

Effective Self Attention-Based Deep Learning Model with Evolutionary Grid Search for Robust Wave Farm Energy Forecasting

Amin Abdollahi Dehkordi^{1,7}, Mehdi Neshat^{2,3}, Natalia Y. Sergienko³, Zahra Ghasemi³, Lei Chen³, John Boland⁴, Hamid Moradkhani⁵, and Amir H. Gandomi^{*2,6}

¹Department of Computer Engineering, Najaf Abad Azad University ,
amin.abdollahi.dehkordi@azad.edu.ir

²Faculty of Engineering and Information Technology, University of Technology Sydney, Ultimo ,
mehdi.neshat@uts.edu.au

³School of Electrical and Mechanical Engineering, University of Adelaide ,
natalia.y.sergienko@adelaide.edu.au, [Zahra.Ghasemi@adelaide.edu.au](mailto:zahra.ghasemi@adelaide.edu.au) ,
[Lei.Chen@adelaide.edu.au](mailto:lei.chen@adelaide.edu.au)

⁴Industrial AI Research Centre, UniSA STEM, University of South Australia ,
John.Boland@unisa.edu.au

⁵Center for Complex Hydrosystems Research (CCHR), The University of Alabama, Tuscaloosa,
 USA , hmoradkhani@ua.edu

⁶University Research and Innovation Center (EKIK), Obuda University

⁷School of Systems & Computing, The University of New South Wales ,
a.abdollahidehkordi@unsw.edu.au

July 15, 2025

Abstract

Achieving carbon neutrality and mitigating greenhouse gas emissions represent critical global imperatives, as emphasized by the United Nations Sustainable Development Goal (SDG) #13. In this context, wave energy emerges as an up-and-coming renewable resource, theoretically capable of producing approximately 30,000 terawatt-hours (TWh) of clean electricity annually—exceeding current global electricity demand. Despite its vast potential, the wave energy sector remains in a nascent stage, constrained by several technical and economic barriers. A primary obstacle is the accurate forecasting of wave farm power output, which is crucial for maintaining grid stability, ensuring investment viability, and facilitating the broader commercial deployment of wave energy technologies.

To address this challenge, the present study investigates the convergence of green technological innovation and wave energy conversion, with a specific emphasis on forecasting the power output of wave energy farms. The goal is to facilitate the seamless integration of wave-generated electricity into the existing power infrastructure, thereby supporting a stable and resilient energy supply. To this end, we introduce a novel predictive framework for wave energy output, leveraging the spatial characteristics of Wave Energy Converter (WEC) locations. The proposed approach employs a hybrid sequential learning architecture—a Self-Attention-enhanced Convolutional Bi-directional Long Short-Term Memory (Bi-LSTM) network, augmented with an efficient hyperparameter optimisation algorithm. This methodology offers a robust and scalable solution for modelling the complex temporal-spatial dynamics of wave power generation. The model is trained and validated on four real-world datasets from operational wave farms along the southern coastline of Australia, specifically in Adelaide, Sydney, Perth, and Tasmania.

The performance of the proposed hybrid sequential predictive model is benchmarked against ten widely adopted machine learning (ML) algorithms. Experimental evaluations reveal that our model consistently outperforms the comparative methods, achieving high coefficients of determination (R^2) of 91.7% (Adelaide), 88.0% (Perth), 82.8% (Tasmania), and 91.0% (Sydney). These results underscore the superior predictive accuracy of the proposed framework across diverse marine environments. Moreover, the model demonstrates enhanced robustness and generalisation capability in forecasting wave farm power output, surpassing the performance of conventional ML and state-of-the-art deep learning approaches. Collectively, these findings affirm the effectiveness of the proposed Self-Attention Convolutional Bi-LSTM model in delivering reliable and accurate power predictions for wave energy systems.

Keywords: Renewable energy , Wave energy , Hybrid deep learning , Convolutional LSTM neural networks , Evolutionary grid search

*Corresponding author: gandomi@uts.edu.au

1 Introduction

The need to urgently address greenhouse gas emissions and achieve carbon neutrality has become a shared responsibility, emphasised by the United Nations Sustainable Development Goal [1]. Concrete actions are paramount in the pursuit of carbon neutrality, with green innovation emerging as a crucial approach closely aligned with this goal [2]. As societies and global supply chains strive to meet net-zero targets, those leading in green innovation and mastering innovative technologies will play a pivotal role in bridging the gap [3]. Recognising the pivotal role of green innovation in achieving carbon neutrality using modern Artificial Intelligence (AI) techniques is significant in exploring its implications and potential.

Energy generation with minimal environmental impact has become increasingly challenging [4] in recent years, prompting governments to support renewable energy systems, especially compared to non-renewable resources in today's energy market. Wave energy [5], a concentrated form of solar energy, holds particular promise among renewable energy sources. Ocean wave energy is not just efficient; it is one of the most efficient renewable sources, offering a high potential for large-scale clean energy production that surpasses other renewables like solar and wind by orders of magnitude. Wave energy converters facilitate the conversion of wave energy into electricity, a complex task in harnessing ocean wave energy [6]. Predicting the power output of wave energy converters in a wave farm is crucial for optimising their performance and reducing computational costs. Various models, including physical-based and statistical models, have been developed to address this challenge [7].

Numerous studies have been conducted on modelling, optimising, and predicting Wave energy and wave energy converters (WEC) performance in the last decade, with researchers constantly seeking ways to improve their efficiency and reliability. Statistical analysis has been utilised to forecast wave power, reducing computational costs. One area of interest is the application of artificial intelligence (AI) in evaluating the production potential of energy systems, providing efficient forecasting in this domain [8, 9]. A hybrid model was proposed by Liu et al. [10], a combination of a radial basis function neural network (RBFNN) and genetic algorithms (GAs) to predict and optimise a scaled oscillating wave surge converter. They numerically investigated these types of converters using the smoothed particle hydrodynamics method and random changes in nine design parameters and capture factors to assess energy conversion efficiency with high accuracy at 83.33%. While this approach shows promise, it is important to note that RBFNNs are computationally expensive, and selecting the optimal number of functions requires time and effort. A 30% increase was achieved in power capture efficiency using A feedforward back-propagation artificial neural network (FBNN) by Li et al. [11] that addressed non-causality in real-time wave energy control by online forecasting of future wave force. The impact of prediction errors on power extraction was investigated, showing that phase error had a greater influence than amplitude error. Additionally, a relationship between power capture efficiency and control constraint was identified.

Another hybrid predictive model was proposed [12] consisting of the long short-term memory (LSTM) algorithm and principal component analysis (PCA) for predicting electrical power generation from a WEC. The results demonstrate the superior performance of this integrated model compared to LSTM alone and other ML models, the support vector machine (SVM), regression tree (RT), Gaussian process regression (GPR) and ensembles of trees (ET). The experiments highlight the significant impact of high-frequency oscillating waves and long-term features on the model's accuracy, showcasing the proposed model's effectiveness in handling time sequence data and high-frequency signals. A comparative study [13] developed in order to predict electric power generated by a two-body hinge-barge WEC using four ML algorithms: Back-propagation Neural Network (BPNN), Long Short-Term Memory (LSTM), Support Vector Machine (SVM), and Radial Basis Neural Network (RBFF). The results demonstrate that BPNN achieved a coefficient of determination (R^2) of up to 0.99% on the test dataset.

In recent work [14], the impact of the ensemble learning model was emphasised and proposed a Meta-learner gradient boosting method (MLGB). The Meta wave learner was trained and validated using wave farm datasets from Australia's southern coast, and its performance was compared with 15 other ML methods. The experimental results demonstrate the proposed model's competitiveness, with accuracy ranging from 84.4% to 90.3% across different locations and improved robustness in predicting wave farm power output. One of the important parameters in wave power prediction is the significant wave height (H_{sig}), which has led to the application of AI models in numerous research studies. Time series models have shown superior performance compared to other ML models like linear regression, support vector machines (SVM), NNs, etc. For instance, Ali et al. [15] developed a hybrid ML model, ICEEMDAN-ELM, for forecasting significant wave height (H_{sig}) in the eastern coastal zones of Australia. The model combined the improved complete ensemble empirical mode decomposition method and the adaptive noise method with the extreme learning machine (ELM) algorithm, incorporating lagged historical H_s data as predictors. Comparisons with other models showed that this hybrid model outperforms them (such as random forest (RF)), providing accurate H_s forecasts. This model can be valuable for decision-making when designing reliable ocean energy converters.

Neural networks possess the capacity to autonomously uncover and depict hierarchical characteristics through the arrangement of interconnected nodes across multiple layers. This notable attribute enables them to grasp intricate and nonlinear associations within input data, rendering them potent models for prediction-oriented tasks. A particular investigation [16] demonstrated the efficacy of employing wavelet neural networks in enhancing the numerical forecasting of significant wave height and peak wave period in a case study focused on a coastal region adjacent to Puducherry, situated on India's east coast. By comparing the suggested prediction method to a classic neural network model trained solely on measured data from Puducherry, it was found that the former approach outperformed the latter, exhibiting sustained prediction performance of superior quality.

A considerable accuracy was achieved [17] in the near-real-time significant wave height forecast model using a hybrid approach combining multiple linear regression (MLR) with covariance-weighted least squares (CWLS) estimation and optimised to forecast Hsig values 30 minutes ahead. The trained and validated data was from the eastern coastal zones of Australia. The MLR-CWLS model is compared to other models, including MARS, M5 Model Tree, and MLR, and the results showed that the hybridised MLR-CWLS model provides reliable forecasts of Hsig compared to the other models ($R^2=0.97$). Pujan et al. [18] proposed a novel approach for forecasting significant wave heights by combining machine learning models, specifically transformer neural networks, with numerical methods such as differencing and WaveWatch III. The study conducted a case study using data from 92 buoys. The experimental results demonstrated that the proposed model achieved excellent performance, with a root mean square error of 0.231 m for forecasting significant wave heights two days in advance. Comparing the proposed method with LSTM, CNN, and Transformer models, the proposed approach outperformed the others, showing superior accuracy in predicting significant wave heights. In another study, a hybrid deep learning model called CLSTM-BiGRU [19] was recently proposed to accurately forecast H_{sig} for WEC sites in Queensland, Australia. The model combines Convolutional Neural Networks (CNNs), LSTM networks, and Bidirectional Gated Recurrent Units (BiGRUs) to predict H_{sig} at different time horizons. Historical wave properties and H_{sig} are utilised to train the hybrid model, presenting improved accuracy compared to benchmark ML models. However, the study did not compare the complexity and training runtime of the proposed model with other ML models.

To tackle the challenges mentioned above related to predicting the power output of wave farms, particularly considering the intricate hydrodynamic interplays between WECs and the inherent variability of wave conditions, precision in power output predictions remains a challenging task. To address these complexities, we introduce a robust and efficient hybrid sequential deep learning model. The primary contributions of this research endeavour are outlined as follows:

- A comprehensive analysis is conducted, contrasting ten established sequential and hybrid machine learning methods to formulate a specialised comparative AI-based framework tailored to precisely anticipate power output in wave farms across diverse real wave scenarios.
- Also, we propose an adept and precise hybrid model by combining a developed convolutional deep model with a multi-layer bidirectional long short-term memory network and a self-attention mechanism, thus fortifying the accuracy of predictions.
- In response to the challenge posed by hyper-parameters and architecture optimisation, the proposed hybrid model implements a swift heuristic algorithm, merging grid search techniques with the 1+1 Evolutionary Algorithm (1+1EA).
- Furthermore, the performance of the hybrid predictive model is meticulously evaluated utilising an extensive dataset reflective of four real wave scenarios, with multiple metrics employed to assess its efficacy.

Through these concerted efforts, this study aims to advance the state-of-the-art predictive modelling for wave farm power output, fostering a deeper understanding of how advanced machine learning methodologies can be harnessed to enhance predicting precision and operational efficiency in green energy conversion technologies.

This study begins with an initial exposition on the description and modelling of the wave energy converter (WEC) system, encompassing the fundamental equations of motion governing its behaviour, as well as a detailed explanation of the chosen deployment site and the performance measures employed for evaluating its effectiveness (detailed in Section 2). Subsequently, a variety of machine learning (ML) and deep learning algorithms are introduced, along with the intricate technical aspects of the proposed Meta ensemble deep learning method, outlined in Section 3. The numerical results obtained through the application of this method are then meticulously presented and discussed in Section 4, facilitating a comprehensive comparison of the efficiency and effectiveness of the proposed approach. Finally, the manuscript concludes by summarizing the key findings and highlighting the advantages offered by the proposed method over existing approaches (Section 5).

2 System description and modelling

The utilisation of submerged spheres as wave power absorbers was initially proposed in the literature[20] and has been implemented in the current study. The wave farm is composed of several Wave Energy Converters (WECs) positioned in proximity to one another. Each sphere has a radius of 5 meters, with the distance from the buoy's centre of mass to the still water level fixed at 8.5 meters. Three mooring lines are employed to anchor the sphere to the seabed. The mooring system's configuration is symmetrical to ensure that each WEC can independently absorb power regardless of the incoming wave direction. By inclining the mooring lines at 54 degrees relative to the vertical, the submerged sphere achieves optimal power absorption, as detailed in [21]. To maintain tension in the mooring lines, the buoys must be positively buoyant, necessitating that the buoy's mass be half that of the displaced volume. The wave farm comprises multiple closely situated WECs, as illustrated in Figure 1.

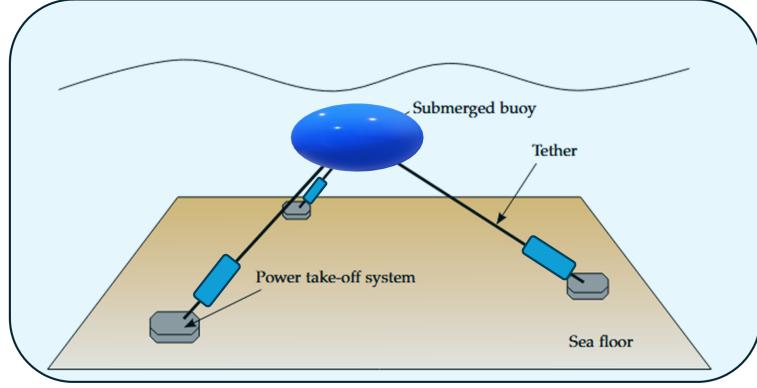


Figure 1: schematic of a fully submerged three tethers wave energy converter.

2.1 Equations of motion

The hydrodynamic loading on the Wave Energy Converters (WECs) is simulated using linear potential flow theory [22]. The power take-off mechanism of each WEC is represented as a simplified linear spring-damper system with adjustable stiffness and damping control parameters. The dynamics of the WEC array are mathematically described in the frequency domain, considering wave directionality. The motion of each WEC is analyzed in the surge, sway, and heave modes, with the rotational movement of the spherical WECs being disregarded. The equations of motion are formulated as follows:

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

In the given system, where ω represents the regular wave frequency, β signifies the wave angle, \mathbf{x} denotes the complex amplitude vector with dimensions of $[N \times 3, 1]$, where N stands for the count of WECs within a farm, \mathbf{M} is the diagonal mass matrix, \mathbf{A} and \mathbf{B} are matrices encapsulating added mass and radiation damping coefficients respectively, accounting for hydrodynamic interactions among WECs in the farm. Moreover, \mathbf{K}_{pto} and \mathbf{B}_{pto} are matrices depicting power take-off stiffness and damping, governing the PTO mechanism, while \mathbf{F}_e represents the wave excitation vector.

