8-23-2023 11:00 AM

Prediction of peak energy demand and timestamping in commercial supermarkets using deep learning

Mengchen Zhao, *Western University*

Supervisor: Capretz, Miriam A.M., *The University of Western Ontario*
Co-Supervisor: Sadhu, Ayan, *The University of Western Ontario*
A thesis submitted in partial fulfillment of the requirements for the Master of Engineering Science degree in Electrical and Computer Engineering

Abstract

Peak demand consumption is an ongoing research topic due to current environmental concerns. Accurate prediction of peak demand leads to improved power storage scheduling and smart grid management. However, existing researches on peak demand in commercial buildings lack focus on the timestamps for the incident of peak consumption. For this reason, this research proposes to label three indexes per day and use the novel Energy Peaks and Timestamping Prediction (EPTP) framework to detect the energy peaks as well as the timestamps for the occurring indexes.

The EPTP framework is proposed with three phases. In the first phase, data preprocessing cleans the raw data into the intended input for the deep learning model, and timestamp labelling creates the expected output for training and evaluation of the model. The second phase focuses on energy consumption prediction using Long Short-Term Memory (LSTM) network, which is dedicated to processing sequential data. The last phase uses Multilayer Perceptron (MLP) for the purpose of timestamp prediction. The EPTP framework is evaluated using various data resolutions and compared to the common label of using block maxima from extreme value theory. Specifically, the two-hour hit rate improves from 21% using the block maxima approach to 52.6% with the proposed EPTP framework, and from 65.3% to 86% for the 1-hour resolution and the 15-minute resolution, respectively. In addition, the average minute deviation decreases from 120 minutes using the block maxima approach to 62 minutes with the proposed EPTP framework for the high-resolution data. The framework shows adequate results from high-resolution data using real-world commercial supermarket energy consumption.

Keywords: Commercial building energy consumption, Peak demand, Timestamp prediction, LSTM.

Summary for Lay Audience

Building energy consumption takes up more than 30% of total global energy use and accounts for 27% of total greenhouse gas emissions. Peak demand management serves the purpose of effective building energy use, which in turn decreases carbon footprint. Accurate peak demand in commercial buildings benefits the suppliers by improving efficiency in electricity production; and the consumers by reducing fluctuations and energy waste. Accurately predicting energy peaks can lead to proactive peak-shaving strategies and battery response schedulings. In the existing research, most of the effort has been dedicated to predicting the intensity or the amount of energy used during peak demand. However, there has been a lack of interest in extracting the timing of peak demand, especially in commercial buildings.

In this research, the effort is dedicated to not only predicting the energy consumption during the peak hours but also the index or the timestamps for the instance of which peak has occurred. A three-phase methodology for Energy Peaks and Timestamping Prediction (EPTP) is proposed in this work. Data preprocessing and timestamp labelling create the intended input for the deep learning model. Energy consumption prediction is performed with a Long Short-Term Memory (LSTM) network, which is dedicated to processing sequential data; followed by the timestamp prediction using Multilayer Perceptron Network. The proposed methodology is validated through experiments using real-world commercial supermarket data. Results show that the 2-hour hit rate reaches 86% with 62 minutes deviation in the peaking index with 15-minute resolution data.

This thesis is dedicated to my late mother.

I miss you more than words can say.

Acknowlegements

First and foremost, I would like to express my sincere gratitude to my supervisors, Dr. Miriam Capretz and Dr. Ayan Sadhu, for their invaluable advice, continuous support, and motivation during my master's study. This research would not have been done without their guidance and patience, especially when obstacles were encountered.

I want to acknowledge the Neelands-Kalder team for providing the dataset. I want to thank Craig Zych and Hooman Nouraei for providing guidance and insights for the operation of the commercial setting. I would also like to acknowledge Mitacs Accelerate Fellowship for providing a platform for collaboration with the industrial partner.

I am sincerely grateful for my friends and colleagues who have encouraged, challenged, and inspired me. I like to extend my gratitude especially to Santiago Gomez-Rosero for giving guidance and providing meaningful feedback.

Last but not least, I would not have been able to accomplish this work without the continued support of my family. I am forever grateful for their unconditional love and unfailing support throughout my years of study.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation

Energy consumption has an attractive trend for research as supply difficulties and global warming have caused rising concerns [1, 2, 3]. Among all areas of energy consumption, buildings account for more than 30% of total global energy use and account for 27% of total greenhouse gas emissions[3, 4]. Therefore, efficiency in building energy management becomes a crucial strategy in the low-carbon economy [5, 6, 7]. Predicting power loads turns out to be a key issue to avoid energy wastage and to build effective power management blueprints [8]. Accurate predictions also benefit both the suppliers and the consumers due to improved power management, grid security, and load control [9]. In addition, accurate power load forecasting is not only the first step in dealing with load management [10] but also can act as a baseline for peak demand prediction [11].

The electrical grid structure faces challenges in balancing power flow between the production and consumption of energy. Peak demand prediction in energy smart grid management aims to improve such issues through power storage scheduling and demand flexibility. The ultimate goal of peak demand management is to balance electricity supply and demand to maximize the benefits of the power. In addition, accurate peak demand prediction help with peak shaving and load smoothing. The improved efficiency of electricity production, as a result of reduced fluctuations in power demand, will also decrease carbon footprints [12].

Peak demand prediction has gained increasing focus in various industries in load management and scheduling. Such topic in residential hours and neighborhood energy management have been studied using ensemble Long Short-Term memory (LSTM) model [13], CNN sequence-to-sequence network [14], and cluster analysis [15]. Saini [16] concluded adaptive learning backpropagation-based artificial neural networks had the best performance in forecasting peak loads in a substation. Lu et al. [17] proposed to use a hybrid complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN) for data decomposition, and extreme gradient boosting (XGBoost) for predicting energy consumption. The proposed CEEMDAN-XGBoost method is tested with daily peak power consumption used by an intake tower. On a greater scale, peak demand forecasting has also been studied with Iran's power system network [11], UK national grids [18], and South African daily electricity demand data [19, 20]. Overall, forecasting accurate peak loads is particularly important in many use cases, including network constraint management, peak shaving, and battery and demand response scheduling.

In this research, commercial supermarket dataset will be analyzed for the task of peak demand prediction. In commercial supermarkets, over 70% of the energy utilized is in the form of electricity, primarily dedicated to powering refrigeration equipment. The remaining portion of the energy is allocated to tasks such as lighting, HVAC (heating, ventilation, and air conditioning), baking, and other supplementary services [21, 22]. In such scenarios, the refrigeration equipment can be used as virtual batteries by increasing energy consumption during low-demand hours and discharging during peak hours to smooth out the electricity consumption. Therefore, accurate predictions of the peak demand as well as the time index are the key preconditions in promoting such load-shaving strategies.

1.2 Scope and Objective

The scope and objectives of this research are detailed as below:

- Conduct a detailed literature review on topics related to energy forecasting and peak demand prediction in commercial applications, and identify the challenges and gaps in existing literature,
- Understand the implementation challenges of various baseline deep learning models including Multilayer Perceptron (MLP), LSTM, and various hyperparameter optimization techniques for large-scale commercial building applications,
- Propose a framework for peak demand prediction in commercial supermarket data and analysis the performance of the framework using various experiments with real-world commercial supermarket data.

1.3 Main Contributions

The main contributions of this thesis are summarized as the following:

- The proposed Energy Peaks and Timestamping Prediction (EPTP) framework as a novel approach for commercial applications in not only predicting the value of peak energy consumptions but also the starting, peaking, and ending timestamps,
- Label the peak indexes with width and base values to avoid long durations in low-frequency data,

- Evaluation of the proposed EPTP framework using the different frequencies of energy consumption data, resulting in improved performance with more granular resolution,
- Benchmarking and baselining the proposed EPTP framework using existing open-source dataset and common indexing practice.

1.4 Thesis Outline

The outline of the thesis is organized as follows:

Chapter 2 introduces the background information and reviews existing literature. Background information covers Multilayer perceptron, Long Short-Term Memory Networks, and hyperparameter optimization techniques.

Chapter 3 presents the proposed Energy Peaks and Timestamping Prediction (EPTP) framework including its three phases.

Chapter 4 describes the experiments with the proposed EPTP framework. It presents three experiments including the performance on open-sourced dataset, baseline model analysis, and the performance of the EPTP framework with higher data frequency.

Chapter 5 concludes this thesis and discusses limitations and future work.

Chapter 2

Background and Literature Review

2.1 Introduction

This chapter provides the background information for the proposed framework, including the overview of Multilayer Perceptron (MLP), Long Short-Term Memory (LSTM), and hyperparameter optimization techniques. Then reviews existing methods for energy consumption prediction and peak energy timestamp labelling in commercial buildings. The research objective and challenges are identified.

2.2 Background Information

This section provides background information for the proposed framework, focusing on deep learning modeling. Specifically, Multilayer Perceptron is introduced followed by an introduction on Long Short-term Memory. This section ends with a discussion of various hyperparameter optimization techniques.

2.2.1 Multilayer Perceptron

Multilayer Perceptron (MLP) is a type of neural network that makes no prior assumptions concerning the data distribution and has been shown to be an effective alternative to traditional statistical techniques [23]. As seen in Figure 2.1, each MLP model consists of an input layer, one or more hidden layers, and an output layer. The input layer is responsible for receiving a vector of values to be processed. The number of neurons in the input layer is determined by the input vector length. Hidden layers are not directly exposed to the input data but connected to the output values from the previous layer. Last, the output layer is the final layer in the MLP model, and it is responsible for outputting a vector of values that correspond to the required format and vector length.

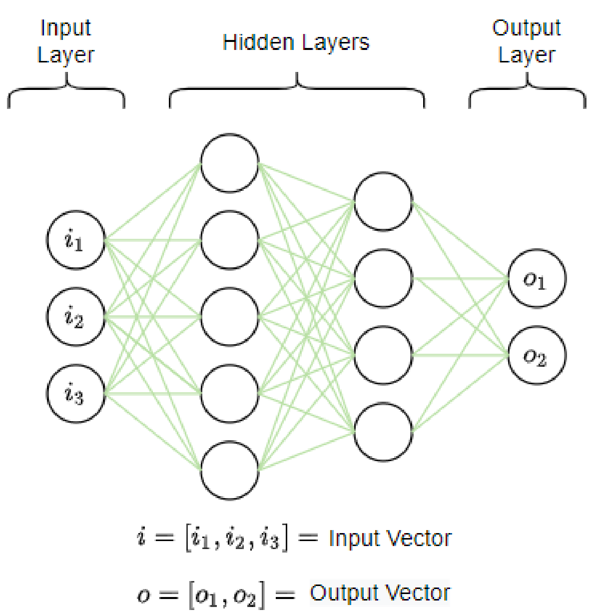


Figure 2.1: A multilayer perceptron model with 2 hidden layers

The MLP model is a system of interconnected neurons where the passing of information is done by the nonlinear transformation of the weighted sum of the inputs as shown in Equation (2.1). A graphical demonstration of a single neuron is included in Figure 2.2, where $\boldsymbol{x} = \{x_1, x_2, ..., x_n\}$ is the input vector, $\boldsymbol{w} = \{w_1, w_2, ..., w_n\}$ is the corresponding weights, and $\boldsymbol{b}$ is the bias term for adjusting the offset. The activation function f is applied to the weighted sum of the inputs.

$$y = f\left(b + \sum_{i=1}^{n} x_i w_i\right) \tag{2.1}$$

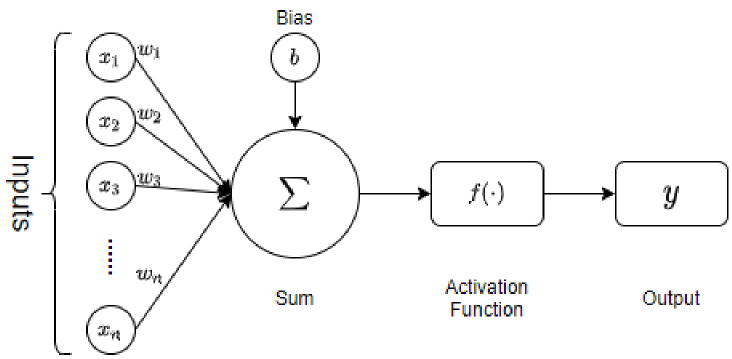


Figure 2.2: Neuron structure

Training for an MLP network is required such that the combination of weights results in the smallest error, defined by a differentiable error function. Common error functions include Mean Square Error (MSE) and Mean Absolute Error (MAE) for regression tasks, and Cross Entropy Loss for classification tasks. The back-propagation training algorithm uses gradient descent in attempting to find the smallest error, known as the global minimum of the error surface. The weights in the MLP network are initially set to small random values, leading to a random point on the error surface. The local gradient is then calculated with the back-propagation algorithm such that the weights are updated in the direction of the steepest downside. This process is continued until the error reaches the desired or smallest value.

2.2.2 Long Short-Term Memory Networks

Long Short-Term Memory (LSTM) network, a type of Recurrent Neural Network (RNN), that is predominantly used to learn and process sequential data. RNN systems are problematic in practice as they suffer from vanishing gradients and exploding gradients as unrolling is presented without justification from the beginning of the sequence to the end [24]. The LSTM network is then introduced to address this problem by incorporating nonlinear controls into the RNN cells that ensure the gradient of the objective function with respect to the state signal does not vanish [25]. Due to the ability in capturing long-term temporal dependencies without suffering from optimization hurdles, LSTM networks have been used widely in language modeling [26], text sentiment analysis [27], and time series forecasting [28, 29, 30].

