Feature Selection Strategy for Short Term Electricity Consumption Forecasting

Marcia Hon
Engineering Applied Science and Management
Toronto Metropolitan University
Toronto, Canada
marcia.hon.29@torontomu.ca

Dr Bala Venkatesh
Electrical Engineering
Toronto Metropolitan University
Toronto, Canada
bala@torontomu.ca

Dr Naimul Khan
Electrical Engineering
Toronto Metropolitan University
Toronto, Ontario
n77khan@torontomu.ca

***Abstract*—There are many variables involved in Electricity Consumption Forecasting. Variables include the actual Electricity Consumption, Date Information, Weather Information, and a combination of the preceding. This paper seeks to determine what are the minimum necessary number of features and processing time for an effective forecasting. Electricity forecasting has been done in the literature. However, the features are usually picked ad-hoc. Here, Pearson Correlation and XAI (Explainable Artificial Intelligence) are systematically evaluated proving that Pearson Correlation and XAI – SHAP Values are good strategies for feature selection for Short Term Electricity Consumption Forecasting and, possibly good, for other forms of Time Series Short-Term Forecasting. The Deep Learning model is CNN-LSTM with data of Toronto Electricity Consumption and corresponding Toronto Weather. The Forecasting is done for the timespans of one day (24 hours), three days (72 hours), one week (168 hours), and one month (720 hours). Here, it is shown that Correlation and XAI – SHAP Values perform optimally for this task.**

Keywords —Time Series, Short-Term Electricity Consumption Forecasting, Pearson Correlation, XAI (Explainable Artificial Intelligence), SHAP Values, and CNN-LSTM.

I. Introduction

Short-Term Electricity Consumption Forecasting is a Time Series procedure that is vital to ensuring that adequate Electricity is created and provided for all consumers. By being able to accurately predict the consumption patterns, we ensure that there are no consumers who are jeopardized. Additionally, this forecasting allows for the efficient operation of generation, transmission, and distribution. From the literature review, there are many different approaches for Short-Term Forecasting. Generally, the features used are Energy, Date, and Weather. However, these are leveraged in an ad-hoc manner without systematically analyzing what features are necessary. In this paper, we seek to find a minimal number of necessary features and processing time to produce accurate Short-Term Forecasting. The techniques we compare are using all of Date and Weather data, Pearson Correlation, and XAI – SHAP Values. These techniques may be used for other forms of Short-Term Forecasting features selection.

II. Related Works

The literature investigated comprises Electricity Consumption forecasting. Forecasting produces impressive results, however, none of the papers directly addressed the feature selection strategy and what is the optimum number and type of features to employ.

In [1] they use Support Vector Regression to produce the optimum power demand forecasting model. Eight aspects of electricity demand were investigated: economic development, urbanization, industrialization, population, industrial structure, household consumption level, electricity price, and energy efficiency. Their approach does not include weather information. Plus, they used all features without determining if each one is needed.

In [2], they perform K-means clustering in order to identify customers with similar behavior. This is important because by grouping like consumption, forecasting is more specific and accurate. There are demographics of households whereby they have very different consumption profiles.

In [3], they used Non-Intrusive Load Monitoring (NILM) to determine the house occupancy, multimedia use, socioeconomic, and health profiles of residents. This is important because by breaking down the total consumption, you can better pinpoint what are the contributing factors.

In [4], LSTM was employed to produce Electricity Price Forecasting using Electricity Load. Price and Load work very much in tangent. Thus, by using Price we can determine the load and vice versa.

In [5], the Electricity Forecasting was performed using Federated Learning and Recurrent Neural Networks. The main reason for this architecture is that creating a model for every Smart Meter is costly, thus, when aggregated with Federated Learning much will be saved. Thus, our feature selection procedure should seek to have minimum processing time by having fewer features.

In [6], they explored forecasting with seven parameters: date, time, voltage, intensity, active power, reactive power, plus three categories of appliances. They learned that these were not sufficient predictors, thus, they recommend adding weather and to do further feature engineering.

