Energy 304 (2024) 132122

Contents lists available at ScienceDirect

Energy

journal homepage: www.elsevier.com/locate/energy

$\mathcal{M}$*eta* $\mathcal{W}$ave $\mathcal{L}$earner: Predicting wave farms power output using effective meta-learner deep gradient boosting model: A case study from Australian coasts

Mehdi Neshat [a,e,*], Nataliia Y. Sergiienko [b], Ashkan Rafiee [c,d], Seyedali Mirjalili [e,f], Amir H. Gandomi [a,f], John Boland [g]

[a] *Faculty of Engineering and Information Technology, University of Technology Sydney, Ultimo, Sydney, 2007, NSW, Australia*
[b] *School of Electrical and Mechanical Engineering, University of Adelaide, Australia*
[c] *Technology Department, INPEX, Perth, Australia*
[d] *Causal Dynamics Pty Ltd, Perth, Australia*
[e] *Center for Artificial Intelligence Research and optimisation, Torrens University Australia, Brisbane, QLD 4006, Australia*
[f] *University Research and Innovation Center (EKIK), Obuda University, Budapest, 1034, Hungary*
[g] *Industrial AI Research Centre, UniSA STEM, University of South Australia, Mawson Lakes, 5095, Australia*

ARTICLE INFO

Keywords:
Renewable energy
Wave energy
Power output prediction
Deep ensemble learning method
Extreme gradient boosting
Transfer learning

ABSTRACT

Precise prediction of wave energy is indispensable and holds immense promise as ocean waves have a power capacity of 30–40 kW/m along the coast. Utilising this energy source does not generate harmful emissions, making it a superior substitute for fossil fuel-based energy. The computational expense associated with simulating and computing intricate hydrodynamic interactions in wave farms restricts optimisation methods to a few thousand evaluations and makes a challenging situation for training in deep neural prediction models. To address this issue, we propose a new solution: a Meta-learner gradient boosting method that employs four multi-layer convolutional dense neural network surrogate models combined with an optimised extreme gradient boosting. In order to train and validate the predictive model, we used four wave farm datasets, including the absorbed power outputs and 2D coordinates of wave energy converters (WECs) located along the southern coast of Australia, Adelaide, Sydney, Perth and Tasmania. Furthermore, the capability of the transfer learning strategy is evaluated. The WECs used in this study are of the fully submerged three-tether converter type, similar to the CETO prototype. The effectiveness of the proposed approach is assessed by comparing it with 15 well-established and effective machine learning (ML) methods. The experimental findings indicate that the proposed model is competitive with other ML and deep learning approaches, exhibiting considerable accuracy of 88.8%, 90.0%, 90.3%, and 84.4% in Adelaide, Perth, Sydney and Tasmania and improved robustness in predicting wave farm power output.

1. Introduction

Ocean wave energy is considered one of the most efficient renewable energy sources. It has a high potential for large-scale clean energy production, sometimes by orders of magnitude larger than other sources of renewable energy like solar and wind [1]. The swift progress of wave power technology hinges on the precise prediction of power generation to ensure an unwavering and dependable power supply to the grid. This entails forecasting variables such as dynamic environmental conditions, wave parameters [2], power take-off [3], and the positioning of WECs [4] in wave farms. For example, accurate predictions of ocean wave characteristics can estimate the power produced by WECs while enhancing these converters' performance [2].

Recently commissioned generators for WECs have shown considerable promise, with a potential power output of up to 1 MW per converter [5]. However, to be commercially viable, wave farms must incorporate multiple converters. Determining the optimal placement of WECs in a wave farm is a complex challenge, as there is no straightforward pattern for their arrangement in real wave scenarios. On the

* Corresponding author at: Faculty of Engineering and Information Technology, University of Technology Sydney, Ultimo, Sydney, 2007, NSW, Australia.
E-mail addresses: mehdi.neshat@torrens.edu.au, mehdi.neshat@uts.edu.au (M. Neshat), nataliia.sergiienko@adelaide.edu.au (N.Y. Sergiienko), ashkan.rafiee@inpex.com.au, ashkan.rafiee@causaldynamics.tech (A. Rafiee), ali.mirjalili@torrens.edu.au (S. Mirjalili), gandomi@uts.edu.au (A.H. Gandomi), John.Boland@unisa.edu.au (J. Boland).

https://doi.org/10.1016/j.energy.2024.132122
Received 1 February 2024; Received in revised form 24 May 2024; Accepted 16 June 2024
Available online 21 June 2024

other hand, when the number of converters increases, optimising the placement of buoys becomes even more difficult due to the complex hydrodynamic interactions between the WECs. These interactions are dependent on the position of the WECs, and the prevailing wave regime can either enhance or diminish the average absorbed power output. The process of modelling these nonlinear interactions for a moderately or large-sized wave farm layout can take anywhere from several minutes to an hour. Additionally, optimising farm layouts presents a multi-modal problem that typically requires a large number of model evaluations to explore the objective space properly. There is potential to expedite this search process by using a reliable and precise prediction model. However, the primary challenge is to train such a model quickly enough to enable a reduction in overall optimisation runtime.

The prediction of wave power generated from WECs has gained attention due to technological advancements. Two approaches, physical and data-driven models [6], have been used for wave power forecasting. Physical-based methods employ numerical models to predict wave parameters using geographical and meteorological data. These numerical wave models (NWMs) are crucial for offshore engineering, providing accurate forecasts and hindcasts. While effective for large-scale areas, NWMs are computationally expensive and may not be suitable for real-time forecasting due to high computational costs [7]. Machine learning (ML) methods have the capability to make accurate predictions about wave power in the short and long term, which can range from minutes to a day in advance. These predictions are made by utilising a combination of previous power generation, wave data and weather variables. In the field of ML models, linear approaches such as auto-regression (AR) and autoregressive moving averages (ARMA) are widely utilised [8] and have consistently shown improvements in terms of reliability and efficiency [9]. However, linear models cannot accurately capture complex and nonlinear relationships in wave power data, resulting in lower forecasting accuracy. In order to tackle the challenges of regression and classification tasks, various data-driven modelling (DDM) approaches have been developed. In the last decade, deep-learning (DL) networks derived from artificial intelligence (AI) have gained immense popularity. This can be attributed to the increased capability of computing power and the availability of vast amounts of data. These DL networks have proven their prowess in a multitude of domains, especially renewable energy forecasting. Their superiority lies in their advanced adaptabilities, computational capacities and recognition abilities that have been demonstrated in recent studies, especially in renewable energy forecasting [10].

One of the most significant applications of DL models in wave energy domains is to predict the optimal power take-off (PTO) parameters in maximising the total absorbed power [11]. However, converting wave energy to electricity requires significant effort in developing efficient PTO control algorithms. Traditional model-based control algorithms use reduced-order models that neglect other subsystem dynamics and can lead to misleading results in practice. Deriving a model-based control for a highly nonlinear/complex system like a wave-to-wire model is challenging. One of the initial studies in applying the deep learning model in [12] Recently, advanced machine and deep learning models have been considered to address various challenges in predicting wave energy converters' characteristics. For instance, developing an efficient PTO control algorithm is complicated due to a highly nonlinear and complex relationship between PTO parameters, wave features and power output. To address this issue, Zou et at [13] proposed a Deep Reinforcement Learning (DRL) control that optimises the performance of Wave Energy Converters (WECs) from wave to wire. The proposed DRL control outperforms conventional model-based controls in numerical simulations, achieving up to 152% advancement in energy production and 84% in power quality.

Time series vision is one of the most popular insights into wave energy forecasting. Therefore, a wide range of sequential machine learning methods has been modelled to predict the absorbed power of WECs. An integrated long short-term memory (LSTM) with the principal component analysis (PCA) proposed [14] to predict the electrical power generation from a WEC, and the findings indicated a remarkable performance compared with the LSTM alone. In another recent study on eastern Australia coastal zones [2], to forecast peak wave energy periods, the application of an extreme learning machine (ELM) was suggested and was demonstrated that ELM is able to beat recurrent neural network (RNN), convolutional neural network (CNN) and Conditional Maximisation with Multiple Linear Regression (MLR-ECM). However, the impact of hyper-parameters tuning was ignored and not discussed. Considering decomposition techniques combined with a CNN featuring bi-directional long short-term memory (Bi-LSTM) [15] presented a competitive improvement (at 13%) in forecasting wave power. In order to predict significant wave height at the North Sea [16], a nested adaptive neuro-fuzzy inference system (ANFIS) integrated with a particle swarm optimisation (PSO) algorithm was proposed and compared with other classical ML models. The ANFIS delivered the most accurate prediction of wave heights. However, the performance of a time-series deep learning model was not compared with the ANFIS predictor. In another successful example of DL models, a deep learning-based model [17] was proposed for modelling the power of hinged-raft WEC and compared with various ML models, including the LM, MLP, CNN, LSTM and the GRU; the validation study showed that the deep model could capture the WEC dynamics more accurately than others.

With regard to the applications of decomposition techniques used for improving the performance of wave power predictor, Ni et al. [18] introduced a hybrid deep learning model which effectively combines the empirical wavelet transform (EWT) technique with a robust CNN for wave power prediction. The employment of EWT allows for the decomposition of wave power observations into sub-bands, each possessing distinct frequencies, enabling a detailed analysis of the underlying wave characteristics. On the other hand, the CNN architecture excels at extracting spatial features from the multi-dimensional grid data, thereby capturing the intricate relationships and patterns within the wave power dataset. By merging these two powerful techniques, the hybrid EWT-CNN model offers a comprehensive and holistic approach to predicting wave power output accurately. The results obtained from this comparative analysis reveal the superiority of the hybrid model in accurately predicting short-term wave power output, leveraging both time and space domain information. The incorporation of the EWT technique and CNN architecture enables the hybrid model to effectively capture the complex dynamics and temporal variations present in the wave power dataset, resulting in enhanced predictive capabilities. Despite the benefits of integrating decomposition techniques into the forecasting model, there is a noticeable augmentation in the overall complexity of the hybrid model [19]. Consequently, this augmentation poses a formidable challenge regarding training and optimising the combined model, especially for larger datasets and extended training durations. Furthermore, the amplified complexity of the model may also give rise to the problem of overfitting or hinder the generalisation of the model to novel data.

This article suggests a new solution to overcome the aforementioned obstacles by introducing a Meta ensemble extreme gradient boosting model. A meta-ensemble model is a sophisticated hybrid machine-learning technique that aims to optimise the combination of predictions generated by multiple base models. This model predicts the overall absorbed power of WECs in comprehensive layouts of the four real wave scenarios located along the southern coast of Australia: Adelaide, Perth, Sydney and Tasmania. The primary contributions of this research are as follows:

- Develop a technical comparative framework for predicting the wave farms' power output using 15 well-known machine-learning methods.

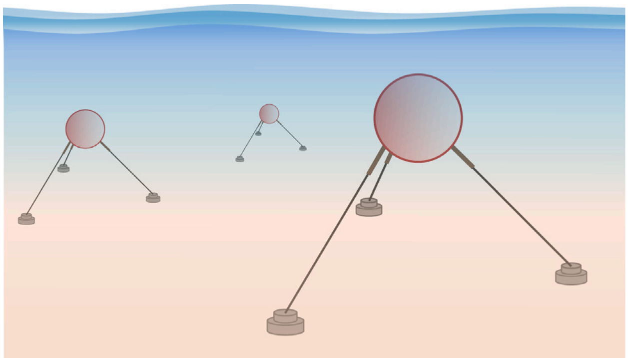


Fig. 1. Artistic impression of a wave farm.

- Propose a new wave farm power output predictor by combining an optimal number of base learners based on a comparative performance between DNN and CDNN with a Meta-learner (best performed of Extreme Gradient Boosting) to reinforce the prediction accuracy.
- Optimise the architecture and hyper-parameters of the proposed Meta ensemble learning model using a grid search technique.
- Develop a transfer learning approach to assess the generalisation ability of the trained models using one wave farm and testing it by other wave farm datasets.
- Evaluate the performance of the hybrid deep predictive model using a comprehensive dataset based on a four real wave scenario.

We show that the hybrid model described it outperforms the other 15 prediction models regarding the accuracy and learning error.

In the following, we commence by providing an initial exposition on the description and modelling of the WECs system, including not only the fundamental equations of motion that govern its behaviour but also a comprehensive explanation of the chosen deployment site and the performance measures that will be employed to evaluate its effectiveness (as expounded upon in Section 2). Following this, we introduce a multitude of ML and deep learning algorithms, as well as the intricate technical details of the proposed Meta ensemble deep learning method, as outlined in Section 3. Subsequently, we meticulously outline the numerical results obtained through the application of the method above and provide an in-depth discussion of the findings in Section 4, thereby facilitating a comprehensive comparison of the efficiency and effectiveness of the proposed approach. Finally, we conclude the manuscript by summarising the main findings and expounding upon the advantages our proposed method (See Section 5) offers over existing approaches.

2. System description and modelling

The submerged sphere as an absorber of wave power was proposed in [20] and is used in this study. The wave farm consists of multiple WECs that operate relatively close to each other. The radius of the sphere is set to 5 m, and the distance between the buoy's centre of mass and the still water level is set to 8.5 m. The sphere is attached to the sea floor by three mooring lines. The arrangement of the mooring system is symmetric to ensure that the power absorption of an isolated WEC is independent of the wave direction. The mooring lines inclined at 54 deg to the vertical lead to the maximum power absorption of the submerged sphere, as shown in [21]. The buoys should be positively buoyant to provide the tension in each mooring line, therefore, the mass of the buoy is set to be half the mass of the displaced volume. The wave farm consists of multiple WECs that operate relatively close to each other as shown in Fig. 1.

2.1. Equations of motion

The WEC hydrodynamic loading is modelled using linear potential flow theory [22]. The behaviour of the power take-off machinery of each WEC is simplified to the linear spring–damper system with tunable stiffness and damping control parameters. The WEC farm dynamics is formulated in the frequency domain also taking into account the wave directionality. Each WEC moves in the surge, sway, and heave modes, while the rotational motion of the spherical WECs is neglected. The equations of motion are written as:

$$\mathbf{x}(\omega,\beta) = \left(-\omega^2(\mathbf{M}+\mathbf{A}(\omega)) + \mathrm{i}\omega\left(\mathbf{B}(\omega)+\mathbf{B}_{pto}\right)+\mathbf{K}_{pto}\right)^{-1}\mathbf{F}_e(\omega,\beta), \tag{1}$$

where ω is the regular wave frequency, β is the wave angle, $\mathbf{x}$ is the vector of complex amplitudes with dimensions of $[N\times 3,1]$, where N is the number of WECs in a farm, $\mathbf{M}$ is the diagonal mass matrix, $\mathbf{A}$ and $\mathbf{B}$ are the matrices of added mass and radiation damping coefficients that take into account the hydrodynamic interaction between WECs in the farm, $\mathbf{K}_{pto}$ and $\mathbf{B}_{pto}$ are the power take-off stiffness and damping matrices that model the PTO action, and $\mathbf{F}_e$ is the wave excitation vector.

