Article

Short and Very Short Term Firm-level Load Forecasting for Warehouses: A Comparison of Machine Learning and Deep Learning Models

Andrea Maria N. C. Ribeiro¹, Pedro Rafael X. do Carmo², Patricia Takako Endo³, Pierangelo Rosati⁴, Theo Lynn⁵

¹ Universidade Federal de Pernambuco; amnc@cin.ufpe.br

² Universidade Federal de Pernambuco; prxc@cin.ufpe.br

³ Universidade de Pernambuco; patricia.endo@upe.br

⁴ Irish Institute of Digital Business, Dublin City University; pierangelo.rosati@dcu.ie

⁵ Irish Institute of Digital Business, Dublin City University; theo.lynn@dcu.ie

Abstract: Commercial buildings are a significant consumer of energy worldwide. Logistics facilities, and specifically warehouses, are a common building type yet under-researched in the demand-side energy forecasting literature. Warehouses have an idiosyncratic profile when compared to other commercial and industrial buildings with a significant reliance on a small number of energy systems. As such, warehouse owners and operators are increasingly entering in to energy performance contracts with energy service companies (ESCOs) to minimise environmental impact, reduce costs, and improve competitiveness. ESCOs and warehouse owners and operators require accurate forecasts of their energy consumption so that precautionary and mitigation measures can be taken. This paper explores the performance of three machine learning models (Support Vector Regression (SVR), Random Forest, and Extreme Gradient Boosting (XGBoost)), three deep learning models (Recurrent Neural Networks (RNN), Long Short-Term Memory (LSTM), and Gated Recurrent Unit (GRU)), and a classical time series model, Autoregressive Integrated Moving Average (ARIMA) for predicting daily energy consumption. The dataset comprises 8,040 records generated over an 11-month period from January to November 2020 from a non-refrigerated logistics facility located in Ireland. The grid search method was used to identify the best configurations for each model. The proposed XGBoost models outperform other models for both very short load forecasting (VSTLF) and short term load forecasting (STLF); the ARIMA model performed the worst.

Keywords: Very short term load forecasting; VSTLF; Short term load forecasting; STLF; deep learning; RNN; LSTM; GRU; machine learning; SVR; random forest; extreme gradient boosting; energy consumption; ARIMA; time series prediction.

1. Introduction

The commercial sector is a large and growing consumer of delivered energy worldwide [1]. The world's stock of warehouses was estimated at 150,000 warehouses in 2020, approximately 2.3 billion square metres of space [2]. Driven by growing global e-commerce consumption, the number of units is set to grow to 180,000 warehouses by 2025 [2]. Warehouses have a distinct energy consumption profile concentrated in a few key systems - lighting (71%), heating and ventilation (16%), battery charging (7%), and other miscellaneous energy consumption [3]. Warehouses are not only an important part of the commercial building sector but are embedded elements of supply chains thus there are a wide range of drivers of greater energy efficiency including corporate social responsibility, legal, competitive, and cost factors [4]. While the relative percentage of greenhouse gases (GHGs) from logistics facilities is small at 0.55%, this still represents over 300 megatonnes of GHG emissions per year [5]. This is particularly salient because electricity costs and usage can be reduced dramatically through a number of small interventions [3]. For example, the Carbon Trust estimate non-refrigerated

warehouses operating with legacy lighting can typically reduce electricity costs by 70% by moving to LED, while an investment in solar photovoltaic (PV) has an estimated payback of 8.8 years [3]. As such, green warehousing can be a significant first step towards both net zero warehousing and supply chain decarbonisation.

Energy load forecasting techniques are essential for effective energy management in both residential and commercial buildings [6,7]. Energy load forecasting applications can be categorised into four main categories based on the time length of the prediction (forecasting horizon): (i) very short-term load forecasting (VSTLF) (minutes to hours), (ii) short-term load forecasting (STLF) (hours to days), (iii) medium term load forecasting (weeks to months), and (iv) long-term load forecasting (months to years) [7,8]. Similarly, load forecasting applications can be classified as either supply- or demand-side based on whether the focus is on energy production or consumption [9]. This paper specifically focuses on short and very short term demand-side load forecasting. While load forecasting has been one of the main research topics in electrical engineering for more than three decades [8], STLF and VSTLF has only become possible in recent years thanks to widespread adoption of Advanced Metering Infrastructure (AMI) and connected sensors that are able to capture electrical consumption at a high level of granularity [7,10]. This real-time, fine-grain consumption data at the point of use in buildings, and associated analysis, can be used by building energy managers and end users for planning and end user behaviour change [11].

The optimisation of energy consumption is not only an important design factor in warehouse operations and intralogistics [4], but is a critical element in the energy performance contracting (EPC)-based business models, a key strategy in combating climate change [12,13]. EPC involves the outsourcing of one or more energy-related services to a third party, typically an ESCO [13,14]. Under an EPC arrangement, the ESCO "implements a project to deliver energy savings, or a renewable energy project, and uses the stream of income from the expense reduction, or the renewable energy produced, to repay the whole or part of the costs of the project, including the costs of the investment" [13]. Through EPC, the ESCO establishes a link between contract payments and equipment performance over a long term period, typically based on energy performance and associated energy and cost savings [13,14]. Warehouses are an ideal target for ESCOs and energy service contracts due to the concentration of energy consumption in a small number of systems, ideally placed for outsourcing i.e. lighting and heating, but also due to their suitability for solar PV installations. While lighting and heating are predictable, energy demand management may be required to mitigate the impact of other elements, e.g. plug-in electric vehicles and other energy storage units. Near-term electricity load forecasting can help in the design of energy performance contracts, building a business case for green warehousing, controlling building energy systems and managing the charging/discharging or energy storage units in an energy-efficient and cost-effective way.

Extant literature on STLF and VSTLF has typically (i) focused on supply-side perspectives, (ii) aggregated energy costs, and (iii) failed to recognise the idiosyncrasies of warehouses. Furthermore, the use of STLF and VSTLF has not been considered from an EPC motivation. We argue more accurate load forecasting allows warehouse operators and ESCOs make better decisions to inform their investment decisions with respect to equipment and renewable energy systems. There is a paucity of studies in demand-side process-related (V)STLF using deep learning and machine learning for warehouses. The limited studies that have been published do not compare deep learning performance against widely used machine learning models, classical time series models, or approaches used in practice. Similarly, few articles compare performance between VSTLF and STLF. In addition to proposing prediction models, we also address this gap.