In [7], they investigate weather data in electricity forecasting and conclude that it is important. In addition to weather, they also state that appliance information is important. They suggest improving upon forecasting by considering different machine learning and deep learning methodologies.

In [8], they perform forecasting in Italy. Weather is taken into consideration and found to be a significant influence on electricity consumption.

In [9], they used energy consumption data from Jiangsu Province from January 2005 to September 2019. Their models include ARIMA, LSTM, Prophet, and N-Beats. They conclude that LSTM and Prophet (both Deep Learning models) were successful and should be further investigated. We improve upon this by using the CNN-LSTM model.

In [10], they conclude that electricity consumption, along with weather, works in tandem with Gross Domestic Product (GDP) in China from 29 provinces during 2006 to 2017. Using LSTM, they prove with 90% confidence that electricity consumption can be a strong predictor of GDP.

In [11], due to lack of statistics in Netherlands, they present a model that leverages dynamic weather variables along with hourly-based electricity consumption. They found that the results of multiple regression and ANN were very close to the actual results. However, their approach involved shuffling the data rather than splitting it by time. Plus, for future investigations, they suggest also performing forecasting with weather data.

In [12], they pursue electricity consumption conservation by automatic scheduling. For example, the occupant could charge his Electric Vehicle by 7AM as well as have the laundry completed by 8PM. This schedule, with weather information, finds the optimal schedule of lowest costs to the consumer. They conclude that if 10 houses use this, there would be, on average, 10% less power usage.

In [13], they perform analysis in order to classify into demographic characteristics such as occupancy and wealth. These are important characteristics to investigate as they also determine how much energy a household will consume.

In [14], they pursue Electricity Consumption statistically. Their data consists of electricity consumption in addition to weather data. They conclude that they have good forecasting but there is no extensive explanation of why and which weather data was chosen.

In [15], they create an ANN-SVM model to predict gas and electricity that incorporates weather data. Their reasoning is that, in use, there are a lot of weather-sensitive electrical appliances such as air conditioners and heating. They find that ANN-SVM combo works better than single ANN or single SVM.

In [16], they investigate energy demand in Qatar for the year of 2012. They conclude that the most influential climate parameter is temperature due to air conditioning which is usually activated at around 22 degrees Celsius. Humidity did not contribute much to forecasting. They suggest further investigating the impacts of population, urban, economic, and industrial growth.

In [17], they perform long term electricity forecasting by investigating UK from 1989 to 2003. They combine climate related as well as socioeconomic patterns. They incorporate temperature because there is an expected 3 degrees Celsius average annual increase by 2080. Thus, they are a testament that temperature is very important.

In [18], they evaluate SHAP to determine if it is a good feature selector. Their conclusion is nuanced whereby it sometimes is good but sometimes does not perform well. They suggest more empirical studies to understand how SHAP works to see if it is appropriate for feature selection.

Overall, there is research that focuses on Day and Weather, while others include Demographics such as Social Economic Status and Appliance Information. The ones who incorporated weather cite Climate Change as a major motivator. In this paper, we seek to systematically determine which Date and Weather information are important. In all papers, there is not a clear distinction of which Day and which Weather information to keep. We seek to find the least computationally intensive number of features. The goal is to find the most effective Feature Selection Strategy for Short-Term Electricity Consumption Forecasting using the CNN-LSTM model.

III. Methodology

The appropriate data were used that incorporate electricity consumption, date information, and weather information. These were split into five datasets – Dataset 01 is Date, Dataset 02 is Weather, Dataset 03 is Date + Weather, Dataset 04 used Pearson Correlation to find the features whereas Dataset 05 used SHAP Values. In addition to the datasets, a Deep Learning model was employed consisting of CNN-LSTM, in sequence, to perform the forecasting tasks.