The average power absorbed by all WECs in a farm as a function of the regular wave frequency and wave direction is evaluated as:

$$\overline{P}(\omega,\beta) = \frac{\omega^2}{2}\mathbf{x}^{\mathrm{T}}(\omega,\beta)\mathbf{B}_{pto}\mathbf{x}(\omega,\beta), \tag{2}$$

2.2. Deployment sites and wave climates

The performance of the wave farm is assessed for the four sea sites in Australia. The hindcast data for these sites is shown in Fig. 2. The data is obtained from the Australian Wave Energy Atlas.

2.3. Performance measures

Based on frequency domain data, it is possible to evaluate how much power a wave farm can potentially absorb in the irregular wave characterised by the significant wave height H_s, peak wave period T_p and wave angle β:

$$P_i(H_s,T_p,\beta) = 2\int_0^\infty S_i(\omega)\overline{P}(\omega,\beta)\mathrm{d}\omega \tag{3}$$

where S_i is the wave spectrum (Bretschneider in this study).

Once the wave farm power output for each ith sea state is evaluated, the average annual power production can be estimated taking into account the probability of occurrence $O_i(H_s,T_p,\beta)$ of each wave condition:

$$P_{AAP} = \sum_i^{N_s} P_i(H_s,T_p,\beta)\cdot O_i(H_s,T_p,\beta) \tag{4}$$

3. Methods and materials

3.1. Wave energy converters dataset

The Wave Energy Converters dataset developed by Neshat et al. and published in UCI Machine Learning Repository [23] used in this study includes measurements of the position and power output of 16 WECs located in a wave farm based on the real wave scenario at the Adelaide sea site. The dataset was obtained from simulations of a WEC system developed by the wave energy team at the University of Adelaide. The WEC model was simulated based on a fully submerged three-tether converter called CETO [24] which has been designed and developed by Carnegie Clean Energy, Australia. In the provided dataset, the Wave Energy Converters (WECs) are positioned in a specific area with x-positions defined as $X = [x_1,\ldots,x_N]$ and corresponding y-positions denoted as $Y = [y_1,\ldots,y_N]$. The maximum number of WECs in the dataset is predefined to be N = 16. Each buoy, represented by index i, is characterised by its position coordinate $[x_i, y_i]$. Thus, a 16-buoy

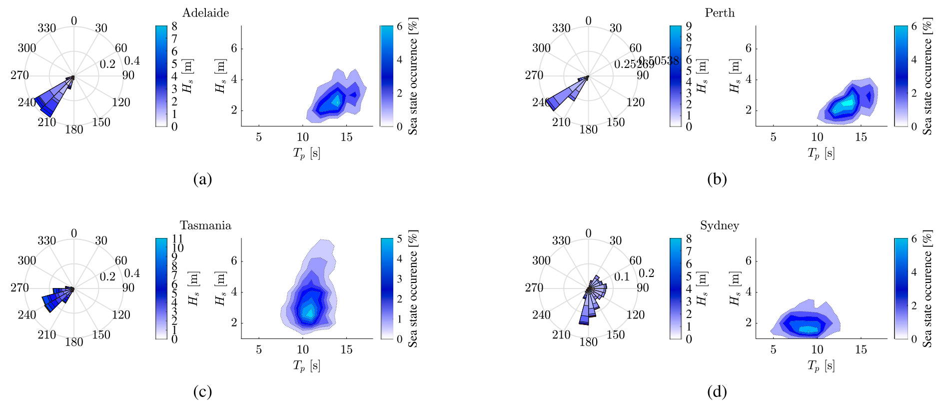


Fig. 2. Wave scatter diagrams and the directional wave roses of four sites: (a) Adelaide, (b) Perth, (c) Tasmania, and (d) Sydney.

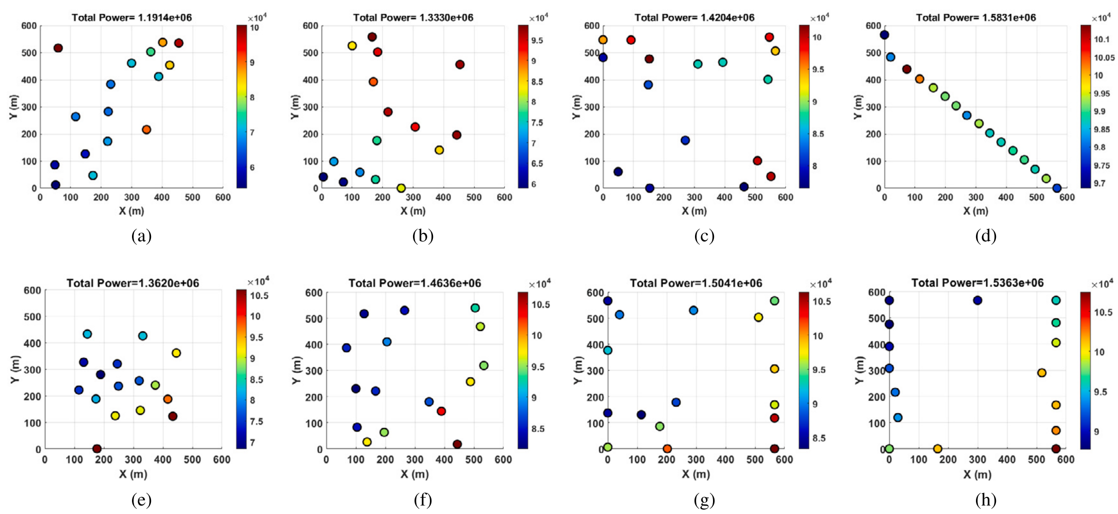


Fig. 3. Some array examples of WECs from the dataset based on the Adelaide (first row) and Sydney (second row) wave scenario. a,e) an array with lowest absorbed power output. (b), (c), (f), and (g) are four random arrays selected from the dataset. (d) and (h) an array with the highest produced total power output. Circles represent the WECs and are highlighted based on the power output of each converter.

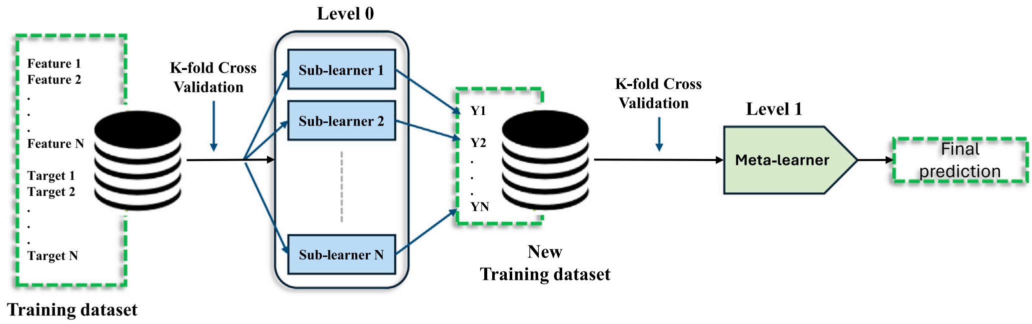
array can be represented as $[(x_1, y_1), (x_2, y_2), \dots, (x_{16}, y_{16})]$. The positions of the WECs are not arbitrary but are confined within a defined area denoted as $\mathcal{A}$. The dimensions of this area are determined by $\mathcal{A} = L{\times}W$, where both the length (L) and width (W) are equal to $\mathcal{A} = L{\times}W$ and $L = W = \sqrt{N * 20000}$ m. Thus, for a configuration with 16 WECs, the length and width of the area are set to $L = W = 566$ m. Providing further technical details of the WECs in this dataset, each buoy has a radius of 5 m, operates at a water depth of 50 m, and has a submergence depth to the buoy centre of 8 m. The buoy itself has a mass of $m = 376 \times 10^3$ kilograms. The Power Take-Off (PTO) system associated with the buoy is characterised by a stiffness coefficient, Kpto, of 2.7×10^5 N/m and a damping coefficient, Bpto, of 1.3×10^5 N/m.

The dataset contains 49 features, including 16 WECs' coordination values based on the X-axis of the farm map. In the following, we can see 16 coordination values of the WECs according to the Y-axis of the farm map. Next, the power output of 16 WECs is listed, and finally, the total power (Watt) out of the wave farm is reported. The dataset includes a total of 72,000 instances. The main objective of the dataset is to predict the total power output of the 16 WECs based on the position of the floats. This dataset can be used for time-series analysis and modelling, as well as for developing machine learning models for regression or forecasting tasks. Fig. 3 shows four samples of 16-WEC arrays of the dataset used with various arrangements and power outputs.

3.2. Meta ensemble model

Meta ensemble model is a type of stacking learning technique that aims to create a diverse set of models by using different kinds of models for training and combining their predictions [25,26]. Stacking involves training a meta-learner to combine the predictions of several base models, which are referred to as first-level learners. The meta-learner is the second-level learner, and it is used to create an ensemble of models that can make more accurate predictions. The first-level learners are also known as level-0 models, while the model that combines their predictions is referred to as a level-1 model. The standard approach involves a two-level hierarchy of models. Still, more layers can be added to the ensemble, such as using multiple level-1 models and a single level-2 model to combine their predictions. By creating a diverse set of models and combining their predictions, stacking can improve the accuracy and robustness of machine learning models.

The motivation for stacking comes from the fact that different base models may have different strengths and weaknesses, and by



(a)

Fig. 4. The framework of the meta ensemble learning method.

combining them, we can leverage their strengths and compensate for their shortcomings, resulting in better performance than any individual model [27]. Another motivation for stacking is to reduce the risk of overfitting. By using multiple base models trained on different subsets of the data or with other algorithms, we can reduce the chances of overfitting the training data. Stacking combines the predictions of these base models, which can help generalise new, unseen data better. Stacking also allows for more flexibility in model selection. Instead of being limited to a single model, we can choose multiple base models specialised for diverse aspects of the problem [28]. For example, we could use a decision tree model to capture non-linear relationships in the data, a linear regression model to capture linear relationships, and a neural network model to capture complex interactions between features. By combining the predictions of these models, we can obtain a more comprehensive and accurate model [29]. The detailed sections of the meta ensemble learning model can be seen in Fig. 4.

3.3. Convolutional deep learning models

Convolutional deep learning models (CNNs) are a type of multi-layer feed-forward artificial neural network particularly well-suited for image recognition tasks, extracting hidden spatial features [30] and multi-dimensional signal modelling. They are based on the concept of convolution, which involves sliding a small filter over an input image and computing the dot product of the filter and the portion of the image it is currently covering. This produces a feature map, representing where the filter detected certain features in the image. The network then applies several layers of these convolutions, followed by pooling layers, which downsample the feature maps, and then fully connected layers, which produce the final classification output. When compared to other neural networks like the deep belief network (DBN) [31], a CNN stands out due to its sparse connectivity and weight-sharing properties. These characteristics significantly decrease the number of parameters a CNN has to learn. In this research, a CNN is employed to discover the potential spatial connection between the coordination elements of a WECs and those of its neighbouring area, with the aim of minimising power prediction errors. The primary CNN architecture includes three sections: convolutional, pooling, and fully connected layers. The CNN computation can be represented as follows [30]:

$$f_{mn}^{uv} = relu(\sum_{l}\sum_{h=0}^{h'-1}\sum_{w=0}^{w'-1} w_{mnl}^{hw}.map_{(m-1)l}^{(u+h)(v+w)} + b_{mn}) \tag{5}$$

The equation uses variables to reference specific elements of the convolutional neural network. Specifically, u and v correspond to the row and column indices of the feature map, while h and w correspond to the row and column indices of the convolution filter. h and w represent the number of rows and columns in the filter, respectively.

3.4. Extreme gradient boosting model

The Extreme Gradient Boosting (XGB) [32] technique, frequently employed in predictive applications, utilises an aggregation of Decision Trees (DT) to create a robust regression model. This large-scale ML approach is designed to leverage multi-threaded parallelism, automatically reducing computation time. Unlike gradient-boosted decision tree (GBDT) models, XGB relies on the second-order Taylor expansion for the loss function. Additionally, the regularisation components, specifically tree depth and leaf node weights, are integrated into the XGB objective function. Consequently, the iterative process minimises and improves tree construction efficiency. To reduce model intricacy, a level-wise decision tree expansion strategy is employed. The additive strategy of the tree model is formulated as follows:

$$y' = \sum_{i=1}^{N} f_i(x_i) \rightarrow f_i \in F \tag{6}$$

where f and N indicate the tree objects and their number. F and x_i are the set of regression trees and the ith eigenvector, respectively. Thus, the fitness function that should be minimised is

$$L(\gamma) = \sum_{i=1}^{N} l(y', y_i) + \sum_{i=1}^{N} \phi(f_i) \tag{7}$$

$$\phi(f) = \alpha T + \frac{1}{2}\beta\|(w)\|^2 \tag{8}$$

where loss function and model complexity penalty factor are indicated by l and ϕ. α and β are L_1 and L_2 regularisation coefficients, and the tree's divided weight is shown by w. The challenging point in learning tree parameters at once results in the developed objective function as follows:

$$L^t = \sum_{i=1}^{N} [y_i, y_i^{'(t-1)} + f_t(x_i)] + \phi(f_i) \tag{9}$$

The second-order Taylor expansion is used to modify the loss function:

$$L^t = \sum_{j=1}^{N} [(\sum_{i\in I_j} g_i)w_j + \frac{1}{2}(\sum_{i\in I_j} h_i + \lambda)w_j^2] + \lambda T \tag{10}$$

where g_i and h_i are the loss function first and second derivatives in the gradient direction.

3.5. Deep dense neural networks (DNNs)

DNNs are a popular artificial neural network type consisting of multiple densely connected layers. A DNN typically contains several hidden layers depending on the complexity level of data patterns [33]. In a DNN, each neuron in a layer is connected to every neuron in the following layer, creating a dense matrix of connections. The input layer receives the input data and then processes it through a series of hidden layers consisting of densely connected neurons. Each neuron in

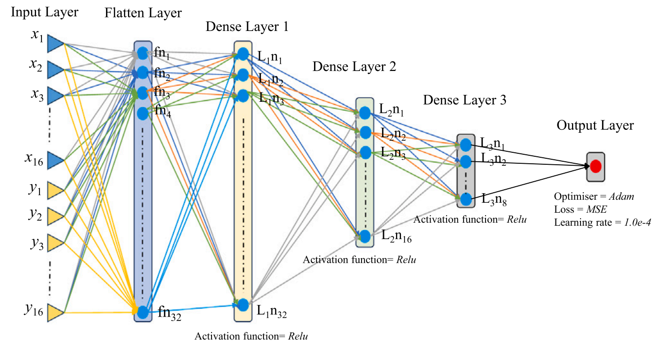


Fig. 5. The architecture of dense learning model.

the hidden layer receives inputs from all the neurons in the previous layer and applies a non-linear activation function to the weighted sum of those inputs. The output of the activation function is then passed on to the next layer as input. deep models (i.e., models with many layers) are better able to capture and represent the intricate relationships and details present in the input data. This is because deeper layers of the model can learn to identify increasingly abstract and complex features, which can help the model make more accurate predictions or classifications. However, training deep neural networks can be challenging due to the vanishing gradient problem, where the gradients of the loss function concerning the weights become very small as they propagate through multiple layers, leading to slow convergence or even stagnation of the learning process [34]. The architecture of the dense model applied in this study can be seen in Fig. 5.

