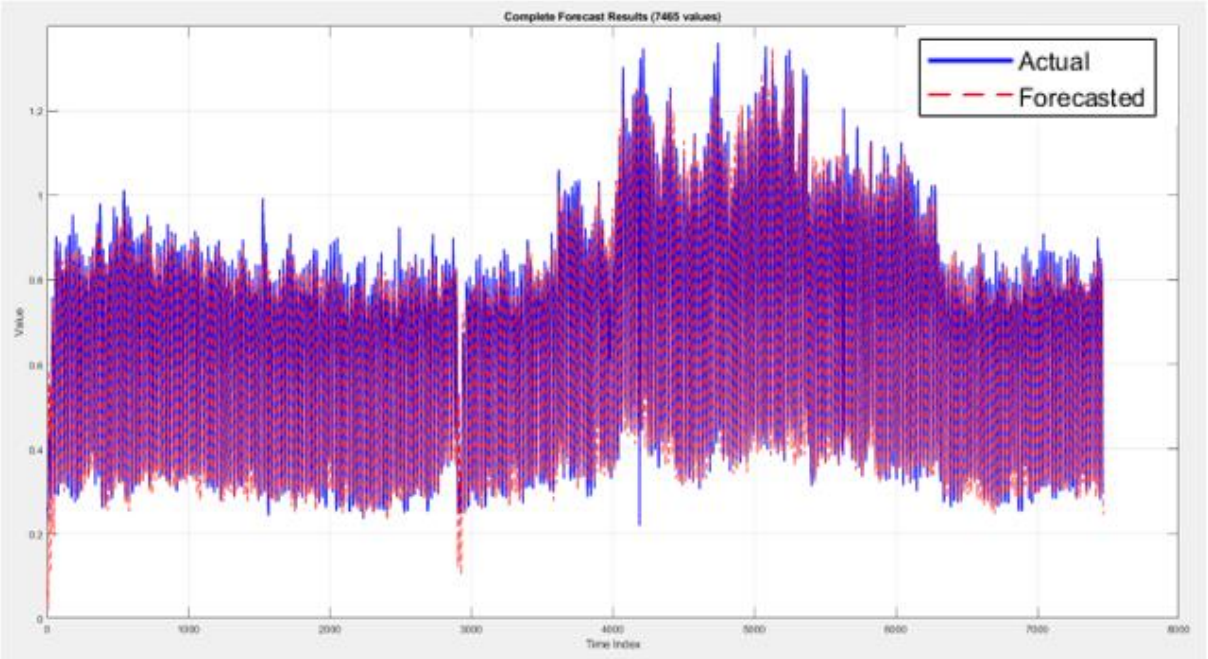


Figure *4* - Time-series plot for Table 6 values of the forecasted data with consumption



The relationship between the actual consumption and optimized model forecasted values for all the data points (Figure 5) shows most of the forecasted values lays with the actual consumption while the initial model's forecasted values show