A. Dataset

The data leveraged is from the IESO [19] and Renewables Ninja [20]. The IESO contains hourly Energy Demand for the city of Toronto. Renewables Ninja contains the hourly weather data consisting of: Air temperature, Precipitation, Snowfall, Snow mass, Air density, Ground-level solar irradiance, Top of atmosphere solar irradiance, and Cloud cover fraction. Toronto is located at 43.6535°N and minus 79.3839°W. From this data, five datasets were created: Date, Weather, Date + Weather, Pearson Correlation, and XAI – SHAP Values. The formulas for Pearson Correlation and SHAP Values are in Table II. The following lists the different variables in each dataset. For Pearson Correlation, the cutoff was 0.10 and for SHAP Values the cutoff was 30.

Dataset 1 (DS01) : Date Information Only

1. Year
2. Month
3. Day
4. Hour
5. Holiday_Holiday
6. Holiday_Working
7. Week Day_Friday
8. Week Day_Monday
9. Week Day_Saturday
10. Week Day_Sunday
11. Week Day_Thursday
12. Week Day_Tuesday
13. Week Day_Wednesday

Dataset 2 (DS02) : Weather Information Only

1. Air temperature
2. Precipitation
3. Snowfall
4. Snow mass
5. Air density
6. Ground-level solar irradiance
7. Top of atmosphere solar irradiance
8. Cloud cover fraction

Dataset 3 (DS03) : Date and Weather Information

1. Year
2. Month
3. Day
4. Hour
5. Holiday_Holiday
6. Holiday_Working

7. Week Day_Friday
8. Week Day_Monday
9. Week Day_Saturday
10. Week Day_Sunday
11. Week Day_Thursday
12. Week Day_Tuesday
13. Week Day_Wednesday
14. Air temperature
15. Precipitation
16. Snowfall
17. Snow mass
18. Air density
19. Ground-level solar irradiance
20. Top of atmosphere solar irradiance
21. Cloud cover fraction

Dataset 4 (DS04) : Pearson Correlation Features (Fig. 1)

1. Top of atmosphere solar irradiance
2. Ground-level solar irradiance
3. Hour
4. Week Day_Sunday
5. Snow mass
6. Air temperature
7. Week Day_Saturday

Dataset 5 (DS05) : XAI - SHAP Features (Fig. 2)

1. Hour
2. Air temperature
3. Day
4. Top of atmosphere solar irradiance
5. Ground-level solar irradiance

Top of atmosphere solar irradiance	0.483509
Ground-level solar irradiance	0.445846
Hour	0.442831
Week Day_Sunday	0.172877
Snow mass	0.160746
Air temperature	0.158721
Week Day_Saturday	0.124194
Holiday_Holiday	0.096713
Holiday_Working	0.096713
Air density	0.095233
Week Day_Wednesday	0.080678
Week Day_Thursday	0.070141
Week Day_Tuesday	0.068170
Month	0.061552
Snowfall	0.049827
Year	0.044243
Week Day_Friday	0.043900
Week Day_Monday	0.033307
Precipitation	0.028876
Day	0.022281
Cloud cover fraction	0.000961

Fig. 1. Pearson Correlation

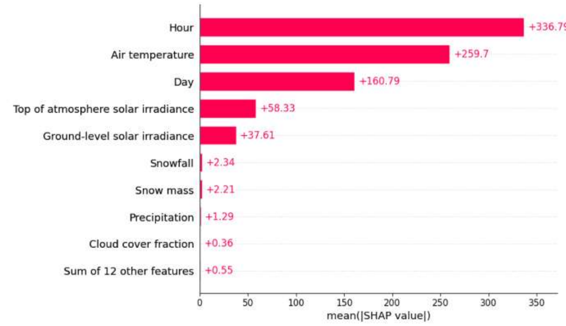


Fig. 2. SHAP Values

B. *CNN (Convolution Neural Netwok)*

CNN is a special Deep Learning architecture that has found tremendous success in image processing via the various ImageNet competitions this architecture has consistently won. Essentially, it consists of various iterations of Convolutional layers followed by Pooling layers. The Convolutional layers extract important features such as sharp edges whereas the Pooling layer consolidates the data to reduce the size. Finally, the results are sent to a Dense layer where it does the classification. The essential behavior of CNN is that it extracts important features and then determines from these features the fundamental patterns that is able to understand the images. In Fig. 3, we can see a CNN architecture with repeating convolutions followed by pooling and then ending with a dense layer that classifies the image.

