Convolutional Neural Network based power generation prediction of wave energy converter

Chenhua Ni [1,2], Xiandong Ma [2], Yang Bai[1]

[1] National Ocean Technology Center, Tianjin 300112, China
nichiva@126.com

[2] Engineering Department, Lancaster University, Bailrigg, Lancaster LA1 4YW, UK
xiandong.ma@lancaster.ac.uk

Abstract— **The prediction of power generation from a marine wave energy converter (WEC) has been increasingly recognized, which needs to be efficient and cost-effective. This paper introduces a four-inputs model based approach that uses convolutional neural network (CNN) to predict the electricity generated from a oscillating buoy WEC device. The CNN works essentially by converting values of the multiple variables into images. The study shows that the proposed model based CNN outperforms both multivariate linear regression and conventional artificial network-based approaches. This model-based approach can furthermore detects changes that could be due to the presence of anomalies of the WEC device by comparing output data obtained from operational device with those predicted by the model. The precise prediction can also be used to control the electricity balance among energy conversion, electrical power production and storage.**

Keywords- Wave Energy Converter; Marine Energy; Predication; Artificial Neural Network; Deep Learning; Convolutional Neural Network

I. INTRODUCTION

Globally, the increasing energy demand and greenhouse gas regarding climate change lead to rapid development of the reliable and alternative energy technologies. Wave energy, taking the advantage of renewable and predictable sources of ocean energy, has remained widely untapped. The speed of utilization of wave energy grows faster due to economic and environmental reasons in recent years. Until now, a plenty of wave energy technologies have bloomed based on the mechanism of absorbing energy from the waves, the water depth and the location where the devices are deployed (shoreline, near-shore, offshore) [1]. Despite the high TRL (level eight) reached by some devices, the power prediction, representing one of the crucial operation elements, still needs to be proven [2].

In order to plan the energy usage and shortage, as well as control the grid balance, a responsible and accurate power prediction should play a crucial role to maximize the profit [3]. With this reason, large-scale wave power can be reliably produced, and the variability of wave generation and its impacts on the grid can be managed. Traditionally, wave forecasting can be performed by statistical techniques or physics-based models [4]. Scientists have developed physical models to perform weather forecasting and disaster warning at global and regional level based on the nested grids [5]. This is because the vast observation data and high-level technologies have led to development of the physics-based models. In brief, the wave prediction technology will take great opportunities to advance further with the significant progresses succeeded over the near future in terms of physical parameter forecasting and power output prediction.

Many novel methods, approaches along with algorithms have been developed in the field of statistical models in the wind power and other renewable energy predictions. Artificial Neural Networks (ANNs) by means of abundant data have been developed to predict solar power for a small capacity solar power system [6]. Neural network models and those parametric and non-parametric machine learning algorithms, such as boosting tree, random forest, support vector machine and k-nearest neighbor, have been developed to perform very short-term wind power predictions by using long term wind parameters and generator variables as the model inputs [7]. In terms of the medium and long-term wind power predictions, the neuro-fuzzy inference approach and multilayer perceptrons based derivatives have been widely acceptable [8]. With the development of machine learning as well as deep learning algorithms being emerged, the CNN, long short term memory (LSTM) networks, deep brief net (DBN) and recurrent neural network (RNN) modellings have attracted more attentions for the dedicated result prediction, including electrical power prediction from power plants. A data-based model derived from RNN was successfully applied to predict 6-12 h electricity load of a city grid [8]. A deep multi-layer neural model was in charge of estimating the electrical generation produced by a wind farm by 24 h in advance [9]. A creative hybrid deep-learning network, which combined the empirical wavelet transformation and RNN algorithm, was proven to be completely competent to predict wind speed with a higher accuracy.