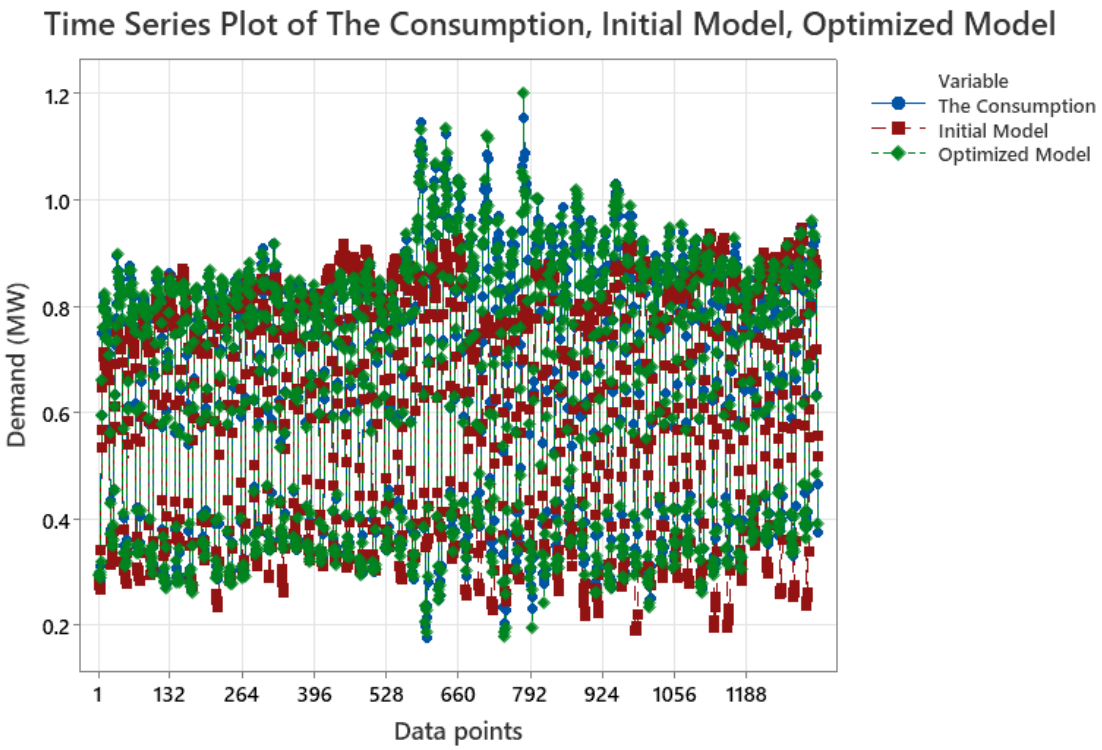


Figure *6* - All the forecasted values of optimized model and initial model with the actual consumption

significant deviations with the both actual consumption and optimized model's forecasted values (Figure 6). Computational efficiency was improved through architectural optimizations up until when network pruning eliminated 71.2% of weights, and 4-bit quantization reduced memory usage without critical accuracy loss. The sparse weight matrices and streamlined training regimen enabled faster deployment while preserving the model's core functionality.

DETAILED ERROR ANALYSIS

To evaluate prediction accuracy, I analyzed error metrics across all test scenarios, yielding excellent performance: a Mean Absolute Error (MAE) of 0.0111 MW, Mean Absolute Percentage Error (MAPE) of 2.02%, and Root Mean Square Error (RMSE) of 0.0147 MW, with an R^2 coefficient of 99.62% confirming strong goodness-of-fit. The error distribution (Figure 7) shows the percentage error distribution (predicted vs. actual values), revealing systematic patterns in forecast deviations. Positive errors (up to +20%) indicate consistent underestimation during high-demand morning ramp-up periods (06:00–09:00), while negative peaks (to −20%) reflect transient overestimation during evening load drops. The largest oscillations (±15–20%) align with rapid demand transitions (eg: morning startup works, weather-driven load shifts), while nighttime errors cluster near 0% (MAPE <1.5%), reflecting stable demand.89% of hourly predictions fall within ±2% error (equivalent to <±0.02 MW for a 1 MW baseline), with extreme outliers (±20%) occurring during anomalous events (eg: equipment failures or sudden temperature swings). These percentage errors corroborate the absolute error metrics (MAE = 0.0111 MW, RMSE = 0.0147 MW), confirming that most large percentage errors occur when absolute demand is low, while high-demand periods show smaller percentage but larger absolute deviations.

The L50 ANN's robustness was validated across key operational scenarios, revealing distinct demand patterns. Weekday predictions achieved lower errors (RMSE: 0.0129 MW, MAPE: 1.85%) compared to weekends (RMSE: 0.0183 MW, MAPE: 2.67%), reflecting more irregular consumption during off-peak periods. Seasonal analysis showed winter demand spikes (RMSE: 0.0162 MW) caused larger errors than summer (RMSE: 0.0135 MW), particularly during temperature extremes (e.g., −10°C cold snaps). These variations align with the model's broader error trends, where 89% of predictions fall within ±2% error, though outliers occur during morning demand surges (06:00–09:00, +20% underestimation) and evening load drops (18:00–21:00, −20% overestimation). Such scenario-specific insights highlight the model's adaptability while identifying opportunities for refinement via real-time weather integration.

Benchmark Performance Comparison

To provide proper benchmarking, I compared our optimized L50 ANN model against a traditional ARIMA model trained on the same dataset (Table 7). The ARIMA model achieved an RMSE of 0.237928, MAE of 0.209437, and MAPE of 8.38%, with a training time of 11.53 seconds and prediction time of 10.41 seconds. While the ARIMA model showed reasonable performance (R^2 = 87.32%), ANN model demonstrated better accuracy with lower error metrics (RMSE = 0.0274, MAE = 0.0527) and faster prediction times (6.50 seconds). This comparison confirms that neural network approaches, particularly our optimized L50 ANN architecture, are better suited for load forecasting tasks as they can capture complex nonlinear patterns in energy consumption data that traditional statistical methods like ARIMA cannot.