In this paper, we focus on performance analyses of deep learning and machine learning models for building-level STLF and VSTLF in a non-refrigerated logistics facility. We use energy consumption data from a real multinational warehouse located in Ireland. We propose three deep learning models - simple Recurrent Neural Networks (RNN), Long Short-Term Memory (LSTM) networks, and Gated Recurrent Unit (GRU) - and three machine learning models - Support Vector Regression (SVR), Random

Forest, and Extreme Gradient Boosting - for predicting daily energy consumption. We use the grid search method to identify the best model configurations. We compare the performance of the deep learning and machine learning models for predicting the energy consumption of the next hour using data from the previous 48 hours (VSTLF) using (i) common metrics – root mean squared error (RMSE), mean absolute percentage error (MAPE), and mean absolute error (MAE), and (ii) a classical time series model, Autoregressive Integrated Moving Average (ARIMA). The best performing models for VSTLF, XGBoost-5-100 and XGBoost-7-100, were further evaluated for predicting energy consumption over longer time horizons, i.e., 12 hours and 24 hours.

The remainder of this paper is organised as follows. Section 2 presents the description of the data, pre-processing, and the evaluation metrics used in our work. Section 4 presents the results of our analysis. Section 3 presents the models identified for evaluation. Section 5 discusses related works in the field of STLF and VSTLF using machine learning and deep learning for warehouse facilities. The paper concludes with a summary of the paper and future avenues for research in Section 6.

2. Material and method

2.1. Dataset

The data used in this study was sourced from an ESCO that offers services to warehouse owners and operators worldwide. Specifically, this ESCO specialises in the replacement of legacy lighting systems with LED lighting and intelligent controls and the installation and generation of electricity through Solar PV. The ESCO operates an EPC arrangement with customers and thus generates its income from reducing the energy consumption and energy costs of client facilities. The dataset comprises 8,040 records of hourly energy consumption over an 11-month period from 1 January 2020 to 30 November 2020 for a non-refrigerated logistics facility located in Ireland. Figure 1 presents the time series of the dataset used in this work while Table 1 presents some descriptive and quantile statistics.

Figure 1. Time series overview of the energy consumption of the focal site (January - November 2020).

Table 1. Quantile and descriptive statistics.

Quantile statistics		Descriptive statistics	
Description	Values	Description	Values
Minimum	0.00	Standard deviation	4.18
Maximum	13.67	Coefficient of variation	0.63
Median	7.69	Mean	6.68
Range	13.67	Median Absoluta Deviation	3.26
Interquartile range	9.02	Variance	17.44

2.2. Data preprocessing

An initial analysis of the dataset revealed that there were no missing values and measurement errors, thus it was not necessary to perform any data cleaning. However, it was necessary to normalize the data so that all inputs to the model had equal weights and a similar range. This was also necessary to reduce forecast errors and training process time. [15]. Normalization reduces the data range to

zero and one [0, 1]. Sklearn's MinMaxScaler function was used to normalize the data in this study as presented in Equation 1.

$$X'_i = \frac{X_i - \min(x)}{\max(x) - \min(x)} \quad (1)$$

Where X'_i is the rescaled value; X_i is the original value; $\min(x)$ is the minimum value of feature x ; $\max(x)$ is the maximum value of feature x .

2.3. Evaluation metric

In this study, we adopt Root mean squared error (RMSE), mean absolute percent error (MAPE), and mean absolute error (MAE) to evaluate and compare the performance of different models as they are the most commonly metrics used in the evaluation the accuracy of energy consumption models [16].

RMSE is defined as the square root of the mean squared error (MSE) [17] (Equation 2), that is, the root of the mean square error of the difference between the prediction (P_i) and the real value (R_i), where n represents the sample size. RMSE is particularly sensitive to outliers as it squares the difference between the predicted value and the observed value. RMSE presents error values in the same scale as the original variable [17] and it has been widely applied in time series analysis [18].

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (P_i - R_i)^2} \quad (2)$$

MAPE is widely used for evaluating prediction models when high-quality forecast is required and is used in numerous energy consumption forecasting studies [19–24]. MAPE is calculated as presented in Equation 3 [19], and expresses the accuracy of the error as a percentage. It can be applied in a wide range of contexts, as it is relatively intuitive to interpret but it can only be used when observed values in the dataset are not equal zero [25].

$$MAPE = \frac{100\%}{n} \sum_{i=1}^n \left| \frac{P_i - R_i}{R_i} \right| \quad (3)$$

MAE is defined as per Equation 4 [21,26]. In contrast to RMSE and MAPE, MAE depends on the scale of the data and it is not sensitive to outliers as it treats all errors (both positive and negative) in the same way. In this study, we use MAE to quantify a model's ability to predict energy consumption.

$$MAE = \frac{1}{n} \sum_{i=1}^n |P_i - R_i| \quad (4)$$

These metrics are commonly used in previous studies related to load forecasting (see, for example, [20], [27], [28], and [9]) and specifically in relation to short term load forecasting at commercial building levels [7,29].

3. Finding models to predict energy consumption

The forecasting techniques used to perform VSTLF and STLF mainly come from the statistics and computational intelligence domains [6]. Regression methods belong to the former and typically assume that a linear relationship exists between load prediction and selected explanatory variables. These models have attracted attention as they are relatively easy to implement and interpret but they return large forecasting errors when an unexpected change in the input variables occurs [30]. Other statistical approaches that have been used for STLF include traditional and adaptive autoregressive moving average [31–33] and stochastic time series [34,35] but they are all subject to the same limitations as regression models.

Machine learning techniques perform significantly better than statistical techniques, particularly when non-linear trends and patterns are present in the data as in the dataset used in this study (see Figure 1). Traditional machine learning models include, for example, regression trees [36], support vector regression (SVR) [37–39], random forest [40], extreme gradient boosting (XGBoost) [41,42] and Artificial Neural Networks (ANNs) [43,44]. More recently, the development of deep learning methods have provided further performance improvements thanks to their capacity to extract a variety of features from large datasets [7]. These models include (i) Convolutional Neural Networks (CNN) which perform particularly well in terms of feature extraction and generalisation [45], and (ii) Recurrent Neural Networks (RNN) which use information and patterns embedded in the time series itself to perform tasks that other ANNs are unable to do [9]. While deep learning models tend to outperform more traditional machine learning algorithms, they are not without limitations. RNNs, for example, struggle with long-term dependencies due to vanishing and exploding gradient issues during training [46,47]. More recently, alternative RNN models have been developed, namely Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) which aim to overcome such limitation [48–50]. LSTM algorithms are able to store the useful information of long-span states and have achieved significantly better performance than other RNNs when predicting energy demand [24,51]. GRU models are simpler than LSTM models as they use only a single gate named update gate and tend to be faster to train [52]. Similarly to LSTM, GRU models have been used for STLTF with excellent performance [53,54].

