

Load Forecasting based on Deep Neural Network and Historical Data Augmentation

Chun Sing Lai ^{1,2}, Zhenyao Mo ³, Ting Wang ³, Haoliang Yuan ¹, Wing W. Y. Ng ³, Loi Lei Lai ^{1,*}

¹ Department of Electrical Engineering, School of Automation, Guangdong University of Technology, Guangzhou, China

² Brunel Institute of Power Systems, Department of Electronic and Computer Engineering, Brunel University London, London, UB8 3PH, UK

³ Guangdong Provincial Key Lab of Computational Intelligence and Cyberspace Information, School of Computer Science and Engineering, South China University of Technology, Guangzhou 510630, China

* l.l.lai@ieee.org

Abstract—Load forecasting is a complex non-linear problem with high volatility and uncertainty. This paper presents a novel load forecasting method known as deep neural network and historical data augmentation (DNN-HDA). The method utilizes HDA to enhance regression by DNN for monthly load forecasting, considering that the historical data to have a high correlation with the corresponding predicted data. To make the best use of the historical data, one year's historical data is combined with the basic features to construct the input vector for a predicted load. In this way, if there is C years' historical data, one predicted load can have C input vectors to create the same number of samples. DNN-HDA increases the number of training samples and enhances the generalization of the model to reduce the forecasting error. The proposed method is tested on daily peak loads from 2006 to 2015 of Austria, Czech and Italy. Comparisons are made between the proposed method and several state-of-the-art models. DNN-HDA outperforms DNN by 44%, 38% and 63% on the three datasets, respectively.

Keyword—deep neural network, data augmentation, load forecasting, regression

Nomenclature

M	Month
D	Day
Y	Year
W	Week
H	Holiday
$y_{(M,D,Y)}$	Daily peak load of M month D day, Y year
$E_{(M,D,Y)}$	Exogenous features of M month D day, Y year

$B_{-L}_{(M,D,Y)}^o$	Neighboring loads before $y_{(M,D,Y)}$ and o denotes the number of these loads
$x_{(M,D,Y)}$	Input vector without historical data augmentation
C	Number of years of historical loads
$B_{-L}_{(M,D,Y-t)}^p$	Adjacent loads before $y_{(M,D,Y-t)}$ and p is the number of these loads, $t \in [1, C]$
$A_{-L}_{(M,D,Y-t)}^q$	Adjacent loads after $y_{(M,D,Y-t)}$ and q is the number of these loads, $t \in [1, C]$
$H_{-D}_{(M,D,Y-t)}$	Historical feature vector of year $Y-t$, $t \in [1, C]$
$S_{(M,D,Y)}$	Input vector for historical data concatenation (HDC)
$X_{(M,D,Y)}$	Input vector for historical data augmentation (HDA)
L	Number of layers in the deep neural network (DNN) model
x_i^l	Input of the l^{th} layer of the i^{th} sample, $l \in [1, L]$
b^l	Biases of the neurons in the l^{th} layer
$\varphi(\cdot)$	Activation function
W^l	Weights matrix of the l^{th} layer
N	Number of training samples
μ	Penalty coefficient of the model complexity

E	Loss of the deep neural network (DNN) model
y_i	True label of the i^{th} sample
\bar{y}_i	Model's output of the i^{th} sample
v	Number of evaluated samples

I. INTRODUCTION

Load forecasting plays an essential role in many areas such as energy trading in power markets [1-3], system security assessments [4] and infrastructure maintenance [5]. Furthermore, load forecasting is getting increasingly important since the beginning of the 21st century because of the development of the smart grid [6].

However, load forecasting is a challenging problem due to the high volatility and uncertainty. For instance, there is a lower electricity demand in rainy days than that of the sunny days in summer. The lower temperature will cause a less number of air-conditioners to operate, which is an electricity intensive appliance. Nevertheless, the weather is irregular and becomes an uncertain factor for the prediction of loads. The emergence of renewable energy (e.g. wind and solar) also increases the difficulty of forecasting due to their non-regular behavior. To encourage the customers to stagger the peak time of electricity consumption, the decision-makers of power market have designed various electricity prices in a different time, which encourages customers to use electricity when the energy cost is low. This strategy also increases the uncertainty of forecasting, since its influence to the consumption behavior of users is difficult to quantify. All of these variables bring volatility and uncertainty to load forecasting, which makes it a complex non-linear problem.

Classical statistical methods like autoregressive integrated moving average (ARIMA) [7], exponential smoothing method [8], multiple linear regression [9] and Kalman filter [10] were proposed as the common methods for load forecasting. Traditional methods explicitly map the input to the output and are easy to implement, but these methods have drawbacks with the highly non-linear problem. More attentions are paid on the machine learning models because of their great ability in building the non-linear mapping. Methods such as radial basis function neural network (RBFNN) [11], artificial neural network (ANN) [12], support vector regression (SVR) [13] and extreme learning machine (ELM) [14] had been applied for load forecasting problem. However, the above machine learning methods suffer from under-fitting and over-fitting issues. Some scholars concentrate on finding a better way to tune parameters; for the model to avoid staying in the local optimum. They combine evolutionary algorithms and machine learning models to reach this goal [6, 15]. Authors in Reference [16] tried the methods including cross-validation, early stopping and

regularization which are commonly applied on machine learning to escape from the local optimum. Considering the implicit relationships between similar but distinct domain data, the authors proposed a two-layer architecture to improve the accuracy of the model based on transfer learning [17]. In addition, an investigation was devoted to aggregating a series of learning models to avoid the over-fitting and strengthen model's ability [18]. To obtain the optimal structure of RBFNN and avoid from the over-fitting and under-fitting problems in the training process, Abedinia and Amjady [19] constructed a stochastic search method to find the suitable number neurons in the hidden layer of RBFNN. In Reference [20], a feature selection method based on information-theoretic criteria was utilized to get the non-linearities and interacting features which improves the forecasting model by modeling interaction, relevancy and redundancy. To create a forecasting model with high accuracy, Reference [21] provided a hybrid prediction model based on a feature selection method which chooses the best candidate inputs and enhanced support vector machine which fine tunes the free parameter of the forecast engine to tackle the prediction of aggregated loads of buildings and the impact of electric vehicle. Another method based on non-linear optimization was introduced in Reference [22] which supports the assessment of load forecasts in local energy markets (LEM) simulations by generating erroneous load profiles and decreases the implications of forecast errors.