The assessment of the mean power absorbed by all WECs within a farm, contingent on the regular wave frequency and wave orientation, is determined as follows:

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

2.2 Deployment sites and wave climates

The efficiency of the wave farm is examined across four marine locations in Australia. The historical data for these specific sites is visualized in Figure 2, sourced from the Australian Wave Energy Atlas.

The deployment sites in Adelaide and Sydney exhibit distinct wave climate characteristics, which are critical in evaluating the performance and generalisability of wave energy forecasting models. As illustrated in the wave rose and joint probability distribution plots, Adelaide's wave climate is predominantly influenced by southwesterly swells, with significant wave heights (H_s) mainly ranging between 1.5 and 4.5 meters, and peak wave periods (T_p) concentrated around 9 to 13 seconds. The sea state occurrence in Adelaide shows a relatively high density of energetic wave conditions, indicating its strong potential for wave energy extraction. In contrast, Sydney is characterized by a broader directional spread, primarily from the southeast, with lower wave energy levels. Here, H_s typically remains below 3 meters, and T_p is generally between 7 and 11 seconds, with a more moderate sea state occurrence distribution. These variations in wave dynamics between the two locations provide valuable insights into the spatial diversity of marine energy resources along the southern Australian coast and underscore the need for site-specific modeling approaches to accurately forecast power output.

2.3 Performance measures

Utilising frequency domain data, it becomes feasible to assess the potential power absorption capacity of a wave farm within an irregular wave environment, defined by parameters such as the significant wave height H_s , peak wave period T_p , and wave angle β :

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

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

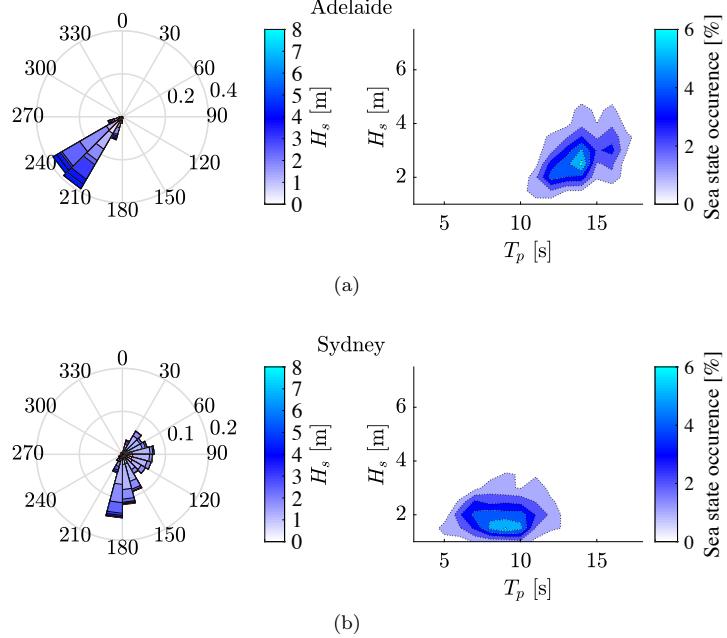


Figure 2: Wave scatter diagrams and the directional wave roses of four sites: (a) Adelaide, (b) Sydney.

After computing the power output of the wave farm for each i -th sea state, the average annual power generation can be approximated by considering the probability of each wave condition's occurrence, denoted as $O_i(H_s, T_p, \beta)$:

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

3 Methods

Recurrent neural networks [23] (RNNs) are an efficient and robust sort of neural network and algorithm with internal storage, making them one of the most intriguing algorithms. RNNs, like many other deep learning techniques, are relatively new. Despite being designed in the 1980s, it has only recently demonstrated its actual capabilities. With internal storage, the RNN can memorize significant aspects of the information it processes, allowing it to anticipate what will happen next with high accuracy. As a result, they are the finest algorithms for time series, voice, text, financial data, audio, video, and weather. Recurrent neural networks, as opposed to other algorithms, can provide a more in-depth comprehension of sequences and their surroundings. The information in an RNN is cycled in a loop. When making a decision, it takes into account both the input information and what it has learned from prior inputs. The vanishing gradient problem happens when the gradient decreases through time as it rises. As a result, the layer that receives a modest gradient does not learn and provides the network with short-term memory. However, using LSTMs as a solution to the problem is another option [24].

LSTM [25] is a sort of recurrent neural network (RNN) that is used to model long-distance sequences. It is a specialized RNN designed to address the gradient vanishing issue. A technique known as a gate allows LSTM to learn long-term reliance. These gates can learn which information in the sequence should be kept or discarded. The LSTM is comprised of three gates: input, forgetting, and output. In summary, the Forget gate chooses which data from the preceding phase should be saved. The input gate selects the necessary information to be added from the current stage. The next concealed state is determined by the output gate. According to the research on the predictions analyzed, we opted to construct and evaluate nine distinct types of LSTM models and a CNN to estimate the power production of a wave farm. Each model has different unique properties. They are as follows: Vanilla LSTM, Stacked LSTM, Bi-directional LSTM, Stacked Bi-LSTM, Gated recurrent unit (GRU), Stacked GRU, CNN, CNN-LSTM, CNN-GRU, CNN-BiLSTM.

3.1 Stacked LSTM

Stacked LSTM networks are a type of recurrent neural network (RNN) that consists of multiple LSTM layers stacked on top of each other. Each layer captures different levels of abstraction and temporal dependencies in the input data, allowing the network to learn even more complex patterns in sequential data. The mathematical equations related to stacked LSTM networks can be summarized as follows. The forward LSTM layer processes the input sequence from left to right. The equations for the forward LSTM are similar to those of the standard LSTM but with the

indices unchanged. The backward LSTM layer processes the input sequence from right to left. The equations for the backward LSTM are similar to those of the standard LSTM but with the indices reversed. Combining Forward and Backward LSTM Outputs: The final output of the BiLSTM network is the weighted sum of the forward and backward LSTM outputs.

$$o_t = \alpha o_{t,f} + (1 - \alpha) o_{t,b} \quad (5)$$

where o_t is the final output, α is the weight given to the forward LSTM output, and $o_{t,f}$ and $o_{t,b}$ are the forward and backward LSTM outputs, respectively. Stacked LSTM networks consist of multiple LSTM layers stacked on top of each other. The output of each LSTM layer is fed as input to the next LSTM layer. The equations for the stacked LSTM can be written as follows:

$$h_t^{(l)} = \text{LSTM}_l(h_t^{(l-1)}) \quad (6)$$

where $h_t^{(l)}$ is the hidden state of the l_{th} LSTM layer at time step t , and LSTM_l is the l_{th} LSTM layer. These equations describe the flow of information through a stacked LSTM network, which allows it to capture complex temporal dependencies in sequential data. Stacked LSTM networks have been demonstrated to be effective in various natural language processing tasks, including language modeling, machine translation, and speech recognition.



Figure 3: The stacked long short-term memory (LSTM) architecture

3.2 Bi-directional LSTM

A Bidirectional Long Short-Term Memory [26] (BiLSTM) network is a type of recurrent neural network (RNN) that learns bidirectional long-term dependencies between time steps of time-series or sequence data. BiLSTM networks are widely used in various applications, such as natural language processing, speech recognition, and time series prediction [27]. The mathematical equations related to BiLSTM architecture can be summarized as follows:

Forward LSTM: The forward LSTM layer processes the input sequence from left to right. The equations for the forward LSTM are as follows:

Forget Gate:

$$f_{t,f} = \sigma_g(W_{fx_t} + U_f h_{t-1,f} + b_f) \quad (7)$$

where $f_{t,f}$ is the forget gate, W_f , U_f , and b_f are the corresponding weight matrices and biases, and σ_g is the sigmoid activation function.

Input Gate:

$$i_{t,f} = \sigma_g(W_{ix_t} + U_i h_{t-1,f} + b_i) \quad (8)$$

where $i_{t,f}$ is the input gate, W_i , U_i , and b_i are the corresponding weight matrices and biases, and σ_g is the sigmoid activation function.

Output Gate:

$$o_{t,f} = \sigma_g(W_{ox_t} + U_o h_{t-1,f} + b_o) \quad (9)$$

where $o_{t,f}$ is the output gate, W_o , U_o , and b_o are the corresponding weight matrices and biases, and σ_g is the sigmoid activation function.

Cell State:

$$c_{t,f} = f_{t,f} \odot c_{t-1,f} + i_{t,f} \odot x_t \quad (10)$$

where $c_{t,f}$ is the cell state, and \odot denotes the element-wise multiplication.

Hidden State:

$$h_{t,f} = o_{t,f} \odot h_{t-1,f} + o_{t,f} \odot c_{t,f} \quad (11)$$

where $h_{t,f}$ is the hidden state, $o_{t,f}$ is the output gate, and $h_{t-1,f}$ is the previous hidden state.

Backward LSTM: The backward LSTM layer processes the input sequence from right to left. The equations for the backward LSTM are similar to the forward LSTM, but with the indices reversed (e.g., t becomes $t+1$, $h_{t-1,f}$ becomes $h_{t+1,f}$)

Combining Forward and Backward LSTM Outputs: The final output of the BiLSTM network is the weighted sum of the forward and backward LSTM outputs.

$$o_t = \alpha o_{t,f} + (1 - \alpha) o_{t,b} \quad (12)$$

where o_t is the final output, α is the weight given to the forward LSTM output, and $o_{t,f}$ and $o_{t,b}$ are the forward and backward LSTM outputs, respectively. These equations describe the flow of information through the BiLSTM network, which allows it to capture bidirectional long-term dependencies in sequential data. The BiLSTM network's ability to learn from both the past and future components of an input sequence makes it a powerful tool for various applications in natural language processing and time series prediction.

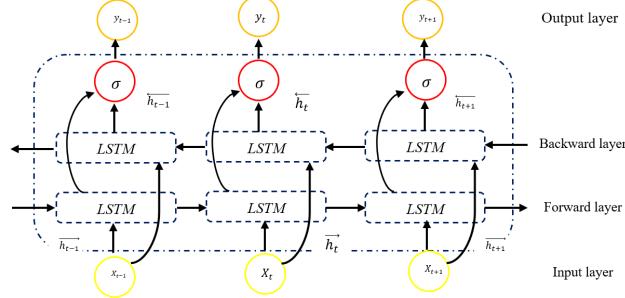


Figure 4: The architecture of basic Bi-directional LSTM (BiLSTM).

3.3 Gated recurrent unit (GRU)

A Gated Recurrent Unit (GRU) [28] network is a type of Recurrent Neural Network (RNN) that addresses the issue of long-term dependencies in sequential data by using gating mechanisms to selectively update the hidden state of the network at each time step. The mathematical equations related to GRU networks can be summarized as follows:

Update Gate: The update gate decides which information should be passed to the output and which information should be discarded. It is defined as follows:

$$z_t = \sigma_g (W_{zx_t} + U_z h_{t-1} + b_z) \quad (13)$$

where z_t is the update gate, W_z , U_z , and b_z are the corresponding weight matrices and biases, and σ_g is the sigmoid activation function.

Reset Gate: The reset gate, often referred to as the barrier gate, is responsible for resetting the hidden state to zero at specific points in the sequence. It is defined as follows:

$$r_t = \sigma_g (W_{rx_t} + U_r h_{t-1} + b_r) \quad (14)$$

where r_t is the reset gate, W_r , U_r , and b_r are the corresponding weight matrices and biases, and σ_g is the sigmoid activation function [2]. **Candidate Hidden State:** The candidate hidden state, $h_t^{(.)}$, is the intermediate memory state before the update. It is defined as follows:

$$h_t^{(.)} = \tanh (W_{hx_t} + U_h h_{t-1} + b_h) \quad (15)$$

where $h_t^{(.)}$ is the candidate hidden state, W_h , U_h , and b_h are the corresponding weight matrices and biases, and \tanh is the hyperbolic tangent activation function. The hidden state, h_t , is updated using the update gate and the candidate hidden state as follows:

$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot h_t^{(.)} \quad (16)$$

where h_t is the hidden state, and \odot denotes the element-wise multiplication. These equations describe the flow of information through a GRU network, which allows it to model sequential data effectively. GRU networks have been shown to be effective in various natural language processing tasks, such as language modelling, machine translation, and speech recognition.

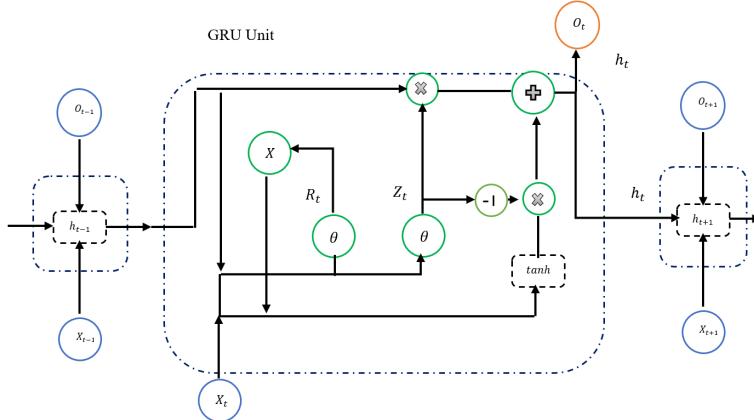


Figure 5: The architecture of basic Gated Recurrent Unit (GRU).

3.4 Self-attention model

The attention mechanism [29], a technique used in neural networks, assigns weights to different input components based on their importance and combines them to create a final representation. This allows the neural network to prioritize important input parts, leading to improved model performance. The SE Block (Squeeze and Excitation Block) is a structural unit that efficiently focuses on the channel dimension and incorporates a channel attention mechanism into the model. By assigning weights to individual feature channels, the SE Block efficiently enhances or suppresses specific channels for different tasks, enabling the extraction of valuable features. The SE attention mechanism involves several steps, including compressing two-dimensional features into real numbers using global average pooling, generating weight values for each feature channel through fully connected layers, and applying normalised weights to the channel features. These steps collectively enhance the model's predictive accuracy by adjusting the importance of different channel features.

3.5 Proposed hybrid model: CNN-BiLSTM-SA with adaptive hyper-parameter optimiser

The primary limitation of fully connected LSTM and BiLSTM when dealing with spatiotemporal data lies in its reliance on full connections for input-to-state and state-to-state transitions [30], lacking the incorporation of spatial information. Moreover, it treats all input features equally, disregarding the sequence's spatial relationships between distinct components. Furthermore, In tabular data such as wave energy converter datasets, the spatial relationships are not as clear, making it harder for CNNs to extract meaningful features. Also, wave energy converter datasets often contain diverse features with complex interactions and dependencies. CNNs may not naturally capture these inter-feature relationships as effectively as other architectures.

