

Anomaly Detection in Electricity Consumption Data using Deep Learning

Mohammad Kardi
Faculty of Informatics
Tishreen University
Latakia, Syria
mohammadkardi1@gmail.com

Tarek AlSkaif
Information Technology Group
Wageningen University & Research
Wageningen, The Netherlands
tarek.alskaif@wur.nl

Bedir Tekinerdogan
Information Technology Group
Wageningen University & Research
Wageningen, The Netherlands
bedir.tekinerdogan@wur.nl

João P. S. Catalão
Faculty of Engineering
University of Porto & INESC TEC
Porto, Portugal
catalao@fe.up.pt

Abstract—Anomaly detection in electricity consumption data is one of the most important methods to identify anomalous events in buildings and electric assets, such as energy theft, metering defect, cyber attacks and technical losses. In this paper, a novel deep learning based approach is presented to detect anomalies in electricity consumption data one hour ahead of time. We address this challenge in two stages. First, we build an Long Short-Term Memory (LSTM) based neural network model to predict the next hour sample. Second, we use another LSTM autoencoder to learn the features of normal consumption. The output of the first stage is used as an input to the LSTM autoencoder. The LSTM autoencoder will learn the features of normal consumption and the input will be similar to output when applied. For anomalies, the input and output will be significantly different. The Exponential Moving Average (EMA) is used as a threshold and two types of anomalies are distinguished, local and global anomalies. Several weather features are considered in this study, such as pressure, cloud cover, humidity, temperature, wind direction and wind speed in addition to temporal and lag features. A feature selection method to find the optimal features that achieve good results is also implemented. We validate the proposed approach by comparing the detected anomalous consumption and the normal consumption within the same period, and the results demonstrate a drastic increase in electricity consumption during the anomalous periods. The results also show that the temporal and lag features have improved the efficiency and performance of the proposed method.

Index Terms—LSTM autoencoder, Deep Learning, Anomaly detection, Electricity consumption, Anomalous consumption.

I. INTRODUCTION

Anomaly detection is a common technique used to identify unusual patterns in data that raise suspicions by differing significantly from normal behavior. This technique is considered of high importance because detecting anomalous events provides a decision making process with information on what should be done to deal with a certain event. This problem has been addressed in a variety of domains and within diverse research areas [1]. For example, an abnormal credit card

transaction may indicate fraud and that the system should prevent the transaction [2].

Anomaly detection in electricity consumption data is one of the most important methods to identify anomalous events in buildings and critical electric infrastructure facilities, such as energy theft, metering defect, and technical losses [3]. Besides, this technique is useful to detect cyber attacks in cyber-physical systems, such as power overloading attacks in smart grids or microgrids [4]. To avoid running into power failure problems, anomalous consumption should be detected before it occurs, and therefore guaranteeing a normal operation of electric assets and networks and avoiding power outages.

Several anomaly detection methods have been proposed in the area of energy analytics and smart grids that aim to find patterns in data which do not conform to the expected behavior. For example, the authors of [5] presented a multi-agent based unsupervised anomaly detection method to detect abnormal energy consumption in a smart campus. The method is applied in two steps. First, they used an ensemble model to label the unlabeled dataset. Second, they proposed a method for detecting anomalies based on combining LSTM networks and autoencoder neural networks. The results obtained from the ensemble model is used to evaluate the performance of the proposed method in the second step. The authors in [6] examined five models to detect anomalies in school electricity consumption data. Finally, they presented a hybrid model that combines polynomial regression and Gaussian distribution to detect anomalies. A real-time detection method of anomalous power consumption with the aim to reduce power consumption in buildings has been proposed in [7]. The authors in [8] employed density-based clustering techniques to detect unusual energy consumption, where clusters have been regarded as normal behavior and noise points as anomalous behavior. The work in [9] proposed a early fault detection method of anomalous electricity load profiles using time series clustering and decision trees.