In recent years, deep learning [23] is considered by the research community. Comparing to other machine learning methods, their deep structure enhances the feature extraction capability and helps studying the complicated non-linear mapping better. To show the effectiveness of deep learning methods, the comparison between two classical deep learning methods (recurrent neural network (RNN) and convolutional neural network (CNN)) and one popular traditional method autoregressive integrated moving average with exogenous inputs (ARIMAX) is exhibited in [24]. The experiments indicate that deep learning methods have improved results compared to traditional statistical methods. In [25], the authors constructed a deep neural network (DNN) model with a restricted Boltzmann machine (RBM) as the pre-training model to initial DNN's weights to get an accurate prediction model. Because the load forecasting is a time series task and long short-term memory (LSTM) does well in abstracting sequence features, [26] proposed an LSTM recurrent neural network-based framework to deal with the task of load forecasting for individual residential households. Dedinec et al. [27] utilized deep belief networks (DBN) to solving the load forecasting problem and obtaining a better performance, as compared with a typical feed-forward multi-layer perceptron neural network and Macedonian system operator. To learn the relationship from the training data more effectively, [28] proposed a model based on deep residual network (DRN) with the ability to integrate

Table 1 Function comparisons between the proposed method and state-of-the-art models (Y: Yes; N: No)

Functions	[13]	[14]	[15]	[16]	[17]	[18]	[19]	[20]	[21]	[22]	[25]	[26]	[27]	[28]	Proposed method
Feature selection	N	N	Y	N	N	N	N	Y	Y	Y	N	N	Y	N	N
Deep learning	N	N	N	N	Y	N	N	N	N	N	Y	Y	Y	Y	Y
Optimizing the hyper parameters of learning model	N	N	Y	Y	N	N	Y	N	Y	N	N	N	N	N	N
Data augmentation	N	N	N	N	N	N	N	N	N	N	N	N	N	N	Y
Multiple outputs	Y	Y	N	N	N	Y	N	N	N	N	N	N	N	Y	Y

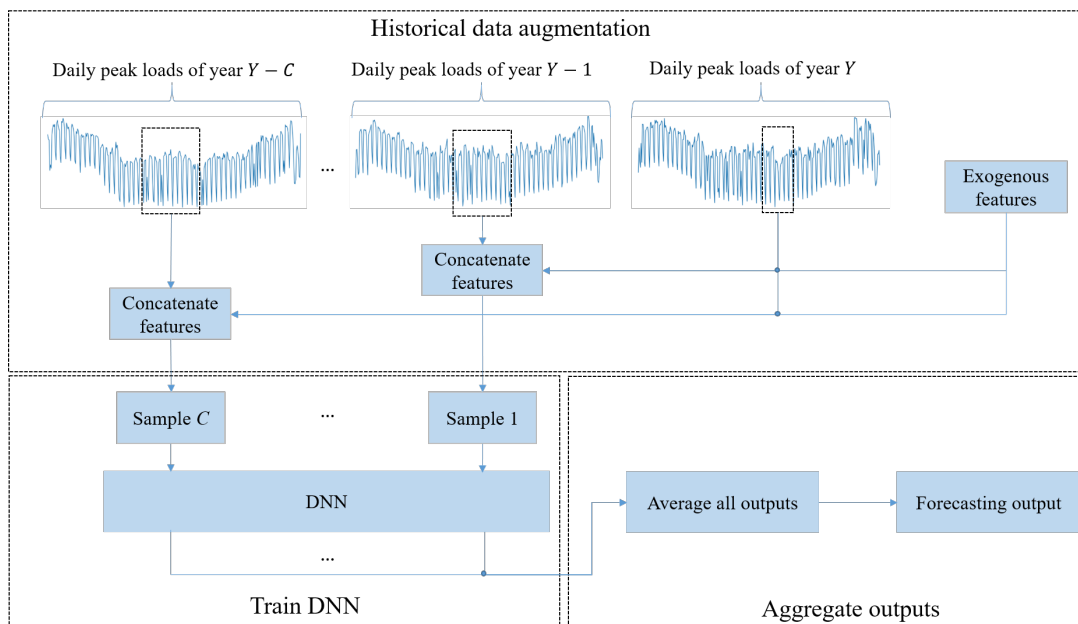


Fig. 1. The framework of the DNN-based regression model and HDA

domain knowledge and researchers' understanding of the task under different neural network building blocks. In our work, DNN is adopted as the regression model because of its extraordinary capability for the non-linear task.

In our study, it is found that loads of different years on the same date (LDYSD), such as the daily peak load of December 1, 2014, and that of December 1, 2015, have a strong correlation based on correlation analysis [29]. The discovery indicates that the corresponding historical loads have some implicit relationships with the load to be predicted. It inspires us to utilize this nature to improve the forecasting model. To make full use of these historical loads, historical data augmentation (HDA) is proposed to strengthen the capability of the model. It uses the concept of data augmentation [30] which focuses on increasing the number samples and provides additional information to the learning model simultaneously. This feature

strengthen the model, its generalization ability and avoid from over-fitting. Many methods use data augmentation to tackle their problems such as sound classification [31-33]. To the best knowledge of the authors, there are few methods about applying data augmentation on load forecasting.

The contributions of this paper can be summarized as follows:

1. Combining the load data of the same date in different years to augment the data for a better performance. This fully utilizes the high correlation of data for the same period in different years.
2. A forecasting model consisting of DNN and HDA is proposed. The DNN can learn the non-linear relationships for load forecasting. The HDA provides extra information to the DNN which strengthens its

robustness and improves the performance. The combination of these two methods yields a more effective load forecasting model.

3. The proposed HDA method has several corresponding samples which bring the same number of outputs after sending them into the trained model. Thus, these outputs can be aggregated to improve the model further. Experimental results show that the proposed method outperforms several existing methods. Comparing with only using DNN, the proposed method outperforms it by 44%, 38% and 63% on the three datasets, respectively.

To demonstrate the contributions of the proposed method as compared to other state-of-the-art models, the function comparisons are summarized in Table 1. It is shown that some researchers focused on the feature selection which aims to find the good input for forecasting model to improve its ability [15, 20-22, 27]. Some studies have applied deep learning to better deal with the non-linear mapping problem in load forecasting [25-28]. To avoid the over-fitting and under-fitting problems in machine learning, authors in references [15-16, 19, 21] made contributions on optimizing the hyper parameters of learning model. There are other studies that employed different methods to generate multiple outputs for one load; to predict and aggregate them to get a higher accuracy [13-14, 18, 28]. In our work, HDA is proposed to increase the number of samples and enhance the generalization of the model. This method also helps a forecasted load to get multiple outputs and improve the model's performance further. Meanwhile, DNN is adopted as the forecasting engine to solve the non-linearity of load forecasting.