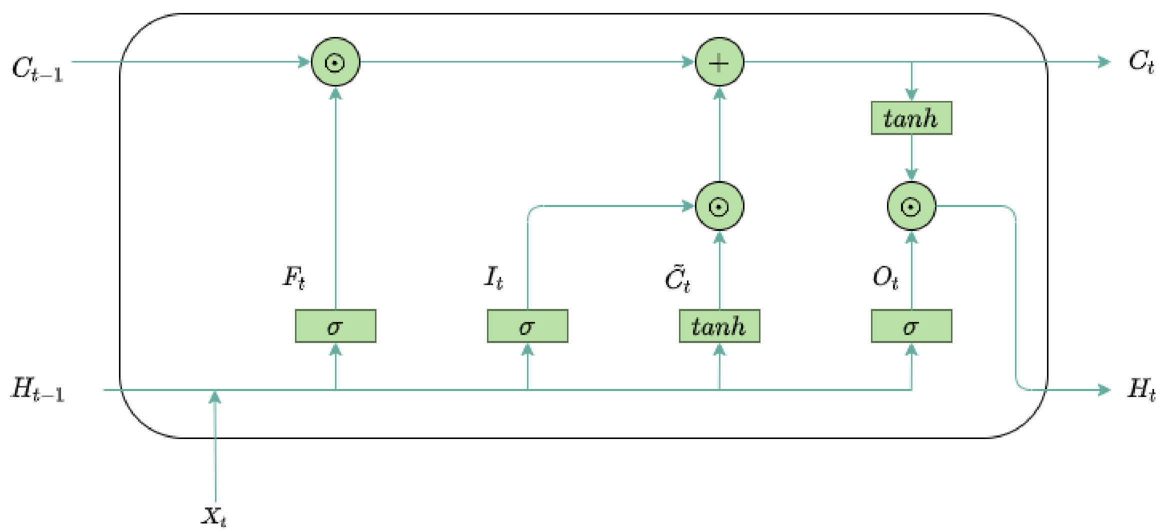


Figure 2.3: Architecture of a long short-term memory unit

Figure 2.3 shows the internal architecture of an LSTM unit which contains three gates, one node, and two states. The input gate, I_t, controls the amount of information to be added to the current memory cell's internal state. The forget gate, F_t, controls the amount of information to keep from the previous state. The output gate, O_t, controls whether the current memory cell influences the output at the current timestep. Overall, all three gates control the flow of information with the sigmoid activation function, which limits the value between 0 and 1. Equation (2.2) shows the calculation of the gate values at time t, in which X is the input value, H is the hidden state, W is the corresponding weight parameter, and B is the corresponding bias parameter.

$$
\begin{aligned}
I_t &= \sigma(X_t W_{xi} + H_{t-1} W_{hi} + b_i) \\
F_t &= \sigma(X_t W_{xf} + H_{t-1} W_{hf} + b_f) \\
O_t &= \sigma(X_t W_{xo} + H_{t-1} W_{ho} + b_o)
\end{aligned}
\tag{2.2}
$$

The input node, $\tilde{C}_t$, combines the information from the current input value and the hidden state from the previous timestamp, which is restricted to a range of -1 to 1 with the tanh activation function. The cell state, C_t, is also known as long-term memory. It uses the elementwise product operator to combine the information from the internal state of the previous timestamp and the current input node. Input gate and forget gate allow the cell to learn how much to respond to the subsequent input, which alleviates the vanishing gradients from a classic RNN cell. The hidden state, H_t, is also known as short-term memory in which the value is restricted to a range of -1 to 1 with the elementwise multiplication. Equation (2.3) shows the calculation of input node, cell state, and hidden state at time t.

$$
\begin{aligned}
\tilde{C}_t &= \tanh(X_t W_{xc} + H_{t-1} W_{hc} + b_c) \\
C_t &= F_t \odot C_{t-1} + I_t \odot \tilde{C}_t \\
H_t &= O_t \odot \tanh(C_t)
\end{aligned}
\tag{2.3}
$$

2.2.3 Hyperparameter Optimization

Hyperparameters are variables in machine learning systems that control the learning process for optimized performance [31]. In a deep learning system, examples of hyperparameters include the number of neurons and hidden layers, learning rate, and activation function. In a model with LSTM layers, sliding window length should also be optimized with justification. The effects of these hyperparameters on the deep learning model are explained as follows:

- The neuron size and number of hidden layers contribute to the fitting of the model. insufficient number of neurons will result in the underfitting of the model, whereas too many neurons will increase the training time and may also lead to overfitting [32].
- Learning rate generally strongly impacts the stability and efficiency of the training process. A large value will cause divergence of the objective function and results whereas a small value will result in slow learning and inefficiency.
- Activation functions improve the ability of the model to extract complex and complicated information from data. Although Rectified Linear Unit (ReLU) is a preferred activation function in the hidden layers, there exist various choices which are context dependent. Examples include sigmoid, tanh, ReLU, and Leaky ReLU; each with distinct advantages and ranges [33].
- Sliding window length will impact the LSTM unit on recognizing the temporal dependencies among features since the historical values are taken as input variables. Optimizing sliding window length is also a trade-off between prediction accuracy and the training speed as increasing the window size will give LSTM unit more information to learn from the past.

As the hyperparameters cannot be directly estimated from data and there exist no analytical formulas to calculate the appropriate values, tuning techniques are needed to optimize model performance [34]. The hyperparameter optimization can be represented as Equation (2.4), where $f(x)$ represents the objective function to minimize, X represents the search spaces of the hyperparameters, and x_{best} is the combination of hyperparameters that yield the lowest value on the score.

$$x_{best} = \arg\min_{x \in X} f(x) \tag{2.4}$$

In general, the basic method for hyperparameter optimization is grid search. Grid search contains user-defined search space which contains a finite set of hyperparameter combinations. Due to the full factorial design, the required number of evaluations grows exponentially as the number of tuning hyperparameters increases [35]. This property makes the grid search algorithm only reliable in low-dimensional spaces due to inefficiencies and the need for computational resources. An improved tuning method is random search where the user only defines the boundary of hyperparameters as the search space. Bergstra and Bengio [36] proved that randomly chosen trials are more efficient in the optimization process as the random points are far more evenly distributed in the subspaces. Both grid search and random search optimization methods can run in parallel as the data samples from the search space do not depend on previous trials.

Bayesian optimization algorithm is also a popular tuning method, which contains two key components: the probabilistic surrogate model and an acquisition function [37]. The probabilistic surrogate model consists of a prior distribution that models the unknown objective function, and the acquisition function is optimized for choosing the next sample point. However, this global optimization strategy in its naive form is not parallel as the acquisition function needs to be updated after each trial, therefore requiring a time-consuming procedure and resulting in high computational cost. Bayesian optimization algorithms also can only work on continuous hyperparameters but not categorical ones such as activation functions. Therefore, due to the proven efficiency of the random search algorithm and the need for tuning the activation function as a categorical variable; the proposed framework incorporates random search as the hyperparameter tuning method.

2.3 Literature Review

This section of the thesis provides an extensive literature review of energy consumption forecasting and peak prediction methods in the application of commercial buildings.

2.3.1 Energy Prediction in Commercial Applications

Non-residential buildings accounted for more than 10% of the global energy use and total global carbon dioxide emissions [4]. In the United States, commercial buildings, including office and retail buildings, educational and health-care buildings, and lodging, accounted for 19% of the energy consumption. Specifically, retail and wholesale trade buildings accounted for 28% of total energy consumption in the non-residential building sector [38]. As briefly introduced in *Chapter 1.1: Motivation* section, the commercial supermarkets will be analyzed for the task of peak demand prediction, where the majority of electricity is consumed by refrigeration systems. The remaining portions are allocated to tasks such lighting, HVAC, baking, and other supplementary services. [39].

Mocanu et al. [40] grouped demand forecasting into three categories: Short-term forecasts, medium-term forecasts, and long-term forecasts. In the definition, short-term forecasts are applied to intervals ranging from one hour to one week, medium-term is usually a week to one year, and long-term forecasts are for ranges longer than one year. Short-term forecast is of specific interest in this research. Therefore, the literature review mainly focuses on short-term forecasts in commercial applications [41, 42]. Upon reviewing some state-of-the-art models, common methods include variations or hybrid models of autoregression integrated moving average (ARIMA) [43, 44], support vector machine (SVM) [45, 46, 47], LSTM [30, 48].

Guo ei al. [44] proposed to use a hybrid model of ARIMA-SVR to predict the electricity consumption with different prediction horizons of an office building. The authors concluded that the proposed hybrid model performed better with a shorter prediction horizon, and the vanilla ARIMA model had better predictions when the horizon is longer. However, the proposed model is only tested on a small-scale dataset containing 117 daily electricity consumption datapoints; and is limited to stationary univariate time series or stationary univariate time series after differencing.

Hong and Fan [46] proposed to use H-EMD-SVR-PSO hybrid model to improve the forecasting accuracy of New South Wales market in Australia. In the proposed model, empirical mode decomposition (EMD) is applied to decompose the electric load data into nine intrinsic mode functions (IMFs). The IMFs are divided into three groups and modeled separately using support vector regression (SVR) with particle swarm optimization (PSO), initially proposed in [45]. The authors concluded that the proposed H-EMD-SVR-PSO model generated high accuracy at 5.83 for MAPE error and can be easily interpreted.

Pallonetto et al. [48] compared the results of LSTM and SVM models on a building with a strong commercial profile. The authors examined the correlation coefficients between historical load and weather profile and proposed to use only historical load data due to low coefficients. The model performance is evaluated using Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE). The authors concluded that the LSTM model has a higher prediction accuracy when the load data is sufficient. However, when the training data is insufficient and the time cost is prioritized, the overall performance favored the SVM model.

Based on the literature review of short-term forecasts in commercial buildings, the proposed methodology will include LSTM as a baseline model for peak demand prediction. Fu et al. [49] also claimed that there are typically three types of independent variables considered for most commercial building energy models: weather, occupancy, and time. Based on the availability of data acquisition, this research will include weather information and time-related variables in the energy prediction modelling. The common metrics for evaluating the model performance include MAE, RMSE, and MAPE. These metrics will be leveraged to evaluate the proposed model.

2.3.2 Peak Demand Prediction

As mentioned in *Chapter 1*, peak demand prediction has many use cases in applications ranging from residential houses to national grids. This section of the literature review will focus on commercial applications [50, 51, 52], as well as methodologies proposed for time index [18, 53].

Fan et al. [50] proposed to use ensemble models for predicting next-day total energy consumption and peak power demand. The proposed method presented a data mining-based approach where variables for the individual models are selected based on recursive feature elimination (RFE), and the weights for the ensemble model are optimized using the genetic algorithm (GA). The authors concluded that the accuracy of the ensemble model is evidently higher than those of individual base models. However, the prediction of the peak power demand is treated as a standalone task where RFE and GA steps are done separately from the energy consumption prediction. In addition, the model does not consider the timing of the peak demand, which is necessary information for peak shaving.

Taheri and Razban [51] proposed using a probabilistic regression model for daily electrical peak demand prediction in commercial and manufacturing applications. In the modeling, the authors proposed to use temperature and occupancy (and production for the manufacturing application) in the prediction of electric demand. This research concluded that the proposed method is more effective than machine learning algorithms (specifically support vector machine, random forest, and multilayer perceptron) for small datasets. However, in both cases (aluminum die casting company as manufacturing application and museum as commercial application), the model is trained with only one datapoint for each feature per day (i.e. predicting only maximum daily consumption) such that the output only considered the value of the peak.

Chae et al. [52] studied the data-driven forecasting model for day-ahead electricity usage of commercial buildings in 15-min intervals. The proposed method for the short-term building energy forecasting model is an artificial neural network model with a Bayesian regularization algorithm. Results show that the proposed adaptive training methods are capable to predict the daily peak electricity usage reasonably well in a test case of a commercial building complex. However, the peak hours are predefined by the suppliers and are only based on the seasons.

The above methodologies used various methods in predicting the value or magnitude of peak demand. However, the index or timestamp for the peak occurrence was not of interest to research. On the other hand, Amara-Ouali et al. [18] considered daily electricity load peak demand as the maximum of the electricity power demand curve over one day. This is also known as the block maxima (BM) approach. This group of researchers proposed to use a multi-resolution approach in forecasting demand peak magnitude and the instant at which it occurs. The proposed algorithm was tested with a half-hourly demand resolution using UK total national demand. The authors also reported a mixed performance on peak timing forecasting, especially with a multi-resolution generalized additive model (GAM).

In addition, Gilbert et al. [53] proposed to use probabilistic forecasting with forecast fusion in the low-voltage network to predict both peak density and timing. The proposed methodology was tested using real smart meter data and a hypothetical low-voltage network hierarchy comprising feeders, secondary, and primary substations. The authors concluded that the proposed framework based on Generalized Additive Models for Location, Scale, and Shape is in contrast to non-parametric methods and produced an average improvement of 5% compared to Kernel Density Estimation.

2.3.3 Challenges and Objectives

From this literature review, it can be concluded that the time index of the peak demand has not been studied enough in the realm of commercial building applications. In most studies, peak demand prediction only considered the magnitude of the peak value. This lack of interest in identifying the timestamps was also a common theme among applications in other areas such as residential houses. In the few studies that consider the timestamp for peak demand prediction, the methodologies were freshly proposed and state-of-the-art. Those methodologies have not been tested with commercial building applications, and either have mixed performance in the timestamp forecasting or using hypothetical networks.

To fill the gap, this research recognizes the importance of predicting both the value and the index of the peak for the purpose of smart grid management and operational planning [54]. The proposed framework predicts not only the magnitude of the peak demand value, but also dedicates effort towards predicting the timestamps for the starting, peaking, and ending index in commercial supermarkets.

