Forecasting Ontario's Electrical Demand Using Machine Learning
ECSE 552 - Deep Learning
Final Report

Group 2
Kuan Wei (260569853) and **Lucas Crea** (260607090)
and **Manuel Sage** (260843979) and **Jiarui Xie** (260961962)

Abstract

Motivation: To maximize the efficiency of electrical grids, overproduction of electricity must be avoided as any excess is discarded. As such, energy suppliers must be able to properly forecast the electrical demand to adjust power generation accordingly. To this end, a search of the most appropriate method to forecast electrical demand is needed. Data from the province of Ontario was used as a case study.
Results: Various machine learning architectures were studied with the ability to forecast the demand for the next hour or 24 hours. The experiments show that LSTM and GRU architectures preformed better with the task of predicting the next hour demand. When it came to predicting the next 24 hours, it was determined that FCNN models were superior.

1 Introduction

During the past ten years, the emission-intensive energy sources such as coal plants have been rapidly phased out across Ontario for greener technologies and smart grids (Weis and Partington, 2011). Although the sustainability of energy generation methods in Canada has been significantly enhanced, renewable energy sources are less controllable and thus, the flexibility to quickly adapt to power demand is lost. Further, energy cannot be efficiently stored in large amount and if generated more than needed, the superfluous energy would be discarded. This would cause considerable waste for energy suppliers and distributors. However, the problem can be remedied by accurately forecasting the electricity demand and adjusting the power generation plans accordingly in advance, on both short-term and long-term bases (Cho et al., 2013). This way, both sides of the market benefit: producers profit from increased plannability by allowing them to reduce (often coal or gas-fired) overcapacities while customers can employ off-peak consumption strategies to avoid high electricity prices. Therefore, many green initiatives proposed by the Ontario energy suppliers and government authorities are related to electricity demand forecasting (Raza and Khosravi, 2015).

The most powerful and popular models for electricity demand forecasting are econometric methods and computational intelligence (CI). Econometric methods combine economic theories and statistical models, such as autoregressive integrated moving average (ARIMA) and partial least square regression (PLSR). They contribute to many demand forecasting applications because the electricity consumption is correlated with economic growth (Lin and Ouyang, 2014). For instance, de Assis Cabral et al. (2017) found that the regional power consumption rates in Brazil are spatially interdependent. They derived the spatiotemporal relationship of electricity consumption in Brazil and used the spatial ARIMA model to forecast demand. The spatial ARIMA model outperformed the traditional ARIMA model and achieved a mean absolute percentage error (MAPE) of 1.85%. PLSR is often used to deal with missing data. Zhang et al. (2009) utilized PLSR to predict the yearly transport energy demand in China and obtained an MAPE of 2.6% with the historical data of gross domestic product (GDP), urbanization rate, passenger turnover and freight turnover. However, the statistical models are incapable of incorporating numerous variables and extracting useful information (Porteiro et al., 2020). This disadvantage becomes noticeable as there exist more data sources nowadays to aid electricity demand forecasting.

Computational intelligence, especially machine learning (ML), are suitable for complicated prediction tasks that have large datasets and many variables. Researchers usually start with shallow ML algorithms to forecast the electricity demand. Fiot and Dinuzzo (2016) created a multi-task learning model with kernel machines to predict the electricity consumption of multiple locations in Ireland. The authors derived kernels that specifically extract the seasonal patterns from the dataset and obtained an MAPE of 4% for day-ahead prediction. Porteiro et al. (2020) used an ensemble method called extra tree regression to forecast the electricity demand in Uruguay. They achieved an MAPE of 5.17% for the load forecasting in 24 hours. Clearly, the performance of shallow ML models is not comparable to econometric models for many cases. Thus, many researchers resort to deep learning (DL), which can catch more complicated and recurrent patterns. Kuo and Huang (2018) built three DL models that predict the next hour electricity consumption in Texas, including fully connected neural networks (FCNN), long-short term memory (LSTM) and convolutional neural networks (CNN). They also constructed three forecasting models with traditional ML algorithms (support vector machine (SVM), random forest (RF) and decision tree (DT)) to compare with the DL models. With the history of electricity consumption as the only input, the best MAPE obtained was 9.8% with CNN, 10.87% with RF, and 11.33% with LSTM. Ugurlu et al. (2018) constructed a three-layer gated recurrent units (GRU) to forecast the electricity price in Turkey, incorporating the information of historical price, demand, supply and temperature. From the predictions of one-hour ahead to six-hour ahead, the MAPE increased from 9.75% to 20.03%. With a bi-directional LSTM, more advanced DL structure was constructed by Mughees et al. (2021) to predict the next day peak demand in Pakistan. They trained the model with the historical electricity consumption and national holidays and compared the LSTM models with SVM and artificial neural network (ANN) models. The best MAPEs they obtained were 4.74% for normal days and 13.34% for holidays with SVM, despite the sophisticated structure and high computational cost of LSTM models. From the above it was determined that the strengths of DL, especially recurrent neural networks (RNN), have not been fully realized in the tasks of electricity load forecasting. In addition, it is clear that there is a need to find the best practices regarding which data sources to be incorporated and which preprocessing steps to utilize.