The paper is organized as follows: Section II describes a novel method for load forecasting. The case studies with real-life data, its experimental results and discussion are presented in Section III. Finally, the conclusion and future work are given in Section IV.

II. DNN WITH HDA FOR LOAD FORECASTING

The proposed DNN-HDA method consists of two parts: the DNN-based regression model and the HDA. Assuming that the loads are forecasted in year Y and there are C -year historical data. In the procedure of HDA, for each load to be predicted, several features could be used to construct C samples. The features include the most recent loads of the predicted load, LDYSD, the neighboring loads of LDYSD and the exogenous features like a holiday, calendar information, weather etc. After generating samples, the deep neural network is trained with the training data. For one predicted load, there are C outputs that will be averaged to generate the final prediction. Fig. 1 shows the whole process of DNN-HDA. In the following subsections, the detail about HDA and the construction of DNN will be given.

A. Historical data augmentation

In this subsection, the proposed method of HDA will be elaborated. The components of input are firstly depicted since this method is based on the construction of input. Subsequently, the proposed HDA is introduced fully.

In this paper, the commonly used features including past load values, calendar information and holiday are selected as the input attributes for three datasets. The calendar information includes year, month, day and day of the week. Assuming that $y_{(M,D,Y)}$ identifies the current load value to be predicted. M , D and Y denote current date's month, day and year, respectively. Let W and H denotes the day of week and holiday information of the predicted day, respectively. Therefore, the exogenous features can be represented as:

$$E_{(M,D,Y)} = (Y, M, D, W, H) \quad (1)$$

Note that all these exogenous data are transformed into one-hot encoding in the experiments.

Suppose $B_L_{(M,D,Y)}^o$ represents the neighboring load values before $y_{(M,D,Y)}$ and o denotes the number of these load values. The common input vector $x_{(M,D,Y)}$ without HDA can be formulated as follows:

$$x_{(M,D,Y)} = (B_L_{(M,D,Y)}^o, E_{(M,D,Y)}) \quad (2)$$

HDA will be described after describing the input features. Many researchers consider features like the past load values, calendar information, weather as the composition of input. Besides these features, there is another effective attribute which is ignored by many studies. That is load values of different years on the same date (LDYSD). For example, when there is each day's peak load value from 2012 to 2014 and to compute $y_{(12,1,2015)}$, not only the most recent load values should be considered, but also the historical same period data (i.e. $y_{(12,1,2012)}$, $y_{(12,1,2013)}$ and $y_{(12,1,2014)}$) should be taken into account. The motivation to include LDYSD into the input is because of the historical samples show a high correlation with the corresponding predicted data. To prove this, the correlation coefficient analysis is made between loads of different years. There are 10 years' loads for the three nations. Applying all the combinations is unnecessary, so 3 years (2009, 2010 and 2014) daily peak load values of Czech are randomly selected to carry out the correlation coefficient analysis with those of 2015. The results are displayed in Fig. 2 to Fig. 4. Because the inputs of the correlation coefficient analysis are two vectors, therefore, each year's load value is divided into 4 parts according to the seasons and set them as input vectors. From the table, the corresponding season's loads of different years always have a positive and high coefficient, especially in the analysis of loads of 2009 and 2015, whose correlation coefficients are almost higher than 0.9. The analysis implies that electricity consumption of the corresponding day from different years are

highly relevant. This phenomenon promotes us to apply LDYSD on the model.

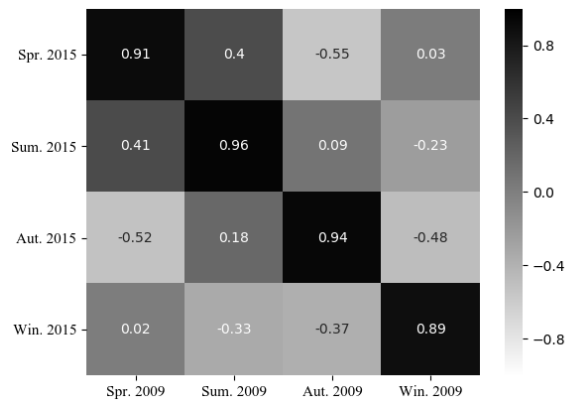


Fig. 2. Correlation coefficient between loads of 2009 and 2015

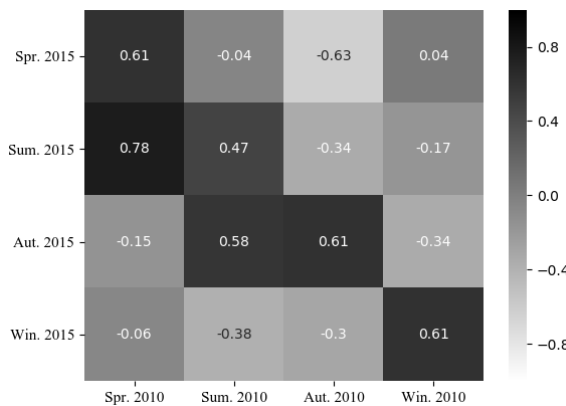


Fig. 3. Correlation coefficient between loads of 2010 and 2015

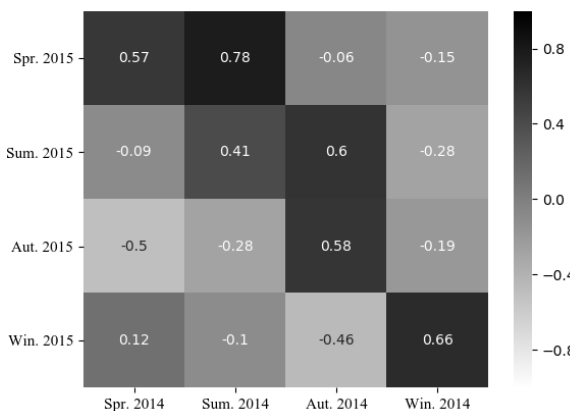


Fig. 4. Correlation coefficient between loads of 2014 and 2015

To make full use of LDYSD, the neighboring points before and after them are considered. For instance, the load values for December 2-10, 2014 and December 2-10, 2013 can be utilized to compute the load value of December 2-10, 2015 i.e. $y_{(12,7,2015)}$. Suppose that there are C -year historical load values and the current load value to be forecasted is $y_{(M,D,Y)}$, then the

corresponding historical load of t years ago can be expressed as $y_{(M,D,Y-t)}$, $t \in [1, C]$. Let $B_{L_{(M,D,Y-t)}}^p$ represents the adjacent loads before $y_{(M,D,Y-t)}$ and p is the number of these loads. $A_{L_{(M,D,Y-t)}}^q$ denotes the adjacent loads after $y_{(M,D,Y-t)}$ and q is the number of these loads. The historical feature vector of year $Y-t$ can be presented as:

$$H_{D_{(M,D,Y-t)}} = (B_{L_{(M,D,Y-t)}}^p, y_{(M,D,Y-t)}, A_{L_{(M,D,Y-t)}}^q, Y-t) \quad (3)$$

