Wave power forecasting using an effective decomposition-based convolutional Bi-directional model with equilibrium Nelder-Mead optimiser

Author

Neshat, Mehdi, Nezhad, Meysam Majidi, Sergiienko, Nataliia Y, Mirjalili, Seyedali, Piras, Giuseppe, Garcia, Davide Astiaso

Published

2022

Journal Title

Energy

Version

Submitted Manuscript (SM)

DOI

https://doi.org/10.1016/j.energy.2022.124623

Wave Power Forecasting Using an Effective Decomposition-based Convolutional Bi-directional Model with Equilibrium Nelder-Mead Optimiser

Mehdi Neshat[1,*], Meysam Majidi Nezhad[2], Nataliia Y. Sergiienko[3], Seyedali Mirjalili[1,4], Giuseppe Piras[2], Davide Astiaso Garcia[5]

[1] *Center for Artificial Intelligence Research and Optimization, Torrens University Australia, Brisbane, QLD 4006, Australia,* neshat.mehdi@gmail.com (* Corresponding), ali.mirjalili@torrens.edu.au

[2] *Department of Astronautics, Electrical and Energy Engineering (DIAEE), Sapienza University of Rome, Italy,* meysam.majidinezhad@uniroma1.it, giuseppe.piras@uniroma1.it

[3] *School of Mechanical Engineering, University of Adelaide, Australia,* nataliia.sergiienko@adelaide.edu.au

[4] *Yonsei Frontier Lab, Yonsei University, Seoul, Republic of Korea*

[5] *Department of Planning, Design, and Technology of Architecture, Sapienza University of Rome, Italy,* davide.astiasogarcia@uniroma1.it

Abstract

Energy industries and governments consider ocean wave power a promising renewable energy source for reaching the net-zero plan by 2050 and restricting the rise in global temperatures. It expects the potential global ocean wave power production to be around 337 GW annually. Although wave energy forecasting critically enables economic dispatch, optimal power system management, and the integration of wave energy into power grids, the forecasting process is complicated by the stochastic, intermittent, and non-stationary nature of waves. Thus, this paper proposes a novel hybrid forecasting model comprising an adaptive decomposition-based method (Nelder-Mead variational mode decomposition) and a convolutional neural network featuring bi-directional long short-term memory. Furthermore, we propose a fast and effective optimiser to adjust the hybrid model's hyper-parameters and evaluate the decomposition technique's role in increasing the accuracy of wave energy flux predictions considering a forecasting period of 6 h. With regard to assessing the proposed model's effectiveness, we use a real wave dataset from a buoy positioned off Favignana Island in the Mediterranean Sea and compare the proposed model with six well-known forecasting methods and five hybrid deep-learning models. According to our findings, the proposed model significantly outperforms existing approaches over extended time periods and compared with the bi-directional long short-term memory, the developed adaptive decomposition method, and new hyper-parameters tuner improve the prediction accuracy at 45% and 13.6%, respectively.

Highlights

- A novel hybrid convolutional model is proposed for wave energy flux prediction.
- An effective hybrid variational mode decomposition method is introduced.
- A new hyper-parameter optimiser is proposed: Equilibrium Nelder-Mead optimisation.
- The proposed model's efficiency is compared with 11 hybrid and popular prediction models

Keywords: Adaptive decomposition method, convolutional deep learning model, equilibrium optimisation, ocean wave power prediction, significant wave height, wave energy flux

Nomenclature

Table 1: A list of all abbreviations used in this research work, and are sorted in alphabetical order:

abbreviation	**full name**
ADMM	Alternate direction method of multipliers
AI	artificial intelligence
ANFIS	Adaptive neuro-fuzzy inference system
ANN	Artificial Neural networks
Bi-LSTM	Bidirectional Long short-term memory network
CR	probability crossover rate
CNN	Convolutional neural network
ConvLSTM	Convolutional Long short-term memory network
DNN	Deep neural networks
DRL	Deep Reinforcement Learning
DWNN	Discrete Wavelet Neural Networks
EC	Evolutionary computation
ENMO	Equilibrium Nelder-Mead optimisation algorithm
EO	Equilibrium optimisation
FC-LSTM	Fully connected long short-term memory network
FFNN	Feed-forward neural networks
FS	Feature selection
GA	Genetic algorithm
HPO	Hyper-parameter Optimisation
IMF	Intrinsic mode functions
LSTM	Long short-term memory network
MAE	Mean absolute error
MSE	Mean square error
NM	Nelder-Mead simplex direct search method
NMVMD	Nelder-Mead Variational mode decomposition
NOA	National Observatory of Athens
NWRN	Italian National Wave Recording Network
PNN	Polynomial neural networks
PSO	Particle Swarm Optimisation
PTO	Power take-off
PI	Proportional integral
RFNN	Recurrent fuzzy neural network
RMSE	Root mean square error
RNN	Recurrent neural networks
SCADA	Supervisory control and data acquisition
VMD	Variational mode decomposition

Table 2: A list of symbols, Greek letters, subscript, and superscript applied in this paper

Symbols	Description	Symbols	Description
T_p	Peak wave period	V	control volume
T_e	Wave energy period	D	atoms gravity
H_s	Significant wave height	ϕ	velocity of flow around and inside of V
D_p	Mean wave direction	β	control parameter
ω	wave frequency	λ	a vector of real numbers from 0 to 1
h	water depth	Max_{iter}	maximum number of evaluations
J_ω	Wave energy flux	γ	exploitation coefficient
ρ	Density of water	θ	exploration coefficient
g	Gravitational acceleration	x_r	best-reflected candidate
$S(\omega)$	monodirectional wave spectrum	x_e	expanded candidate
C_g	group velocity	δ_{ie}	inside contraction rate
D_c	interval between sea site and the coast	δ_{oc}	outside contraction rate
D_p	interval between sea site and adjacent port	δ_e	expansion rate
t	time step	w_s	centre frequency
$h_t^{\rightarrow}, h_t^{\leftarrow}$	forward and backward hidden layers	$D(t)$	The Dirac distribution
W_{HH}	weight at recurrent neuron	α	fidelity constraint
W_{IH}	weight at input neuron	F_e	estimated value
W_{HO}	weight at output layer	F_t	true value
b	bias value	K	Entire number of the modes
$tanh$	hyperbolic-tangent	u_k	signifies the k_{th} mode
σ	sigmoid (non-linear activation function)	UB	Upper bound of the decision variables
$*$	convolution operator	LB	Lower bound of the decision variables
$\odot$	Hadamard product operator	B_S	batch size
I_t	input gate	L_R	LSTM learning rate
C_t	modulation of input gate	N_h	Number of layers
H_t	hidden gate	N_n	Number of neurons in each layer
F_t	forget gate	Op	LSTM optimiser
x_t	input layers	PF	Penalty factor
O_t	output gate	L_m	Lagrange Multiplier

1. Introduction

Exacerbated global environmental problems and the shortcomings of conventional energy sources (e.g. oil and natural gas) have propelled the renewable energy sector's dramatic growth in recent decades [1]. Various global plans have been developed to address climate change, including the European Green Deal project [2], which intends to replace all traditional energy sources with renewable energy (zero greenhouse gas emissions) by 2050. Ocean wave power has emerged as a strong contender in the renewable energy [3] due to its high power density and minimal environmental impact [4]. Although it has yet to be widely deployed, ocean wave power could provide remote island communities with an off-grid solution, fulfil the offshore power needs of small industrial projects, and serve the rapidly expanding aquaculture industry.

However, the rapid development of wave power technology [5] depends accurately predicting power production to ensure a reliable and stable supply of power to the grid [6]. Therefore, this requires forecasting dynamic environmental conditions [7] and the parameters of incoming waves, including wave power flux. Ocean wave predictions can both estimate the power generated by wave energy converters (WECs) [8], and improve the performance of those converters. In particular, tuning the power take-off parameters of WECs to changing wave conditions [9] and developing efficient strategies for controlling WECs [10] can significantly increase power generation. It is critical to consider the design of the WEC damping controller given that the complex interaction between WEC components and dynamic sea conditions [11] challenges the prediction of optimal settings. For Carnegie's CETO design, a sensitivity analysis demonstrated that optimising the damping controller could boost the energy harnessed by up to 6% [12]. Moreover, such predictions are also critical to preventing damage to wave power units in extreme wave conditions [13].

However, the forecasting process is complicated by the nonlinear and stochastic nature of waves, which means developing accurate and reliable predictors of wave energy flux and wave parameters is critical to the wave power industry's evolution [14]. Furthermore, time scales for forecasts depend substantially on the

application. Predicting wave elevation 10-30 s in advance is essential for advanced control strategies that maximise power output from WECs [15]. However, the highly dynamic nature of wave data, the considerable model training time, and overall model complexity complicate making accurate real-time predictions, with wave forecasting at the 'sea-state' time scale (up to 6 hours) required to reliably integrate wave power technology into the existing electricity grid [16]. 'weather-' (up to 7 days) and 'seasonal-' (beyond 1 month) predictions of wave conditions are critical for vessel routing [17], the aquaculture industry, and the security of the residents and businesses of coastal regions [18]. Long-term wave forecasting (days and months) usually involves hindcast wave models [18] that are accurate but computationally expensive. Hence, state-of-the-art machine-learning and deep-learning mechanisms which are relatively fast and can handle large datasets can usefully predict wave conditions over shorter periods of time (i.e., hours) [16].

Various researchers have developed wave parameter forecasting using soft computing approaches. For example, Hashim et al. [19] proposed an advanced version of the adaptive neural fuzzy inference model, the Takagi-Sugeno-based fuzzy model. As a first step [19], they proposed a feature selection mechanism to find the best subset of a complete set of model input parameters for producing forecast targets. These parameters included sea surface wind speed, wind-direction, atmospheric temperature, and temperature of sea surface. However, they found that training the membership functions was a time-consuming process. Elsewhere, an early study [20] designed a model forecasting significant wave height near Mangalore (on the west coast of India), over 48-hour intervals using a discrete type of wavelet neural network (DWNN). As input, the DWNN model used a multi-resolution time set, with the prediction results demonstrating better accuracy and consistency than traditional artificial neural networks (ANN) models featuring one-time resolution. Meanwhile, a recent study by Bento et al. [16] developed a deep learning-based model for predicting wave energy flux and certain other wave parameters using 13 wave datasets collected from the Pacific and Atlantic oceans. Bento et al. used a bio-inspired optimisation algorithm (moth-flame optimisation) to adjust the hyper-parameters of a feed-forward neural network (FFNN) model. However, the optimised model did not perform well in terms of forecasting time-series data.

With regard to optimise the energy produced by a WEC, a power take-off controller has frequently been used. However, despite the a considerable research concerning different control strategies and the design of optimal and smart WEC controllers, efficient controller implementation remains challenging, mainly because the process depends on the incoming wave forces. For example, Lu et al. [21] proposed combining a hybrid recurrent wavelet and Elman neural network with an improved gravitational search algorithm in order to control integration of offshore wind and wave energy operations. Their comparison of the hybrid model with a recurrent fuzzy neural network (RFNN) and a classical proportional-integral (PI) network demonstrated that the hybrid model overwhelmed both PI and RFNN. In another work, a hybrid neural model was proposed to predict the short-term wave forces [22]; this multi-layer fully connected model was confirmed via comparison with the random wave model that 60% and 80% more energy was absorbed. However, Li et al. [22] did not consider a proper decomposition method for decreasing the signal noise and trained the model with the components of the main signal. Another study [23] introduced an online forecasting model comprising a sequence-to-sequence learning model mixed with a derivative-free optimisation method. Although it aimed to develop a new resampling method that could forecast short and long-term intervals for wind, solar, and ocean wave power. However, the superiority of that work [23] compared just with a few traditional machine learning methods. Recently, a two-modular convolutional deep learning model was hybridised by a gradient-based optimisation [24] in order to predict the aerodynamic attributes and the physical fields in a flapping foil, and the modelling results revealed that the two-modular CNN model was more effective than other predictive methods, especially in terms of computational cost.In another practical study, Zou et al. [25] proposed a Deep Reinforcement Learning (DRL) model and simulated a point absorber with a direct-drive PTO for predicting the wave power. The authors reported a considerable improvement in the prediction accuracy from 24% up to 152% compared with other model-based controls.

With regard to recapitulating the principal challenges in developing long-term predictive wave characteristics models, the main research gaps can be detailed as follows:

1. One of the most important features in order to improve the performance of the deep learning models is hyper-parameters tuning [16]. Recently, various meta-heuristics were proposed to find the optimal

hyper-parameter settings; however, in the majority of case studies, their performance is not considerable due to improper control parameters initialisation and updating during the optimisation process.

