# Machine Learning Approach towards Engineering Vegetation as Coastal Protection Measures

N Hari Ram*1 [0009-0000-7750-4790], S Sangeetha*2 [0009-0007-3102-4573], V Sriram1 [0000-0003-3586-9577] and Abhilash Somayajula2 [0000-0002-5654-4627]

Note: *Joint First Authors

[1] Department of Ocean Engineering, Indian Institute of Technology Madras, Chennai, Tamil Nadu 600036, India

[2] Marine Autonomous Vessels (MAV) Laboratory, Department of Ocean Engineering, Indian Institute of Technology Madras, Chennai, 600036, Tamil Nadu, India

**Abstract.** The present study investigates the application of Artificial Neural Networks (ANN) to address the problem of dissipative coefficient ($C_D$) prediction in the modeling of the wave-vegetation interaction process. The key parameters involved in the wave vegetation studies were identified through systematic correlation studies from the experimental datasets. Pearson, Spearman, and Principal Component Analysis (PCA) were used to determine the contribution of the input features toward $C_D$ prediction for the ANN model. Based on the understanding from our experimental studies, the modified submergence ratio, single-cylinder inertial coefficient ($C_m^*$), and drag coefficient ($C_d^*$) are included in the input features from the literature. The correlation coefficient and PCA analysis suggest that these features contribute significantly to the overall variance in the experimental datasets. This provides motivation to further investigate their potential impact on the predictive model and assess whether incorporating them enhances model performance. The machine learning (ML) approach was leveraged to analyze the effect of including these additional features as the model inputs in the prediction of dissipative coefficient $C_D$. This study shows the performance and effect of these features included in the model prediction compared to the existing empirical model. Furthermore, this model can be a potential tool supporting the macroscopic numerical studies related to wave vegetation interaction studies.

**Keywords:** Dissipative coefficient, Vegetation, Coastal Protection, Machine Learning.

## 1. Introduction

The coastal vegetation has been regarded as an eco-based coastal protection measure. In recent decades, investigations of the protective role of vegetation patches have been extensively carried out owing to field observation, which reported the significant role in protecting the coastline from extreme waves (Danielsen et al., 2005; Sundar et al., 2007; Tanaka, 2009). The studies related to the wave-vegetation interaction started with Dalrymple et al. (1984). An analytical formulation for evaluating the reduction in wave height as a wave passes through the vegetation belt on a horizontal bed was proposed. (Mendez & Losada, 2004) extended the works of (Dalrymple et al., 1984) by including the variation of bed slope, irregular waves and breaking waves. The significance of the vegetal parameters has been invested by several researchers, which includes vegetation density (Nepf, 1999; Struve et al., 2003; W. C. Wu & Cox, 2015), length of vegetation patch (Hari Ram, Sriram, et al., 2022; Noarayanan et al., 2012; Phan et al., 2019) and submergence (Augustin et al., 2009; Busari & Li, 2015; Maza et al., 2016). A compiled review of the vegetal characteristics, hydrodynamic conditions and the effectiveness of vegetation has been provided in detail in the literature (Hari Ram & Sundar, 2024; Sriram et al., 2023; Tanaka, 2009).