2.4 Summary

This chapter provides background information for the proposed framework, focusing on deep learning modeling. Specifically, Multilayer Perceptron is introduced followed by an introduction on Long Short-term Memory and a discussion of various hyperparameter optimization techniques. The second section of this chapter conducted a detailed literature review on the topics of energy forecasting and peak demand prediction mainly in commercial buildings. The chapter ends with a summary of the current research challenges and opportunities.

Chapter 3

Methodology

3.1 Introduction

This chapter of the thesis introduces the proposed Energy Peaks and Timestamping Prediction (EPTP) framework as a novel approach for peak demand detection using electricity consumption data, and weather information. The flow of this chapter includes an overview of the proposed EPTP framework, followed by a detailed description of each step in the subsequent sections.

3.2 Framework Overview

Figure 3.1 shows an overview of the proposed EPTP framework. This framework can be mainly divided into three phases, namely data preprocessing and timestamp labelling, energy consumption prediction with LSTM model, and timestamp prediction with MLP model.

The flow of the data through the framework is described as the following. The raw data

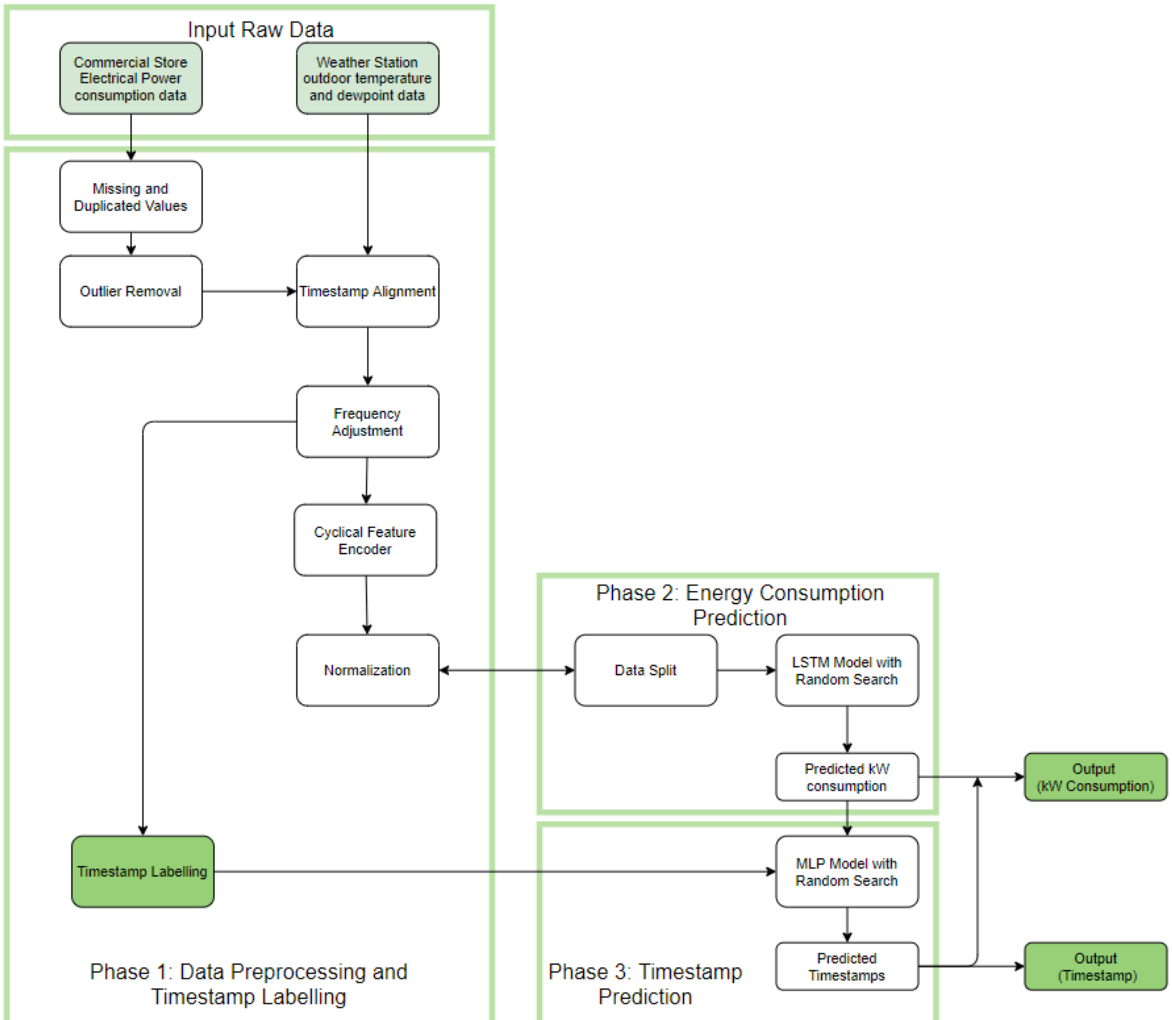


Figure 3.1: The proposed EPTP framework

is obtained from two sources: a commercial supermarket and its closest weather station. In the commercial supermarket, the energy consumption is obtained from the sensor reading, in the unit of kW. Braun et al. [55] studied the energy consumption pattern in commercial supermarkets, and concluded that half of the electricity usage in commercial supermarkets is directly related to the weather. Therefore from the weather station, outdoor temperature (unit: $^{\circ}C$) and dewpoint temperature (unit: $^{\circ}CTd$) are extracted. The raw data then goes through the data preprocessing steps to sanitize the input required for the model. Time-related variables are feature engineered as part of the input. Then, three timestamps per day are labelled as starting, peaking, and ending indexes. In the second stage, the sanitized dataset passes through an LSTM model to predict the energy consumption of the commercial supermarket for the next 24 hours. Lastly, an MLP model is trained with the predicted 24-hour energy consumption as input to obtain the indexes which are compared to the label from the original dataset. As the end goal of the proposed EPTP framework, the prediction should not only contain the value of the peak consumption but also the indexes for which it occurs 24 hours in advance.

3.3 Phase 1: Data Preprocessing and Timestamp Labelling

Phase 1 in the proposed EPTP framework is to preprocess the raw dataset and label the starting, peaking, and ending indexes. Each step will be explained in detail in the following section.

3.3.1 Data Preprocessing

Data preprocessing is a necessary step in building machine learning models. When obtaining raw data from real-world applications, impurities are naturally included. Data preprocessing serves the purpose of removing impurities in the dataset and improving accuracy and reliability in model training. The data preprocessing section is further divided into several steps, namely dealing with missing and duplicated values, outlier removal, timestamp alignment, frequency adjustment, cyclical feature encoder, and data normalization. Each step will be discussed in detail as follow:

(a). Missing and Duplicated Values

During the data collection stage, unexpected situations may happen causing the occurrence of missing or duplicated values. These unexpected situations include unstable sensor connections, data corruption, or hardware failures. To mitigate the problem of missing values, short-term missing datapoints are linearly interpolated with the average of the previous and next datapoints. In case of long-term missing values have occurred, the daily kW consumption information is removed from the dataset. In the proposed model, a two-hour benchmark has been used to distinguish between long-term and short-term missing values. To ensure that each timestamp contains only one datapoint, any duplicated entries are eliminated.

(b). Outlier Removal

In data collection for commercial building energy consumption, observation is recorded based on the sensor reading. However, it is possible for the observations to contain noise, errors, or unwanted data due to potential sensor faults or connection issues [56]. These observations are considered as outliers that should be removed to improve the data quality for better prediction accuracy.

The outliers in the commercial building kW consumption information are detected and replaced with Hampel filter [57]. The Hampel filter is a type of decision-based filter that implements the moving window of the Hampel identifier. In the Hampel identifier [58], the sliding window is considered to have a window length of $(2N + 1)$ where N represents a positive integer called the window half-width. The median absolute deviation (MAD) can be represented as ψ_k and the median value of the time series data window can be represented as $w_k = \text{median}\{x_{k-N}, ..., x_k, ..., x_{k+N}\}$. The outliers are determined using Equation (3.1), where m_k represents the median value of the sliding window, τ represents the threshold value, and ψ_k represents MAD of the sliding window.

$$\text{Outlier } (O_k) = \begin{cases} \text{True}, & \text{if } |x_k - m_k| > \tau * \psi_k, \\ \text{False}, & \text{if } |x_k - m_k| \leq \tau * \psi_k. \end{cases} \tag{3.1}$$

Once the outliers are detected in the time series kW consumption dataset, the outliers are replaced with the median value of the sliding window, as shown in Equation (3.2).

$$y_k = \begin{cases} m_k, & \text{if Outlier } (O_k) = \text{True}, \\ x_k, & \text{if Outlier } (O_k) = \text{False}. \end{cases} \tag{3.2}$$

When utilizing the Hampel filter, N and τ are user-defined parameters that can be adjusted based on the specific outlier detection and replacement requirements. Specifically, N controls the size of the sliding window that is used to calculate m_k and ψ_k; τ determines the boundary hence the number of the outliers detected in the dataset. In this framework, N is chosen to be the number of datapoints per day for the kW consumption prediction, i.e., 1440 if the prediction frequency is one datapoint per minute. τ uses the default value of 3 based on Pearson's rule [59].

(c). Timestamp Alignment

Given that the commercial building kW consumption dataset and the weather information dataset originate from distinct sources, a two-step adjustment is necessary for the timestamp alignment. The modification involves both converting from Coordinated Universal Time (UTC) to local time (LT) and accommodating any daylight saving time (DST) changes. In the conversion from UTC to LT, the timezone is determined to obtain the offset which indicates the time difference. The offset is typically expressed in hours and minutes as a positive or negative value. DST also affects the offset where the clock is adjusted forward by one hour in the spring and then adjusted back by one hour in the fall. After adjusting the UTC time by adding the offset if the LT is ahead of UTC or subtracting the offset if LT is behind UTC, the kW consumption dataset and the weather information dataset are synchronized.

For the observations that are collected during the DST changes, the data is treated as missing and duplicated values. Specifically, when the DST starts in the spring, the observations at 2 a.m. are treated as missing values, thus replaced with the values using interpolation as explained in the first step in data preprocessing. On the other hand, when the DST ends in the fall, the values that are collected at 2 a.m. are treated as duplicated values, thus removed during the transition back to standard time.

(d). Frequency Adjustment

In the proposed framework, kW consumption information of the commercial building and the weather information are collected from different sources. Hence, the frequencies for the available data are different. For example, kW consumption from the commercial building is collected at one data sample per minute whereas the available weather information is based on one data sample per hour. Based on the prediction frequency requirement, data samples are either averaged to a lower frequency or interpolated to a higher frequency.

(e). Cyclical Feature Encoder

Commercial building electricity consumption is affected by the hour of the day [42]. This is especially true in the retail industry as customer footprint increases during the store's operational hours. Therefore, in this proposed EPTP framework, the hour of the day (i.e. ranges between 0 and 23) is included as an independent variable for the purpose of predicting energy consumption. By assigning increasing equidistant values to the successive time intervals, the similarity of adjacent pairs is consistent. However, this encoding of the cyclical feature sees the first and the last instance (i.e. 0 and 23) as being far away from each other, leading to inconsistent similarity.

A common method for encoding the cyclical feature is to transform the data into two dimensions with a trigonometric encoder. To ensure that each interval is uniquely represented, both sine and cosine transformations are included as shown in Equation (3.3).

$$
\begin{aligned}
x_{sin} &= \sin\left(\frac{2 \cdot \pi \cdot x}{max(x)}\right) \\
x_{cos} &= \cos\left(\frac{2 \cdot \pi \cdot x}{max(x)}\right)
\end{aligned} \tag{3.3}
$$

(f). Normalization

Normalization is a common practice in data preprocessing as the raw attributes typically lack sufficient quality for obtaining accurate predictive models. Normalization aims at transforming the original attributes to enhance the predictive capability of the model [60]. In the proposed EPTP framework, the dataset is normalized using the min-max scaling technique to transform the features into the same unit of measure that the deep learning network can easily learn.

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}} \tag{3.4}$$

As shown in Equation (3.4), X represents an observation of the feature, X_{max} and X_{min} are the maximum and minimum values of the feature, and X' is the normalized value for the observation. The normalization is performed for all individual features of the dataset. The MinMaxScaler is fitted on the training set and transformed on the full dataset. The train-test split will be introduced in *Phase 2: Energy Consumption Prediction*.

3.3.2 Timestamp Labelling

A major component of this work is to label and predict the timestamps for the starting, peaking, and ending instances within a specified duration. The peak consumption is defined using the block maxima (BM) approach in the extreme value theory (EVT). BM consists of dividing the observation period into non-overlapping blocks of equal size and retrieving the maximum value within each block [61]. Considering the observation period as P which can be divided into B blocks of size M in which $b = [1, 2, ..., B]$ and $m = [1, 2, ..., M]$. Equation (3.5) shows the retrieval of the peak value at block b where $\tilde{X}_b$ represents all the datapoints in block b.

$$X_b = \underset{(b-1)m<j\leqslant bm}{max} \tilde{X}_b. \tag{3.5}$$

After defining the peak consumption, the main contribution of this study is to label the starting, peaking, and ending indexes in each block b. The peaking index is defined as the timestamp such that the maximum energy consumption has occurred in the predefined window. In this proposal, the window is defined to be twenty-four hours (i.e.: $M = 24$ if the frequency is 1 datapoint per hour, and $M = 96$ if the frequency is 4 datapoints per hour). Algorithm 1 shows the definition of the starting, peaking, and ending indexes for a single block b, the process is repeated for all blocks in B. In the Algorithm, line 2 to line 6 extracts the starting index of the peak using the base value on the left-hand side. The occurrence of the peak is led by an increasing trend where the start timestamp of the increment is defined as the starting index of the peak. Similarly, line 7 to line 11 defines the ending index using the base value on the right-hand side.