2. Most of the popular meta-heuristic optimisation algorithms have been invented in order to handle the numerical benchmarks rather than the hyper-parameters tuning. Therefore, they need to be modified for optimising the hyper-parameters.

3. Proposing the optimal decomposition methods' parameters is challenging [26]. It has been reported that these parameters strongly involve the performance of the decomposition methods. Moreover, an insufficient decomposition setting leads to the low performance of the wave predictors.

To address the aforementioned difficulties in forecasting long-term wave parameters, we developed a new convolutional bi-directional learning-based framework. This proposed model features an efficient decomposition-based method that can predict long-term wave power with considerable precision and employs, an architecture that can be considered a parallel bi-directional recurrent neural model (Bi-LSTM) [27]. Thus, this work improves on current wave forecasting research via several principal contributions:

1. First, we introduce a novel hybrid forecasting model comprising an adaptive decomposition-based method (Nelder-Mead variational mode decomposition [NMVMD]) and a convolutional neural network (CNN) featuring bi-directional long short-term memory (BiLSTM).

2. In order to overcome the shortcomings of the decomposition techniques (such as requiring tuning parameters), an adaptive decomposition approach is proposed. This effective decomposition method includes variational mode decomposition and a Nelder-Mead search algorithm to de-compose time-series data into sub-signals.

3. We also propose a new hybrid optimisation algorithm that combines the equilibrium optimisation (EO) algorithm with the Nelder-Mean (NM) simplex direct search to optimise the hyper-parameters of the proposed hybrid wave forecasting model.

4. Next, our comparative analysis considers Bi-LSTM against stacked LSTM and vanilla LSTM in terms of modelling stochastic wave data, demonstrating that Bi-LSTM outperforms the other LSTM models.

5. After applying the proposed hybrid neural model to real data collected from a buoy in the Mediterranean Sea off Favignana Island between 1999 and 2013, we evaluate its performance based on the six-hours-ahead long term forecast interval.

6. Finding from a systematic comparison of the hybrid model with the adaptive neuro-fuzzy inference system (ANFIS), feed-forward neural network (FFNN), polynomial neural networks (PNNs), vanilla LSTM, stacked LSTM, the Bi-LSTM model, and five hybrid decomposition-based models. reveal that the proposed hybrid model outperformed the other models in terms of both accuracy and the convergence curve.

The remainder of this work is structured as follows: Section 2 includes some subsections: the wave energy resource (2.1), the details of the case study with a summary of the collected data(2.2, and in the following methods (2.3, 2.4, 2.5) used in order to design and develop the proposed forecasting wave model. In Section 3, we formulate and compare various forecasting strategies with the proposed framework. Then, in Section 4, we characterise the principles and associations specified by each significant finding and place them in the proper standpoint. Finally, we summarise the results obtained by this work and present a few thoughts for future research plans in in Section 5.

2. Methods

In this section, initially we describe the technical details of wave energy resource modelling and assessing. Next, the collected wave dataset is introduced. In the following, the primary methods which were applied

in this study are discussed comprehensively, including the standard Bi-directional Long short-term memory network (Bi-LSTM) and convolutional LSTM. Furthermore, one of the modern and effective meta-heuristics is described entitled equilibrium optimisation method and its hybrid version with Nelder-mead as a hyper-parameter tuner. The applied decomposition method in this research is a combination of variational mode decomposition and a fast local search discussed in detail in the following.

2.1. Wave energy resource

Different methodologies are employed to assess wave energy's potential use as a renewable energy resource. These include in-situ wave buoy measurements, satellite altimetry, and wave hindcasts using spectral wave models [28]. Measurements available from wave buoy records include significant wave height (H_{m0}, or H_s), characteristic wave period, (e.g. peak wave period T_p or wave energy period T_e), and mean wave direction (D_p). The resulting datasets can be used to estimate wave power (or wave energy flux) at a specific location. Thus, for a sea state characterised by H_s and T_p (which propagates in a single direction), wave power can be calculated by applying the principle of superposition, which considers wave power at each independent frequency component ω [29]:

$$J = \int_0^\infty J(\omega)d\omega = \int_0^\infty \rho g S(\omega) \cdot C_g(\omega)d\omega \tag{1}$$

where $\rho g S(\omega)$ represents the energy in the wave, and $C_g(\omega)$ represents the group velocity, or the velocity at which the wave energy is propagating, ρ is the density of water, g is the gravitational acceleration, and $S(\omega)$ is the monodirectional wave spectrum. It should be noted that the wave energy flux due to the combination of two waves $S_1(\omega)$, $C_{g1}(\omega)$ and $S_2(\omega)$, $C_{g2}(\omega)$, arriving simultaneously to the same point from two different directions is not well captured by Equation (1).

The group wave speed depends on the wave frequency ω and water depth h [29]:

$$C_g(\omega) = \frac{1}{2}\frac{\omega}{k(\omega)}\left(1 + \frac{2k(\omega)h}{\sinh 2k(\omega)h}\right), \tag{2}$$

where $k(\omega)$ is the wave number that is defined by the dispersion relation [29]:

$$\omega^2 = gk(\omega)\tanh k(\omega)h. \tag{3}$$

As a result, the monodirectional wave energy flux for a known wave spectrum $S(\omega)$ is calculated as:

$$J = \int_0^\infty \rho g S(\omega) \cdot \frac{1}{2}\frac{\omega}{k(\omega)}\left(1 + \frac{2k(\omega)h}{\sinh 2k(\omega)h}\right)d\omega. \tag{4}$$

If to assume a Pierson-Moskowitz shaped spectrum, $S(\omega)$ is defined by [30]:

$$S(\omega) = \frac{5\pi^4 H_{m0}^2}{\omega^5 T_p^4}\exp\left(\frac{-20\pi^4}{\omega^4 T_p^4}\right), \tag{5}$$

where the peak wave period T_p can be related to the wave energy period as $T_e = 0.857T_p$.

2.2. Wave data from the Mediterranean Sea off Favignana Island

This paper utilises data collected at a buoy installed in the Mediterranean Sea off Favignana Island by the Italian National Wave Recording Network (NWRN) and administered by the Agency for Environmental Protection and Technical Services [31] over a period of 14 years from 1999 to 2013. The geographical location of the installed buoy (HS5) and the sea depth off the northwest coast of Sicily, in meters under the medium sea level and surrounding Favignana Island [32] can be seen in Figure 1.

As we can see that the site $HS5$ is situated close to the western shore of Favignana Island. We chose this site for the greater prevalence of energetic waves compared to neighbouring locations [31]. The average

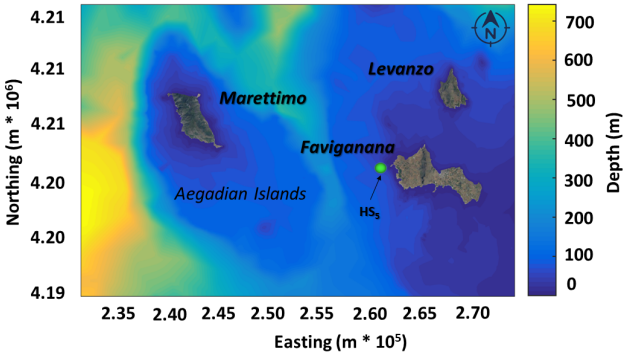

Figure 1: Bathymetry of the studied Favignana Island and surrounding. The sea depth shows by the color bar.

wave energy flux reported to be around 6.88 kW/m. Most of the high-energy waves have a frequency of 25.97% (roughly 95 days per year). The wave climate of $HS5$ is demonstrated in Figure 2, where the colour scale represents the wave energy per meter of wave front [kWh/m], the numbers indicate the sea state's occurrence [number of hours per year], and the contour lines indicate the wave power levels [kW/m].

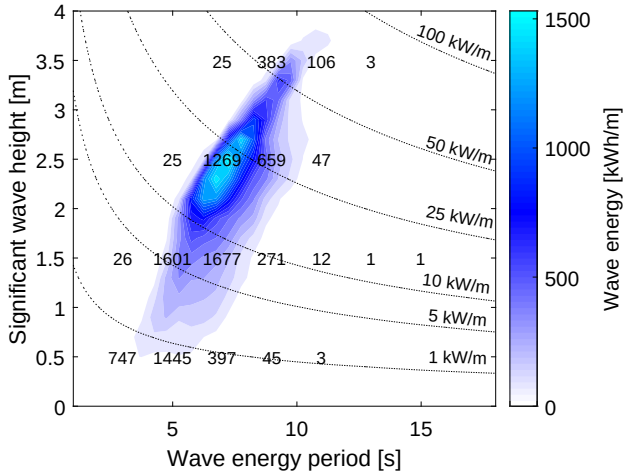

Figure 2: Joint probability density estimation of the wave buoy data for the $HS5$ site.

As Figure 3 makes apparent, wave energy is generally focused on fewer bins than neighbouring sea sites. The figure also shows the wave energy period distribution with respect to the wave direction for the selected sea site. The predominant wave directions are between 330^o and 350^o N by a frequency of 26.12%. Nevertheless, peak seasonal variation reaches around 75% between the summer and winter months. Table 3 presents the characteristics of the $HS5$ site, including its geographical coordinates, the sea depth, annual average wave power and wave energy. D_c and D_p denote the distance from the $HS5$ site to the coast, and to the most adjacent port, respectively.

Figure 4 shows the extensive fluctuations of the wave data. The first observation from the zoomed version of this figure is that the fluctuation of the wave direction is less than other wave parameters compared with wave energy period and the mean energy period.

Table 3: The details information of the applied sea site including the geographical coordinates; the sea depth; annual average wave power and wave energy.

Parameter	Value
Site	HS_5
Coordinates (long, lat)	12.27^o, -37.94^o
Water depth (m)	10.00
Average wave power (kW/m)	6.88
Average wave energy (MWh/m)	60.27
D_c (km)	0.50
D_p (km)	7.00
Port	Faviganana

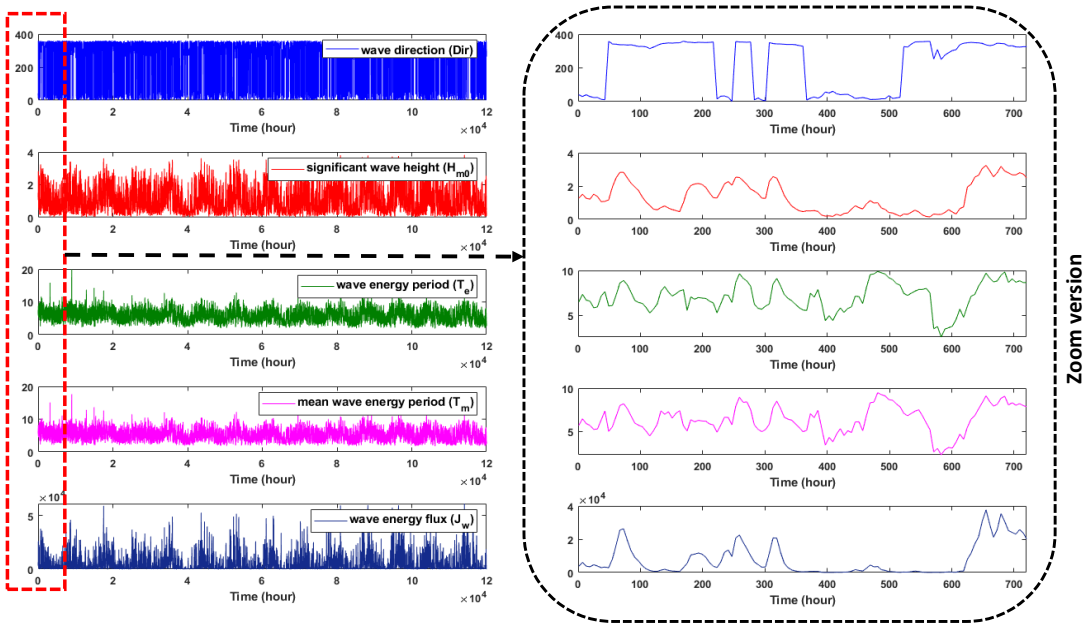

Figure 4: The features of the collected dataset for Favignana Island. The right figure is regard to one month of the recorded wave data.

