EXPERIMENTAL INVESTIGATION ON LONG-TERM FORECASTING OF AREA POWER LOAD USING DEEP LEARNING AND MACHINE LEARNING METHODS

Paresh S Chaudhari[1]

[1]Lecturer, B and B Institute of Technology, Vallabh Vidyanagar -388120, India. Contact information: pareshchaudhariee@gmail.com.

Abstract:
Climate change and smart grid advancements have heightened the need for accurate electricity demand forecasts. This study develops and compares district-level load prediction models using machine learning and deep learning techniques. Results show deep learning outperforms traditional methods with an R-Squared of 0.93–0.96 and MAPE of 4–10%. The model benefits grid management and expansion for municipalities and utility companies, and aids households in adopting renewable energy technologies.

Keywords: neural networks (RNN); random forest; support vector machine (SVM) long short-term memory (LSTM); deep learning; machine learning; non-linear auto-regressive exogenous (NARX)
DOI Number: 10.48047/nq.2020.18.6.NQ20193 NeuroQuantology 2020;18(6):128-134

1. Introduction

Human activities and the combustion of fossil fuels for energy production in previous years have precipitated a dire global climate crisis. This crisis, characterized by rising temperatures worldwide, has spurred changes in how urban areas consume energy and raised significant concerns for power providers and utility sectors globally. For instance, a study focusing on the UK's office heating and cooling consumption anticipates a 2–4 kWh/m2 increase in annual cooling usage by 2030 due to climatic shifts. Additionally, factors like urbanization and industrialization have profoundly influenced energy consumption patterns and demand. As energy needs continue to rise, there is a growing imperative for accurate forecasting to inform long- and medium-term strategies and regulate energy usage in buildings effectively [1]. Electric power demand forecasting offers various benefits, including enhancing energy efficiency, detecting faults, and developing reliable budget forecasts, especially in the face of population growth. Consequently, medium- and long-term energy forecasting has become critical for resource management, integral to feasibility assessments for constructing new power plants. Load forecasting, the practice of predicting future energy demand, is categorized into long-term, medium-term, and short-term forecasting, each serving distinct planning purposes, from expansion to operational efficiency and cost minimization[2].

The growing incorporation of renewable energy sources into power grids, coupled with their intermittent nature, presents new challenges and heightens the demand for more precise and granular load forecasting.

Maintaining a short- and medium-term balance between supply and demand is crucial for effectively managing modern grids. Additionally, limitations in energy storage at the district level further emphasize the critical importance of accurate load prediction, despite its inherent challenges. Traditionally, energy production and consumption forecasts rely on meteorological data like temperature, humidity, and wind speed, as well as past power measurements. While statistical methods such as ARIMA are commonly used for load forecasting across different time horizons, Machine Learning (ML) techniques have gained popularity due to their efficacy, precision, and adaptability, especially with the increased availability of digitalized data and computational resources.

In recent years, various machine learning and deep learning algorithms have been developed and employed to forecast future load demands with greater accuracy. These newer algorithms leverage more intricate mathematical models to achieve superior predictive performance compared to simpler methods. This study aims to assess the suitability of such techniques for predicting hourly energy consumption over a medium-term horizon. It explores and compares the concepts of ML and deep learning, proposing their application to support the development of sustainable strategies for electrical generation units and utilities.Historical data on electrical load, temperature, and wind speed are utilized to train the models, with the forecasted load demand serving as their output. Four distinct models are evaluated, each representing a different data mining method: Random Forest (RF) as an ensemble learning approach, Support Vector Machines (SVM) as a representative of ML, and Recurrent Neural Network (RNN) and non-linear auto-regressive exogenous (NARX) Neural Network as representatives of deep learning algorithms. While this comparison doesn't generalize to a broader assessment of data mining methods, the selection aims to encompass representatives from the simplest to the most complex techniques available [3].

This paper presents the findings of an extensive investigation into selecting the most suitable model for medium-term electrical load forecasting. The novelty of the proposed approach lies in systematically selecting a model based on data behavior and optimizing parameters to achieve the highest possible accuracy in load forecasting at the district level, while maintaining reasonable computational complexity. The paper is organized into five sections. In Section 2, the topic is introduced along with a review of literature on various types of load forecasting and techniques for predicting electricity consumption. Section 3 outlines the methodology employed in the research. A case study is presented in Section 4, detailing the results, performance evaluations, and comparisons of all techniques used. Section 5 discusses the optimization of hyperparameters aimed at enhancing the accuracy of each technique. Finally, Section 6 summarizes the paper and provides suggestions for future research directions.

