DOI: https://doi.org/10.24113/ijoscience.v4i5.136

Correlation Enhanced Machine Learning Approach based Wave Height Prediction

Ruchi Shrivastava
DDIPG Scholar
Department of Electrical & Electronics Engineering
UTD RGPV
Bhopal India
ruchishrivastva@gmail.com

Dr. Krishna Teerth Chaturvedi
Professor
Department of Electrical & Electronics Engineering
UTD RGPV
Bhopal, India

Abstract: **The prediction of wave height is one of the major problems of coastal engineering and coastal structures. In recent years, advances in the prediction of significant wave height have been considerably developed using flexible calculation techniques. In addition to the traditional prediction of significant wave height, soft computing has explored a new way of predicting significant wave heights. This research was conducted in the direction of forecasting a significant wave height using machine learning approaches. In this paper, a problem of significant wave height prediction problem has been tackled by using wave parameters such as wave spectral density. This prediction of significant wave height helps in wave energy converters as well as in ship navigation system. This research will optimize wave parameters for a fast and efficient wave height prediction. For this Pearson's, Kendall's and Spearman's Correlation Coefficients and Particle Swarm Optimization feature reduction techniques are used. So reduced features are taken into consideration for prediction of wave height using neural network. In this work, performance evaluation metrics such as MSE and RMSE values are decreased and gives better performance of classification that is compared with existing research's implemented methodology. From the experimental results, it is observed that proposed algorithm gives the better prediction as compared to PSO feature reduction technique. So, it is also concluded that Co-relation enhanced neural network is better as compared to PSO based neural network with increased number of features.**

Keywords: Machine Learning, Classification, Wave height Prediction, Correlation Co-efficient, Accuracy.

I. INTRODUCTION

Wave energy transfer provides a convenient and natural concentration of wind energy within the waves. Wave energy generation refers to the energy of ocean surface waves and the utilization of that energy to generate electricity. The energy within a wave is dependent on the following factors; wind speed, duration of the wind blowing, the distance of open water that the wind has blown over (fetch), and water depth [1]. Wave power could be determined by wave height, wavelength, and water density. This mathematically could be described as in [2]:

$$P = \frac{\rho g}{64\pi} H^2 T \approx \frac{1}{2} H^2 T \; kW/m$$

Where, P is the wave energy flux per unit wave crest length (kW/m); q the mass density of the water (kg/m^3); g the gravitational gravity (m/s^2); H the wave height (m) and T is the wave time cycle (s). For example: for a 1.6 m wave and 10 s period, the power produced is approximately 12.8 kW/m.

Ocean waves' kinetic energy can be transformed into electricity by means of Wave Energy Converters (WEC), contributing this way to reduce our deep dependence on fossil fuels [3]. This type of marine facilities to obtain energy shows a clear potential for sustainable growth [4]: marine energy resources do not generate CO2 and reduce oil imports, a crucial geo-economical issue.

However, in spite of this potential, the use of marine energy sources is nowadays still minor at global level. In spite of this, wave energy plays a key role for sustainable development in several offshore islands because it provides not only technical and economical benefits (to satisfy the demands of clean electricity) but also without significant environmental impact, a key concern in offshore islands, committed to the protection of ecological systems [5]. Some interesting reviews of the most important issues involved in the generation of electricity from oceans (including converters, their related economical aspects, and the potential of a number of ocean regions to be exploited worldwide) can be found in [6]. Ocean waves are usually produced by wind action and are therefore an indirect form of solar energy. Wave energy uses Wave Energy Converters (WECs) to convert ocean energy into electricity.

WECs transform the kinetic energy of wind-generated waves into electricity by means of either the vertical oscillation of waves or the linear motion of waves, and exhibit some important advantages when compared to alternatives based on tidal converters, for example. Note however, that not all of the available wave energy resources can be realized as usable power, mainly due to various factors including socio-economics, the severe ocean environment, power conversion losses, and cost. Moreover, ocean waves are difficult to characterize, because its generation and propagation can be modelled as nonlinear processes. The real sea is a superposition of irregular waves trains which differ in period, height and direction. The local behavior of the sea estate can be represented by a spectrum, which specifies how the wave energy is distributed in terms of frequency and direction. As a consequence of this complexity, both the design, deployment, and control of WECs [7], become key topics that require a proper characterization and prediction of waves. Maybe, the most important wave parameters in this regard to characterize waves is the significant wave height (Hs), in which prediction this paper is focused on.