In this paper we assess and compare the performance of different machine learning and deep learning models for building-level STLTF and VSTLTF. More specifically, we will compare SVR, Random Forest and XGBoost models with RNN, LSTM and GRU. We will also use an AutoRegressive Integrated Moving Average (ARIMA) model as our baseline benchmark. Each of these models will be presented in more details in the next sections. All our models use energy consumption data from the previous 2 days (48 hours) as input and have a single output, the predicted energy consumption for the next hour.

3.1. Machine learning models

For this study, we selected SVR, Random Forest and XGBoost as suitable machine learning models as they have been implemented in a number of STLTF studies (see, for example, [38], [39], [40], [42]).

While ANNs have become extremely popular in the load forecasting literature over the last decade, they face significant challenges in when it come to STLTF and VSTLTF real-life applications mostly due to model overfitting and the exponential increase in complexity associated with high dimensionality [37]. SVR is a regression technique based on Support Vector Machines (SVM) [55] (a machine-learning technique based on statistical learning theory [56]) and has demonstrated to perform well in forecasting time series [57]. SVR models perform a linear regression in the high-dimensional feature space created by a kernel function using a epsilon-insensitive loss function while also minimising the model coefficients to reduce complexity [55].

Random Forest models were initially proposed by Breiman [58] as potential solution to the generalisability and and overfitting issues typical of decision trees [59]. Random Forest models are based on the Bagging ensemble learning theory [60] and the random subspace method [61]. They integrate several weak classification decision trees into a strong classifier. More specifically, each decision tree generates an independent classification and the final result is the one that received the majority of the votes among all the decision trees [62,63].

Similarly to Random Forest, XGBoost models are based on ensemble learning theory which combine gradient boosting decision trees models and second-order Taylor expansion on the loss function to speed up the optimisation process while avoiding overfitting [64]. XGBoost models also support parallel processing and therefore are faster to train and deploy than traditional decision trees.

In order to identify the best hyperparameters for each model, we performed a grid search. The parameters and levels used for the grid search are listed in Table 2.

Table 2. Machine learning parameters and levels.

Technique	Parameter	Levels
SVR	Number of C	0.1, 1 and 10
SVR	Type of kernel	Polinomial, RBF, Sigmoid and Linear
Random Forest	Number of max. depth	From 3 to 9, step 2
Random Forest	Number of trees	From 50 to 200, step 50
XGBoost	Number of max. depth	From 3 to 9, step 2
XGBoost	Number of trees	From 50 to 200, step 50

A learning rate of 0.07 was fixed. For SVR, a gamma of 0.002, and the cost and the type of kernel were used as parameters. The maximum depth of trees and the number of trees were used for Random Forest and XGBoost. These parameters were chosen empirically.

80% of the original dataset (1 January 2020 to 24 September 2020) was used as a training dataset while the remaining 20% (25 September 2020 to 30 November 2020) was used as a test dataset using the holdout process.

Figures 2, 3 and 4 present the grid search results for RMSE, MAPE, and MAE, respectively, for each model. For SVR, the model with a configuration where the C value is 10 and uses the rbf kernel (SVR-10-rbf) generated the best results for RMSE and MAPE; the model configuration with a C value of 10 using the linear kernel (SVR-10-linear), generated the best results for MAE. For Random Forest, the best configuration across the three metrics were those with a maximum depth of nine and with 200 trees (Random Forest-9-200). For XGBoost, the models with configurations with a maximum depth of five and with 100 trees (XGBoost-5-100) generated the best results for RMSE; the model configuration with a maximum depth of seven and with 100 trees (XGBoost-7-100) generated the best results for MAPE and MAE. Consequently, these five model configurations were used in the benchmark evaluation.

Figure 2. RMSE grid search result for (a) SVR, (b) Random Forest, and (c) XGBoost.

3.2. Deep learning models

As mentioned in Section 3, we implement three deep learning models, namely RNN, LSTM and GRU. RNNs are designed to recognise patterns in sequential data streams. Such models process an input sequence at a time and maintain hidden units in their state vector that contain information

Figure 3. MAPE grid search result for (a) SVR, (b) Random Forest, and (c) XGBoost.

Figure 4. MAE grid search result for (a) SVR, (b) Random Forest, and (c) XGBoost.

about the history of all the past elements of the sequence [65]. This means that the the decision, classification, or learning done at a given time will influence the decision, classification, or learning at a subsequent time [9]. However, RNNs suffer from gradient vanishing issues which means that the weights propagated forward and backward through the layer tend to decrease, and therefore the algorithm cannot preserve long-range dependencies [46].

LSTM overcomes the main limitation of RNNs by introducing cell state and gates into RNN cells which preserve weights propagated through time and different layers [66]. More specifically, an LSTM network uses three main gates, namely a forget gate, an input gate, and an output gate. Figure 5 provides an overview of the typical LSTM architecture. The forget gate is responsible for deleting information that is no longer useful in the unit [67]. At each time step, the input $x(t)$ and the output from the previous unit ($h(t-1)$) are multiplied by the weight matrix. The result is then passed through an activation function $f(t)$ that provides a binary output that causes the data to be forgotten $c(t-1)$. The input gate instead provides useful information to the unit's status. The information is initially adjusted by the sigmoid function σ and then the $\tanh$ function is used to create a vector whose values range between -1 and +1. Finally, the output gate completes the extraction of information from the current state by applying a $\tanh$ function to a cell. Through these steps, LSTM models are able predict time series with time intervals of unknown duration [66]. However, LSTM is not without limitations. As LSTM models take a long time to train [68], GRU models are increasingly used as alternatives.

Figure 5. LSTM architecture [69]

GRUs are faster to train and are also capable of reaching comparable performance to LSTMs as they are able to capture long- and short-term dependencies in time series data [52,68]. GRUs are less complex than LSTMs as they only use two gates, namely update and reset gate [52] (see Figure 6). GRU models transfer time dependencies in the data between different time steps by a single hidden state and are trained to selectively filter out any irrelevant information while maintaining what is useful [9].

Figure 6. GRU architecture [69]

Following the same approach presented in Section 3.1, we implemented a grid search method to determine the most suitable configuration for each deep learning model [70–74]. The hyperparameters considered in the grid search were (i) the number of layers, and (ii) the number of nodes in each layer as summarised in Table 3. As per the machine learning models, all deep learning models were tested using 80% of the original dataset as a training set and the remaining 20% as a test dataset.

Table 3. Parameter and levels of used in grid search.

Parameter	Levels
Number of nodes	From 100 to 400, step 100
Number of layers	From 1 to 4, step 1

The following parameters were fixed across different models: 1000 epochs with early stopping function, a batch size of 256, Sigmoid [75] as the activation function, MSE as the loss function, and a method for stochastic optimisation (Adam) as the optimiser. Due to the stochastic nature of the optimisation process, the grid search was performed 30 times using RMSE, MAPE and MAE were calculated to evaluate the models' performance.