2. Literature Review

In literature, the utilization of Machine Learning (ML) in electricity demand analysis is broadly categorized into unsupervised and supervised learning methods. Unsupervised methods, like clustering and anomaly detection, are often employed for descriptive analytics or preprocessing tasks. However, our focus in this paper is on predictive modeling, hence we've concentrated on supervised learning algorithms. Among these, Support Vector Machine (SVM) has shown consistent performance in addressing both linear and non-linear problems, making it a prominent choice in both research and industry applications.For example, Li et al. (2010) utilized SVM, alongside other ML techniques, to forecast annual residential energy demand in China. SVM exhibited the highest accuracy on test samples, with errors around 2%, outperforming competitors like General Regression Neural Network. Jain et al. (2013) investigated the impact of temporal and spatial granularity on SVM's predictability, finding it most effective in forecasting residential floor energy demand on an hourly scale [4].

Other studies have also explored various ML methods for energy demand forecasting. Ruiz-Abellón et al. (2018) used regression tree methods to predict short-term electrical consumption at a university campus in Spain, highlighting the effectiveness of the random forest method. Ahmad et al. (2017) compared Random Forest with Artificial Neural Networks (ANN) for HVAC energy consumption forecasting in Madrid, showing both models' effectiveness in hourly load demand prediction [5].

Furthermore, Khan et al. (2020) proposed hybrid ML methods combining random forest, extreme gradient boosting, and categorical boosting, achieving enhanced accuracy in energy demand forecasting [6].Moreover, researchers have explored specific types of neural networks, like Non-linear Autoregressive Exogenous (NARX) and Convolutional Neural Networks (CNN), to improve time-series data handling. Bendaoud and Farah (2020) proposed a new form of CNN for short-term load forecasting, reporting promising results in a case study in Algeria. Thokala et al. (2018) compared SVR and NARX neural network methods, finding SVR outperforming NARX in various scenarios.

Additionally, the ability of ML and deep learning methods for load forecasting at the building scale has been investigated by several researchers. Ahmad et al. (2018) proposed precise medium- and long-term district level prediction using ANNs and multivariate linear regression based on environmental and historical data, enhancing forecasting accuracy in the smart grid environment.In this paper, we aim to fill a gap in the literature by examining Long Short-Term Memory (LSTM) as a method for medium- and long-term load prediction. We compare LSTM against SVM and Random Forest, two robust ML models, to provide a reference point for evaluating LSTM's performance.

3. Materials and Methods

The high-level method employed in the current study involves several steps of data preprocessing before training the models. The data preprocessing consists of four primary steps: feature selection, outlier detection, replacing missing values, and normalization.Firstly, feature selection was applied to reduce the size of the raw data, particularly focusing on attributes in the climatic data that are logically unrelated to load consumption.Secondly, missing values were replaced by the average of their immediately preceding and succeeding one hour's values to ensure continuity in the dataset.Thirdly, negative demand values were removed from the database as inconsistencies. These instances were deemed implausible surplus production by prosumers based on contextual information gathered, indicating data collection errors.As for large magnitude load demand values, while they might appear as statistical outliers, none were removed from the dataset. These values, occurring frequently around the same time each year, were considered consistent data points possibly reflecting common behavioral patterns within the chosen municipality.The last step involved normalizing the data to bring all feature values into a common scale, aiding model training and convergence.Following data cleansing, the processed data was used to train the models, as further detailed in subsequent sections.

3.1. Random Forest (RF)

Decision Trees (DT) have long been favored in load forecasting due to their simplicity and interpretability. They operate by traversing a flowchart-like structure, where each node represents a decision based on historical load data and building characteristics, leading ultimately to a prediction at the leaf node.To enhance the performance of traditional decision trees, ensemble techniques like Random Forest (RF) have been developed. RF combines multiple decision trees and aggregates their predictions through voting, yielding more robust results. Each tree in the forest is trained using bootstrap resampling, generating random training sets. The trees are then constructed by selecting the best splits among randomly chosen input variables, and the process continues until stopping criteria are met.In this study, the Random Forest model was trained to forecast electrical load, with temperature identified as the most influential feature. The splitting of nodes was based on the least-square residuals criterion, minimizing the squared distance between node averages and actual values. This approach resulted in improved load prediction accuracy [7].