Further, several studies on wave vegetation interaction were carried out through macroscopic numerical modelling approaches. Maza et al. (2015) modelled the propagation of solitary waves over vegetation patches using IHFOAM in OpenFOAM. Chen and Zou (2019) studied the effects of suspended/ floating canopies on wave-driven currents using SWASH (a non-hydrostatic model). Divya et al. (2020) carried out studies by modelling the vegetation patch as a porous medium using a particle-based numerical method - Improved Meshless Local Petrov Galerkin using Rankine source (IMLPG-R). Yin et al. (2024) used an intercom solver in OpenFOAM to investigate breaking waves over vegetated slopes. Apart from these studies several numerical models such as Coastal and Estuarine Storm Tide (CEST), Xbeach, Delft3D, Advanced CIRCulation (ADCIRC), Finite Volume Community Ocean Model (FVCOM) have been widely used to evaluate the protective role of vegetation (Figueroa-Alfaro et al., 2022; K. Hu et al., 2015; Liu et al., 2013; Zhang & Zhang, 2023). All the above studies investigate vegetation's effects by explicitly modelling its effects through dissipative coefficients. However, predicting this dissipative coefficient ($C_D$) is not straightforward. Previous studies have relied on empirical equations or trial-and-error methods to evaluate it. A comprehensive overview of the empirical relationships presented in the literature for the prediction of the dissipative coefficient is listed in Table 1. However, the limitation of the empirical approach is the high degree of deviation (Maza et al., 2013; Wang et al., 2023), whereas the limitation of trial-and-error approach is dependent on experimental data for its evaluation (Rosenberger & Marsooli, 2022; Suzuki et al., 2019; Zhang et al., 2013). Furthermore, the dissipative coefficient itself is influenced by multiple interdependent factors, such as vegetation density, length of the vegetation patch, submergence, and flow characteristics. As a result, there remains a significant gap in achieving accurate, reliable predictions of the dissipative coefficient.

**Table 1.** Proposed empirical equations for predicting the dissipative coefficient in the literature with $R^2$ value**.**

| Literature | Empirical Formula | $R^2$ Value \| Remarks |
|---|---|---|
| Wu et al. (1999) | $C_D = \frac{(3.44 \times 10^6)S^{0.5}}{R}$<br>$C_D = \frac{(4.50 \times 10^{10})S^{0.89}T^{1.09}}{R^{1.7}}$ | 0.998 \| Emerged condition<br>0.958 \| Submerged condition |
| Mendez & Losada (2004) | $C_D = 0.47exp(-0.052K)$<br>$C_D = exp(-0.0138Q)Q^{0.3}$ | 0.76 \| KC number<br>0.92 \|Modified KC number |
| Maza et al. (2013) | $C_D = \left(\frac{\gamma}{Re}\right)^{\beta} + \alpha$ | 0.49 \| for $\alpha, \beta, \gamma =$ $0.87, 0.88, 2200$ respectively |
| Hu et al. (2014) | $C_D = \left(\frac{\gamma}{Re}\right)^{\beta} + \alpha$ | 0.49 \| for $\alpha, \beta, \gamma =$ $1.04, 1.37, 730$ respectively |
| He et al. (2019) | $C_D = 18.025\, exp(-0.043KC)$<br>$C_D = \frac{51.456\, exp(-0.02QC)}{QC^{0.457}}$<br>$C_D = \frac{23.326\, exp(-0.601Wr)}{Wr^{0.24}}$ | 0.735 \| KC number<br>0.817 \| Modified KC number<br>0.85 \| Modified Ursell number |
| Wang et al. (2022) | $C_D = \left(\frac{0.77}{KC_{rv}}\right)^{0.41} + 0.42$ | 0.73 \| Modified KC number |
| Wang et al. (2023) | $C_D = 0.57\left(\frac{H}{L}\right)^{-0.16}\left(\frac{d}{L}\right)^{-0.21}\lambda_f^{-0.13}\left(\frac{hv}{d}\right)^{0.41}\left(\frac{L}{Lv}\right)^{0.41}$ | 0.42 \| Based on Multivariate Non-linear regression |