Meanwhile, Table 4 also shows important statistical information about wave features and wave energy flux including mean, median, standard deviation, and upper and lower bounds. In order to calculate the correlation coefficients between wave energy flux (J_w) and significant wave height (H_{m0}), wave energy period (T_e), mean energy period (T_m) (also known as the "zero cross mean wave period"), and wave direction (Dir), we used the correlation matrix presented in Table 5. As expected, the biggest correlation coefficient is regarding between wave energy period (T_e) and mean energy period (T_m) at 99%. Furthermore, the top-ranking correlation concerns significant wave height and wave energy flux by 91%.

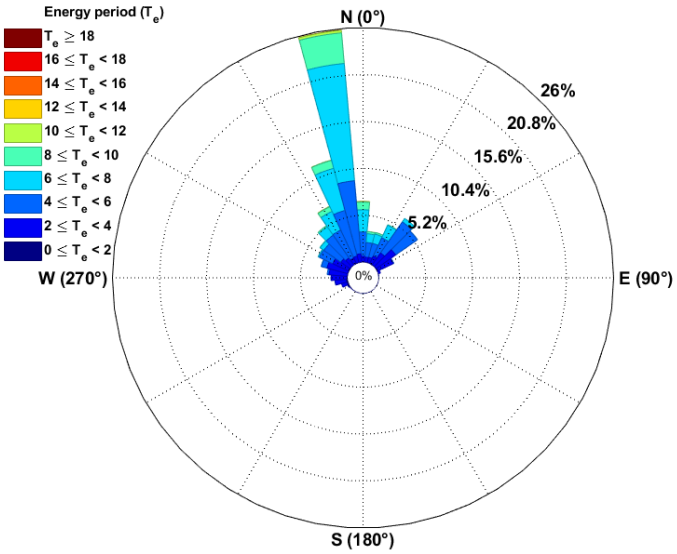

Figure 3: The wave climate of energy period (T_e) and wave direction (Dir)

Table 4: Statistical information of wave measurements dataset

	H_{m0}[m]	T_m[s]	Dir [°N]	T_e [s]	J_w [W/m]
Mean	1.089	5.945	245.216	5.355	5036.158
Median	0.858	5.913	322.93	5.297	1477.969
Min	0.0133	1.875	0.0065	1.727	0.626
Max	3.816	19.88	360	17.633	61122.49
STD	0.795	1.793	136.324	1.729	7511.134

Table 5: The correlation among the features of the collected wave dataset.

	H_{m0}[m]	T_m[s]	Dir [°N]	T_e [s]	J_w [W/m]
H_{m0}[m]	1	0.78	033	0.77	0.91
T_m[s]	0.78	1	0.26	0.99	0.74
Dir [°N]	0.33	0.26	1	0.27	0.30
T_e [s]	0.77	0.99	0.27	1	0.74
J_w [W/m]	0.91	0.74	0.30	0.74	1

2.3. Bi-directional long short-term memory network

The main drawback of regular recurrent neural networks (RNN) is that they pay particular attention to previous correlations of sequential data during the training phase. This attribute produces vanishing and exploding gradient problems in the context of long prediction sequence. Thus, the model might be trained with null weights (un-trained) or exploding weights. In order to deal with this deficiency, it is also recommended to explore future correlations of the sequential data [33].

One successful development of LSTM architectures is Bi-LSTM. Bi-LSTM is a modified formation of LSTM that includes both forward and backward layers of LSTM. Using both types of connection before and after updating the weights of sequential neurons, Bi-LSTM can simultaneously analyse past and future

time series data [27]. This bi-directional structure means every memory block comprises a couple of LSTM layers. The forward and backward layers both feature two series of hidden layer elements with reverse time processes. Furthermore, the states of the two hidden layers are joined to produce an identical output. The forward LSTM is responsible for obtaining prior information on the input sequence and the backward LSTM evaluates future features of the sequential data [27].

The forward and backward hidden layers are denoted by $h_t^{\rightarrow}$ and $h_t^{\leftarrow}$, respectively, at the current time step t. Equation 6 shows how the hidden states of the forward process are estimated.

$$h_t^{\rightarrow} = f_H(W_{IH\rightarrow} x_t + W_{HH\rightarrow} h_{t-1}^{\rightarrow} + b_{h\rightarrow}) \tag{6}$$

where $h_t^{\rightarrow}$ is repeated over $t = (1, ..., T)$. In the following, the computation of backward layer can be seen in Equation 7.

$$h_t^{\leftarrow} = f_H(W_{IH\leftarrow} x_t + W_{HH\leftarrow} h_{t-1}^{\leftarrow} + b_{h\leftarrow}) \tag{7}$$

Finally, to aggregate and compute the output sequence y_t at each time step t, Equation 8 is applied.

$$y_t = W_{HO\rightarrow} h_t^{\rightarrow} + W_{HO\leftarrow} h_t^{\leftarrow} + b_o \tag{8}$$

This research used MATLAB R2020a to develop the whole implementation. For the Bi-LSTM and LSTM, we used the deep learning with time series toolbox, and BiLSTM (*bilstmLayer*). to update the cell and hidden state, the activation function set the hyperbolic tangent function (*tanh*) and used the *sigmoid* function to engage the gates. The '*unit-forget-gate*' was used initialise the bias. There were 200 and 150 in the first and second Bi-LSTM units, respectively.

2.4. Convolutional long short-term memory model (ConvLSTM)

The main motivation for proposing a convolutional LSTM is the considerable number of redundancy of spatial data, that a fully connected LSTM (FC-LSTM) can engender. To deal with this issue, Shi et al. [34] first introduced this approach with an LSTM comprising convolutional architectures with recurrent connections. That convolutional LSTM model utilised the operator of the convolution to restrict the coming state of a singular cell before, the inputs recursively defined the future state together with the prior states of its connected neighbours. The ConvLSTM model can be formulated [34] as follows:

$$\begin{aligned}
F_t &= \sigma\ (w_{(XF)}\ *\ x_t + w_{(HF)}\ *\ h_{(t-1)} + B_F) \\
I_t &= \sigma\ (w_{(XI)}\ *\ x_t + w_{(HI)}\ *\ h_{(t-1)} + B_I) \\
\check{C}_t &= tanh\ (w_{(X\check{C})}\ *\ x_t + w_{(H\check{C})}\ *\ h_{(t-1)} + B_{\check{C}}) \\
O_t &= \sigma\ (w_{(XO)}\ *\ x_t + w_{(HO)}\ *\ h_{(t-1)} + B_O) \\
C_t &= F_t\ \odot\ C_{(t-1)} + I_t\ \odot\ \check{C}_t \\
h_t &= O_t\ \odot\ \tanh\ (C_t)
\end{aligned} \tag{9}$$

where both *tanh* (hyperbolic-tangent) and σ (sigmoid) are non-linear activation functions that let the model design complex mappings among the inputs and outputs of the model applied. The convolution operator shows by $*$, and $\odot$ denotes the Hadamard product operator [35]. The I_t, C_t, H_t, and F_t are input gate, modulation of input gate, hidden states, and forget gates, respectively. Meanwhile, input layers and output gates show by x_t, and O_t, respectively. In convLSTM, the most significant module is the memory cell (C_t) that plays the role of a state information aggregator. This can be controlled by the gates.

2.5. Hyper-parameter tuning

This study used the grid search method, one of the most widely employed strategies for hyper-parameter optimisation (HPO) problems [36]. Due mainly to its straightforward implementation [37] and status as a complete search method [38]. That is a grid search can guarantee finding the optimal solution with-in

a finite amount of time in the search space. However, the curse of dimensionality [39] can undermine the approach if the number of hyper-parameters rises, the grid search runtime will increase exponentially. A grid search operates as an exhaustive search that depends on pre-defined subsets of the hyper-parameter search space. The hyper-parameters are designated by combinations of the lower bound, upper bound and the resolution of the grids. Upon assessing all combinations, the most accurate model should represent the optimal hyper-parameters configuration. Three scale types determine the size of steps: logarithmic, quadratic, and linear scales. This study used the linear scale to combine the hyper-parameters.

2.5.1. Equilibrium optimisation (EO) algorithm

The recently introduced (EO) algorithm [40] is among the most popular population-based optimisation methods, performing adequately in comparison to other meta-heuristics. The EO algorithm was developed based on the mass balance control volume in order to determine whether various situations should be dynamic or equilibrial. Equation 10 represents the first-order differential formula of the general mass-balance calculation.

$$V\frac{dD}{dt} = \phi D_{eq} - \phi D + g \tag{10}$$

where V expresses the control volume and D denotes the atoms gravity near the V, and also $V\frac{dD}{dt}$ defines the ratio of mass change. In order to determine the velocity of flow around and inside of V, ϕ is calculated. g signifies the mass formation ratio interior the V. D_{eq} implies the particles' density towards the V through an equilibrium phase. If we assume $V\frac{dD}{dt} = 0$, a regular equilibrium phase is obtained and the movement is represented via $\lambda = \frac{\phi}{V}$. Finally, Eq 10 is recalculated as follows:

$$\frac{dD}{\lambda Deq - \lambda D + \frac{g}{V}} = dt \longrightarrow \int_{D_0}^{D} \frac{dD}{\lambda D_{eq} - \lambda D + \frac{g}{V}} = \int_{t_0}^{t} dt \tag{11}$$

Eq. 12 is achieved after calculating Eq. 11 as follows:

$$D = D_{eq} + (D_0 - D_{eq})\exp[-\lambda(t - t_0)] + \tfrac{g}{\lambda V}(1 - (\exp[-\lambda(t - t_0)])) \tag{12}$$

where t_0 and D_0 are the commencement time and concentration-dependent on the integration interim.

The final convergence stage involves the EO algorithm indicating the equilibrium situation, which might be a local optimum or, in the best-case scenario, a global optimum. The EO method involves a vector characterising the equilibrium pool to present a set of candidate solutions. The main exploration operator introduces four best-found particles during the optimisation search process. After that, the average of four particles is computed to produce a fifth particle.

The EO uses this exploration procedure in order to increase robustness and search ability. Furthermore, this proportion of best-found particles simultaneously increases the exploitation stage. The equilibrium pool vector is presented as follows:

$$\vec{D}_{eq,pool} = \left\{\vec{D}_{eq(1)}, \vec{D}_{eq(2)}, \vec{D}_{eq(3)}, \vec{D}_{eq(4)}, \vec{D}_{eq(ave)}\right\} \tag{13}$$

The β is one of the most notable control parameters in the EO algorithm that is related to an exponential term for developing a fit equivalence between both exploitation and exploration. The β reckoning stands as follows:

$$\vec{\beta} = e^{-\vec{\lambda}(t - t_0)} \tag{14}$$

where λ implies a vector of real numbers from zero to one uniformly distributed and the iteration of the fitness function is shown by t that should be reduced throughout the optimisation process as follows.

$$t = \left(1 - \frac{i}{Max_i}\right)^{\gamma\frac{i}{Max_i}} \tag{15}$$

where the current and the maximum number of evaluations depicted by *iter* and Max_{iter}, respectively. γ indicates a fixed cost in order to accomplish EO's exploitation capacity. One effective mechanism is proposed in the EO algorithm to control the exploration and exploitation ability that can guarantee a convergence situation by decreasing the search step size as follows.

$$\vec{t_0} = \frac{1}{\vec{\lambda}} \ln(-\theta sign(\vec{r} - 0.5)[1 - e^{-\vec{\lambda}t}]) + t \tag{16}$$

where θ denotes a fixed value that can be effective in improving the EO's exploration ability.

2.5.2. Equilibrium Nelder-Mead optimisation (ENMO) algorithm

The standard EO algorithm adaptively updates the control parameters which improve exploratory ability among the initial generations, and enhances exploitation behaviour during the final iterations. Although substantially balancing these two search phases produces a noteworthy optimisation performance. In the multi-modal search space, however, the EO algorithm cannot perform adequately and converge with the optimal solutions derived by the exploitation mechanism during the final iterations. The controlling parameter used by the EO, algorithm cannot properly decrease the search step size for particles during the last iterations. To address this shortcoming, we employ a fast local search namely, the Nelder-Mead (NM) simplex direct search method is applied.

NM method [41] is a downhill and unconstrained optimisation method that has mostly been applied to nonlinear optimisation problems without considering the derivatives. Although the NM method can demonstrate a high convergence rate during the initial iterations and proposes a local or global optimum. However, a premature convergence represents one of the approach's main disadvantages of the NM. The NM method commences through forming an origin simplex and assessing the fitness function at each point of the simplex, and the function rates the candidates. Next, in order to compute the simplex during the next replication, three procedures are performed: expansion, reflection and contraction. Otherwise, all candidates will be updated based on the best solution.