3.2. Support Vector Machine (SVM)

Support Vector Machines (SVM) are a supervised learning technique commonly used for classification and regression tasks. In the context of predicting electricity demand time series, Support Vector Regression (SVR) is applied. SVR operates by identifying support vectors, which are training samples situated on the ε-tube defining the decision surface. Residual values within this tube do not influence predicted values, while residuals beyond it are penalized during optimization [8].

The SVR model approximates the hourly demand forecast, represented by function Y, as:

$$Y = V \cdot \theta(X) + b$$

Here, θ(X) denotes the nonlinear kernel function, which implicitly maps the input space X (comprising historical consumption and climatic data) to a lower-dimensional space, facilitating maximal class distinction. This kernel function efficiently computes the dot product of transformed vectors, particularly advantageous in complex and nonlinear scenarios or high-dimensional data.

The coefficients V and the bias term b are determined by minimizing the primal objective function:

$$\frac{1}{2}||\omega||^2 + C\sum_{i=1}^{k}(\xi_i + \xi_i^*)$$

Where ω represents the weight vector optimized for generalization, and the parameter C controls the trade-off between minimizing errors and maximizing the margin. The objective function accounts for residuals beyond the tolerance ε through slack variables ξi and ξi^*. Optimization is subject to constraints ensuring that the discrepancy between observed and predicted values, as determined by the kernel function, remains within ε.

In this study, the chosen kernel function is the scalar (dot) product:

$$k(x, y) = x \cdot y = ||x|| \times ||y|| \times \cos\theta$$

Here, ||x|| and ||y|| denote the magnitude of vectors x and y, respectively, while θ represents the angle between them.

3.3. Deep Learning

A traditional multilayer perceptron (MLP) artificial neural network (ANN) constructs nonlinear models to understand and learn the connections between input and output values. Each node in the network calculates its response based on a weighted sum of inputs from previous nodes, along with a bias term, which is then transformed by an activation function.After obtaining all outputs, the model calculates the error between predicted and recorded energy consumption for each training example. This error is minimized through backpropagation and weight adjustments to iteratively reduce the overall cost function **[**9**]**.

Recurrent Neural Networks (RNNs), a type of deep ANN, use feedback loops to remember values from previous time steps, making them suitable for time-series data. Long Short-Term Memory (LSTM) networks address RNNs' challenge in handling long-term dependencies by retaining information over extended periods. The LSTM architecture includes a forget layer, which determines the relevance of past data, and mechanisms for updating cell states and generating outputs.Weights in each layer are optimized through training iterations, with optimizers like Adam and RMSprop used for efficiency. Dropout regularization prevents overfitting by temporarily removing contributions from randomly selected neurons during training[10].

Prediction model performance is evaluated using metrics like root mean square error (RMSE), coefficient of determination (R2), and mean absolute percentage error (MAPE). RMSE measures the model's error, MAPE quantifies the average percentage error, and R2 indicates how much variance in the dependent variable is explained by the model.

4. Results

4.1. Case Study

The selection of Bruce County as the case study for this research provides valuable insights into electricity demand forecasting in a small rural region. With an area of 4079 km² and a population of 68,147 as of 2016, Bruce County represents a typical rural setting in Ontario, Canada.Ontario relies on a mix of energy sources for power generation, including nuclear, solar, wind, hydropower, and fossil fuels. While nuclear power is predominant, the presence of other sources highlights the region's diverse energy landscape [11].

The initial dataset used in this study combines hourly weather data obtained from the government of Canada's environmental and natural resources database with electricity consumption data for Bruce County. The electricity consumption dataset spans from 2010 to 2019 and is sourced from the Independent Electricity System Operator (IESO), which manages Ontario's meter data.Among the various attributes in the weather dataset, temperature and wind speed are identified as the most significant for electricity demand prediction, based on their common usage in literature and their tested contribution to prediction accuracy. Temperature exhibits regular seasonal behavior, with summer temperatures ranging from 28.2°C to 32.5°C and winter temperatures ranging from -20.1°C to -30.6°C. The correlation between exterior temperature and electricity consumption, particularly in residential buildings, underscores the importance of weather variables in load forecasting. Additionally, wind speed, especially during winter, influences load demand, potentially due to the wind chill effect.By focusing on Bruce County and leveraging weather and electricity consumption data, this research aims to develop accurate forecasting models to enhance understanding and management of electricity demand in rural regions [12].