The simplest way to use these historical data is to concatenate all $H_{D_{(M,D,Y-t)}}$ and append them to the original input $x_{(M,D,Y)}$. The new input vector $s_{(M,D,Y)}$ can be written as:

$$s_{(M,D,Y)} = (x_{(M,D,Y)}, H_{D_{(M,D,Y-1)}}, \dots, H_{D_{(M,D,Y-t)}}, \dots, H_{D_{(M,D,Y-C)}}) \quad (4)$$

This method is named as historical data concatenation (HDC). However, this method has a defect that only the load values of year Y have enough historical data to construct the samples to ensure that all inputs have the same dimension. In other words, it cannot use load values before year Y as a label to generate samples. This leads to the number of samples to be much less.

HDA is proposed to overcome the shortage of HDC. HDA inserts one $H_{D_{(M,D,Y-t)}}$ into the original input to generate a new sample, instead of aggregating the historical data of all previous years into the one input. C -year historical data can create C new samples for one load to be predicted. The process of HDA is shown in Fig. 1. The inputs set $X_{(M,D,Y)}$ of $y_{(M,D,Y)}$ can be written as:

$$X_{(M,D,Y)} = \{x|x = (x_{(M,D,Y)}, H_{D_{(M,D,Y-t)}})\}, t \in \{1, \dots, C\} \quad (5)$$

HDA increases the number of samples and brings historical features to the learning model at the same time. In this way, HDA enhances the diversity of samples and helps the model be able to have a better performance. Furthermore, because the predicted load value has several corresponding inputs, several outputs could be obtained after feeding the inputs to the regression model. These outputs can be aggregated to get an enhanced result. In this paper, these outputs are averaged as the final result.

It is obvious that HDA has more advantages than HDC. To examine HDA and prove its effectiveness, the comparison between two methods will be shown in Section III.

B. Deep Neural Network

Deep learning [23] is gaining popularity in the past few years because of its excellent ability to construct complicated relationships between two domains. Deep learning greatly promotes the development of artificial intelligent (AI) and makes extraordinary successes in many fields, such as computer vision, natural language processing and speech recognition. Load forecasting is a non-linear problem, while many researchers have demonstrated that deep learning is good at

dealing with this type of task in recent years. Therefore in this work, a DNN model is constructed to predict the load values.

To build the DNN model, it is necessary to determine the number of hidden layers and neurons. However, it is impossible to test all the combinations of layers and neurons. A simple and effective way is adopted to determine the model's parameters as follows: 50 is set as the initial value for the number of neurons, which is increased or decreased in each layer and then the combination of layers and neurons is enumerated to create the models. With sensitivity analysis, the optimal performance of the proposed DNN has the parameters of 4 hidden layers with 100 neurons per layer.. The structure of the DNN is shown in Fig. 5.

Suppose that L is the number of layers in the model. Let $x_i^l = (x_{i,1}^l, x_{i,2}^l, \dots, x_{i,n}^l)$ identifies the input of the l^{th} layer of the i^{th} sample, $l \in [1, L]$. The mapping function in the l^{th} layer can be

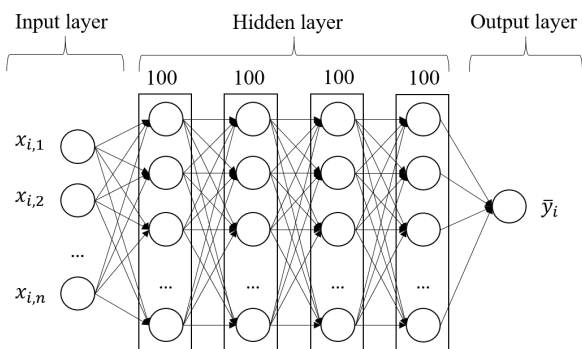


Fig. 5. Structure of DNN

written as:

$$x_i^{l+1} = \varphi(W^l x_i^l + b^l), l \in [1, L] \quad (6)$$

where b^l represents the biases of the neurons in the l^{th} layer and $\varphi(\cdot)$ denotes the activation function. W^l is the weights matrix of the l^{th} layer.

To optimize the parameters of the model, generally, the backpropagation is applied to minimize the following loss function:

$$E = \frac{1}{N} \sum_{i=1}^N \|y_i - \bar{y}_i\| + \mu \sum_{l=1}^L \|W^l\|_2^2 \quad (7)$$

where N identifies the number of samples and μ is the penalty coefficient of the model complexity. y_i and \bar{y}_i are the true label and the model's output of the i^{th} sample, respectively. The first part of a loss function $\frac{1}{N} \sum_{i=1}^N \|y_i - \bar{y}_i\|$ represents the empirical risk, while the second part $\mu \sum_{l=1}^L \|W^l\|_2^2$ is the regularization term to control the model complexity.

III. CASE STUDY

In this section, the details about the experiment is introduced and the description of the dataset, the process of implementation and the results on the three datasets are included.

A. Dataset

In this work, experiments based on the daily peak loads of the three nations (Austria, Czech and Italy) are carried out. The data can be downloaded from the website open power system data platform [34]. The website provides 37 European national loads from 2006 to 2015 in hourly data format. The maximization of the hourly loads per day as the daily peak load is considered. For each dataset, there are ten years of load values. Among these datasets, load values in December 2015 are selected as the test set, while the others are used for the model training. For the training part, at each run, 25% of the samples are randomly chosen as the validation set, to decide the best model and the rest is the training set. Each experiment is run for 10 times.

Note that all the following experiments are monthly forecasting since a month data is chosen as the test set. All the loads in December 2015 are regarded as the unknown data. To obtain the daily peak load values in December 2015, the known data is firstly used to forecast the load on December 1, 2015. Then the load value for December 2, 2015 can be predicted with the load value on December 1, 2015. In this way, all the predictions are obtained for December 2015.