This article presents a data-drive model-based CNN to predict the electrical generation of a WEC device for the contribution to accurate prediction and further condition monitoring. For these reasons, this article is organized as follows: the methodology of the multi-input CNN algorithm is described in Section II. Model performance and prediction results are demonstrated in Section III. Finally, Section IV summarizes the conclusions and discussions from this work.
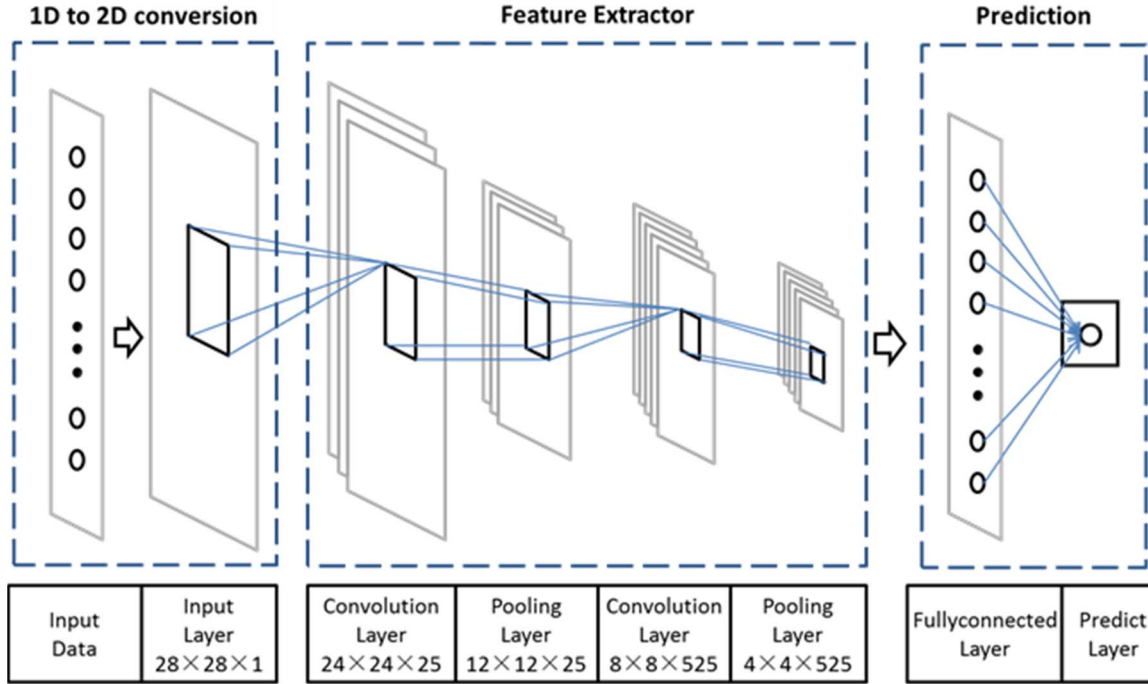
Figure 1.   The fundamental structure of the multi-input CNN model

II.   METHODOLOGY

A.   Convolutional Neural Networks

Traditionally, both the inputs and power output are time-series signals, implying the connections between time intervals might be lost. The CNN algorithm is able to convert time-series input data into array data (image format), thus connecting the relationship between time intervals and among the variables. Instead of the linear maps learned by ANNs, CNN considers a specialized type of ANN featuring with convolutional layers (also referred as to convolutional filters). Convolutional filters are linear functions that are applied to the input data in a sliding-window fashion [10]. Among the state-of-the-art deep neural networks, CNN is adept at image processing because it is able to extract features between layers [11]. In summary, there are plenty of benefits to applying CNNs, including (i) the neurons inside a convolution layer are related to only a small area of the convolution layer before it, known as receptive field, which could reduce plenty of parameters; (ii) the filters are replicated from others during the entire visual field and they share the same weight and bias; (iii) the brilliant usage of pooling layer makes the translation invariant features to be extracted more effectively [12].