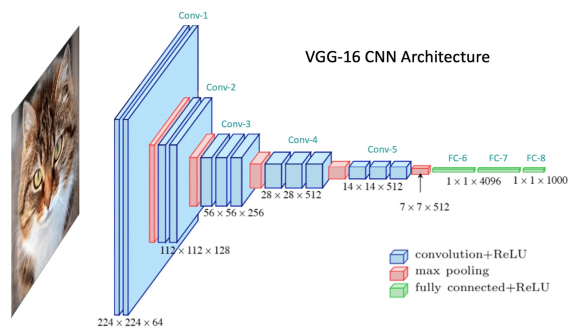
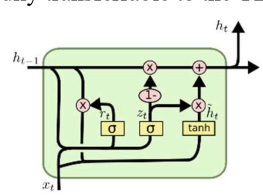


Fig. 3. Architecture for CNN, Source: Adapted from [21]

C. *LSTM (Long-Short Term Memory)*

The LSTM is the Deep Learning model leveraged here. It is a very popular and successful model that has superseded RNN (Recurrent Neural Network) type architectures due to its special way of handling past data. LSTM originally was designed to handle sequential data having its first successful applications in Natural Language Processing. Thus, it is readily transferrable to the Time Series Forecasting domain.



$$z_t = \sigma\left(W_z \cdot [h_{t-1}, x_t]\right)$$
$$r_t = \sigma\left(W_r \cdot [h_{t-1}, x_t]\right)$$
$$\tilde{h}_t = \tanh\left(W \cdot [r_t * h_{t-1}, x_t]\right)$$
$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t$$

Fig. 4. Architecture for LSTM, Source: Adapted from [22]

The LSTM, in Fig. 4, includes several gates: forget gate, cell state, input gate, and output gate. These gates ensure that LSTM can avoid gradient problems that were systemic to RNNs.

D. CNN-LSTM (Deep Learning)

In this project, we combine CNN with LSTM, thus, getting the benefits of both worlds. Essentially, CNN extracts the important features and LSTM performs sequence analysis. In Table I, the summary of the architecture of our CNN-LSTM model is presented. There is one 1D-Convolutional layer followed by two LSTM layers. Finally, the result is flattened and sent to a dense layer to process the Energy consumption.

TABLE I. ARCHITECTURE OF CNN-LSTM MODEL

Layer (type)	Output Shape	Param #
conv1d (Conv1D)	(None, 11, 32)	128
lstm (LSTM)	(None, 11, 64)	24832
lstm_1 (LSTM)	(None, 64)	33024
flatten (Flatten)	(None, 64)	0
dense (Dense)	(None, 1)	65

Total params: 58,049
Trainable params: 58,049
Non-trainable params: 0

TABLE II. FORMULAS FOR MEASUREMENTS

Measurement	Formula
MAPE: Mean Absolute Percentage Error	$\frac{1}{n}\sum_{i=1}^{n}\left\lvert\frac{Y_i-\hat{Y}_i}{Y_i}\right\rvert$
Pearson Correlation	$\frac{\sum(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum(x_i-\bar{x})^2\sum(y_i-\bar{y})^2}}$
SHAP Values	$\frac{1}{p}\sum_{S}\frac{[v(S\cup\{m\})-v(S)]}{\binom{p-1}{k(S)}}$ p – members, S – subteam, k(S) – size of S, v(S) – value by S, m – number of joins

E. Pearson Correlation

This is a ratio of the covariance between two samples with the product of their standard deviation. The goal is to measure the linear relationship between the two samples. The values range from -1 to +1 whereby -1 means a negative correlation, +1 is a positive correlation, and 0 is no correlation at all. The formula is in Table II. In this paper, we found the Pearson Correlation of the 21 variables (Date + Weather) and compared it with the Energy values. We set a cutoff of 0.10, thus, only the features above 0.10 are incorporated into the Dataset 04. Thus, we reduced from 21 to 7 features in total.