The present study focuses on providing a solution methodology for engineering coastal vegetation by incorporating knowledge from traditional modeling approaches for input feature selection and using Machine Learning (ML) techniques for prediction. By providing the input features based on the understanding of the correlation between input and output features, the ML model becomes a powerful tool. Unlike traditional methods, ML is a data-driven approach that learns input-output relationships from patterns in data, often without incorporating underlying physical principles. Traditional methods, including experimental and numerical models, have significantly contributed to understanding the wave-vegetation interaction process. However, due to the limitations of these approaches, as discussed above, machine learning approaches are being explored to make predictions of the dissipative coefficient using measurements of other known parameters. ML approaches have also shown promising results in various ocean and coastal engineering applications by utilizing their ability to learn from data, identify patterns, and make predictions (J Chen et al., 2021; Elmar Balas et al., 2004; Panda, 2023; Sangeetha et al., 2024). Specifically for regression problems, various traditional ML algorithms such as Linear regression, LASSO regression and Support Vector Machine (SVM) have shown promising results in engineering applications (Deogaonkar et al., 2023; Kim & Lee, 2022). However, capturing non-linear relationships between input features and the output quantities is challenging. ML-based Artificial Neural Networks (ANNs), such as Feedforward Neural Networks (FFNN), have the potential to achieve reliable predictions. The ANN models are flexible and are ideal for problems that are associated with complex parameters having random distribution (Fan et al., 2020). ANNs have been shown to be universal function approximators; hence, this study focuses on using the ANN model for the prediction of dissipative coefficients ($C_D$). Previous studies, such as (Hari Ram & Sundar, 2024; Sriram et al., 2023), have demonstrated the importance of incorporating various hydrodynamic and vegetation characteristics into predictive models for coastal processes. Additionally, works by (Wang et al., 2023; Malvin et al., 2020) have explored the application of artificial neural networks (ANNs) in predicting dissipative coefficients, highlighting their potential for improving coastal protection strategies. Building upon these efforts, this study proposes the inclusion of three additional parameters as input features in an ANN model, providing a systematic analysis through correlation coefficients and principal component analysis (PCA). While PCA identifies features based on variance rather than direct predictive relevance, including these parameters among the top features highlights their significant contribution to the dataset's variability. Correlation analysis helps uncover linear and monotonic relationships between input parameters, minimizing redundancy and motivating further investigation into the role of these three additional parameters in model performance and whether their inclusion leads to any improvement.

A feedforward neural network (FFNN) is employed to analyze the effect of incorporating these additional features on $C_D$ prediction. The network is trained with different sets of input features to examine the influence of hydrodynamic effects, vegetation characteristics, and their combined impact on $C_D$. The model is optimized through grid search for hyperparameter tuning, with early stopping applied to prevent overfitting. Neural network models are trained with cross-validation, and to evaluate the model's generalization and predictive performance, a separate testing dataset, never encountered during training, is used. The goal is to assess whether incorporating these additional input parameters enhances model performance. By integrating machine learning with a structured feature selection approach, this study aims to improve the prediction of the dissipative coefficient, ultimately contributing to better coastal vegetation management and more effective coastal protection strategies.

The paper is organized as follows: Section 2 outlines the methodology of the present study, including datasets and feature selection. Section 3 describes the architecture and implementation of the FFNN model. Section 4 presents and discusses the results. The final section provides a summary and suggests future directions for the study.

## 2. Methodology

This study employs an FFNN to predict the vegetation belt's vegetation dissipative coefficient ($C_D$) in wave-vegetation interactions. FFNNs are a type of deep learning model that can effectively capture nonlinear relationships and patterns within the data. The dissipative coefficient ($C_D$) derived from the experimental studies (Hu et al., 2021; Ozeren, 2011) was input for FFNN model training. Fig. 1 shows the significant parameters associated with the dissipative coefficient involved in the wave-vegetation interaction process under regular waves. Several non-dimensional parameters associated with vegetation belts' attenuation process are derived. The functional expression for the dissipative coefficient is given in Eq. (1).