B.   Network Architecture

The network structure, comprising from input layers, convolution layers, pooling layers and fully connected layers, is shown in Fig.1. It is worth mentioning that the parameters chosen in this model have been optimized based on the trial and error method after numerous experiments. Table 1 lists hyper-parameters and their values set in this network. The input layer is organized by four series of variable collected from an operating WEC. The 1D to 2D conversion is in charge of rearranging 2D input from four time-series observations. The size of 2D input is set to 28 × 28 pixels because the value of input data from a traditional CNN regularly is adapted as 28 pixels. The convolution operations are conducted at the convolution layer, where a 5×5 kernel size is set by feature maps. The $24 \times 24 \times 25$ indicates the input image size is 24×24 while the number of first convolution layer is 25. After that, an activation function, called rectified linear unit (ReLU), is connected to each convolution layer. The application of a ReLU function makes the convolutions be more effective and is able to reduce likelihood of vanishing gradient compared with the sigmoid function. The average pooling size is set to 2 × 2, which means half number of input samples will be pooled by calculating the average values. Finally, the prediction section formed by a fully connected layer and a predict layer is required. The dimension of the fully connected layer is 40.

A linear function is used as the activation function of the regression predict layer because of the unconstrained values during the regression process.

The least absolute deviations (L1) is applied here to train this network because it can reduce the absolute differences between predictions and practical measurements. The loss function L1 is defined as

$$S = \sum_{j=0}^{n} \left| d_0^{(j)} - d_t^{(j)} \right| \qquad (1)$$

where n represents the number of the dataset. The $d_0^{(j)}$ represents the jth target value and the $d_t^{(j)}$ represents the jth estimated values of this network.

- Convolution Layer

TABLE I.   LIST OF THE VALUES OF HYPER-PARAMETERS IN THIS NETWORK

Hyper Parameters	Values
Input variables	4
CNN Layers	25
Fully Connected Layer	40
Predict Layer	1
Batch size	20
Number of Epochs	100

The convolution layer is a feed-forward neural network and it uses a convolution process to map the local features into global features [13]. As mentioned above, the neurons always kept unconnected while in the own layer and share weight in the different layers in order to simplify the process in the both feed forward and back propagation. Clearly, the feature maps in the previous layer are convolved with the shared weights. The convolution layer is described as follows,

$$a_{i,j} = f\left(\sum_{m=0}^{2}\sum_{n=0}^{2} w_{m,n} x_{i+m,j+n} + w_b\right) \qquad (2)$$

where, $x_{i,j}$ represents a typical pixel in the input image, $w_{m,n}$ represents the weight, m and n mean the row and column number respectively. The w_b represents bias of the filter while $a_{i,j}$ denotes the feature map. In addition, the output activation function f is selected to be the ReLU.

- Pooling Layer

Pooling layer is usually combined with convolution layer when it appears. Classical convolution layer is normally intersected with a pooling layer for reasons not only to build up the high-level invariant structures without sacrifice of the characteristic, but also to minimize the calculation time. The maximum and average pooling methods are normally adopted based on the pooling prospective. We use average pooling in this paper because it is able to show the connection with multi-resolution analysis. Given an input $x = (x_0, x_1, \cdots, x_{n-1}) \in R_n$, average pooling outputs a vector of a fewer components $y = (y_0, y_1, \cdots, y_{m-1}) \in R_m$, as

$$y_j = \frac{1}{p}\sum_{k=0}^{p-1} x_{p,j+k} \qquad (3)$$

where p denotes the number of pooling target and $m = n/p$.

- Fully connected Layer

In a typical CNN architecture, the last hidden layer needs to be fully connected. It is a linear function, which concentrates all targets on a single result that can be seen as the features extracted from the original inputs. More specifically, for the highest order representations, $P_1^h, \ldots, P_{K_h}^h$ (assume $P_k^h \in \mathbb{R}^{d \times p}$), firstly flat them into a vector $P \in \mathbb{R}^{K_h \times d \times p}$, then transform it with a dense matrix $H \in \mathbb{R}^{(K_h \times d \times p) \times n}$ and finally apply the non-linear activation:

$$\hat{x} = \alpha(p^T H) \qquad (4)$$

where $\hat{x} \in \mathbb{R}^n$ denotes the final extracted feature vector. The parameters in matrix H are optimized during training [14].