Figure 5 provides an example of the NM algorithm, where x_c is the centroid point (solution), which is the opposite of the worst point. To summarise NM search performance, it should be noted that it includes a comparison of f^r, which is the value of the evaluation function at the reflected object with the value of the best, the second-worst, and worst simplex model ($f^k_{best} \leq f(y_{n-1}) \leq f(y_n)$). The NM algorithm's main steps are as follows.

1. The worst candidate will be substituted using the best-reflected candidate of (x_r) and the x_e (expanded candidate) if $f(x_r) < f^k_{best}$.
2. The worst candidate should be substituted using the x_r (reflected candidate) if $f^k_{best} \leq f(x_r) < f(y_{n-1})$.
3. The worst candidate will be substituted using the best-performed candidate between x_oc and x_r if $f(y_{n-1}) \leq f(x_r) < f(y_n)$.
4. The worst candidate will be substituted using x_{ic}, if $f(y_n) \leq f(x_r) < f(x_{ic})$.
5. if $f(y_n) \leq f(x_r) < f(x_{ic})$ next the whole simplex should be shrunk and just the best candidate (y_0) will be fixed.

2.6. Decomposition approaches

Decomposition is a broad technique in order to solve a problem with high complexity by splitting it into more diminutive sup-problems and figuring each one out individually, in the form of parallel or consecutively. However, if the decomposition is carried out sequentially, the runtime complexity of the problem increases exponentially or quadratically [42].

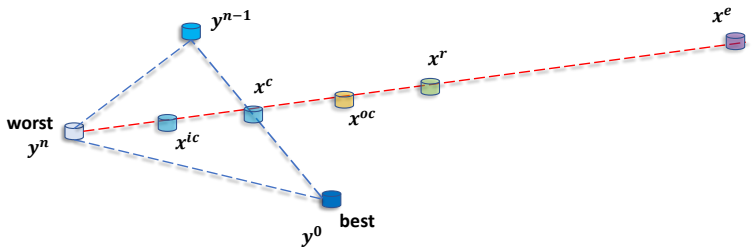

Figure 5: A schematic search behaviour of the Nelder-Mead algorithm when $\delta_{ic} = -0.5$ (inside contraction rate) ,$\delta_{oc} = 0.5$ (outside contraction rate) and $\delta_e = 2$ (expansion rate)

2.6.1. Variational mode decomposition (VMD)

Variational mode decomposition (VMD) is one of the modern decomposition methods, that shows a high performance with regard to decompose the climatic time-series parameters effectively [43] and embedded with a single step or multi-step hybrid deep learning models for wind and wave power forecasting [44]. More technical details of VMD method can be seen in Appendix A.

2.6.2. Nelder-Mead Variational mode decomposition (NMVMD)

Despite the advantages of Variational Mode Decomposition (VMD) compared to traditional decomposition techniques which include noise suppression, high-speed estimate, substantial numerical back grounding and a non-recursive sifting method the efficiency of VMD depends on the initial values of the control parameters. The main parameters are the number of intrinsic mode functions (IMFs), the value of the penalty factor to push the process in the feasible area, the type of approach required to adjust the central frequencies, and (perhaps most importantly) the value of the update frequency for the Lagrange multiplier. In order to handle the issue of hyper-parameter tuning, this paper proposes an effective optimiser to address this problem, namely, the iterative NM simplex direct search algorithm detailed in the previous subsection. This approach represents a rapid polytope search method, particularly useful for low-dimensions optimisation problems and often employed to address nonlinear problems where derivatives cannot be known [45]. Figure 6 shows the performance of the Nelder-mead algorithm, in terms of convergence rates, in the context of optimising VMD in ten independent experiments.

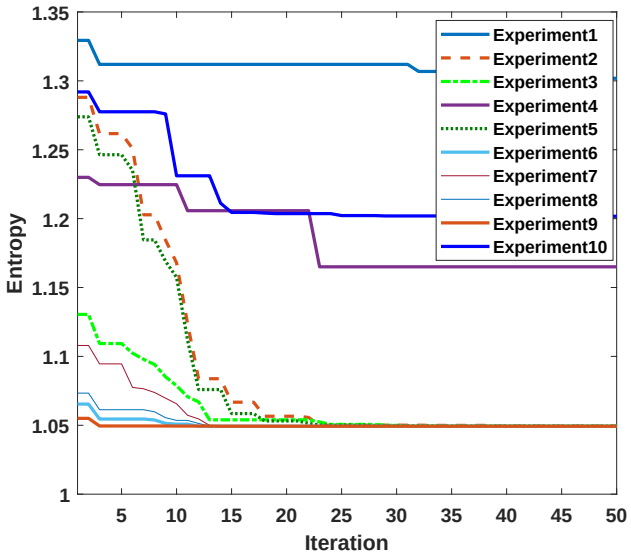

Figure 6: A comparison of 10 independent runs of iterative Nelder-mead to optimise the performance of VMD.

To evaluate the VMD's performance, we applied permutation entropy (PE) [46], a reliable and robust time-series method that presents a quantification parameter of a dynamic system's complexity through ordering relations between states of a time series and deriving a probability distribution of the ordinal patterns. Consequently, having better VMD's settings produces a small entropy value meaning the NM approach explores the search space of the VMD's hyper-parameters to minimise the entropy value. The convergence rate of ten independent experiments can be seen in Figure 6.

2.7. Proposed novel hybrid wave forecasting model

Building on the components of the previous subsections, introduces a hybrid deep learning-based model for forecasting wave energy flux based on time-series wave data with system inputs including significant wave height (H_{m0}), wave energy period (T_e), mean energy period (T_m), and wave direction (Dir). Development and functioning of the proposed model (NMVMD-BiLSTM) can be understood as follows:

1. To evaluate the efficiency of the wave parameters for forecasting wave energy flux, six Bi-LSTM models with different wave inputs were adopted.
2. After comparing the performance of the six models described in Section 3, we selected the best-performing model to use as the kernel of the hybrid system.
3. To decompose the original wave data into sub-series of wave data, we applied the VMD method.
4. subsequently, the NM simplex search method was utilised to address hyper-parameters issues existing during the VMD's initialisation phase (i.e. number of intrinsic mode functions, penalty factor, method for adjusting central frequencies, and update frequency for the Lagrange multiplier). Table 6 shows the upper and lower bounds of the VMD parameters. To assess the NM-VMD algorithm's performance, we applied permutation entropy [46], the value of which should be minimised to converge with the superior VMD configuration. Figure 7 illustrates the best-found settings identified by the NM-VMD.
5. To improve the performance of the Bi-LSTM model, a CNN was embedded to the make a hybrid model (CNN-BiLSTM) capable of extracting and learning the features from the sequential wave data. The complete architecture of the CNN-BiLSTM wave forecasting model is presented in Figure 8.
6. Finally, in order to address the issue of hyper-parameters tuning, we proposed a new optimisation method, namely the Equilibrium NM optimisation (ENMO) algorithm which we then compared with three other well-known optimisation methods. The ENMO algorithm considerably outperformed the other models in terms of robustness and convergence rate.

Table 6: The upper and lower bounds of VMD's hyper-parameters.

	IMFs number	Penalty factor	Lagrange multiplier	Initialise Method
Min	3	100	10^{-6}	1
Max	10	2000	10^{-2}	3

3. Experiments and results

In order to analyse the effectiveness of the proposed hybrid deep learning-based model, we developed a comprehensive comparative framework. The first step of which involved developing six sequential Bi-LSTM models and comparing them based on the wave parameters addressed in Section 3.1.

Section 3.2 discusses the influence of hyper-parameters within the Bi-LSTM forecasting model and emphasises the importance of hyper-parameters and, the whole prediction, the stability, and accuracy analysis outcomes are analysed in Section 3.3. The descriptions of the performance indices of forecasting models can be seen in Appendix B.

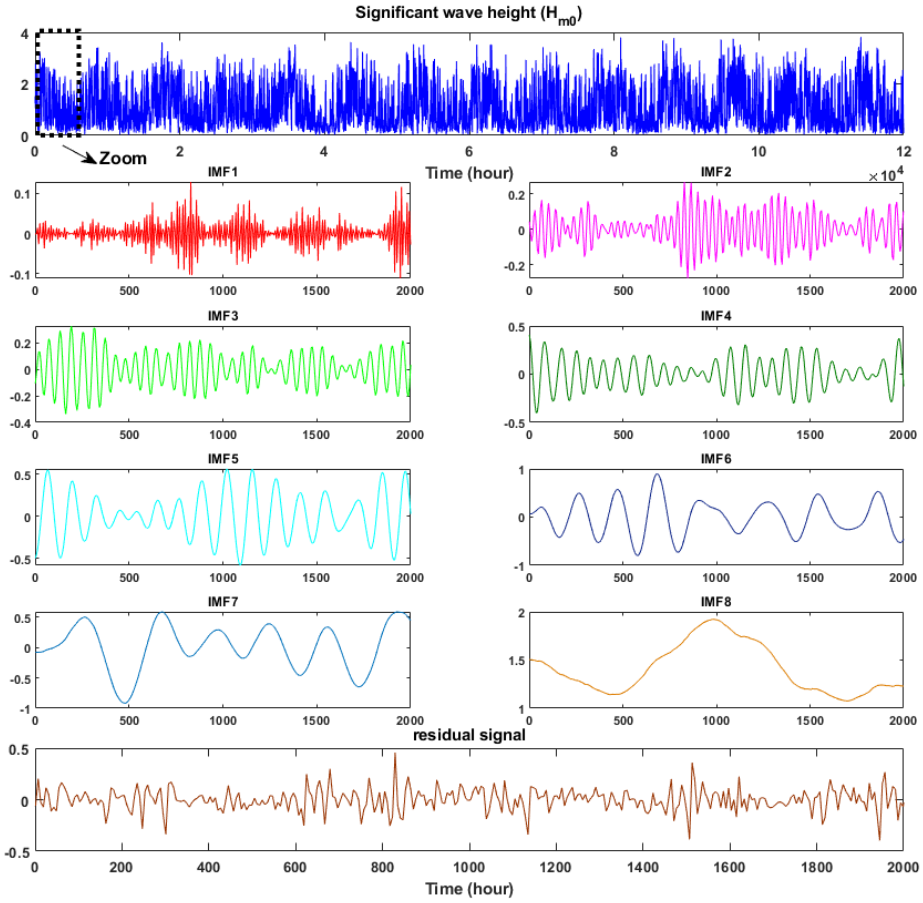

Figure 7: Significant wave height (H_{m0}) components decomposed by NM-VMD for 11 years recorded dataset.

3.1. Bi-LSTM performance models

According to the correlation coefficients values among wave features that can be seen in Table 5, we developed six independent wave forecasting models as follows;

- $Model_1$: has one input that is the significant wave height (H_{m0}) because H_{m0} has the highest correlation (0.91%) with the wave energy flux (J_w).
- $Model_2$: has two inputs including significant wave height (H_{m0}) and the wave direction.
- $Model_3$: has two inputs; significant wave height (H_{m0}) and wave energy period (T_e). Both parameters can play an important role to predict the J_w.
- $Model_4$: the number of inputs is three including the significant wave height (H_{m0}), wave energy period (T_e), and mean energy period (T_m).
- $Model_5$: has the same number of inputs as Model 4, but instead of the mean energy period (T_m), and the new input is wave direction.
- $Model_6$: includes four inputs such as H_{m0}, T_e, T_m, and wave direction.