Wave height is one of the most important factors in coastal processes and coastal engineering studies. In case of energy extraction from waves, sediment movements, harbour design and soil erosion, wave height plays a vital role in these. Long- term observed data's are needed for all the practical applications. There are different methods for finding wave heights such as field measurements, theoretical studies and numerical stimulation [9,10,11]. But in most of these cases there won't be long-term measurements, so prediction of wave height is essential. Now day's soft computing based models have been used for wave prediction.

Soft computing based methods give results with high accuracy and time taken for training and prediction is very less compared to other traditional methods.

As mentioned above, the stochastic nature of the waves makes it very difficult to predict the wave energy resource, so research on this topic has been intense in recent years. From the point of view of machine learning approaches, one of the first wave height prediction algorithms proposed using artificial neural networks is used to obtain an accurate prediction of wave height. Alternative proposals based on different approaches have been recently proposed, with various soft calculation techniques being tested for wave height prediction [12, 13].

The objective of this research is to maximize the prediction speed of wave height by minimizing and selecting the optimal properties. Weather forecasts and energy production will benefit from the assessment of wave height with emphasis on the characteristics that result from this analysis.

II. PROPOSED METHODOLOGY

In the current scenario, an algorithm is proposed which provide a way to predict the significant wave height. Figure 1 shows the location of five buoy location i.e. "42056", "42057", "42058", "42059" and "42060" taken from National Oceanic and Atmospheric Administration (NOAA) [8]. This research work is intended to predict the significant wave height at 42058 buoy center. The yellow dots represents the buoy location. The wave describing parameters have been formally computed on the basis of spectral wave density measured hourly by NOAA at each buoy centers during 2017.

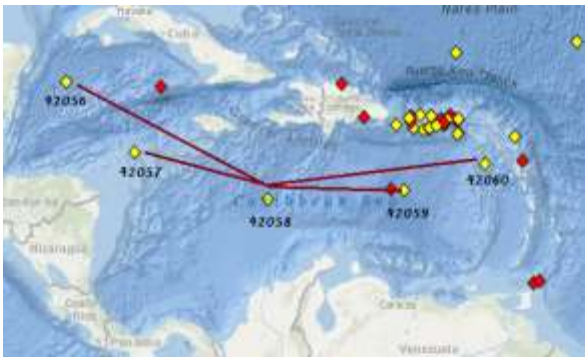

Figure 1: Map Representation of Buoy Location in Caribbean Sea

A. Wave Parameters

As mentioned above based on the hourly wave spectral density data collected at five different buoys, following 15 wave parameters are calculated in order to predict significant wave height. The main wave parameters are [14]:

Wave Spectral Moments

The wave spectral moments i.e. m_{-1}, m_0, m_1 and m_2, is calculated by using spectral moment as below:

$$m_n = \int_0^{\infty} f^n * S(f)df \qquad (1)$$

Where S(f) is the spectral density data from given buoy and n=-1, 0, 1, 2………..

They provide information on different statistical and physical characteristics of waves. For instance, m_0 is the variance of the wave elevation.

Wave Height

The significant wave height is calculated as :

$$H_s = 4(m_0)^{\frac{1}{2}} \qquad (2)$$

H_s the parameter related to the wave height that is most used in wave energy and in the design of ships and marine structures and coastal protection.

3.2.3 Peak Period

The peak period is defined as:

$$T_p = \frac{1}{f_p}$$

f_p is the maximum spectral peak frequency.

Wave Mean Period

It can be computed by using, among others, the estimator T_{xy} as:

$$T_{xy} = \frac{m_x}{m_y}$$

T_{xy} is an estimate of the average period used in the design of turbines for wave energy conversion. In this research work T_{m01}, T_{m02}, Tm_{-10} are calculated.

Goda's Peakedness Parameter

It is computed as:

$$Q_p = \frac{2}{m^2}\int_0^{\infty} f.S^2(f)df$$

Q_p has the potential to describe the statistical features of consecutive wave heights.

Longuet–Higgins Spectral Bandwidth

It is computed as:

$$v = \sqrt{\frac{m_2 * m_o}{m_1^2} - 1}$$

Quantifies the degree to which spectral energy spreads over the frequency range.