3.6. Transfer learning approach

In traditional deep learning, when faced with new tasks, it is necessary to thoroughly analyse the specific characteristics of these tasks in order to formulate appropriate models. This process also requires a substantial amount of data annotations. Additionally, it is crucial that the training and test data are scheduled in the same distribution to ensure an accurate evaluation of the model's performance [35]. However, in order to enhance efficiency and save time, transfer learning presents a different approach. The primary objective of transfer learning is to identify similarities between a new problem and a previously solved problem, allowing for the acquisition of new knowledge through the process of feature transferring. Transfer learning [36] encompasses two fundamental concepts, namely domain and task. The domain refers to learning, which primarily consists of data and probability distributions. On the other hand, the task represents the goal of learning, composed of the labels and the corresponding functions.

3.7. Meta Ensemble model with extreme gradient boosting (MLGBM)

This study considers some critical steps to propose an effective Meta ensemble model to predict the total power putout of a wave farm with 16 WECs that is able to reduce bias and variance of the prediction error. These steps include:

- The first step is carefully selecting the appropriate sub-learners based on specific criteria. In the case of regression tasks, predictive accuracy is a common criterion. After conducting a literature review [11], it was determined that the most successful deep learning algorithms for WEC power output generation forecasts are DNN, FFNN, MLP, Linear regressions and decision tree methods. Therefore, we developed a comprehensive WEC power prediction framework to compare and find the best-performed base learner. The statistical analysis shows that DNN and CDNN outperform other ML models considerably (See Table 3).
- After selecting the sub-models, we trained them on various folds of the training WECs data using k-fold cross-validation. In the beginning, we selected just two sub-models for training and validation. Next, we increased the number of sub-models until the average performance improved in a greedy format.
- The next step involves defining the levelling of the base models. In this case, only one level of stacked modelling was selected due to the computational heaviness of deep learning. Adding more layers would make the system more complex, and multi-level deep learning stacking may not offer enough benefit in accuracy relative to the computational cost.
- Selecting an appropriate meta-learner is significant in integrating base learners' outputs as the final decisions. Various meta-learners were tested in this study to maximise diversity, including AdaBoost, LightBoost, and XGBoost. The predominant method for selecting the best meta-learner is to estimate the accuracy of the problem. The results of the experiments showed that XGBoost outperformed all other learning algorithms. The role of XGBoost as the meta-learner is to learn the optimal way to weigh or combine the predictions of the base learners. It learns the relationship between the base learners' outputs and the target variable, identifying patterns or dependencies that can improve the overall predictive performance. By leveraging its gradient boosting algorithm, XGBoost optimises the combination of base learners' predictions and generates a more accurate and robust final prediction.
- In the final step, a grid search was used to find the optimal hyper-parameters of the base learners and meta-learner.

The details of the proposed WEC prediction model are as follows and can be seen in Fig. 6. The feature combination serves a crucial role in our ensemble model. It fuses the predictions of the sub-learners into a unified representation that can be utilised by the meta-learner, in this case, XGBoost. This combined representation acts as a higher-level feature representation that encapsulates the collective knowledge of the sub-learners, significantly enhancing the model's ability to capture complex relationships and make accurate predictions. This is where the true power of our ensemble model lies.

The primary reason for selecting DNN models is that they are well known for their excellent flexibility and ability to generalise well in various types of data. This can be attributed to their ability to understand detailed patterns within the data and establish connections between different parts of the input. More than that, DNNs possess the amazing and interesting ability to gain representations of the data in a self-ruling way, allowing them to extract and translate significant features at many levels of data. However, it is important to note that dense layers have certain negative effects. One such effect is their need for a large memory, which can result in insignificant computational expense during training. Also, dense layers are easily influenced by overfitting, an important event in which the model simply memorises the training data instead of learning patterns that can be applied to new, hidden data.

We applied CDNNs as an alternative base-learner because they employ convolutional layers and pooling operations to harness spatial invariance and translational equivariance characteristics in data like images. Furthermore, they make use of shared weights in convolutional layers, which decreases the number of trainable parameters in comparison to DNNs. This unique property of parameter-sharing enables CDNNs to effectively learn from vast datasets without falling into the trap of overfitting. Moreover, CDNNs incorporate multiple convolutional layers to gradually extract hierarchical features from the input, allowing them to grasp both local and global patterns present in the data. Nevertheless, focusing on the computationally demanding aspects is crucial, particularly when working with extensive datasets or complex architectures. CDNNs are susceptible to overfitting, especially in cases where the dataset is limited, or the model contains an excessive

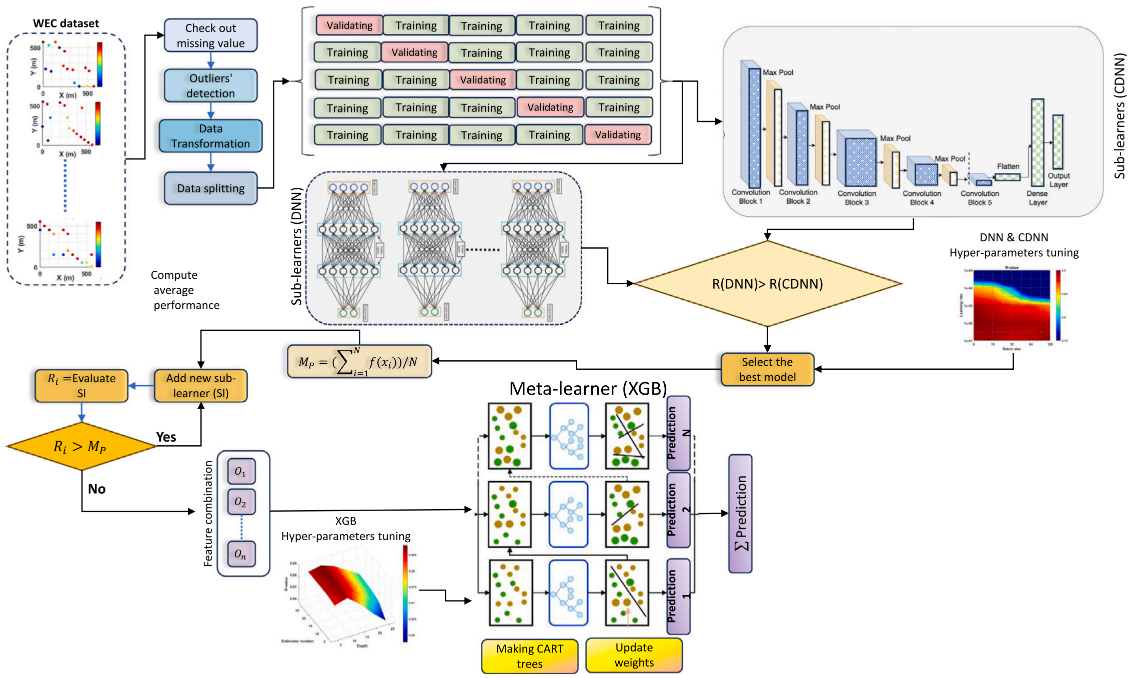


Fig. 6. The schematic of the proposed stacked ensemble model for predicting the total power out of a wave farm.

number of parameters. Therefore, it is essential to utilise effective regularisation methods like dropout or weight decay to address this challenge.

In our dataset [37], each feature corresponds to the coordinates (x or y) of a specific Wave Energy Converter (WEC). When predicting the total power output of the wave farm, we made a deliberate decision to retain all the features. This choice is primarily motivated by the application of the developed predictor in real-world scenarios within the wave energy industry. By including all the coordinate features, we aim to capture the spatial information and preserve the holistic representation of the wave farm. This ensures that the predictor can effectively handle the complexity and interplay between the WECs' positions, which is crucial for accurate power output estimation in practical wave energy applications.

3.8. Outlier detection

In this study, outlier detection was performed using the Local Outlier Factor (LOF) [38] method to identify and remove outliers in the dataset. LOF is a density-based method that assesses the local density of data points compared to their neighbours, allowing it to effectively detect anomalies. The outliers in this study were characterised by layouts exhibiting unusually high or very low total power output. The application of LOF resulted in the identification and removal of outliers from the dataset. It is worth noting that the number of layouts that were flagged as outliers and subsequently removed accounted for less than 1% of the total dataset. This indicates that the dataset contained a relatively small proportion of anomalous layouts that deviated significantly from the majority of the data points.

3.9. Transformation technique

Regarding the standardisation of data, we have implemented one of the most renowned transformation techniques known as Min–Max Normalisation (MMN). MMN stands as a highly favoured method [39] that effectively adjusts the unprocessed data to fit within a predefined lower and upper limit. Generally, this process involves re-scaling the data to be confined within the spectrum of 0 to 1 or −1 to 1. The mathematical expression for MMN is outlined by the proportion of the disparity between the value of the feature and its minimum value to the span between the maximum and minimum values. The boundaries set for the re-scaling of the data are denoted by a_{Min} and a_{Max} correspondingly.

$$a'_{i,n} = \frac{a_{i,n} - \min\left(a_i\right)}{\max\left(a_i\right) - \min\left(a_i\right)}(a_{Max} - a_{Min}) + a_{Min} \tag{11}$$

Through the utilisation of MMN, a guarantee is established that the altered data falls within the anticipated range, thereby streamlining consistent and comparable depictions of the characteristics. This methodology ensures that the transformed data is standardised and uniform, enhancing the accuracy and reliability of the analysis conducted on the dataset. By employing MMN, the data normalisation process becomes more efficient and effective, enabling a seamless comparison of various features within the dataset. The utilisation of MMN aids in enhancing the interpretability and reliability of the data analysis results, as it ensures that the data is uniformly scaled and standardised for accurate comparisons and evaluations. Consequently, adopting MMN contributes to enhancing data quality and the precision of analytical outcomes, fostering a more robust and insightful data analysis process.

3.10. Hyper-parameters tuning

Grid search is a widely used technique for hyper-parameter tuning in machine learning [40]. The idea behind grid search is to create a grid of possible hyper-parameter values and train the model with each combination of values to find the best set of hyper-parameters that result in the highest performance metric. The grid is constructed by specifying a range of hyper-parameter values for each hyper-parameter of interest. The range can be discrete or continuous. Grid search is an exhaustive search method and can be computationally expensive, especially when dealing with a large number of hyper-parameters or a large dataset. However, it is a simple and effective way to find the optimal hyper-parameters for a given model and dataset.

In this study, we consider the depth and number of estimator parameters in meta-learner (XGB), which are crucial hyper-parameters, and optimising them can significantly improve the model's overall performance [40,41]. The depth of a tree in XGBoost controls the complexity of the model and its ability to capture intricate relationships between

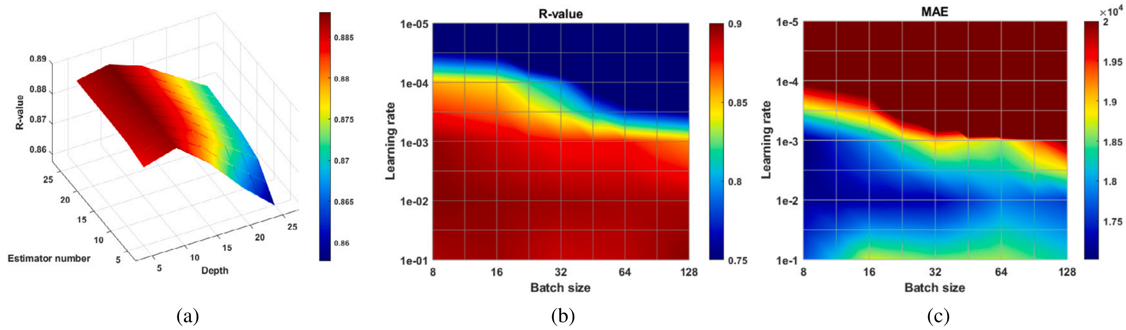


Fig. 7. Adjusting hyper-parameters of (a) XGBoost and (b) DNN in Meta-learner model using grid search.

the features. A deeper tree can learn more complex interactions but can also lead to overfitting. Therefore, finding the optimal depth is crucial in balancing the trade-off between model complexity and generalisation performance. The number of estimators in XGBoost controls the number of trees trained and combined to make the final prediction. Increasing the number of estimators can improve the model's performance by reducing the variance and expanding the model's ability to capture complex patterns in the data. However, increasing the number of estimators beyond a certain point can lead to overfitting, increased computational cost, and decreased model interpretability. Fig. 7 shows the landscape of the grid search for tuning the XGB's hyper-parameters. For this case study, the best-performed parameters are depth= 10, and the number of estimators= 20.

To optimise the performance of our sub-models (DNN or CDNN), we conducted a hyper-parameter tuning experiment using a grid search based on the Perth wave scenario. The experiment involved tuning several key hyper-parameters, including the number of neurons in each layer, the learning rate, the batch size, and the kernel initialisation method. In particular, we focused on the learning rate, which controls the step size for weight updates during training. We explored a range of learning rates, from $1e-01$ to $1e-05$, as it can significantly impact the model's performance and convergence. A careful choice between small and large learning rates is crucial. Another important hyper-parameter we considered was the batch size. This parameter determines the number of samples processed before weight updates occur in each training iteration. We experimented with batch sizes ranging from 8 to 128. The reason for this range is that smaller batch sizes can lead to faster convergence, while larger batch sizes can provide a more accurate estimate of the gradient. Fig. 7 depicts the landscape analysis of sub-model performance using different hyper-parameter settings. The analysis was based on the R-value (7(b)) and mean absolute error (MAE) (7(c)). We observed that smaller batch sizes, combined with higher learning rates, led to more accurate power predictions and lower validation errors. In addition to the previous findings, Fig. 8 provides insights into the R-value of prediction validation results for different configurations of three Dense layers in the sub-models. We varied the number of neurons in these layers within a range of 8 to 512. Significantly, we discovered a direct relationship between the prediction accuracy and the number of neurons in the first layer. Figs. 8(a) to 8(e) vividly demonstrate this, showing a consistent increase in accuracy as the number of neurons in the first layer escalates from 8 to 128. This underscores the practical importance of the number of units in the initial dense layer for achieving enhanced performance in real-world scenarios. The colour scheme in Fig. 8 represents the accuracy levels, where dark red and dark blue indicate the highest and lowest accuracy in terms of the R-value, respectively. It is evident that the relationship between the number of neurons in the second and third layers is nonlinear, resulting in a multi-modal landscape. This observation implies that the optimal configuration for these layers may not necessarily follow a linear progression. By analysing the landscape of the R-value across different configurations of the Dense layers, we gain valuable insights into the impact of varying the number of neurons in each layer.