4.2. Implementation

The methodology employed in this research involves training predictive models for electricity demand forecasting using a tenfold cross-validation approach. Each fold of the cross-validation process involves using nine years of historical data for training and one year for testing. For example, one fold uses data from 2010 to 2018 for training and data from 2019 for testing. This process is repeated for each combination of training and test data, allowing for comprehensive model evaluation.Before training the models, the dataset undergoes preprocessing steps, including handling missing values and normalization. Missing values in the electricity load data are replaced with the average of the nearest loads preceding and succeeding them. Weather and load data are then cleansed and preprocessed before being used for model training.

Three models are trained and fine-tuned for comparison: Random Forest (RF) regression, Support Vector Regression (SVR), and Long Short-Term Memory (LSTM) neural network. Each model has specific hyperparameters that are optimized to maximize accuracy and minimize error during training.For the RF regression model, hyperparameters such as tree size and depth are tuned to prevent overfitting. The optimal tree size is determined to be 150, while the depth is limited to 30 to avoid overfitting [13].

In SVR, hyperparameters like misclassification tolerance (C) and kernel type are optimized to balance model generalization and computational efficiency. The dot kernel is selected as the most suitable kernel function, while a C value of 0 is chosen to set strict boundaries.For LSTM neural networks, hyperparameters such as optimizer, batch size, and number of epochs are optimized using GridSearchCV from the Scikit-learn library. The Adam optimizer, a batch size of 32, and 10 epochs yield the best performance.Additionally, dropout regularization is applied to prevent overfitting in both LSTM and NARX networks. In the NARX network, the number of hidden layers and delay parameters are manually set to 10 and 168, respectively.By carefully tuning the hyperparameters and applying appropriate regularization techniques, the models are optimized to accurately forecast electricity demand in Bruce County, Ontario. The evaluation of model performance is based on metrics such as mean squared error (MSE), root mean square error (RMSE), coefficient of determination (R2), and mean absolute percentage error (MAPE) [14].

5. Discussion

The performance evaluation of the three predictive models—Random Forest (RF), Support Vector Regression (SVR), Long Short-Term Memory (LSTM)—and the Nonlinear Autoregressive Exogenous (NARX) neural network indicates distinct strengths and weaknesses in predicting electricity load demand in Bruce County, Ontario.Starting with the RF model, it achieves an R2 value of 0.871 and an RMSE of 11.925 MW for predicting the electricity load in 2019. While these metrics fall within an acceptable range compared to the literature, the RF model struggles to accurately predict peak load demands, often underestimating lows and overestimating highs.The SVR model exhibits similar challenges in predicting peak load demands, but it achieves a slightly higher R2 of 0.877 and an RMSE of 12.308 MW for 2019. However, its MAPE is higher compared to RF, indicating a slightly less accurate prediction overall.In contrast, the LSTM model demonstrates superior performance, accurately predicting peak load demands with an R2 of 0.93 and an RMSE of about 8.3 MW for 2019. Its MAPE is also competitive at 10.21%, showing consistent performance across various load demand levels.The NARX neural network outperforms all other models with an RMSE of 5.81, an R2 of 0.96, and a MAPE of 4.2% for forecasting load demand in 2019. It shows the lowest error and highest accuracy among the models evaluated.A 10-fold cross-validation is conducted to address bias in testing and evaluate the models' performance comprehensively. The results show that the accuracy ranges achieved by SVM and LSTM models are very close and slightly higher than RF models. However, the NARX model exhibits the best performance in terms of both accuracy and error metrics.Overall, while all models can reasonably predict overall load demand, the main challenge lies in accurately predicting peaks. RF and SVM models consistently fail to match peak load demand, while LSTM performs better but still has mismatches, particularly for minimum peaks. The NARX model demonstrates considerably lower mismatches, making it the preferred choice for accurate load demand forecasting in Bruce County, Ontario. Additionally, the NARX model offers acceptable computational time, making it a practical choice for deployment in real-world applications.