Figures 7, 8 and 9 present the normalised results of the grid search. For RNN, the model with configuration with four layers and 200 nodes (RNN-4-200) presented the best result based on the MAE average metrics; the model configuration with three layers and 400 nodes (RNN-3-400) presented the best result based on the RMSE average and the MAPE average metrics. For LSTM, the models with configurations (a) with three layers and 300 nodes (LSTM-3-300), (b) four layers and 400 nodes (LSTM-4-400), and (c) three layers and 200 nodes (LSTM-3-200) generated the best results for RMSE,

MAPE, and MAE, respectively. For GRU, the model with configuration with three layers and 100 nodes (GRU-3-100) presented the best result based on the RMSE average and the MAE average metrics; the model configuration with four layers and 300 nodes (GRU-4-300) presented the best result based on the MAPE average metrics. Thus, these seven models were considered in the benchmark evaluation.

Figure 7. RMSE grid search result for (a) RNN, (b) LSTM, and (c) GRU.

Figure 8. MAPE grid search result for (a) RNN, (b) LSTM, and (c) GRU.

Figure 9. MAE grid search result for (a) RNN, (b) LSTM, and (c) GRU.

3.3. Benchmarks

An ARIMA model was selected as a benchmark due to its widespread use in building-level energy forecasting (see, for example, [33], [29], and [76]). The choice of the ARIMA model for this study was based on the time-series nature of our dataset (data numbers and the output variable relates to your past data). Equation (5) [77] represents the mathematical expression for the autoregressive component of the model.

$$x(t) = \sum_{i=1}^p \alpha_i x(t-i) \quad (5)$$

where t is the index represented by an integer, $x(t)$ is the estimated value, p is the number of autoregressive terms, and α is the polynomial related to the autoregressive operator of order p .

Equation (6) [77] represents the time dependency of the errors of previous estimates, i.e., the errors of the forecast are taken into account when estimating the next value in the time series.

$$x(t) = \sum_{i=1}^q \beta_i \varepsilon(t-i) \quad (6)$$

where q is the number of moving average terms, β is the polynomial related to the moving average operator of order q , and ε is the difference between the estimated and observed values of $x(t)$.

Equation (7) [77] combines Equations (5) and (6) and summarises the ARIMA model (p and q) used as a benchmark for this study.

$$x(t) = \sum_{i=1}^p \alpha_i x(t-i) - \sum_{i=1}^q \beta_i \varepsilon(t-i) \quad (7)$$

After empirical analysis, the selected ARIMA model presented the order of the autoregressive ($p = 1$), the degree of differencing ($d = 0$), and the order of the moving average ($q = 1$).

4. Results and discussions

As described, the models presented in Section 3 were trained to predict the energy consumption of the next hour using data from the previous 48 hours. By definition, this prediction is classified as very short-term load forecasting (VSTLF) [78]. For comparison and analysis purpose, we use the same models' configuration to predict the next 12 and 24 hours, in order to explore short-term load forecasting (STLF) performance [78]. This is done as follows, once defined which is the best model to predict the next 1h (the single-output regression model), this model will be used to create a chained multi-output regression, that is, a linear sequence of models capable of performing multi-output regression. This means that the first model in the sequence takes an input and predicts the output, then the second model uses the same input as the first model by adding the output of the first model to make a prediction, and so on (A schematic can be seen at Figure 10). This approach is used to define the VSTLF model, to predict the next 12h and 24h. This technique can end up propagating residual errors throughout the prediction, but despite this, it allows to use of the same trained model to carry out larger predictions, in a simple way and without needing to train new models. Both VSTLF and STLF can be used for purchase and production of electric power, but STLF has more broadly applications, such as transmission, transfer and distribution of electric power, management and maintenance of electric power sources, and management of the daily electric load demand [79]. Furthermore, it can be used by ESCOs to inform energy performance contract design and implementation.

Figure 10. Chained multi-output regression

4.1. Very short-term load forecasting

Table 4 presents the RMSE, MAPE, and MAE results of the five machine learning models (SVR-10-rbf, SVR-10-linear, Random Forest-9-200, XGBoost-5-100 and XGBoost-7-100) and the seven deep learning models (RNN-3-400, RNN-4-200, LSTM-3-200, LSTM-3-300, LSTM-4-400, GRU-3-100 and GRU-4-300) defined in previous section.

Table 4. RMSE, MAPE, and MAE results for the selected analysed model.

Models	RMSE	MAPE (%)	MAE
ARIMA	0.1114	28.8016	0.0615
SVR-10-rbf	0.1001	42.8923	0.0705
SVR-10-linear	0.1028	42.9745	0.0610
Random Forest-9-200	0.0863	22.771	0.0469
XGBoost-5-100	0.0844	21.659	0.0461
XGBoost-7-100	0.0847	21.573	0.0453
RNN-3-400	0.0947	28.325	0.0592
RNN-4-200	0.0909	29.122	0.0552
LSTM-3-200	0.0912	28.805	0.0553
LSTM-3-300	0.0911	29.03	0.0560
LSTM-4-400	0.0919	28.117	0.0564
GRU-3-100	0.0918	28.94	0.0558
GRU-4-300	0.0933	28.297	0.0585

The traditional method for evaluating the performance of prediction models is defined through comparative analysis of the metrics of each model. Based on this analysis, we conclude that the XGBoost models presented the best results among all the analyzed models. The ARIMA benchmark model presented the worst RMSE result, and the SVR models presented the worst MAPE result and MAE result among all the analyzed models.

Figures 11 and 12 illustrate the hourly load forecasts for the machine-learning and deep-learning models compared against the ground truth data. These clearly show that the proposed XGBoost models (Figure 11d–e) are very similar to the ground truth data compared to the another models.

Figure 11. Hourly load forecasting using (a) the SVR-10-rbf model, (b) the SVR-10-linear model, (c) the Random Forest-9-200 model, (d) the XGBoost-5-100 model and (e) the XGBoost-7-100 model.

Figure 12. Cont.

Figure 12. Hourly load forecasting using (a) the RNN-3-400 model, (b) the RNN-4-200 model, (c) the LSTM-3-200 model, (d) the LSTM-3-300 model, (e) the LSTM-4-400 model, (f) the GRU-3-100 model, and (g) the GRU-4-300 model.

4.2. Short-term load forecasting

As per the previous section, XGBoost-5-100 and XGBoost-7-100 were the best performing models tested. These models were then tested for predicting energy consumption over longer time horizons, i.e., 12 hours and 24 hours, consistent with STLTF. Table 5 presents the RMSE, MAPE and MAE results of the STLTF model (12 hours and 24 hours) compared with the result of the VSTLTF model.