To address the above shortages of LSTM, BiLSTM and CNNS, various solutions were suggested [31]; however, each one suffered from generalisation and robust ability to solve different problems with various characteristics. In this way, we proposed a hybrid model that stands out by featuring inputs (X_1, \dots, X_t) , cell outputs (C_1, \dots, C_t) , hidden states (H_1, \dots, H_t) , and gates (i_t, o_t, f_t) in the hybrid CNN-BiLSTM as 3D tensors, with the last two dimensions representing spatial aspects. Through the CNN-BiLSTM, the prediction of a cell's future state within the grid hinges on its neighbouring cells' inputs and previous states, forming a cohesive spatially-aware framework. The integration of CNN and BiLSTM models combines the advantages of both architectures to tackle the spatial and temporal complexities inherent in WEC layout datasets. CNNs are adept at extracting spatial features by utilising convolutional layers, which are proficient in identifying and learning patterns within spatial data, making them well-suited for processing the spatial coordinates of WECs. On the other hand, BiLSTM networks are specifically designed to capture long-term dependencies in sequential data by processing information from both past and future contexts within a sequence, which is essential for understanding the temporal dynamics of wave energy data. By merging CNN and BiLSTM, the hybrid model is able to simultaneously capture spatial dependencies through CNN layers and temporal dependencies through BiLSTM layers, thereby ensuring a more comprehensive understanding and precise prediction of wave farm power output.

For each sequence, vector D , a convolutional operation is conducted to extract high-level features and a hyperbolic tangent linear activation function is employed. Ultimately, the output generated by the convolutional neural network is represented by a new vector $X = x_1 \oplus x_2 \oplus x_3 \oplus \dots \oplus x_t \oplus \dots \oplus x_k$, which can be expressed as:

$$X = f(\text{conv}(\mathbf{W}_c, D)) \quad (17)$$

where \mathbf{W}_c denotes the weight associated with convolution, and $f(0)$ represents the hyperbolic linear unit:

$$f(x) = \begin{cases} x, & x > 0 \\ \frac{\alpha x}{1-x}, & x \leq 0 \end{cases} \quad (18)$$

α is a hyper-parameter.

Following the execution of the CNN operation, the Bidirectional Long Short-Term Memory (BiLSTM) [26] with a hidden neuron size of l is utilized to effectively capture the high-level contextual information, which is synthesized from multiple LSTM layers, specifically the forward LSTM and the backward LSTM. Subsequently, the output of the BiLSTM is represented as a $k \times nl$ dimensional vector, with the intricate details articulated in Equations (25)–(32), wherein x_t signifies the input to the t_{th} LSTM cell, $O_{lstm_{t-1}}$ represents the output from the preceding LSTM cell; σ indicates the logistic regression function known as Sigmoid; $*$ denotes element-wise multiplication; and W_f , W_i , W_c , W_o , along with b_f , b_i , b_c , b_o , are designated as the weights and biases corresponding to the t_{th} LSTM cell.

$$\mathbf{FG}_t = \sigma(W_f [OC_{lstm_{t-1}}, x_t] + b_f), \quad (19)$$

where \mathbf{FG}_t denotes the forget gate within an LSTM cell, which possesses the capability to ascertain the specific information that necessitates elimination via x_t and $OC_{lstm_{t-1}}$.

$$ig_t = \sigma(W_i [OC_{lstm_{t-1}}, x_t] + b_i) \quad (20)$$

$$\tilde{c}_t = \tanh(\mathbf{W}_c [OC_{lstm-1}, \mathbf{x}_t] + \mathbf{b}_c), \quad (21)$$

ig_t represents the input gate of the LSTM cell, possessing the authority to determine which information ought to be modified through \mathbf{x}_t and OC_{lstm-1} ; subsequently, the candidate cell information \hat{c}_t can be derived as articulated in Equation (27).

$$ci_t = FG_t * ci_{t-1} + ig_t * \hat{c}_t \quad (22)$$

ci_{t-1} signifies the previous cell information, which subsequently influences the decision regarding the information to be discarded through the forget gate, while \hat{c}_t dictates the information to be refreshed via the input gate, culminating in the derivation of the new cell information ci_t .

$$OG_t = \sigma(\mathbf{W}_o [OC_{lstm-1}, \mathbf{x}_t] + \mathbf{b}_o), \quad (23)$$

$$OC_{lstm} = OG_t * \tanh(ci_t) \quad (24)$$

OG_t constitutes the output gate of the LSTM cell, which is capable of evaluating the conditions pertinent to the output based on \mathbf{x}_t and OC_{lstm} , thereby allowing for the generation of the LSTM cell output OC_{lstm} through the output gate in conjunction with the cell information. Ultimately, the resultant output of BiLSTM O can be denoted as:

$$O = o_1 \oplus o_2 \oplus \dots \oplus o_t \oplus \dots \oplus o_k \quad (25)$$

$$o_t = [OC_{lstm}^f, OC_{lstm}^b] \quad (26)$$

where OC_{lstm}^f represents the output generated by the $t\mathbf{h}$ cell within the forward Long Short-Term Memory (LSTM) architecture, and OC_{lstm}^b signifies the output produced by the $t\mathbf{h}$ cell within the backward LSTM framework.

These equations describe the flow of information through the BiLSTM network, which allows it to capture bidirectional long-term dependencies in sequential data.

Furthermore, the self-attention mechanism [29] is employed to allocate varying weights to the output features generated by the BiLSTM, with the resultant output being regarded as the single-channel coding features. This allows the BiLSTM to prioritize important input parts, leading to improved model performance. The SE Block (Squeeze and Excitation Block) is a structural unit that efficiently focuses on the channel dimension and incorporates a channel attention mechanism into the model. By assigning weights to individual feature channels, the SE Block efficiently enhances or suppresses specific channels for different tasks, enabling the extraction of valuable features. The specifics are enumerated as follows:

$$F = \tanh(S_k O^T) \quad (27)$$

$$\begin{aligned} A &= \text{softmax}(S_a F), \\ O_{AL} &= AO \end{aligned} \quad (28)$$

where O^T represents the transposed matrix of O , O_{AL} indicates the resultant output of the attention layer, and S_k , S_a are parameters that are subject to training. The preceding exposition has elaborated on the conventional architecture of a single channel.

The incorporation of self-attention with CNN and BiLSTM layers is anticipated to yield several significant enhancements, including heightened predictive accuracy, improved generalisation across diverse sea sites, and more efficient training through adaptive hyper-parameter tuning. The advantages of the combination of these methods can be listed as follows:

- The incorporation of CNN and BiLSTM enables the model to comprehend both spatial and temporal relationships, resulting in notably improved prediction accuracy in comparison to individual models.
- Adaptability and Versatility: The model's structure can be readily adjusted to different wave farm setups and datasets, making it a versatile tool for predicting wave energy.

The proposed hybrid CNN and BiLSTM model, enriched with a self-attention mechanism and adaptive hyper-parameter optimization, signifies notable progress in forecasting the power output of wave farms. By integrating spatial and temporal feature extraction capabilities with targeted attention on significant data points, this model addresses the challenges associated with wave energy power prediction, thus facilitating more efficient integration of wave energy into the power grid. Figure 6 demonstrates a schematic of various components of the proposed hybrid model.

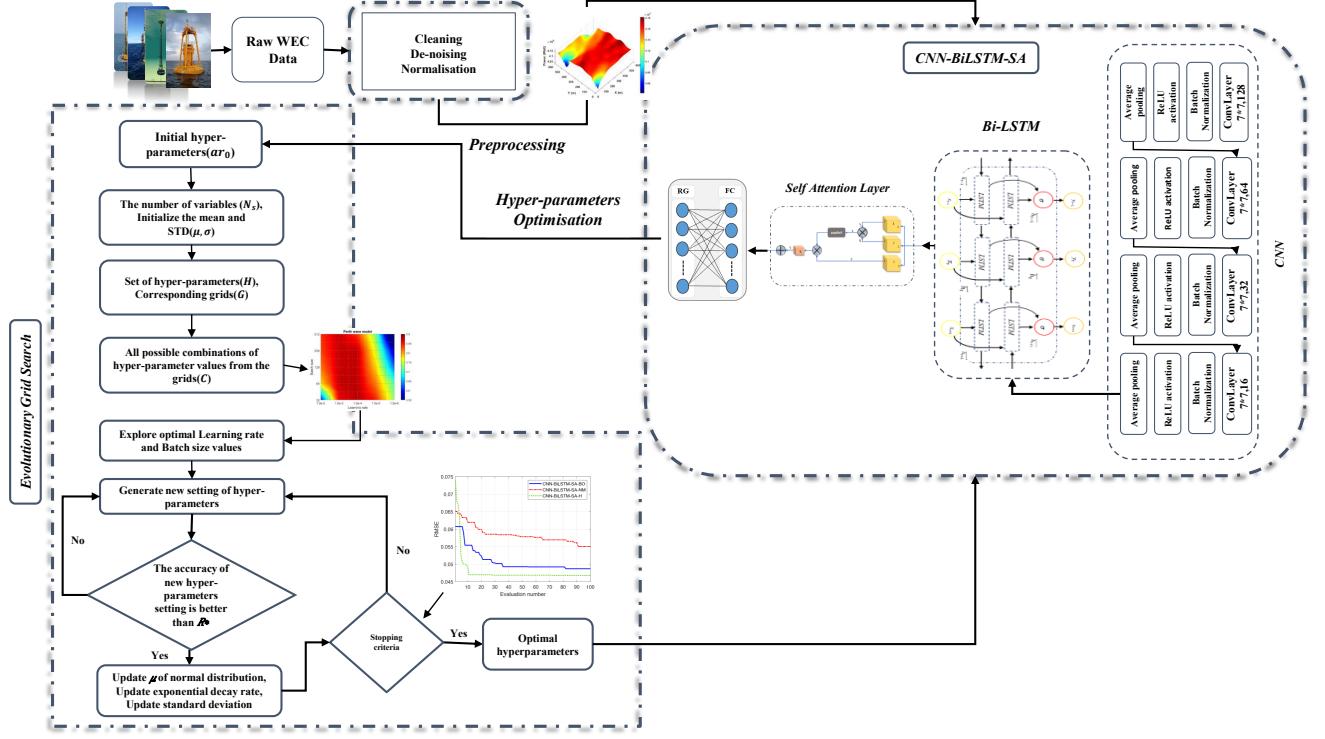


Figure 6: The proposed hybrid deep model: CNN-BiLSTM-SA with hyper-parameters tuning method (EGS)

3.5.1 Tuning the Hyper-parameters of deep neural networks

Hyper-parameters refer to a collection of parameters utilized in the training and testing stages of the learning process. When training a deep learning model, it is essential for the model to learn various instances and weights that correspond to different feature combinations as patterns [32]. Common hyper-parameters consist of the learning rate, number of iterations, hidden layers, batch size, activation functions, momentum, and regularization. In the context of convolutional neural networks used for image classification, factors such as the field size of convolutional and pooling layers and the stride parameter for handling step size are taken into account. These parameters can be represented as integers or continuous or categorical variables within specified lower and upper bounds. Intermediate layers may introduce additional hyper-parameters, leading to potential variation in the number of neurons across layers.

During the training process, it is essential to select the optimal hyper-parameters to effectively train the model, as this can significantly impact its performance. Inadequate learning rates may inadvertently omit crucial patterns. The combination of parameters should be carefully chosen to minimize the loss function or maximize the model's performance and accuracy. Tuning hyper-parameters can be viewed as an optimization problem akin to fuzzy optimization [33, 34] applied in raw data. The hyperparameters of a deep learning model remain constant throughout training, enhancing model accuracy while also considering memory and time constraints to reduce the loss function value. The selection of hyper-parameters is specific to each model and is determined by the problem being addressed. There is no universally optimal set of hyper-parameters applicable to all models [35]. This study has focused on the learning rate, momentum, regularization, and network depth. The learning rate aids in recognizing general patterns in images, while momentum assists in thoroughly exploring the entire search space without overlooking important points. Regularization supports the model in achieving better generalization, thereby preventing overfitting and enhancing prediction accuracy.

3.5.2 Bayesian optimization

Deep learning optimization involves solving black-box optimization problems where the objective function, denoted as $f(y)$, is not explicitly known. It is crucial to minimize the number of samples used across the layers. In cases where human expertise cannot significantly contribute to improving accuracy, Bayesian optimization proves to be highly beneficial. This approach incorporates prior information about the function f and continually updates posterior information, leading to reduced loss and enhanced model accuracy and has successful application in optimising renewable energy systems [36].

Bayesian optimization efficiently uncovers the global optima of the black box function within neural networks. It effectively handles noisy data and non-continuous spaces to achieve global minima while making optimal use of available resources.

The optimization process relies on Bayes' Theorem as described by [37] in the equation that pertains to a model

Z and an observation Y .

$$P(Z | Y) = (P(Y | Z)P(Z))/P(Z) \quad (29)$$

Bayesian optimization involves finding the minimum of a function $f(y)$ within a bounded set Y , using the posterior probability $P(Z | Y)$, likelihood $P(Y | Z)$, prior probability $P(Z)$, and marginal probability $P(Z)$.

Although Bayesian optimization presents a realm of efficient exploration, reduced function evaluations, and adeptness in navigating through noisy functions and diverse objectives, it can pose a challenge with its computational intricacy, susceptibility to hyper-parameters, restrictions in scaling up in high-dimensional realms, and susceptibility to biases in initialization. In this paper, the Root Mean Square Error (RMSE) was used as an objective function for the Bayesian optimization algorithm to optimize the proposed model's hyper-parameters.

3.5.3 Nelder-mead optimisation

The pioneering Nelder-Mead approach is widely acclaimed as the foremost direct search technique for optimising unconstrained objective functions. This strategy involves assessing function values at the $n + 1$ vertices x_i of a simplex, where n indicates the problem's dimensionality. The simplex, initially positioned at x_0 randomly, adheres to a particular guideline for determining its magnitude, as stipulated by [38]. By continuously comparing function values and adjusting the simplex, the Nelder-Mead method strives to move closer to the problem's optimal solution [39].

$$x_i = x_0 + pe_i + \sum_{\substack{k=1 \\ k \neq i}}^n q e_k, \quad i = 1, n := \begin{cases} p = \frac{a}{n\sqrt{2}}(\sqrt{n+1} + n - 1) \\ q = \frac{a}{n\sqrt{2}}(\sqrt{n+1} - 1) \end{cases} \quad (30)$$

where e_i are the unit base vectors. Throughout the implementation of the NM technique, the simplex's vertices are consistently altered through *reflection*, *expansion*, and *contraction* manoeuvres. These manoeuvres are designed to pinpoint a better configuration that steers the algorithm towards the most favourable solution. The sequence persists until the values of the simplex vertices reach a level of similarity that signals convergence, which is assessed through an inequality standard as follows.

$$\sqrt{\sum_{i=1}^{n+1} (f_i - \bar{f})^2 / n} < \varepsilon, \bar{f} = \frac{1}{n+1} \sum_{i=1}^{n+1} f_i \quad (31)$$

where ε dances as a tiny optimistic scalar, the combined impact of the actions performed on the simplex can be likened to elongating the form in the paths of descent and swiftly navigating around nearby peaks. In a nutshell, the NM optimisation technique provides a blend of simplicity, resilience, and adaptability for enhancing various functions. Nevertheless, it could face challenges such as sluggish convergence, absence of assured global convergence, vulnerability to initial conditions, and restricted scalability in complex spaces with many dimensions.