In this study, we present a novel unsupervised deep learning

based approach to detect anomalies in electricity consumption data one hour ahead of time. We address this challenge in two stages. First, we build an Long Short-Term Memory (LSTM) to predict the next hour sample. Second, we use an LSTM autoencoder neural network to learn the features of normal consumption. The output of the first stage is fed as an input for the LSTM autoencoder. The LSTM autoencoder will learn the features of normal consumption and the input will be similar to output when applied. For anomalies, the input and the output will be significantly different since it will be considered as unexpected data. The Exponential Moving Average (EMA) is used as a threshold and two types of anomalies are distinguished in our method, namely: i) local anomaly, when a single metric loss crosses a threshold, and ii) a global anomaly, when the mean loss of all metrics crosses a threshold. Further, external weather variables are considered in this study, such as pressure, cloud cover, humidity, temperature, wind direction and wind speed. The temporal and lag features will be created from the historical electricity consumption data. Further, a feature selection method is used to find the optimal features that achieve better results.

The rest of this paper is organized as follows. Section II provides a brief background on anomaly detection and the relationship between electricity demand and weather variables. Section III covers the unsupervised method being implemented to detect abnormal consumption. In section IV the results of research are presented. Finally, The paper is concluded in Section V.

II. BACKGROUND

A. Anomaly Detection on Machine Learning

In the past, manual anomaly detection was a viable option. There were only a few metrics to track and the data sets were manageable enough. However, currently in the world of digitization, the amount of data exceeds the human ability to study it manually [10]. Hence, automated anomaly detection becomes a necessity where machine learning algorithms are widely being used to automate anomaly detection. Anomaly detection algorithms on machine learning automatically analyze datasets and identify breaches in the data patterns that signal an anomaly. Anomalous data can indicate critical incidents, such as a technical glitch, fraud, or network intrusion. Detection of anomalous data can help point out where an error is occurring and quickly get tech support on the issue by informing the responsible parties to act.

B. Time Series on Anomaly Detection

Time series data is a sequence of data points taken at successive equally spaced time intervals, giving us the ability to track changes over time [11]. In this study, we analyzed a time series of electricity demand for a given consumption over a certain period. Time series analysis is very useful to identify and analyze the natural changes that occur in electricity consumption, so that any significant deviation from the natural change is most likely to be considered anomalous. The electricity demand data used in this study are recorded

with hourly resolution for the period between 2017 and 2019. Fig. 1 shows one week's electricity consumption from Monday to Sunday. As shown, consumption is highest on weekdays and lowest on weekends. The data for one week contains 168 data instances. In Fig. 1, each data point represents the electricity demand in the current hour.

C. The Impact of Weather Variables on Electricity Demand

Many weather variables, such as solar irradiance, cloud cover, humidity and wind speed, typically have an interdependency with temperature. The interdependency between weather variables could be very complex [12]. Hence, these variables can lead to an increase or a decrease in the temperature. An increase or decrease in temperature leads to growth in the electricity demand due to cooling and heating requirements, respectively. The weather variables interdependency may differ from one location to another [13]. Indeed, several works in prior literature have already shown the influences of weather variables on electricity demand in different locations. For instance, the authors in [12], analysed the influence of climate variables on electricity demand in the state of New South Wales, Australia. Their study showed that climatic variables which are heating degree days, evaporation, humidity, and wind speed greatly affect the electricity demand. The authors in [14], concluded that daily electricity demand in Niamey varies both seasonally and from year to year, showing that temperature, humidity, and solar radiation have significant influence on electricity consumption. They also observed a very low coherence between wind speed and daily electricity consumption. The authors in [15], highlighted the impact of weather variables on electricity demand in Thailand. Therefore, it is important to consider the impact of these variables on this proposed method.