Table 5. RMSE, MAPE, and MAE results for the VSTLTF and STLTF models.

Models	Prediction (hour)	RMSE	MAPE (%)	MAE
XGBoost-5-100	1	0.0844	21.659	0.0461
XGBoost-7-100	1	0.0847	21.573	0.0453
XGBoost-5-100	12	0.1835	21.580	0.1009
XGBoost-7-100	12	0.1717	21.749	0.0926
XGBoost-5-100	24	0.2342	21.603	0.1370
XGBoost-7-100	24	0.2149	21.639	0.1232

When comparing the results obtained with the STL and VSTL models, we observed a maximum increase of 177% and 202% in RMSE and MAE, respectively. This increase can be explained by implementing the STL model since there is a dependency between the outputs.

5. Related work

There is a substantial and increasing literature base on the use of machine learning and deep learning for load forecasting by forecasting horizon, target use case, and sector [80]. A significant focus of this literature remains supply-side energy consumption and demand forecasting from the perspective of the management and optimisation of power systems and electricity grids, and typically with forecasting horizons longer than minutes and hours. The motivations of warehouse owners and operators, and indeed ESCOs who manage their energy systems, are significantly different than those operating utilities. Furthermore, and as discussed, the energy profile of warehouses is somewhat idiosyncratic compared to other commercial and industrial buildings. Despite this, there are relatively few studies addressing load forecasting in the warehouse context [80]. Similarly, reflecting the focus on power systems and grids, a wider set of parameters (e.g. seasonal weather and special events) and longer forecasting horizons (short-to-long) are typically adopted in studies. Very short term load forecasting is arguably the least addressed due to the relatively narrow focus on extrapolating recently observed load patterns to the nearest future; modelling the relationship between load, time, weather conditions, special events and other load affecting factors is less important [81]. At the same time, Guan et al. [82] argue that effective VSTL is further complicated by a noisy data collection process, possible malfunctioning of data gathering devices and complicated load features. Indeed, studies on very short term horizons are largely absent from reviews [80].

Machine learning and deep learning has been used for building energy predictions in a wide range of studies [80,83]. In their recent review, Gassar et al. [80] note the most prominent energy prediction techniques for large-scale buildings as artificial neural network (ANN), support vector machines (SVM), multiple linear regression, i.e., multiple linear regression (MLR), gradient boosting (GB), and random forests (RF). However, none of the studies studied warehouses as a building type. Similarly, of the fourteen papers identified by Li et al. [83] using LSTM for building energy predictions, none dealt with warehouses.

A number of studies have sought to deploy artificial neural networks (ANN) for building energy prediction in short time horizons [29,84–86], however yet again largely warehouses are not addressed. Escrivá et al. [84], Gonzalez et al. [86] and Neto et al. [85] all use university-related buildings as a type of commercial building. Chae et al. [29] propose an ANN for forecasting day-ahead electricity usage of commercial buildings in 15-minute resolution. As variables, they selected day type indicator, time-of-day, HVAC set temperature schedule, outdoor air dry-bulb temperature, and outdoor humidity as the most important predictors for electricity consumption [29]. The ANN model was a conventional multi-layered feedforward network using a backpropagation algorithm. Correlation coefficient and CV(RMSE) were used as metrics to compare against a simple naive model and a variety of machine learning models including SVM, linear regression, Gaussian process, K-star classifier, nearest neighbour ball tree. The ANN outperformed all models and results suggest that the ANN could provide a day-ahead electricity usage profile with sub-hourly intervals and satisfactory accuracy for daily peak consumption. It is important to note the specific commercial buildings are not identified by the authors and thus the applicability of these findings to warehouses is uncertain.

In one of the few papers investigating load forecasting for commercial buildings that included warehouses, Chitalia et al. [7] compared nine deep learning algorithms (LSTM, BiLSTM, Encoder-Decoder, LSTM with attention, BiLSTM with attention, CNN+LSTM, CNN+BiLSTM, ConvLSTM, and ConvBiLSTM) for one-hour ahead and 24-hour ahead forecasting. The models were tested against peak loads and weather conditions. The algorithms delivered a 20-45% improvement in load forecasting performance compared to benchmarks and found that hybrid deep learning models could deliver satisfactory hour-ahead load prediction with as little as one month of data. The metrics

used were RMSE, MAPE, co-efficient of variance (CV). As Chitalia et al. [7] sought to compare different building types, root mean square algorithmic error was also used to allow fair comparison among buildings. A grid search was not employed.

As can be seen from above, there is a dearth of research on demand-side STLTF and VSTLTF for warehouses in general, and specifically using machine learning and deep learning models. Studies of commercial buildings typically involve universities or office buildings whose energy profile is significantly different from warehouses. Indeed, within the commercial building sector, warehouses are idiosyncratic and consequently require discrete consideration. Even within the small number of studies that could be identified, it is not clear whether one is directly comparable [29] on the basis of building type and the other does not compare performance between machine learning and deep learning models. Finally, no papers identified compare the performance of machine learning and deep learning models for both STLTF and VSTLTF. As such, we seek to address these gaps in the literature.

6. Conclusion and future work

This work explores short term and very short term electrical load forecasting for an under-researched building type, warehouses. This study compares the performance of machine learning and deep learning models to forecast the energy consumption of the upcoming hour based on data from the previous 48 hours, and benchmarks this performance against a classical time series forecasting technique, ARIMA. Unlike existing studies, we consider the data not only from the perspective of a warehouse owner and operator but an ESCO operating under an energy performance contract, and the data available to such a firm. Our results suggest that the XGBoost models outperformed all other machine learning and deep learning models, as well as ARIMA, for very short term load forecasting. All machine learning and deep learning models outperformed ARIMA when using RMSE as a measure; the SVR models presented the worst MAPE and MAE results. Unsurprisingly, XGBoost was less accurate for longer time horizons, i.e., 12 hour and 24 hours.

Accurate local forecasting close to real-time decision making can be used by warehouse owners and operators and ESCOs to design energy performance contracts, build a business case for green warehousing, control building energy systems and manage the charging/discharging or energy storage units in an energy-efficient and cost-effective way. It can also be used for anomaly detection and proactive plant management, including renewable energy sources such as Solar PV, and potentially transacting on future open carbon trading systems. While this study suggests machine learning models may be sufficient, ensemble solutions combining machine learning and deep learning may provide better results and is worthy of exploration. This study only examined building energy consumption. It may be fruitful for future research to consider the timing of battery charging/discharging and energy consumed by other operating activities. Furthermore, the site for this study did not feature high levels of automation or refrigeration, and was located in a country with moderate climate thus negating the need for substantial ventilation and air conditioning. Rudiger et al. [87] provide a useful overview of logistics services activities that contribute to GHG emissions including transshipment activities, and the need for cold or ambient storage, and the need for automated or manual order-picking, amongst others. Similarly, warehouses are only one type of logistics facility, transshipment terminals and distribution centres may also prove to be fruitful units of analysis. The addition of parameters that reflect different scenarios would provide greater robustness and generalisability of the proposed methods against variations in warehouse configuration and use.