F. XAI – SHAP Values

XAI (Explainable AI) is a methodology that was created to better explain how AI works. SHAP (Shapley Additive explanation) is one such formula to accomplish this. It is based on Shapley Values which is a Game Theory concept. A SHAP Value is an amount that determines how much a certain feature contributes to the total outcome. The formula is in Table II. Here, we set a cutoff was +30.00, thus, making Dataset 05 with only 5 features instead of the total 21.

G. Training

All datasets used Toronto + Weather data for the full 2019 year as the training data. The test data was for the full 2020 2020, and we considered only test data for the following timelines: one day (24 hours), three days (72 hours), one week (168 hours), and one month (720 hours). Then, the MAPE scores for these timelines were averaged.

IV. EXPERIMENTAL RESULTS

The overall processing was computationally intensive. Accordingly, Digital Research Alliance of Canada was used. It is a free service for academic institutes across Canada. The CPUs used were 2 x Intel Silver 4216 Cascade Lake @ 2.1GHz and the GPUs were 4 x NVIDIA V100 Volta (32G HBM2 memory). The Python was version 3.9 on Jupyter Lab. ChatGPT was consulted.

The five datasets used the CNN-LSTM model at the intervals of one day (24 hours), three days (72 hours), one week (168 hours), and one month (720 hours). The following figures (5 to 9) are the results for each timeframe comparing the Actual with the Predicted energy consumption. Visually, we can observe that all the results are comparable, but it seems that DS04 and DS05 perform the best whereby the Actual and Predicted curves overlap each other the best.