The duration of the peak is also validated on line 12. This is for the purpose of avoiding long durations of peak occurrence, especially for low-frequency data. In this proposed Algorithm, four hours is used as a benchmarking value. If the peak lasts more than four hours based on the definition, the ending index or the starting index is updated based on the higher consumption, as seen on line 14 and line 18, respectively.

Algorithm 1 Timestamp labelling

Require: $\tilde{X}_b \wedge X_b \wedge M$
PeakIndex $\leftarrow m$ where $\tilde{X}_b^m = X_b$
for StartIndex = PeakIndex to 1 **do**
  if $X_{StartIndex-1} > X_{StartIndex}$ **then**
    Break For loop
  end if
end for
for EndIndex = PeakIndex to M **do**
  if $X_{EndIndex+1} > X_{EndIndex}$ **then**
    Break For loop
  end if
end for
duration = EndIndex - StartIndex
if duration > 4 hours **then**
  if $X_{StartIndex} > X_{EndIndex}$ **then**
    $X_{EndIndex} \leftarrow X_{StartIndex}$
    $EndIndex \leftarrow k$ where $\tilde{X}_b^k < X_{EndIndex}$
  end if
  if $X_{EndIndex} > X_{StartIndex}$ **then**
    $X_{StartIndex} \leftarrow X_{EndIndex}$
    $StartIndex \leftarrow k$ where $\tilde{X}_b^k < X_{StartIndex}$
  end if
end if
return StartIndex, PeakIndex, EndIndex

3.4 Phase 2: Energy Consumption Prediction

The second phase of the proposed EPTP framework is to predict energy consumption using the given features with LSTM modelling. As discussed in *Phase 1*, the input features include energy consumption in the unit of kW, outdoor temperature in the unit of $^\circ C$, dewpoint temperature in the unit of $^\circ CTd$, and the cyclical transformation of the hour information which are unitless. The complete dataset is then split into a training set, validation set, and test set based on roughly a 70-15-15 ratio. The aforementioned features are normalized using a MinMaxScaler which is fitted to the training set and then used on the validation and test sets.

The input layer of the LSTM modelling accepts 3-dimensional input in the shape of batch size $\times$ window length $\times$ number of features. Since the dataset contains time series sequential data, the input should not be shuffled to maintain the relationship between timesteps. In the LSTM modelling of the proposed EPTP framework, batch size and window length are kept as tuning hyperparameters. The output of the LSTM model depends on the frequency at which the data is processed. The goal of this framework is to predict not only the value of the peak consumption but also the timestamps, therefore it is necessary to predict the consumption for the next day at once. In cases where the data frequency is at one datapoint per hour, the output shape is batch size $\times$ 24. Similarly, when predicting with sub-hourly data at 1 datapoint per 30-minute interval, the output shape is batch size $\times$ 48; and for 15-minute intervals, the output shape is batch size $\times$ 96. Figure 3.2 demonstrates the input shape and output shape of the LSTM model.

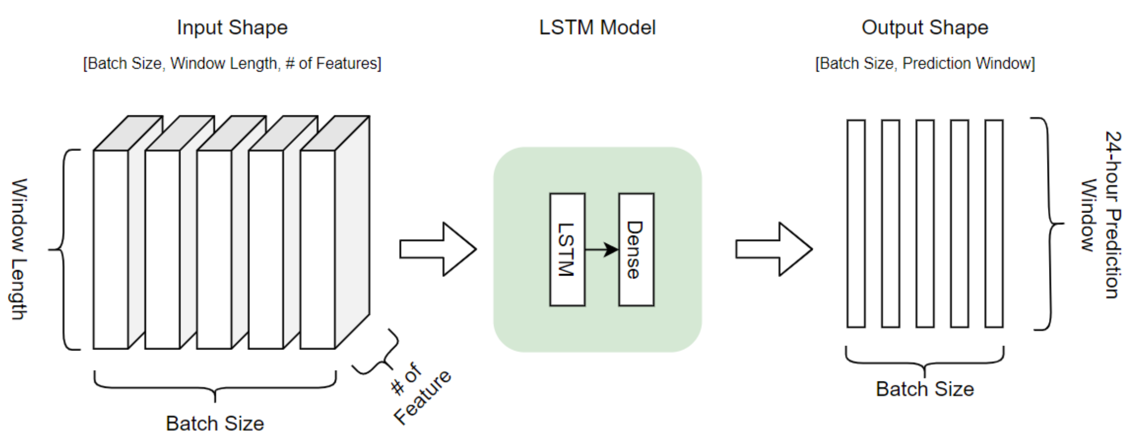


Figure 3.2: Input and output shapes of the LSTM model

As briefly touched upon, the tuning hyperparameters for the LSTM modelling include the sliding window length, activation function, number of LSTM layers, batch size, and learning rate. As the hyperparameters include the learning rate as a continuous variable and the activation function as a categorical variable, Bayesian Optimization does not work well as a tuning method. Since the tuning is done at a high dimension with five tuning parameters, the grid search method leads to wasted computational power and resources due to the full factorial design. Therefore, random search is chosen as the hyperparameter tuning method.

3.5 Phase 3: Timestamp Prediction

In the final phase of the EPTP framework, an MLP model is proposed to predict the timestamps of the daily starting, peaking, and ending indexes. MLP models are proven to show success in detecting delays and time deviations in flight and traffic controls [62, 63]. In the proposed framework, the MLP input layer accepts a 2D input in the shape of batch size x units. As the model aims at predicting the daily peak indexes, the input should contain the electricity consumption for the full day. Therefore, depending on the frequency of the data sample, the input shape for Phase 3 of the proposed EPTP framework is batch size x 24 for hourly frequency, batch size x 48 for a 30-minute frequency, or batch size x 96 for a 15-minute frequency.

The output of the MLP model is batch size x 3 as each neuron represents a single index. In the proposed EPTP framework, starting, peaking, and ending indexes need to be predicted for upcoming peak shaving strategies in commercial applications, hence the output layer contains three neurons. The predicted indexes are then compared with the timestamp labelling from *Phase 1* to identify the model performance. Similarly to LSTM modelling, hyperparameters are also tuned to optimize the performance of the MLP model. These hyperparameters include the number of hidden layers and neuron size, activation function, and learning rate. As the MLP model predicts the timestamps on a daily frequency (i.e.: contains only 365 samples per year) and MLP is less computationally heavy compares to the LSTM modelling, batching is not required in this phase.

3.6 Summary

This chapter of the study introduced the proposed Energy Peaks and Timestamping Prediction (EPTP) framework as a novel approach to detect not only the value/intensity of the energy peak demand in commercial building applications but also the timestamps/duration of the peak occurring indexes. Specifically, this chapter is divided into three phases, namely data preprocessing and timestamp labelling, energy consumption prediction, and timestamp prediction.

Chapter 4

Evaluation for the EPTP Framework

4.1 Introduction

This chapter describes the evaluation of the proposed EPTP framework. First, the dataset, the experimental setup, and the evaluation metrics are introduced. Then three experiments are performed to introduce the performance baselines, tackling the analysis of the real-world commercial supermarket data with various resolutions, and benchmarking the performance with an open-sourced dataset. Finally, the performance of the proposed model is discussed in detail.

4.2 Dataset Information

This section introduces two datasets that are used for the evaluation of the proposed EPTP framework. These two datasets are: a private commercial supermarket dataset provided by an industrial partner, and a commercial retail building from the Building Data Genome Project 2 (BDG2) [64, 65].

4.2.1 Private Supermarket Dataset

In this study, the electricity consumption from a commercial supermarket is used to evaluate the proposed EPTP framework. The supermarket is located in Quebec, Canada. This store opens seven days a week from 8:00 a.m. to 10:00 p.m. (excluding national holidays) and offers a variety of products and services including in-house bakeries, fish and seafood departments, pastry-shop counters, and bistro express. The available data range from August 17, 2020, to January 24, 2023, for a duration of 890 days. The electricity consumption data is collected with a resolution of one datapoint per minute. The weather information is extracted from the closest station listed on the Environment Canada website, at a frequency of 1 datapoint per hour. Since the electricity consumption data is collected at a high resolution, three cases are created to evaluate the model performance. Specifically, the three cases are 1-hour resolution, 30-minute resolution, and 15-minute resolution. An example of the weekly energy consumption data is included in Figure 4.1. It shows that the energy consumption graph contains more details and spikiness with higher resolution. As the resolution decreases, the graph shows a smoother trend with less information.

As briefly touched upon *Chapter 3 Phase 2*, the complete dataset is split into a training set, validation set, and test set based on roughly a 70-15-15 ratio. In this specific commercial supermarket dataset, the training set contains 620 days of data, and validation and test sets each contain 135 days of data. As the experiments in predicting the energy consumption values are treated as sequential data in the LSTM modelling, shuffling is rejected during the split. However, in the last phase of proposed EPTP framework with MLP modelling in predicting the timestamps, each day is treated as a separate sample. Therefore, shuffling is introduced after the LSTM prediction.

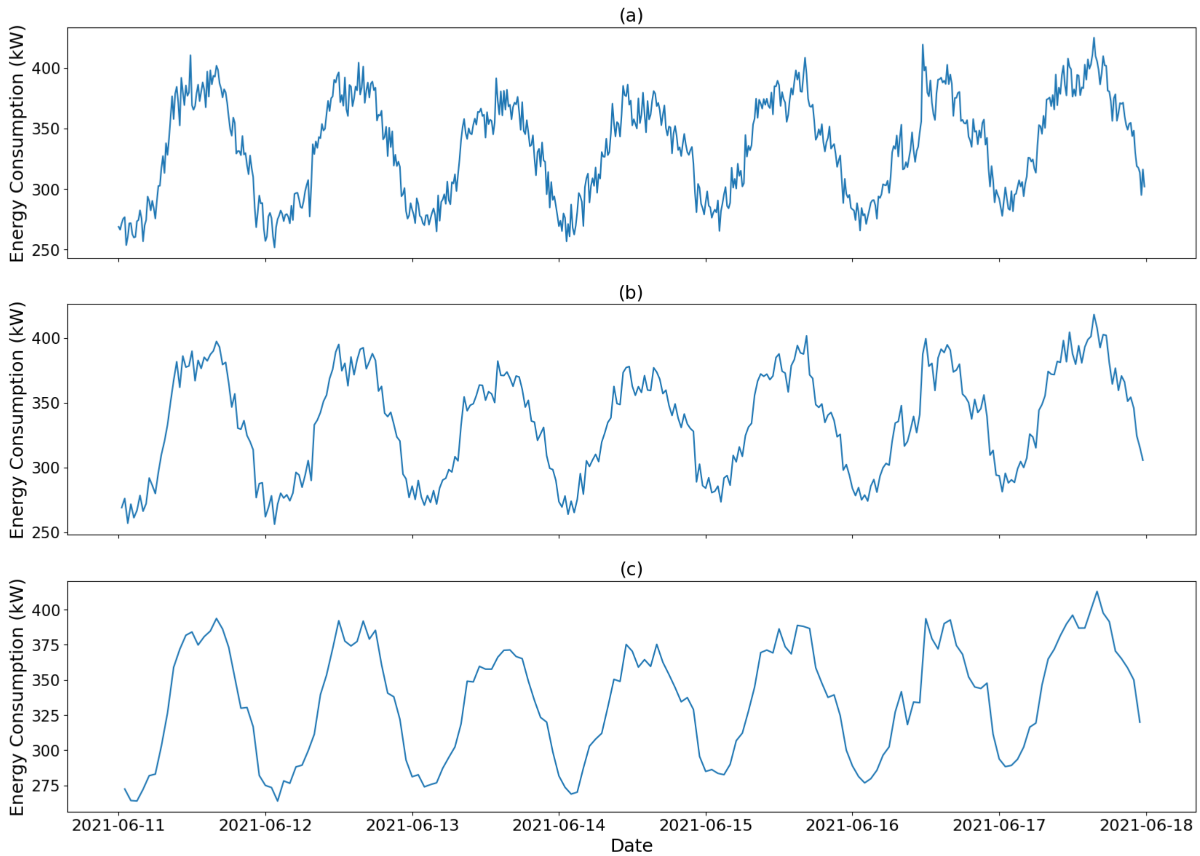


Figure 4.1: Sample weekly energy consumption data for the private commercial supermarket dataset: (a) 15-minute resolution, (b) 30-minute resolution, (c) 1-hour resolution.

4.2.2 Benchmark Dataset

The BDG2 open dataset contains 1636 non-residential buildings with a range of two full years (2016 - 2017) at an hourly frequency. Among the 1636 non-residential buildings, only 12 are categorized into ”Retail” for primary use type. For the purpose of benchmarking, the Fox Retail Manie building was chosen to model the peak values and timestamps. Figure 4.2 shows the example of weekly energy consumption for all 12 ”Retail” buildings in the BDG2 dataset. Upon close inspection of the BDG2 open dataset, the Fox Retail Manie building had better data quality compared to the other retail buildings. 4.2. Fox Retail Manie also has the highest correlation, at 0.56, with the private dataset that was previously introduced. All other retail buildings from BDG2 have a correlation of below 0.47 with the private dataset.