B. Implementation

The code is implemented by PyTorch platform [35]. The value of hyper-parameters like epoch, batch size, learning rate and L_2 regularization coefficient μ are displayed in Table 2. For all the datasets, stochastic gradient descent (SGD) is adopted as the optimizer.

Table 2 Setting of parameters

	Epoch	Batch size	Learning rate	μ
Austria	200	128	0.01	0.001
Czech	200	128	0.005	0.001
Italy	200	128	0.01	0.001

For the evaluation criteria, the popular measurement mean absolute percentage error (MAPE) is employed to evaluate the proposed method in this study. The equation of the index is:

$$\text{MAPE} = \frac{\sum_{i=1}^v \frac{|y_i - \bar{y}_i|}{y_i}}{v} \times 100\% \quad (8)$$

where ν is the number of evaluated samples, y_i denotes the true load of the i^{th} sample and \bar{y}_i represents the output from the model of the i^{th} sample.

C. Experiment on the three datasets

The results of the proposed method on the three datasets are shown in this subsection. Firstly, a sensitivity analysis is carried out to decide the values of o , p and q as mentioned in Section

Table 3 MAPE (%) performances of different p and q based on DNN-HDA

		Austria			Czech			Italy		
p	q	Mean±Std	Max	Min	Mean±Std	Max	Min	Mean±Std	Max	Min
5	0	3.55±0.52	4.43	2.78	4.76±0.84	6.09	3.54	4.84±0.67	5.75	3.90
7	0	3.19±0.37	3.76	2.66	4.59±0.65	5.72	3.86	3.76±0.66	4.82	2.50
9	0	3.53±0.45	4.18	2.78	4.50±0.50	5.25	3.78	3.62±0.48	4.38	2.87
11	0	3.52±0.60	4.57	2.70	4.52±0.48	5.60	4.02	4.17±0.74	5.76	3.32
0	5	3.28±0.91	5.56	2.58	4.01±0.45	5.55	2.98	3.73±0.57	4.99	3.20
0	7	2.85±0.40	4.17	2.71	3.69±0.48	5.16	3.33	3.16±0.43	3.65	2.45
0	9	3.06±0.45	3.88	2.26	3.77±0.45	5.79	3.14	3.21±0.33	3.95	2.88
0	11	3.11±0.49	3.81	2.34	3.80±0.70	5.22	3.07	3.19±0.47	3.94	2.67
5	5	2.96±0.56	4.20	2.36	3.86±0.79	5.27	2.52	3.18±0.36	3.81	2.79
7	7	2.79±0.47	3.60	2.14	3.48±0.57	4.83	2.87	2.97±0.66	4.24	2.33
9	9	3.01±0.38	4.00	2.62	3.80±0.81	5.25	2.83	3.15±0.34	3.43	2.30
11	11	2.93±0.35	3.54	2.45	3.60±0.96	5.70	2.31	3.05±1.04	5.81	2.16

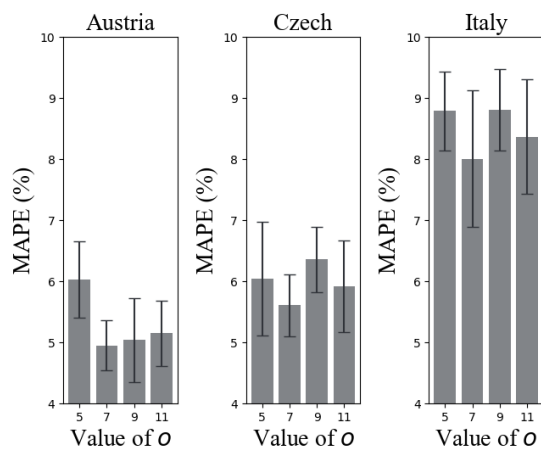


Fig. 6. Performances of different o values on the three datasets

1) Sensitivity analysis

The structure of the model input is depicted in Section II. The input vector consists of several variables. Among these variables, the most recent loads of the predicted load $B_L^o_{(M,D,Y)}$,

II. Subsequently, the discussion for the effect of historical loads' amount on the proposed method is provided. Next, a comparison with other methods is made to demonstrate the performance of the proposed method. Finally, the influence of seasonality on the method will be discussed.

the adjacent points before and after the LDYSD (i.e. $B_L^p_{(M,D,Y-t)}$ and $A_L^q_{(M,D,Y-t)}$) are considered to be important. To utilize these features, the values of their parameters o , p and q have to be calculated first.

In the experiments, the parameter o is examined first. Since the authors only test the values of o and do not include the other two parameters, several o values based on DNN without using the proposed HDA are examined. For the real daily peak loads, the load values of one week forms a cycle because the load demand on the weekend is lower than that on a weekday. Generally, the load data of one week before the load to be predicted is suitable for the input. Therefore, 7 is a suitable value for o and also different values are chosen. Meanwhile, 5, 7, 9 and 11 are selected as the value of o to examine its performance on the three datasets. The comparison results are presented in Fig. 6. These results show that when the value of o is 7, the lowest mean and standard deviation for both Austria and Czech datasets can be achieved. However, such value could not be obtained on Italy dataset. As an evaluation index, mean is a better choice to represent the performance of a model comparing to the standard deviation. Therefore, 7 is regarded as the most suitable value of o for Italy dataset, since it obtains the

best average MAPE. According to the experiments, it is shown that one week's data before the predicted load can actually give the satisfactory features for the forecasting.

The determination of p and q is required after determining o . 5, 7, 9 and 11 are selected as the sensitivity values for p and q . To carry out the sensitivity analysis, 12 combinations of four selected values are chosen for comparison. Table 3 gives the MAPE results of different p and q values for DNN-HDA, with the lowest value of each column is highlighted in bold. From

Table 3, it is shown that no combination of values can outperform others on all measurements on three datasets. However, the lowest mean MAPE could be obtained when setting 7 for both p and q . Therefore 7 is regarded as the most suitable value for p and q for the three datasets.