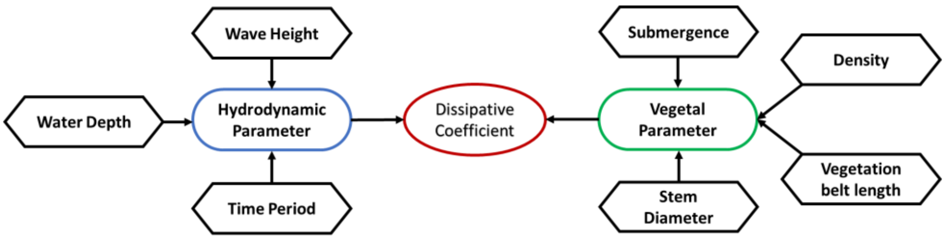



**Fig. 1.** Parameters involved in wave-vegetation interaction contributing towards the dissipative coefficient**.**

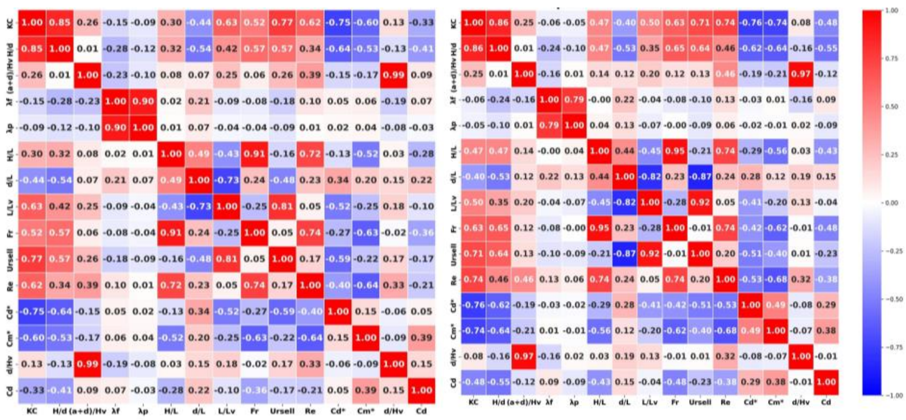
$$C_D = \begin{cases} \left(\frac{H}{d}, \frac{H}{L}, \frac{d}{L}, Fr,\ Ur\right) - Hydrodynamic \\ (\lambda_p, \lambda_f) - Vegetal \\ \left(Re,\ Kc, \frac{(a+d)}{H_v}, \frac{L}{L_v}, C_d^*, C_m^*\right) - Hydrodynamic\ + Vegetal \end{cases} \tag{1}$$

where $H$ = wave height, $d$ = water depth, $L$ = wave length, $\frac{H}{d}$ = relative wave height, $\frac{H}{L}$ = relative steepness, $\frac{d}{L}$ = relative water depth, $Fr$ = Froude number, $Ur$ = Ursell number, $\lambda_p$ = vegetation plan area $\lambda_f$= vegetation surface area, $Re$ = Reynolds number, $Kc$ = Keulegan Carpenter number, $H_v$ = height of vegetation, $a$ = wave amplitude, $\frac{(a+d)}{H_v}$ = proposed submergence ratio, $L_v$ = length of vegetation belt, $\frac{L}{L_v}$ = relative wavelength, $C_d^*$= drag coefficient of a single cylinder, $C_m^*$= inertial coefficient of a single cylinder.

## 2.1. Dataset and Feature Selection

From the obtained experimental dataset, a subset of input features was used in training different ML models for the prediction of dissipative coefficient to understand how the inclusion of these features affects the model prediction. These features from Eq. 1 are all non-dimensional parameters, and they are considered in such a way that the influence of hydrodynamic effects, vegetation and its combined effects are captured through these features. Prior studies by Wang et al. (2023) have included most of these features from the above-mentioned list, apart from $C_d^*$, $C_m^*$ and the proposed submergence ratio. This study provides a systematic analysis of the inclusion of these three additional parameters using correlation coefficients and principal component analysis.

**Correlation Coefficient analysis:** Three correlation coefficients were analyzed to comprehensively understand the relationship between the input features and the dissipative coefficient. The Pearson correlation coefficient was used to capture the linear relationships, the Spearman Rank coefficient was used to capture the monotonic relationships, and Principal Component Analysis (PCA) was used for ranking the input features with respect to the dissipative coefficient.
Fig. 2 shows the heatmaps of both the correlation coefficients.