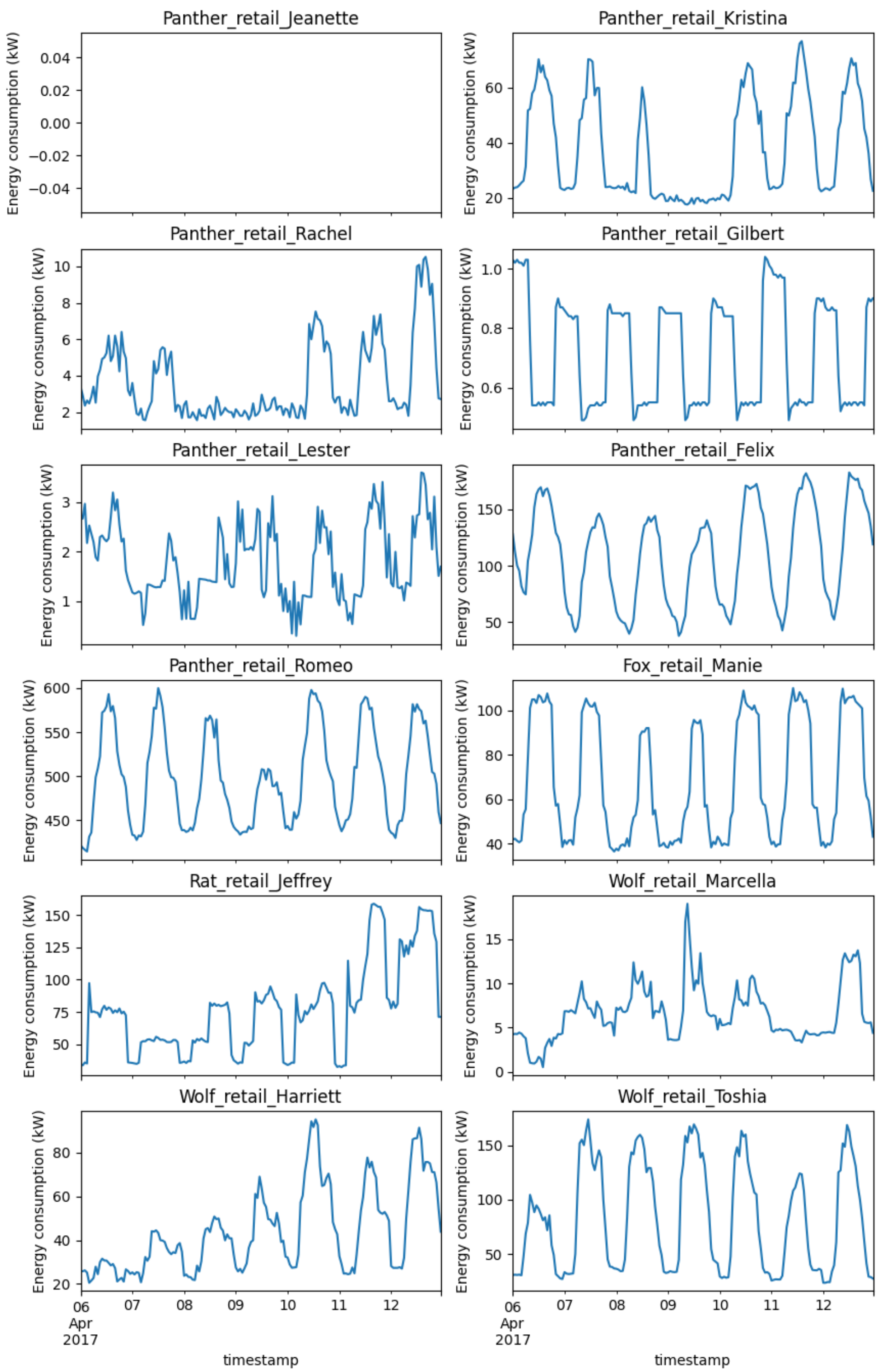


Figure 4.2: Sample weekly energy consumption data for the BDG2 dataset with buildings for retail as primary use type.

The BDG2 dataset contains only two-year of data, therefore 560 days are used in the training set, and 85 days each are used in the validation and the test set. As all the data are provided in 1-hour resolution, the benchmark dataset will only be used to model the low-frequency application. Same to the private commercial supermarket dataset, the energy consumption is treated as sequential data in the LSTM modelling with rejected shuffling during the split. Each day is treated as a separate sample in the MLP modelling in predicting the timestamps, and thus shuffling is introduced after the LSTM prediction.

4.3 Experiments Setup

The models and experiments ran on a Linux server with an Intel® Core™ i7-6850K Processor and 64 GB RAM, with 2 GPUs GeForce GTX 1080 Ti and OS Ubuntu 20.4. The algorithm and framework were implemented in Python version 3.9.0. Pandas and Numpy libraries are implemented for dataset exploration and manipulation. The deep learning models in the case study are implemented using the PyTorch library with Ray Tune hyperparameter optimization. Visualizations are generated with the Matplotlib library.

The list of hyperparameters and their search spaces of the LSTM modelling is provided in Table 4.1. To generalize the results, Ray tune optimization is performed with 20 trials. Since random search is used as the optimization method with the learning rate as a continuous hyperparamater, regenrating the searching process with the exact learning rate using seeding is impossible. Therefore, the hyperparameters from the top three trials of the tuning process are extracted, and each trial is run for the second time to compare the averaged result. With the best energy consumption prediction results, the model moves to the third phase of predicting the timestamps where the list of hyperparameters and their search spaces for the MLP modelling is provided in Table 4.2. The MLP hyperparameter optimization process is also performed with 20 trials with the top three trials repeated for the second time to generate the averaged results.

Table 4.1: List of hyperparameters and the search space in the LSTM modelling for electricity consumption prediction using different resolutions.

Hyperparameter	Description	Search Space
Sliding Window	Historical Window Size	Sample [1] $\times$ [5, 7, 10]
Activation	The activation function for each layer	['tanh', 'relu', 'sigmoid']
Learning Rate	Learning rate for the optimizer	log uniform [0.01, 0.0001]
Batch	Number of samples processed before model update	[256, 512, 1024]
LSTM layer	number of LSTM layers	[1, 2]

Table 4.2: List of hyperparameters and the search space in the MLP modelling for timestamp prediction using different resolutions.

Hyperparameter	Description	Search Space
Hidden Layer	Number of hidden layers in the model	[1, 2, 3] [2]
Neuron Size 1	The number of hidden neurons for the first layer	1-hour: random integer [18, 23] 30-min: random integer [34, 47] 15-min: random integer [66, 95]
Neuron Size 2	The number of hidden neurons for the second layer	1-hour: random integer [11, 17] 30-min: random integer [19, 33] 15-min: random integer [35, 65]
Neuron Size 3	The number of hidden neurons for the third layer	1-hour: random integer [4, 10] 30-min: random integer [4, 18] 15-min: random integer [4, 34]
Activation	The activation function for each layer	['tanh', 'relu', 'sigmoid']
Learning Rate	Learning rate for the optimizer	log uniform [0.01, 0.0001]

4.4 Evaluation Metrics

Various metrics can be used to evaluate a predictive model. In this study, mean absolute error (MAE), root mean squared error (RMSE), and mean absolute percentage error (MAPE) will be used to evaluate the energy consumption prediction. RMSE measures the average difference between the predicted values and the actual values whereas MAE measures the average of the

[1]Sample refers to the number of datapoints per day. In this case, Sample = 24 in the one-hour resolution modelling, sample = 48 in the 30-minute resolution modelling, and sample = 96 in the 15-minute resolution modelling.

[2]Neuron Size 1, 2, and 3 refers to the situation where three hidden layers are used in the modelling. Only neuron size 2 will be used if hidden layer = 1, and neuron size 1 and 3 will be used if the hidden layer = 2.

absolute errors. Although a lower value represents a more reliable result in both metrics, both RMSE and MAPE are not normalized regarding the scale of the predictions. On the other hand, MAPE provides a better understanding of the error scale as it represents the average of the absolute percentage of errors.

$$\text{RMSE} = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(t_i - \hat{t}_i)^2} \tag{4.1}$$

$$\text{MAE} = \frac{1}{m}\sum_{i=1}^{m}|t_i - \hat{t}_i| \tag{4.2}$$

$$\text{MAPE} = \frac{100\%}{m}\sum_{i=1}^{m}\left|\frac{t_i - \hat{t}_i}{t_i}\right| \tag{4.3}$$

Equations (4.1) to (4.3) shows the calculation of RMSE, MAE, and MAPE values where m represents the sample size, $t = \{t_1, t_2, ..., t_i\}$ and $\hat{t} = \{\hat{t}_1, \hat{t}_2, ..., \hat{t}_i\}$ represent the grounded truth and predicted values, respectively.

In the timestamp prediction evaluation, hit rate (HR) metric [66] will be used in addition to the RMSE and MAE metrics. This is due to the fact that deviations from the discrete timestamp cannot be directly translated to percentages. Equation 4.4 shows the calculation of the HR metic where δ represents the tolerance residual for the expected timestamp and h is a flag representing whether the predicted timestamp is within the tolerance interval of $[t_i - \delta, t_i + \delta]$. In this study, the tolerance interval is chosen to be one hour (HR-1)and two hours (HR-2) for a better comparison. This means that δ is [1, 2] for the hourly frequency, [2, 4] for the half-hourly frequency, and [4, 8] for the 15-min frequency.

$$\mathrm{HR} = \frac{100\%}{m}\sum_{i=1}^{m} h_i \text{ where } h_i = \begin{cases} 1, & \hat{t}_i \in [t_i - \delta, t_i + \delta], \\ 0, & \text{otherwise.} \end{cases} \tag{4.4}$$

4.5 Experiments and Results

This section of the thesis presents three experiments to prove the performance of the proposed EPTP framework. In the first experiment, two performance baselines are established. In the second experiment, the proposed EPTP framework is performed on the private commercial supermarket dataset to compare with the performance baseline. The third experiment shows the evaluation of the model results on the prediction with BDG2 Fox Retail Manie data.

4.5.1 Experiment 1: Baseline Comparison

In this experiment, two performance baselines are established to compare with the results of the proposed EPTP framework. The first baseline is created such that the energy consumption prediction is based on Support Vector Regression (SVR), followed by MLP for the timestamp predictions. The second baseline is created using the common method in extracting the peak index for commercial building applications: Block Maxima (BM) with Extreme Value Theory (EVT).

Experiment Process

In the baseline using SVR modelling for commercial supermarket energy consumption prediction. The raw data goes through three phases as explained in *Chapter 3* except that the energy consumption prediction is performed with SVR instead of LSTM architecture. Following the same structure for data preprocessing, the cleaned data then passes through SVR with hyperparameter tuning. The hyperparameters of the SVR model are listed in Table 4.3. With the best possible energy consumption predictions, the predictions are grouped into days based on the number of datapoints. This means that the samples are grouped by 24 datapoints at an hourly resolution; 48 datapoints at a 30-minute resolution, and 96 datapoints at a 15-minute resolution. The grouped data then goes through the MLP model for timestamp prediction of the starting, peaking, and ending indexes.

Table 4.3: List of hyperparameters and the search space in the SVR modelling for commercial supermarket electricity consumption prediction.

Hyperparameter	Description	Search Space
C	Regularization for the penalty	random value between [1, 100]
Kernel	Kernel function used in the algorithm	['rbf', 'poly', 'sigmoid']

In the second baseline, the timestamps are extracted based on the BM. This means that instead of modelling for the indexes, the timestamp for the highest energy consumption for each day is treated as the peak index. However, since the model will not be trained to learn the starting and ending indexes in the second baseline, only the peak index will be compared to the proposed EPTP framework in the later experiments.

Results Analysis

The objective of this experiment is to investigate the performance of the timestamp prediction using SVR for energy consumption prediction and the BM for timestamp extraction as a common method. Table 4.4 summarizes the results of the two baseline models for the purpose of comparing timestamp indexes. The MAE error represents the deviation of the predicted timestamps from the true label. However, since the model works with different resolutions, the MAE error has been translated to minute deviation values for a better comparison. In the subsequent sections, the performance of the proposed EPTP framework will be compared to the HR-1, HR-2, and Minute deviations specifically to validate the results.

Table 4.4: Results of the performance baselines for commercial supermarket electricity consumption data in different resolutions.

Method	Resolution	Index	RMSE	MAE	HR-1 (%)	HR-2 (%)	Minutes
SVR+MLP	15 minute	Start	11.68	9.78	20.71	47.17	146
		Peak	11.7	9.89	18.57	49.28	148
		End	11.46	9.66	22.86	51.43	145
	30 minute	Start	5.95	5.15	19.38	47.86	154
		Peak	5.51	4.83	21.43	47.14	145
		End	5.26	4.55	33.57	50	136
	1 hour	Start	3.16	2.76	25.2	47.86	165
		Peak	2.86	2.5	30	46.43	150
		End	3.8	3.19	25.72	42.86	191
LSTM+BM	15 minute	Peak	10.69	8.01	41.43	65.52	120
	30 minute	Peak	5.11	3.77	50.88	78.10	113
	1 hour	Peak	2.68	2.32	19.74	21.04	139

4.5.2 Experiment 2: EPTP Framework Performance over the Private Dataset

In this experiment, the proposed EPTP framework is examined using commercial supermarket electricity consumption data with various frequencies. Initially, the model is proposed to work with data of 15-minute resolution based on the industrial specifications. However, in situations where high-frequency data is inaccessible (such as the BDG2 dataset and the weather information), prediction using lower-frequency data is deem necessary. Therefore, three cases are used to evaluate the model performance: 1-hour resolution, 30-minute resolution, and 15-minute resolution.

Experiment Process

As explained in *Dataset Information*, the energy consumption of the commercial supermarket is collected at 1 datapoint per minute frequency whereas the weather information is retrieved at 1 datapoint per hour. In this experiment, the raw data is preprocessed such that high-frequency data is averaged into the expected resolution, and the low-frequency data is interpolated if necessary. In the *Phase 2* of the proposed ETPT framework, LSTM for energy consumption prediction is optimized with 20 trials, and the hyperparameters for the top three trials are listed in Table 4.5. The top three trials are repeated for a second training process to avoid lucky shots.