3.5.4 Evolutionary Grid Search

Widely adopted strategies like local and global optimisation methods involve tuning hyper-parameters and architectures to enhance the performance of predictive ML and deep learning models. This task proves challenging and time-consuming due to the vast number of possible variable combinations. A solution set, denoted as X , is defined alongside a real-valued function, f , mapping X to real numbers. The objective is to explore an optimal solution, x^* , where $x^* \in X$ with $f(x^*) \leq f(x), \forall x \in X$. As these combinatorial problems grow exponentially with expanded variables, a rapid and efficient optimisation method becomes necessary to converge towards a suitable solution while minimising the number of evaluations required. However, local search methods encounter the drawback of being susceptible to local optima, restricting their ability to explore substantial portions of the search space and hindering their capability to discover the global optimum in multimodal problems. Conversely, global search methods, such as evolutionary algorithms, exhibit slow convergence due to the necessity for numerous evaluations to attain a satisfactory solution.

In this paper, a fast and adaptive Evolutionary Grid Search (EGS) algorithm is proposed to develop the performance of the proposed hybrid forecasting model by tuning the hyperparameters. EGS is a combination of a grid search with an iterative local search (1+1EA) with an adaptive exploration step size in which a local optimum is in the neighbourhood of the assembled solution. EGS has three steps, including *i*) grid search, *ii*) exploration, and *iii*) exploitation phase. In the first step, in order to speed up the optimisation process, we used a grid search algorithm and specified a set of ranges for both learning rate and batch size hyper-parameters, which play a significant role in training deep learning models. After finding the optimal configuration of the learning rate and batch size, they are kept fixed. Then, we selected another pair of hyper-parameters, such as the number of neurons and filters, and ran an evolutionary local search. This procedure is followed to ensure all parameters are tuned. The technical details of the EGS Algorithm 1 can be seen in the following.

4 Experimental results

The results obtained from the previously mentioned methods on our datasets are discussed in this section.

Algorithm 1 Evolutionary Grid Search(EGS)

```

1: procedure EGS ( $solution_0 = h_1, h_2, \dots, h_n$  )
2: Initialisation
3:    $ar_0 = Eval(solution_0)$                                       $\triangleright$  Evaluate model by initial hyper-parameters
4:    $N_s = len(solution_0)$                                       $\triangleright$  The number of variables
5:    $\mu = array, \sigma = \sigma_1, \sigma_2, \dots, \sigma_{N_s}, \lambda = -0.04$   $\triangleright$  Initialise the mean and STD
6:   Grid Search
7:    $H = \{h_1, h_2, \dots, h_{N_s}\}$                                       $\triangleright$  Set of hyper-parameters
8:    $G = \{G_1, G_2, \dots, G_{N_g}\}$                                       $\triangleright$  Corresponding grids
9:    $C = (c_1, c_2, \dots, c_{N_c}) \in G_1 \times G_2 \times \dots \times G_{N_g}$   $\triangleright$  All possible combinations of hyper-parameter values from the grids
10:  for  $i$  in  $[1, \dots, N_{lr}]$  do                                          $\triangleright$  Explore optimal Learning rate and Batch size values
11:    for  $j$  in  $[1, \dots, N_{bs}]$  do
12:       $R(c_{i,j}) = Train(Model(c_{i,j}))$ 
13:    end for
14:  end for
15:   $< c^*, R^* > = argmax(R(c_{i,j})), where(c_{i,j}) \in C$             $\triangleright$  Best-performed configuration of the Grid search
16:  Adaptive 1+1EA
17:   $\mu = c^*$ 
18:  while  $t \neq stopping criteria$  do
19:    for  $i$  in  $[1, \dots, N_s]$  do                                          $\triangleright$  Generate new setting of hyper-parameters
20:      if  $rand \leq \frac{2}{N_s}$  then  $solution_i = \mathcal{N}(\mu_i, \sigma_i^2)$ 
21:      end if
22:    end for
23:     $\langle R_t \rangle = Train(Model(solution))$                                 $\triangleright$  The accuracy of new hyper-parameters setting is better than  $R^*$ 
24:    if  $R_t > R^*$  then                                          $\triangleright$  Update  $\mu$  of normal distribution
25:       $\mu = solution$                                                   $\triangleright$  Update exponential decay rate
26:       $\lambda = \lambda - r_d$                                               $\triangleright$  Update standard deviation
27:       $\sigma = \sigma_0 \times e^{-\lambda t}$ 
28:       $R^* = R_t$ 
29:    end if
30:  end while                                                        $\triangleright$  Optimal configuration
31: return  $solution, R^*$ 
32: end procedure

```

4.1 Datasets and analysis

In order to execute statistical modelling, a considerable amount of data needs to be available for analysis [40], and to collect and publish the current model wave energy dataset. This dataset was obtained from UCI's machine learning library. The data set offers information on the positioning of WEC along Australia's southern coast in four natural wave cycles (Sydney, Adelaide, Perth, and Tasmania). The converter model used is a completely submerged three-tether converter known as CETO. In a space-constrained context, the 16 WEC spots are positioned and optimized. The first 32 variables of the data set are the positions (X coordinates and Y coordinates) of the 16 WEC on a continuous scale ranging from 0 to 566 (m) of the wave farm. The following 16 attributes represent the absorbed power of each Wave Energy Converter (WEC) in the wave farm. The final attribute is the total power output from the wave farm, which is calculated as the sum of all 16 absorbed powers. The placement of WECs in relation to their generated power is used to predict the wave farm's power output. Each dataset consists of 72000 different combinations of WEC placements and their associated absorption powers. To assess the predictive model's accuracy, we split the dataset into 70% test data and 30% validation data. Moreover, to construct the predictive model, the data is normalized and scaled.

4.2 Wave farm landscape analysis

In the landscape analysis that was conducted, the focus was on exploring the relationship between the position adjustments of individual WECs and their impact on the total power output in the wave farms located in Adelaide and Sydney sea sites. Two specific layouts were selected for evaluation: one with four WECs and another with eight WECs. The objective was to understand how modifying the position of a single WEC while keeping the remaining WECs fixed would influence the overall power production. To carry out the experiments, a systematic approach was employed. The positions of the $N - 1$ buoys (B_1, B_2, \dots, B_{N-1}) were kept constant, representing the fixed layout, while the position of the last buoy (B_N) was iteratively adjusted. The separation distance between B_N and the other buoys within the wave farm was varied across a range of distances, starting from a minimum of 50 m (refer to safe distance) and extending to all feasible locations. This strategy allowed for a comprehensive analysis of the impact of relative positions on power output.

The results of the experiments, depicted in Figure 7, showcased the relationship between the total power output production and the relative positions of all buoys and wave interference in the wave farms. The analysis focused on two distinct wave climates, namely Adelaide and Sydney, providing insights into how the variations in wave characteristics in different locations can influence the power generation efficiency of the wave farms. Interestingly, the

power landscapes observed in the wave farms of Sydney and Adelaide exhibit distinct constructive and destructive wave interaction patterns. In the case of the four and eight WEC layouts in Adelaide, constructive wave interactions manifest as an unimodal region characterised by two edges that possess the highest potential for wave power extraction (see Figure 7 a-d). Besides, the absorbed power from a layout situated in Adelaide attains its peak value at a relatively short separation distance and a specific angle between the converters. These findings highlight the intricate relationship between wave interactions and power generation efficiency, emphasising the importance of carefully considering wave energy converters' spatial arrangement and orientation in optimising wave farm performance. The main difference between the analysis for the Adelaide and Sydney sites can be explained by differences in the directionality of the wave climates. The Adelaide site has narrowly spread seas, resulting in a clear directional preference for the potential installation of the next WEC for both array configurations. On the contrary, the Sydney site has wide directional spreading of the incoming waves, and there is no clear directional pattern for the last buoy installation.

It is interesting to acknowledge that the influence of separation distance on power generation varies depending on wave climates. For the Sydney site, an increase in the separation distance positively affects the power generation efficiency. This behaviour can be attributed to the complex nature of wave directions observed in Sydney, leading to a multifaceted power landscape depicted in Figure 7 (e-h). The power landscape exhibits a multimodal characteristic, shaped by the adjustments in the position of the last WEC in both the four and eight-WEC layouts.

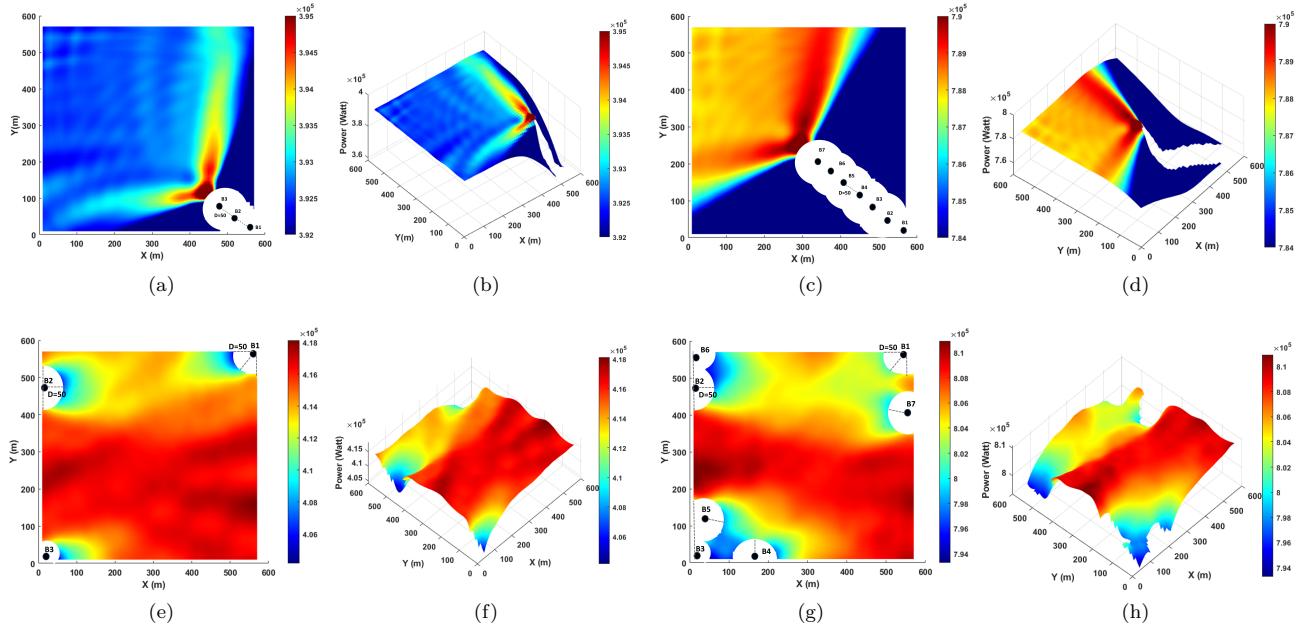


Figure 7: Total power output Landscape analysis of the farm best-configuration based on Adelaide and Sydney sea site. The evaluation power of the fourth WEC through a whole feasible wave farm location with the initial three fixed WECs can be seen in (a and b) in the Adelaide site and (e and f) in the Sydney site. Various power potentials of the eighth WEC with the initial seven fixed WECs are shown by (c and d) in Adelaide and (g and h) based on the Sydney sea site. The black circles show the WECs with a safe distance equal to 50 m.

4.3 Results and Finding

In this study, nine distinct long-short-term memory (*LSTM*) models and a CNN were constructed, and their outputs were compared. Prior to the model's training, the number of input dimensions was adjusted to 32, and the total power value was predicted. Various parameter alterations were made to construct a model with a high degree of accuracy and to combat the overfitting problem. The data was scaled and normalized using the Min-Max formula before training, which can be seen in the following.

$$x_{norm}^i = \frac{x^i - \mathbf{x}_{\min}}{\mathbf{x}_{\max} - \mathbf{x}_{\min}} \quad (32)$$

We used eight evaluation metrics, including The Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Loss function, Mean absolute error (MAE), Coefficient of Determination (R^2), Mean Squared Logarithmic Error (MSLE), Median Absolute Error (MedAE), and Max_{Error} values of the training and test data were analyzed to assess the goodness-of-fit of the model. The technical details of these metrics can be seen in Table 1.

Table 2 presents the statistical results of two different LSTM models: a one-layer LSTM (Vanilla) model and a stacked model with two layers. To ensure a comprehensive evaluation, the average, minimum, maximum, and standard deviation of predicted power outputs were considered. These results were obtained through a 10-fold cross-validation

Table 1: Performance metrics and formulations.

Parameters	Formula
Mean Squared Error (MSE)	$MSE = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y})^2$, Where, \hat{y} is the predicted value of y
Root Mean Squared Error (RMSE)	$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y})^2}$
LOSS	$Loss = \frac{1}{2N} \sum_{i=1}^N (y_i - \hat{y})^2$
Mean absolute error (MAE)	$MAE = \frac{1}{N} \sum_{i=1}^N y_i - \hat{y} $
Coefficient of Determination (R^2)	$R^2 = 1 - \frac{\sum(y_i - \bar{y})^2}{\sum(y_i - \hat{y})^2}$, Where, \bar{y} is the mean value of y
Mean Squared Logarithmic Error (MSLE)	$MSLE = \frac{1}{N} \sum_{i=1}^N ((\log(1 + y_i)) - (\log(1 + \hat{y})))^2$
Median Absolute Error (MedAE)	$MedAE = median(y - \hat{y})$
MAX_{ERROR}	$Max_{Error} = \frac{1}{N} \sum_{i=1}^N \hat{y} - y_i $

method, enhancing the accuracy of the comparison. An important observation is that increasing the number of LSTM layers does not guarantee superior performance in predicting wave farm power output across all wave scenarios. For instance, when examining the Tasmania dataset, the stacked LSTM model failed to outperform the LSTM model in terms of R^2 , with the LSTM outperforming it by 22%. However, in other wave farms, such as Adelaide, Perth, and Sydney, the stacked LSTM model exhibited better performance than the LSTM model by 8.3%, 0.3%, and 8.65%, respectively.