Wave Height Correlation Coefficient

It is computed as:

$$\gamma = \frac{\mathcal{E}(k) - \frac{(1-k^2)K(k)}{2} - \frac{\pi}{4}}{1 - \frac{\pi}{4}} \quad (7)$$

Where k, parameter which is calculated as:

$$k = |\frac{1}{m_0}\int_0^\infty S(f) * e^{i2\pi f T_c} df| \quad (8)$$

T_c is the estimator. For k_{01} and γ_{01} parameter, $T_c = T_{m01}$ and k_{02} and γ_{02} parameter, $T_c = T_{m02}$ are estimated.

B. Correlation Analysis

A bivariate analysis used for measuring the degree of association amongst two vectors say A and B is known as Correlation. In data mining, the value obtained after doing Correlation analysis varies between ±1. When this value is greater than 0, then a positive correlation exists and if this value is less than zero, then a negative correlation exists. If the value is 0, then the relationship between them is weak. For the proposed work that correlation value is selected whose value is positive one.

In this research work for feature selection Correlation Analysis is performed using Pearson, Spearman and Kendall coefficients which are explained in algorithm 1, algorithm 2, algorithm 3 and algorithm 4.

Pearson Correlation Analysis

Pearson correlation coefficient ρ is calculated by the formula as given below:

$$\rho = \frac{E[AD] - E[A]E[D]}{\sqrt{E[A^2] - (E[A])^2}\sqrt{E[D^2] - (E[D])^2}}$$

Where:
A stands for the Attribute Vector
D stands for the Decision Vector
E[A] stands for the sum of the elements in A

Algorithm 1: Pearson Correlation Analysis

procedure PEARSON(Dataset)
cols ← ncols(Dataset)
cols ← cols − 1
rows ← nrows(Dataset)
PearsonVector ← c()
for each i in 1:cols do
product ← 1
$SumA_i$ ← 0
Sum ← 0
$SumSquareA_i$ ← 0
SumSquare ← 0
for each j in 1:rows do
product ← product + (Dataset[j, 61] ∗ Dataset[j, i])
$SumA_i$ ← $SumA_i$ + Dataset[j, i]
Sum ← Sum + Dataset[j, i]
$SumSquareA_i$ ← $SumSquareA_i$ + $Dataset[j, i]^2$
SumSquare ← SumSquare + $Dataset[j, 61]^2$
end for

$$P = \frac{product - (SumAi * Sum)}{\sqrt{\text{SumSquareAi} - SumAi^2 * \sqrt{\text{SumSquare} - \text{Sum}^2}}} \quad (6)$$

PearsonVector ← append(PearsonVector, p)
end for
end procedure

Spearman Correlation Analysis

Spearman Correlation coefficient σ is calculated by the formula mentioned below:

$$\sigma = 1-(6\Sigma d_i^2)/n(n^2-1)$$

Where,
d_i stands for the difference between the ranks of variables P and Q
n stands for the sample size

Algorithm 2: Spearman Correlation Analysis

procedure SPEARMAN(Dataset)
n ← Number of rows
$d_d e$ ← $Dataset[, decision]_r$
SpearmanVector ← c()
for each i in 1 : ncol(Dataset) do
d ← 0
for each j in 1 : n do
d ← d + $(Dataset[, i]_r - d_d e)^2$
end for
spearman ← $\frac{6*d}{n(n^2-1)}$
spearman ← 1 − spearman
SpearmanVector ← append(SpearmanVector, spearman)
end for
end procedure

Kendall Correlation Analysis

Kendall Correlation coefficient τ is calculated by the formula as given below:

$$\tau = (n_c - n_d)/(1/2n(n-1))$$

Where,
d_i stands for the difference between the ranks of variables P and Q
n stands for the sample size

Algorithm 3: Kendall Correlation Analysis

procedure KENDALL(Dataset)
n ←Number of rows
vec ← Dataset[, decision]
n_c ← 0
n_d ← 0
for each attribute columns of Dataset do
for each i in 1 : n do
for each j in 1 : n do
if (($attr_i > attr_j$)AND($vec_i > vec_j$))
OR(($attr_i < attr_j$)AND($vec_i < vec_j$)) then
$n_c \leftarrow n_c + 1$
else
$n_d \leftarrow n_d + 1$
end if
end for end for end for
kendall ← $n_c - n_d$
kendall ← $\frac{2*kendall}{n(n-1)}$

end procedure

After doing Pearson Correlation by Algorithm 1, Spearman Correlation using Algorithm 2 and Kendall-rank Correlation by Algorithm 3, we get a list of attributes that satisfy the respective correlation criteria. After obtaining the three individual results which reduces the number of features using Algorithm 4 discussed below:

Algorithm 4: Attribute Selection after Correlation

procedure ATTRIBUTESELCTION(Dataset)
rows ← nrows(Dataset)
cols ← ncols(Dataset)
pearsonVector ← pearson(Dataset)
spearmanVector ← spearman(Dataset)
kendallVector ← kendall(Dataset)
for each i in 1:cols do
if pearsonVector[i]>0 AND spearmanVector[i]>0 AND kendallVector[i]>0 then
Selection ← true
else
Selection ← false
end if
end for
return dataset[,Selection]
end procedure

C. Particle Swarm Optimization Feature Reduction

The basic process of the PSO algorithm is given by:

Step 1: (Initialization) Randomly generate initial particles. For the PSO algorithm, the complete set of features is represented by a string of length N.

Step 2: (Fitness) Measure the fitness of each particle in the population. The selection of this fitness function is a crucial point in using the PSO algorithm, which determines what a PSO should optimize. Here, the task of the PSO algorithm is to find the global minimum value according to the definition of the fitness function. The definition of the fitness function for the basic method is simply the accuracy of detection.

Step 3: (Update) Compute the velocity of each particle.

Step 4: (Construction) For each particle, move to the next position.

Step 5: (Termination) Stop the algorithm if the termination criterion is satisfied; return to Step 2 otherwise.

D. Proposed Methodology

The proposed model is designed to predict the wave height at buoy location 42058 by analyzing the significant wave features at four different buoy locations i.e. 42056, 42057, 42059, and 42060. To predict the significant wave height at particular location proposed algorithm flow chart is shown in figure 2.

The proposed algorithm is divided into three sections i.e. data collection and preprocessing, feature reduction and finally forecasting using neural network. In first stage the real time data is collected from National Oceanic and Atmospheric Administration (NOAA) for five different buoy centers which is discussed above. Further the data is used to extract 15 different features using wave parameters. And further the extracted features are reduced by using co-relation analysis which determines that which wave parameters are more efficient to determine the significant wave height. And in last step neural network is used to forecast the wave height at buoy location 42058 in jan-feb 2018.

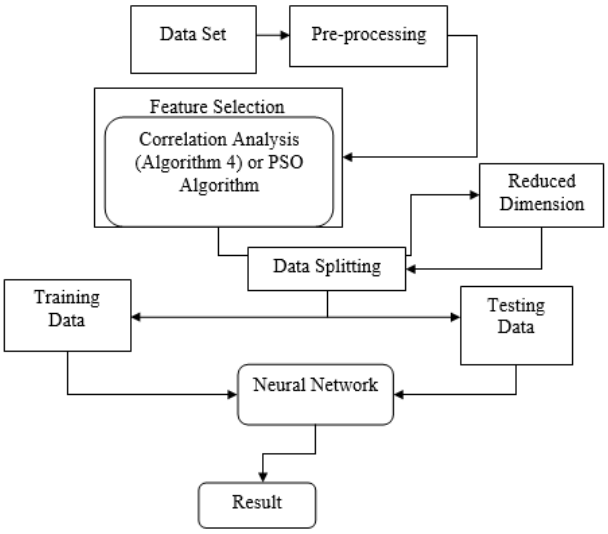

Figure 2: Flow Diagram of Proposed Methodology for Wave Height Prediction

E. Neural Network Forecasting

Neural networks are a class of nonlinear flexible models that can adaptively recognize models from data. In theory, it has been shown that neural networks can learn from experience in an appropriate number of nonlinear processing units and can estimate any complex functional relationship with great accuracy. Empirically, many successful applications have established their role in pattern recognition and prediction. Although many types of neural network models have been proposed, the most common model for time series prediction is the direct acting network model. Figure 3.3 shows a typical three-layer feedforward model used for prediction purposes. The input nodes are the previous delayed observations, while the output provides the forecast for the future value.