Table 4.5: Top three trials of the LSTM hyperparameter optimization for commercial supermarket electricity consumption prediction using different resolutions.

Resolution	Trial	Sliding Window	Activation Function	Learning Rate	Batch	Layer
15 minute	M1	960	'tanh'	0.000482	256	2
	M2	480	'relu'	0.000699	256	1
	M3	672	'tanh'	0.000310	256	2
30 minute	M1	336	'relu'	0.00095	256	2
	M2	336	'relu'	0.000961	256	2
	M3	336	'relu'	0.00033	256	2
1 hour	M1	240	'tanh'	0.000955	512	1
	M2	120	'tanh'	0.000886	512	1
	M3	168	'relu'	0.000703	512	1

Once the best hyperparameters are retrieved from *Phase 2*, the predicted electricity consumption data are used to predict the timestamps for starting, peaking, and ending indexes by comparing the predicted indexes to the timestamp labelling in *Phase 1*. Similar to the LSTM training, the top three trials (listed in Table 4.6) are repeated again to generalize the results and avoid lucky shots. As explained in Table 4.2, only neuron size 1 and neuron size 3 are used if the hidden layer is 2 in the random search. Therefore, neuron size 2 is not included as the value is not related to the tuning process.

Results Analysis

The hardware requirement for training the proposed EPTP framework is given in *Chapter 4.3: Experiments setup*. MLP modelling is less computationally heavy compared to the LSTM modelling as the MLP model predicts the timestamps on a daily frequency (i.e.: contains only 365 samples per year). Therefore, the training time focuses on the computational power for

Table 4.6: Top three trials of the MLP hyperparameter optimization for timestamp prediction with commercial supermarket electricity consumption data using different resolutions.

Resolution	Trial	Activation	Learning Rate	Layer	Hidden1	Hidden2	Hidden3
15 minute	M1	'relu'	0.00620	3	91	52	16
	M2	'relu'	0.00968	3	93	46	18
	M3	'sigmoid'	0.00913	3	95	64	8
30 minute	M1	'relu'	0.00814	2	42	-	16
	M2	'relu'	0.00772	2	34	-	16
	M3	'relu'	0.00911	2	36	-	17
1 hour	M1	'relu'	0.00257	3	18	12	6
	M2	'relu'	0.00214	3	18	14	7
	M3	'relu'	0.00229	3	23	12	9

the LSTM modelling. The training time for each LSTM random search model is around 0.52 hours, 2.45 hours, and 5.33 hours for the 1-hour, 30-minute, and 15-minute resolutions, respectively. Table 4.7 shows the results of the LSTM model in commercial supermarket energy consumption prediction. The model is generating comparable results across different resolutions as the MAPE error on the energy consumption is all around 4.5% in all three cases. Table 4.8 shows the metrics in evaluating the predicted starting, peaking, and ending timestamps. Note that the Minutes deviation metric is a direct translation from the MAE error in timestamp deviations.

Table 4.7: Final result of the LSTM model in energy consumption prediction with commercial supermarket electricity consumption data in different resolutions.

Resolution	Trial	RMSE	MAE	MAPE(%)
15 minute	M1	20.33	15.32	4.48
30 minute	M2	21.21	15.05	4.20
1 hour	M1	22.54	16.52	4.88

From the timestamping comparison, it is clear that as the frequency of the input data increases (i.e., more samples per hour), the predictive results present better accuracy. Figure 4.3 shows the boxplot of the differences between predicted indexes and real indexes in all three resolutions. It demonstrates that the range between the first and third quartiles is narrower for higher-resolution data. However, the boxplot of the 15-minute resolution data shows some outliers in the predictions. Figure 4.4 shows the comparison between the density plot of different

Table 4.8: Final result of the MLP model in timestamp prediction with commercial supermarket electricity consumption data in different resolutions.

Resolution	Trial	Index	RMSE	MAE	HR-1 (%)	HR-2 (%)	Minutes
15 minute	M3	Start	7.08	6.22	57.14	76.03	93
		Peak	5.72	4.16	56.31	86.05	62
		End	6.16	4.91	54.82	82.14	74
30 minute	M3	Start	5.20	4.39	33.11	53.25	132
		Peak	4.98	4.22	33.26	54.55	127
		End	5.01	4.06	37.66	57.78	122
1 hour	M1	Start	3.42	2.86	32.46	42.86	172
		Peak	3.09	2.47	30.52	52.60	148
		End	3.33	2.72	32.47	53.91	163

resolutions. It reveals that as the frequency of the data increases, the predicted results better capture the bi-modal distribution in the real label. This prediction of the bimodal timestamp label may have contributed to the outliers in the boxplot.

The EPTP framework generated better results, especially for the 15-minute resolution data when compared to the performance baselines. The LSTM modelling outperformed the SVR modelling in following the trends and fluctuations in the commercial supermarket consumption data, hence better timestamp prediction results. However, using LSTM with BM provided higher HR for the 30-minute resolution, in which the proposed EPTP framework needs to be further improved.

4.5.3 Experiment 3: EPTP Performance over Commercial Building

In this experiment, benchmark analysis is performed using open access dataset from BDG2 to evaluate the proposed EPTP framework. Specifically Fox Retail Manie building is chosen as explained in the *Dataset Information* section.

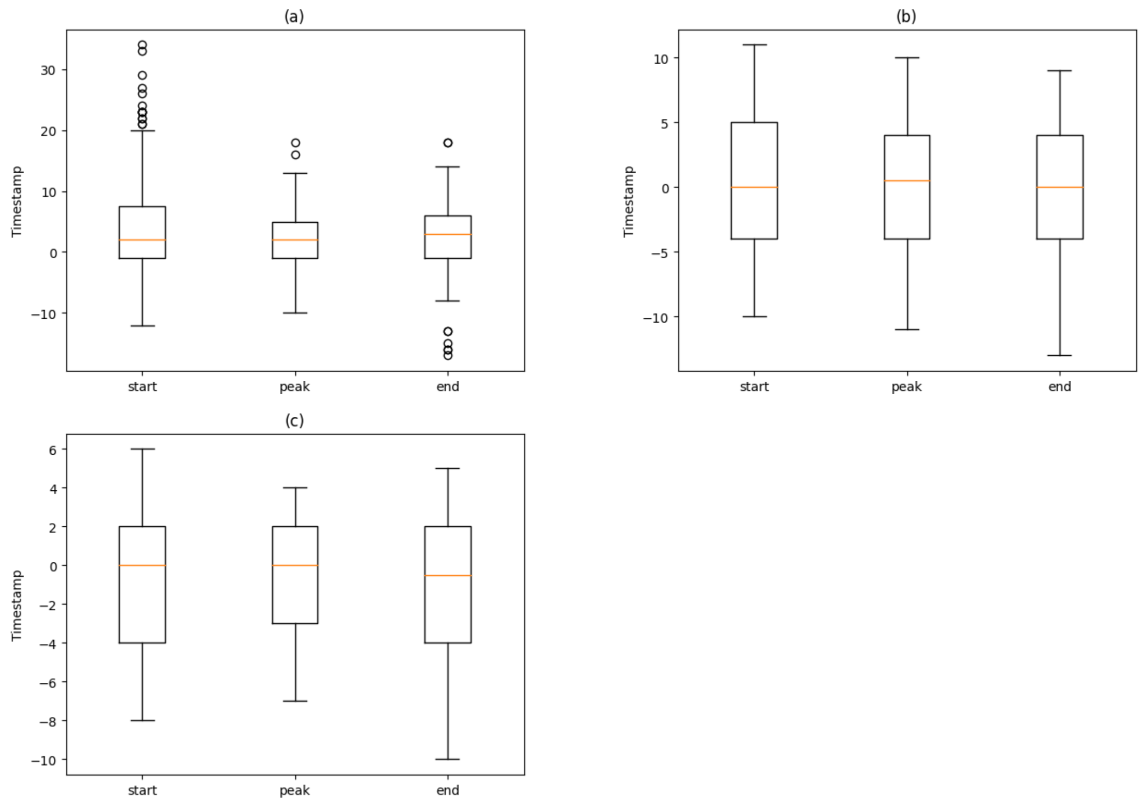
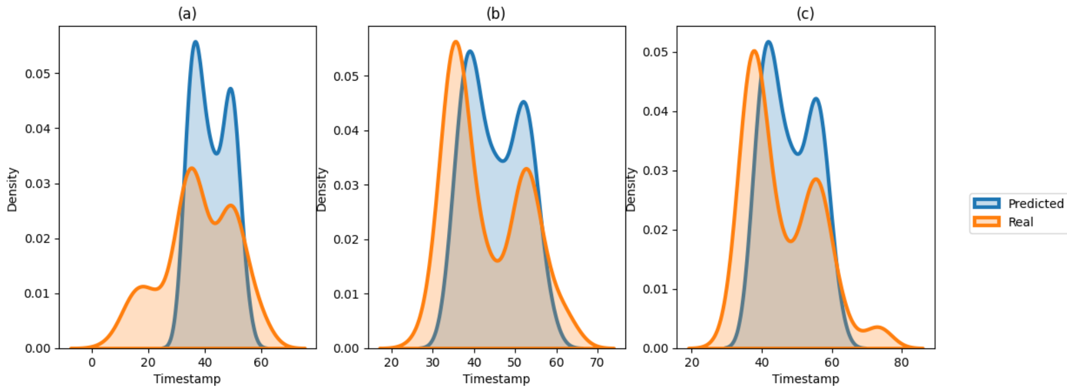


Figure 4.3: Comparison of the boxplot in predicting the starting, peaking, and ending indexes with MLP modelling: (a) 15-minute resolution, (b) 30-minute resolution, (c) 1-hour resolution.

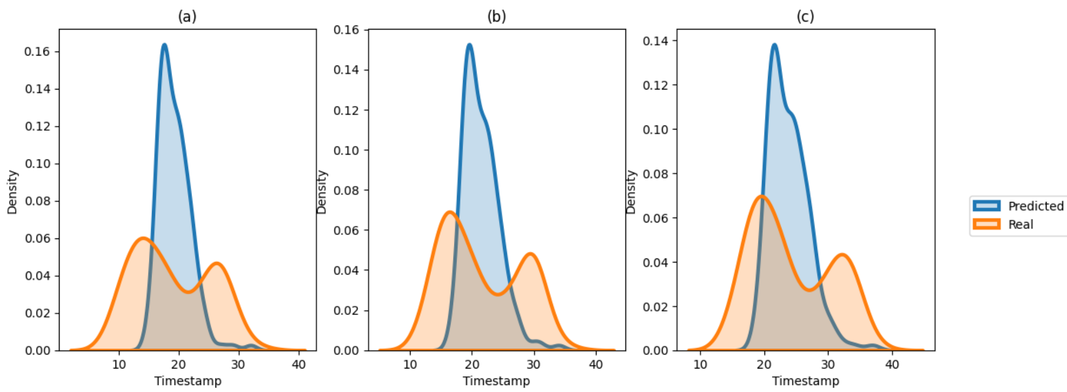
Experiment Process

The Fox Retail Manie building consumption data goes through the same procedure as listed in the proposed EPTP framework. However, since the dataset contains only two years of data at an hourly resolution, only hourly modelling is only performed to avoid excessive data interpolation. The same search space is used for both energy consumption prediction (Table 4.1) and timestamp prediction (Table 4.2) to maintain consistency. As explained, to avoid lucky shots and provide generalization of the model performance, the top three trials, as listed in Table 4.9, are extracted to run for the second time and compare the averaged results.

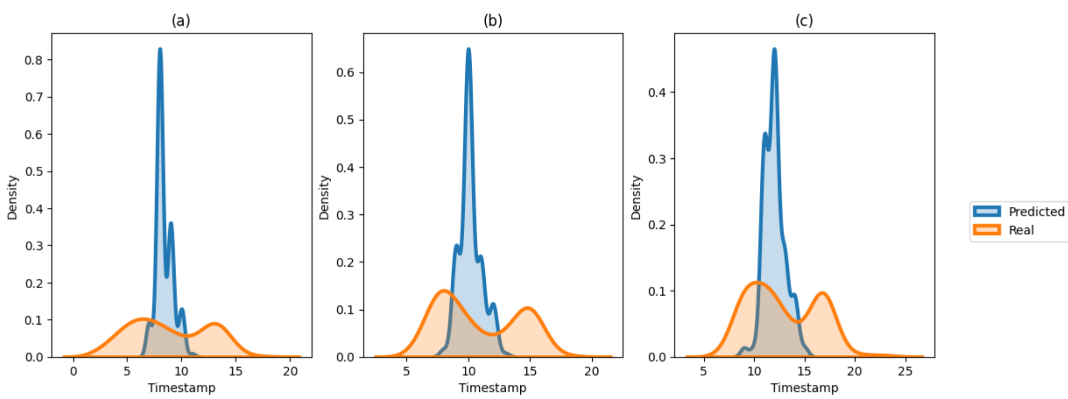
Following the same procedure as in *Experiment 2*, the predicted electricity consumption data from the best LSTM hyperparameter combination is used to predict the timestamps. Table 4.10 lists the top three combinations of the MLP modelling. Note that only hidden neuron 1



(i) 15-minute resolution



(ii) 30-minute resolution



(iii) 1-hour resolution

Figure 4.4: Comparison of the density plot on the test set between real indexes and predicted indexes: (a) starting index, (b) peaking Index, (c) ending Index.

Table 4.9: Top three trials of the LSTM hyperparameter optimization for Fox Retail Manie energy consumption prediction.