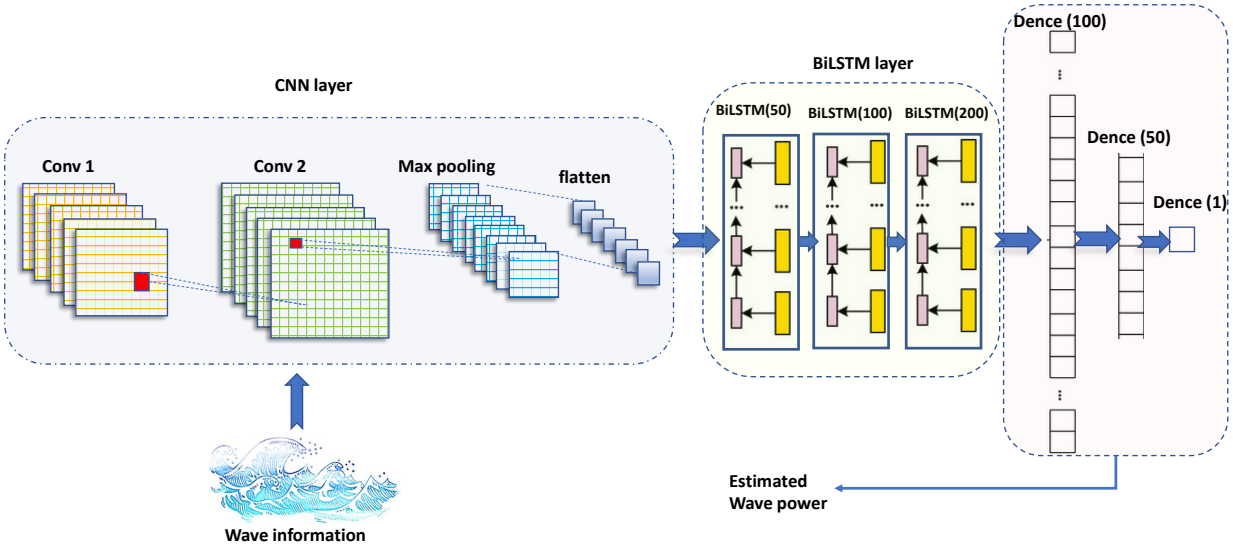

Figure 8: The architecture of the proposed CNN-BiLSTM wave deep model. filters=128, kernel size=1, activation='relu',

Figure 9 presents the results derived from the models described, importantly revealing that the forecasting model with two inputs (Model2) can estimate wave power flux better than the other models. Meanwhile, the box plot (Figure 9) shows that the system with only one input of significant wave height (H_{m0}) performed better than the models with two, three, and four inputs, making it the second best-performing model. Meanwhile, a combination of significant wave height (H_{m0}) and wave energy period (T_e) as inputs (Model3) produced the third-highest accuracy.

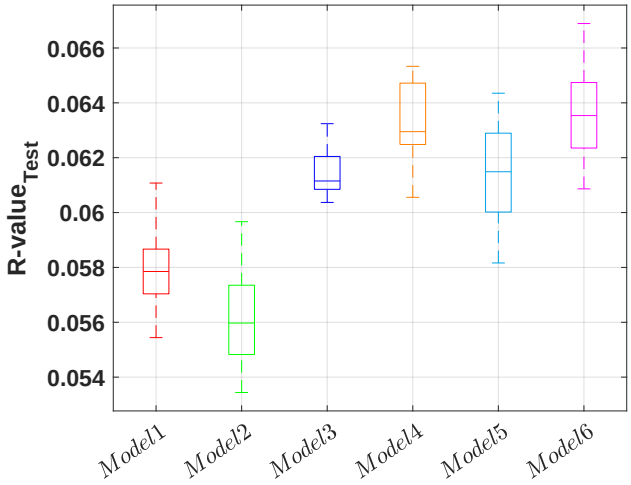

Figure 9: The statistical results of the six forecasting wave flux models with different inputs.

3.2. Bi-LSTM hyper-parameters tuning

Given the challenge of configuring the Bi-LSTM, and the absence of a reliable theory [47] regarding establishing the hyper-parameters and architecture, this study used a popular grid search method. Regarding tuning these configurations, we should consider both the dynamic neural model's behaviour and the

computational budget. This meant assuming a fixed architecture for the Bi-LSTM model and focusing on the influence of batch size and learning rate. Evaluation produced batch sizes in the range $2^6 \leq B_S \leq 2^9$ and learning rates in the range $10^{-6} \leq L_R \leq 10^{-4}$.

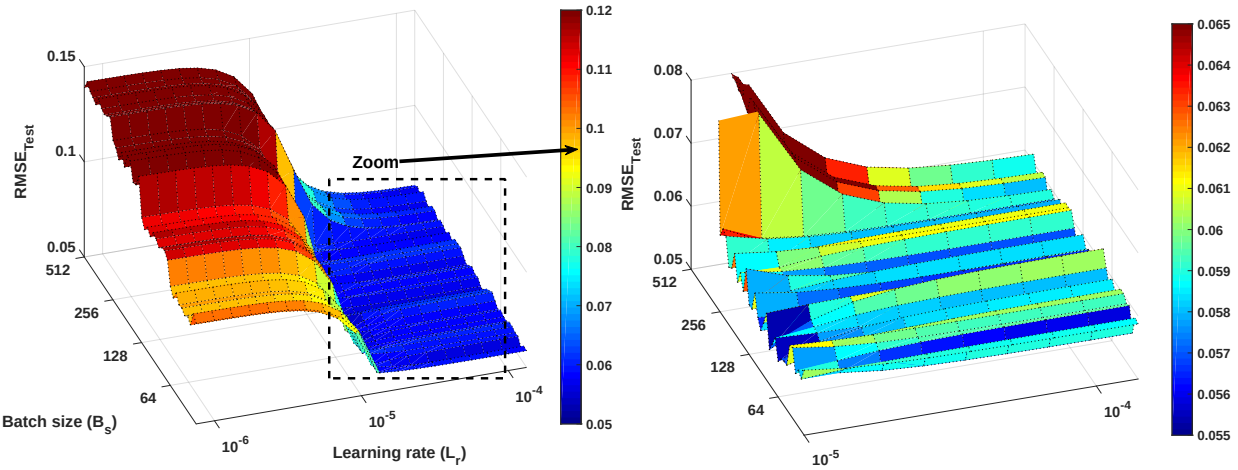

Figure 10: Batch size and learning rate hyper-parameters tuning using the grid search for the BiLSTM wave forecasting model with one input (significant wave height (H_{m0})). The zoomed version can be seen in (b).

Figure 10 provides the landscape analysis of the Bi-LSTM model's hyper-parameters (batch size and learning rate). The most prominent observation of Figure 10 revealed that the relationship between the accuracy of the model and the values of the learning rate is meaningful. The figure also indicates a learning rate around 10^{-4} obtained the best forecasting results for smaller batch sizes. Notably, batch size moderates the estimation precision of the error gradient when training neural models [48]. The outcomes presented confirm that adopting smaller batch sizes (e.g. 2^6 or 2^7) can deliver the most desirable training stability and generalisation effectiveness (for a given computational cost) over an extensive range of learning rates.

Table 7 presents the hyper-parameters sensitivity results for the Bi-LSTM forecasting model. According to training outcomes, the Bi-LSTM model exhibited the most stable forecasts with a learning rate of 10^{-4} and batch size of 2^7, and the Bi-LSTM model demonstrated the highest sensitivity to large batch sizes.

3.3. Hybrid wave power forecasting model's results

Thus, the first step, compared the performance of the hybrid wave forecasting model without hyper-parameters optimisation with various popular machine learning-based forecasting techniques. Figure 11 (a) visually summarises the statistical prediction results including the minimum, maximum and median RMSE values for each model. From Figure 11, we can see a comparison of eight models: NMVMD-CNNBiLSTM (the proposed hybrid model), LSTM, stacked LSTM, Bi-LSTM, ANFIS [49], PNN, and FFNN. Although the fewest training and testing errors were obtained using the proposed model (NMVMD-CNNBiLSTM). Furthermore, the Bi-LSTM model performed competitively. To further investigate the proposed model's performance, Table 8 summarises the outcomes of wave energy flux forecasting based on four indices. It is obvious that the NMVMD-CNNBiLSTM hybrid model without hyper-parameter optimiser is able to outperform other popular models and produce more reliable forecasting outcomes.

Our second step compared the effectiveness of using the proposed hybrid model with hyper-parameter optimisation (NMVMD-CNNBiLSTM with ENMO), comparing the performance of this approach with that of three other hybrid models and the hybrid model without hyper-parameter optimisation. Figure 11 (b) demonstrates that the performance of the hybrid model with hyper-parameter optimisation not only performs better than the other hybrid models but also performs significantly better than the proposed model without ENMO. Table 9 details the statistical prediction results for the four hybrid wave models.

Table 7: The average performance metrics for the Bi-LSTM forecasting wave model with one input (significant wave height (H_{m0})) with different batch sizes (B_s) and learning rate (L_r) values.

	BS=64, LR=1e-04									**BS=128, LR=1e-04**							
	MSE		**RMSE**		**MAE**		**R**			**MSE**		**RMSE**		**MAE**		**R**	
	Train	**Test**	**Train**	**Test**	**Train**	**Test**	**Train**	**Test**		**Train**	**Test**	**Train**	**Test**	**Train**	**Test**	**Train**	**Test**
Mean	3.450E-03	3.479E-03	3.510E-02	5.897E-02	3.510E-02	3.521E-02	8.834E-01	8.851E-01	***Mean***	3.452E-03	3.506E-03	3.508E-02	5.918E-02	3.508E-02	3.500E-02	8.840E-01	8.817E-01
Max	3.599E-03	3.626E-03	3.615E-02	6.021E-02	3.615E-02	3.605E-02	8.849E-01	8.887E-01	***Max***	3.505E-03	3.917E-03	3.566E-02	6.259E-02	3.566E-02	3.731E-02	8.849E-01	8.878E-01
Min	3.385E-03	3.126E-03	3.454E-02	5.591E-02	3.454E-02	3.428E-02	8.809E-01	8.789E-01	***Min***	3.401E-03	3.117E-03	3.461E-02	5.583E-02	3.461E-02	3.328E-02	8.819E-01	8.763E-01
Std	5.804E-05	1.447E-04	4.954E-04	1.245E-03	4.954E-04	5.670E-04	1.125E-03	2.818E-03	***Std***	3.891E-05	2.399E-04	3.617E-04	2.030E-03	3.617E-04	1.092E-03	9.379E-04	3.406E-03
	BS=256, LR=1e-04									**BS=512, LR=1e-04**							
	MSE		**RMSE**		**MAE**		**R**			**MSE**		**RMSE**		**MAE**		**R**	
	Train	**Test**	**Train**	**Test**	**Train**	**Test**	**Train**	**Test**		**Train**	**Test**	**Train**	**Test**	**Train**	**Test**	**Train**	**Test**
Mean	3.447E-03	3.511E-03	3.490E-02	5.924E-02	3.490E-02	3.521E-02	8.837E-01	8.829E-01	***Mean***	3.466E-03	3.462E-03	3.509E-02	5.882E-02	3.509E-02	3.526E-02	8.833E-01	8.839E-01
Max	3.495E-03	3.753E-03	3.507E-02	6.126E-02	3.507E-02	3.598E-02	8.848E-01	8.888E-01	***Max***	3.520E-03	3.748E-03	3.538E-02	6.122E-02	3.538E-02	3.630E-02	8.847E-01	8.891E-01
Min	3.393E-03	3.238E-03	3.463E-02	5.690E-02	3.463E-02	3.422E-02	8.822E-01	8.782E-01	***Min***	3.419E-03	3.209E-03	3.476E-02	5.665E-02	3.476E-02	3.433E-02	8.821E-01	8.795E-01
Std	3.817E-05	1.884E-04	1.332E-04	1.590E-03	1.332E-04	5.814E-04	8.220E-04	3.490E-03	***Std***	3.892E-05	1.852E-04	2.080E-04	1.574E-03	2.080E-04	6.263E-04	8.121E-04	3.337E-03
	BS=64, LR=1e-05									**BS=128, LR=1e-05**							
	MSE		**RMSE**		**MAE**		**R**			**MSE**		**RMSE**		**MAE**		**R**	
	Train	**Test**	**Train**	**Test**	**Train**	**Test**	**Train**	**Test**		**Train**	**Test**	**Train**	**Test**	**Train**	**Test**	**Train**	**Test**
Mean	3.469E-03	3.409E-03	3.502E-02	5.836E-02	3.502E-02	3.474E-02	8.836E-01	8.838E-01	***Mean***	3.468E-03	3.380E-03	3.497E-02	5.812E-02	3.497E-02	3.456E-02	8.834E-01	8.861E-01
Max	3.542E-03	3.745E-03	3.535E-02	6.120E-02	3.535E-02	3.628E-02	8.847E-01	8.872E-01	***Max***	3.533E-03	3.730E-03	3.535E-02	6.108E-02	3.535E-02	3.550E-02	8.845E-01	8.895E-01
Min	3.401E-03	3.093E-03	3.463E-02	5.561E-02	3.463E-02	3.316E-02	8.830E-01	8.792E-01	***Min***	3.390E-03	3.074E-03	3.473E-02	5.544E-02	3.473E-02	3.325E-02	8.825E-01	8.817E-01
Std	4.831E-05	2.237E-04	2.595E-04	1.909E-03	2.595E-04	9.367E-04	5.817E-04	2.818E-03	***Std***	4.445E-05	2.003E-04	1.737E-04	1.714E-03	1.737E-04	6.815E-04	7.184E-04	2.932E-03
	BS=256, LR=1e-05									**BS=512, LR=1e-05**							
	MSE		**RMSE**		**MAE**		**R**			**MSE**		**RMSE**		**MAE**		**R**	
	Train	**Test**	**Train**	**Test**	**Train**	**Test**	**Train**	**Test**		**Train**	**Test**	**Train**	**Test**	**Train**	**Test**	**Train**	**Test**
Mean	3.759E-03	3.747E-03	3.613E-02	6.118E-02	3.613E-02	3.621E-02	8.802E-01	8.817E-01	***Mean***	9.862E-03	1.002E-02	7.176E-02	1.001E-01	7.176E-02	7.185E-02	8.702E-01	8.698E-01
Max	3.838E-03	4.074E-03	3.647E-02	6.383E-02	3.647E-02	3.746E-02	8.809E-01	8.863E-01	***Max***	1.016E-02	1.046E-02	7.278E-02	1.023E-01	7.278E-02	7.302E-02	8.713E-01	8.734E-01
Min	3.685E-03	3.438E-03	3.573E-02	5.864E-02	3.573E-02	3.509E-02	8.790E-01	8.781E-01	***Min***	9.563E-03	9.535E-03	7.050E-02	9.765E-02	7.050E-02	7.029E-02	8.692E-01	8.659E-01
Std	5.152E-05	2.481E-04	2.535E-04	2.031E-03	2.535E-04	8.505E-04	5.111E-04	2.614E-03	***Std***	1.544E-04	2.997E-04	6.370E-04	1.501E-03	6.370E-04	7.988E-04	6.877E-04	2.556E-03
	BS=64, LR=1e-06									**BS=128, LR=1e-06**							
	MSE		**RMSE**		**MAE**		**R**			**MSE**		**RMSE**		**MAE**		**R**	
	Train	**Test**	**Train**	**Test**	**Train**	**Test**	**Train**	**Test**		**Train**	**Test**	**Train**	**Test**	**Train**	**Test**	**Train**	**Test**
Mean	1.006E-02	1.011E-02	7.339E-02	1.005E-01	7.339E-02	7.343E-02	8.684E-01	8.684E-01	***Mean***	1.321E-02	1.313E-02	7.911E-02	1.146E-01	7.911E-02	7.898E-02	8.679E-01	8.657E-01
Max	1.023E-02	1.115E-02	7.392E-02	1.056E-01	7.392E-02	7.508E-02	8.702E-01	8.722E-01	***Max***	1.340E-02	1.363E-02	7.971E-02	1.167E-01	7.971E-02	8.029E-02	8.704E-01	8.716E-01
Min	9.763E-03	9.500E-03	7.181E-02	9.747E-02	7.181E-02	7.226E-02	8.674E-01	8.612E-01	***Min***	1.306E-02	1.245E-02	7.847E-02	1.116E-01	7.847E-02	7.762E-02	8.664E-01	8.559E-01
Std	1.463E-04	5.479E-04	6.387E-04	2.707E-03	6.387E-04	8.863E-04	8.196E-04	3.175E-03	***Std***	1.133E-04	4.033E-04	3.345E-04	1.766E-03	3.345E-04	8.124E-04	1.148E-03	4.543E-03
	BS=256, LR=1e-06									**BS=512, LR=1e-06**							
	MSE		**RMSE**		**MAE**		**R**			**MSE**		**RMSE**		**MAE**		**R**	
	Train	**Test**	**Train**	**Test**	**Train**	**Test**	**Train**	**Test**		**Train**	**Test**	**Train**	**Test**	**Train**	**Test**	**Train**	**Test**
Mean	1.708E-02	1.745E-02	7.294E-02	1.321E-01	7.294E-02	7.339E-02	8.671E-01	8.658E-01	***Mean***	1.970E-02	2.012E-02	7.560E-02	1.418E-01	7.560E-02	7.666E-02	8.664E-01	8.667E-01
Max	1.739E-02	1.878E-02	7.359E-02	1.370E-01	7.359E-02	7.716E-02	8.689E-01	8.711E-01	***Max***	1.988E-02	2.084E-02	7.580E-02	1.444E-01	7.580E-02	7.827E-02	8.668E-01	8.678E-01
Min	1.659E-02	1.619E-02	7.199E-02	1.273E-01	7.199E-02	7.116E-02	8.657E-01	8.599E-01	***Min***	1.960E-02	1.932E-02	7.542E-02	1.390E-01	7.542E-02	7.503E-02	8.660E-01	8.648E-01
Std	2.844E-04	9.832E-04	5.117E-04	3.725E-03	5.117E-04	1.759E-03	9.986E-04	3.604E-03	***Std***	1.020E-04	5.722E-04	1.555E-04	2.018E-03	1.555E-04	1.181E-03	3.567E-04	1.404E-03