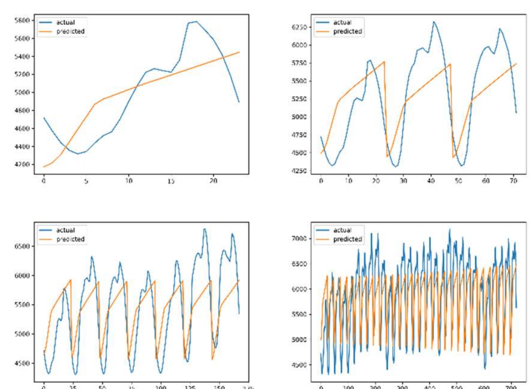


Fig. 5. DS01 – Actual vs Predicted – Hours: 24, 72, 168, 720

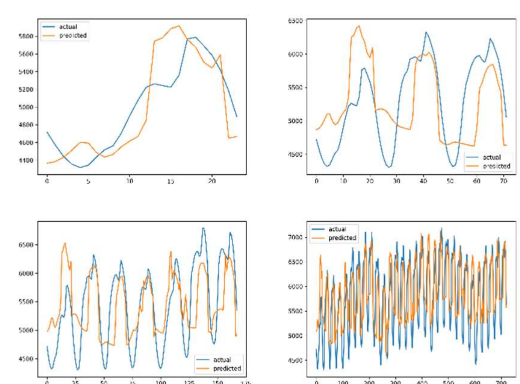


Fig. 6. DS02 – Actual vs Predicted – Hours: 24, 72, 168, 720

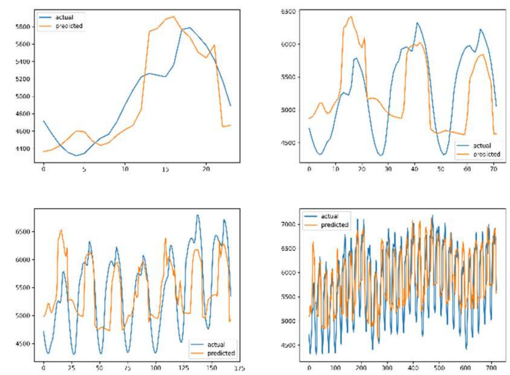


Fig. 7. DS03 – Actual vs Predicted – Hours: 24, 72, 168, 720

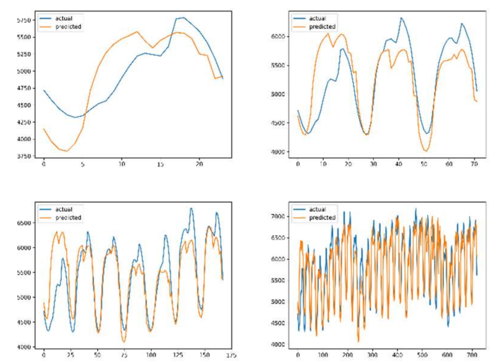


Fig. 8. DS04 – Actual vs Predicted – Hours: 24, 72, 168, 720

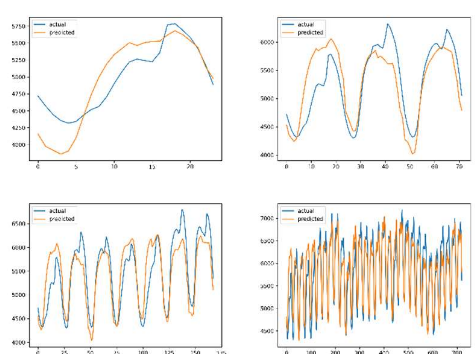


Fig. 9. DS05 – Actual vs Predicted – Hours: 24, 72, 168, 720

According to the MAPE results (Table III), DS01 and DS02 performed the worst, which is expected as Date and Weather are both needed together for Energy Consumption. We can see that for 24 hours DS03 was the best, for 72 it was DS05, and for 168 and 720 it was DS04. However. Across the board the values are comparable with a 3% difference between the maximum and the minimum MAPE scores for any given time. Regarding execution time, DS05 performed the fastest with DS03 performing the slowest. This difference is significant, being 7 minutes. Additionally, with regards to calculating the Pearson Correlation and the SHAP Values, Correlation took less than one second whereas SHAP Values took 1 hour and 40 minutes.

TABLE III. MAPE RESULTS

Hour	DS01	DS02	DS03	DS04	DS05	Max-Min
24	5.430%	5.216%	4.300%	7.496%	5.665%	3.196%
72	7.599%	9.416%	8.844%	6.473%	5.912%	3.504%
168	7.967%	8.661%	8.726%	5.387%	5.759%	3.339%
720	7.708%	7.790%	7.725%	3.799%	4.619%	3.994%
Training Time	00:04:20	00:02:40	00:06:52	00:02:18	00:01:42	00:06:52

V. DISCUSSIONS

In this paper, CNN-LSTM was the forecasting model utilized that compared the performance of four different datasets with features for Short-Term Electricity Consumption Forecasting. The four datasets were: Date, Weather, Date + Weather, Pearson Correlation, and SHAP Values. Alone, DS01 and DS02 performed inadequately with DS03, DS04, and DS05 performing relatively well. Across board, the differences between the maximum and minimum MAPE scores were insignificant meaning that choosing any of 3, 4, and 5 would be comparable and acceptable. DS03 required 21 variables whereas DS04 used 7 and DS05 used 5. Plus, the running time was fastest for DS05 followed by DS04. It took a significant amount of time, more than one hour, to determine the SHAP Values whereas Pearson Correlation took less than 1 second. Additionally, this paper shows that XAI - SHAP Values can be used to successfully determine the minimum necessary features for Short-Term Load Forecasting. Thus, XAI – SHAP Value is effective as compared to Pearson Correlation and compared to using all features. Overall, SHAP Value was able to perform very well and the fastest with the least number of features.