III. METHODS

A. Long Short-Term Memory Regular Network

Recurrent Neural Networks (RNNs) are powerful tools for modeling sequences of data. They are extensible and capable of learning order dependence in sequence prediction problems [16]. RNNs can remember important features about the input they received, due to the internal memory, which makes it applicable to tasks such as natural languages processing and

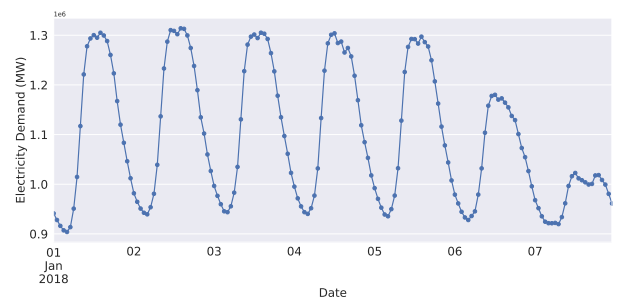


Fig. 1: Electricity demand over one week.

time series forecasting. In this study, when predicting the electricity demand ahead of time, the previous electricity demand data are required. Hence, there is a need to remember the previous inputs. RNNs are different from Feed-forward Neural Networks (FNNs) because they include cyclic connections, while FNNs have no cyclic connection between nodes [5]. RNNs use the same parameters for each input and the output of the previous state will be fed as the input of the next state (time step). A visual example of the architecture of an RNN is shown in Fig. 2. For each time step, activation $a^{<t>}$ and output $y^{<t>}$ are expressed using (1) and (2).

$$a^{<t>} = g_1(W_{aa}a^{<t-1>} + W_{ax}x^{<t>} + b_a), \quad (1)$$

$$y^{<t>} = g_2(W_{ya}a^{<t>} + b_y), \quad (2)$$

where $W_{aa}, W_{ax}, W_{ya}, b_a, b_y$ are coefficients that are shared temporally, and g_1, g_2 are activation functions.

However, RNN suffers from short-term memory, which hampers learning of long data sequences [17]. The problem is that at each step, an RNN updates its state [5]. During back propagation, gradients are values used to update a neural network's weights. Due to the many updates when the sequence is long enough, the gradient becomes smaller and smaller. The gradient also carries information used in the RNN parameter update and when the gradient becomes too small, RNN stops learning. Therefore, it will have a problem carrying information from earlier time steps to later ones and will leave out important information from the beginning. This is a major drawback in RNNs, which is called vanishing gradient. This problem is solved by LSTM networks. LSTMs are an extension for RNNs (i.e., with short-term memory) and also have a long-term memory. Therefore, it is well suited to remember inputs over a long period of time. An LSTM cell contains three gates: input, forget and output gate. With these gates, LSTM has solved the vanishing gradient problem.

B. Long Short-Term Memory Autoencoder Network

An autoencoder is a type of artificial neural network, which is mainly used to learn efficient data coding in an unsupervised manner. It is designed to encode the input into a compressed representation, and then decode it back such that the reconstructed input is as similar as possible to the original one [18]. The difference between the original input and the decoded one is measured by the reconstruction error. In this study, we used an autoencoder to learn a representation of the data for normal electricity consumption. A well-trained autoencoder will reconstruct the data that is coming from the

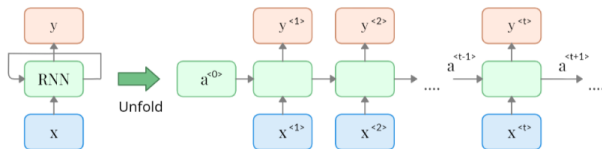


Fig. 2: RNN architecture.

normal electricity consumption such that the reconstruction error is small. Regarding the data from the abnormal electricity consumption, the autoencoder network may not be able to reconstruct it well because it only saw normal instances during training. This will lead to a higher reconstruction error, enabling us to detect such high reconstruction errors and label them as abnormal electricity consumption. When an autoencoder is trained to reconstruct the original input data, the data is processed through the following steps:

- The data will be fed by an encoder from a high-dimensional input to a bottleneck layer, where the dimension becomes lower because the number of neurons is the least in order to compress the input data and extract the most important attributes out of it.
- The decoder fetches the compressed input data from the bottleneck layer to convert it back to the original input data shape. In other words, reconstructing the original input data.
- The previous steps are repeated until the autoencoder network is able to best reconstruct the original input from an "encoded" state.