- Predict Layer

The predict layer usually uses a linear regression layer to perform the final results after obtaining the feature vector $\hat{x}_{ir}$,

$$y_{ir} = [1, \hat{x}^T] \cdot W \qquad (5)$$

The parameter values in vector W can be optimized during training.

- Back Propagation Algorithm

The back propagation (BP) algorithm is a training method to solve the prediction issues, which always train weights and biases with stochastic gradient descent every iterations. Basically, BP is used to minimize the squared-error between the outputs and targets. The loss function E_m is calculated as follows,

$$E_m = \frac{1}{m}\sum_{i=1}^{m}\sum_{j=1}^{d}\left(h_j^i - y_j^i\right)^2 \qquad (6)$$

Thus, the weights W and biases b, β, c should be updated during the training, using the rules defined as follows,

$$W = W - \eta \cdot \partial E_m / \partial W \qquad (7)$$

$$b = b - \eta \cdot \partial E_m / \partial b \qquad (8)$$

$$\beta = \beta - \eta \cdot \partial E_m / \partial \beta \qquad (9)$$

$$c = c - \eta \cdot \partial E_m / \partial c \qquad (10)$$

here, $\partial E_m / \partial W$, $\partial E_m / \partial b$, $\partial E_m / \partial \beta$ and $\partial E_m / \partial c$ are the partial derivatives of loss function with regards to perturbations of W, b, β and c.

C. Model Performance Metrics

Three performance metrics are selected to assess forecasting accuracy. The root mean square error (RMSE) is used here because it is more sensitive to large difference between forecasts and targets. The mean absolute error (MAE), on the other hand, is the absolute value of the difference between the forecasted value and the actual value. The RMSE and MAE can be calculated by eqs. (11) and (12) for signals of N samples.

$$\text{RMSE} = \frac{1}{\sqrt{N}}\sqrt{\sum_{i=1}^{N}\left(I_{(pred,i)} - I_{meas,i}\right)^2} \qquad (11)$$

$$\text{MAE} = \frac{1}{N}\sum_{I=1}^{N}\left|I_{(pred,i)} - I_{meas,i}\right| \qquad (12)$$

The coefficient of determination (R^2) is organized here as a reference to measure how the model structure is adapted optimally during training, and is defined,

$$R_T^2 = 1 - \frac{\sigma_e^2}{\sigma_y^2} \qquad (13)$$

where σ_e^2 denotes the sample variance of the differences, which means the residuals between the predicted and the actual outputs; and σ_y^2 denotes the variance of the actual output. The R_T^2 tells us how close between the predicted values and actual outputs. A unity value elaborates a perfect modelling fit while a value tending to be zero shows a poor prediction [15].

III. RESULTS AND DISCUSSIONS

A. Data acquisition

The principle of this WEC contains a dual-buoy hydraulic oscillating body device with ten kW level capacity, which experienced open sea condition in nearshore area. The operation period continued from February to April 2017. The WEC power system is comprised by three main parts: a power capture system that is hidden inside the buoy, a hydraulic motor and a generator accommodated in the power conversion system; and a power transmission system, as shown in Fig. 2. The buoy uses up-and-down oscillations to capture wave energy and convert the kinetic energy into electrical power by the generator. Finally, the electricity is transferred from ocean to seashore through undersea cables. There are two conversions through the entire process; the buoy firstly captures the wave energy in the primary conversion while a hydraulic pressure system is then deployed to generate hydraulic power in the second conversion. The power take-off system includes a hydraulic ram, which is installed inside the oscillating buoy. The significant readings associated with power output are hydraulic pressure and flow, motor speed and motor torque that will be used in this study. Examples of the acquired data from the hydraulic system of this WEC are illustrated in Fig.3, where the four internal parameters are shown.