Trial	Sliding Window	Activation Function	Learning Rate	Batch	Layer
M1	168	'relu'	0.00095	512	2
M2	240	'relu'	0.000986	512	2
M3	168	'relu'	0.000920	512	2

Table 4.10: Top three trials of the MLP hyperparameter optimization for Fox Retail Manie energy consumption prediction.

Trial	Activation	Learning Rate	Layer	Hidden1	Hidden2	Hidden3
M1	'sigmoid	0.00551	2	22	-	6
M2	'sigmoid	0.00603	2	18	-	10
M3	'sigmoid'	0.00633	2	18	-	4

and hidden neuron 3 are used in the cases where the hidden layer is randomly chosen as two. Therefore hidden neuron 2 is not included as the value does not relate to the tuning process. Each combination of the best hyperparameters are run for the second time to compare the averaged results.

Results Analysis

Table 4.11 shows the results of the LSTM model in Fox Retail Manie in comparison to the private commercial supermarket dataset that is previously used in *Experiment 2*. Although the RMSE and MAE errors for Fox Retail Manie are lower compared to the private commercial building dataset, the MAPE error is higher. This is mainly contributed by two factors. First, the range of energy consumption for Fox Retail Manie is on a lower scale compared to the commercial supermarket private dataset. The highest reading from Fox Retail Manie is around 118 kW, whereas the highest reading from the private commercial supermarket dataset is around 532 kW. Second, the commercial supermarket opens seven days a week (excluding national holidays), meaning that the variation in energy consumption per day is minimal. Upon visual inspection of the energy consumption graphics from Fox Retail Manie, it is clear that the building is not in full operation seven days a week as the consumption is substantially lower on some days compared to others. Since the detail of the operating schedule is not disclosed with the dataset, it is impossible to generate a model result as good as the commercial supermarket that opens seven days a week.

Table 4.11: Comparison of the energy consumption prediction between Fox Retail Manie and the private supermarket dataset at an hourly resolution.

Building	Trial	RMSE	MAE	MAPE(%)
Fox retail Manie	M1	13.97	10.25	14.83
Private Supermarket	M1	22.54	16.52	4.88

On the other hand, in the timestamp prediction with MLP modelling as seen in Table 4.12, the results are on par with the commercial supermarket. This is also contributed by two factors. First, even though the energy consumption prediction deviates from the actual consumption, the general trend per day remains similar. Second, Figure 4.5 shows that the real indexes that need to be predicted have a unimodal distribution, as compared to a bimodal distribution in the private commercial supermarket dataset. This feature makes the model easier to train, thus leading to improvements in the HR metrics.

Table 4.12: Comparison of the timestamp prediction between Fox Retail Manie and the private supermarket dataset at an hourly resolution.

Building	Trial	Index	RMSE	MAE	HR-1 (%)	HR-2 (%)	Minutes
Fox retail Manie	M1	Start	3.57	2.76	36.25	56.25	166
		Peak	2.79	2.13	43.75	66.25	128
		End	4.24	3.01	38.75	56.25	181
Private Supermarket	M1	Start	3.42	2.86	32.46	42.86	172
		Peak	3.09	2.47	30.52	52.60	148
		End	3.33	2.72	32.47	53.91	163

4.6 Summary

In this Chapter, the proposed EPTP framework is evaluated with three experiments using real-world commercial building data. The performance of the model in all three experiments is summarized as the following:

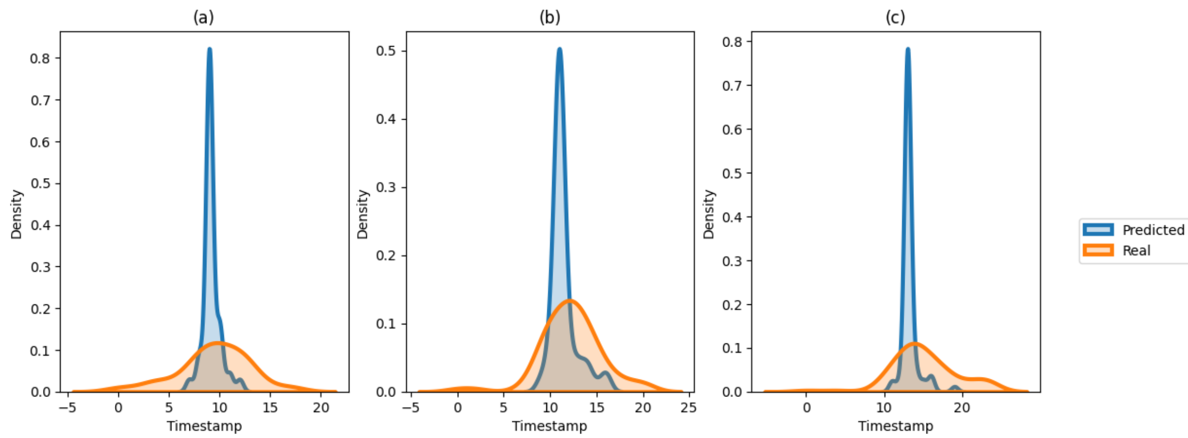


Figure 4.5: Comparison of the density plot for Fox Retail Manie test set between real indexes and predicted indexes on an hourly resolution: (a) starting index, (b) peaking index, (c) ending index.

- The purposed EPTP framework performed well with data of high resolution. The model is able to generate 86% 2-hour hit rate on the peak index with the 15-minute resolution data. The proposed EPTP framework results in 62 minutes deviation from the real peak index.

- The purposed EPTP framework produces comparable results for multiple commercial building datasets. The experiments with the Fox Retail Manie building resulted in an MAE error of 2.13 on the peak index and 2.47 for the commercial supermarket using hourly resolution. The difference is primary driven by the density distribution of the energy consumption where the Fox Retail Manie building shows an unimodal distribution on the index value and the commercial supermarket shows a bimodal distribution.

- The proposed EPTP framework outperforms the SVR model for energy consumption prediction followed by MLP for timestamp prediction. Although SVR is a common method in existing research for predicting energy consumption in commercial buildings, the LSTM model better captures the fluctuations in high-frequency data. It is also validated that using LSTM alone with BM to extract the index is insufficient for decreasing the MAE error.

Chapter 5

Conclusion and Future Work

5.1 Conclusions and Contributions

Energy consumption management has been a rising concern due to supply difficulties and environmental consequences. Peak demand prediction aims to improve such issues to balance electricity supply and demand to maximize power benefits. Thus, accurate peak predictions help with peak shaving and load smoothing and therefore lead to efficient power storage scheduling and demand flexibility. To achieve the goal of peak predictions, the area of interest is not only energy consumption intensity but also the timestamp for which the peak happens. However, existing methods in peak demand prediction for commercial applications show a lack of focus on the timestamp matter.

Consequently, this thesis proposes to use a three-phase Energy Peaks and Timestamping Prediction (EPTP) framework to predict not only the peak consumption but also the corresponding starting, peaking, and ending indexes. In the first phase of the proposed EPTP framework, energy consumption data and weather information is acquired and preprocessed for the deep learning model. In addition, the timestamps for the peaks are labeled with starting, peaking, and ending indexes. The second phase uses LSTM to predict the next day energy consumption for the commercial application. Finally, predicted energy consumption is passed to the MLP model for the output of timestamps. The performance of the model is evaluated using the label from the real data in the first phase.

The proposed EPTP framework is evaluated with three experiments using a commercial supermarket located in Quebec, Canada, and open-source low-frequency data. In the first experiment, the EPTP framework is examined with commercial supermarket data using different frequencies. It concludes that the model results in better performance with higher frequency. In the second experiment, the framework is trained with a retail building in the BDG2 dataset. Although the retail building is exposed to fluctuations due to operation patterns, the predictions of the index values result in similar accuracy to commercial supermarket with the same frequency. The higher hit rate with the retail building is contributed by the unimodal distribution of the real peaks. In the last experiment, the timestamp prediction is compared to two baselines. The first baseline is established by using SVR for the energy consumption prediction, followed by MLP for index modelling. The second baseline is extracted using the most common method in peak demand prediction: block maxima. The results reveal that the proposed model shows better performance, except for the 30-minute resolution using block maxima.

Overall, the proposed EPTP framework is capable of predicting the timestamps for starting, peaking, and ending indexes, especially with high-frequency data. On average, the predicted starting, peaking, and ending indexes have a deviation of 1.5 hours, 1 hour, and 1.25 hours from the real indexes respectively based on the 15-minute resolution. The predicted indexes provide building managers with easy-to-understand results and therefore proactively manage energy use.

5.2 Limitations and Future Research

Future work will mainly be focused on improving the performance of the proposed EPTP framework. The future work includes:

- As discussed, the proposed EPTP framework is initially established to work with high-resolution data, i.e.: energy consumption dataset with 15-minute intervals. It shows degradation in performance with stretched resolution. Therefore, areas of improvement could work around improving index prediction accuracy in low-resolution data.
- During the benchmarking process of using the open-sourced dataset, the proposed ETPT framework shows areas of improvement in energy consumption prediction with varied operating patterns. Therefore, the LSTM model could be further improved for better accuracy during non-business days.
- Investigate whether including historical peak information and store operation features, such as occupancy and production, could improve the timestamp prediction results. The current model uses only commercial building consumption, weather information, and cyclical engineered variables. Hence, further investigation is needed to improve the prediction, especially with low resolution.
- Investigate whether computing with higher resolution could further improve the accuracy of the model. The private commercial supermarket dataset is currently computed with 1-hour, 30-minute, and 15-minute resolutions. The results show that accuracy improves with shortened interval. Hence, further modelling is needed to test out the framework performance with even higher frequency.
- Edge computing refers to the execution of machine learning tasks closer to the source of the data itself. Implementing edge computing reduces drawbacks such as high latency, security, and privacy issues, as well as power wastage. Future research also focuses on

applying the model in real-time with Edge devices such that it is more accessible to the commercial industry.

Bibliography

[1] K. K. Wan, D. H. Li, W. Pan, J. C. Lam, Impact of climate change on building energy use in different climate zones and mitigation and adaptation implications, Applied Energy 97 (2012) 274–282, energy Solutions for a Sustainable World - Proceedings of the Third International Conference on Applied Energy, May 16-18, 2011 - Perugia, Italy. doi:https://doi.org/10.1016/j.apenergy.2011.11.048.

[2] M. Santamouris, Innovating to zero the building sector in europe: Minimising the energy consumption, eradication of the energy poverty and mitigating the local climate change, Solar Energy 128 (2016) 61–94, special issue: Progress in Solar Energy. doi:https://doi.org/10.1016/j.solener.2016.01.021.

[3] U. Berardi, A cross-country comparison of the building energy consumptions and their trends, Resources, Conservation and Recycling 123 (2017) 230–241. doi:https://doi.org/10.1016/j.resconrec.2016.03.014.

[4] IEA. Buildings, Technical report, IEA (2022).

[5] A. Allouhi, Y. El Fouih, T. Kousksou, A. Jamil, Y. Zeraouli, Y. Mourad, Energy consumption and efficiency in buildings: current status and future trends, Journal of Cleaner Production 109 (2015) 118–130, special Issue: Toward a Regenerative Sustainability Paradigm for the Built Environment: from vision to reality. doi:https://doi.org/10.1016/j.jclepro.2015.05.139.

[6] J. C. Lam, K. K. Wan, C. Tsang, L. Yang, Building energy efficiency in different climates, Energy Conversion and Management 49 (8) (2008) 2354–2366. doi:https://doi.org/10.1016/j.enconman.2008.01.013.

[7] L. Pérez-Lombard, J. Ortiz, C. Pout, A review on buildings energy consumption information, Energy and Buildings 40 (3) (2008) 394–398. doi:https://doi.org/10.1016/j.enbuild.2007.03.007.

[8] S. Mahjoub, L. Chrifi-Alaoui, B. Marhic, L. Delahoche, Predicting energy consumption using lstm, multi-layer gru and drop-gru neural networks, Sensors 22 (11) (2022). doi:10.3390/s22114062.

[9] K. Yan, X. Wang, Y. Du, N. Jin, H. Huang, H. Zhou, Multi-step short-term power consumption forecasting with a hybrid deep learning strategy, Energies 11 (11) (2018) 3089.

[10] M. Elsaraiti, G. Ali, H. Musbah, A. Merabet, T. Little, Time series analysis of electricity consumption forecasting using arima model, in: 2021 IEEE Green Technologies Conference (GreenTech), 2021, pp. 259–262. doi:10.1109/GreenTech48523.2021.00049.

[11] N. Amjady, Short-term hourly load forecasting using time-series modeling with peak load estimation capability, IEEE Transactions on Power Systems 16 (4) (2001) 798–805. doi:10.1109/59.962429.

[12] S. B. Taieb, J. W. Taylor, R. J. Hyndman, Hierarchical probabilistic forecasting of electricity demand with smart meter data, Journal of the American Statistical Association 116 (533) (2021) 27–43. doi:10.1080/01621459.2020.1736081.

[13] S. Ai, A. Chakravorty, C. Rong, Evolutionary ensemble lstm based household peak demand prediction, in: 2019 International Conference on Artificial Intelligence in Information and Communication (ICAIIC), 2019, pp. 1–6. doi:10.1109/ICAIIC.2019.8668971.

[14] M. Aouad, H. Hajj, K. Shaban, R. A. Jabr, W. El-Hajj, A cnn-sequence-to-sequence network with attention for residential short-term load forecasting, Electric Power Systems Research 211 (2022) 108152. doi:https://doi.org/10.1016/j.epsr.2022.108152.

[15] A. Satre-Meloy, M. Diakonova, P. Grünewald, Cluster analysis and prediction of residential peak demand profiles using occupant activity data, Applied Energy 260 (2020) 114246. doi:https://doi.org/10.1016/j.apenergy.2019.114246.