The autoencoder network architecture is shown in Fig. 3. As soon as the data is reconstructed, it is possible to compare the reconstructed data with the original data, compute the difference, and calculate the loss, which can then be minimized. The loss function of the autoencoder network is calculated by (3):

$$L(\theta, \phi) = \frac{1}{N} \sum_{i=1}^N (x^i - f_{\theta}(g_{\phi}(x^i)))^2. \quad (3)$$

According to (3), the loss function depends on θ and ϕ , which are the parameters that define the encoder and the decoder, x^i is the i -th feature, the encoder is represented by g_{ϕ} , and the decoder is represented by f_{θ} . Equation (3) sums up the difference between the original data x , and the reconstructed data $f_{\theta}(g_{\phi}(x^i))$ over N which is the number of input features.

In our case, the electricity consumption data are time correlated, thus to handle these data we used an LSTM based autoencoder. Stacking LSTM networks in an autoencoder fashion enables the method to learn more complex patterns inherent in the data [5]. The LSTM encoder learns to map

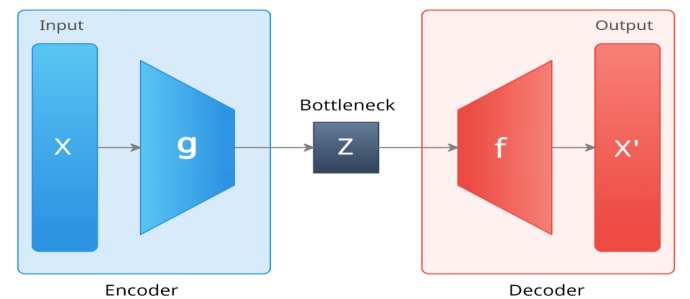


Fig. 3: Autoencoder architecture.

compressed representations of the normal time series and the LSTM decoder uses this representation to reconstruct the normal time series. This is particularly useful in scenarios when anomalous data is not available or is sparse, making it difficult to learn a classification model over the normal and anomalous sequences [19].

C. Model Design

The proposed model consists of two major components mentioned previously: i) an LSTM regular network and II) an LSTM autoencoder network. This model uses the LSTM regular network for predicting the sample's electricity demand one hour ahead of time, while the LSTM autoencoder network is used to address the predicted values from the LSTM regular network to detect if the electricity demand sample will be anomalous or not. In this study, the steps of the method for detecting anomalies are outlined as:

- 1) The first step is to preprocess the electricity consumption data. The data preprocessing includes two procedures: i) data standardization, and ii) feature selection. Data standardization is a crucial part of the data preprocessing where the features have drastically different ranges. This method rescales the range of the features to a similar scale. Having features on a similar scale can help our algorithm to converge faster and reduce training time [10]. The features can be normalized using (4):

$$\mathbf{X}' = \frac{\mathbf{X} - \mu}{\sigma}, \quad (4)$$

where \mathbf{X} is the matrix of the feature values, \mathbf{X}' is the matrix of the normalized feature values, μ is the mean of the feature values and σ is the standard deviation of the feature values. On the other hand, we applied the feature selection procedure to find the optimal and the most important features for electricity demand prediction. Selecting the optimal features improves the performance of machine learning algorithms and reduces the complexity and computational cost of the problem [20].

- 2) After the preprocessing step, the LSTM regular network is trained on the electricity consumption data in order to predict future changes in electricity demand. In addition to predicting the electricity demand, the important features of the electricity demand are also predicted.
- 3) Then, we will train the LSTM autoencoder network on the normal electricity consumption data. The objective is to learn the attributes of the normal electricity consumption. After that the LSTM autoencoder network will be able to reconstruct instances of normal time series well.
- 4) The LSTM regular network predicts the electricity consumption sample one hour ahead of time. Then, determining the likelihood of a point in a time series being anomalous or not is done by the LSTM autoencoder network. Mean Square Error (MSE) is used as a metric for the model evaluation. The proposed model architecture is illustrated in Fig. 4.