(a) Pearson correlation (b) Spearman correlation

**Fig. 2.** Correlation analysis heat map for the dissipative coefficient with the non - dimensional parameters. (a) shows the correlation based on Pearson correlation and (b) shows the correlation based on Spearman correlation. The colour gradient from red – white – blue shows the nature of the correlation, positive – negative

It can be inferred from the heat map of Pearson Correlation that there exists a significant correlation (> 0.25) between the dissipative coefficient with Keulegan Carpenter number (Kc ), Relative wave steepness, Relative wave height, Froude Number and $C_m^*$. Several studies related to wave vegetation interaction have also reported the significance of these parameters (Hari Ram & Sriram, 2024; He et al., 2019; Mendez & Losada, 2004).

The Spearman correlation study provides us with the monotonic relationship between two given parameters. The Spearman correlation strength (> 0.25) for the $Kc$ , Relative wave height, Relative wave steepness, Froude Number, Reynolds Number, $C_d^*$ and $C_m^*$ also proves a strong dependence on the dissipative coefficient. The experimental study by Hari Ram and Sriram (2024) observed the same and reported the significance of $C_d^*$ and $C_m^*$. The correction studies also support the experimental observation. However, the Spearman correlation for submergence is very low, which signifies there is no monotonic relation between the dissipative coefficient. This aligns with the observed nature of wave attenuation characteristics of vegetation belts. The dissipative of wave energy shows a positive trend for emergent conditions and is higher near the tip of the vegetation (Augustin et al., 2009; Busari & Li, 2015; Hari Ram, Sriram, et al., 2022). As the water level increases and the flow occurs over submerged vegetation, the dissipative characteristics of the vegetation belt are reduced (Hari Ram & Sundar, 2024; Li et al., 2018; Wu et al., 2023).

It can also be noted from the Spearman Correlation analysis that the proposed submergence ratio $(a + d)/H_v$ by Hari Ram et al. (2022) has a slightly higher correlation with the dissipative coefficient than $d/H_v$. The proposed submergence ratio considers the effect of the wave along with water depth and vegetation height. However, this is in contrast to the observation from Pearson Correlation, where the value for the proposed submergence ratio is slightly lower than that of $d/H_v$. This suggests that the relation between the submergence ratio and dissipative coefficient is more towards nonlinearity. The observation from the correlation analysis thus motivates the inclusion of $C_d^*$ , $C_m^*$ and the proposed submergence ratio in the prediction of the dissipative coefficient.

**Principal Component analysis (PCA):** PCA is a dimensionality reduction technique which aims to capture the direction of maximum variance from the input feature space (Rahmat et al., 2023). In this study, PCA was used to calculate the top principal component modes required to explain 95% variance in the input data using all the input features. The importance of each of the original input features was then estimated by calculating the contribution of variance due to each feature in the top principal modes. This importance was calculated using the weighted sum of the loadings of each of the input features on the principal modes, weighted by the explained variance ratio of the corresponding Principal Component (PC) as shown in Eq. (2). The top 7 features based on their importance rankings were $Fr$ , $Re$ , $L/L_v$, $d/L$, $C_m^*$, $C_d^*$ and $Ur$.

$$Feature\ Importance = \sum_{i=1}^{7} (|Loading\ for\ Feature\ in\ PCi\ x\ Explained\ Variance\ Ratio\ of\ PC \quad (2)$$

Even though this analysis does not guarantee that $C_m^*$ and $C_d^*$ will directly improve the prediction of $C_D$, as PCA identifies variance in the input space rather than predictive relevance. However, their inclusion in the top seven features suggests that they contribute significantly to the overall variance in the dataset. This also provides motivation to further investigate their potential impact on the predictive model and assess whether incorporating them enhances model performance.

The next section details the ML approach used for analyzing the effect of including these additional features in the prediction of dissipative coefficient $C_D$.

## 3. Machine learning approach

The neural network model is trained to map the input non-dimensional parameters associated with the attenuation process in vegetation belts to the dissipative coefficient. This function mapping enables the estimation of dissipative coefficients without relying on empirical equations or trial-and-error methods. The network is trained with different sets of input features to examine the influence of hydrodynamic effects, vegetation characteristics, and their combined impact on the dissipative coefficient. For neural network training, this study uses the same input features as those considered in prior research by Wang et al. (2023). Additionally, this study incorporates three extra parameters into the input feature set as discussed in the previous section. A detailed description of the data points and input features used for training the neural network model is provided in Table 2.