Table 4 MAPE (%) of different number of historical years for DNN-HDA

Years	Austria			Czech			Italy		
	Mean±Std	Max	Min	Mean±Std	Max	Min	Mean±Std	Max	Min
3	3.62±0.45	4.34	3.22	5.23±0.58	6.13	4.49	4.64±0.40	5.20	4.08
5	3.22±0.31	3.72	2.75	3.81±0.88	4.96	2.46	4.10±0.35	4.71	3.62
7	2.89±0.48	3.33	2.34	3.61±0.56	4.67	2.86	3.54±0.50	4.25	2.88
9	2.79±0.60	3.60	2.14	3.48±0.57	4.83	2.67	2.97±0.66	4.24	2.33

Table 5 MAPE (%) comparison of various models on the three datasets

Model	Austria			Czech			Italy		
	Mean±Std	Max	Min	Mean±Std	Max	Min	Mean±Std	Max	Min
ARIMA	6.43±0	6.43	6.43	6.43±0	6.43	8.32	10.53±0	10.53	10.53
DNN	4.95±0.41	5.69	4.55	5.61±0.51	6.44	4.68	8.01±1.12	10.19	6.60
LSTM [26]	4.80±0.62	6.11	5.46	5.78±1.26	8.03	4.21	6.40±1.71	9.39	4.57
DRN [28]	3.21±0.35	3.50	3.01	5.22±0.23	5.52	4.90	9.09±0.51	9.81	8.24
DNN-HDC	3.19±0.48	4.40	2.60	5.05±0.55	6.01	4.29	5.46±0.50	6.10	4.62
DNN-HDA	2.79±0.60	3.60	2.14	3.48±0.57	4.83	2.87	2.97±0.66	4.24	2.33

Through the experiments, it is shown that the best value for all three parameters on all datasets is 7. The results showed that using the load values of one cycle as an input feature is suitable for the prediction. Also, using both $B_L^p_{(M,D,Y-t)}$ and $A_L^q_{(M,D,Y-t)}$ can reduce the MAPE, which indicates that the performance of adding both of these two features into the input is better than that of only utilizing one of them. Furthermore, the proposed method is robust and less sensitive to the values of p and q in the case of applying both $B_L^p_{(M,D,Y-t)}$ and $A_L^q_{(M,D,Y-t)}$, because different values have similar performances on the three datasets.

2) Discussion on the number of historical loads

In Section II, it is mentioned that the number of samples could be increased with HDA. More samples could be created

with additional historical data; to create more information for the learning model and to get a more accurate forecast. To verify the aforementioned statement, different amounts of historical loads are examined, and their performances are tested with DNN-HDA. Table 4 shows the results of utilizing historical data of 3, 5, 7 and 9 years.

It is shown that the lowest average MAPE is obtained by using 9-year historical loads on three datasets. The forecasting result obtained with the larger historical dataset ranks top two on the other two measurements (i.e. max and min). On the other hand, the experiments also imply that there will be little model improvement when the amount of historical loads outnumber a threshold.

3) Comparison with existing methods

To demonstrate the effectiveness of the proposed method, two benchmark models: an autoregressive integrated moving average (ARIMA) and deep neural network (DNN); and two state-of-the-art models: long short-time memory (LSTM) [26]

and deep residual neural network (DRN) [28] are selected for comparisons. The results of HDC combined with DNN (DNN-HDC) are also exhibited in this subsection. Table 5 gives the comparison results and the best result of each column is highlighted in bold.

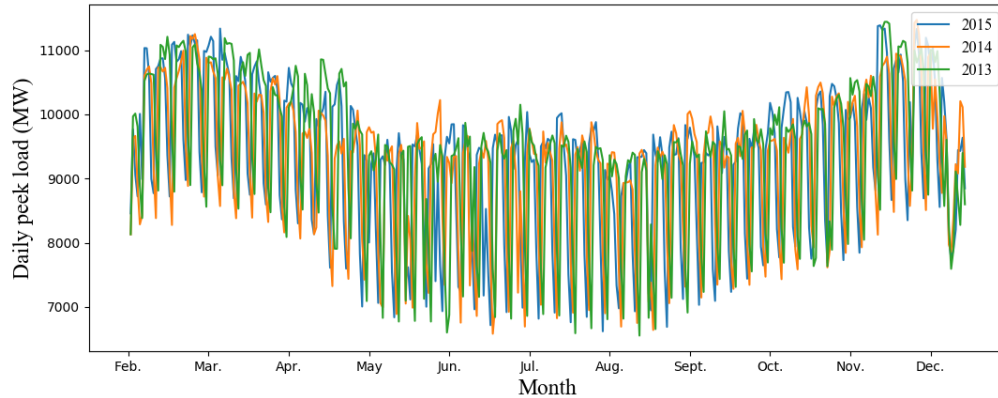


Fig. 7. Daily peak load value of Austria

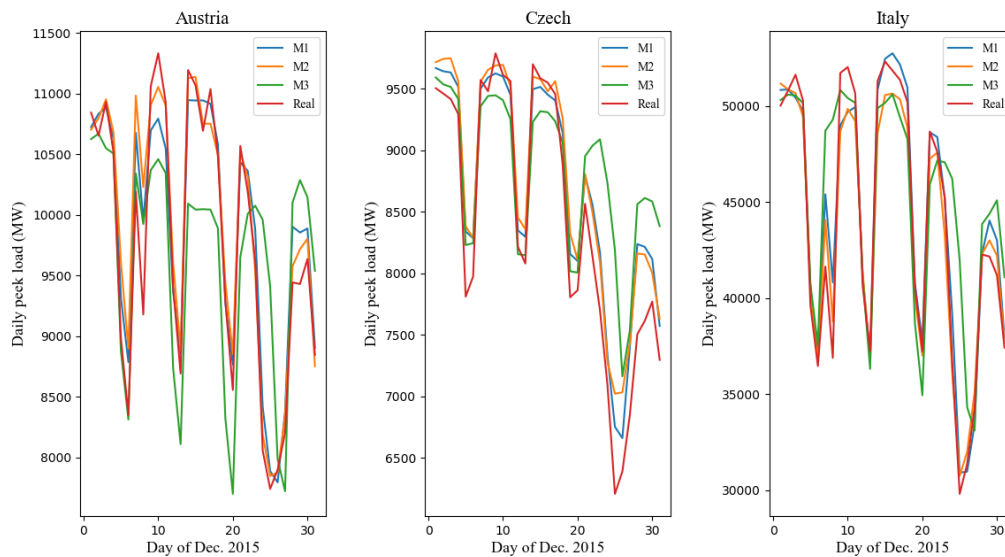


Fig. 8. Analysis of Seasonal Effect on the Proposed Method DNN-HDA (M1: Use all training data, M2: Use only training data of winter, M3: Use all training data except those of winter)

The results of the proposed methods DNN-HDC and DNN-HDA are also given in Table 5. It shows that DNN-HDA has a better performances for all datasets, and proves that HDA is a better method of utilizing historical same period load values. Moreover, the MAPE is the lowest for DNN-HDA and achieves the best results on all datasets. The comparison of DNN-HDA and DNN is reported in Table 5. The MAPE average values of DNN-HDA are 44%, 38% and 63% lower than those of DNN on three datasets, respectively. The proposed method shows enhanced improvements for DNN on three datasets especially on Italian data, because 1) of the worst performance of DNN in this dataset; 2) of the outstanding result of DNN-HDA. Comparing with the data of Austria and Czech, Italian data are

more complex which brings about the bad performance of DNN. From the results, it is shown that even with the complex data, HDA method is very effective and brings a great improvement to the learning model. It fully implies that the hybrid of DNN and HDA has enhanced performance than only using deep neural network.