With the availability of Ontario's historical electricity demand provided by the Independent Electricity System Operator (IESO), this research project aims to construct an Ontario electricity demand forecasting framework with four goals: 1) to train machine learning models that accurately predict the future electricity demands; 2) to incorporate auxiliary information such as temperature and date; 3) to investigate the impact of time scale difference between 1h and 24h forecasting; and 4) to compare among some selected shallow machine learning and deep learning algorithms, such as SVM, RF, FCNN, LSTM and GRU.

2 Methodology

2.1 Dataset

A new dataset was created for demand forecasting for the province of Ontario, Canada. Ontario's historic hourly electricity demand (in MW) was obtained from IESO[1] for the years 2017 to 2020. Besides the time-related information included in this data, two features with the potential of facilitating forecasts were added: holidays and air temperature. The holiday information was believed to be relevant as reduced industrial usages may occur during these days. Air temperature was found to be relevant as increased electricity demand was observed in both hot or cold days (Figure S1). Statutory holidays were added as a binary column to the dataset, where 1 represents holidays and 0 stands for regular days. Hourly averaged air temperature values (in °C) were downloaded[2] for six weather stations in these major population centers across the province: Hamilton, Kitchener, London, Ottawa, Toronto and Windsor. An analysis shows closely related temperatures over the stations, with Ottawa being slightly colder than the other cities (Figure S2). To keep the dataset consistent at a province-wide level, a single temperature feature was composed by computing the weighted average of the six weather stations with respect to the cities' population sizes.

The final dataset comprises 35,064 hourly values for the years 2017-2020 for these features:

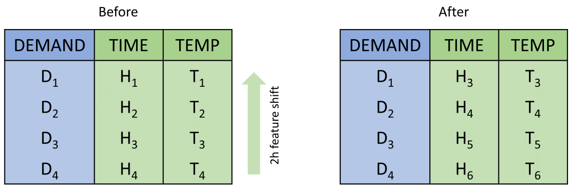
- Ontario electricity demand (MW)
- Hour
- Day of month
- Day of week
- Month
- Holiday
- Average Ontario air temperature (°C)

The first 2.5 years of this data form the training set for all implemented models. The following 6 months were utilized as a validation set to evaluate different preprocessing techniques and tune model-specific hyperparameters. The test set consists of the last year's data and was solely used to report the final performances for each model and task.

2.2 Preprocessing

As a next step, several preprocessing methods were included into the project pipeline that are tested during model training regarding their effect on each model's performance:

- **One-hot encoding:** One-hot-encoding was experimented with for the day of week and month features in the dataset.
- **Sine-cosine encoding:** Computing autocorrelation and discrete Fourier transforms for the demand time-series shows that the most significant frequencies embedded in the demand have wavelengths of 12h, 24h, one week and two years (Figure S3). These frequencies were encoded each by a sine and cosine wave (London, 2016) and all other time-related features except holidays were dropped. With this approach, the notion of similarity in periodical features were accounted for. For example, 4pm - 5pm have the same distance as 23pm – 0am.
- **Normalization:** For the demand and temperature features the effects of standardization (zero mean and unit variance) and min-max scaling to a range of (0, 1) were examined.
- **Feature shift:** In a separate set of experiments, the information related to date, time and temperature were shifted backwards relative to the demand values in order to allow the models to use future information for predictions (see Figure 1 for a visualization). While this approach is unproblematic for date and time, it raises the concern whether it is appropriate to include temperature measurements rather than temperature forecasts in the time shift. It was hypothesized that weather forecasts for the short prediction intervals examined are fairly accurate and averaging over an entire province significantly reduces potential noise in weather models. In addition, historic hourly weather forecasts for Ontario were not freely available online.



Before

DEMAND	TIME	TEMP
D_1	H_1	T_1
D_2	H_2	T_2
D_3	H_3	T_3
D_4	H_4	T_4

After

DEMAND	TIME	TEMP
D_1	H_3	T_3
D_2	H_4	T_4
D_3	H_5	T_5
D_4	H_6	T_6

Figure 1: Example for a feature shift by 2 hours.

[1] https://www.ieso.ca/en/Power-Data/Data-Directory

[2] https://climate.weather.gc.ca

> The investigation was conducted to determine if a feature shift improves predictions, as this would indicate that demand forecasts using machine learning algorithms profit from weather forecasts.

The last step of the data preparation and preprocessing pipeline is sample creation with a sliding window. The sliding window defines how many subsequent hourly values form the input and how many hours following the last input define the output of each sample. Output sequences of length 1 and 24 were used for next hour and next 24 hours demand predictions. The length of the input sequence was treated as a hyperparameter and tested for each model on the validation set. The tested span ranges from 1 to 96 hours. The recurrent networks can retain and process the input sequences produced by the sliding window. For the other models, each sample's input was flattened to a single vector.

2.3 Model Implementation

To begin, two baselines for the 1h and the 24h predictions were established. For 1h forecasts, a persistence model naively used the demand of the current hour as prediction for the next hour, i.e. $D_t = D_{(t+1)}$. For 24h forecasts, the previous 24h sequence was copied and utilized as a prediction for the next 24 values, i.e. $\{D_{(t-23)}, \ldots, D_t\} = \{D_{(t+1)}, \ldots, D_{(t+24)}\}$.