**Table 2.** Details of data points and input features used for FFNN model training and testing

| Input feature selection | Number of inputs features | Input features | Total number of data points | Train and test split (ratio) |
|---|---|---|---|---|
| Reference (Wang et al. 2023) | 11 | $Kc, \frac{H}{d}, \frac{d}{H_v}, \lambda_f, \lambda_p, \frac{H}{L}, \frac{d}{L}, \frac{L}{L_v}, Fr, Ur, Re$ | 564 | 9:1 |
| Proposed | 13 | $Kc, \frac{H}{d}, \frac{(a+d)}{H_v}, \lambda_f, \lambda_p, \frac{H}{L}, \frac{d}{L}, \frac{L}{L_v}, Fr, Ur, Re, C_d^*, C_m^*$ | 564 | 9:1 |

The FFNN model was implemented using the PyTorch framework. The datasets collected from the experimental study (Z. Hu et al., 2021; Ozeren, 2011) were split into training and testing sets with random shuffle, with the train-test ratio provided in Table 2. Before training, the inputs were normalized using MinMax scaler to scale the features within a specified range of [0, 1].

K-fold cross-validation was applied to the training data to assess the model's robustness to avoid overfitting. The training data were randomly shuffled before splitting into k folds. Specifically, the training dataset was partitioned into training and validation sets k times in a (k−1):1 ratio, ensuring that each iteration used a different validation subset. The model was trained on the training portion and validated against the corresponding validation set. In this study, 10-fold cross-validation was performed, and the corresponding loss and performance metrics were recorded. The model's accuracy during cross-validation was assessed using average Mean Squared Error (MSE) and $R^2$ over the 10 folds.

Furthermore, to prevent overfitting, an early stopping mechanism with a patience of 20 epochs was incorporated. If the validation loss did not improve for 20 consecutive epochs, training was halted, and the best-performing model state was saved. This ensured optimal generalization by selecting the model configuration that achieved the lowest validation loss.

The network was tuned using grid search hyperparameter optimization with 10-fold cross-validation, exploring various configurations of hidden layer sizes, learning rates, dropout rates, activation functions and batch sizes. The best FFNN architecture was selected based on the configuration that yielded the lowest validation loss (MSE) and highest performance metric ($R^2$) thereby optimizing the model's performance. The specifications of the hyperparameters used for grid search and training of the neural network are given in Table 3. The chosen architecture was then trained on a complete training dataset and evaluated with the test dataset, as shown in Fig. 3.

| **Hyperparameters** | **Configuration** | **Best Configuration** |
|---|---|---|
| Size of the first hidden layer 1 | 8, 16, 32, 64 | 64 |
| Size of the second hidden layer 2 | 4, 8, 16, 32 | 32 |
| Size of the third hidden layer 3 | 2, 4, 8 | 8 |
| Non-Linear Activation Function | ReLU and LeakyReLU | LeakyReLU |
| Optimization algorithm | Adam | Adam |
| Learning rate for the optimizer | 0.01, 0.001 | 0.001 |
| Weight initialization scheme | Glorot | Glorot |
| Batch size | 16, 32, 64 | 32 |
| Max Number of training epochs | 5000 | 5000 |
| Loss function | MSE | MSE |
| Dropout | 0, 0.1, 0.2 | 0.1 |
| Early Stopping Patience | 20 epochs | 20 epochs |