Table 2: Statistical results of total power output prediction for Vanilla-LSTM and Stacked-LSTM models based on four wave scenario

Stacked-LSTM (2 layers)							Vanilla-LSTM								
Adelaide							Adelaide								
Metric	RMSE	LOSS	MAE	R2	MSLE	MEDAE	Max-error	Metric	RMSE	LOSS	MAE	R2	MSLE	MEDAE	Max-error
Mean	1.16E-01	6.73E-03	9.27E-02	5.49E-01	5.82E-03	7.95E-02	4.92E-01	Mean	1.20E-01	7.28E-03	9.64E-02	5.07E-01	6.28E-03	8.19E-02	5.02E-01
Min	9.83E-02	4.84E-03	7.89E-02	5.23E-01	4.08E-03	6.76E-02	4.62E-01	Min	1.02E-01	5.23E-03	8.15E-02	4.60E-01	4.40E-03	6.92E-02	4.62E-01
Max	1.28E-01	8.14E-03	1.03E-01	5.88E-01	7.19E-03	8.88E-02	5.24E-01	Max	1.32E-01	8.71E-03	1.05E-01	5.87E-01	7.72E-03	8.95E-02	5.40E-01
STD	8.36E-03	9.49E-04	6.73E-03	2.30E-02	1.02E-03	6.28E-03	1.79E-02	STD	8.00E-03	9.34E-04	6.56E-03	3.44E-02	9.98E-04	6.53E-03	2.49E-02
Perth							Perth								
Mean	1.26E-01	7.91E-03	9.94E-02	3.51E-01	7.01E-03	8.30E-02	5.27E-01	Mean	1.26E-01	7.97E-03	1.00E-01	3.50E-01	7.07E-03	8.41E-02	5.36E-01
Min	1.16E-01	6.72E-03	9.13E-02	2.49E-01	5.38E-03	7.49E-02	4.96E-01	Min	1.16E-01	6.77E-03	9.14E-02	3.05E-01	5.42E-03	7.44E-02	5.11E-01
Max	1.42E-01	1.00E-02	1.13E-01	4.42E-01	9.25E-03	9.74E-02	5.62E-01	Max	1.38E-01	9.57E-03	1.11E-01	3.90E-01	8.87E-03	9.44E-02	5.58E-01
STD	7.65E-03	9.86E-04	6.64E-03	6.66E-02	1.08E-03	6.84E-03	2.08E-02	STD	7.16E-03	9.08E-04	6.11E-03	3.27E-02	1.06E-03	6.30E-03	1.40E-02
Tasmania							Tasmania								
Mean	1.16E-01	6.70E-03	9.02E-02	2.05E-01	6.14E-03	7.43E-02	5.13E-01	Mean	1.13E-01	6.40E-03	8.79E-02	2.50E-01	5.88E-03	7.21E-02	5.14E-01
Min	1.03E-01	5.32E-03	8.12E-02	1.83E-01	4.55E-03	6.77E-02	4.91E-01	Min	1.01E-01	5.05E-03	7.90E-02	2.25E-01	4.33E-01	6.43E-02	5.01E-01
Max	1.28E-01	8.13E-03	9.96E-02	2.45E-01	7.90E-03	8.22E-02	5.38E-01	Max	1.25E-01	7.83E-03	9.73E-02	2.70E-01	7.63E-03	8.04E-02	5.22E-01
STD	7.45E-03	8.62E-04	5.89E-03	1.83E-02	1.03E-03	4.88E-03	1.46E-02	STD	7.70E-03	8.66E-04	6.04E-03	1.62E-02	1.02E-03	4.92E-03	6.38E-03
Sydney							Sydney								
Mean	1.32E-01	8.75E-03	1.04E-01	2.89E-01	6.81E-03	8.75E-02	6.59E-01	Mean	1.34E-01	8.98E-03	1.05E-01	2.66E-01	6.98E-03	9.01E-02	6.63E-01
Min	1.15E-01	6.61E-03	9.24E-02	2.22E-01	4.74E-03	7.99E-02	6.14E-01	Min	1.16E-01	6.70E-03	9.31E-02	2.23E-01	4.80E-03	8.25E-02	6.15E-01
Max	1.51E-01	1.14E-02	1.17E-01	3.99E-01	9.33E-03	9.54E-02	7.11E-01	Max	1.51E-01	1.14E-02	1.17E-01	3.19E-01	9.35E-03	9.59E-02	7.25E-01
STD	1.18E-02	1.56E-03	8.43E-03	5.70E-02	1.46E-03	6.23E-03	3.05E-02	STD	1.10E-02	1.46E-03	7.52E-03	3.17E-02	1.40E-03	4.09E-03	3.30E-02

One of the most popular sequential deep learning models is Bidirectional LSTM (BiLSTM) due to its ability to handle long-term dependencies better than traditional LSTM. By capturing information from both the past and future, BiLSTM provides an improved understanding of context. Therefore, we conducted a comparison between BiLSTM and its stacked version with two layers in predicting the power output at four sea sites. The results are presented in Table ???. The stacked BiLSTM model outperformed the standard BiLSTM model among the sea sites examined, except for the Tasmania wave model. Specifically, in the Adelaide, Perth, and Sydney sea sites, the stacked BiLSTM model performed better by 3.29%, 4.10%, and 10.38%, respectively.

In the evaluation of wave farm power prediction models, the Gated Recurrent Unit (GRU) was tested using the following machine learning model. GRU incorporates a simplification of the LSTM architecture by combining the input and forgets gates into a single update gate. This modification reduces computational complexity and enhances the training speed of GRU models. The statistical analysis presented in Table 3 compares the performance of both GRU and stacked versions' performance using eight metrics. While the stacked model demonstrated greater accuracy in terms of the coefficient of determination (R^2), interestingly, similar to the LSTM and Bi-LSTM models, a standalone GRU model produced more precise results compared to the stacked GRU model in the Tasmania wave model.

Upon conducting tests and comparisons of LSTM, Bi-LSTM, and GRU models alongside their stacked versions, our focus shifted towards Convolutional deep learning models. These models employ convolutional layers and pooling operations, enabling them to effectively capture spatial or temporal patterns inherent in the data. To enhance the capabilities of CNNs, we devised a hybrid model that integrates both CNN and LSTM architectures. The rationale behind this approach stems from the fact that combining CNNs with LSTMs allows for further processing and extraction of hierarchical representations acquired by the CNN layers, transforming them into higher-level features through the LSTM layers. Consequently, this amalgamation enables the model to leverage the local pattern recognition abilities of CNNs as well as the sequential modelling capabilities of LSTMs. Table 4 presents the prediction outcomes of both the CNN and the hybrid CNN-LSTM model for four distinct wave case studies. Notably, the CNN exhibits a

Table 3: Statistical results of total power output prediction for Vanilla-GRU and Stacked-GRU models based on four wave scenario

Stacked-GRU							Vanilla-GRU								
Adelaide							Adelaide								
Metric	RMSE	LOSS	MAE	R2	MSLE	MEDAE	Max-error	Metric	RMSE	LOSS	MAE	R2	MSLE	MEDAE	Max-error
Mean	1.12E-01	6.27E-03	8.94E-02	5.93E-01	5.46E-03	7.69E-02	4.76E-01	Mean	1.23E-01	7.62E-03	9.87E-02	4.88E-01	6.55E-03	8.43E-02	5.13E-01
Min	9.34E-02	4.36E-03	7.49E-02	5.80E-01	3.71E-03	6.49E-02	4.52E-01	Min	1.11E-01	6.15E-03	8.84E-02	4.45E-01	5.12E-03	7.48E-02	4.88E-01
Max	1.26E-01	7.92E-03	1.01E-01	6.12E-01	7.01E-03	8.72E-02	4.95E-01	Max	1.36E-01	9.28E-03	1.09E-01	5.36E-01	8.22E-03	9.49E-02	5.53E-01
STD	9.40E-03	1.04E-03	7.48E-03	9.90E-03	1.08E-03	6.63E-03	1.64E-02	STD	8.42E-03	1.05E-03	6.92E-03	2.94E-02	1.11E-03	7.23E-03	1.82E-02
Perth							Perth								
Metric	RMSE	LOSS	MAE	R2	MSLE	MEDAE	Max-error	Metric	RMSE	LOSS	MAE	R2	MSLE	MEDAE	Max-error
Mean	1.18E-01	6.99E-03	9.33E-02	4.64E-01	6.27E-03	7.84E-02	5.09E-01	Mean	1.24E-01	7.72E-03	9.84E-02	3.82E-01	6.87E-03	8.25E-02	5.26E-01
Min	1.09E-01	5.99E-03	8.65E-02	3.96E-01	4.83E-03	7.07E-02	4.81E-01	Min	1.15E-01	6.64E-03	9.02E-02	3.15E-01	5.31E-03	7.30E-02	5.03E-01
Max	1.29E-01	8.28E-03	1.03E-01	5.40E-01	7.79E-03	8.85E-02	5.34E-01	Max	1.34E-01	9.01E-03	1.07E-01	4.38E-01	8.40E-03	9.31E-02	5.44E-01
STD	6.67E-03	7.96E-04	5.61E-03	4.07E-02	9.31E-04	5.80E-03	1.51E-02	STD	6.38E-03	7.97E-04	5.56E-03	4.07E-02	9.49E-04	6.16E-03	1.33E-02
Tasmania							Tasmania								
Metric	RMSE	LOSS	MAE	R2	MSLE	MEDAE	Max-error	Metric	RMSE	LOSS	MAE	R2	MSLE	MEDAE	Max-error
Mean	1.13E-01	6.44E-03	8.81E-02	2.47E-01	5.92E-03	7.20E-02	5.06E-01	Mean	1.12E-01	6.24E-03	8.67E-02	2.75E-01	5.74E-03	7.09E-02	5.03E-01
Min	9.99E-02	4.99E-03	7.82E-02	1.57E-01	4.27E-03	6.45E-02	4.73E-01	Min	9.96E-02	4.96E-03	7.82E-02	2.59E-01	4.25E-03	6.41E-02	4.91E-01
Max	1.29E-01	8.32E-03	1.01E-01	2.94E-01	8.06E-03	8.36E-02	5.39E-01	Max	1.24E-01	7.64E-03	9.62E-02	3.02E-01	7.45E-03	7.90E-02	5.22E-01
STD	8.33E-03	9.53E-04	6.68E-03	3.74E-02	1.10E-03	5.61E-03	2.11E-02	STD	7.39E-03	8.24E-04	5.82E-03	1.52E-02	9.89E-04	4.85E-03	8.34E-03
Sydney							Sydney								
Metric	RMSE	LOSS	MAE	R2	MSLE	MEDAE	Max-error	Metric	RMSE	LOSS	MAE	R2	MSLE	MEDAE	Max-error
Mean	1.28E-01	8.34E-03	1.00E-01	3.45E-01	6.53E-03	8.41E-02	6.41E-01	Mean	1.32E-01	8.73E-03	1.04E-01	2.96E-01	6.81E-03	8.85E-02	6.60E-01
Min	1.08E-01	5.82E-03	8.66E-02	1.35E-01	4.20E-03	7.58E-02	6.00E-01	Min	1.16E-01	6.68E-03	9.21E-02	2.70E-01	4.79E-03	8.04E-02	6.21E-01
Max	1.63E-01	1.32E-02	1.27E-01	4.13E-01	1.07E-02	1.08E-01	6.81E-01	Max	1.47E-01	1.08E-02	1.13E-01	3.39E-01	8.86E-03	9.50E-02	7.12E-01
STD	1.56E-02	2.10E-03	1.17E-02	8.26E-02	1.85E-03	9.14E-03	2.89E-02	STD	1.14E-02	1.48E-03	7.99E-03	2.47E-02	1.41E-03	4.79E-03	2.80E-02

Table 4: Statistical results of total power output prediction for CNN and CNN-LSTM models based on four wave scenario

CNN Adelaide								CNN-LSTM Adelaide							
Metric	RMSE	LOSS	MAE	R2	MSLE	MEDAE	Max-error	Metric	RMSE	LOSS	MAE	R2	MSLE	MEDAE	Max-error
Mean	8.83E-02	3.93E-03	6.72E-02	6.97E-01	3.59E-03	5.16E-02	4.61E-01	Mean	7.44E-02	2.78E-03	5.48E-02	7.97E-01	2.68E-03	4.03E-02	4.31E-01
Min	7.60E-02	2.89E-03	5.64E-02	6.46E-01	2.64E-03	4.05E-02	4.19E-01	Min	6.63E-02	2.20E-03	4.76E-02	7.80E-01	2.08E-03	3.34E-02	3.58E-01
Max	1.01E-01	5.13E-03	7.76E-02	7.38E-01	4.97E-03	6.34E-02	5.31E-01	Max	8.20E-02	3.36E-03	6.02E-02	8.28E-01	3.35E-03	5.07E-02	4.96E-01
STD	8.27E-03	7.33E-04	3.22E-02	7.19E-04	7.46E-03	3.70E-02	STD	5.18E-03	3.86E-04	4.62E-03	1.35E-02	4.25E-04	5.82E-03	4.21E-02	
Perth								Perth							
Metric	RMSE	LOSS	MAE	R2	MSLE	MEDAE	Max-error	Metric	RMSE	LOSS	MAE	R2	MSLE	MEDAE	Max-error
Mean	8.83E-02	3.90E-03	6.46E-02	6.59E-01	3.71E-03	4.66E-02	4.95E-01	Mean	7.45E-02	2.78E-03	5.46E-02	7.62E-01	2.67E-03	3.94E-02	4.20E-01
Min	8.36E-02	3.49E-03	6.06E-02	6.25E-01	3.18E-03	4.26E-02	4.05E-01	Min	7.03E-02	2.47E-03	5.33E-02	7.53E-01	2.24E-03	3.80E-02	3.82E-01
Max	9.67E-02	4.68E-03	7.08E-02	7.06E-01	4.57E-03	5.11E-02	5.94E-01	Max	7.90E-02	3.12E-03	5.80E-02	7.73E-01	3.05E-03	4.22E-02	4.69E-01
STD	4.41E-03	3.93E-04	3.22E-03	2.76E-02	4.21E-04	2.65E-03	5.62E-02	STD	2.63E-03	1.97E-04	1.76E-03	5.90E-03	2.32E-04	1.46E-03	3.12E-02
Tasmania								Tasmania							
Metric	RMSE	LOSS	MAE	R2	MSLE	MEDAE	Max-error	Metric	RMSE	LOSS	MAE	R2	MSLE	MEDAE	Max-error
Mean	8.63E-02	3.74E-03	6.56E-02	5.21E-01	3.54E-03	5.05E-02	4.66E-01	Mean	7.46E-02	2.80E-03	5.69E-02	6.73E-01	2.71E-03	4.42E-02	3.75E-01
Min	7.82E-02	3.05E-03	5.96E-02	4.67E-01	2.81E-03	4.56E-02	3.63E-01	Min	6.70E-02	2.25E-03	5.11E-02	6.59E-01	2.05E-03	3.87E-02	3.35E-01
Max	1.01E-01	5.14E-03	7.67E-02	5.80E-01	5.09E-03	5.91E-02	5.45E-01	Max	8.68E-02	3.76E-03	6.62E-02	7.04E-01	3.78E-03	5.17E-02	4.19E-01
STD	6.41E-03	5.77E-04	4.89E-03	3.54E-02	6.25E-04	4.37E-03	5.81E-02	STD	6.92E-03	5.33E-04	5.27E-03	1.28E-02	5.96E-04	4.25E-03	3.18E-02
Sydney								Sydney							
Metric	RMSE	LOSS	MAE	R2	MSLE	MEDAE	Max-error	Metric	RMSE	LOSS	MAE	R2	MSLE	MEDAE	Max-error
Mean	8.31E-02	3.50E-03	5.87E-02	7.28E-01	2.80E-03	3.95E-02	5.58E-01	Mean	6.40E-02	2.12E-03	4.59E-02	8.50E-01	1.72E-03	3.22E-02	4.00E-01
Min	6.46E-02	2.08E-03	4.53E-02	6.81E-01	1.63E-03	2.99E-02	4.19E-01	Min	5.23E-02	1.37E-03	3.83E-02	8.45E-01	1.04E-03	2.61E-02	3.06E-01
Max	9.38E-02	4.40E-03	6.64E-02	7.87E-01	3.60E-03	4.57E-02	9.04E-01	Max	8.30E-02	3.44E-03	6.01E-02	8.54E-01	2.89E-03	4.27E-02	4.74E-01
STD	9.72E-03	7.83E-04	7.20E-03	3.19E-02	6.94E-04	5.49E-03	1.35E-01	STD	1.21E-02	8.09E-04	8.78E-03	3.09E-03	7.18E-04	6.49E-03	6.06E-02