Data pre-processing is necessary to eliminate those abnormal data recorded and make the prediction more accurate when the WEC is in inactive or malfunction conditions. There are gaps between data because the harsh sea condition makes the generator working under inactive situation. Plenty of abnormal values also exist within the data caused due to signal disturbance and other abnormal conditions.

For a CNN modelling, the datasets used are usually divided into training set, validation set and test set. Training set is used for model training, which is to fit the parameters of the classifier. Validation set is a set of samples applied to unify and verify the parameters. Test
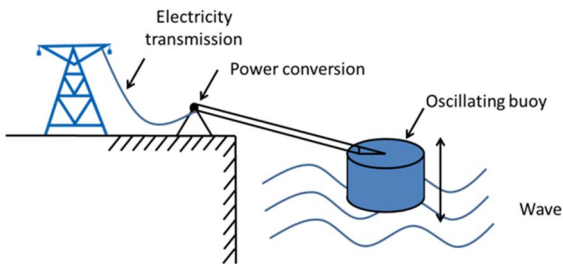

Figure 2. The schematic diagram of the wave power generation system
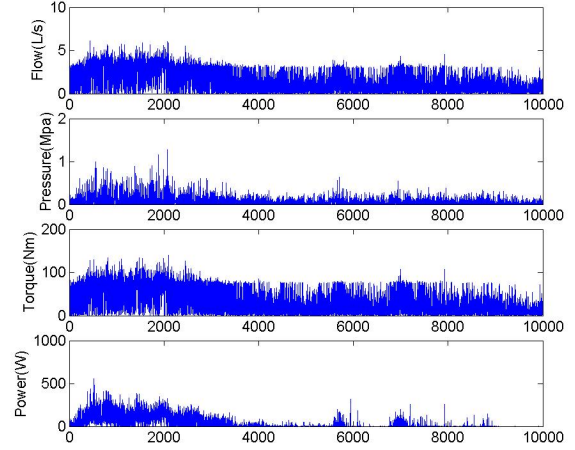

Figure 3. The example of raw data acquired from the WEC

set is a set of samples to estimate the performance of a fully specified classifier. Here, 100,352 samples in total were applied to the CNN network proposed above, dated from February 2017 for three months. These samples were divided sequentially into three sets in terms of the percentage, i.e., training dataset (80%), validation dataset (5%) and test dataset (15%).

B. Results

The four time series input variables are firstly converted to image format before CNN training based on four different regulations. There are four possible different conversions to perform different rearrangements in order to improve accuracies. The 1st and 2nd conversion methods are to rearrange inputs sequentially while the 3rd and 4th methods form the images by sequentially arranging sub-images and by mixing sub-images of the input variables respectively. Furthermore, different image sizes (28×28 pixels, 20×20 pixels, 14×14 pixels, 10×10 pixels) have been tried to investigate which size could produce the best results. The curve-fitting tools are used for each conversion method to reveal the fitting details. Furthermore, other representative supervised modelling approaches, including ANN, support vector machine and robust linear regression, were also trained in order to compare the CNN performance with the same datasets. Finally, the model performance metrics introduced above are used as the metrics for comparison to evaluate the prediction performance.

Table II shows the results from different conversion methods, image sizes and modelling algorithms. It can be seen from RMSE and MAE that the 3rd and 4th methods perform much better than the 1st and 2nd methods with lower residual values, which means they predict values in a higher accuracy. In terms of all metrics, a bigger image size demonstrates a better fit than the small size inputs while other parameters remain constant. Clearly, the 4th conversion method incorporating with 28×28 image size outperforms all other combinations, with a R2 value of 0.96 being achieved. The result also reveals reason why the large images are able to contain more information than the medium and small ones. In addition, the 3rd conversion method also achieves lower RMSE and MAE values with a higher R2 value at a comparable level with the 4th method.