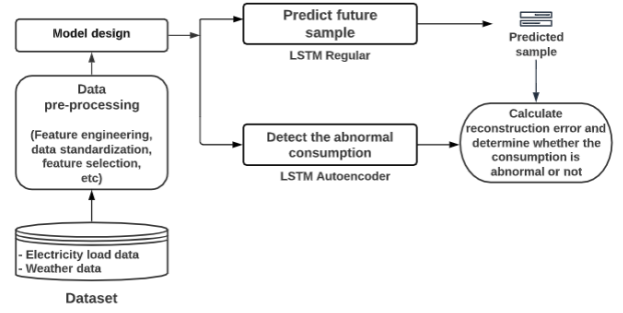


Fig. 4: The general model of the proposed method.

IV. RESULTS AND DISCUSSION

A. Predictor Variables

When dealing with time sensitive data, there is a collection of useful engineered features that describe the time series index of a time based data set. When predicting time series that contain common seasonal and trend patterns, the engineered time-related features can improve the model performance. Fig. 5 depicts the electricity consumption in May 2017. It clearly shows daily and weekly seasonality in electricity consumption. Much less energy is consumed on the weekends while weekdays are the busy days where the consumption is highest.

Fig. 6 shows the hourly electricity consumption over the period of March 2017 – December 2019, which clearly exhibits a yearly seasonality. It can also be shown in Fig. 6 that the electricity consumption in the considered dataset is splitted into two clusters, one with oscillations centered roughly around 1.3 megawatt, and another one with more scattered data points centered roughly around 1.1 megawatt. Another important point that becomes apparent is the drastic decrease in electricity consumption in early January and late December, during the holidays.

Since our electricity consumption data have daily, weekly and yearly seasonality. Hence, the hour of the day and the day of the week will be important features. In Table I, all the features that we considered in our study are summarized.

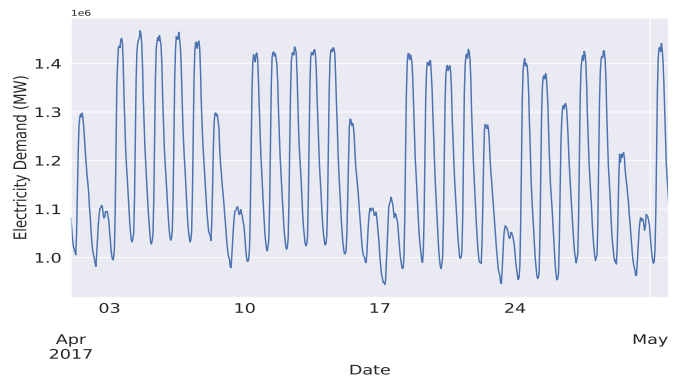


Fig. 5: Hourly electricity consumption during May 2017.

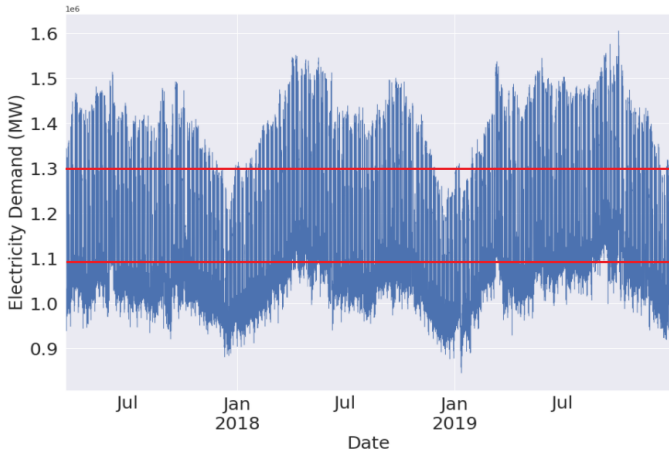


Fig. 6: Hourly electricity consumption during the considered period of March 2017 – December 2019.