VI. CONCLUSIONS

This paper proves that it is possible to have excellent Short-Term Electricity Consumption Forecasting using CNN-LSTM by having less than the total features of Date + Weather (21 total features). Using Pearson Correlation and XAI – SHAP Values, the total features were reduced to 7 and 5 respectively with excellent performance even beating the performance that includes all the Date + Weather data. Additionally, although SHAP Values took the longest to calculate as compared to Pearson Correlation, it was the fastest to train once all the variables were known. Given all the results, for this experiment, it has been determined that Pearson Correlation and SHAP Values are the best feature selection strategies for Short-Term Electricity Forecasting. SHAP, being a new concept, requires further investigation. Additionally, Date + Weather, Pearson Correlation, and XAI – SHAP perform better than just Date or just Weather Information. Additionally, our feature selection strategy is simple and performs very well.

Further research includes determining if other Deep Learning models perform well such as Transformers or Deep Reinforcement Learning. Additionally, it would be interesting to determine how other features may be strategically applied such as Demographics and Appliance Information.

Keeping up with the spirit of reproducible research, all our models, dataset, and code can be accessed through the repository at: https://github.com/marciahon29/Feature-

Selection-Strategy-for-Short-Term-Electricity-Consumption-Forecasting .

Acknowledgments

We would like to thank our families and Toronto Metropolitan University for their on-going support.

References

[1] L. Liu, J. Yang, and W. Chen, "Electricity Consumption Forecasting in Jiangsu Province Based on Machine Learning Model," no. 2, pp. 613–616, 2023, doi: 10.1109/iaecst57965.2022.10061984.

[2] G. E. Okereke, M. C. Bali, C. N. Okwueze, E. C. Ukekwe, S. C. Echezona, and C. I. Ugwu, "K-means clustering of electricity consumers using time-domain features from smart meter data," *J. Electr. Syst. Inf. Technol.*, vol. 10, no. 1, 2023, doi: 10.1186/s43067-023-00068-3.

[3] P. A. Schirmer and I. Mporas, "On the non-intrusive extraction of residents' privacy- and security-sensitive information from energy smart meters," *Neural Comput. Appl.*, vol. 35, no. 1, pp. 119–132, 2023, doi: 10.1007/s00521-020-05608-w.

[4] B. Wang, W. Wei, and W. Su, "Short-term Electricity Price Forecasting Based on Data Mining," *2022 2nd Int. Conf. Algorithms, High Perform. Comput. Artif. Intell. AHPCAI 2022*, pp. 394–397, 2022, doi: 10.1109/AHPCAI57455.2022.10087616.

[5] M. N. Fekri, K. Grolinger, and S. Mir, "Distributed load forecasting using smart meter data: Federated learning with Recurrent Neural Networks," *Int. J. Electr. Power Energy Syst.*, vol. 137, no. January, pp. 2006–2007, 2022, doi: 10.1016/j.ijepes.2021.107669.

[6] J. Gaboitaolelwe, A. M. Zungeru, A. Yahya, and C. K. Lebekwe, "Electricity Load Prediction Using Machine Learning," *2022 Int. Conf. Smart Appl. Commun. Networking, SmartNets 2022*, pp. 22–25, 2022, doi: 10.1109/SmartNets55823.2022.9993990.

[7] G. Ben Brahim, "Weather Conditions Impact on Electricity Consumption in Smart Homes: Machine Learning Based Prediction Model," *2021 8th Int. Conf. Electr. Electron. Eng. ICEEE 2021*, no. 1, pp. 93–98, 2021, doi: 10.1109/ICEEE52452.2021.9415917.

[8] M. Contu *et al.*, "A weather temperature methodology on the Italian electricity demand," *2021 AEIT Int. Annu. Conf. AEIT 2021*, 2021, doi: 10.23919/AEIT53387.2021.9626983.