Table 1 shows the hyper-parameters and settings of the ML methods applied in this study. It is noted that changing hyper-parameters may lead to different prediction results.

3.11. Setup of implementations

The implementation was done through *Spyder (4.1.5)*, a cross-platform integrated development environment under the *Anaconda (2.1.1)* distribution of *Python (3.7.9)*. Developing all 15 machine learning models required using a modular and extensible open-source library for *Keras (2.12)* [42] and *scikit-learn (1.2.2)* [43]. Keras works on top of *Tensorflow (2.12.0)* [44], an open-source machine learning platform developed by Google as an infrastructure layer for differential programming. The XGBoost gradient boosting library was utilised from Ref. [45] to develop the meta-learner.

4. Numerical results and discussions

In this section, we provide an overview of the ML models and their settings employed in our study to predict the power output of wave farms at four sea sites around Australia. Initially, we introduce all the ML models utilised, including their configurations and parameters (See Table 1). Afterwards, we thoroughly analyse and compare the performance of our proposed model with other ML models employed in the study. We assess the effectiveness of the proposed model by evaluating its performance in comparison to alternative models. Additionally, we investigate the impact of varying the number of sub-learners on the overall performance of the proposed predictor model. This analysis allows us to determine the optimal sub-learner configuration for achieving the best predictive results. Finally, we assess the potential of the proposed model for training based on a wave model dataset and subsequently test it with different wave data.

4.1. Evaluation metrics

Several metrics can be considered when comparing the effectiveness of a proposed prediction model with other models in a regression problem. Choosing the appropriate metrics for the specific task and data is essential to validate the model's performance on a held-out test set to ensure the model is balanced with the training data (no over-fitting or under-fitting issues). This study considers a wide range of metrics categorised into accuracy and bias. A list of applied evaluation metrics in this study can be seen in Table 2.

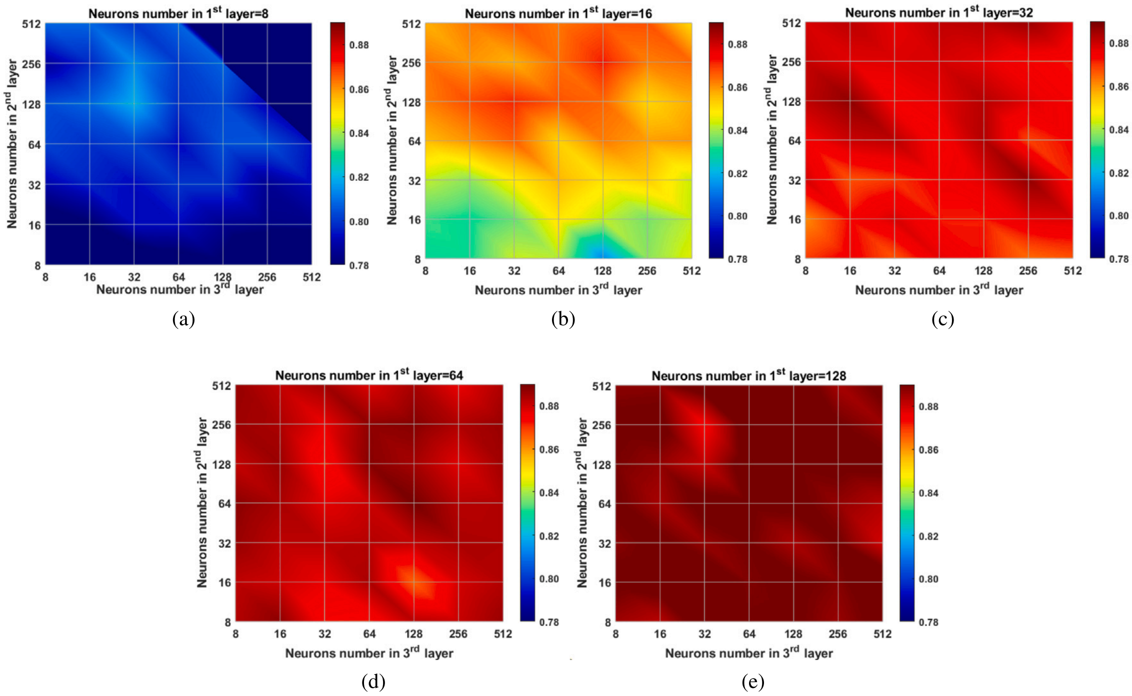


Fig. 8. Adjusting hyper-parameters of DNN in Meta-learner model using grid search.

Table 1
The technical settings of the Machine and Deep learning methods.

#	Acronym	Full name	Hyper-parameters
1	KNN	K Nearest Neighbours	(K=Number of neighbours)
2	LoR	Logistic Regression	solver=‘lbfgs’, penalty=‘l2’,tol=0.0001, C=1.0, maxiter=100
3	LR	Linear Regression	pre-defined settings (scikit-learn)
4	Lasso	Lasso Regression	α=1.0, fit intercept=True, precompute=False, max iter=1000, tol=0.0001
5	EN	Elastic Net	α=1.0, L1 ratio=0.5, max iter=1000, tol=0.0001
6	DT	Decision Tree Regressor	criterion=‘squared error’, splitter=‘best’, max depth= D, min samples split=2, min samples leaf=1, min weight fraction leaf=0.0,
7	MLP	Multi-layer Perceptron	solver=‘adam’, activation=‘relu’, alpha=1e−4, hidden layer sizes=(200,20,), maxiter=1000
8	PAR	Passive Aggressive Regressor	C=1.0, fit intercept=True, max iter=1000, tol=0.001, early stopping=False, validation fraction=0.1, n iter no change=5, shuffle=True, verbose=0, loss=‘epsilon insensitive’, epsilon=0.1
9	BR	Bayesian Regression	niter=300, tol=0.001, α_1=1e−06, α_2=1e−06, λ_1=1e−06, λ_2=1e−06
10	SGD	Stochastic gradient descent	loss=‘squared error’, penalty=‘l2’, alpha=0.0001, l1 ratio=0.15, max iter=1000, tol=0.001, shuffle=True, epsilon=0.1, learning rate=‘invscaling’, eta0=0.01, power t=0.25,
11	AdaB	AdaBoost	number estimators=50, learning rate=1.0, loss=‘linear’, base estimator=‘deprecated’
12	XGB	XGBOOST	Number of estimators=10,max depth=10 , $\gamma = 2$, $\eta = 0.99$, $reg_\alpha = 0.5$, $reg_\lambda = 0.5$
13	LightGBM	LightBoost	metric= ‘rmse’, num iterations=50, num leaves= 100, learning rate= 0.001, feature fraction= 0.9, max depth= 10
14	DNN	Dense Neural networks	Neuron number=32, 16, and 8, kernel initialiser=‘normal’, activation=‘relu’, lr = 0.0001, Optimiser=Adam
15	CDNN	Convolutional Dense Neural networks	filters=64, kernel size=3, kernel initialiser=‘normal’, activation=‘relu’, lr = 0.0001, loss=‘mean squared error’, Optimiser=Adam

4.2. Proposed model evaluations and comparisons

For a detailed comparison, we selected seven well-known linear and extended linear regression models, including Lasso, Linear regression, Logistic regression, Passive aggressive regression, Ridge regression, Bayesian regression and Elastic net method, to predict 16 WECs power output using a linear relationship estimation between dependent variables (coordination of WECs) and independent variables (Power Output (Watt)). Furthermore, to extract nonlinear and complex relationships among the features and target, we tested the performance of five advanced machine learning models such as SGD, SVM, MLP, DNN and Decision tree. Finally, due to several sharp characteristics of ensemble models, such as high accuracy, robustness to overfitting, and better generalisation, we evaluated and compared the effectiveness of three popular ensemble models, XGBoost [32], Adaboost [46] and LightGBM [47]. The analyses presented in Tables 3, and 4 provide

Table 2
The performance evaluation metrics for the prediction models.

Metrics	Definition	Equation
R-value	Pearson correlation coefficient	$R = \frac{\frac{1}{N_s}\sum_{k=1}^{N_s}(f_e(k)-\overline{f}_e)(f_t(k)-\overline{f}_t)}{\sqrt{\frac{1}{N_s}\sum_{k=1}^{N_s}(f_e(k)-\overline{f}_e)^2}\times\sqrt{\frac{1}{N_s}\sum_{k=1}^{N_s}(f_e(k)-\overline{f}_t)^2}}$
EVS	Explained variance score	$EVS = 1 - \frac{\sum_{k=1}^{N_s} Variance(f_t(k)-f_e(k))}{Variance(f_t(k))}$
MAE	Mean absolute error	$\text{MAE} = \frac{1}{N_s}\sum_{k=1}^{N_s} \lvert f_e(k) - f_t(k)\rvert$
MSLE	Mean squared log error	$\text{MSLE}= \frac{1}{N_s}\sum_{k=1}^{N_s}(log_e(1+f_t(k)) - log_e(1+f_e(k)))^2$
RMSE	Root mean square error	$\text{RMSE} = \sqrt{\frac{1}{N_s}\sum_{k=1}^{N_s}(f_e(k)-f_t(k))^2}$
SMAPE	Symmetric mean absolute percentage error	$\text{SMAPE}= \frac{1}{N_s}\sum_{k=1}^{N_s}\frac{\lvert f_t(k)-f_e(k)\rvert}{(\frac{1}{2}(f_t(k)+f_e(k)))}\times 100$

Table 3
The statistical results of seven regression methods, Lasso, Bayesian, Linear, Logistic, Ridge, Passive Aggressive performance, and also Elastic-net, SGD, SVM, MLP, DNN, and Decision Trees performance for ten independent runs with 5-fold cross-validation based on the Adelaide wave site.

	Lasso								Bayesian Regression						
	MSE	RMSE	MAE	MSLE	SMAPE	EVS	R-value		MSE	RMSE	MAE	MSLE	SMAPE	EVS	R-value
Min	2.518E+09	5.018E+04	4.028E+04	1.264E−03	2.854E+00	1.703E−01	4.130E−01	Min	2.523E+09	5.023E+04	4.027E+04	1.265E−03	2.852E+00	1.732E−01	4.163E−01
Max	2.610E+09	5.109E+04	4.099E+04	1.309E−03	2.904E+00	1.896E−01	4.360E−01	Max	2.609E+09	5.108E+04	4.105E+04	1.307E−03	2.906E+00	1.890E−01	4.350E−01
Mean	2.571E+09	5.071E+04	4.068E+04	1.289E−03	2.881E+00	1.795E−01	4.238E−01	Mean	2.561E+09	5.061E+04	4.059E+04	1.283E−03	2.874E+00	1.815E−01	4.261E−01
Median	2.572E+09	5.071E+04	4.069E+04	1.289E−03	2.881E+00	1.805E−01	4.248E−01	Median	2.560E+09	5.060E+04	4.054E+04	1.281E−03	2.871E+00	1.815E−01	4.260E−01
STD	2.469E+07	2.436E+02	2.023E+02	1.235E−05	1.422E−02	4.785E−03	5.647E−03	STD	2.220E+07	2.192E+02	2.074E+02	1.106E−05	1.464E−02	3.805E−03	4.495E−03
	Linear Regression								Logistic Regression						
	MSE	RMSE	MAE	MSLE	SMAPE	EVS	R-value		MSE	RMSE	MAE	MSLE	SMAPE	EVS	R-value
Min	2.507E+09	5.007E+04	4.030E+04	1.260E−03	2.856E+00	1.738E−01	4.170E−01	Min	1.699E+09	4.122E+04	3.184E+04	8.633E−04	2.262E+00	3.375E−01	6.072E−01
Max	2.610E+09	5.109E+04	4.096E+04	1.308E−03	2.899E+00	1.882E−01	4.341E−01	Max	2.063E+09	4.542E+04	3.473E+04	1.031E−03	2.459E+00	4.650E−01	7.014E−01
Mean	2.576E+09	5.075E+04	4.071E+04	1.292E−03	2.883E+00	1.807E−01	4.252E−01	Mean	1.762E+09	4.196E+04	3.292E+04	9.155E−04	2.336E+00	4.277E−01	6.757E−01
Median	2.582E+09	5.081E+04	4.071E+04	1.294E−03	2.883E+00	1.813E−01	4.259E−01	Median	1.705E+09	4.129E+04	3.278E+04	8.998E−04	2.327E+00	4.353E−01	6.825E−01
STD	2.451E+07	2.421E+02	1.780E+02	1.210E−05	1.226E−02	3.727E−03	4.420E−03	STD	9.318E+07	1.086E+03	8.543E+02	4.609E−05	5.812E−02	3.644E−02	2.583E−02
	Passive Aggressive Regression								Elastic net						
	MSE	RMSE	MAE	MSLE	SMAPE	EVS	R-value		MSE	RMSE	MAE	MSLE	SMAPE	EVS	R-value
Min	2.565E+09	5.065E+04	4.019E+04	1.283E−03	2.846E+00	1.613E−01	4.019E−01	Min	2.573E+09	5.072E+04	4.051E+04	1.285E−03	2.868E+00	1.608E−01	4.105E−01
Max	2.673E+09	5.170E+04	4.113E+04	1.337E−03	2.913E+00	1.789E−01	4.234E−01	Max	2.646E+09	5.144E+04	4.114E+04	1.323E−03	2.912E+00	1.729E−01	4.309E−01
Mean	2.613E+09	5.111E+04	4.057E+04	1.306E−03	2.873E+00	1.706E−01	4.132E−01	Mean	2.610E+09	5.109E+04	4.083E+04	1.306E−03	2.890E+00	1.671E−01	4.210E−01
Median	2.607E+09	5.106E+04	4.052E+04	1.304E−03	2.868E+00	1.713E−01	4.141E−01	Median	2.612E+09	5.111E+04	4.090E+04	1.307E−03	2.895E+00	1.665E−01	4.207E−01
STD	3.022E+07	2.953E+02	2.814E+02	1.546E−05	1.999E−02	4.136E−03	5.065E−03	STD	2.381E+07	2.332E+02	2.118E+02	1.204E−05	1.520E−02	3.739E−03	6.067E−03
	Ridge Regression								SGD						
	MSE	RMSE	MAE	MSLE	SMAPE	EVS	R-value		MSE	RMSE	MAE	MSLE	SMAPE	EVS	R-value
Min	2.524E+09	5.024E+04	4.031E+04	1.264E−03	2.853E+00	1.696E−01	4.121E−01	Min	2.526E+09	5.026E+04	4.038E+04	1.266E−03	2.859E+00	1.648E−01	4.078E−01
Max	2.615E+09	5.114E+04	4.108E+04	1.313E−03	2.910E+00	1.883E−01	4.340E−01	Max	2.626E+09	5.125E+04	4.109E+04	1.316E−03	2.909E+00	1.889E−01	4.348E−01
Mean	2.571E+09	5.070E+04	4.069E+04	1.288E−03	2.881E+00	1.794E−01	4.236E−01	Mean	2.578E+09	5.078E+04	4.074E+04	1.292E−03	2.884E+00	1.768E−01	4.208E−01
Median	2.567E+09	5.067E+04	4.064E+04	1.286E−03	2.877E+00	1.801E−01	4.244E−01	Median	2.581E+09	5.080E+04	4.072E+04	1.293E−03	2.883E+00	1.776E−01	4.216E−01
STD	2.635E+07	2.599E+02	2.322E+02	1.357E−05	1.645E−02	5.291E−03	6.180E−03	STD	2.255E+07	2.222E+02	1.889E+02	1.123E−05	1.328E−02	5.626E−03	6.564E−03
	Decision Tree								MLP						
	MSE	RMSE	MAE	MSLE	SMAPE	EVS	R-value		MSE	RMSE	MAE	MSLE	SMAPE	EVS	R-value
Min	9.670E+08	3.110E+04	2.299E+04	5.058E−04	1.650E+00	6.709E−01	8.198E−01	Min	9.967E+08	3.157E+04	2.475E+04	5.155E−04	1.768E+00	4.602E−01	6.784E−01
Max	1.028E+09	3.206E+04	2.363E+04	5.379E−04	1.696E+00	6.889E−01	8.304E−01	Max	1.710E+09	4.135E+04	3.311E+04	8.701E−04	2.353E+00	6.828E−01	8.264E−01
Mean	1.005E+09	3.170E+04	2.342E+04	5.255E−04	1.682E+00	6.793E−01	8.247E−01	Mean	1.295E+09	3.587E+04	2.858E+04	6.635E−04	2.035E+00	5.889E−01	7.663E−01
Median	1.009E+09	3.177E+04	2.349E+04	5.277E−04	1.687E+00	6.784E−01	8.241E−01	Median	1.288E+09	3.588E+04	2.855E+04	6.615E−04	2.033E+00	5.950E−01	7.713E−01
STD	1.798E+07	2.843E+02	1.881E+02	9.083E−06	1.324E−02	5.424E−03	3.232E−03	STD	2.039E+08	2.837E+03	2.383E+03	1.015E−04	1.668E−01	6.396E−02	4.199E−02
	SVM								DNN						
	MSE	RMSE	MAE	MSLE	SMAPE	EVS	R-value		MSE	RMSE	MAE	MSLE	SMAPE	EVS	R-value
Min	3.086E+09	5.555E+04	4.356E+04	1.532E−03	3.083E+00	1.326E−02	6.442E−01	Min	8.881E+08	2.980E+04	2.314E+04	4.562E−04	8.030E−01	6.211E−01	8.030E−01
Max	3.219E+09	5.674E+04	4.471E+04	1.598E−03	3.163E+00	1.362E−02	6.601E−01	Max	1.191E+09	3.451E+04	2.675E+04	6.068E−04	8.535E−01	7.239E−01	8.535E−01
Mean	3.165E+09	5.626E+04	4.428E+04	1.570E−03	3.133E+00	1.342E−02	6.524E−01	Mean	9.782E+08	3.125E+04	2.438E+04	5.007E−04	8.350E−01	6.881E−01	8.350E−01
Median	3.174E+09	5.633E+04	4.434E+04	1.571E−03	3.137E+00	1.338E−02	6.520E−01	Median	9.583E+08	3.096E+04	2.412E+04	4.903E−04	8.372E−01	6.944E−01	8.372E−01
STD	3.206E+07	2.854E+02	3.072E+02	1.701E−05	2.168E−02	1.208E−04	4.261E−03	STD	9.299E+07	1.451E+03	1.106E+03	4.623E−05	1.510E−02	3.049E−02	1.510E−02