For all configurations of model and predictive tasks, the data preprocessing techniques described above were tested with regards to their effect on validation set performance which was measured via root-mean-squared error (RMSE). Furthermore, the model specific hyperparameters were tuned. For random forest and linear SVR, this was implemented with random searches of 100 iterations. In both cases, the identified best configuration did not outperform scikit-learn's default setting of the models and was hence not used for test set evaluation. The tested parameter spans defined as well as the best configurations found are represented in Table S1 and S2.

The deep learning models FCNN, LSTM and GRU were implemented with Adam as the optimizer and utilized mean-squared error (MSE) as the loss function. The initial learning rate was treated as a tuneable hyperparameter and a scheduler was deployed to adjust the learning rate dynamically during training. Early stopping with a patience of 5 epochs was used as a regularization technique and to shorten training time. For GRU and LSTM, the input was fed into at least one layer of recurrent units. The FCNN consist of a fully connected input layer with ReLU activations, followed by a varying number of hidden layers with the same activation function. All three models end with a fully connected layer without activation function. The number of hidden layers and the number of units per hidden layer were treated as hyperparameters for all three models. Additionally, different values for batch size and dropout regularization were also tested.

The first stage of tuning experiments was conducted manually. It was determined that networks with a single hidden layer yield best results and that the most influential hyperparameter were the number of units in the hidden layer and the length of the input sequence. The best values for these two parameters were then identified in a grid search for each of the tasks in question. Details about the tested configurations during hyperparameter tuning can be found in Table S2 in the appendix.

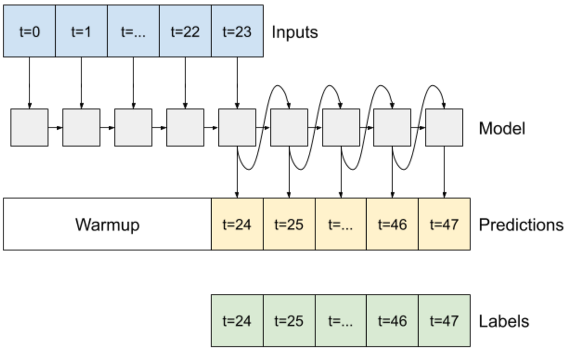


Figure 2: Graphical Illustration of AutoRegressive Model in GRU and LSTM (Google, 2021). During the "Warmup" stage, the model uses the input features to adjust the weights of LSTM/GRU units shown as grey blocks. After 24 hours, the model uses the electricity demand predicted from the previous time step, combined with the other features to predict the electricity demand of the next time step.

Finally, an auto-regressive (AR) approach was implemented for 24h predictions as comparison to 24h sequence predictions. The AR model iteratively predicts the next hours demand which is then fed with the feature values (time and temperature) of the next hour together with its own prediction to make a new prediction for the following hour. This process is repeated 24 times to obtain a sequence of predictions for the next 24h. For RF and FCNN, the best model on 1h prediction was deployed for the AR task. For GRU and LSTM, new models were trained to optimize the loss over the entire produced sequence (Figure 2). With information from future time-steps being used in the AR models, this approach was evaluated under the same circumstances as the sequence predictions with feature shift.

2.4 Performance Evaluation

To compare the different model architectures, Root Mean Squared Error (RMSE) and Mean Absolute Percentage Error (MAPE) were used as performance metrics. The RMSE is defined by the following equation:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} (y_{true} - y_{pred})^2}{N}},$$

where N is the total number of samples, y_{true} is the true target value, and y_{pred} is the predicted target value. The MAPE is defined as follow:

$$MAPE = \frac{1}{n} \sum_{i=1}^{N} \frac{(y_{true} - y_{pred})}{y_{pred}}.$$

To obtain statistically meaningful results, each model was trained 5 times and the mean performance was assessed on the test set. T-tests were used when comparing different performances to determine whether the differences were statistically significant. A P value of 0.05 was chosen as the threshold for T-tests.

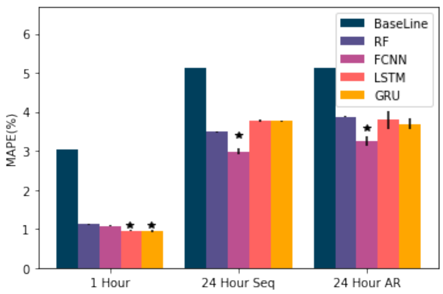


Figure 3: Performance comparison between different models. The bar heights represent the means and the black lines on the top of the bars represent the standard deviation. The "⋆" sign indicates the corresponding bar is significantly less than its counterpart (P < 0.05)