Acknowledgments: This research has been partially financial supported by Irish Institute of Digital Business. Authors would like to thank Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq); Fundação de Amparo a Ciência e Tecnologia do Estado de Pernambuco (FACEPE); and Universidade de Pernambuco (UPE), an entity of the Government of the State of Pernambuco focused on the promotion of Teaching, Research and Extension.

Abbreviations

The following abbreviations were used in this manuscript:

AMI	Advanced Metering Infrastructure
ANN	Artificial Neural Network
ARIMA	Autoregressive Integrated Moving Average
CV	Co-efficient of Variance
EPC	Energy Performance Contract
ESCO	Energy Service Company
GB	Gradient Boosting
GHG	Greenhouse Gas
GRU	Gated Recurrent Unit
LSTM	Long Short-Term Memory
MAE	Mean Absolute Error
MLR	Multiple Linear Regression
MAPE	Mean Absolute Percent Error
MSE	Mean Squared Error
PV	Solar Photovoltaic
RF	Random Forest
RMSE	Root Mean Squared Error
RNN	Recurrent Neural Networks
STLF	Short-Term Load Forecasting
SVM	Support Vector Machines
SVR	Support Vector Machine Regression
VSTLF	Very Short-Term Load Forecasting
XGBoost	Extreme Gradient Boosting

References

1. EIA. International Energy Outlook 2021, 2021. <https://www.eia.gov/outlooks/ieo/>, Last accessed on 2021-12-13.
2. Depreaux, J. 28,500 Warehouses To Be Added Globally To Meet E-Commerce Boom, 2021. <https://www.interactanalysis.com/28500-warehouses-to-be-added-globally-to-meet-e-commerce-boom/>, Last accessed on 2021-12-13.
3. Trust, C. Warehousing and Logistics - Energy Opportunities for Warehousing and Logistics Companies, 2019. <https://www.carbontrust.com/resources/warehousing-and-logistics-guide>, Last accessed on 2021-12-13.
4. Lewczuk, K.; Kłodawski, M.; Gepner, P. Energy Consumption in a Distributional Warehouse: A Practical Case Study for Different Warehouse Technologies. *Energies* **2021**, *14*, 2709.
5. World Economic Forum.; Accenture. Supply Chain Decarbonisation: The Role of Logistics and Transport in Reducing Supply Chain Carbon Emissions. World Economic Forum and Accenture Geneva, 2009.
6. Dagdougui, H.; Bagheri, F.; Le, H.; Dessaint, L. Neural network model for short-term and very-short-term load forecasting in district buildings. *Energy and Buildings* **2019**, *203*, 109408.
7. Chitalia, G.; Pipattanasomporn, M.; Garg, V.; Rahman, S. Robust short-term electrical load forecasting framework for commercial buildings using deep recurrent neural networks. *Applied Energy* **2020**, *278*, 115410.
8. Hippert, H.S.; Pedreira, C.E.; Souza, R.C. Neural networks for short-term load forecasting: A review and evaluation. *IEEE Transactions on power systems* **2001**, *16*, 44–55.
9. Ribeiro, A.M.N.; do Carmo, P.R.X.; Rodrigues, I.R.; Sadok, D.; Lynn, T.; Endo, P.T. Short-Term Firm-Level Energy-Consumption Forecasting for Energy-Intensive Manufacturing: A Comparison of Machine Learning and Deep Learning Models. *Algorithms* **2020**, *13*, 274.
10. Quilumba, F.L.; Lee, W.J.; Huang, H.; Wang, D.Y.; Szabados, R.L. Using smart meter data to improve the accuracy of intraday load forecasting considering customer behavior similarities. *IEEE Transactions on Smart Grid* **2014**, *6*, 911–918.