a comparison of 15 ML models with the proposed Meta ensemble model performance based on seven evaluation metrics for a case study of the Adelaide wave site. The prediction accuracy shown by the R-value should be maximised, and a wide range of metrics for measuring the error of learning (MSE, RMSE, MAE, MSLE, and SMAPE) should be minimised. Moreover, the EVS values near 1.0 are the best and represent better squares of standard deviations of learning errors. As can be seen, the highest average validation accuracy (R-value) achieved by the proposed Meta ensemble model (MLGBM) at 88.8% and followed by XGBoost, LightGBM and Decision tree at 85.6%, 84.4%, and 82.5%, respectively. Therefore, the proposed model, which is a combination of DNN and XGBoost as a stacked ensemble meta-learner model, performed better than both DNN and XGBoost at 5.5% and 3.3%, respectively. Among seven linear and extended regression models, Logistic regression performance was considerably better than other regression models because it can capture nonlinear relationships by using nonlinear transformations of the independent variables, such as polynomial or interaction terms [48].

A more comprehensive comparative analysis is visualised to show the correlation coefficient between the predicted power output of WECs and the true values based on the Adelaide wave scenario that can be seen in Fig. 9. From this boxplot, we can observe that the prediction accuracy of the proposed ensemble model (MLGBM) significantly outweighs other ML models. The second important observation is that the performance of 20 independent runs of the MLGBM provides prediction accuracy with low variance. A low variance indicates that the proposed model is stable and consistent in its predictions across several runs. This is a desirable characteristic for machine learning models, as it is a sign that the model is reliable and can be trusted to make accurate predictions on new and unseen data.

Table 4
The statistical results of three ensemble ML models, including XGBoost, AdaBoost, LightGBM and also the proposed meta-learner model (MLGBM) performance for ten independent runs with 5-fold cross-validation based on the Adelaide wave site.

XGBoost	MSE	RMSE	MAE	MSLE	SMAPE	EVS	R-value
Min	8.156E+08	2.856E+04	2.107E+04	4.294E−04	1.516E+00	7.210E−01	8.501E−01
Max	8.805E+08	2.967E+04	2.193E+04	4.639E−04	1.576E+00	7.437E−01	8.642E−01
Mean	8.492E+08	2.914E+04	2.153E+04	4.475E−04	1.548E+00	7.301E−01	8.557E−01
Median	8.484E+08	2.913E+04	2.152E+04	4.465E−04	1.547E+00	7.301E−01	8.557E−01
STD	1.964E+07	3.374E+02	2.489E+02	1.034E−05	1.774E−02	5.725E−03	3.577E−03

LightGBM	MSE	RMSE	MAE	MSLE	SMAPE	EVS	R-value
Min	2.878E+09	5.365E+04	4.318E+04	1.436E−03	3.055E+00	6.304E−02	8.411E−01
Max	2.981E+09	5.460E+04	4.399E+04	1.487E−03	3.113E+00	6.431E−02	8.497E−01
Mean	2.936E+09	5.419E+04	4.356E+04	1.465E−03	3.082E+00	6.363E−02	8.443E−01
Median	2.944E+09	5.425E+04	4.355E+04	1.469E−03	3.083E+00	6.354E−02	8.431E−01
STD	3.417E+07	3.158E+02	2.513E+02	1.747E−05	1.799E−02	4.633E−04	3.092E−03

AdaBoost	MSE	RMSE	MAE	MSLE	SMAPE	EVS	R-value
Min	1.336E+09	3.655E+04	3.024E+04	6.700E−04	2.147E+00	6.622E−01	8.148E−01
Max	1.592E+09	3.990E+04	3.324E+04	8.047E−04	2.368E+00	6.978E−01	8.434E−01
Mean	1.465E+09	3.826E+04	3.173E+04	7.365E−04	2.255E+00	6.775E−01	8.280E−01
Median	1.470E+09	3.834E+04	3.177E+04	7.394E−04	2.259E+00	6.746E−01	8.278E−01
STD	5.525E+07	7.235E+02	6.670E+02	2.915E−05	4.911E−02	9.571E−03	7.642E−03

MLGBM	MSE	RMSE	MAE	MSLE	SMAPE	EVS	R-value
Min	6.504E+08	2.550E+04	1.970E+04	3.369E−04	1.408E+00	7.849E−01	8.859E−01
Max	6.815E+08	2.610E+04	2.010E+04	3.525E−04	1.437E+00	7.949E−01	8.916E−01
Mean	6.651E+08	2.579E+04	1.991E+04	3.442E−04	1.423E+00	7.886E−01	8.880E−01
Median	6.651E+08	2.579E+04	1.992E+04	3.442E−04	1.424E+00	7.875E−01	8.874E−01
STD	8.732E+06	1.692E+02	1.239E+02	4.321E−06	8.763E−03	3.938E−03	2.213E−03

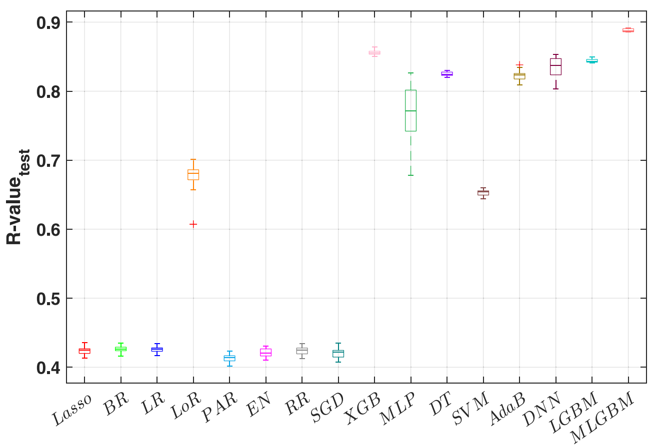


Fig. 9. Statistical analysis of 15 machine learning models compared with the proposed predictive model (MLGBM) in terms of prediction accuracy (R-value) of the total power output of 16 WECs based on the Adelaide wave scenario. Each method runs 20 times with random initialisation settings.

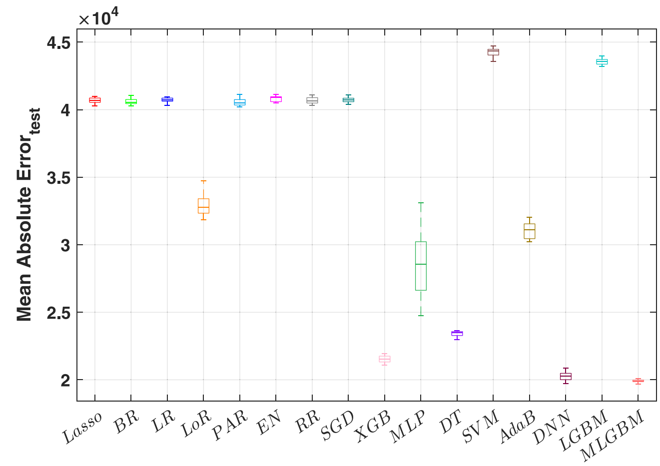


Fig. 10. Statistical analysis of 15 machine learning models compared with the proposed predictive model (MLGBM) in terms of mean absolute evaluation error of the total power output of the wave farm based on Adelaide wave scenario. Each method runs 20 times with random initialisation settings.

Analysing the learning error of the machine learning models and comparing it with other models is essential because it assists in identifying shortages and potential sources of learning error in the ML models. By understanding the learning error, we are able to identify problem spaces where the ML models are not performing well and improve their accuracy and reliability. In this regard, Fig. 10 shows the statistical analysis of 15 ML models compared with the proposed ensemble model based on the mean absolute validation error based on the Adelaide wave scenario. It can be seen that the minimum MAE recorded for MLGBM has a low variance. Moreover, the MAE of DNN and XGBoost is significantly lower than DT, AdaBoost and LightGBM. In order to evaluate effectively the proposed prediction model's capacity to extrapolate and excel in a multitude of distinct situations, we conducted extensive experiments on a wide range of wave farm datasets (Adelaide, Perth, Tasmania, and Sydney) with varying characteristics in terms of their geographical disposition, wave dynamics, and spatial arrangement. Consequently, subjecting the model to rigorous testing using diverse wave farm datasets enables us to ascertain its proficiency in delivering accurate predictions across many real-world scenarios, strengthening its applicability and reliability in practical settings.

Through these experimental analyses, we selected seven best-performed ML models from 15 ones to compare their effectiveness with the MLGBM.

Table 5 represents the statistical prediction results of the proposed model compared with others based on the Perth wave farm datasets. In terms of accuracy (R-value) metric, MLGBM greatly outweighs other ensemble and deep learning models and proposed a high level of accuracy of 90% on average. Moreover, MLGBM's mean absolute learning error is considerably less than AdaBoost, LightGBM, CDNN, DNN, and XGBoost at 85.1%, 28.6%, 25.7%, 17.7%, and 7.4%.

The modelling statistical results of eight wave farm power predictors for the Sydney wave site can be seen in Table 6. Similar to Adelaide and Perth wave layout datasets, we can see a significant superiority of the MLGBM compared with other popular models in seven various evaluation metrics. As the Sydney wave scenario (See Fig. 2d) exhibits different wave direction patterns and variability, we can evaluate the robustness of the proposed deep model and assess how well it handles different wave characteristic variations. These prediction results help ensure that the proposed model performs reliably under various conditions and is not overly sensitive to specific wave characteristics.

Finally, Table 7 shows the detailed prediction results of eight ML models for modelling the total power output of 16 WEC layouts located at the Tasmania sea site. Tasmania is exposed to the Southern Ocean, which can result in powerful and consistent wave energy. These waves can travel long distances before reaching the coastline, leading to larger swells. The highest power prediction accuracy is related to MLGBM at 84.4%, where CDNN, XGBoost and DNN accuracy are 74%, 79.6%, and 81%, respectively.

Fig. 11 serves as an illustration, visually depicting the summary statistics that are associated with the wave power predictors across three distinct wave farms, which are specifically referred to as Perth (a and b), Sydney (c and d), and Tasmania (e and f). To conduct a comprehensive evaluation, two performance metrics, namely Accuracy and MAE, are taken into consideration. It is significant to note that summary statistics encompass fundamental measures such as median,

Table 5
The statistical results of seven advanced machine learning methods compared with the proposed meta-learner model (MLGBM) performance for ten independent runs with 5-fold cross-validation based on the Perth wave site.