3 Results & Discussion

3.1 1 Hour and 24 Hour Demand Predictions

As distinct models are used in this project to predict the electricity demand, the hyperparameters of the models were optimized separately and the resulting performances were compared. Great improvements in prediction accuracy can be found using machine learning models compared to using the naive baseline persistence model (Figure 3). The linear SVR performed significantly worse than the other four models on 1h predictions (larger than 1.5% MAPE). Therefore, it was excluded from the remaining experiments and the results are not presented. As shown in the figure, LSTM and GRU architectures preformed similarly and achieve better results on the 1h forecast, with a same MAPE of 0.95%. Surprisingly, FCNN outperforms significantly for 24h predictions, with a MAPE of 2.99% for sequence predictions and 3.24% for AR predictions. These observations deviate from the common understanding that LSTM and GRU perform better on longer sequence predictions. One potential reason is that the electricity demand prediction in the context of this project is a relatively simple task. The number of input features is small enough that a FCNN or even a RF can be trained within a reasonable amount of time. Since there is a greater number of parameters in FCNN than in LSTM or GRU, it can study and capture more information from the input features. However, one critical drawback of FCNN is that once the number of input features becomes large, the number of samples required to effectively train a FCNN becomes exponentially higher. For those cases, it is expected that LSTM and GRU will out preform FCNN. However, the result also suggests that blindly believing in recurrent neural networks such as LSTM or GRU may lead to sub-optimal performance for simpler tasks.

Figures S4 and S5 in the appendix graphically illustrate how well the LSTM model forecasts the demand 1 hour ahead and how well the FCNN model forecasts the next 24 hour demand, respectively.

When comparing sequence predictions and AR predictions, it was observed that the means are very similar but the AR predictions have larger standard deviations. The major difference between the two techniques is that models trained for sequence predictions only utilize the input features, while models trained for AR predictions reuse its own predicted electricity demands. The predicted values may never be exactly the same as the true values, and as a result, using predicted values can introduce additional variance. This variance will propagate for future predictions and eventually cause a larger deviation. However, despite that AR models did not obtain better performance than sequence models, they are much more flexible in terms of the target time range. For example, if a 48h prediction needs to be made, a new sequence model has to be trained, while the same AR model can be used. This advantage of AR models can be beneficial when the electricity demand in various ranges of time needs to be predicted.

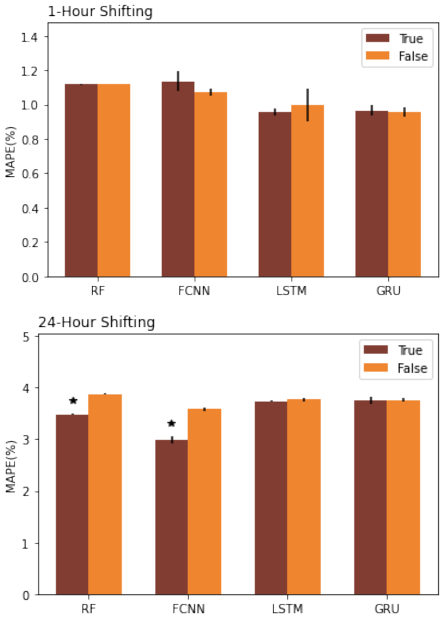


Figure 4: A study of how feature shifting affects model performances. 1h shifting is used for 1h predictions and 24h shifting is used for 24h sequence predictions. The bar heights represent the means and the black lines on the top of the bars represent the standard deviation. The "⋆" sign indicates the corresponding bar is significantly less than its counterpart (P < 0.05)

3.2 Feature Time Shifting

As feature shifting was a novel concept that was conceived for this project, it is worthwhile to study its effect individually. Figure 4 shows that feature shifting does not significantly change the model performance for 1h predictions. However, both RF and FCNN receive a performance boost using 24h feature shifting for 24h sequence predictions. Indeed, as the length of the shift period increases, the difference between feature values with and without shifting becomes more significant. For example, the temperature may vary by only 1 degree Celsius in 1 hour but can change by more than 12 degree Celsius in 24 hours. Since the feature values after shifting are more relevant to the electricity demand to be predicted, it is reasonable to believe that feature shifting can increase model performance. Another observation is that feature shifting does not have an impact on LSTM and GRU no matter the length of the shift. A plausible explanation is that both LSTM and GRU try to capture the trends of features within a period, instead of looking for connections between the exact values. Unlike

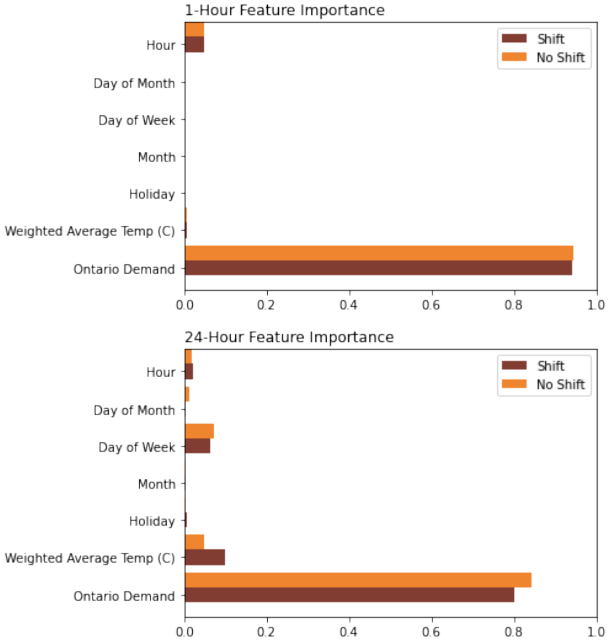


Figure 5: Tree-based feature importance plots for RF model on the test set.

the values, the trends for the features, such as time and temperature, can be similar between two consecutive days. For example, the temperature is almost always low during the night while high around noon. As a result, LSTM and GRU may capture very similar information whether the features are shifted or not.