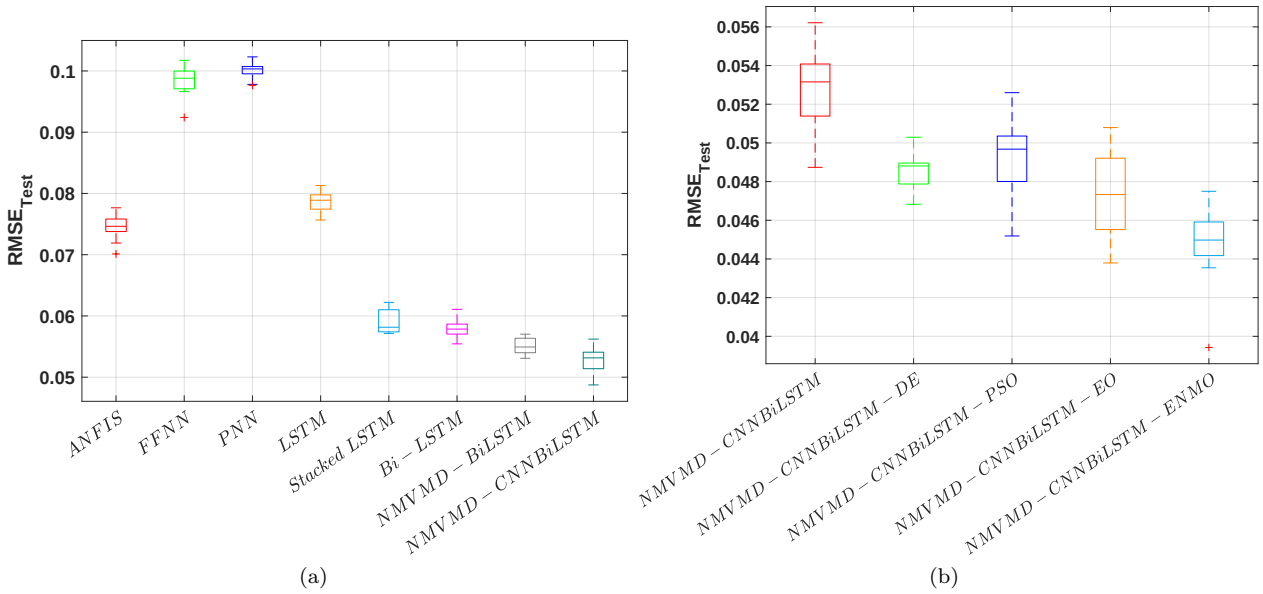

Figure 11: The comparison of the statistical results of the NMVMD-CNNBiLSTM wave model and other seven forecasting wave flux models (a). (b) And also a comparison among without and with hyper-parameters tuner, DE, PSO, EO, and ENMO.

4. Discussions

A fast and effective wave energy flux forecasting model was proposed. In order to evaluate the performance of the model, we used the wave features from Favignana Island, located in the Mediterranean Sea. The prediction interval is 6-hour ahead (long term). According to the forecasting results, the introduced wave predictive model performed the best compared to the other six standard machine learning methods. Five hybrid models were developed regarding accuracy and resilience. The second most significant observation is that combining the convolutional neural model (CNN) with the Bi-LSTM network improved the accuracy of the wave power prediction. This is mainly because CNN benefits from local spatial coherence in the sequential input data, which qualified CNNs to share fewer weights, especially fitting to extract relevant features with a lower computational budget than fully-connected networks. Furthermore, the proposed adaptive decomposition method enhanced the total performance of the prediction results considerably without imposing large demand on the computation. Finally, in order to find an optimal set of hyper-parameters of the proposed hybrid wave model, a new tuner (Equilibrium Nelder-Mead optimisation algorithm (ENMO)) was suggested. The prediction results demonstrated that ENMO provided a better training error than DE, PSO, and EO at 8.43%, 11.15%, and 6.00%, respectively.

Table 8: The statistical performance metrics for the proposed NMVMD-CNNBiLSTM model.

	ANFIS							
	MSE		**RMSE**		**MAE**		**R**	
	Train	**Test**	**Train**	**Test**	**Train**	**Test**	**Train**	**Test**
Mean	5.828E-03	5.535E-03	3.579E-02	7.437E-02	3.579E-02	3.532E-02	8.396E-01	8.467E-01
Max	5.955E-03	6.030E-03	3.622E-02	7.766E-02	3.622E-02	3.643E-02	8.438E-01	8.588E-01
Min	5.669E-03	4.918E-03	3.548E-02	7.013E-02	3.548E-02	3.345E-02	8.368E-01	8.363E-01
Std	7.528E-05	3.158E-04	1.970E-04	2.140E-03	1.970E-04	8.177E-04	2.242E-03	6.659E-03
	FFNN							
	MSE		**RMSE**		**MAE**		**R**	
	Train	**Test**	**Train**	**Test**	**Train**	**Test**	**Train**	**Test**
Mean	9.87E-03	9.69E-03	6.38E-02	9.84E-02	6.38E-02	6.36E-02	5.90E-01	5.96E-01
Max	1.04E-02	1.03E-02	6.46E-02	1.02E-01	6.46E-02	6.46E-02	5.97E-01	6.16E-01
Min	9.59E-03	8.54E-03	6.29E-02	9.24E-02	6.29E-02	6.16E-02	5.83E-01	5.79E-01
Std	2.15E-04	5.11E-04	5.55E-04	2.63E-03	5.55E-04	9.32E-04	4.56E-03	1.10E-02
	PNN							
	MSE		**RMSE**		**MAE**		**R**	
	Train	**Test**	**Train**	**Test**	**Train**	**Test**	**Train**	**Test**
Mean	9.862E-03	1.002E-02	7.176E-02	1.001E-01	7.176E-02	7.185E-02	8.202E-01	8.198E-01
Max	1.016E-02	1.046E-02	7.278E-02	1.023E-01	7.278E-02	7.302E-02	8.213E-01	8.234E-01
Min	9.563E-03	9.535E-03	7.050E-02	9.765E-02	7.050E-02	7.029E-02	8.192E-01	8.159E-01
Std	1.544E-04	2.997E-04	6.370E-04	1.501E-03	6.370E-04	7.988E-04	6.877E-04	2.556E-03
	LSTM							
	MSE		**RMSE**		**MAE**		**R**	
	Train	**Test**	**Train**	**Test**	**Train**	**Test**	**Train**	**Test**
Mean	6.011E-03	6.166E-03	5.065E-02	7.850E-02	5.065E-02	5.111E-02	7.757E-01	7.720E-01
Max	6.116E-03	6.609E-03	5.112E-02	8.129E-02	5.112E-02	5.275E-02	7.772E-01	7.846E-01
Min	5.893E-03	5.724E-03	5.023E-02	7.566E-02	5.023E-02	4.931E-02	7.726E-01	7.658E-01
Std	7.010E-05	2.891E-04	2.984E-04	1.847E-03	2.984E-04	9.465E-04	1.406E-03	5.639E-03
	Stacked LSTM							
	MSE		**RMSE**		**MAE**		**R**	
	Train	**Test**	**Train**	**Test**	**Train**	**Test**	**Train**	**Test**
Mean	3.570E-03	3.493E-03	5.975E-02	5.907E-02	3.518E-02	3.484E-02	8.824E-01	8.830E-01
Max	3.631E-03	3.871E-03	6.026E-02	6.221E-02	3.597E-02	3.742E-02	8.834E-01	8.879E-01
Min	3.497E-03	3.265E-03	5.914E-02	5.714E-02	3.442E-02	3.363E-02	8.813E-01	8.759E-01
Std	3.993E-05	2.356E-04	3.342E-04	1.980E-03	4.978E-04	1.097E-03	7.215E-04	3.958E-03
	Bi-LSTM							
	MSE		**RMSE**		**MAE**		**R**	
	Train	**Test**	**Train**	**Test**	**Train**	**Test**	**Train**	**Test**
Mean	3.468E-03	3.380E-03	3.497E-02	5.812E-02	3.497E-02	3.456E-02	8.834E-01	8.861E-01
Max	3.533E-03	3.730E-03	3.535E-02	6.108E-02	3.535E-02	3.550E-02	8.845E-01	8.895E-01
Min	3.390E-03	3.074E-03	3.473E-02	5.544E-02	3.473E-02	3.325E-02	8.825E-01	8.817E-01
Std	4.445E-05	2.003E-04	1.737E-04	1.714E-03	1.737E-04	6.815E-04	7.184E-04	2.932E-03
	NMVMD-BiLSTM							
	MSE		**RMSE**		**MAE**		**R**	
	Train	**Test**	**Train**	**Test**	**Train**	**Test**	**Train**	**Test**
Mean	2.77E-03	3.14E-03	5.26E-02	5.59E-02	2.63E-02	2.97E-02	9.10E-01	8.99E-01
Max	3.26E-03	4.26E-03	5.71E-02	6.53E-02	2.86E-02	3.25E-02	9.14E-01	9.07E-01
Min	2.53E-03	2.82E-03	5.03E-02	5.31E-02	2.47E-02	2.81E-02	9.01E-01	8.90E-01
Std	2.52E-04	4.20E-04	2.34E-03	3.54E-03	1.28E-03	1.50E-03	4.66E-03	5.45E-03
	NMVMD-CNN-BiLSTM							
	MSE		**RMSE**		**MAE**		**R**	
	Train	**Test**	**Train**	**Test**	**Train**	**Test**	**Train**	**Test**
Mean	2.51E-03	2.79E-03	5.01E-02	5.28E-02	2.53E-02	2.68E-02	9.16E-01	9.09E-01
Max	2.61E-03	3.16E-03	5.11E-02	5.62E-02	2.73E-02	2.93E-02	9.18E-01	9.19E-01
Min	2.43E-03	2.38E-03	4.93E-02	4.87E-02	2.39E-02	2.39E-02	9.13E-01	9.02E-01
Std	5.25E-05	2.29E-04	5.23E-04	2.18E-03	1.03E-03	1.42E-03	1.47E-03	4.73E-03