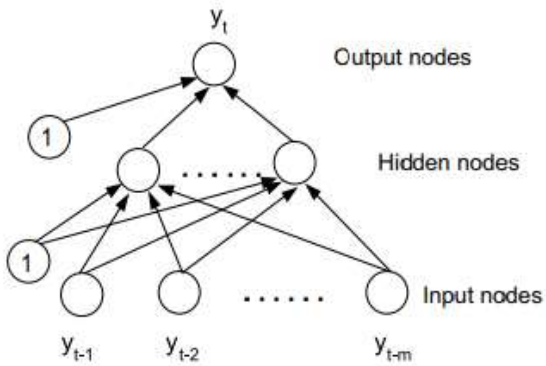

Figure 3: Neural Network

Hidden nodes with appropriate non-linear transfer functions are used to process the information received from the input nodes. The model can be written as:

$$y_t = \alpha_0 + (\sum_{i=1}^{m} \beta_{ij} y_{t-i} + \beta_{0j}) + \varepsilon_t \tag{9}$$

where m is the number of input nodes, n is the number of hidden nodes, f is a sigmoid transfer function, such as logistics:

$$f(x) = \frac{1}{1+\exp(-x)} * \{\alpha_j, j = 0,1,2,\ldots\ldots,n\} \quad (10)$$

is a vector of weights from the hidden to output nodes and $\{\beta_{ij}$, i=0,1, ... , m; j =1,2,….,n$\}$ are the weights of the input to the hidden nodes α0 and β0j the weights of the arcs resulting from the polarization terms whose values are always equal to 1. It should be noted that the linear transfer function it is used in the root node as desired for forecasting problems.

III. Result Analysis

A. Dataset Description

A news dataset from National Oceanic and Atmospheric Administration (NOAA) is prepared for this research work. The data is collected from buoy centers from Caribbean Sea. The location of five buoy location i.e. "42056", "42057", "42058", "42059" and "42060" taken from National Oceanic and Atmospheric Administration (NOAA). This research work is intended to predict the significant wave height at 42058 buoy centres.

B. Performance Parameters

Mean Square Error (MSE)

MSE of any estimator (classifier) measures the average squares of errors or deviations, i,e. the difference between the estimator and what is estimated. MSE is a risk function corresponding to the expected value of the squared error loss.

$$MSE = \frac{1}{N}(Target_{value} - Obtained_{value}) \quad (11)$$

Root Mean Square Error (RMSE)

RMSE is a parameter that determines the difference in squares between the output and the input.

$$RMSE = \sqrt{MSE}$$

As mentioned before, the evaluation parameters are mean square error (MSE) as well as root mean square error (RMSE). In this proposed algorithm the feature optimization or reduction technique is used with neural network which is applied on our real dataset. The result analysis describes that out of 15 parameters only seven parameters are co-related i.e. which have impact on deciding significant wave height of the waves.

After selecting efficient feature from the wave parameters, we have analysed by increasing number of features and calculated the MSE and RMSE. Table I shows the MSE as well as RMSE result analysis using neural network.

Table I:MSE and RMSE Parameter Evaluation under varying Features using Co-relation Analysis

MSE				
3 Features	4 Features	5 Features	6 Features	7 Features
0.149	0.1241	0.1244	0.1265	0.1236
RMSE				
0.386	0.3522	0.3527	0.3557	0.3481

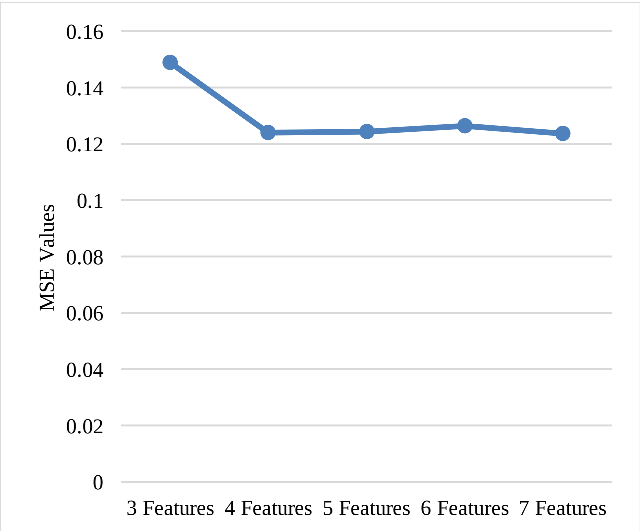

Figure 4: MSE Parameter Evaluation under Varying Features using Co-relation Analysis