3.3 Effects of Preprocessing Techniques

As was mentioned above, new features along with preprocessing methods were proposed in an attempt to improve the forecasting capabilities of the models in more general cases. To identify which features are useful for the models on predicting unseen data, tree-based feature importance plots were generated on the test set using RF model. As seen from Figure 5, the models rely mostly on the previous demand values for both tasks considered in this report. It was also found that the introduction of temperature data, holiday encoding or one-hot encoding of days, weeks, and months had little to no effect on the performance for predicting the demand of the next hour. The only additional feature that seemed to have an effect on the results was the "hour" information. For the task of predicting the demand for the next 24 hours, the additional features seem to have an overall greater impact on the results and can be deemed to have improved the performance. Interestingly, time and weather features became more significant when feature shifting was implemented. The phenomenon is especially obvious for 24h predictions. The observation echos the results found in Figure 4 and provides evidence to demonstrate that feature shifting does at least help the RF model to utilize time and weather information more effectively.

3.4 Test vs. Validation Set Performance

During testing, significant differences between model performances on validation set and test set were observed. As seen from Figure 6, the RF model has a much lower RMSE on the validation set. Other models also behave similarly. Usually, if the validation set and the test set have same distributions, the models should perform similarly on both datasets. A hypothesis for the below par performance on the test set is that it may be related to the way the data was split. The test data is from 2020-01-01 to 2020-12-31, a period of time during the COVID-19 pandemic. However it doesn't appear that the pandemic is the sole contributor to the discrepancy between the validation and testing losses. Table 1 shows that the data for the year 2020 is certainly different to the years prior, but the difference is not enormous. That being said, there are other potential causes for the differences such as the fact that the training and validation set were not combined to retrain the models a final time. This can lead to models with lower generalization.

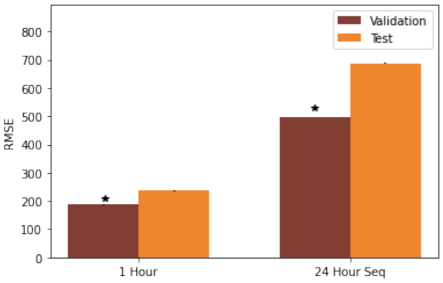


Figure 6: RMSE values for the random forest model on validation set and test set. The bar heights represent the means and the black lines on the top of the bars represent the standard deviation. The "⋆" sign indicates the corresponding bar is significantly less than its counterpart ($P < 0.05$)

	2017	2018	2019	2020
2017	0	-	-	-
2018	1848	0	-	-
2019	1795	1647	0	-
2020	1971	1973	1628	0

Table 1: RMSE Between Years

4 Conclusion

In conclusion , LSTM and GRU architectures were found to be superior to predict the electricity demand for the next hour with a MAPE of 0.95% while the FCNN architecture was demonstrated to work best for 24h predictions with MAPE of 2.99%. In addition, the results show that preprocessing step proposed and the additional features introduced were important for the 24h predictions. Overall, the observed performance of the models studied are comparable to the ones reported in existent literature with similar tasks.

References

Haeran Cho, Yannig Goude, Xavier Brossat, and Qiwei Yao. 2013. Modeling and forecasting daily electricity load curves: a hybrid approach. *Journal of the American Statistical Association*, 108(501):7–21.

Joilson de Assis Cabral, Luiz Fernando Loureiro Legey, and Maria Viviana de Freitas Cabral. 2017. Electricity consumption forecasting in brazil: A spatial econometrics approach. *Energy*, 126:124–131.

Jean-Baptiste Fiot and Francesco Dinuzzo. 2016. Electricity demand forecasting by multi-task learning. *IEEE Transactions on Smart Grid*, 9(2):544–551.

Google. 2021. Time series forecasting.

Ping-Huan Kuo and Chiou-Jye Huang. 2018. A high precision artificial neural networks model for short-term energy load forecasting. *Energies*, 11(1):213.

Boqiang Lin and Xiaoling Ouyang. 2014. Energy demand in china: Comparison of characteristics between the us and china in rapid urbanization stage. *Energy conversion and management*, 79:128–139.

Lan London. 2016. Encoding cyclical continuous features - 24-hour time.

Neelam Mughees, Syed Ali Mohsin, Abdullah Mughees, and Anam Mughees. 2021. Deep sequence to sequence bi-lstm neural networks for day-ahead peak load forecasting. *Expert Systems with Applications*, page 114844.

Rodrigo Porteiro, Luis Hernández-Callejo, and Sergio Nesmachnow. 2020. Electricity demand forecasting in industrial and residential facilities using ensemble machine learning. *Revista Facultad de Ingenieria Universidad de Antioquia*.

Muhammad Qamar Raza and Abbas Khosravi. 2015. A review on artificial intelligence based load demand forecasting techniques for smart grid and buildings. *Renewable and Sustainable Energy Reviews*, 50:1352–1372.

Umut Ugurlu, Ilkay Oksuz, and Oktay Tas. 2018. Electricity price forecasting using recurrent neural networks. *Energies*, 11(5):1255.

Tim Weis and PJ Partington. 2011. *Behind the switch: pricing Ontario electricity options*. Pembina Institute.