[16] L. M. Saini, Peak load forecasting using bayesian regularization, resilient and adaptive backpropagation learning based artificial neural networks, Electric Power Systems Research 78 (7) (2008) 1302–1310. doi:https://doi.org/10.1016/j.epsr.2007.11.003.

[17] H. Lu, F. Cheng, X. Ma, G. Hu, Short-term prediction of building energy consumption employing an improved extreme gradient boosting model: A case study of an intake tower, Energy 203 (2020) 117756. doi:https://doi.org/10.1016/j.energy.2020.117756.

[18] Y. Amara-Ouali, M. Fasiolo, Y. Goude, H. Yan, Daily peak electrical load forecasting with a multi-resolution approach, International Journal of Forecasting 39 (3) (2023) 1272–1286. doi:https://doi.org/10.1016/j.ijforecast.2022.06.001.

[19] M. E. Lebotsa, C. Sigauke, A. Bere, R. Fildes, J. E. Boylan, Short term electricity demand forecasting using partially linear additive quantile regression with an application to the unit commitment problem, Applied Energy 222 (2018) 104–118. doi:https://doi.org/10.1016/j.apenergy.2018.03.155.

[20] J. Boano-Danquah, C. Sigauke, K. Kyei, Analysis of extreme peak loads using point processes: An application using south african data, IEEE Access PP (2020) 1–1. doi:10.1109/ACCESS.2020.3015259.

[21] S. Tassou, Y. Ge, A. Hadawey, D. Marriott, Energy consumption and conservation in food retailing, Applied Thermal Engineering 31 (2) (2011) 147–156. doi:https://doi.org/10.1016/j.applthermaleng.2010.08.023.

[22] L. Timma, R. Skudritis, D. Blumberga, Benchmarking analysis of energy consumption in supermarkets, Energy Procedia 95 (2016) 435–438, international Scientific Conference “Environmental and Climate Technologies”, CONECT 2015. doi:https://doi.org/10.1016/j.egypro.2016.09.056.

[23] M. W. Gardner, S. Dorling, Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences, Atmospheric environment 32 (14-15) (1998) 2627–2636.

[24] A. Sherstinsky, Fundamentals of recurrent neural network (rnn) and long short-term memory (lstm) network, Physica D: Nonlinear Phenomena 404 (2020) 132306. doi:https://doi.org/10.1016/j.physd.2019.132306.

[25] S. Hochreiter, J. Schmidhuber, Long Short-Term Memory, Neural Computation 9 (8) (1997) 1735–1780. doi:10.1162/neco.1997.9.8.1735.

[26] D. Soutner, L. Müller, Application of lstm neural networks in language modelling, in: Text, Speech, and Dialogue, Springer Berlin Heidelberg, 2013, pp. 105–112.

[27] D. Li, J. Qian, Text sentiment analysis based on long short-term memory, in: 2016 First IEEE International Conference on Computer Communication and the Internet (ICCCI), 2016, pp. 471–475. doi:10.1109/CCI.2016.7778967.

[28] J. Cao, Z. Li, J. Li, Financial time series forecasting model based on ceemdan and lstm, Physica A: Statistical Mechanics and its Applications 519 (2019) 127–139. doi:https://doi.org/10.1016/j.physa.2018.11.061.

[29] V. K. R. Chimmula, L. Zhang, Time series forecasting of covid-19 transmission in canada using lstm networks, Chaos, Solitons Fractals 135 (2020) 109864. doi:https://doi.org/10.1016/j.chaos.2020.109864.

[30] J. Jang, J. Han, S.-B. Leigh, Prediction of heating energy consumption with operation pattern variables for non-residential buildings using lstm networks, Energy and Buildings 255 (2022) 111647. doi:https://doi.org/10.1016/j.enbuild.2021.111647.

[31] M. Feurer, F. Hutter, Hyperparameter Optimization, Springer International Publishing, 2019, pp. 3–33. doi:10.1007/978-3-030-05318-5_1.

[32] J. Heaton, Introduction to neural networks with Java, Heaton Research, Inc., 2008.

[33] S. Sharma, S. Sharma, A. Athaiya, Activation functions in neural networks, Towards Data Sci 6 (12) (2017) 310–316.

[34] M. Kuhn, K. Johnson, et al., Applied predictive modeling, Vol. 26, Springer, 2013.

[35] D. C. Montgomery, Design and analysis of experiments, eight edition, Joh Wiley and Sons (2013).

[36] J. Bergstra, Y. Bengio, Random search for hyper-parameter optimization., Journal of machine learning research 13 (2) (2012).

[37] B. Shahriari, K. Swersky, Z. Wang, R. P. Adams, N. De Freitas, Taking the human out of the loop: A review of bayesian optimization, Proceedings of the IEEE 104 (1) (2015) 148–175.

[38] Y. Demirel, 5.2 energy conservation, in: I. Dincer (Ed.), Comprehensive Energy Systems, Elsevier, Oxford, 2018, pp. 45–90. doi:https://doi.org/10.1016/B978-0-12-809597-3.00505-8.

[39] M. Glavan, D. Gradišar, S. Moscariello, ani Juričić, D. Vrančić, Demand-side improvement of short-term load forecasting using a proactive load management–a supermarket use case, Energy and Buildings 186 (2019) 186–194. doi:https://doi.org/10.1016/j.enbuild.2019.01.016. URL `https://www.sciencedirect.com/science/article/pii/S0378778818327014`

[40] E. Mocanu, P. H. Nguyen, M. Gibescu, W. L. Kling, Deep learning for estimating building energy consumption, Sustainable Energy, Grids and Networks 6 (2016) 91–99. doi:https://doi.org/10.1016/j.segan.2016.02.005.

[41] I. K. Nti, M. Teimeh, O. Nyarko-Boateng, A. F. Adekoya, Electricity load forecasting: a systematic review, Journal of Electrical Systems and Information Technology 7 (13) (2020). doi:10.1186/s43067-020-00021-8.

[42] B. Yildiz, J. Bilbao, A. Sproul, A review and analysis of regression and machine learning models on commercial building electricity load forecasting, Renewable and Sustainable Energy Reviews 73 (2017) 1104–1122. doi:https://doi.org/10.1016/j.rser.2017.02.023.

[43] B. Nepal, M. Yamaha, A. Yokoe, T. Yamaji, Electricity load forecasting using clustering and arima model for energy management in buildings, JAPAN ARCHITECTURAL REVIEW 3 (1) (2020) 62–76. doi:https://doi.org/10.1002/2475-8876.12135.

[44] C.-H. Chen, N. Guo, W. Chen, M. Wang, Z. Tian, H. Jin, Appling an improved method based on arima model to predict the short-term electricity consumption transmitted by the internet of things (iot), Wireless Communications and Mobile Computing (2021). doi:10.1155/2021/6610273.

[45] G.-F. Fan, L.-L. Peng, X. Zhao, W.-C. Hong, Applications of hybrid emd with pso and ga for an svr-based load forecasting model, Energies 10 (11) (2017). doi:10.3390/en10111713.

[46] W.-C. Hong, G.-F. Fan, Hybrid empirical mode decomposition with support vector regression model for short term load forecasting, Energies 12 (6) (2019). doi:10.3390/en12061093.

[47] F. Zhang, C. Deb, S. E. Lee, J. Yang, K. W. Shah, Time series forecasting for building energy consumption using weighted support vector regression with differential evolution optimization technique, Energy and Buildings 126 (2016) 94–103. doi:https://doi.org/10.1016/j.enbuild.2016.05.028.

[48] F. Pallonetto, C. Jin, E. Mangina, Forecast electricity demand in commercial building with machine learning models to enable demand response programs, Energy and AI 7 (2022) 100121. doi:https://doi.org/10.1016/j.egyai.2021.100121.

[49] H. Fu, J.-C. Baltazar, D. E. Claridge, Review of developments in whole-building statistical energy consumption models for commercial buildings, Renewable and Sustainable Energy Reviews 147 (2021) 111248. doi:https://doi.org/10.1016/j.rser.2021.111248.

[50] C. Fan, F. Xiao, S. Wang, Development of prediction models for next-day building energy consumption and peak power demand using data mining techniques, Applied Energy 127 (2014) 1–10. doi:https://doi.org/10.1016/j.apenergy.2014.04.016.

[51] S. Taheri, A. Razban, A novel probabilistic regression model for electrical peak demand estimate of commercial and manufacturing buildings, Sustainable Cities and Society 77 (2022) 103544. doi:https://doi.org/10.1016/j.scs.2021.103544.

[52] Y. T. Chae, R. Horesh, Y. Hwang, Y. M. Lee, Artificial neural network model for forecasting sub-hourly electricity usage in commercial buildings, Energy and Buildings 111 (2016) 184–194. doi:https://doi.org/10.1016/j.enbuild.2015.11.045.

[53] C. Gilbert, J. Browell, B. Stephen, Probabilistic load forecasting for the low voltage network: Forecast fusion and daily peaks, Sustainable Energy, Grids and Networks 34 (2023) 100998. doi:https://doi.org/10.1016/j.segan.2023.100998.

[54] A. Soman, A. Trivedi, D. Irwin, B. Kosanovic, B. McDaniel, P. Shenoy, Peak forecasting for battery-based energy optimizations in campus microgrids, in: Proceedings of the Eleventh ACM International Conference on Future Energy Systems, Association for Computing Machinery, New York, NY, USA, 2020, p. 237–241. doi:10.1145/3396851.3397751.

[55] M. Braun, H. Altan, S. Beck, Using regression analysis to predict the future energy consumption of a supermarket in the uk, Applied Energy 130 (2014) 305–313. doi:https://doi.org/10.1016/j.apenergy.2014.05.062.

[56] A. Blázquez-García, A. Conde, U. Mori, J. A. Lozano, A review on outlier/anomaly detection in time series data, ACM Comput. Surv. 54 (3) (apr 2021). doi:10.1145/3444690.

[57] R. Pearson, Y. Neuvo, J. Astola, M. Gabbouj, Generalized hampel filters, EURASIP Journal on Advances in Signal Processing 2016 (08 2016). doi:10.1186/s13634-016-0383-6.

[58] L. Davies, U. Gather, The identification of multiple outliers, Journal of the American Statistical Association 88 (423) (1993) 782–792. doi:10.1080/01621459.1993.10476339.

[59] R. Pearson, Outliers in process modeling and identification, IEEE Transactions on Control Systems Technology 10 (1) (2002) 55–63. doi:10.1109/87.974338.

[60] S. García, J. Luengo, F. Herrera, Data Preparation Basic Models, Springer International Publishing, Cham, 2015, pp. 39–57. doi:10.1007/978-3-319-10247-4_3.

[61] A. Ferreira, L. Haan, On the block maxima method in extreme value theory: Pwm estimators, The Annals of Statistics 43 (10 2013). doi:10.1214/14-AOS1280.

[62] H. Alla, L. Moumoun, Y. Balouki, A multilayer perceptron neural network with selective-data training for flight arrival delay prediction, Hindawi Scientific Programming (2021). doi:https://doi.org/10.1155/2021/5558918.

[63] Y. Jiang, Y. Liu, D. Liu, H. Song, Applying machine learning to aviation big data for flight delay prediction, in: 2020 IEEE Intl Conf on Dependable, Autonomic and Secure Computing, Intl Conf on Pervasive Intelligence and Computing, Intl Conf on Cloud and Big Data Computing, Intl Conf on Cyber Science and Technology Congress (DASC/PiCom/CBDCom/CyberSciTech), 2020, pp. 665–672. doi:10.1109/DASC-PICom-CBDCom-CyberSciTech49142.2020.00114.

[64] C. Miller, A. Kathirgamanathan, B. Picchetti, P. Arjunan, J. Y. Park, Z. Nagy, P. Raftery, B. W. Hobson, Z. Shi, F. Meggers, The building data genome project 2, energy meter data from the ashrae great energy predictor iii competition, Scientific Data 7 (368) (2020). doi:10.1038/s41597-020-00712-x.

[65] C. Miller, P. Arjunan, A. Kathirgamanathan, C. Fu, J. Roth, J. Y. Park, C. Balbach, K. Gowri, Z. Nagy, A. D. Fontanini, J. Haberl, The ashrae great energy predictor iii competition: Overview and results, Science and Technology for the Built Environment 26 (10) (2020) 1427–1447. doi:10.1080/23744731.2020.1795514.

[66] J. Xue, Z. Xu, J. Watada, Building an integrated hybrid model for short-term and mid-term load forecasting with genetic optimization, International Journal of Innovative Computing, Information and Control 8 (2012) 7381–7391.

Curriculum Vitae

Name:	Mengchen Zhao
Post-Secondary Education and Degrees:	University of Western Ontario London, Ontario, Canada 2021 - 2023 M.E.Sc, Electrical and Computer Engineering Collaborative Specialization in Artificial Intelligence
	University of Western Ontario London, Ontario, Canada 2016 - 2021 B.E.Sc Mechatronic Systems Engineering
	University of Western Ontario London, Ontario, Canada 2016 - 2021 B.A Honors Business Administration
Honours and Awards:	Undergraduate Research Dean's Award 2021
Related Work Experience:	Mitacs Accelerate Program 2022 - 2023
	Graduate Teaching Assistant The University of Western Ontario 2021 - 2023

Publications:

Zhao, M., Sadhu, A. and Capretz, M. (2022). Multiclass anomaly detection in imbalanced structural health monitoring data using convolutional neural network. Journal of Infrastructure Preservation and Resilience. 3(10).

Zhao, M., Sadhu, A. and Capretz, M. (2022, August). Deep Learning-based Anomaly Detection in Structural Health Monitoring. In 11th International Conference on Structural Health Monitoring of Intelligent Infrastructure (SHMII-11).