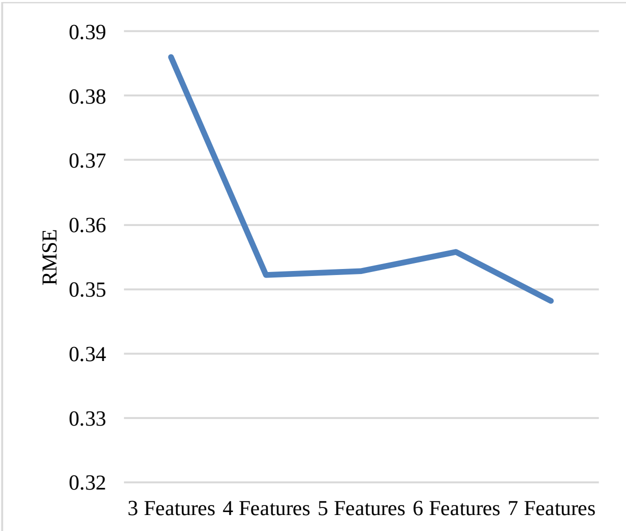

Figure 5: RMSE Parameter Evaluation under Varying Features using Co-relation Analysis

After selecting efficient feature from the wave parameters, we have analysed by increasing number of features and calculated the MSE and RMSE. Table II shows the MSE as well as RMSE result analysis using PSO-neural network.

Table II:MSE and RMSE Parameter Evaluation under varying Features using PSO

MSE				
3 Features	4 Features	5 Features	6 Features	7 Features
0.1156	0.1175	0.1165	0.1301	0.1316
RMSE				
0.386	0.3401	0.3428	0.3414	0.3607

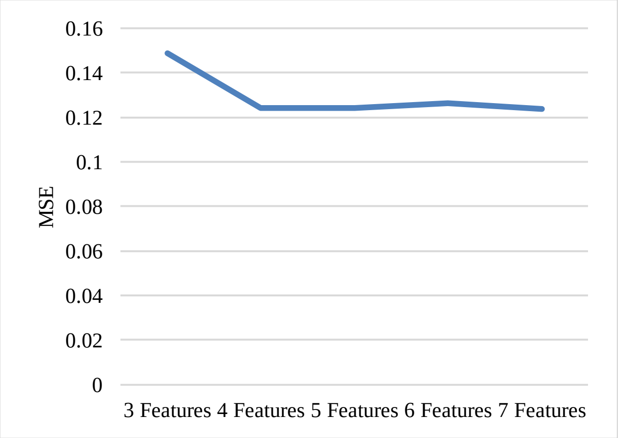

Figure 6: MSE Parameter Evaluation under Varying Features using PSO

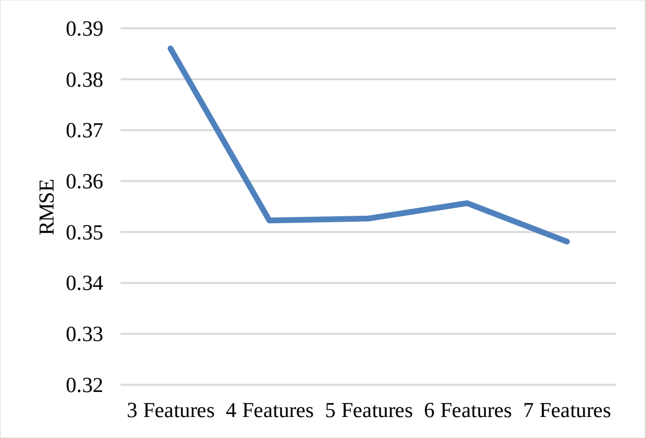

Figure 7: RMSE Parameter Evaluation under Varying Features using PSO

After analysing the two feature reduction techniques with neural network following conclusions are derived:

i. With increased number of features in PSO algorithm, MSE and RMSE increases.

ii. With increased number of features, Co-relation feature reduction technique reduces the MSE as well as RMSE.

iii. So, it is concluded that Co-relation enhanced neural network is better as compared to PSO based neural network with increased number of features.

IV. CONCLUSION

Ocean energy is clean and renewable, but has less environmental reserves. In this work, the use of artificial intelligence (AI) in the field of wave height prediction is studied. The advantages of applying KI techniques to energy problems are their great potential to handle a plethora of inaccurate or missing data. The research focused on the use of machine learning techniques to predict significant wave heights. The predictive power of a machine learning approach depends on the quality and size of all available data. Hybrid models offer better results than simple models. wind speed, air temperature, sea surface temperature and wind direction with minimal impact on wave height prediction. The combination of wind speed, air temperature and wind direction is all the most influential input parameters.