Table 5: Statistical results of total power output prediction for CNN-BiLSTM and CNN-GRU models based on four wave scenario

CNN-BiLSTM Adelaide								CNN-GRU Adelaide							
Metric	RMSE	LOSS	MAE	R2	MSLE	MEDAE	Max-error	Metric	RMSE	LOSS	MAE	R2	MSLE	MEDAE	Max-error
Mean	7.12E-02	2.56E-03	5.27E-02	8.12E-01	2.46E-03	3.89E-02	4.03E-01	Mean	7.26E-02	2.66E-03	5.36E-02	7.84E-01	2.50E-03	3.91E-02	4.36E-01
Min	5.71E-02	1.63E-03	4.18E-02	7.85E-01	1.54E-03	3.03E-02	3.45E-01	Min	6.71E-02	2.25E-03	4.87E-02	7.76E-01	2.10E-03	3.42E-02	3.61E-01
Max	8.15E-02	3.32E-03	6.07E-02	8.57E-01	3.35E-03	5.10E-02	5.06E-01	Max	8.70E-02	3.78E-03	6.39E-02	7.96E-01	3.78E-03	5.07E-02	5.00E-01
STD	8.56E-03	6.02E-04	7.04E-03	1.86E-02	6.35E-04	7.10E-03	4.87E-02	STD	6.67E-03	5.08E-04	5.57E-03	6.04E-03	5.60E-04	5.97E-03	3.97E-02
Perth															
Metric	RMSE	LOSS	MAE	R2	MSLE	MEDAE	Max-error	Metric	RMSE	LOSS	MAE	R2	MSLE	MEDAE	Max-error
Mean	7.50E-02	2.83E-03	5.58E-02	7.85E-01	2.73E-03	4.16E-02	3.96E-01	Mean	7.60E-02	2.89E-03	5.64E-02	7.51E-01	2.75E-03	4.14E-02	4.06E-01
Min	6.53E-02	2.13E-03	4.86E-02	7.62E-01	2.04E-03	3.59E-02	3.24E-01	Min	7.18E-02	2.57E-03	5.36E-02	7.36E-01	2.32E-03	3.78E-02	3.72E-01
Max	8.57E-02	3.68E-03	6.54E-02	8.26E-01	3.64E-03	5.15E-02	4.49E-01	Max	8.16E-02	3.33E-03	5.98E-02	7.60E-01	3.25E-03	4.39E-02	4.43E-01
STD	6.05E-03	4.59E-04	4.85E-03	2.92E-02	5.09E-04	4.77E-03	3.33E-02	STD	3.72E-03	2.85E-04	2.02E-03	8.63E-03	3.66E-04	1.84E-03	2.15E-02
Tasmania															
Metric	RMSE	LOSS	MAE	R2	MSLE	MEDAE	Max-error	Metric	RMSE	LOSS	MAE	R2	MSLE	MEDAE	Max-error
Mean	7.19E-02	2.61E-03	5.46E-02	6.88E-01	2.52E-03	4.22E-02	3.75E-01	Mean	7.54E-02	2.85E-03	5.75E-02	6.52E-01	2.73E-03	4.47E-02	3.94E-01
Min	6.25E-02	1.95E-03	4.84E-02	6.59E-01	1.79E-03	3.83E-02	3.27E-01	Min	6.69E-02	2.24E-03	5.16E-02	6.17E-01	2.04E-03	4.03E-02	3.64E-01
Max	8.54E-02	3.64E-03	6.49E-02	7.08E-01	3.67E-03	5.06E-02	4.28E-01	Max	8.48E-02	3.59E-03	6.40E-02	6.69E-01	3.58E-03	4.87E-02	4.36E-01
STD	7.66E-03	5.63E-04	5.37E-03	1.42E-02	6.31E-04	3.99E-03	3.24E-02	STD	5.62E-03	4.30E-04	3.87E-03	1.57E-02	5.07E-04	3.12E-03	2.45E-02
Sydney															
Metric	RMSE	LOSS	MAE	R2	MSLE	MEDAE	Max-error	Metric	RMSE	LOSS	MAE	R2	MSLE	MEDAE	Max-error
Mean	6.26E-02	2.02E-03	4.49E-02	8.55E-01	1.63E-03	3.17E-02	3.80E-01	Mean	6.55E-02	2.20E-03	4.69E-02	8.37E-01	1.77E-03	3.28E-02	4.23E-01
Min	5.24E-02	1.38E-03	3.74E-02	8.49E-01	1.05E-03	2.58E-02	3.17E-01	Min	5.45E-02	1.48E-03	3.99E-02	8.24E-01	1.13E-03	2.84E-02	3.23E-01
Max	8.12E-02	3.30E-03	6.00E-02	8.63E-01	2.75E-03	4.50E-02	4.74E-01	Max	8.61E-02	3.71E-03	6.24E-02	8.48E-01	3.11E-03	4.47E-02	5.14E-01
STD	1.15E-02	7.56E-04	8.67E-03	4.57E-03	6.67E-04	6.93E-03	6.06E-02	STD	1.13E-02	7.86E-04	8.03E-03	6.97E-03	7.05E-04	5.83E-03	7.35E-02

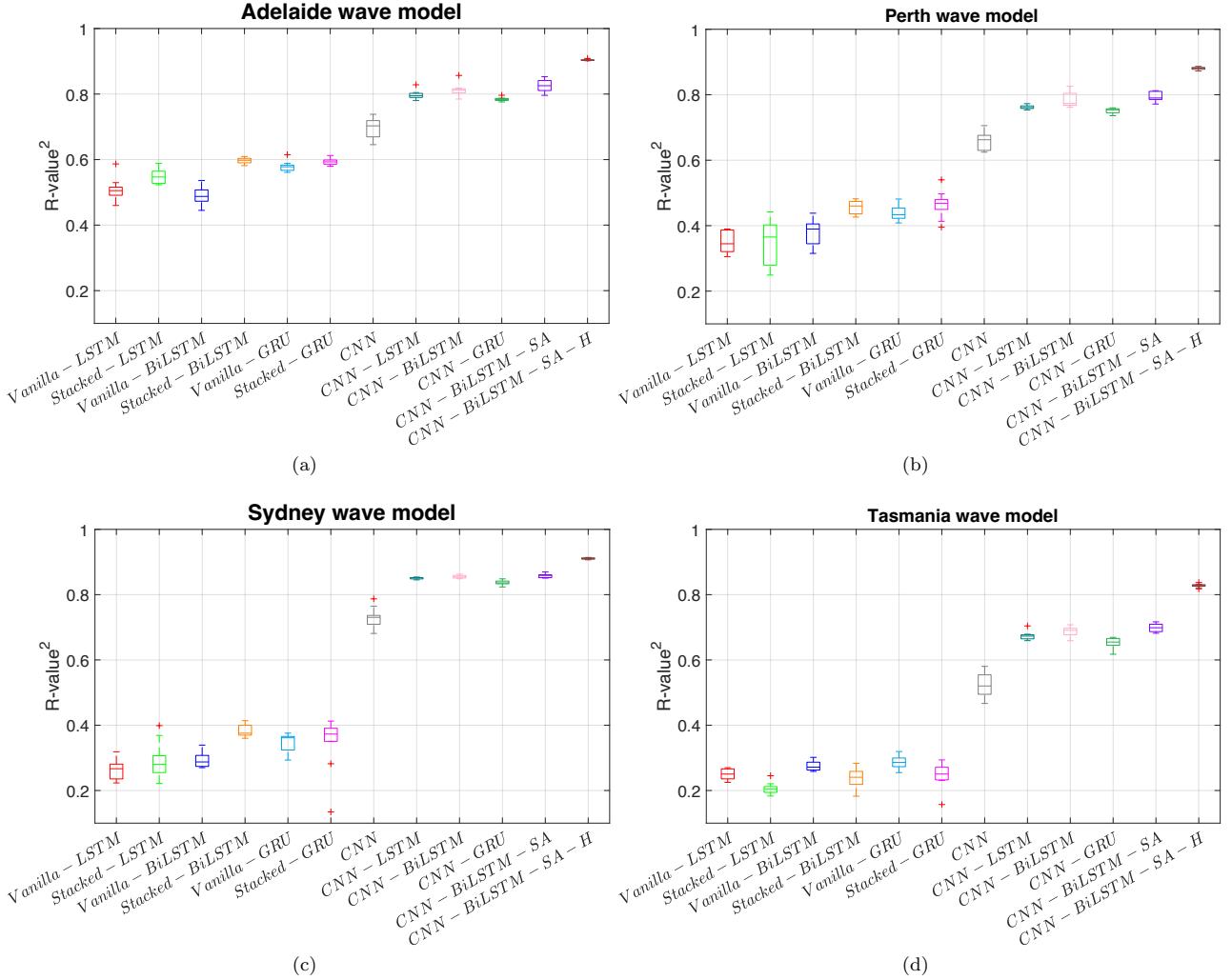


Figure 8: Statistical details of wave power prediction models comparison based on R^2 (accuracy) for four-wave models, (a) Adelaide, (b) Perth, (c) Sydney, (d) Tasmania sea sites.

remarkable capability to discern the complex and nonlinear relationship between the position of the WECs and the power generated by the wave farm. Consequently, the CNN model delivers more precise and robust results compared to all previously tested models. Moreover, incorporating LSTM layers significantly enhances the performance of the CNN, as demonstrated in Table 4. The CNN-LSTM model surpasses the CNN model by margins of 14.4%, 15.62%, 29.18%, and 16.80% in the respective locations of Adelaide, Perth, Tasmania, and Sydney in terms of R^2 .

This study evaluated the performance of three hybrid models: CNN-LSTM, CNN-BiLSTM, and CNN-GRU. Among these models, the highest level of performance was observed with the CNN-BiLSTM architecture across four distinct sea sites, as demonstrated in Table 5. The CNN-BiLSTM model exhibited exceptional accuracy in predicting the power output, achieving impressive percentages of 85.5%, 81.2%, 78.5%, and 68.8% in Sydney, Adelaide, Perth, and Tasmania, respectively. These results highlight the superior predictive capabilities of the CNN-BiLSTM hybrid model, positioning it as a promising choice for accurate power output predictions in the context of wave energy applications.

Traditional recurrent neural networks (RNNs) like LSTM and GRU encounter difficulties capturing long-range dependencies within sequential data. However, including self-attention layers in hybrid models proves advantageous in addressing this limitation. Self-attention allows each position in the input sequence to attend to all other positions, enabling the hybrid model to capture long-distance relationships and dependencies effectively. This capability enhances the model's contextual understanding and overall performance. Table 6 presents the prediction results of the combined CNN-BiLSTM model with self-attention (SA). A comparison between CNN-BiLSTM-SA and CNN-BiLSTM reveals a notable improvement attributed to the benefits of the self-attention mechanism, resulting in performance gains of 1.72%, 1.02%, 1.45%, and 0.47% in the respective locations of Adelaide, Perth, Tasmania, and Sydney. These results further demonstrate the efficacy of self-attention layers in enhancing the performance of hybrid models.

In order to further enhance the performance of the hybrid model (CNN-BiLSTM-SA), we introduced a fast and adaptive hyper-parameter optimiser named CNN-BiLSTM-SA-H. The outcomes of this modified hybrid model are presented in Table 6. Containing the hyper-parameter optimiser had a significant impact, resulting in substantial

Table 6: Statistical results of total power output prediction for CNN-BiLSTM-SA and CNN-BiLSTM-SA-E models based on four wave scenario

CNN-BiLSTM-SA								CNN-BiLSTM-SA-H							
Adelaide								Adelaide							
Metric	RMSE	LOSS	MAE	R2	MSLE	MEDAE	Max-error	Metric	RMSE	LOSS	MAE	R2	MSLE	MEDAE	Max-error
Mean	6.99E-02	2.47E-03	5.21E-02	8.26E-01	2.36E-03	3.92E-02	4.03E-01	Mean	4.521E-02	1.023E-03	3.506E-02	9.041E-01	9.175E-04	2.837E-02	2.626E-01
Min	5.75E-02	1.65E-03	4.24E-02	7.96E-01	1.57E-03	3.07E-02	3.51E-01	Min	4.294E-02	9.219E-04	3.292E-02	9.016E-01	8.307E-04	2.583E-02	2.289E-01
Max	8.00E-02	3.20E-03	6.23E-02	8.53E-01	3.20E-03	5.72E-02	4.65E-01	Max	4.641E-02	1.077E-03	3.600E-02	9.087E-01	9.752E-04	2.994E-02	2.934E-01
STD	7.85E-03	5.44E-04	6.35E-03	1.89E-02	5.91E-04	7.63E-03	3.55E-02	STD	9.317E-04	4.154E-05	1.004E-03	2.198E-03	3.663E-05	1.465E-03	2.124E-02
Perth								Perth							
Metric	RMSE	LOSS	MAE	R2	MSLE	MEDAE	Max-error	Metric	RMSE	LOSS	MAE	R2	MSLE	MEDAE	Max-error
Mean	7.06E-02	2.51E-03	5.31E-02	7.93E-01	2.38E-03	3.98E-02	3.87E-01	Mean	4.776E-02	1.142E-03	3.715E-02	8.808E-01	1.024E-03	3.022E-02	2.839E-01
Min	6.27E-02	1.97E-03	4.85E-02	7.72E-01	1.77E-03	3.54E-02	3.39E-01	Min	4.595E-02	1.056E-03	3.577E-02	8.724E-01	9.376E-04	2.906E-02	2.246E-01
Max	8.67E-02	3.76E-03	6.61E-02	8.12E-01	3.75E-03	5.13E-02	4.21E-01	Max	5.029E-02	1.264E-03	3.844E-02	8.867E-01	1.186E-03	3.193E-02	3.403E-01
STD	6.67E-03	5.01E-04	4.93E-03	1.43E-02	5.72E-04	4.36E-03	2.99E-02	STD	1.477E-03	7.081E-05	1.092E-03	4.594E-03	8.431E-05	1.156E-03	3.594E-02
Tasmania								Tasmania							
Metric	RMSE	LOSS	MAE	R2	MSLE	MEDAE	Max-error	Metric	RMSE	LOSS	MAE	R2	MSLE	MEDAE	Max-error
Mean	6.83E-02	2.35E-03	5.22E-02	6.98E-01	2.25E-03	4.07E-02	3.52E-01	Mean	4.698E-02	1.104E-03	3.693E-02	8.280E-01	1.001E-03	3.052E-02	2.539E-01
Min	6.20E-02	1.92E-03	4.77E-02	6.82E-01	1.77E-03	3.73E-02	3.02E-01	Min	4.532E-02	1.027E-03	3.560E-02	8.177E-01	9.313E-04	2.902E-02	2.095E-01
Max	8.09E-02	3.27E-03	6.14E-02	7.17E-01	3.26E-03	4.77E-02	3.87E-01	Max	4.897E-02	1.199E-03	3.918E-02	8.372E-01	1.078E-03	3.362E-02	3.418E-01
STD	6.31E-03	4.49E-04	4.52E-03	1.26E-02	5.04E-04	3.51E-03	3.13E-02	STD	1.045E-03	4.926E-05	1.098E-03	5.505E-03	4.183E-05	1.435E-03	4.133E-02
Sydney								Sydney							
Metric	RMSE	LOSS	MAE	R2	MSLE	MEDAE	Max-error	Metric	RMSE	LOSS	MAE	R2	MSLE	MEDAE	Max-error
Mean	6.39E-02	2.09E-03	4.59E-02	8.59E-01	1.71E-03	3.24E-02	3.88E-01	Mean	4.066E-02	8.273E-04	3.063E-02	9.105E-01	6.223E-04	2.355E-02	2.451E-01
Min	4.89E-02	1.19E-03	3.60E-02	8.51E-01	9.05E-04	2.48E-02	2.72E-01	Min	3.939E-02	7.757E-04	2.935E-02	9.067E-01	5.844E-04	2.116E-02	1.958E-01
Max	8.20E-02	3.37E-03	6.04E-02	8.70E-01	2.85E-03	4.50E-02	4.72E-01	Max	4.320E-02	9.330E-04	3.381E-02	9.140E-01	6.921E-04	2.843E-02	2.954E-01
STD	1.11E-02	7.27E-04	8.29E-03	5.98E-03	6.55E-04	6.63E-03	6.38E-02	STD	1.389E-03	5.750E-05	1.547E-03	2.568E-03	3.936E-05	2.270E-03	3.383E-02