TABLE II. THE PREDICTION PERFORMANCE OF THE CNN MODEL

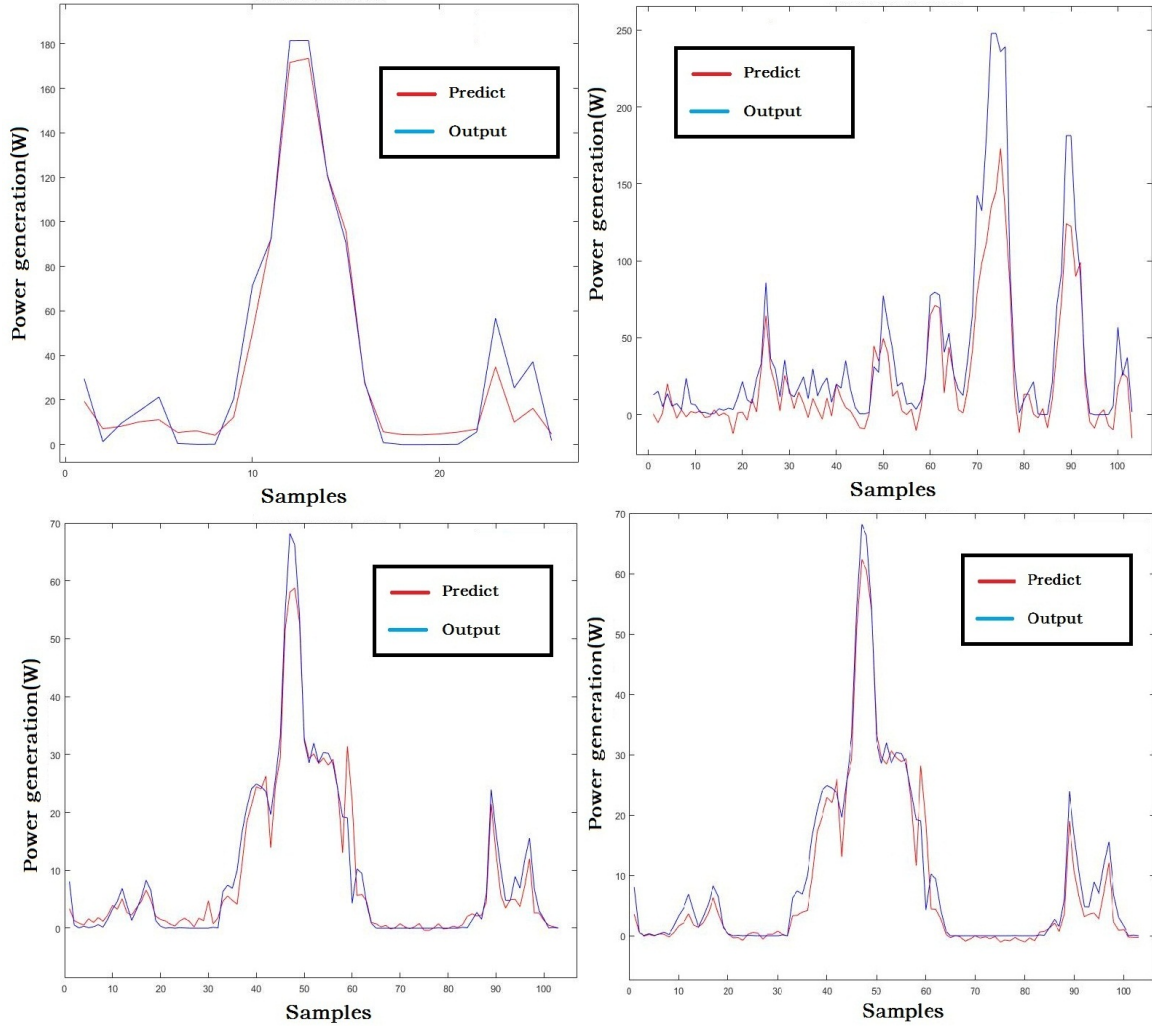| Image Size	1st Conversion Method			2nd Conversion Method			3rd Conversion Method			4th Conversion Method		
	RMSE	MAE	R^2	RMSE	MAE	R^2	RMSE	MAE	R^2	RMSE	MAE	R^2
28×28	10.02	8.05	0.95	23.48	21.19	0.85	3.37	2.23	0.94	3.11	1.92	0.96
20×20	16.43	8.64	0.91	29.46	23.67	0.77	3.63	1.84	0.94	3.76	2.14	0.93
12×12	20.48	9.63	0.87	25.21	19.82	0.83	4.45	2.81	0.91	4.25	2.45	0.92