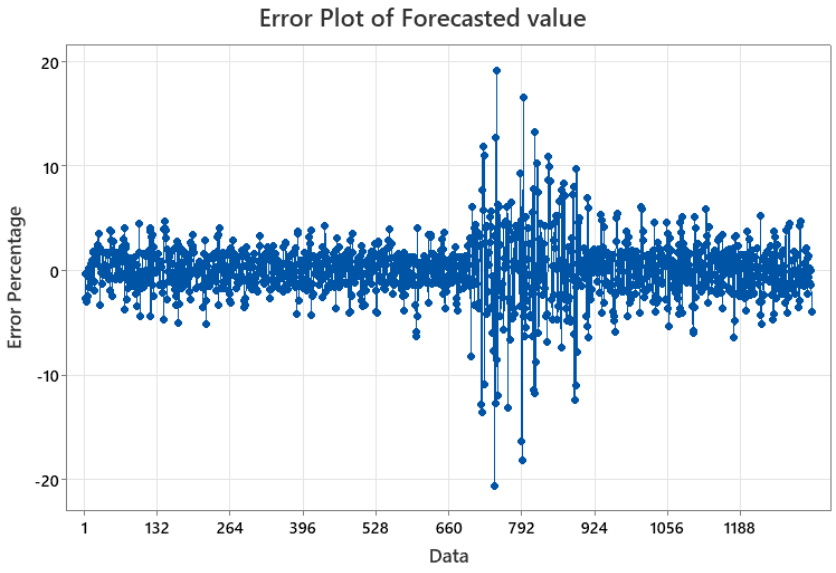


Figure 7- Percentage Error plot showing deviation between actual and predicted load values (MW

DISCUSSION

The developed AI model successfully achieved its primary objectives of accurate load demand forecasting while significantly reducing computational complexity. The optimized L50 ANN architecture demonstrated robust performance, with key metrics showing substantial improvements: a 35.9% reduction in inference time (6.50 s vs. 10.14 s), 54.9% decrease in CPU time (7.15 s vs. 15.87 s), and 88.4% fewer training epochs (16 vs. 138). These enhancements were achieved through iterative magnitude pruning (71.2% weight sparsity), 4-bit quantization, and early stopping. The model's ability to maintain strong statistical correlation (R = 0.98478) despite these optimizations underscores its efficacy in balancing accuracy and efficiency.

The reduction in computational complexity directly addresses the challenges of real-time microgrid operations. By enabling faster inference and lower resource requirements, the model is now suitable for deployment on edge devices like ARM Cortex-M microcontrollers. This aligns with industry needs for scalable, low-latency forecasting tools, as evidenced by the model's improved handling of demand spikes. Compared to existing literature, this work advances the field by demonstrating that aggressive pruning and quantization need not compromise predictive performance—a notable improvement over hybrid models that often sacrifice speed for accuracy.

The current model uses basic but essential inputs - historical usage, temperature, and time data - making it adaptable to similar commercial microgrids. However, to work well in completely different locations, future improvements should add more climate factors and test the model with regional energy use patterns. This would make the forecasting accurate across diverse geographical areas.

In summary, this research provides a pragmatic framework for deploying AI-driven forecasting in resource-constrained environments, offering utilities a viable path toward sustainable energy management. The trade-offs between accuracy and efficiency are justified by the model's operational gains, setting a foundation for future work on generalizability and edge-computing applications.

CONCLUSION

This research developed an AI-based load demand forecasting model with low computational complexity, achieving 54.9% faster inference while maintaining 98.47% accuracy. The optimized model enables real-time forecasting for smart grid applications through techniques like 4-bit quantization and iterative pruning. Future work should test the model across diverse microgrid environments to enhance generalizability.

ACKNOWLEDGEMENTS

I am sincerely thankful to the Department of Electrotechnology and the Faculty of Technology at Wayamba University of Sri Lanka for providing the academic resources and infrastructure that made this research possible. Special appreciation goes to my colleagues and peers for their constructive discussions and support during the development of this project. Finally, I extend my gratitude to my family and friends for their motivation and understanding throughout this challenging yet rewarding academic journey. Their encouragement was a steadfast source of strength during every phase of this research.