Ming Zhang, Hailin Mu, Gang Li, and Yadong Ning. 2009. Forecasting the transport energy demand based on plsr method in china. *Energy*, 34(9):1396–1400.

5 Appendix

Table S1: Best preprocessing parameters for each model

Model	Task	Time Encoding			Scaling			Shifting		Window Size
		None	One-Hot	Sin-Cos	None	Standard	Min-Max	True	False	1-96
SVR	1h	One-Hot			None			False		24
RF	1h	None			None			True		18
	24h sequence	None			None			True		18
	24h AR	None			None			False		18
FCNN	1h	None			Standard			False		24
	24h sequence	None			Standard			True		24
	24h AR	None			Standard			False		24
LSTM	1h	Sin-Cos			Standard			True		24
	24h sequence	Sin-Cos			Standard			True		48
	24h AR	Sin-Cos			Standard			True		24
GRU	1h	Sin-Cos			Standard			True		18
	24h sequence	Sin-Cos			Standard			True		48
	24h AR	Sin-Cos			Standard			True		48

Table S2: Tested ranges and best parameters found for each model and task

Model	Task / Tested & Best Parameters				
SVR	Parameters	tolerance	regularization param. 'C'	epsilon	max. iterations
	Values tested	1e-2, 1e-3, 1e-4, 1e-5	0.01, 0.1, 0.5, 1, 2, 5	0, 0.01, 0.1, 1	100, 1000, 10000
	1h	1e-4	1	0	1000
RF	Parameters	# of estimators	max. depth	min. samples /split	min. samples /leaf
	Values tested	10, 50, 100, 500	5, 10, 25, 50, 100, None	2, 3, 4, 5	1, 2, 3, 4
	1h	100	None	2	1
	24h sequence	500	None	2	1
	24h AR	100	None	2	1
FCNN	Parameters	batch size	# of layers	# of neurons per layer	learning rate
	Values tested	16, 32, 128	2, 4	50, 100, 200, 500, 1000	1e-2, 1e-3, 1e-4
	1h	32	2	100	1e-3
	24h sequence	32	2	100	1e-3
	24h AR	32	2	100	1e-3
LSTM	Parameters	batch size	# of layers	# of neurons per layer	learning rate
	Values tested	16, 32, 128	1, 2, 3	32, 64, 96, 128	1e-2, 1e-3, 1e-4
	1h	32	1	96	1e-3
	24h sequence	32	1	128	1e-3
	24h AR	32	1	64	1e-3
GRU	Parameters	batch size	# of layers	# of neurons per layer	learning rate
	Values tested	16, 32, 128	1, 2, 3	32, 64, 96, 128	1e-2, 1e-3, 1e-4
	1h	32	1	64	1e-3
	24h sequence	32	1	128	1e-3
	24h AR	32	1	96	1e-3

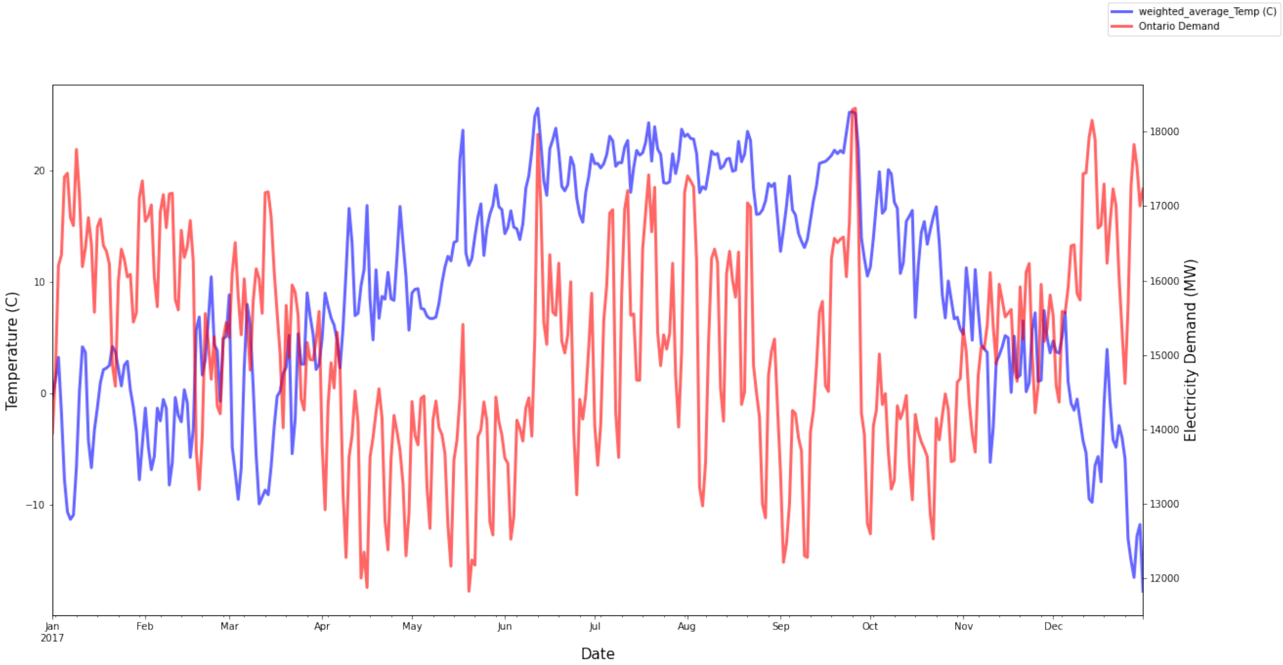


Figure S1: Daily average of weighted-average-temperature and electricity demand over time in 2017. High electricity demands are observed when the temperature drops down below 0 degrees and above 25 degrees