Table 9: The statistical prediction results of the proposed hybrid deep learning model compared with other three hybrid forecasting wave models.

	NMVMD-CNNBiLSTM-DE							
	MSE		**RMSE**		**MAE**		**R**	
	Train	**Test**	**Train**	**Test**	**Train**	**Test**	**Train**	**Test**
Mean	2.403E-03	2.360E-03	4.900E-02	4.857E-02	2.446E-02	2.399E-02	9.199E-01	9.218E-01
Max	2.595E-03	2.529E-03	5.095E-02	5.029E-02	2.650E-02	2.630E-02	9.244E-01	9.263E-01
Min	2.255E-03	2.193E-03	4.748E-02	4.683E-02	2.219E-02	2.230E-02	9.155E-01	9.181E-01
Std	1.117E-04	9.416E-05	1.140E-03	9.705E-04	1.512E-03	1.204E-03	3.131E-03	2.837E-03
	NMVMD-CNNBiLSTM-PSO							
	MSE		**RMSE**		**MAE**		**R**	
	Train	**Test**	**Train**	**Test**	**Train**	**Test**	**Train**	**Test**
Mean	2.457E-03	2.434E-03	4.956E-02	4.930E-02	2.394E-02	2.420E-02	9.181E-01	9.175E-01
Max	2.589E-03	2.767E-03	5.089E-02	5.260E-02	2.464E-02	2.595E-02	9.214E-01	9.244E-01
Min	2.305E-03	2.042E-03	4.802E-02	4.519E-02	2.336E-02	2.208E-02	9.160E-01	9.092E-01
Std	9.541E-05	1.991E-04	9.654E-04	2.038E-03	4.689E-04	1.075E-03	2.017E-03	4.658E-03
	NMVMD-CNNBiLSTM-EO							
	MSE		**RMSE**		**MAE**		**R**	
	Train	**Test**	**Train**	**Test**	**Train**	**Test**	**Train**	**Test**
Mean	2.349E-03	2.227E-03	4.846E-02	4.714E-02	2.353E-02	2.248E-02	9.215E-01	9.251E-01
Max	2.559E-03	2.580E-03	5.059E-02	5.079E-02	2.614E-02	2.489E-02	9.244E-01	9.344E-01
Min	2.227E-03	1.918E-03	4.720E-02	4.379E-02	2.242E-02	2.106E-02	9.159E-01	9.138E-01
Std	9.881E-05	2.105E-04	1.011E-03	2.230E-03	1.043E-03	1.211E-03	2.698E-03	6.242E-03
	NMVMD-CNNBiLSTM-ENMO							
	MSE		**RMSE**		**MAE**		**R**	
	Train	**Test**	**Train**	**Test**	**Train**	**Test**	**Train**	**Test**
Mean	2.216E-03	2.006E-03	4.707E-02	4.473E-02	2.214E-02	1.991E-02	9.263E-01	9.323E-01
Max	2.317E-03	2.256E-03	4.814E-02	4.750E-02	2.423E-02	2.182E-02	9.295E-01	9.447E-01
Min	2.116E-03	1.554E-03	4.600E-02	3.942E-02	2.068E-02	1.775E-02	9.237E-01	9.227E-01
Std	6.517E-05	1.947E-04	6.939E-04	2.239E-03	1.154E-03	1.322E-03	1.822E-03	5.937E-03

However, some potential concerns still merit more engagement in the suggested hybrid wave model. The described contemplations are summarised in the following.

1. Despite the fact that the adaptive decomposition method could decompose wave time-series data into optimal sub-signals with various frequencies efficiently and improve the developed model's total performance, there are many modern meta-heuristic algorithms that can also be used concerning speeding up the convergence velocity. Hence, these strategies can be evaluated in prospective studies.

2. According to the experimental prediction results, the proposed forecasting wave power model got the best accuracy compared with the other six machine learning methods and five hybrid predictive deep learning models. However, recently, various modern architectures of CNN have been proposed with a considerable performance, such as EfficientNet [50] and MobileNet [51]. Using such modern CNN architectures may improve the prediction accuracy of the wave energy flux. It should be noted that there is no straightforward approach to selecting an adequate combination of CNN layers and recurrent neural layers in order to develop a wave power predictor.

3. One of the most critical advantages of deep learning models is the ability to work in novel unseen datasets without any over or under-fitting issues. To test the generalisation capability of the proposed model, we used the multiple train-test splits by selecting an arbitrary split point for training and evaluation. However, the dataset applied in this study has been collected from one buoy. The predictor's performance should be assessed by other collected wave datasets from various geographies to extend the evaluation procedure.

5. Conclusions and perspectives

This study adapted the systematic strategy of developing a novel end-to-end deep learning-based model to forecast wave energy flux. The model combines a CNN with Bi-LSTM, a hybrid decomposition method (i.e.

NMVMD, and hyper-parameter optimisation (ENMO). The main contributions of this work are as follows: (1) a comparative framework was developed that incorporates six well-known machine learning methods such as ANFIS, FFNN, PNN, vanilla LSTM, stacked LSTM, and Bi-LSTM. The initial forecasting results indicated that Bi-LSTM performed better than other models. to identify the dominant wave parameters for predicting wave energy flux, we proposed six Bi-LSTM models featuring different input combination. The comparative analysis showed that the model using the two inputs of significant wave height (H_{m0}) and wave direction most accurately predicted wave energy flux. (3) We utilised VMD to decompose the original wave data into a series of sub-layers; and incorporated the NM search mechanism to optimise the performance of the VMD, and an improvement was reported at 25% compared with the Bi-LSTM. (4) The combined BiLSTM and NMVMD was developed using the customised convolutional layers because CNNs are highly noise-resistant prototypes and are able to extract instructive and profound features that are self-dependent from time. At the same time, (5) to address the hyper-parameters issue in the context of the CNN-BiLSTM model, we subsequently developed a new hyper-parameter optimisation method (ENMO), and compared its effectiveness with three other well-known optimisation methods (DE, PSO and EO). The achievements demonstrated that optimising the hyper-parameters of the proposed model using the ENMO algorithm could provide more accurate wave forecasting than employing other optimisation methods at 8.43%, 11.15%, and 6.00%, respectively.

This study's experimental results, allow the conclusion that the ENMO proposed novel hybrid wave forecasting model not only outperforms the original Bi-LSTM models but also delivers more reliable prediction of wave energy flux than the other six forecasting models for 6-hour forecasting intervals. To summarise with some perspectives, our new wave power forecasting approach is indeed flexible, potable, and paves the way for developing different predictors of wave characteristics. At the same time, what we accomplished for the wave energy flux can undoubtedly be transformed into other types of renewable energies, such as wind speed, wind power and solar radiation. Toward further investigations, research should consider incorporating the influence of climate factors by considering them as parameters in the proposed models. In order to evaluate the generalisation ability of the proposed model, we will consider various wave collected datasets for training and validating. Furthermore, combining other types of RNNs (e.g. gated recurrent unit and Hierarchical RNNs) and combined with CNNs may further improve the prediction capacity of this model framework.

Acknowledgements

The authors would like to represent their gratitude to the Civil Engineering Department of Catania University and Favignana Municipality for their cooperation to provide all data.

Appendix A. Variational mode decomposition (VMD)

The VMD is quasi-orthogonal and fully-intrinsic decomposition signal analysis approach, where VMD is able to decompose a sequential signal toward a set of modes including distinct bandwidth in spectral-domain non-recursively [52]. Every mode is compressed a core pulsation interpreted throughout the decomposition procedure. In order to adjust the modes bandwidth, there are some suggestions [53]: Firstly, applying a Hibert transform to harmonise a one-sided frequency spectrum toward every mode. Secondly, the assignment frequency spectrum of mode to base-band for combining the exponential tune to the identical estimated centre of frequency. Finally, concerning all modes, manage the bandwidth by employing the Gaussian smoothness of the demodulated signal. The main restrained variational problem is discussed in the following [54]

$$\min_{u_s, w_s} \left\{ \sum_{s=1}^{K} \left\| \partial_t \left[\left(D(t) + \frac{j}{\pi t} \right) \otimes u_s(t) \right] e^{-j w_s t} \right\|_2^2 \right\} \tag{A.1}$$

$$s.t. \sum_{s=1}^{K} u_s = F(t) \tag{A.2}$$

where $F(t)$ denotes the fundamental signal, t is related to the time script, s shows the whole figure of the modes, u_s implies the s_{th} mode, $D(t)$ describes a specific type of distribution called 'Dirac'. Moreover, the centre frequency signifies w_s, and also the convolution operator shows using $\otimes$. Besides, the low-frequency of sub-layers includes the mode with high-order. To reconstruct the considered optimisation problem to unconstrained, Lagrangian multipliers and the penalty function are concerned, which can be designated in Equation A.3.

$$\begin{aligned} L(u_s, w_s, \lambda) =& \alpha \sum_{s=1}^{K} \left\| \partial_t \left[\left(D(t) + \frac{j}{\pi t} \right) \otimes u_s(t) \right] e^{-jw_s t} \right\|_2^2 \\ &+ \left\| F(t) - \sum_{s=1}^{K} u_s(t) \right\|_2^2 + \left\langle \lambda(t).F(t) - \sum_{s=1}^{K} u_s(t) \right\rangle \end{aligned} \tag{A.3}$$

where the balancing parameter which is associated with the demanded data fidelity constraint can be denoted by α. The same unconstrained formulation in Equation A.3 is estimated by the Alternate Direction Method of Multipliers (ADMM). This method can get the saddle point of the augmented Lagrangian. A couple of directions to make the analysis of the VMD depends on the ADMM are applied to update both u_k and w_k. To figure out the mentioned optimisation problem for u_k, Equation A.4 is developed as follows:

$$\hat{u}_s^{n_d+1} = \frac{\hat{F}(w) - \sum_{i \neq s} \hat{u}_i(w) + (\hat{\lambda}(w)/2)}{1 + (2 \times \alpha(w - w_s)^2)} \tag{A.4}$$

where the total number of iterations represents by n_d, $\hat{F}(w)$, $\hat{u}_i(w)$,$\hat{u}_s^{n+1}$ and $\hat{\lambda}(w)$ are related to the Fourier transforms of $u(t)$, $F(t)$, $\lambda(t)$, and $u_s^{n_d+1}(t)$, respectively.