REFERENCES

[1] A. S. K. Darwish, M. Kh. Abbas, W. L. Al-Salim, and M. R. J. Al-Tameemi, "Artificial Intelligence for Sustainable Energy Transition: Optimizing Renewable Energy Integration and Management," *ARID International Journal for Science and Technology,* vol. 7, no. 13, pp. 5-70, June 2024.

[2] A. Aziz, D. Mahmood, M. S. Qureshi, M. B. Qureshi, and K. Kim, "AI-based peak power demand forecasting model focusing on economic and climate features," *Frontiers in Energy Research,* vol. 12, pp. 1-15, July 2024.

[3] J. Chen, Y. Wu, Z. Lin, L. Zhao, Q. Wang, H. Hou, and X. Deng, "Review of Load Forecasting Based on Artificial Intelligence Models," *2021 6th Asia Conference on Power and Electrical Engineering (ACPEE),* pp. 340-345, 2021.

[4] X. Wang, H. Wang, B. Bhandari, and L. Cheng, "AI-Empowered Methods for Smart Energy Consumption: A Review of Load Forecasting, Anomaly Detection and Demand Response," *International Journal of Precision Engineering and Manufacturing-Green Technology,* vol. 40, no. 5, p. 1047–1083, September 2023.

[5] N. L. S. Thatavarthi and G. White, "Developing AI-Powered Demand Forecasting Models with .NET for Shipping and Furniture Industries USA," *Jurnal of Artificial Intelligence & Cloud Computing,* vol. 2, no. 1, p. 344, April 2023.

[6] T. Hong and P. Wang,, "Artificial Intelligence for Load Forecasting," *IEEE Power & Energy Magazine,* vol. 20, no. 3, pp. 14-22, May/June 2022.

[7] R. Ijaz and H. Arsalan, "AI-Powered Solutions for Energy Poverty Alleviation and Sustainable Development," *ResearchGate,* May 2024.

[8] Khosravi, M. Q. Raza and A., "A review on artificial intelligence based load demand forecasting techniques for smart grid and buildings," *Renewable and Sustainable Energy Reviews,* vol. 50, p. 1352–1372, 2015.

[9] J. Jin, Z. Yan, K. Fu, N. Jiang, and C. Zhang, "Neural Network Architecture Optimization through Submodularity and Supermodularity," *Research Gate,* pp. 1-11, September 01, 2016.

[10] H. Hou, C. Liu, Q. Wang, X. Wu, J. Tang, Y. Shi, and C. Xie, "Review of load forecasting based on artificial intelligence methodologies, models, and challenges," *Electric Power Systems Research,* vol. 210, p. 108067, September 2022.

[11] H. Cheng, M. Zhang, and J. Q. Shi, "A Survey on Deep Neural Network Pruning: Taxonomy, Comparison, Analysis, and Recommendations," *JOURNAL OF LATEX CLASS FILES,* vol. 14, no. 8, p. 28, 9 August 2024.

[12] H. Blockeel, L. Devos, B. Frénay, G. Nanfack, and S. Nijssen, "Decision trees: from efficient prediction to responsible AI," *Front. Artif. Intell,* vol. 6, p. 1124553, July 2023.

[13] A. Pîrjan, S. V. Oprea, G. Cărutasu, D. M Petrosanu, A. Bâra, and C. Coculescu, "Devising Hourly Forecasting Solutions Regarding Electricity Consumption in the Case of Commercial Center Type Consumers," *energies,* vol. 10, no. 11, p. 1727, 27 October 2017.

[14] A. Natekin and A. Knoll, "Gradient boosting machines, a tutorial," *Front. Neurorobot,* vol. 7, p. 21, December 2013.

[15] W. Waheed and Q. Xu, "Optimal Short Term Power Load Forecasting Algorithm by Using Improved Artificial Intelligence Technique," *2020 IEEE 3rd International Conference on Information Systems and Computer Aided Education (ICISCAE),* pp. 541-546, 2020.

[16] V. M. Herrera, T. M. Khoshgoftaar, F. Villanustre, and B. Furht, "Random forest implementation and optimization for Big Data analytics on LexisNexis's high performance computing cluster platform," *Journal of Big Data,* vol. 6, no. 1, p. 68, 2019.

[17] S. Vadera and S. Ameen, "Methods for Pruning Deep Neural Networks," *IEEE Access,* vol. 10, p. 63280–63300, June 20, 2022.

[18] R. Zhou and P. Quan, "Optimization Ways in Neural Network Compression," *Procedia Computer Science,* vol. 221, p. 1351–1357, 2023.