In summary, HDA increases the number of samples which provide extra information to the learning model and strengthen its generalization. The results demonstrate its significant improvement to the model.

4) Influence of seasonality

The influence of season on energy demand is intuitive. The daily peak load values of Austria are examined as an example. Fig. 7 displays the 3-year daily peak loads curve of Austria from 2013 to 2015. It is shown that the monthly electricity consumption changes periodically. In summer, the load demand is the lowest in the year. In contrast, winter has the highest load demand. Also, the electricity demand gradually decreases and increases in spring and autumn, respectively. The reason for the regular change is due to the requirement of heating raised as the temperature descends. To account for this effect in the model, the training data should give prior knowledge about the season. For all the three datasets, the load values that belongs to winter in December 2015 are selected as the test set. The training data is reorganized and three new training sets for each dataset are constructed. They are M1 (Use all training data), M2 (Use only training data of winter) and M3 (Use all training data except those of winter). The three models based on the reorganized training sets for each dataset are trained. Fig. 8 shows the performances of these models.

As expected, the season influences the capability of the model. If winter data is not used in the training process, the capability of the model is the worst on all the three datasets. By comparing in using different combinations of data, it is shown that only using winter data will have better performances on the dataset of Italy. While for loads of Czech and Austria, utilizing all the data will be better than only using the winter data. The differences between these two strategies are small in all datasets. The experiments indicate that loads of the identical season as the test data are essential for the learning model. The model cannot be able to yield a good result without these load values. Also, load values of different seasons are not important for the model since they cannot effectively improve the performance of the model.

IV. CONCLUSIONS AND FUTURE WORK

This paper develops a monthly load forecasting model based on historical data augmentation. The model consists of two parts known as HDA and DNN. HDA is based on determining the load values of different years, for the same date that is highly correlated. Such method increases the number of samples and brings additional information for the model training, which enhances the generalization of the learning model and reduces the forecasting error. Meanwhile, DNN is adopted to deal with the complex non-linear relationships of load forecasting. The combination of these two methods devises an effective model for monthly load forecasting. To examine the performance of DNN-HDA, the daily peak loads of Austria, Czech and Italy have been investigated. The experiments present the comparisons between the model and several existing methods, such as ARIMA, DNN, LSTM and DRN. The method achieves the best results in all the three datasets and gains the MAPE of 2.79%, 3.48% and 2.97% on the dataset of Austria,

Czech and Italy, respectively.

As HDA increases the number of training samples, there could be an issue with producing redundant and even negative information to the learning model. One of the future works will focus on developing a sample selection algorithm, to reduce ineffective or negative samples and enhance the ability of HDA. Another important piece of future work will be on uncertainty. Inevitably, model uncertainty occurs due to uncertainties for the model parameters, lack of appropriate knowledge, or assumptions and simplification made. The uncertainty can be simply quantified based on the difference between observed and predicted values. Although the present main focus of the study is not on uncertainty, however, there is a need to investigate the uncertainty present in the model and how to refine the model.

Acknowledgments

This work was supported by National Natural Science Foundation of China under Grants 61876066 and 61572201; Guangzhou Science and Technology Plan Project 201804010245; Brunel University London BRIEF Funding, UK; the Education Department of Guangdong Province: New and Integrated Energy System Theory and Technology Research Group [Project Number 2016KCXTD022].

REFERENCES

- [1] B. Huang, D. Wu, C. S. Lai, X. Cun, H. L. Yuan, F. Y. Xu, L. L. Lai, K.F.Tsang, "Load Forecasting based on Deep Long Short-term Memory with Consideration of Costing Correlated Factor," *2018 IEEE 16th International Conference on Industrial Informatics (INDIN)*, July 2018, Porto, Portugal.
- [2] F. Y. Xu, X. Cun, M. X. Yan, H. L. Yuan, Y. Wang and L. L. Lai, "Power Market Load Forecasting on Neural Network with Beneficial Correlated Regularization," *IEEE Transactions on Industrial Informatics*, vol 14, no. 11, pp. 5050-5059, Nov. 2018.
- [3] R. Zhang, Z. Y. Dong, Y. Xu, K. Meng and K. P. Wong, "Short-term Load Forecasting of Australian National Electricity Market by an Ensemble Model of Extreme Learning Machine," *IET Generation, Transmission & Distribution*, vol. 7, no. 4, pp. 391-397, April 2013.
- [4] N. Yorino, M. Abdillah, Y. Sasaki and Y. Zoka, "Robust Power System Security Assessment under Uncertainties using Bi-Level Optimization," *IEEE Transactions on Power Systems*, vol. 33, no. 1, pp. 352-362, Jan. 2018.
- [5] A. Ahmad, N. Javaid, M. Guizani, N. Alrajeh and Z. A. Khan, "An Accurate and Fast Converging Short-term Load Forecasting Model for Industrial Applications in a Smart Grid," *IEEE Transactions on Industrial Informatics*, vol. 13, no. 5, pp. 2587-2596, Oct. 2017.
- [6] A. Lahouar and J. Ben Hadj Slama, "Day-ahead Load Forecast using Random Forest and Expert Input Selection," *Energy Conversion and Management*, vol. 103, pp. 1040-1051, Oct. 2015.
- [7] S. Barak and S. Saeedeh Sadegh, "Forecasting Energy Consumption using Ensemble ARIMA-ANFIS Hybrid Algorithm," *International Journal of Electrical Power & Energy Systems*, vol. 82, pp. 92-104, Nov. 2016.
- [8] A. Garulli, S. Paoletti and A. Vicino, "Models and Methods for Electric