B. Optimal Features Selection

Pearson correlation coefficient has been applied in this study to measure the strength of association between the electricity consumption data and the considered predictor variables. Pearson correlation is used to calculate the linear correlation between two variables [21]. Thus, the correlation between the electricity consumption and each predictor variable is measured independently. The correlation results are shown in Fig. 7, where the correlation lies in the range between -1 and 1. A value of 0 indicates that there is no association between the electricity consumption and a certain variable. A value greater than 0 indicates a positive association. A value less than 0 indicates a negative association. From Fig. 7, it can be noticed that more than one variable affects the electricity demand. As shown, CC(t) has a low correlation with the electricity demand, followed by P(t), Month, AL(M-12) and AL(D-365). Thus, these features are dropped out and not included in the training process, while the other features are kept.

TABLE I: Predictor variables considered in our anomaly detection method in electricity consumption data.

Category	Number	Acronym	Description
Weather	1	P(t)	Pressure at t.
	2	CC(t)	Cloud Cover at t.
	3	H(t)	Humidity at t.
	4	T(t)	Temperature at t.
	5	WD(t)	Wind Direction at t.
	6	WS(t)	Wind Speed at t.
Temporal	7	HoD	Hour of the day.
	8	DoW	Day of the week.
	9	Month	Index of the month.
	10	WD	Weekend day.
Lag	11	AL(D-1)	Average Load one day ago.
	12	AL(D-7)	Average Load same day last week.
	13	AL(W-1)	Average Load last week.
	14	AL(M-1)	Average Load last month.
	15	AL(M-12)	Average Load same month in the last year.
	16	AL(D-365)	Average Load same day in the last year.

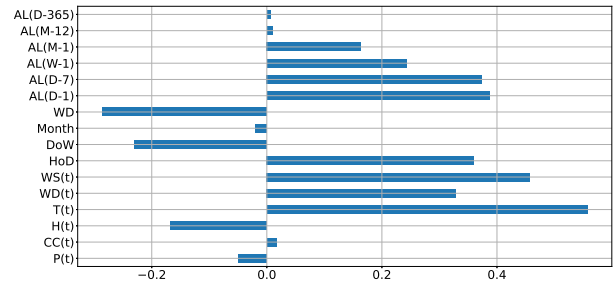


Fig. 7: Pearson correlation between the electricity demand and the predictor variables.

C. Anomaly Detection Results

In this proposed study two different anomaly detection techniques are considered: global and local anomaly detection. Fig. 8 shows all the anomalies detected, local and global, in a sample month of the dataset, April 2018. Local anomaly detection looks for anomalies in one metric and is triggered when the mean loss of the single metric crosses a threshold, while global anomaly detection introduces interaction with all the important features considered and is triggered when the mean loss of all features crosses a threshold. Either way, the anomaly in the energy consumption data is a sign of energy waste whether due to user behaviour, human error or a technical problem that caused losses. Therefore, anomalous consumption should be identified to reduce peak power consumption or change undesirable user behavior. In this study, a baseline of anomaly detector was based on reconstruction error and was constructed by the following steps. First, calculate reconstruction error for each sample in time series which is evaluated to decide what is abnormal. Second, calculate the mean μ and standard deviation σ of the single metric in the case of local anomaly, while of all metrics in the case global anomaly. The mean and standard deviation are calculated over the monitored time series. Third, calculate the corresponding $\mu \pm 3\sigma$ which measures how far standard deviations are from the metric average. Fourth, classify points whether they are anomalous or not. For any given x_i , if $x_i \notin (\mu + 3\sigma, \mu - 3\sigma)$, then x_i is classified as anomalous.