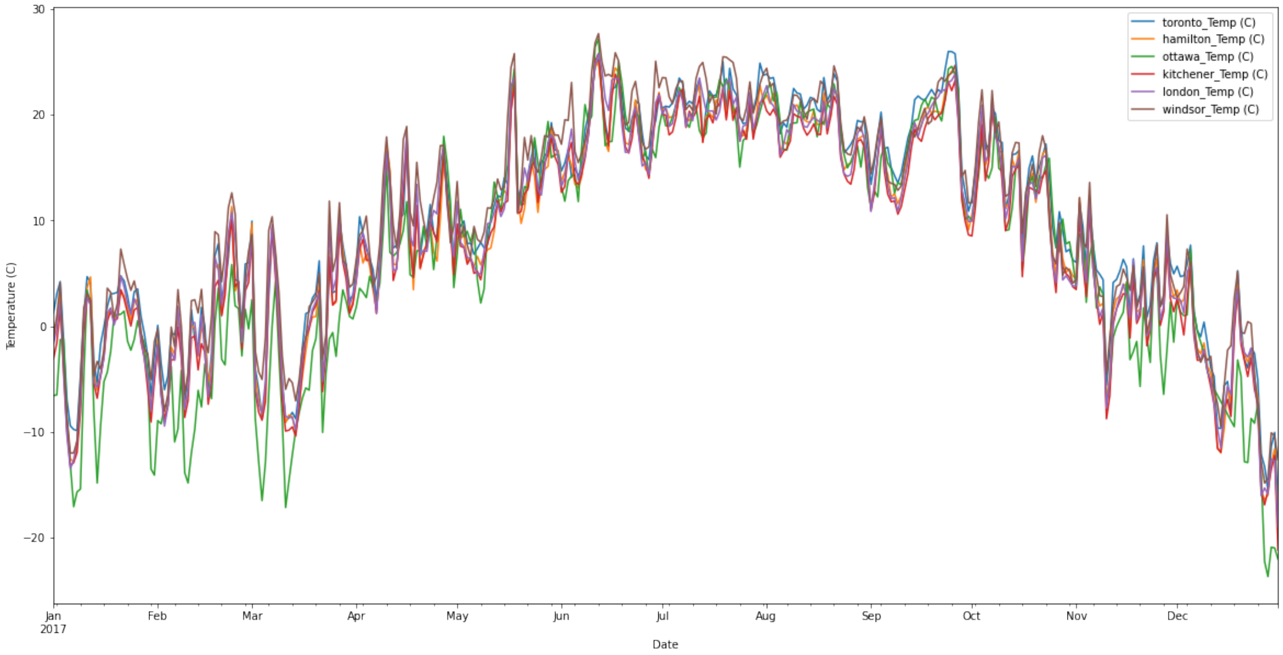


Figure S2: Daily average of temperatures over time in each city in 2017.