	XGBoost								MLP						
	MSE	RMSE	MAE	MSLE	SMAPE	EVS	R-value		MSE	RMSE	MAE	MSLE	SMAPE	EVS	R-value
Min	6.682E+08	2.585E+04	1.859E+04	3.612E−04	1.354E+00	7.372E−01	8.588E−01	Min	1.025E+09	3.201E+04	2.521E+04	5.369E−04	1.816E+00	4.872E−01	6.983E−01
Max	7.120E+08	2.668E+04	1.921E+04	3.859E−04	1.401E+00	7.575E−01	8.707E−01	Max	1.393E+09	3.732E+04	2.889E+04	7.325E−04	2.079E+00	6.266E−01	7.916E−01
Mean	6.900E+08	2.627E+04	1.882E+04	3.735E−04	1.371E+00	7.476E−01	8.649E−01	Mean	1.229E+09	3.503E+04	2.732E+04	6.449E−04	1.968E+00	5.498E−01	7.412E−01
Median	6.908E+08	2.628E+04	1.882E+04	3.740E−04	1.372E+00	7.484E−01	8.653E−01	Median	1.211E+09	3.480E+04	2.715E+04	6.359E−04	1.955E+00	5.515E−01	7.428E−01
STD	1.179E+07	2.245E+02	1.480E+02	6.647E−06	1.102E−02	4.661E−03	2.717E−03	STD	94 894 205	1362.107	1047.855	4.93E−05	0.074509	0.035772	0.024025
	Decision Tree								CDNN						
	MSE	RMSE	MAE	MSLE	SMAPE	EVS	R-value		MSE	RMSE	MAE	MSLE	SMAPE	EVS	R-value
Min	9.849E+08	3.138E+04	2.312E+04	5.266E−04	1.679E+00	6.105E−01	7.887E−01	Min	7.323E+08	2.706E+04	2.000E+04	3.936E−04	1.453E+00	6.550E−01	8.093E−01
Max	1.056E+09	3.249E+04	2.394E+04	5.630E−04	1.736E+00	6.398E−01	8.047E−01	Max	9.522E+08	3.086E+04	2.323E+04	5.051E−04	1.680E+00	7.345E−01	8.571E−01
Mean	1.027E+09	3.205E+04	2.358E+04	5.489E−04	1.711E+00	6.238E−01	7.963E−01	Mean	8.586E+08	2.928E+04	2.198E+04	4.576E−04	1.592E+00	6.878E−01	8.292E−01
Median	1.033E+09	3.213E+04	2.358E+04	5.526E−04	1.712E+00	6.232E−01	7.959E−01	Median	8.653E+08	2.942E+04	2.218E+04	4.612E−04	1.606E+00	6.853E−01	8.279E−01
STD	1.849E+07	2.892E+02	2.273E+02	9.565E−06	1.590E−02	8.744E−03	4.848E−03	STD	8.097E+07	1.403E+03	1.212E+03	4.105E−05	8.565E−02	3.004E−02	1.804E−02
	AdaBoost								DNN						
	MSE	RMSE	MAE	MSLE	SMAPE	EVS	R-value		MSE	RMSE	MAE	MSLE	SMAPE	EVS	R-value
Min	1.428E+09	3.779E+04	3.118E+04	7.315E−04	2.239E+00	5.891E−01	7.739E−01	Min	6.813E+08	2.610E+04	2.029E+04	3.564E−04	1.462E+00	7.232E−01	8.526E−01
Max	1.653E+09	4.065E+04	3.383E+04	8.500E−04	2.433E+00	6.366E−01	8.012E−01	Max	7.381E+08	2.717E+04	2.117E+04	3.860E−04	1.525E+00	7.510E−01	8.675E−01
Mean	1.536E+09	3.918E+04	3.248E+04	7.890E−04	2.334E+00	6.094E−01	7.866E−01	Mean	7.065E+08	2.658E+04	2.068E+04	3.706E−04	1.491E+00	7.394E−01	8.612E−01
Median	1.525E+09	3.905E+04	3.238E+04	7.823E−04	2.326E+00	6.077E−01	7.848E−01	Median	7.046E+08	2.654E+04	2.064E+04	3.701E−04	1.488E+00	7.398E−01	8.617E−01
STD	6.936E+07	8.830E+02	8.153E+02	3.640E−05	5.972E−02	1.378E−02	8.755E−03	STD	2.112E+07	3.967E+02	3.135E+02	1.092E−05	2.234E−02	8.858E−03	4.903E−03
	LightGBM								MLGBM						
	MSE	RMSE	MAE	MSLE	SMAPE	EVS	R-value		MSE	RMSE	MAE	MSLE	SMAPE	EVS	R-value
Min	8.162E+08	2.857E+04	2.216E+04	4.343E−04	1.599E+00	6.845E−01	8.549E−01	Min	5.059E+08	2.249E+04	1.740E+04	2.675E−04	1.257E+00	8.082E−01	8.990E−01
Max	8.610E+08	2.934E+04	2.266E+04	4.595E−04	1.637E+00	7.036E−01	8.667E−01	Max	5.320E+08	2.306E+04	1.774E+04	2.808E−04	1.281E+00	8.167E−01	9.037E−01
Mean	8.414E+08	2.901E+04	2.246E+04	4.483E−04	1.622E+00	6.925E−01	8.594E−01	Mean	5.164E+08	2.272E+04	1.753E+04	2.726E−04	1.266E+00	8.109E−01	9.005E−01
Median	8.412E+08	2.900E+04	2.245E+04	4.481E−04	1.621E+00	6.931E−01	8.592E−01	Median	5.155E+08	2.270E+04	1.752E+04	2.721E−04	1.266E+00	8.107E−01	9.004E−01
STD	1.087E+07	1.875E+02	1.204E+02	6.237E−06	8.965E−03	4.477E−03	2.721E−03	STD	5.967E+06	1.310E+02	9.312E+01	3.215E−06	6.829E−03	2.254E−03	1.250E−03

Table 6
The statistical results of seven advanced machine learning methods compared with the proposed meta-learner model (MLGBM) performance for ten independent runs with 5-fold cross-validation based on the Sydney wave site.

	XGBoost								MLP						
	MSE	RMSE	MAE	MSLE	SMAPE	EVS	R-value		MSE	RMSE	MAE	MSLE	SMAPE	EVS	R-value
Min	1.640E+08	1.280E+04	9.128E+03	7.600E−05	6.180E−01	6.750E−01	8.277E−01	Min	2.386E+08	1.545E+04	1.177E+04	1.097E−04	7.949E−01	1.290E−01	3.591E−01
Max	1.759E+08	1.326E+04	9.464E+03	8.162E−05	6.410E−01	6.892E−01	8.358E−01	Max	4.678E+08	2.163E+04	1.731E+04	2.136E−04	1.166E+00	5.517E−01	7.428E−01
Mean	1.685E+08	1.298E+04	9.250E+03	7.813E−05	6.264E−01	6.839E−01	8.323E−01	Mean	3.110E+08	1.758E+04	1.366E+04	1.426E−04	9.218E−01	4.167E−01	6.401E−01
Median	1.677E+08	1.295E+04	9.244E+03	7.777E−05	6.259E−01	6.846E−01	8.321E−01	Median	3.040E+08	1.743E+04	1.362E+04	1.393E−04	9.190E−01	4.292E−01	6.551E−01
STD	3.188E+06	1.224E+02	7.704E+01	1.520E−06	5.272E−03	3.930E−03	2.241E−03	STD	52774057.87	1448.251034	1340.344529	2.38E−05	0.089713864	0.098886	0.086181
	Decision Tree								CDNN						
	MSE	RMSE	MAE	MSLE	SMAPE	EVS	R-value		MSE	RMSE	MAE	MSLE	SMAPE	EVS	R-value
Min	1.371E+08	1.171E+04	5.074E+03	6.384E−05	3.451E−01	7.074E−01	8.552E−01	Min	1.232E+08	1.110E+04	8.219E+03	5.660E−05	5.554E−01	7.449E−01	8.631E−01
Max	1.565E+08	1.251E+04	5.525E+03	7.290E−05	3.757E−01	7.423E−01	8.719E−01	Max	1.409E+08	1.187E+04	8.823E+03	6.487E−05	5.959E−01	7.679E−01	8.763E−01
Mean	1.465E+08	1.210E+04	5.297E+03	6.826E−05	3.602E−01	7.258E−01	8.634E−01	Mean	1.313E+08	1.146E+04	8.525E+03	6.035E−05	5.758E−01	7.560E−01	8.695E−01
Median	1.467E+08	1.211E+04	5.295E+03	6.829E−05	3.601E−01	7.247E−01	8.625E−01	Median	1.312E+08	1.145E+04	8.503E+03	6.032E−05	5.742E−01	7.548E−01	8.688E−01
STD	4.891E+06	2.021E+02	1.059E+02	2.313E−06	7.242E−03	9.226E−03	4.295E−03	STD	3.966E+06	1.729E+02	1.658E+02	1.856E−06	1.114E−02	6.408E−03	3.682E−03
	AdaBoost								DNN						
	MSE	RMSE	MAE	MSLE	SMAPE	EVS	R-value		MSE	RMSE	MAE	MSLE	SMAPE	EVS	R-value
Min	3.123E+08	1.767E+04	1.535E+04	1.417E−04	1.034E+00	6.101E−01	8.283E−01	Min	1.735E+08	1.317E+04	9.364E+03	8.023E−05	6.340E−01	6.418E−01	8.028E−01
Max	3.613E+08	1.901E+04	1.669E+04	1.641E−04	1.125E+00	6.673E−01	8.400E−01	Max	1.911E+08	1.383E+04	9.678E+03	8.850E−05	6.554E−01	6.672E−01	8.179E−01
Mean	3.411E+08	1.846E+04	1.610E+04	1.548E−04	1.085E+00	6.393E−01	8.348E−01	Mean	1.806E+08	1.344E+04	9.487E+03	8.346E−05	6.424E−01	6.560E−01	8.113E−01
Median	3.436E+08	1.854E+04	1.616E+04	1.559E−04	1.089E+00	6.399E−01	8.353E−01	Median	1.793E+08	1.339E+04	9.486E+03	8.285E−05	6.422E−01	6.578E−01	8.123E−01
STD	1.528E+07	4.158E+02	4.271E+02	6.957E−06	2.892E−02	1.466E−02	3.240E−03	STD	4.822E+06	1.784E+02	8.606E+01	2.287E−06	5.872E−03	7.849E−03	4.617E−03
	LightGBM								MLGBM						
	MSE	RMSE	MAE	MSLE	SMAPE	EVS	R-value		MSE	RMSE	MAE	MSLE	SMAPE	EVS	R-value
Min	1.549E+08	1.245E+04	9.479E+03	7.157E−05	6.404E−01	6.938E−01	8.715E−01	Min	9.363E+07	9.676E+03	5.635E+03	4.347E−05	3.817E−01	8.106E−01	9.004E−01
Max	1.640E+08	1.281E+04	9.671E+03	7.586E−05	6.534E−01	7.063E−01	8.786E−01	Max	9.934E+07	9.967E+03	5.802E+03	4.615E−05	3.932E−01	8.232E−01	9.073E−01
Mean	1.600E+08	1.265E+04	9.565E+03	7.400E−05	6.463E−01	6.995E−01	8.748E−01	Mean	9.696E+07	9.847E+03	5.727E+03	4.506E−05	3.880E−01	8.167E−01	9.038E−01
Median	1.596E+08	1.263E+04	9.564E+03	7.380E−05	6.462E−01	6.998E−01	8.746E−01	Median	9.743E+07	9.871E+03	5.728E+03	4.528E−05	3.882E−01	8.165E−01	9.036E−01
STD	2.392E+06	9.463E+01	5.505E+01	1.146E−06	3.777E−03	2.730E−03	1.713E−03	STD	1.837E+06	9.353E+01	4.410E+01	8.680E−07	3.020E−03	3.604E−03	1.979E−03

quartiles, and the identification of potential outliers. It is worth highlighting that MLGBM, an advanced machine learning model, showcases exceptional superiority in terms of accuracy when compared to other advanced machine learning models across all three wave scenarios. This remarkable capability of MLGBM, in turn, leads to the generation of the lowest mean absolute error predictions for the wave models in Perth and Tasmania. However, the decision tree (DT) model displays the lowest average learning error in Sydney. As a result, these significant findings provide invaluable insights into the model performances' central tendency, spread, and variability. The statistical prediction results of the proposed model were compared with others based on the Perth wave farm datasets can be seen in Fig. 11(a and b). In terms of accuracy (R-value) metric, MLGBM greatly outweighs other ensemble and deep learning models and proposed a high level of accuracy of 90% on average. Moreover, MLGBM's mean absolute learning error is considerably less than AdaBoost, LightGBM, CDNN, DNN, and XGBoost at 85.1%,

Table 7
The statistical results of seven advanced machine learning methods compared with the proposed meta-learner model (MLGBM) performance for ten independent runs with 5-fold cross-validation based on the Tasmania wave site.