Figure 4. Prediction performance comparison from different conversion methods based on 28×28 dimension image uisng conversion methods from 1st to 4th from top left to bottom right

Fig. 4 shows the prediction results in time series as the examples. As can be seen, the predicting curves fit the real outputs well in all the four plots except the top right subplot. The top left subplot shows that the two curves match nearly perfectly during the high power output period than the low power output level. Both bottom subplots perform considerably good fitting results, demonstrating their capabilities of forecasting such distinct fluctuations during power generation from the WEC.

We also applied different supervised modelling approaches to the CNN model with an image size of (28×28) and the results are shown in Table III. It proved that the CNN can make remarkable achievements for electrical power prediction from the WEC among these machine learning algorithms. The CNN model demonstrates a best fitting performance between the actual outputs and the predicted ones. The SVM and RLR show the poorest performance, as evidenced by their higher MAE value (more than twice than others). The R^2 value of the ANN, MT and BT indicate these models yield fair fitting results. However, in terms of the training time consumption, the ANN and CNN would take a little bit longer time than other networks, with almost 43 s in this case. By analyzing further the time consumption, the number of the selected hidden layers and epochs could contribute significantly to the time consumed by the networks.

TABLE III.   THE PERFORMANCE COMPARISON OF THE CNN WITH DIFFERENT SUPERVISED MODELLING APPROACHES

Networks	RMSE	MAE	R^2	Time(s)
Artificial neural network	2144.83	11.38	0.83	39.19
Support vector machine	34.88	27.10	0.69	583
Robust linear regression	35.15	27.30	0.69	4.68
Medium tree	23.36	12.92	0.86	7.21
Boosted tree	20.83	12.49	0.89	11.26
CNN	3.11	1.92	0.96	42.85

C. Discussions

In this case, a Dell work station with 3.6 GHz CPU and 16 GB RAM was implemented to perform the network. The table III shows the time consumption to train different model with different networks. Obviously the SVM takes the longest time among them with more than 10 times higher than the CNN network. The time to train CNN will take almost 43 s while using the parameters given in Table I. To train MT and BT will take 7.21 s and 11.26 s. The RLR and MT take the shortest time in the experiment, both under 10 seconds. The time to train ANN seems much close to CNN due to the similar architectures. Further work is currently carried out in order to reduce the training time and boost the effectiveness of the model.

It has been well known that CNN take the advantage of image process, which makes us of 2D matrix data [16]. One of the features to train deep learning network is that its performance widely depends on feature extraction. The more features extracted from the training dataset, the better performance obtained from the network. The use of image patterns from different data conversion methods to train the CNN models reveals that different distinctive features can be learned from different image patterns by the model. Results have shown that the network can catch more information from large size of input than small size ones. Moreover, the prediction performance is also affected by the layout of the input image and connections between these input variables.

Furthermore, the residual signals can be used as a trigger for fault diagnose of a WEC when the residual values exceed a given threshold. When an appropriate threshold is confirmed, a function is created to determine the value between residual and threshold, then a potential fault will be alarmed and precaution can be made in advance.

IV.   CONCLUSIONS

In this article, a deep learning CNN network was developed to make the accurate electrical power prediction of a WEC. The CNN network takes great advantages of image process. One of the significant contributions of this paper is to develop the methods to convert time series data into image data.