Appendix B. Performance indices of forecasting models

To evaluate and compare the efficiency of the hybrid forecasting models applied, four popular performance metrics were applied: the average square error (MSE), the root average square error (RMSE), average absolute error (MAE), and the Pearson correlation coefficient (R-value) [55]. The formulations for RMSE, MAE and R-value can be seen in the following:

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^{N} |F_e(i) - F_t(i)| \tag{B.1}$$

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (F_e(i) - F_t(i))^2} \tag{B.2}$$

$$R = \frac{\frac{1}{N} \sum_{i=1}^{N} (F_e(i) - \overline{F}_e)(F_t(i) - \overline{F}_t)}{\sqrt{\frac{1}{N} \sum_{i=1}^{N} (F_e(i) - \overline{F}_e)^2} \times \sqrt{\frac{1}{N} \sum_{i=1}^{N} (F_e(i) - \overline{F}_t)^2}} \tag{B.3}$$

where $F_e(i)$ and $F_t(i)$ ARE the estimated and true values at the i^{th} data sample. N is the total number of observed data samples. The variables $\overline{F}_e$ and $\overline{F}_t$ are the means of the estimated and observed power measures, respectively. With regard to develop the effectiveness of the estimated model, MSE, RMSE and MAE should be minimised, while R-value should be maximised.

References

[1] N. Markovska, N. Duić, B. V. Mathiesen, Z. Guzović, A. Piacentino, H. Schlör, H. Lund, Addressing the main challenges of energy security in the twenty-first century–contributions of the conferences on sustainable development of energy, water and environment systems, Energy 115 (2016) 1504–1512.

[2] B. Hrnčić, A. Pfeifer, F. Jurić, N. Duić, V. Ivanović, I. Vušanović, Different investment dynamics in energy transition towards a 100% renewable energy system, Energy 237 (2021) 121526.
[3] P. A. Østergaard, N. Duic, Y. Noorollahi, H. Mikulcic, S. Kalogirou, Sustainable development using renewable energy technology, Renewable Energy 146 (2020) 2430–2437.
[4] B. Jiang, X. Li, S. Chen, Q. Xiong, B.-f. Chen, R. G. Parker, L. Zuo, Performance analysis and tank test validation of a hybrid ocean wave-current energy converter with a single power takeoff, Energy Conversion and Management 224 (2020) 113268.
[5] D. Magagna, A. Uihlein, Ocean energy development in europe: Current status and future perspectives, International Journal of Marine Energy 11 (2015) 84–104.
[6] J. Widén, N. Carpman, V. Castellucci, D. Lingfors, J. Olauson, F. Remouit, M. Bergkvist, M. Grabbe, R. Waters, Variability assessment and forecasting of renewables: A review for solar, wind, wave and tidal resources, Renewable and Sustainable Energy Reviews 44 (2015) 356–375.
[7] C.-w. Zheng, Dynamic self-adjusting classification for global wave energy resources under different requirements, Energy 236 (2021) 121525.
[8] A. G. Majidi, B. Bingölbali, A. Akpınar, E. Rusu, Wave power performance of wave energy converters at high-energy areas of a semi-enclosed sea, Energy 220 (2021) 119705.
[9] S. Jin, R. J. Patton, B. Guo, Enhancement of wave energy absorption efficiency via geometry and power take-off damping tuning, Energy 169 (2019) 819–832.
[10] P. Hardy, B. Cazzolato, B. Ding, Z. Prime, A maximum capture width tracking controller for ocean wave energy converters in irregular waves, Ocean Engineering 121 (2016) 516–529.
[11] M. Neshat, The application of nature-inspired metaheuristic methods for optimising renewable energy problems and the design of water distribution networks, Ph.D. thesis (2020).
[12] B. Ding, B. S. Cazzolato, M. Arjomandi, P. Hardy, B. Mills, Sea-state based maximum power point tracking damping control of a fully submerged oscillating buoy, Ocean Engineering 126 (2016) 299–312.
[13] F. C. da Fonseca, R. Gomes, J. Henriques, L. Gato, A. Falcão, Model testing of an oscillating water column spar-buoy wave energy converter isolated and in array: Motions and mooring forces, Energy 112 (2016) 1207–1218.
[14] L. Wang, J. Isberg, E. Tedeschi, Review of control strategies for wave energy conversion systems and their validation: the wave-to-wire approach, Renewable and Sustainable Energy Reviews 81 (2018) 366–379.
[15] Y. Wang, L. Wang, Towards realistically predicting the power outputs of wave energy converters: Nonlinear simulation, Energy 144 (2018) 120–128.
[16] P. Bento, J. Pombo, R. Mendes, M. Calado, S. Mariano, Ocean wave energy forecasting using optimised deep learning neural networks, Ocean Engineering 219 (2021) 108372.
[17] P. C. Chu, S. E. Miller, J. A. Hansen, Fuel-saving ship route using the navy's ensemble meteorological and oceanic forecasts, The Journal of Defense Modeling and Simulation 12 (1) (2015) 41–56.
[18] R. Bell, B. Kirtman, Seasonal forecasting of winds, waves and currents in the north pacific, Journal of operational oceanography 11 (1) (2018) 11–26.
[19] R. Hashim, C. Roy, S. Motamedi, S. Shamshirband, D. Petković, Selection of climatic parameters affecting wave height prediction using an enhanced takagi-sugeno-based fuzzy methodology, Renewable and Sustainable Energy Reviews 60 (2016) 246–257.
[20] P. C. Deka, R. Prahlada, Discrete wavelet neural network approach in significant wave height forecasting for multistep lead time, Ocean Engineering 43 (2012) 32–42.
[21] K.-H. Lu, C.-M. Hong, Q. Xu, Recurrent wavelet-based elman neural network with modified gravitational search algorithm control for integrated offshore wind and wave power generation systems, Energy 170 (2019) 40–52.
[22] L. Li, Z. Yuan, Y. Gao, Maximization of energy absorption for a wave energy converter using the deep machine learning, Energy 165 (2018) 340–349.
[23] M. Pirhooshyaran, K. Scheinberg, L. V. Snyder, Feature engineering and forecasting via derivative-free optimization and ensemble of sequence-to-sequence networks with applications in renewable energy, Energy 196 (2020) 117136.
[24] Y. Li, T. Liu, Y. Wang, Y. Xie, Deep learning based real-time energy extraction system modeling for flapping foil, Energy 246 (2022) 123390.
[25] S. Zou, X. Zhou, I. Khan, W. W. Weaver, S. Rahman, Optimization of the electricity generation of a wave energy converter using deep reinforcement learning, Ocean Engineering 244 (2022) 110363.
[26] S. Yang, Z. Deng, X. Li, C. Zheng, L. Xi, J. Zhuang, Z. Zhang, Z. Zhang, A novel hybrid model based on stl decomposition and one-dimensional convolutional neural networks with positional encoding for significant wave height forecast, Renewable Energy 173 (2021) 531–543.
[27] T. Peng, C. Zhang, J. Zhou, M. S. Nazir, An integrated framework of bi-directional long-short term memory (bilstm) based on sine cosine algorithm for hourly solar radiation forecasting, Energy 221 (2021) 119887.
[28] M. A. Hemer, S. Zieger, T. Durrant, J. O'Grady, R. K. Hoeke, K. L. McInnes, U. Rosebrock, A revised assessment of australia's national wave energy resource, Renewable Energy 114 (2017) 85–107, wave and Tidal Resource Characterization. doi:https://doi.org/10.1016/j.renene.2016.08.039.
[29] M. Folley, The wave energy resource, Springer International Publishing, Cham, 2017, pp. 43–79. doi:10.1007/978-3-319-39889-1_3.
[30] The Specialist Committee on Waves, Final report and recommendations to the 23rd ITTC, in: Proceedings of the 23rd International Towing Tank Conference, Vol. II, 2002, pp. 505–736.
[31] C. Iuppa, L. Cavallaro, D. Vicinanza, E. Foti, Investigation of suitable sites for wave energy converters around sicily (italy), Ocean Science 11 (4) (2015) 543–557.

[32] C. Lo Re, G. Manno, G. Ciraolo, G. Besio, Wave energy assessment around the aegadian islands (sicily), Energies 12 (3) (2019) 333.
[33] H. Salehinejad, S. Sankar, J. Barfett, E. Colak, S. Valaee, Recent advances in recurrent neural networks, arXiv preprint arXiv:1801.01078 (2017).
[34] S. Xingjian, Z. Chen, H. Wang, D.-Y. Yeung, W.-K. Wong, W.-c. Woo, Convolutional lstm network: A machine learning approach for precipitation nowcasting, in: Advances in neural information processing systems, 2015, pp. 802–810.
[35] S. K. Yadav, K. Tiwari, H. M. Pandey, S. A. Akbar, Skeleton-based human activity recognition using convlstm and guided feature learning, Soft Computing (2021) 1–14.
[36] J. Bergstra, R. Bardenet, Y. Bengio, B. Kégl, Algorithms for hyper-parameter optimization, Advances in neural information processing systems 24 (2011).
[37] B. Shekar, G. Dagnew, Grid search-based hyperparameter tuning and classification of microarray cancer data, in: 2019 Second International Conference on Advanced Computational and Communication Paradigms (ICACCP), IEEE, 2019, pp. 1–8.
[38] M. Neshat, M. M. Nezhad, E. Abbasnejad, S. Mirjalili, D. Groppi, A. Heydari, L. B. Tjernberg, D. A. Garcia, B. Alexander, Q. Shi, et al., Wind turbine power output prediction using a new hybrid neuro-evolutionary method, Energy 229 (2021) 120617.
[39] A. Hinneburg, D. A. Keim, Optimal grid-clustering: Towards breaking the curse of dimensionality in high-dimensional clustering, 1999.
[40] A. Faramarzi, M. Heidarinejad, B. Stephens, S. Mirjalili, Equilibrium optimizer: A novel optimization algorithm, Knowledge-Based Systems 191 (2020) 105190.
[41] J. A. Nelder, R. Mead, A simplex method for function minimization, The computer journal 7 (4) (1965) 308–313.
[42] H. Liu, C. Yu, H. Wu, Z. Duan, G. Yan, A new hybrid ensemble deep reinforcement learning model for wind speed short term forecasting, Energy 202 (2020) 117794.
[43] J. Duan, P. Wang, W. Ma, X. Tian, S. Fang, Y. Cheng, Y. Chang, H. Liu, Short-term wind power forecasting using the hybrid model of improved variational mode decomposition and correntropy long short-term memory neural network, Energy 214 (2021) 118980.
[44] V. Krishna, S. Mishra, J. Naik, P. Dash, Adaptive vmd based optimized deep learning mixed kernel elm autoencoder for single and multistep wind power forecasting, Energy (2021) 122585.
[45] K.-H. Chang, Stochastic nelder–mead simplex method–a new globally convergent direct search method for simulation optimization, European journal of operational research 220 (3) (2012) 684–694.
[46] C. Bandt, B. Pompe, Permutation entropy: a natural complexity measure for time series, Physical review letters 88 (17) (2002) 174102.
[47] N. Somu, G. R. M R, K. Ramamritham, A hybrid model for building energy consumption forecasting using long short term memory networks, Applied Energy 261 (2020) 114131.
[48] D. Masters, C. Luschi, Revisiting small batch training for deep neural networks, arXiv preprint arXiv:1804.07612 (2018).
[49] H. M. I. Pousinho, V. M. F. Mendes, J. P. d. S. Catalão, A hybrid pso–anfis approach for short-term wind power prediction in portugal, Energy Conversion and Management 52 (1) (2011) 397–402.
[50] M. Tan, Q. Le, Efficientnet: Rethinking model scaling for convolutional neural networks, in: International conference on machine learning, PMLR, 2019, pp. 6105–6114.
[51] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, H. Adam, Mobilenets: Efficient convolutional neural networks for mobile vision applications, arXiv preprint arXiv:1704.04861 (2017).
[52] J. Duan, P. Wang, W. Ma, X. Tian, S. Fang, Y. Cheng, Y. Chang, H. Liu, Short-term wind power forecasting using the hybrid model of improved variational mode decomposition and correntropy long short-term memory neural network, Energy 214 (2021) 118980.
[53] H. Liu, X. Mi, Y. Li, Smart multi-step deep learning model for wind speed forecasting based on variational mode decomposition, singular spectrum analysis, lstm network and elm, Energy Conversion and Management 159 (2018) 54–64.
[54] X. Chen, Y. Yang, Z. Cui, J. Shen, Vibration fault diagnosis of wind turbines based on variational mode decomposition and energy entropy, Energy 174 (2019) 1100–1109.
[55] C. Zhang, J. Zhou, C. Li, W. Fu, T. Peng, A compound structure of elm based on feature selection and parameter optimization using hybrid backtracking search algorithm for wind speed forecasting, Energy Conversion and Management 143 (2017) 360–376.