	XGBoost								MLP						
	MSE	RMSE	MAE	MSLE	SMAPE	EVS	R-value		MSE	RMSE	MAE	MSLE	SMAPE	EVS	R-value
Min	4.521E+09	6.724E+04	4.989E+04	3.297E−04	1.339E+00	6.258E−01	7.914E−01	Min	6.257E+09	7.910E+04	6.167E+04	4.468E−04	1.645E+00	1.404E−01	3.748E−01
Max	4.687E+09	6.846E+04	5.072E+04	3.431E−04	1.362E+00	6.381E−01	7.992E−01	Max	1.091E+10	1.045E+05	8.154E+04	7.724E−04	2.169E+00	4.943E−01	7.031E−01
Mean	4.602E+09	6.784E+04	5.036E+04	3.363E−04	1.352E+00	6.326E−01	7.957E−01	Mean	8.538E+09	9.204E+04	7.153E+04	6.070E−04	1.905E+00	3.201E−01	5.542E−01
Median	4.596E+09	6.780E+04	5.036E+04	3.359E−04	1.352E+00	6.335E−01	7.961E−01	Median	8.119E+09	9.010E+04	6.987E+04	5.778E−04	1.862E+00	3.566E−01	5.971E−01
STD	4.774E+07	3.517E+02	2.255E+02	3.760E−06	6.251E−03	3.641E−03	2.315E−03	STD	1566279906	8413.646357	6674.533896	0.00011	0.176017302	0.123267	0.117169
	Decision Tree								CDNN						
	MSE	RMSE	MAE	MSLE	SMAPE	EVS	R-value		MSE	RMSE	MAE	MSLE	SMAPE	EVS	R-value
Min	6.663E+09	8.163E+04	6.092E+04	4.825E−04	1.632E+00	4.197E−01	6.863E−01	Min	5.097E+09	7.140E+04	5.498E+04	3.673E−04	1.470E+00	5.135E−01	7.168E−01
Max	7.319E+09	8.555E+04	6.342E+04	5.306E−04	1.701E+00	4.707E−01	7.100E−01	Max	6.104E+09	7.813E+04	5.977E+04	4.374E−04	1.596E+00	6.029E−01	7.765E−01
Mean	6.999E+09	8.366E+04	6.223E+04	5.076E−04	1.668E+00	4.435E−01	6.975E−01	Mean	5.672E+09	7.526E+04	5.764E+04	4.069E−04	1.539E+00	5.507E−01	7.418E−01
Median	6.956E+09	8.340E+04	6.221E+04	5.038E−04	1.668E+00	4.458E−01	6.978E−01	Median	5.688E+09	7.540E+04	5.778E+04	4.083E−04	1.543E+00	5.434E−01	7.370E−01
STD	1.882E+08	1.124E+03	6.942E+02	1.386E−05	1.902E−02	1.488E−02	5.978E−03	STD	4.298E+08	2.864E+03	1.957E+03	3.022E−05	5.168E−02	3.752E−02	2.515E−02
	AdaBoost								DNN						
	MSE	RMSE	MAE	MSLE	SMAPE	EVS	R-value		MSE	RMSE	MAE	MSLE	SMAPE	EVS	R-value
Min	9.980E+09	9.990E+04	8.296E+04	7.060E−04	2.213E+00	4.272E−01	6.602E−01	Min	4.197E+09	6.479E+04	5.033E+04	3.020E−04	1.345E+00	6.301E−01	7.956E−01
Max	1.238E+10	1.113E+05	9.306E+04	8.890E−04	2.494E+00	5.026E−01	7.098E−01	Max	4.626E+09	6.802E+04	5.252E+04	3.337E−04	1.404E+00	6.701E−01	8.194E−01
Mean	1.105E+10	1.051E+05	8.765E+04	7.872E−04	2.343E+00	4.632E−01	6.838E−01	Mean	4.443E+09	6.665E+04	5.136E+04	3.203E−04	1.373E+00	6.484E−01	8.064E−01
Median	1.088E+10	1.043E+05	8.666E+04	7.775E−04	2.318E+00	4.666E−01	6.856E−01	Median	4.448E+09	6.669E+04	5.138E+04	3.211E−04	1.373E+00	6.515E−01	8.083E−01
STD	7.179E+08	3.395E+03	3.029E+03	5.293E−05	8.286E−02	2.049E−02	1.275E−02	STD	1.079E+08	8.114E+02	6.914E+02	7.911E−06	1.863E−02	9.775E−03	5.867E−03
	LightGBM								MLGBM						
	MSE	RMSE	MAE	MSLE	SMAPE	EVS	R-value		MSE	RMSE	MAE	MSLE	SMAPE	EVS	R-value
Min	1.091E+10	1.045E+05	8.240E+04	7.728E−04	2.191E+00	1.070E−01	7.956E−01	Min	3.510E+09	5.925E+04	4.550E+04	2.535E−04	1.218E+00	7.030E−01	8.386E−01
Max	1.149E+10	1.072E+05	8.435E+04	8.130E−04	2.243E+00	1.117E−01	8.127E−01	Max	3.778E+09	6.147E+04	4.709E+04	2.735E−04	1.260E+00	7.197E−01	8.484E−01
Mean	1.119E+10	1.058E+05	8.341E+04	7.909E−04	2.217E+00	1.094E−01	8.042E−01	Mean	3.629E+09	6.024E+04	4.636E+04	2.624E−04	1.240E+00	7.120E−01	8.439E−01
Median	1.120E+10	1.058E+05	8.352E+04	7.916E−04	2.220E+00	1.093E−01	8.046E−01	Median	3.631E+09	6.026E+04	4.648E+04	2.624E−04	1.243E+00	7.116E−01	8.437E−01
STD	1.233E+08	5.826E+02	4.371E+02	8.605E−06	1.151E−02	1.069E−03	4.721E−03	STD	7.244E+07	6.006E+02	4.495E+02	5.363E−06	1.217E−02	4.538E−03	2.690E−03

28.6%, 25.7%, 17.7%, and 7.4%. In order to compare the performance of the models used in this study with the proposed model (MLGBM), we developed a statistical test called the Friedman test. This test is an extension of the Wilcoxon signed-rank test and serves as a non-parametric version of a one-way repeated measures analysis. As can be seen in Fig. 12, it is noteworthy that the MLGBM model stood out as the top-ranked model in four distinct wave situations, indicating its exceptional performance. Also, we can see the performance of both XG-Boost and LightGBM was considerable. Moreover, both CDNN and DNN accuracy were competitive; in particular, the CDNN model achieved impressive results in the Sydney and Adelaide wave scenarios, while the DNN model excelled in the Perth and Tasmania wave scenarios. These discoveries emphasise the effectiveness and competitiveness of MLGBM, XGBoost, LightGBM, CDNN, and DNN models in meeting our study's goals. The unique strengths demonstrated by each model in various wave scenarios offer valuable insights into their capabilities and potential uses.

4.3. Impact of sub-learners number

In stacking ensemble models, the number of sub-learners is crucial because it can impact the effectiveness of the whole model. Extending the stack with more sub-learners can boost the variousness of the models and diminish the risk of overfitting. However, the complexity and computational cost can be increased [49]. Generally, the stacked model may not extract the various patterns in the data and may be biased towards individual models. Conversely, if we select a large number of sub-learners, the stacked model may become too complicated and overfit the training data, leading to poor generalisation ability. In order to choose a suitable number of sub-learners, we followed a greedy strategy. We added the sub-learners until the average performance of all sub-models was better than the newly added sub-model. Fig. 13 indicates the prediction accuracy of the proposed model with different sub-learners numbers. It is obvious that four sub-learners could develop more reliable collaborative performance compared with fewer or more learners.

4.4. Transfer learning analysis

In this particular section of our study, we have conducted a series of transfer learning experiments with the aim of assessing the potential of utilising the knowledge acquired from training on a wave farm dataset and applying it to a distinct dataset without any form of fine-tuning. The results of these experiments, as presented in Table 8, and Fig. 14 exhibits the statistical analysis of the training process of the proposed predictor, employing four diverse wave farm datasets and subsequently testing this pre-trained model with three additional datasets. Notably, the pre-trained model was not subjected to any form of retraining concerning the target dataset; instead, it was solely tested. Significantly, the highest level of accuracy was attained when the MLGBM was trained using the Adelaide dataset and subsequently evaluated based on the Perth dataset, yielding an accuracy rate of approximately 60%. This impressive achievement can be attributed to the significant correlation observed between the wave characteristics at the Adelaide and Perth sea sites. On the other hand, the transfer learning model exhibited the lowest level of performance when the Sydney dataset was employed as the training dataset. This discrepancy in performance can be explained by the fact that the wave directions at the Sydney sea site differ significantly in comparison to the other three sites.

The above analysis on transfer learning demonstrates the inability of a pre-trained model, in the absence of fine-tuning, to achieve competitive performance when compared to the proposed model that does not employ transfer learning. It is important to note that the correlation between the wave scenarios of the source and target greatly influences the effectiveness of the machine learning (ML) models in predicting the power output of wave farms. This correlation serves as a crucial factor in enhancing the overall performance of the ML model with transfer learning. While fine-tuning is a beneficial approach, it is indeed worth exploring alternative sub-learner choices to enhance the transfer learning process further in our future works. Furthermore, we recognise the substantial role that techniques like Domain Adaptation [50] can play in enhancing transfer learning performance. Approaches such as adversarial training [51] or discrepancy-based methods can significantly boost transfer performance by effectively aligning feature distributions and addressing the disparity between source and target domains. These techniques present exciting avenues for us to explore and incorporate into our ongoing efforts to optimise the transfer learning process.

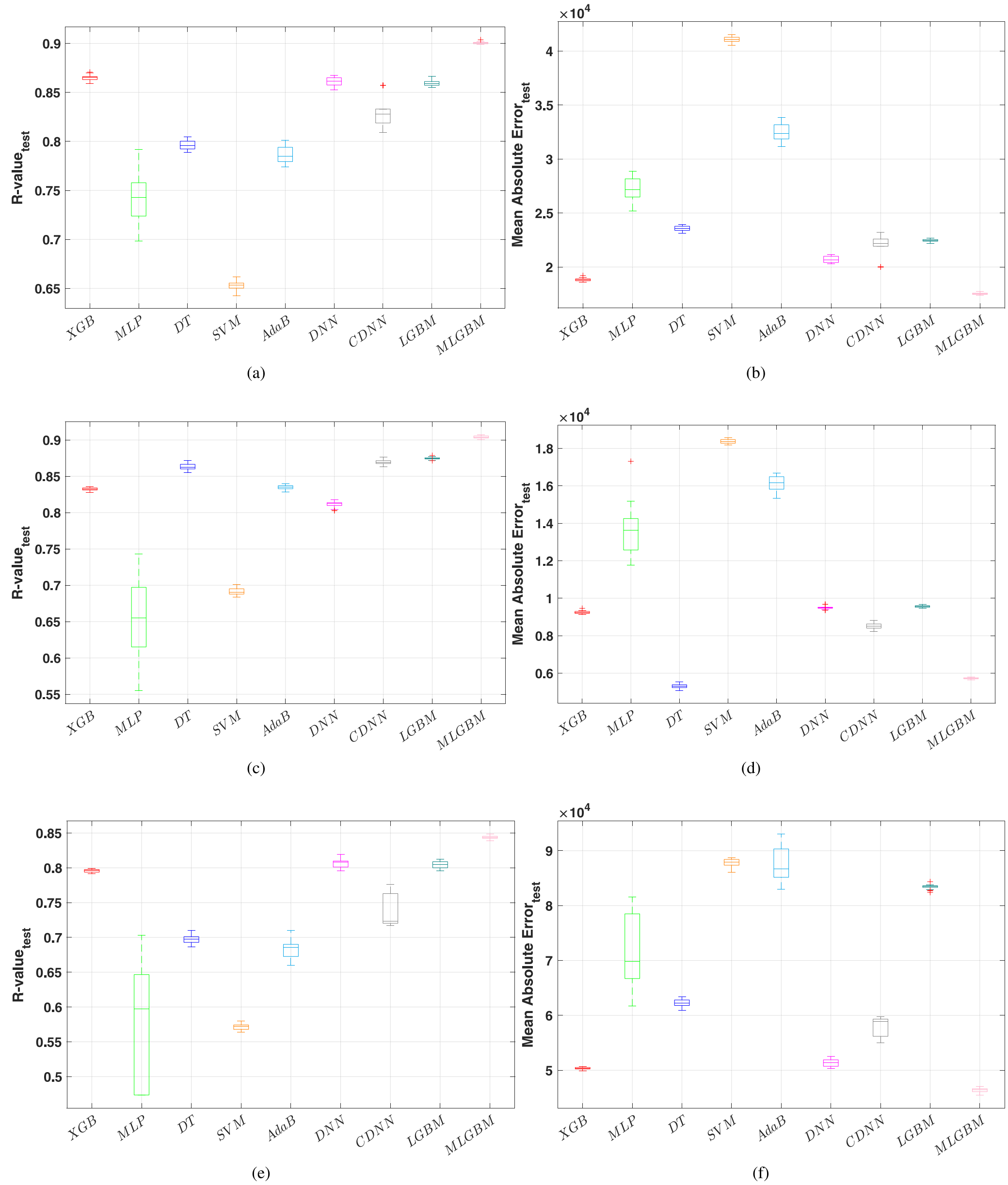


Fig. 11. A comparison performance analysis of the eight advanced ML models in terms of a, c, and (e) Accuracy (R-value) and b, d, and (f) MAE with the proposed predictor (MLGBM) based on the Perth (a and b), Sydney (c and d), Tasmania (e and f) wave scenario.

Table 8
The statistical results of transfer learning based on the proposed meta-learner model (MLGBM) performance for ten independent runs with 5-fold cross-validation.

Train	Test	MSE	RMSE	MAE	MSLE	SMAPE	EVS	R-value
Perth	Adelaide	3.243E+09	5.694E+04	4.379E+04	1.597E−03	3.103E+00	3.131E−01	5.597E−01
	Tasmania	5.702E+12	2.388E+06	2.386E+06	1.014E+00	9.291E+01	2.123E−01	5.163E−01
	Sydney	1.684E+10	1.298E+05	1.265E+05	8.354E−03	8.896E+00	−5.827E−01	3.559E−01
Adelaide	Perth	1.777E+09	4.212E+04	3.285E+04	8.970E−04	2.346E+00	3.565E−01	5.995E−01
	Tasmania	5.639E+12	2.375E+06	2.373E+06	9.945E−01	9.217E+01	2.148E−01	5.218E−01
	Sydney	1.349E+10	1.161E+05	−1.802E−01	6.616E−03	7.933E+00	−1.802E−01	4.429E−01
Tasmania	Perth	5.432E+12	2.331E+06	2.330E+06	9.672E−01	9.106E+01	−3.230E−01	5.230E−01
	Adelaide	5.361E+12	1.158E+06	2.315E+06	9.457E−01	9.018E+01	−1.015E−01	5.453E−01
	Sydney	4.840E+12	1.100E+06	2.199E+06	8.252E−01	8.505E+01	−2.699E+00	2.802E−01
Sydney	Perth	8.606E+09	9.276E+04	8.039E+04	4.306E−03	5.659E+00	−1.333E−01	3.960E−02
	Adelaide	7.059E+09	8.401E+04	7.184E+04	3.492E−03	5.029E+00	−1.945E−01	−2.212E−02
	Tasmania	5.287E+12	2.299E+06	2.296E+06	8.906E−01	8.786E+01	−7.445E−02	−4.110E−02

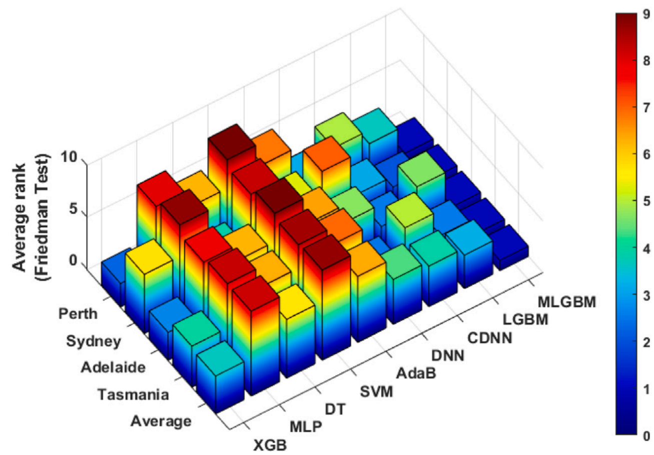


Fig. 12. A comparison performance (R-value) of the MLGBM with other ML models using the Friedman test.