In this research, a problem of the prediction problem of the height of significant waves was solved using wave parameters such as wave spectral density. This significant wave height prediction helps both wave power converters and onboard navigation systems. This research will optimize wave parameters for a fast and efficient wave height prediction. For this Pearson's, Kendall's and Spearman's Correlation Coefficients and Particle Swarm Optimization feature reduction techniques are used. So reduced features are taken into consideration for prediction of wave height using neural network. In this work, performance evaluation metrics such as MSE and RMSE values are decreased and gives better performance of classification that is compared with existing research's implemented methodology. From the experimental results, it is observed that proposed algorithm gives the better prediction as compared to PSO feature reduction technique. So, it is also concluded that Co-relation enhanced neural network is better as compared to PSO based neural network with increased number of features. The future work will be intended to predict some other wave parameters as well as may be applied in order to predict the amount of energy generated by waves at buoy centers.

REFERENCES

[1] Adem Akpınar , Mehmet O¨ zger, Murat I˙hsanKo¨mu¨rcu, 2014. Prediction of wave parameters by using fuzzy inference system and the parametric models along the south coasts of the Black Sea. J Mar SciTechnol (2014) 19:1–14DOI 10.1007/s00773-013-0226-1.

[2] S. Salcedo-Sanza,n, J.C.Nieto Borge, L.Carro-Calvo a, L.Cuadra a, K.Hessner b, E. Alexandre, 2015. Significant wave height estimation using SVR algorithms and shadowing information from simulated and real measured X-band radar images of the sea surface. Ocean Engineering101(2015)244–253

[3] Pradnya Dixita,b, ShreenivasLondhec,n, YogeshDandawate, 2015. Removing prediction lag in wave height forecasting using Neuro - Wavelet modeling technique. Ocean Engineering93(2015)74–83.

[4] J.C. Fernández a, S. Salcedo-Sanzb,n, P.A. Gutiérrez a, E. Alexandre b, C. Hervás-Martínez, 2015. Significant wave height and energy flux range forecast with machine learning classifiers. Engineering Applications of Artificial Intelligence 43 (2015) 44–53.

[5] L. Cornejo-Bueno, J.C. Nieto Borgea, E. Alexandre, K. Hessner, S. Salcedo-Sanz, 2016. Accurate estimation of significant wave height with Support Vector Regression algorithms and marine radar images. Coastal Engineering 114 (2016) 233–243.

[6] L. Cornejo-Bueno a, J.C. Nieto-Borge a, P. García-Díaz a, G. Rodríguez b, S. Salcedo-Sanz, 2016. Significant wave height and energy flux prediction for marine energy applications: A grouping genetic algorithm e Extreme Learning Machine approach. Renewable Energy 97 (2016) 380e389.

[7] RoslanHashima, Chandrabhushan Roy, Shervin Motamedia, Shahaboddin Shamshirband, DaliborPetković , 2016. 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.

[8] http://www.ndbc.noaa.gov/

[9] Christos Stefanakos, 2016. Fuzzy time series forecasting of nonstationary wind and wave data. Ocean Engineering121(2016)1–12.

[10] Pradnya Dixita, Shreenivas Londhe, 2016. Apor Prediction of extreme wave heights using neuro wavelet technique. Applied Ocean Research 58 (2016) 241–252.

[11] S. Salcedo-Sanza,n, J.C.Nieto Borge, L.Carro-Calvo a, L.Cuadra a, K.Hessner b, E. Alexandre, 2015. Significant wave height estimation using SVR algorithms and shadowing information from simulated and real measured X-band radar images of the sea surface. Ocean Engineering101(2015)244–253.

[12] Christos Stefanakos, 2016. Fuzzy time series forecasting of nonstationary wind and wave data. Ocean Engineering121(2016)1–12.

[13] Pradnya Dixita, Shreenivas Londhe, 2016. Apor Prediction of extreme wave heights using neuro wavelet technique. Applied Ocean Research 58 (2016) 241–252.

[14] E. Alexandre, L. Cuadra, J.C. Nieto-Borge, G. Candil-García, M. del Pino, S. Salcedo-Sanz, “A hybrid genetic algorithm—extreme learning machine approach for accurate significant wave height reconstruction”, sevier, (2015).