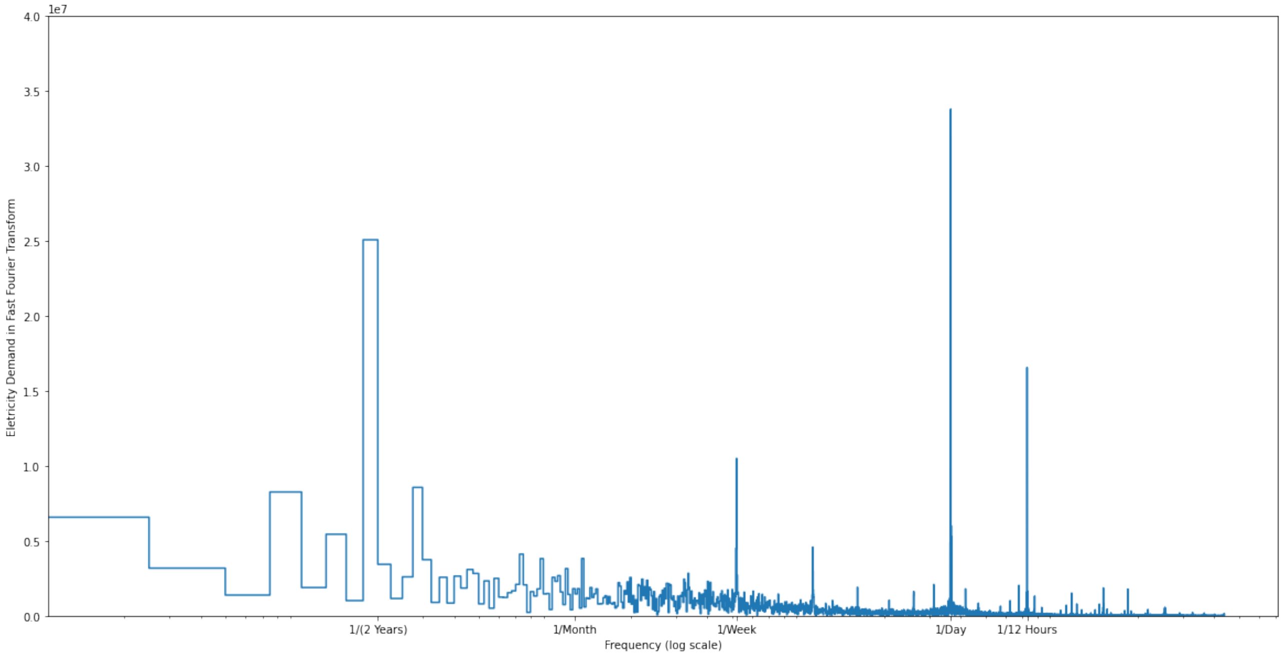


Figure S3: Autocorrelation for significant frequencies for sin-cos encoding

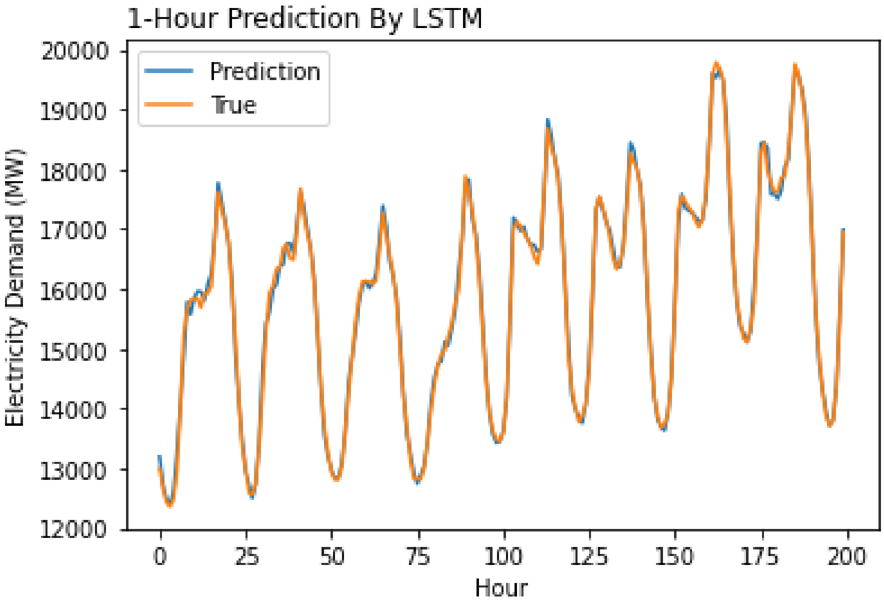


Figure S4: Sample 1-Hour Predictions For LSTM Model

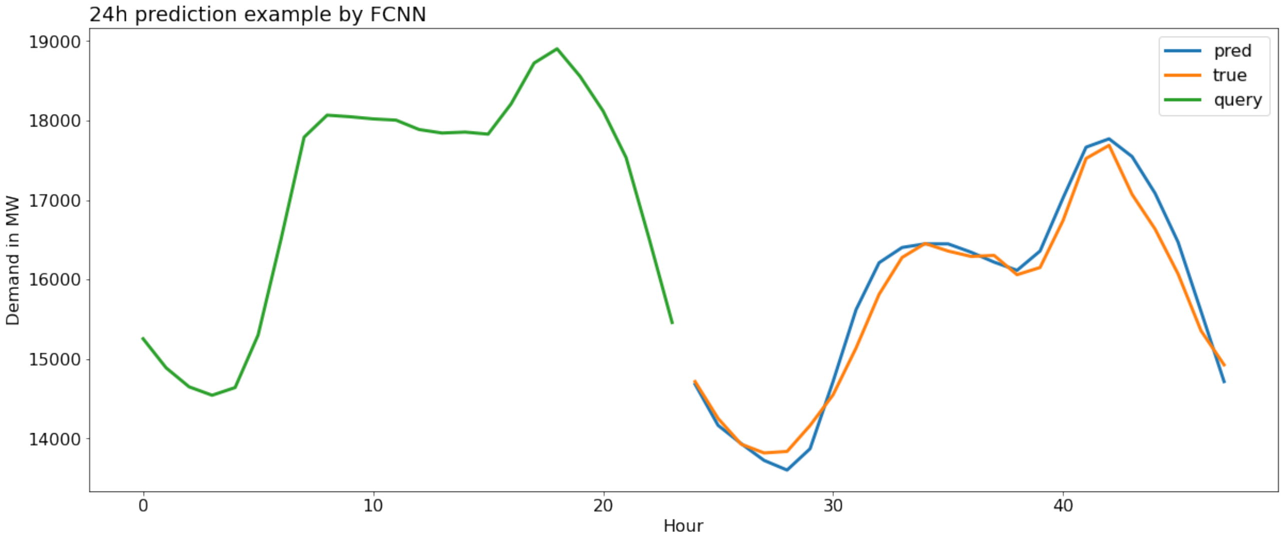


Figure S5: Sample 24-Hour Predictions For FCNN Model