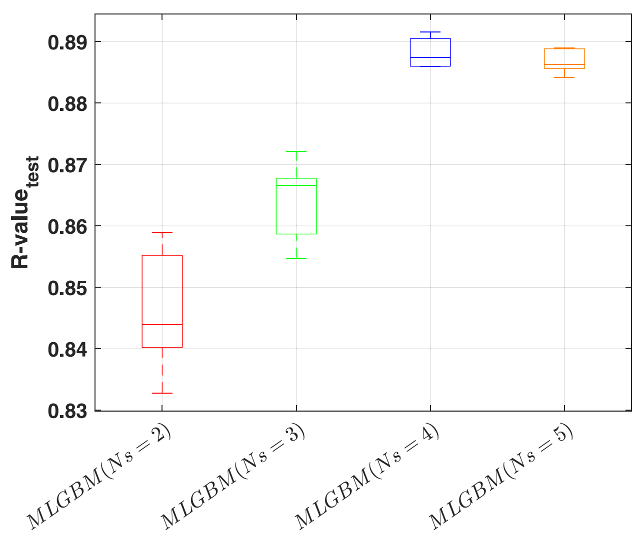


Fig. 13. A comparison performance analysis of the meta-learner using various sub-models number (N_s).

5. Conclusions

The prediction of power output in a wave farm is a complex task due to the unpredictable and chaotic nature of wave characteristics, as well as the intricate and non-linear hydrodynamic interaction between wave energy converters (WECs). In this study, we propose an effective Meta ensemble deep learning model that comprises an optimal number of convolutional dense neural networks combined with the extreme gradient boosting (XGBoost) method as a Meta-learner. The main objective of this model is to accurately predict the total absorbed power output of 16 WECs. In order to evaluate the performance of our proposed model, we compare it with 15 well-known machine-learning models.

To ensure the reliability and accuracy of our findings, we train and test our model using four real wave energy farm datasets obtained from the southern coast of Australia. Prior to the analysis, we conducted various pre-processing studies to enhance the effectiveness of the dataset. These studies include addressing missing values, normalising the data, and removing outliers. To determine an optimal number of sub-learners in our stacked ensemble model, we develop a greedy search method. This method allows us to find the best combination of sub-learners that would yield the most accurate predictions. Also, in order to fine-tune the hyper-parameters of the meta-learner, we employ a grid search technique. This technique can identify the optimal tree depth and the number of estimator parameters that would maximise the performance of our model.

Furthermore, we assess the transfer learning ability of our proposed model without fine-tuning. However, our analysis reveals that fine-tuning and re-training are crucial steps in order to improve the accuracy of the predictions. The experimental results obtained from our study provide strong evidence that our proposed model outperforms the other 15 machine-learning models that were considered. Our model demonstrates exceptional prediction capability and adaptability when it comes to handling complex and non-linear patterns present in wave energy converter data. These findings highlight our proposed model's effectiveness and potential in accurately predicting wave farms' power output, ultimately contributing to the optimisation and efficiency of wave energy conversion systems.

As part of our future strategic objectives, we intend to expand the transfer learning framework by employing sophisticated methodologies such as domain adaptation. This approach concentrates explicitly on transferring expertise from a source domain to a target domain where labelled data is scarce or absent. The ultimate goal is to bridge the disparity between the source and target domains by aligning their characteristic distributions or acquiring domain-invariant representations through learning.

CRediT authorship contribution statement

Mehdi Neshat: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Nataliia Y. Sergiienko:** Writing – review & editing, Writing – original draft, Visualization, Validation, Investigation, Data curation, Conceptualization. **Ashkan Rafiee:** Writing – review & editing, Investigation, Formal analysis, Conceptualization. **Seyedali Mirjalili:** Writing – review & editing, Supervision, Investigation, Conceptualization. **Amir H. Gandomi:** Writing – review & editing, Supervision, Resources, Investigation, Conceptualization. **John Boland:** Writing – review & editing, Supervision, Investigation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The dataset related to Wave Energy Converters for four Australian sea sites is available in the UCI Machine Learning Repository [37], providing easy access and availability to researchers and practitioners. The source code of the applied hydrodynamic simulator that was developed by MATLAB is available in [52].

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the authors used ChatGPT in order to enhance the writing quality and ensure grammatical accuracy. After using this tool, the authors carefully reviewed and edited the content as needed and take full responsibility for the content of the publication.

Acknowledgment

This research was funded by the Australian Government through the Australian Research Council (project number ARC IE230100545).

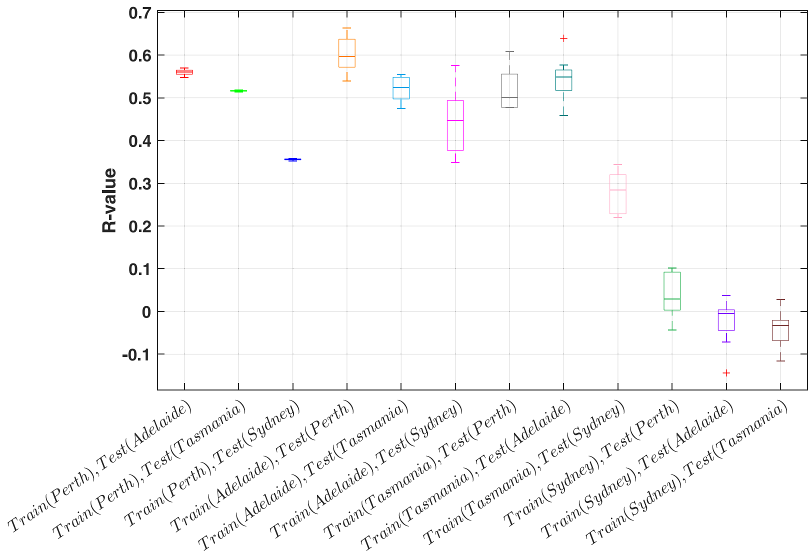


Fig. 14. Transfer learning impact in predicting the power output of the wave farm based on four real wave scenarios and the proposed meta-learner model (MLGBM).

References

[1] Lehmann M, Karimpour F, Goudey CA, Jacobson PT, Alam M-R. Ocean wave energy in the United States: Current status and future perspectives. Renew Sustain Energy Rev 2017;74:1300–13.

[2] Ali M, Prasad R, Xiang Y, Sankaran A, Deo RC, Xiao F, et al. Advanced extreme learning machines vs. deep learning models for peak wave energy period forecasting: A case study in Queensland, Australia. Renew Energy 2021;177:1031–44.

[3] Liu Z, Wang Y, Hua X. Prediction and optimization of oscillating wave surge converter using machine learning techniques. Energy Convers Manage 2020;210:112677.

[4] Sarkar D, Contal E, Vayatis N, Dias F. Prediction and optimization of wave energy converter arrays using a machine learning approach. Renew Energy 2016;97:504–17.

[5] Mann L, Burns A, Ottaviano M. CETO, a carbon free wave power energy provider of the future. In: Proceedings of the 7th European wave and tidal energy conference, vol. 108. 2007.

[6] Silva JM, Vieira SM, Valério D, Henriques JC. Model predictive control based on air pressure forecasting of OWC wave power plants. Energy 2023;284:129217.

[7] Huang Y, Tao J, Zhao J, Sun G, Yin K, Zhai J. Graph structure embedded with physical constraints-based information fusion network for interpretable fault diagnosis of aero-engine. Energy 2023;283:129120.

[8] Yuan X, Tan Q, Lei X, Yuan Y, Wu X. Wind power prediction using hybrid autoregressive fractionally integrated moving average and least square support vector machine. Energy 2017;129:122–37.

[9] Wang J, Hu J. A robust combination approach for short-term wind speed forecasting and analysis–Combination of the ARIMA (autoregressive integrated moving average), ELM (extreme learning machine), SVM (support vector machine) and LSSVM (least square SVM) forecasts using a GPR (Gaussian process regression) model. Energy 2015;93:41–56.

[10] Klaiber J, van Dinther C. Deep learning for variable renewable energy: A systematic review. ACM Comput Surv 2023.

[11] Wang H, Lei Z, Zhang X, Zhou B, Peng J. A review of deep learning for renewable energy forecasting. Energy Convers Manage 2019;198:111799.

[12] Li L, Yuan Z, Gao Y. Maximization of energy absorption for a wave energy converter using the deep machine learning. Energy 2018;165:340–9.

[13] Zou S, Zhou X, Weaver W, Abdelkhalik O. Deep reinforcement learning control of wave energy converters. IFAC-PapersOnLine 2022;55(27):305–10.

[14] Ni C, Ma X, Wang J. Integrated deep learning model for predicting electrical power generation from wave energy converter. In: 2019 25th international conference on automation and computing. IEEE; 2019, p. 1–6.

[15] Neshat M, Nezhad MM, Sergiienko NY, Mirjalili S, Piras G, Garcia DA. Wave power forecasting using an effective decomposition-based convolutional Bi-directional model with equilibrium Nelder-Mead optimiser. Energy 2022;256:124623.

[16] Mahdavi-Meymand A, Sulisz W. Application of nested artificial neural network for the prediction of significant wave height. Renew Energy 2023;209:157–68.

[17] Zhang J, Zhao X, Greaves D, Jin S. Modeling of a hinged-raft wave energy converter via deep operator learning and wave tank experiments. Appl Energy 2023;341:121072.

[18] Ni C, Peng W. An integrated approach using empirical wavelet transform and a convolutional neural network for wave power prediction. Ocean Eng 2023;276:114231.

[19] Mbuli N, Mathonsi M, Seitshiro M, Pretorius J-HC. Decomposition forecasting methods: A review of applications in power systems. Energy Rep 2020;6:298–306.

[20] Srokosz M. The submerged sphere as an absorber of wave power. J Fluid Mech 1979;95(4):717–41.

[21] Sergiienko N, Cazzolato B, Ding B, Arjomandi M. An optimal arrangement of mooring lines for the three-tether submerged point-absorbing wave energy converter. Renew Energy 2016;93:27–37.

[22] Wu G. The interaction of water waves with a group of submerged spheres. Appl Ocean Res 1995;17(3):165–84.

[23] Neshat M, Alexander B, Wagner M, Xia Y. A detailed comparison of meta-heuristic methods for optimising wave energy converter placements. In: Proceedings of the genetic and evolutionary computation conference. 2018, p. 1318–25.

[24] Carnegie Clean Energy. CETO 6 design update@ ONLINE. 2017.

[25] Khan W, Walker S, Zeiler W. Improved solar photovoltaic energy generation forecast using deep learning-based ensemble stacking approach. Energy 2022;240:122812.

[26] Guermoui M, Benkaciali S, Gairaa K, Bouchouicha K, Boulmaiz T, Boland JW. A novel ensemble learning approach for hourly global solar radiation forecasting. Neural Comput Appl 1–23.

[27] Li C, Lin W, Wu H, Li Y, Zhu W, Xie C, et al. Performance degradation decomposition-ensemble prediction of PEMFC using CEEMDAN and dual data-driven model. Renew Energy 2023;118913.

[28] Ganaie MA, Hu M, Malik A, Tanveer M, Suganthan P. Ensemble deep learning: A review. Eng Appl Artif Intell 2022;115:105151.

[29] Liu Y, Sun Y, Gao D-c, Tan J, Chen Y. Stacked ensemble learning approach for PCM-based double-pipe latent heat thermal energy storage prediction towards flexible building energy. Energy 2024;294:130955.

[30] Chen Y, Zhang S, Zhang W, Peng J, Cai Y. Multifactor spatio-temporal correlation model based on a combination of convolutional neural network and long short-term memory neural network for wind speed forecasting. Energy Convers Manage 2019;185:783–99.

[31] Zhang Y, Le J, Liao X, Zheng F, Li Y. A novel combination forecasting model for wind power integrating least square support vector machine, deep belief network, singular spectrum analysis and locality-sensitive hashing. Energy 2019;168:558–72.

[32] Chen T, Guestrin C. Xgboost: A scalable tree boosting system. In: Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining. 2016, p. 785–94.

[33] Menghani G. Efficient deep learning: A survey on making deep learning models smaller, faster, and better. ACM Comput Surv 2023;55(12):1–37.

[34] Tan HH, Lim KH. Two-phase switching optimization strategy in deep neural networks. IEEE Trans Neural Netw Learn Syst 2020;33(1):330–9.

[35] Wang Y-X, Chen Z, Zhang W. Lithium-ion battery state-of-charge estimation for small target sample sets using the improved GRU-based transfer learning. Energy 2022;244:123178.

[36] Pan SJ, Yang Q. A survey on transfer learning. IEEE Trans Knowl Data Eng 2009;22(10):1345–59.

[37] Neshat M. Wave Energy Converter (WEC) Dataset. 2019, https://archive.ics.uci.edu/dataset/494/wave+energy+converters UC Irvine Machine Learning Repository.
[38] Breunig MM, Kriegel H-P, Ng RT, Sander J. LOF: identifying density-based local outliers. In: Proceedings of the 2000 ACM SIGMOD international conference on management of data. 2000, p. 93–104.
[39] Singh D, Singh B. Investigating the impact of data normalization on classification performance. Appl Soft Comput 2020;97:105524.
[40] Claus HM. The importance of hyperparameter optimisation for facial recognition applications. In: Proceedings of the AAAI conference on artificial intelligence, vol. 36, no. 11. 2022, p. 13130–1.
[41] Srinivas P, Katarya R. hyOPTXg: OPTUNA hyper-parameter optimization framework for predicting cardiovascular disease using XGBoost. Biomed Signal Process Control 2022;73:103456.
[42] Chollet F, et al. Keras: The python deep learning library. 2018, p. ascl–1806, Astrophysics source code library.
[43] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine learning in Python. J Mach Learn Res 2011;12:2825–30.
[44] Abadi M, Barham P, Chen J, Chen Z, Davis A, Dean J, et al. Tensorflow: a system for large-scale machine learning. In: Osdi, vol. 16, no. 2016. Savannah, GA, USA; 2016, p. 265–83.
[45] Brownlee J. Xgboost with python: Gradient boosted trees with XGBoost and scikit-learn. 2016, Machine Learning Mastery.
[46] Freund Y, Schapire RE. A decision-theoretic generalization of on-line learning and an application to boosting. In: Computational learning theory: second European conference, euroCOLT'95 Barcelona, Spain, March 13–15, 1995 proceedings 2. Springer; 1995, p. 23–37.
[47] Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, et al. Lightgbm: A highly efficient gradient boosting decision tree. Adv Neural Inf Process Syst 2017;30.
[48] Zhao L, Chen Y, Schaffner DW. Comparison of logistic regression and linear regression in modeling percentage data. Appl Environ Microbiol 2001;67(5):2129–35.
[49] Webb GI, Zheng Z. Multistrategy ensemble learning: Reducing error by combining ensemble learning techniques. IEEE Trans Knowl Data Eng 2004;16(8):980–91.
[50] He J, Wu L. Cross-conditions capacity estimation of lithium-ion battery with constrained adversarial domain adaptation. Energy 2023;277:127559.
[51] Dong X, Sun Y, Dong L, Li J, Li Y, Di L. Transferable wind power probabilistic forecasting based on multi-domain adversarial networks. Energy 2023;285:129496.
[52] Sergiienko N. Wave Energy Converter (WEC) Array Simulator. 2024, https://www.mathworks.com/matlabcentral/fileexchange/71840-wave-energy-converter-wec-array-simulator MATLAB Central File Exchange.