The results show that the multi-input CNN performs much improved prediction outcomes compared with other supervised modelling networks and the highest R^2 value achieved 0.96 for this case. The results also show that the most important factors can affect the accuracy of results are image size and conversion methods. Evidently, a larger image size can extract more distinctive features from input images. It can also be seen from the network architecture, the CNN is deep enough to perform such a prediction problem. The connection of data between inputs and outputs can affect the architecture of network. The more complex the input data are, the deeper the network architecture can be.

Furthermore, the model can be developed as an early warning system when the threshold is given and the residual is exceeded. This can significantly contribute to the electrical power generation management, power transformation and energy storage organization of the renewable energy systems.

REFERENCES

[1] S. Zou, O. Abdelkhalik, R. Robinett, G. Bacelli, D. Wilson. Optimal control of wave energy converters. Renewable Energy 103 (2017) 217-225

[2] D. Magagna, R. Monfardini, A. Uihlein. JRC Ocean Energy Status Report 2016 Edition. Luxembourg: Publications Office of the European Union, 2016

[3] F. Kara. Time domain prediction of power absorption from ocean waves with wave energy converter arrays. Renewable Energy 92 (2016) 30-46

[4] G. Reikard,P. Pinson, J-R Bidlot. Forecasting ocean wave energy: the ECMWF wave model and time series methods. Ocean Eng 2011; 38:1089–99. http://dx.doi.org/10.1016/j.oceaneng.2011.04.009.

[5] A. Uihlein, D. Magagna. Wave and tidal current energy- A review of the current state of research beyond technology. Renewable and Sustainable Energy Reviews 58 (2016)1070–1081

[6] E. Izgi, A.Oztopal, B. Yerli. Short–mid-term solar power prediction by using artificial neural networks. Solar Energy 86 (2012) 725–733

[7] L. Vargas, G. Paredes, G. Bustos. Data mining techniques for very short term prediction of wind power. IREP Symposium-Bulk Power System Dynamics and Control-VIII; 2010:1-7.

[8] H. Shi, M. Xu, Q. Ma, C.Zhang. A Whole System Assessment of Novel Deep Learning Approach on Short-Term Load Forecasting. 9th International Conference on Applied Energy, ICAE2017, 21-24 August 2017, Cardiff, UK

[9] J. M. Torres, R. M. Aguilar, K. V. Zu niga-Meneses. Deep learning to predict the generation of a wind farm. JOURNAL OF RENEWABLE AND SUSTAINABLE ENERGY 10, 013305 (2018)

[10] A.Mozo, B.Ordozgoiti, S.GoÂmez-Canaval. Forecasting short-term data center network traffic load with convolutional neural networks. https://doi.org/10.1371/journal.pone.0191939. Feb, 2018

[11] T.Shimobaba, T.Kakue, T.Ito. Convolutional neural network-based regression for depth prediction in digital holography. Feb, 2018. https://arxiv.org/abs/1802.00664 .

[12] D.Suryani, P.Doetsch, H.Ney. On the Benefits of Convolutional Neural Network Combinations in Offline Handwriting Recognition. 2016. https://www-i6.informatik.rwth-aachen.de/publications/download/1021/Suryani-ICFHR-2016.pdf

[13] H.Wang, G.Li, G.Wang. Deep learning based ensemble approach for probabilistic wind power forecasting. Applied Energy 188 (2017) 56–70

[14] K.Zhao, C.Wang. Sales Forecast in E-commerce using Convolutional Neural Network. https://arxiv.org/abs/1708.07946

[15] P.Cross, X.Ma. Nonlinear system identification for model-based condition monitoring of wind turbines. Renewable Energy 71 (2014) 166-175

[16] A. Zhu, X. Li, Z. Mo, H.Wu. Wind Power Prediction Based On a Convolutional Neural Network. 2017 International Conference on Circuits, Devices and Systems.