Table 7: Optimal hyper-parameters configurations of CNN-BiLSTM-SA-H models based on four wave scenario

Tasmania															
CNF ₁	CNF ₂	CNF ₃	CNF ₄	NHU ₁	NHU ₂	PDO ₁	PDO ₂	BS	LR	AT	L2Reg	RMSE	R ²		
512	256	128	64	64	32	5.33E-02	5.33E-02	256	2.11E-03	16	1.03E-04	0.049	0.821		
512	256	128	64	16	8	6.93E-02	6.93E-02	1024	6.80E-04	64	1.18E-04	0.049	0.819		
512	256	128	64	64	32	5.62E-02	5.62E-02	512	5.32E-04	32	1.19E-04	0.049	0.821		
256	128	64	32	32	16	5.00E-02	5.00E-02	32	1.00E-04	32	1.00E-04	0.047	0.834		
Perth															
CNF ₁	CNF ₂	CNF ₃	CNF ₄	NHU ₁	NHU ₂	PDO ₁	PDO ₂	BS	LR	AT	L2Reg	RMSE	R ²		
256	128	64	32	32	16	5.00E-02	5.00E-02	32	1.00E-04	32	1.00E-04	0.047	0.880		
256	128	64	32	8	4	5.46E-02	5.46E-02	128	2.00E-04	16	1.11E-04	0.046	0.882		
512	256	128	64	32	16	7.05E-02	7.05E-02	256	1.43E-04	64	1.09E-04	0.046	0.881		
512	256	128	64	8	4	5.52E-02	5.52E-02	512	2.33E-04	80	1.35E-04	0.046	0.881		
Adelaide															
CNF ₁	CNF ₂	CNF ₃	CNF ₄	NHU ₁	NHU ₂	PDO ₁	PDO ₂	BS	LR	AT	L2Reg	RMSE	R ²		
256	128	64	32	32	16	5.00E-02	5.00E-02	32	1.00E-04	32	1.00E-04	0.044	0.905		
256	128	64	32	16	8	5.34E-02	5.34E-02	128	8.11E-04	32	1.05E-04	0.045	0.907		
512	256	128	64	8	4	7.19E-02	7.19E-02	128	4.89E-05	16	1.05E-04	0.044	0.909		
512	256	128	64	64	32	1.96E-01	1.96E-01	128	5.69E-05	80	1.01E-04	0.044	0.909		
Sydney															
CNF ₁	CNF ₂	CNF ₃	CNF ₄	NHU ₁	NHU ₂	PDO ₁	PDO ₂	BS	LR	AT	L2Reg	RMSE	R ²		
256	128	64	32	32	16	5.00E-02	5.00E-02	32	1.00E-04	32	1.00E-04	0.040	0.911		
128	64	32	16	16	8	1.58E-01	1.58E-01	64	1.11E-04	64	2.71E-04	0.042	0.899		
256	128	64	32	64	32	8.24E-02	8.24E-02	64	5.89E-05	16	3.28E-04	0.041	0.906		
512	256	128	64	128	64	2.08E-01	2.08E-01	64	2.83E-05	48	1.23E-04	0.041	0.904		

improvements in power prediction accuracy. Specifically, the accuracy rates reached 91%, 90.4%, 88%, and 82.8% in Sydney, Adelaide, Perth, and Tasmania, respectively. These results highlight the efficacy of the CNN-BiLSTM-SA-H model in optimising hyper-parameters and further enhancing the accuracy of power predictions in wave energy applications.

Figure 8 presents a distribution summary of 12 deep learning models, showcasing key statistical measures such as the median, quartiles, and outliers (indicated by red crosses). The accuracy metric (R^2) is used to evaluate the wave farm power predictions across four sea sites. This plot enables comparison and visualisation of the predictions' spread and central tendency.

Among the evaluated models, the proposed hybrid model, CNN-BiLSTM-SA-H, exhibited superior performance compared to standard, stacked, and other hybrid models. Remarkably, the combination with Bi-LSTM yielded more accurate power predictions among the hybrid CNN models. Additionally, CNNs exhibited intriguingly better performance than other Vanilla models (LSTM, BiLSTM, and GRU). This can be attributed to specific characteristics of CNNs, including their ability to extract local features, share parameters efficiently, exhibit translation invariance, and learn hierarchical representations.

Figure 9 presents the statistical prediction results, further emphasising the considerable improvement achieved by the proposed hybrid model across all four wave scenarios, as measured by the mean absolute error (MAE). These findings have significant practical implications for wave energy prediction and deep learning model development, enhancing the accuracy and reliability of wave farm power predictions.

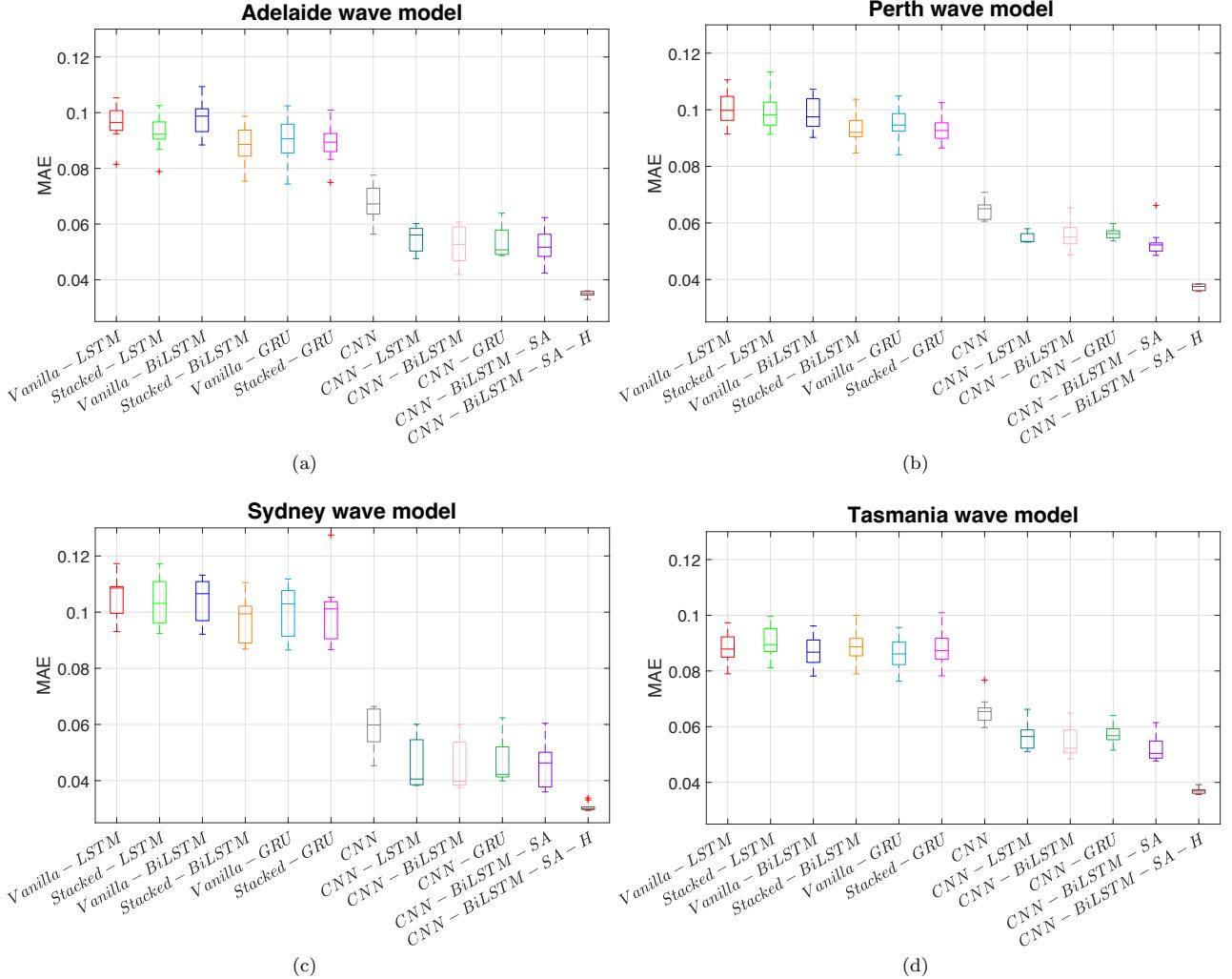


Figure 9: Statistical details of wave power prediction models comparison based on MAE for four-wave models, (a) Adelaide, (b) Perth, (c) Sydney, (d) Tasmania sea sites.

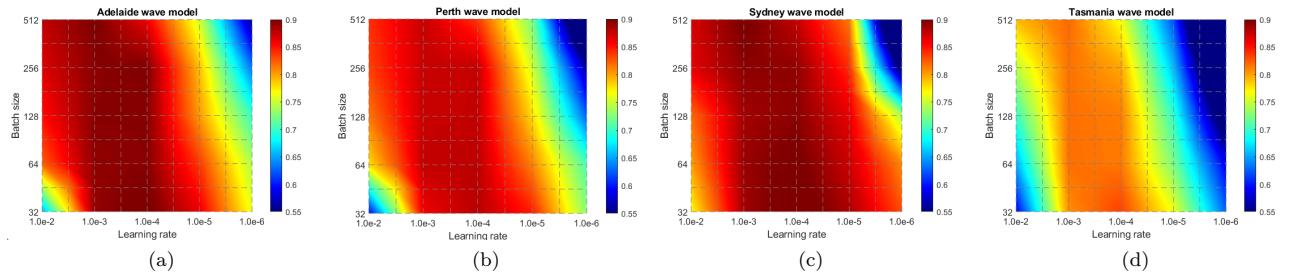
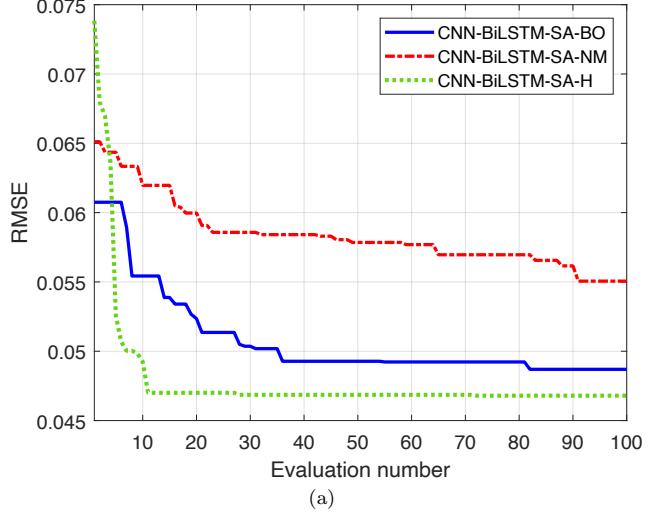


Figure 10: CNN-BiLSTM-SA hyper-parameters tuning using grid search for four wave models, a) Adelaide, b) Perth, c) Sydney, d) Tasmania sea sites.

4.4 Hyper-parameters tuning results

In Section 3.8, we introduced an efficient and fast hyper-parameter optimiser called the Evolutionary Grid Search (EGS) to fine-tune the hyper-parameters of the hybrid model (CNN-BiLSTM-SA). To gain insights into the impact of the learning rate and batch size on the power prediction performance, we conducted a grid search, and the results are depicted in Figure 10. The best-performing hyper-parameter combinations are highlighted in dark red. It is observed that the optimal range for the learning rate lies between $1.0e-3$ and $1.0e-4$, while a batch size greater than 64 yields the best results across all wave farms.

Figure 11 provides a comparison of the convergence rates among three different hyper-parameter tuners and underscores the superior exploration capability of the proposed method (CNN-BiLSTM-SA-H) in identifying the



(a)

Figure 11: Convergence rate comparison for CNN-BiLSTM-SA hyper-parameters tuning based on the Tasmania wave model.