- Load Forecasting in the Presence of Demand Response," *IEEE Transactions on Control Systems Technology*, vol. 23, no. 3, pp. 1087-1097, May 2015.
- [9] T. Fang and R. Lahdelma, "Evaluation of a Multiple Linear Regression Model and SARIMA Model in Forecasting Heat Demand for District Heating System," *Applied Energy*, vol. 179, pp. 544-552, Oct. 2016.
- [10] H. Takeda, Y. Tamura and S. Sato, "Using the Ensemble Kalman Filter for Electricity Load Forecasting and Analysis," *Energy*, vol. 104, pp. 184-198, June 2016.
- [11] S. Yu, K. Wang and Y.-M. Wei, "A Hybrid Self-adaptive Particle Swarm Optimization-Genetic Algorithm-Radial Basis Function Model for Annual Electricity Demand Prediction," *Energy Conversion and Management*, vol. 91, pp. 176-185, Feb. 2015.
- [12] J.-R. Dong, C.-Y. Zheng, G.-Y. Kan, M. Zhao, J. Wen, and J. Yu, "Applying the Ensemble Artificial Neural Network-based Hybrid Data-driven Model to Daily total Load Forecasting," *Neural Computing and Applications*, vol. 26, pp. 603-611, April 2015.
- [13] G.-F. Fan, L.-L. Peng, W.-C. Hong, F. Sun, "Electric Load Forecasting by the SVR Model with Differential Empirical Mode Decomposition and Auto Regression," *Neurocomputing*, vol. 173, pp. 958-970, Jan. 2016.
- [14] S. Li, L. Goel and P. Wang, "An Ensemble Method for Short-term Load Forecasting by Extreme Learning Machine," *Applied Energy*, vol. 170, pp. 22-29, May 2016.
- [15] S. Kouhi, F. Keynia and S. N. Ravadanegh, "A New Short-term Load Forecast Method based on Neuro-evolutionary Algorithm and chaotic feature selection," *International Journal of Electrical Power & Energy Systems*, vol. 62, pp. 862-867, Nov. 2014.
- [16] Z. S. H. Chan, H. W. Ngan, A. B. Rad, A. K. David and N. Kasabov, "Short-term ANN Load Forecasting from Limited Data using Generalization Learning Strategies," *Neurocomputing*, vol. 70, no. 1, pp. 409-419, Dec. 2006.
- [17] L. Cai, J. Gu and Z. Jin, "Two-layer Transfer-learning-based Architecture for Short-Term Load Forecasting," *IEEE Transactions on Industrial Informatics*, vol.16, no.3, pp. 1722-1732, March 2020.
- [18] A.S. Khwaja *et al.*, "Boosted Neural Networks for Improved Short-term Electric Load Forecasting," *Electric Power Systems Research*, vol. 143, pp. 431-437, Feb. 2017.
- [19] O. Abedinia and N. Amjady, "Short-term Load Forecast of Electrical Power System by Radial Basis Function Neural Network and New Stochastic Search Algorithm," *International Transactions on Electrical Energy Systems*, vol. 26, no. 7, pp.1511-1525, 2016.
- [20] O. Abedinia, N. Amjady and H. Zareipour, "A New Feature Selection Technique for Load and Price Forecast of Electrical Power Systems," *IEEE Transactions on Power Systems*, vol. 32, no. 1, pp. 62-74, Jan. 2017.
- [21] M. Duan, A. Darvishan, R. Mohammaditab, K. Wakil and O. Abedinia, "A Novel Hybrid Prediction Model for Aggregated Loads of Buildings by considering the Electric Vehicles," *Sustainable Cities and Society*, vol. 41, pp. 205-219, 2018.
- [22] S. Schreck, I. Prieur de La Comble, S. Thiem and S. Niessen, "A Methodological Framework to support Load Forecast Error Assessment in Local Energy Markets," *IEEE Transactions on Smart Grid*, vol. 11, no. 4, pp. 3212-3220, July 2020.
- [23] L. Yann, Y. Bengio and G. Hinton, "Deep Learning," *Nature*, vol. 521, pp. 436-444, May 2015.
- [24] M. Cai, M. Pipattanasomporn and S. Rahman, "Day-ahead Building Level Load Forecasts using Deep Learning vs. Traditional Time-series Methods." *Applied Energy*, vol. 236, pp. 1078-1088, Feb. 2019.
- [25] S. Ryu, J. Noh and H. Kim, "Deep Neural Network based Demand Side Short Term Load Forecasting," *Energies*, vol. 10, Dec. 2016. <https://doi.org/10.3390/en10010003>
- [26] W. Kong, Z. Y. Dong, Y. Jia, D. J. Hill, Y. Xu and Y. Zhang, "Short-term Residential Load Forecasting Based on LSTM Recurrent Neural Network," *IEEE Transactions on Smart Grid*, vol. 10, no. 1, pp. 841-851, Jan. 2019.
- [27] A. Dedinec, S. Filiposka, A. Dedinec and L. Kocarev, "Deep Belief Network based Electricity Load Forecasting: An Analysis of Macedonian Case." *Energy*, vol. 115, pp. 1688-1700, Nov. 2016.
- [28] K. Chen, K. Chen, Q. Wang, Z. He, J. Hu and J. He, "Short-term Load Forecasting with Deep Residual Networks," *IEEE Transactions on Smart Grid*, vol. 10, no. 4, pp. 3943-3952, July 2019.
- [29] C. S. Lai, Y. Tao, F. Xu, W.W. Ng, Y. Jia, H. Yuan, C. Huang, L. L. Lai, Z. Xu, and G. Locatelli, "A Robust Correlation Analysis Framework for Imbalanced and Dichotomous Data with Uncertainty," *Information Sciences*, vol. 470, pp. 58-77, Jan. 2019.
- [30] D. A. van Dyk and X.-L. Meng, "The Art of Data Augmentation," *Journal of Computational and Graphical Statistics*, vol. 10, no. 1, pp. 1-50, 2001.
- [31] J. Ding, B. Chen, H. Liu and M. Huang, "Convolutional Neural Network With Data Augmentation for SAR Target Recognition," *IEEE Geoscience and Remote Sensing Letters*, vol. 13, no. 3, pp. 364-368, March 2016.
- [32] N. G. Polson and S. L. Scott, "Data Augmentation for Support Vector Machines," *Bayesian Analysis*, vol. 6, no. 1, pp. 1-23. 2011.
- [33] J. Salamon and J. P. Bello, "Deep Convolutional Neural Networks and Data Augmentation for Environmental Sound Classification," *IEEE Signal Processing Letters*, vol. 24, no. 3, pp. 279-283, March 2017.
- [34] https://open-power-system-data.org/data-sources#11_European_load_data (Visited on 3rd August 2020)
- [35] <https://pytorch.org/> (Visited on 3rd August 2020)