In Fig. 9, the estimated bounds defines the range of acceptable deviation, which represents the bounds of the anomaly

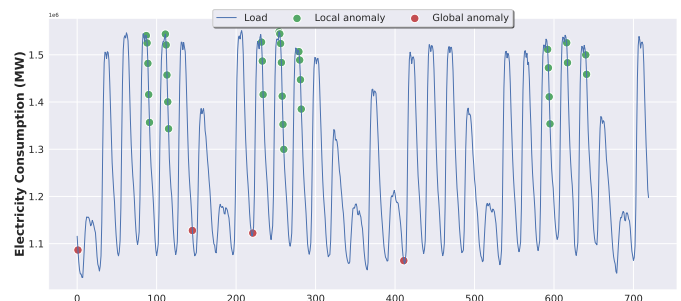


Fig. 8: Local and global anomalies detected in April 2018.

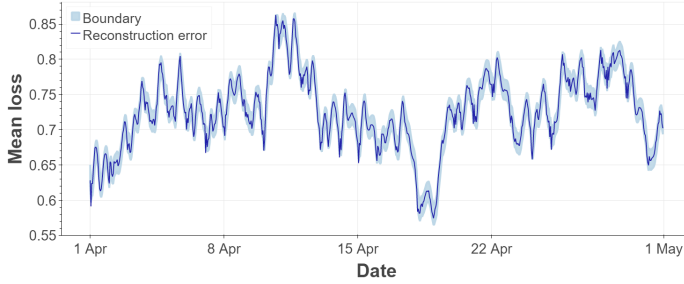


Fig. 9: The estimated bounds for the anomaly detector using the dynamic threshold.

detector. Any actual value that is outside of the upper or lower acceptable range is flagged as anomalous.

The mean that has been calculated in this study to estimate anomaly detector bounds is a moving average which is called EMA. The aim of using a moving average is to use a dynamic threshold which is sensitive to the behavior of data and changes more smoothly over time considering that the noise is seasonal and therefore anomalies are better classified. Fig. 10, illustrates the difference in electricity consumption between a day containing anomalous consumption and a random day with the same weekday value does not contain anomalous consumption. It shows a large gap between the normal and abnormal electricity consumption where a drastic increase in electricity consumption is noticed in the day that has a global anomaly period.

V. CONCLUSION

This paper presented an unsupervised deep learning based two-step approach to detect anomalies in electricity consumption data one hour ahead of time. An LSTM regular neural network was used to predict the next hour sample and an LSTM autoencoder was developed to learn the features of normal consumption. The output of the LSTM regular is fed as an input for the LSTM autoencoder. The two-step approach combines two different anomaly detection techniques: global and local anomaly detection. Several external features were considered and a feature selection method was applied to find the optimal features that achieve better results. According to the evaluation, the temporal and lag features have improved

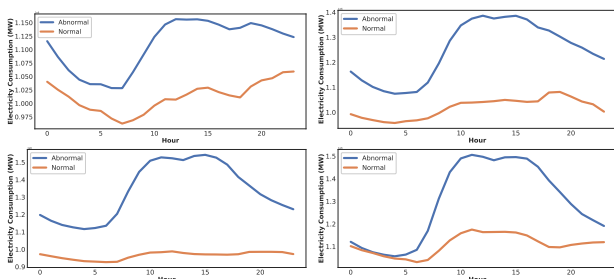


Fig. 10: The difference between the normal and abnormal electricity consumption.

the efficiency of our method to identify anomalies in the data due to the seasonality in electricity consumption data.

In the future, proposing another method for evaluating the performance of anomaly detection needs to be considered. Such a method should be able to label the data. By labeling the data, the quality of the general model will likely to improve and hence a higher accuracy could be obtained.