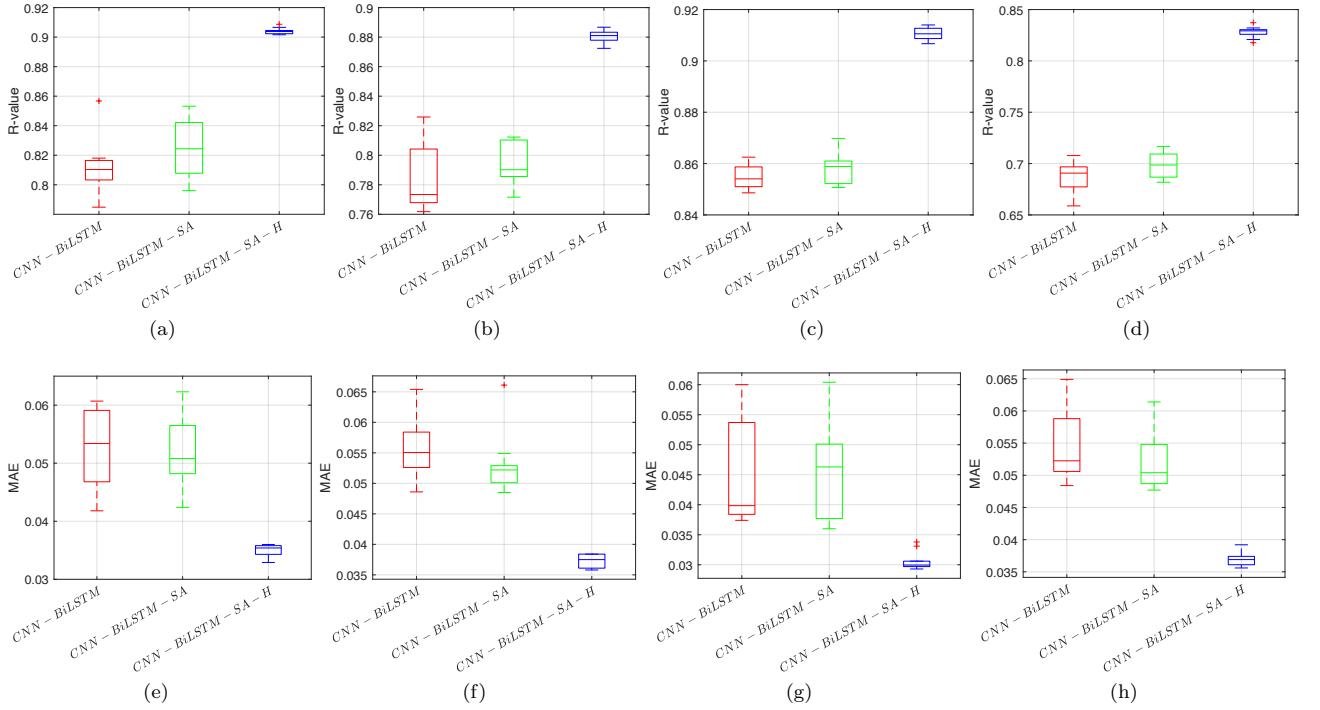


Figure 12: A comparison between CNN-BiLSTM, CNN-BiLSTM combined with self-attention, and hybrid of CNN-BiLSTM-SA with an evolutionary hyper-parameter tuner for four wave models, (a, e) Adelaide, (b, f) Perth, (c, g) Sydney, (d, h) Tasmania sea sites.

optimal configurations within the initial evaluation phase. As summarised in Table 7, the best-found hyperparameters encompass a range of settings crucial to achieving high performance in predicting the total wave farm power output. CNF and NHU denote the number of filters and hidden units in the convolutional and BiLSTM layers, respectively. PDO represents the dropout probability for the dropout layers. BS and LR indicate the optimal batch size and learning rate coefficient, respectively. Additionally, L2Reg denotes the adoption of L2 regularisation, while AT represents the Attention rate. The comprehensive range of hyper-parameter settings showcased in Table 7 emphasises the significance of thorough hyper-parameter tuning investigations. By identifying the most effective hyper-parameter configurations, we can maximise the model's predictive performance in estimating the total wave farm power output.

To gain a comprehensive understanding of the proposed hybrid model's evolution process and the influence of the self-attention strategy and hyper-parameter optimisation, Figure 12 is generated. The boxplots provide valuable insights into the performance improvements achieved by the proposed hyper-parameter optimiser. The results demonstrate its significant role in enhancing power prediction accuracy in the Adelaide, Perth, Sydney, and Tasmania sea sites

by approximately 9.8%, 12.8%, 5.8%, and 18.6%, respectively. These achievements emphasise the hyper-parameter optimisation approach's effectiveness in enhancing the hybrid model's predictive capabilities.

5 Conclusion

In this article, we introduce the innovative Self-attention Convolutional Bi-LSTM Technique as a robust strategy for forecasting power generation in wave farms. Through the utilisation of the strengths of Convolutional Neural Networks (CNNs), Bidirectional Long Short-Term Memory networks (Bi-LSTMs), and self-attention combined with a fast and effective hyper-parameters tuning method, our methodology adeptly captures the intricate spatial and temporal relationships present in wave farm datasets, resulting in remarkably precise predictions of power output. The flexibility of the self-attention mechanism empowers the model to dynamically allocate optimal weights to various temporal characteristics, thereby enhancing the accuracy of predictions. To assess the performance and generalisation ability of the proposed model, we tested it using four different wave farms on the Australian coasts, including Adelaide, Perth, Sydney, and Tasmania.

Our empirical findings, which are grounded in authentic wave farm data, clearly showcase the supremacy of the proposed predictive technique when contrasted with other cutting-edge models. The integration of this sophisticated approach holds significant promise for optimising energy distribution and streamlining the seamless assimilation of wave farms into existing electricity networks. Subsequent studies could explore adapting this technique to diverse renewable energy sectors and explore additional innovations to enhance predictions' precision. Overall, this research contributes to the advancement of wave energy technologies and provides valuable insights for commercialising and integrating wave power into the renewable energy landscape.

As a result, this assignment can be extended to improve the performance of renewable energy systems by integrating transfer learning and hyper-parameters adjustment for hybrid models. The pre-trained model can accept all types of wave conditions thanks to transfer learning. As a result, models may be created by collecting the acquired knowledge while addressing the issue and applying the many wave scenarios that are related to the exact same issue.

References

- [1] Yuekuan Zhou. Worldwide carbon neutrality transition? energy efficiency, renewable, carbon trading and advanced energy policies. *Energy Reviews*, 2(2):100026, 2023.
- [2] Leven J Zheng, Justin Zuopeng Zhang, Louisa Yee Sum Lee, Sajjad M Jasimuddin, and Muhammad Mustafa Kamal. Digital technology integration in business model innovation for carbon neutrality: An evolutionary process model for smes. *Journal of Environmental Management*, 359:120978, 2024.
- [3] Vishwas Dohale, Priya Ambilkar, Sachin Kumar Mangla, and Balkrishna Eknath Narkhede. Critical factors to sustaileanant innovations for net-zero achievement in the manufacturing supply chains. *Journal of Cleaner Production*, 455:142295, 2024.
- [4] Weiliang Tao, Malin Song, Shimei Weng, Xueli Chen, and Lianbiao Cui. Assessing the impact of environmental regulation on ecological risk induced by pm2. 5 pollution: Evidence from china. *Journal of Cleaner Production*, 451:142029, 2024.
- [5] Ahmad Shah Hizam Md Yasir, Mohamad Alif Omar, Nurul Afiqah Mohd Azhar, Azfarizal Mukhtar, and Mohd Zamri Yusoff. A comprehensive review of recent advancements and challenges in breakwater oscillating water column technology for wave energy conversion. *Energy Reports*, 13:4095–4113, 2025.
- [6] Danial Golbaz, Rojin Asadi, Erfan Amini, Hossein Mehdipour, Mahdieh Nasiri, Bahareh Etaati, Seyed Taghi Omid Naeeni, Mehdi Neshat, Seyedali Mirjalili, and Amir H Gandomi. Layout and design optimization of ocean wave energy converters: A scoping review of state-of-the-art canonical, hybrid, cooperative, and combinatorial optimization methods. *Energy Reports*, 8:15446–15479, 2022.
- [7] Zulfiqar Ahmad Khan, Tanveer Hussain, Ijaz Ul Haq, Fath U Min Ullah, and Sung Wook Baik. Towards efficient and effective renewable energy prediction via deep learning. *Energy Reports*, 8:10230–10243, 2022.
- [8] PMR Bento, JAN Pombo, RPG Mendes, MRA Calado, and SJPS Mariano. Ocean wave energy forecasting using optimised deep learning neural networks. *Ocean Engineering*, 219:108372, 2021.
- [9] Chenhua Ni and Xiandong Ma. Prediction of wave power generation using a convolutional neural network with multiple inputs. *Energies*, 11(8):2097, 2018.
- [10] Zhenqing Liu, Yize Wang, and Xugang Hua. Prediction and optimization of oscillating wave surge converter using machine learning techniques. *Energy conversion and management*, 210:112677, 2020.

- [11] L Li, Y Gao, DZ Ning, and ZM Yuan. Development of a constraint non-causal wave energy control algorithm based on artificial intelligence. *Renewable and Sustainable Energy Reviews*, 138:110519, 2021.
- [12] Chenhua Ni, Xiandong Ma, and Ji Wang. Integrated deep learning model for predicting electrical power generation from wave energy converter. In *2019 25th International Conference on Automation and Computing (ICAC)*, pages 1–6. IEEE, 2019.
- [13] LiGuo Wang, ChangWen Wen, ShiXuan Wu, and Sheng Wu. Electric power prediction of a two-body hinge-barge wave energy converter using machine learning techniques. *Ocean Engineering*, 305:117935, 2024.
- [14] Mehdi Neshat, Natalia Y Sergienko, Ashkan Rafiee, Seyedali Mirjalili, Amir H Gandomi, and John Boland. Meta wave learner: Predicting wave farms power output using effective meta-learner deep gradient boosting model: A case study from australian coasts. *Energy*, 304:132122, 2024.
- [15] Mumtaz Ali and Ramendra Prasad. Significant wave height forecasting via an extreme learning machine model integrated with improved complete ensemble empirical mode decomposition. *Renewable and Sustainable Energy Reviews*, 104:281–295, 2019.
- [16] Aditya N Deshmukh, MC Deo, Prasad K Bhaskaran, TM Balakrishnan Nair, and KG Sandhya. Neural-network-based data assimilation to improve numerical ocean wave forecast. *IEEE Journal of Oceanic Engineering*, 41(4):944–953, 2016.
- [17] Mumtaz Ali, Ramendra Prasad, Yong Xiang, and Ravinesh C Deo. Near real-time significant wave height forecasting with hybridized multiple linear regression algorithms. *Renewable and Sustainable Energy Reviews*, 132:110003, 2020.
- [18] Pujan Pokhrel, Elias Ioup, Julian Simeonov, Md Tamjidul Hoque, and Mahdi Abdelguerfi. A transformer-based regression scheme for forecasting significant wave heights in oceans. *IEEE Journal of Oceanic Engineering*, 47(4):1010–1023, 2022.
- [19] Abul Abrar Masrur Ahmed, S Janifer Jabin Jui, Mohanad S Al-Musaylh, Nawin Raj, Reepa Saha, Ravinesh C Deo, and Sanjoy Kumar Saha. Hybrid deep learning model for wave height prediction in australia’s wave energy region. *Applied Soft Computing*, 150:111003, 2024.
- [20] MA Srokosz. The submerged sphere as an absorber of wave power. *Journal of Fluid Mechanics*, 95(4):717–741, 1979.
- [21] NY Sergienko, BS Cazzolato, B Ding, and M Arjomandi. An optimal arrangement of mooring lines for the three-tether submerged point-absorbing wave energy converter. *Renewable Energy*, 93:27–37, 2016.
- [22] GX Wu. The interaction of water waves with a group of submerged spheres. *Applied ocean research*, 17(3):165–184, 1995.
- [23] Hojjat Salehinejad, Sharan Sankar, Joseph Barfett, Errol Colak, and Shahrokh Valaei. Recent advances in recurrent neural networks. *arXiv preprint arXiv:1801.01078*, 2017.
- [24] Mehdi Neshat, Meysam Majidi Nezhad, Seyedali Mirjalili, Davide Astiaso Garcia, Erik Dahlquist, and Amir H Gandomi. Short-term solar radiation forecasting using hybrid deep residual learning and gated lstm recurrent network with differential covariance matrix adaptation evolution strategy. *Energy*, 278:127701, 2023.
- [25] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.
- [26] Mike Schuster and Kuldip K Paliwal. Bidirectional recurrent neural networks. *IEEE transactions on Signal Processing*, 45(11):2673–2681, 1997.
- [27] Lionel P Joseph, Ravinesh C Deo, Ramendra Prasad, Sancho Salcedo-Sanz, Nawin Raj, and Jeffrey Soar. Near real-time wind speed forecast model with bidirectional lstm networks. *Renewable Energy*, 204:39–58, 2023.
- [28] Kyunghyun Cho, Bart Van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. Learning phrase representations using rnn encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078*, 2014.
- [29] Peter Shaw, Jakob Uszkoreit, and Ashish Vaswani. Self-attention with relative position representations. *arXiv preprint arXiv:1803.02155*, 2018.
- [30] Jikai Duan, Mingheng Chang, Xiangyue Chen, Wenpeng Wang, Hongchao Zuo, Yulong Bai, and Bolong Chen. A combined short-term wind speed forecasting model based on cnn-rnn and linear regression optimization considering error. *Renewable Energy*, 200:788–808, 2022.

- [31] Xingjian Shi, Zhourong Chen, Hao Wang, Dit-Yan Yeung, Wai-Kin Wong, and Wang-chun Woo. Convolutional lstm network: A machine learning approach for precipitation nowcasting. *Advances in neural information processing systems*, 28, 2015.
- [32] Shahram Hanifi, Andrea Cammarono, and Hossein Zare-Behtash. Advanced hyperparameter optimization of deep learning models for wind power prediction. *Renewable Energy*, 221:119700, 2024.
- [33] Plamen Angelov. A generalized approach to fuzzy optimization. *International Journal of Intelligent Systems*, 9(3):261–268, 1994.
- [34] Rashmi Dutta Baruah and Plamen Angelov. Dec: Dynamically evolving clustering and its application to structure identification of evolving fuzzy models. *IEEE transactions on cybernetics*, 44(9):1619–1631, 2013.
- [35] Neethu Elizabeth Michael, Ramesh C Bansal, Ali Ahmed Adam Ismail, A Elnady, and Shazia Hasan. A cohesive structure of bi-directional long-short-term memory (bilstm)-gru for predicting hourly solar radiation. *Renewable Energy*, 222:119943, 2024.
- [36] Ali Baheri, Praveen Ramaprabhu, and Christopher Vermillion. Iterative 3d layout optimization and parametric trade study for a reconfigurable ocean current turbine array using bayesian optimization. *Renewable energy*, 127:1052–1063, 2018.
- [37] Oliver Kramer, David Echeverría Ciaurri, and Sławomir Koziel. Derivative-free optimization. In *Computational optimization, methods and algorithms*, pages 61–83. Springer, 2011.
- [38] J F M Barthelemy and Raphael T Haftka. Approximation concepts for optimum structural design—a review. *Structural optimization*, 5:129–144, 1993.
- [39] Marco Antonio Luersen, Rodolphe Le Riche, and Frédéric Guyon. A constrained, globalized, and bounded nelder-mead method for engineering optimization. *Structural and Multidisciplinary Optimization*, 27:43–54, 2004.
- [40] Mehdi Neshat, Bradley Alexander, Nataliaia Y Sergienko, and Markus Wagner. A hybrid evolutionary algorithm framework for optimising power take off and placements of wave energy converters. In *Proceedings of the Genetic and Evolutionary Computation Conference*, pages 1293–1301, 2019.