[9] Q. Gao, Y. Liu, J. Yang, and Y. Hong, "Comparative Research on Electricity Consumption Forecast Based on Deep Learning," *Proc. - 2021 2nd Int. Conf. Artif. Intell. Educ. ICAIE 2021*, pp. 213–217, 2021, doi: 10.1109/ICAIE53562.2021.00052.

[10] X. Wu, H. Chang, J. Li, Z. Zhang, and Q. Huang, "Electricity Consumption and Weather Reflect Macro-Economic Status," *2019 IEEE PES Innov. Smart Grid Technol. Asia, ISGT 2019*, pp. 3729–3734, 2019, doi: 10.1109/ISGT-Asia.2019.8881168.

[11] A. Prabakar, L. Wu, L. Zwanepol, N. Van Velzen, and D. Djairam, "Applying Machine Learning to Study the Relationship between Electricity Consumption and Weather Variables Using Open Data," *Proc. - 2018 IEEE PES Innov. Smart Grid Technol. Conf. Eur. ISGT-Europe 2018*, 2018, doi: 10.1109/ISGTEurope.2018.8571430.

[12] T. Ishtiak, R. M. Orpon, N. Mashnoor, M. Ahmed, and M. A. Nazim, "An advanced application to decrease household power consumption and save energy detecting the weather condition," *2017 8th IEEE Annu. Inf. Technol. Electron. Mob. Commun. Conf. IEMCON 2017*, pp. 622–627, 2017, doi: 10.1109/IEMCON.2017.8117168.

[13] M. Sun, I. Konstantelos, and G. Strbac, "Analysis of diversified residential demand in London using smart meter and demographic data," *IEEE Power Energy Soc. Gen. Meet.*, vol. 2016-Novem, no. 1, 2016, doi: 10.1109/PESGM.2016.7741076.

[14] A. Sahebalam, S. Beheshti, W. Khreich, and E. W. Nidoy, "A novel approach in household electricity consumption forecasting," *Can. Conf. Electr. Comput. Eng.*, vol. 2016-Octob, pp. 4–7, 2016, doi: 10.1109/CCECE.2016.7726768.

[15] Q. Zeng *et al.*, "An optimum regression approach for analyzing weather influence on the energy consumption," *2016 Int. Conf. Probabilistic Methods Appl. to Power Syst. PMAPS 2016 - Proc.*, no. 51325702, 2016, doi: 10.1109/PMAPS.2016.7764178.

[16] A. Gastli, Y. Charabi, R. A. Alammari, and A. M. Al-Ali, "Correlation between climate data and maximum electricity demand in Qatar," *2013 7th IEEE GCC Conf. Exhib. GCC 2013*, pp. 565–570, 2013, doi: 10.1109/IEEEGCC.2013.6705841.

[17] C. L. Hor, S. J. Watson, and S. Majithia, "Analyzing the impact of weather variables on monthly electricity demand," *IEEE Trans. Power Syst.*, vol. 20, no. 4, pp. 2078–2085, 2005, doi: 10.1109/TPWRS.2005.857397.

[18] D. Fryer, I. Strumke, and H. Nguyen, "Shapley Values for Feature Selection: The Good, the Bad, and the Axioms," *IEEE Access*, vol. 9, 2021, doi: 10.1109/ACCESS.2021.3119110.

[19] IESO, "IESO - DemandZonal Data," 2023. http://reports.ieso.ca/public/.

[20] "Renewables.ninja," 2023. https://www.renewables.ninja/.

[21] LearnOpenCV, "Understanding Convolutional Neural Networks (CNNs) : A Complete Guide." https://learnopencv.com/understanding-convolutional-neural-networks-cnn/.

[22] C. Blog, "Understanding LSTM Networks," 2015. http://colah.github.io/posts/2015-08-Understanding-LSTMs/.