REFERENCES

- [1] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," *ACM computing surveys (CSUR)*, vol. 41, no. 3, pp. 1–58, 2009.
- [2] M. Rezapour, "Anomaly detection using unsupervised methods: credit card fraud case study," *Int J Adv Comput Sci Appl*, vol. 10, no. 11, 2019.
- [3] S.-C. Yip, W.-N. Tan, C. Tan, M.-T. Gan, and K. Wong, "An anomaly detection framework for identifying energy theft and defective meters in smart grids," *International Journal of Electrical Power & Energy Systems*, vol. 101, pp. 189–203, 2018.
- [4] A. A. Korba, N. Tamani, Y. Ghamri-Doudane *et al.*, "Anomaly-based framework for detecting power overloading cyberattacks in smart grid ami," *Computers & Security*, vol. 96, p. 101896, 2020.
- [5] Y. Weng, N. Zhang, and C. Xia, "Multi-agent-based unsupervised detection of energy consumption anomalies on smart campus," *IEEE Access*, vol. 7, pp. 2169–2178, 2018.
- [6] W. Cui and H. Wang, "A new anomaly detection system for school electricity consumption data," *Information*, vol. 8, no. 4, p. 151, 2017.
- [7] J.-S. Chou and A. S. Telaga, "Real-time detection of anomalous power consumption," *Renewable and Sustainable Energy Reviews*, vol. 33, pp. 400–411, 2014.
- [8] W. Hurst, C. A. C. Montañez, and N. Shone, "Time-pattern profiling from smart meter data to detect outliers in energy consumption," *IoT*, vol. 1, no. 1, pp. 92–108, 2020.
- [9] X. Liu, Y. Ding, H. Tang, and F. Xiao, "A data mining-based framework for the identification of daily electricity usage patterns and anomaly detection in building electricity consumption data," *Energy and Buildings*, vol. 231, p. 110601, 2021.
- [10] M. Braei and S. Wagner, "Anomaly detection in univariate time-series: A survey on the state-of-the-art," *arXiv preprint arXiv:2004.00433*, 2020.
- [11] J. Lye and J. Hirschberg, "Time series plots: A compendium of time series plots." 2020.
- [12] D. H. Vu, K. M. Muttaqi, and A. P. Agalgaonkar, "Assessing the influence of climatic variables on electricity demand," in *2014 IEEE PES General Meeting— Conference & Exposition*. IEEE, 2014, pp. 1–5.
- [13] T. AlSkaif, S. Dev, L. Visser, M. Hossari, and W. van Sark, "A systematic analysis of meteorological variables for pv output power estimation," *Renewable Energy*, vol. 153, pp. 12–22, 2020.
- [14] A. L. Bonkany, I. Seidou Sanda, and A. A. Balogun, "Wavelet analysis of daily energy demand and weather variables," *Journal of Energy*, vol. 2019, 2019.
- [15] K. Chapagain, S. Kittipiyakul, and P. Kulthanavit, "Short-term electricity demand forecasting: Impact analysis of temperature for thailand," *Energies*, vol. 13, no. 10, p. 2498, 2020.
- [16] T. Donkers, B. Loepp, and J. Ziegler, "Sequential user-based recurrent neural network recommendations," in *Proceedings of the Eleventh ACM Conference on Recommender Systems*, 2017, pp. 152–160.
- [17] Y. Su and C.-C. J. Kuo, "On extended long short-term memory and dependent bidirectional recurrent neural network," *Neurocomputing*, vol. 356, pp. 151–161, 2019.
- [18] D. Bank, N. Koenigstein, and R. Giryes, "Autoencoders," 2020.
- [19] P. Malhotra, A. Ramakrishnan, G. Anand, L. Vig, P. Agarwal, and G. Shroff, "Lstm-based encoder-decoder for multi-sensor anomaly detection," *arXiv preprint arXiv:1607.00148*, 2016.
- [20] L. Visser, T. AlSkaif, and W. Van Sark, "The importance of predictor variables and feature selection in day-ahead electricity price forecasting," in *2020 International Conference on Smart Energy Systems and Technologies (SEST)*. IEEE, 2020, pp. 1–6.
- [21] H. H. Bosman, G. Iacca, A. Tejada, H. J. Wörtche, and A. Liotta, "Spatial anomaly detection in sensor networks using neighborhood information," *Information Fusion*, vol. 33, pp. 41–56, 2017.