11. Alahmad, M.; Peng, Y.; Sordiashie, E.; El Chaar, L.; Aljuhaishi, N.; Sharif, H. Information technology and the smart grid-A pathway to conserve energy in buildings. 2013 9th International Conference on Innovations in Information Technology (IIT). IEEE, 2013, pp. 60–65.
12. Bertoldi, P.; Boza-Kiss, B.; Toleikytė, A. Energy Service Market in the EU. *Publications Office of the European Union: Luxembourg* **2019**.
13. European Commission. A Renovation Wave for Europe - greening our buildings, creating jobs, improving lives. https://ec.europa.eu/energy/sites/ener/files/eu_renovation_wave_strategy.pdf, 2020. Accessed: 2021-04-12.
14. Sorrell, S. The economics of energy service contracts. *Energy policy* **2007**, *35*, 507–521.
15. Sola, J.; Sevilla, J. Importance of input data normalization for the application of neural networks to complex industrial problems. *IEEE Transactions on nuclear science* **1997**, *44*, 1464–1468.
16. Debnath, K.B.; Mourshed, M. Forecasting methods in energy planning models. *Renewable and Sustainable Energy Reviews* **2018**, *88*, 297–325.
17. Willmott, C.J.; Matsuura, K. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate research* **2005**, *30*, 79–82.
18. Kolomvatsos, K.; Papadopoulou, P.; Anagnostopoulos, C.; Hadjiefthymiades, S. A Spatio-Temporal Data Imputation Model for Supporting Analytics at the Edge. Conference on e-Business, e-Services and e-Society. Springer, 2019, pp. 138–150.
19. Luo, J.; Hong, T.; Yue, M. Real-time anomaly detection for very short-term load forecasting. *Journal of Modern Power Systems and Clean Energy* **2018**, *6*, 235–243.
20. Ryu, S.; Noh, J.; Kim, H. Deep neural network based demand side short term load forecasting. *Energies* **2017**, *10*, 3.
21. Berriel, R.F.; Lopes, A.T.; Rodrigues, A.; Varejao, F.M.; Oliveira-Santos, T. Monthly energy consumption forecast: A deep learning approach. 2017 International Joint Conference on Neural Networks (IJCNN). IEEE, 2017, pp. 4283–4290.
22. Azadeh, A.; Ghaderi, S.; Sohrabkhani, S. Annual electricity consumption forecasting by neural network in high energy consuming industrial sectors. *Energy Conversion and management* **2008**, *49*, 2272–2278.
23. Kuo, P.H.; Huang, C.J. A high precision artificial neural networks model for short-term energy load forecasting. *Energies* **2018**, *11*, 213.
24. Kong, W.; Dong, Z.Y.; Jia, Y.; Hill, D.J.; Xu, Y.; Zhang, Y. Short-term residential load forecasting based on LSTM recurrent neural network. *IEEE Transactions on Smart Grid* **2017**, *10*, 841–851.
25. De Myttenaere, A.; Golden, B.; Le Grand, B.; Rossi, F. Mean absolute percentage error for regression models. *Neurocomputing* **2016**, *192*, 38–48.
26. Hsieh, T.J.; Hsiao, H.F.; Yeh, W.C. Forecasting stock markets using wavelet transforms and recurrent neural networks: An integrated system based on artificial bee colony algorithm. *Applied soft computing* **2011**, *11*, 2510–2525.
27. Chen, C.; Liu, Y.; Kumar, M.; Qin, J. Energy consumption modelling using deep learning technique—a case study of EAF. *Procedia CIRP* **2018**, *72*, 1063–1068.
28. Li, C.; Tao, Y.; Ao, W.; Yang, S.; Bai, Y. Improving forecasting accuracy of daily enterprise electricity consumption using a random forest based on ensemble empirical mode decomposition. *Energy* **2018**, *165*, 1220–1227.
29. Chae, Y.T.; Horeh, R.; Hwang, Y.; Lee, Y.M. Artificial neural network model for forecasting sub-hourly electricity usage in commercial buildings. *Energy and Buildings* **2016**, *111*, 184–194.
30. Raza, M.Q.; Khosravi, A. A review on artificial intelligence based load demand forecasting techniques for smart grid and buildings. *Renewable and Sustainable Energy Reviews* **2015**, *50*, 1352–1372.
31. Chen, J.F.; Wang, W.M.; Huang, C.M. Analysis of an adaptive time-series autoregressive moving-average (ARMA) model for short-term load forecasting. *Electric Power Systems Research* **1995**, *34*, 187–196.
32. Huang, S.J.; Shih, K.R. Short-term load forecasting via ARMA model identification including non-Gaussian process considerations. *IEEE Transactions on power systems* **2003**, *18*, 673–679.
33. Kim, Y.; Son, H.g.; Kim, S. Short term electricity load forecasting for institutional buildings. *Energy Reports* **2019**, *5*, 1270–1280.
34. Huang, C.M.; Huang, C.J.; Wang, M.L. A particle swarm optimization to identifying the ARMAX model for short-term load forecasting. *IEEE Transactions on Power Systems* **2005**, *20*, 1126–1133.

35. Chakhchoukh, Y.; Panciatici, P.; Mili, L. Electric load forecasting based on statistical robust methods. *IEEE Transactions on Power Systems* **2010**, *26*, 982–991.
36. Yang, J.; Stenzel, J. Short-term load forecasting with increment regression tree. *Electric Power Systems Research* **2006**, *76*, 880–888.
37. Ceperic, E.; Ceperic, V.; Baric, A. A strategy for short-term load forecasting by support vector regression machines. *IEEE Transactions on Power Systems* **2013**, *28*, 4356–4364.
38. Chen, Y.; Tan, H. Short-term prediction of electric demand in building sector via hybrid support vector regression. *Applied energy* **2017**, *204*, 1363–1374.
39. Chen, Y.; Xu, P.; Chu, Y.; Li, W.; Wu, Y.; Ni, L.; Bao, Y.; Wang, K. Short-term electrical load forecasting using the Support Vector Regression (SVR) model to calculate the demand response baseline for office buildings. *Applied Energy* **2017**, *195*, 659–670.
40. Li, Q.; Zhang, L.; Xiang, F. Short-term load forecasting: A case study in Chongqing factories. 2019 6th International Conference on Information Science and Control Engineering (ICISCE). IEEE, 2019, pp. 892–897.
41. Zhu, K.; Geng, J.; Wang, K. A hybrid prediction model based on pattern sequence-based matching method and extreme gradient boosting for holiday load forecasting. *Electric Power Systems Research* **2021**, *190*, 106841.
42. Zheng, H.; Yuan, J.; Chen, L. Short-term load forecasting using EMD-LSTM neural networks with a Xgboost algorithm for feature importance evaluation. *Energies* **2017**, *10*, 1168.
43. Grolinger, K.; L'Heureux, A.; Capretz, M.A.; Seewald, L. Energy forecasting for event venues: Big data and prediction accuracy. *Energy and buildings* **2016**, *112*, 222–233.
44. Singh, S.; Hussain, S.; Bazaz, M.A. Short term load forecasting using artificial neural network. 2017 Fourth International Conference on Image Information Processing (ICIIP). IEEE, 2017, pp. 1–5.
45. Huang, Q.; Li, J.; Zhu, M. An improved convolutional neural network with load range discretization for probabilistic load forecasting. *Energy* **2020**, *203*, 117902.
46. Bengio, Y.; Simard, P.; Frasconi, P. Learning long-term dependencies with gradient descent is difficult. *IEEE transactions on neural networks* **1994**, *5*, 157–166.
47. Pascanu, R.; Mikolov, T.; Bengio, Y. On the difficulty of training recurrent neural networks. International conference on machine learning. PMLR, 2013, pp. 1310–1318.
48. Sundermeyer, M.; Schlüter, R.; Ney, H. LSTM neural networks for language modeling. Thirteenth annual conference of the international speech communication association, 2012.
49. Marino, D.L.; Amarasinghe, K.; Manic, M. Building energy load forecasting using deep neural networks. IECON 2016-42nd Annual Conference of the IEEE Industrial Electronics Society. IEEE, 2016, pp. 7046–7051.
50. Kuan, L.; Yan, Z.; Xin, W.; Yan, C.; Xiangkun, P.; Wenxue, S.; Zhe, J.; Yong, Z.; Nan, X.; Xin, Z. Short-term electricity load forecasting method based on multilayered self-normalizing GRU network. 2017 IEEE Conference on Energy Internet and Energy System Integration (EI2). IEEE, 2017, pp. 1–5.
51. He, F.; Zhou, J.; Feng, Z.k.; Liu, G.; Yang, Y. A hybrid short-term load forecasting model based on variational mode decomposition and long short-term memory networks considering relevant factors with Bayesian optimization algorithm. *Applied energy* **2019**, *237*, 103–116.
52. Chung, J.; Gulcehre, C.; Cho, K.; Bengio, Y. Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv preprint arXiv:1412.3555* **2014**.
53. Wang, Y.; Liu, M.; Bao, Z.; Zhang, S. Short-term load forecasting with multi-source data using gated recurrent unit neural networks. *Energies* **2018**, *11*, 1138.
54. Wu, W.; Liao, W.; Miao, J.; Du, G. Using gated recurrent unit network to forecast short-term load considering impact of electricity price. *Energy Procedia* **2019**, *158*, 3369–3374.
55. Drucker, H.; Burges, C.J.; Kaufman, L.; Smola, A.; Vapnik, V.; others. Support vector regression machines. *Advances in neural information processing systems* **1997**, *9*, 155–161.
56. Vapnik, V. The nature of statistical learning theory (p. 189). New York: Springer-Verlag. doi **1995**, *10*, 978–1.
57. Sapankevych, N.I.; Sankar, R. Time series prediction using support vector machines: a survey. *IEEE Computational Intelligence Magazine* **2009**, *4*, 24–38.
58. Breiman, L. Random forests. *Machine learning* **2001**, *45*, 5–32.
59. Huang, N.; Lu, G.; Xu, D. A permutation importance-based feature selection method for short-term electricity load forecasting using random forest. *Energies* **2016**, *9*, 767.