6. Conclusions

Your study on predicting electricity demand at a regional scale using machine learning techniques makes significant contributions to the field, particularly in addressing the challenges posed by decentralized and intermittent energy production. Here are some key points summarizing the contributions and limitations of your study, as well as suggestions for future research:

1. Feasibility of Machine Learning Techniques: Your study demonstrates the feasibility of using machine learning techniques, including ensemble learning and deep learning, for accurately predicting electricity demand at a municipal level on an hourly basis.
2. Superior Algorithms: Through benchmarking and cross-comparison of various models, you identify LSTM and NARX neural networks as superior algorithms compared to traditional methods like SVM and RF, especially in predicting peak electrical consumption.
3. Adaptability of LSTM and NARX: LSTM and NARX models exhibit superior adaptability in predicting both peak and off-peak load values with maximum accuracy and minimum error, making them key for predicting time-series load demand.
4. Novel Methodology: Your study introduces tuned recurrent neural networks, specifically LSTM, for district-level load prediction, along with the development of the NARX model to evaluate the impact of recurrent dynamic networks on prediction accuracy.
5. Benchmarking and Cross-Comparison: By evaluating models based on accuracy and computational cost, you provide valuable insights for decision-

making in selecting appropriate models for predicting load demand at the district scale.

6.

References

1. Braun, M.R.; Altan, H.; Beck, S.B.M. Using regression analysis to predict the future energy consumption of a supermarket in the UK. Appl. Energy 2014, 130, 305–313. [CrossRef]
2. Ahmad, T.; Chen, H. Potential of three variant machine-learning models for forecasting district level medium-term and long-term energy demand in smart grid environment. Energy 2018, 160, 1008–1020. [CrossRef]
3. Karthika, S.; Margaret, V.; Balaraman, K. Hybrid short term load forecasting using ARIMA-SVM. Innov. Power Adv. Comput. Technol. i-PACT 2017, 2017, 1–7. [CrossRef]
4. Li, Q.; Ren, P.; Meng, Q. Prediction Model of Annual Energy Consumption of Residential Buildings. In Proceedings of the 2010 International Conference on Advances in Energy Engineering, ICAEE, Beijing, China, 19–20 June 2010; IEEE: New York, NY, USA, 2010; pp. 223–226.
5. Ahmad, M.W.; Mourshed, M.; Rezgui, Y. Trees vs Neurons: Comparison between random forest and ANN for high-resolution prediction of building energy consumption. Energy Build. 2017, 147, 77–89. [CrossRef]
6. 14. Khan, P.W.; Byun, Y.C.; Lee, S.J.; Park, N. Machine learning based hybrid system for imputation and efficient energy demand forecasting. Energies 2020, 13, 2681. [CrossRef]
7. Li, C.; Tao, Y.; Ao, W.; Yang, S.; Bai, Y. Improving forecasting accuracy of daily enterprise electricity consumption using a random forest based on ensemble empirical mode decomposition. Energy 2018, 165, 1220–1227. [CrossRef]
8. Grolinger, K.; L'Heureux, A.; Capretz, M.A.M.; Seewald, L. Energy forecasting for event venues: Big data and prediction accuracy. Energy Build. 2016, 112, 222–233. [CrossRef].
9. Hochreiter, S. Untersuchungen zu Dynamischen Neuronalen Netzen. Diploma Thesis, Technische Universität München, München, Germany, 1991.
10. Menezes, J.M.P.; Barreto, G.A. Long-term time series prediction with the NARX network: An empirical evaluation. Neurocomputing 2008, 71, 3335–3343. [CrossRef]
11. Census Profile, 2016 Census—Bruce, County [Census Division], Ontario and Newfoundland and Labrador [Province], (n.d.).
12. Data Directory, (n.d.). Available online: https://www.ieso.ca/en/Power-Data/Data-Directory (accessed on 8 December 2020).
13. Lahouar, A.; Ben Hadj Slama, J. Day-ahead load forecast using random forest and expert input selection. Energy Convers. Manag. 2015, 103, 1040–1051. [CrossRef]
14. Cheng, Y.; Xu, C.; Mashima, D.; Thing, V.L.L.; Wu, Y. PowerLSTM: Power Demand Forecasting Using Long Short-Term Memory Neural Network. In Proceedings of the Advanced Data Mining and Applications: 13th International Conference, Singapore, 5–6 November 2017; Springer: Berlin/Heidelberg, Germany, 2017; Available online: https://link.springer.com/chapter/10.1007/978-3-319-69179-4_51 (accessed on 5 April 2020).