60. Breiman, L. Bagging predictors. *Machine learning* **1996**, *24*, 123–140.
61. Ho, T.K. The random subspace method for constructing decision forests. *IEEE transactions on pattern analysis and machine intelligence* **1998**, *20*, 832–844.
62. Nadi, A.; Moradi, H. Increasing the views and reducing the depth in random forest. *Expert Systems with Applications* **2019**, *138*, 112801.
63. Hammou, B.A.; Lahcen, A.A.; Mouline, S. An effective distributed predictive model with Matrix factorization and random forest for Big Data recommendation systems. *Expert Systems with Applications* **2019**, *137*, 253–265.
64. Wang, Y.; Sun, S.; Chen, X.; Zeng, X.; Kong, Y.; Chen, J.; Guo, Y.; Wang, T. Short-term load forecasting of industrial customers based on SVM and XGBoost. *International Journal of Electrical Power & Energy Systems* **2021**, *129*, 106830.
65. Zhang, B.; Wu, J.L.; Chang, P.C. A multiple time series-based recurrent neural network for short-term load forecasting. *Soft Computing* **2018**, *22*, 4099–4112.
66. Gers, F.A.; Schmidhuber, J.; Cummins, F. Learning to forget: continual prediction with LSTM. 1999 Ninth International Conference on Artificial Neural Networks ICANN 99. (Conf. Publ. No. 470), 1999, Vol. 2, pp. 850–855 vol.2. doi:10.1049/cp:19991218.
67. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural computation* **1997**, *9*, 1735–1780.
68. Jozefowicz, R.; Zaremba, W.; Sutskever, I. An empirical exploration of recurrent network architectures. International conference on machine learning, 2015, pp. 2342–2350.
69. Yu, Y.; Si, X.; Hu, C.; Zhang, J. A review of recurrent neural networks: LSTM cells and network architectures. *Neural computation* **2019**, *31*, 1235–1270.
70. Bergstra, J.; Bengio, Y. Random search for hyper-parameter optimization. *Journal of Machine Learning Research* **2012**, *13*, 281–305.
71. Liao, J.M.; Chang, M.J.; Chang, L.M. Prediction of Air-Conditioning Energy Consumption in R&D Building Using Multiple Machine Learning Techniques. *Energies* **2020**, *13*, 1847.
72. Yoon, H.; Kim, Y.; Ha, K.; Lee, S.H.; Kim, G.P. Comparative evaluation of ANN-and SVM-time series models for predicting freshwater-saltwater interface fluctuations. *Water* **2017**, *9*, 323.
73. Kavaklioglu, K. Modeling and prediction of Turkey's electricity consumption using Support Vector Regression. *Applied Energy* **2011**, *88*, 368–375.
74. Samsudin, R.; Shabri, A.; Saad, P. A comparison of time series forecasting using support vector machine and artificial neural network model. *Journal of applied sciences* **2010**, *10*, 950–958.
75. Han, J.; Moraga, C. The influence of the sigmoid function parameters on the speed of backpropagation learning. From Natural to Artificial Neural Computation; Mira, J.; Sandoval, F., Eds.; Springer Berlin Heidelberg: Berlin, Heidelberg, 1995; pp. 195–201.
76. Vaghefi, A.; Jafari, M.A.; Bisse, E.; Lu, Y.; Brouwer, J. Modeling and forecasting of cooling and electricity load demand. *Applied Energy* **2014**, *136*, 186–196.
77. Pushp, S. Merging Two Arima Models for Energy Optimization in WSN. *arXiv preprint arXiv:1006.5436* **2010**.
78. Hsiao, Y.H. Household electricity demand forecast based on context information and user daily schedule analysis from meter data. *IEEE Transactions on Industrial Informatics* **2014**, *11*, 33–43.
79. Mele, E.; others. A review of machine learning algorithms used for load forecasting at microgrid level. Sinteza 2019-International Scientific Conference on Information Technology and Data Related Research. Singidunum University, 2019, pp. 452–458.
80. Gassar, A.A.A.; Cha, S.H. Energy prediction techniques for large-scale buildings towards a sustainable built environment: A review. *Energy and Buildings* **2020**, *224*, 110238.
81. Charytoniuk, W.; Chen, M.S. Very short-term load forecasting using artificial neural networks. *IEEE transactions on Power Systems* **2000**, *15*, 263–268.
82. Guan, C.; Luh, P.B.; Michel, L.D.; Wang, Y.; Friedland, P.B. Very short-term load forecasting: wavelet neural networks with data pre-filtering. *IEEE Transactions on Power Systems* **2012**, *28*, 30–41.
83. Li, G.; Zhao, X.; Fan, C.; Fang, X.; Li, F.; Wu, Y. Assessment of long short-term memory and its modifications for enhanced short-term building energy predictions. *Journal of Building Engineering* **2021**, *43*, 103182.

84. Escrivá-Escrivá, G.; Álvarez-Bel, C.; Roldán-Blay, C.; Alcázar-Ortega, M. New artificial neural network prediction method for electrical consumption forecasting based on building end-uses. *Energy and Buildings* **2011**, *43*, 3112–3119.
85. Neto, A.H.; Fiorelli, F.A.S. Comparison between detailed model simulation and artificial neural network for forecasting building energy consumption. *Energy and buildings* **2008**, *40*, 2169–2176.
86. Gonzalez, P.A.; Zamarreno, J.M. Prediction of hourly energy consumption in buildings based on a feedback artificial neural network. *Energy and buildings* **2005**, *37*, 595–601.
87. Rüdiger, D.; Schön, A.; Dobers, K. Managing greenhouse gas emissions from warehousing and transshipment with environmental performance indicators. *Transportation Research Procedia* **2016**, *14*, 886–895.