**Table 3.** Hyperparameters used for grid search of the FFNN model

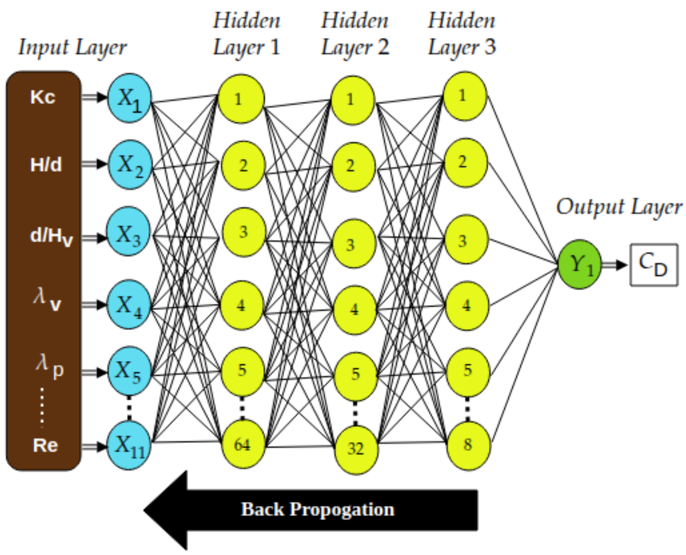



**Fig. 3.** Architecture of Feed Forward Neural Network (FFNN) for reference study

## 4. Results and Discussions

To evaluate the model's generalization and predictive performance, a separate testing dataset never encountered during training was used. To account for the variance across different training instances due to randomness of weight initialization and dropout, the models were trained and their performance was evaluated for 1000 iterations. The average values of the performance metrics such as MSE and $R^2$ were calculated over these 1000 iterations and the training and testing loss history for the best instance during this training was recorded. Different sets of input features have been considered in this study and the results for the best models out of the 1000 iterations are presented further. As discussed earlier in section 2.1, the study provides a systematic analysis of inclusion of three additional parameters in the prediction of $C_D$ to investigate their potential impact on the predictive model and assess whether incorporating them enhances model performance.

Firstly, this study presents the results using the 11 input features from Wang et al. (2023), as shown in Table 2. Fig. 4 shows the total loss curve and the comparison between model predictions and true values. The loss curve evolution for the training and test datasets was recorded over 450 epochs, with early stopping mechanisms applied to prevent model overfitting. The plot compares the predicted drag coefficient values with the actual ones for the test case, showing the alignment of the model's predictions with actual data. The green deviation band represents an acceptable margin of error (±20%), and most points lie within this range, indicating good accuracy.

The $R^2$ value of 0.91 provides a measure of how closely the predicted values follow the measured values, reflecting the model's overall performance.

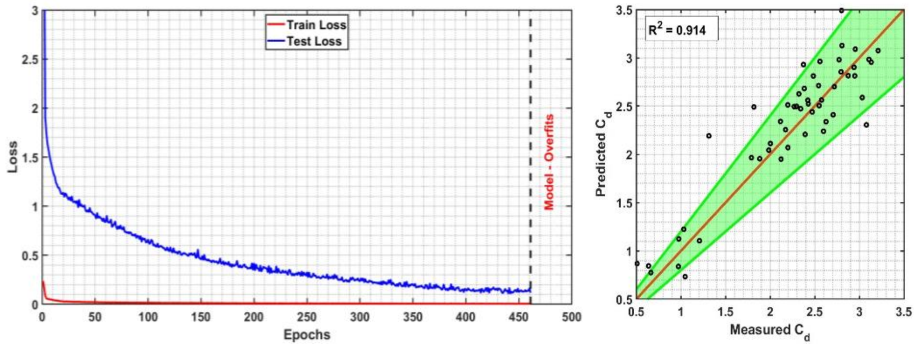



**Fig. 4.** Loss history (left) and prediction deviation (right) band plot for features included from reference study

Next, this study considers the same set of 11 input features; however, the submergence ratio $(d/H_v)$ from the reference study is replaced with the proposed submergence ratio $(a+d)/H_v$ for model training. Fig. 5 presents the training and testing loss curves over 320 epochs, along with the prediction plot, which compares the predicted values against the measured (actual) values. The deviation bands (±20%) illustrate the tolerance, and the model achieves an $R^2$ value of 0.921. This shows that incorporating the proposed submergence ratio as an input feature can potentially improve model performance by approximately 1% compared to the submergence ratio used in the reference study for predicting the dissipative coefficient.

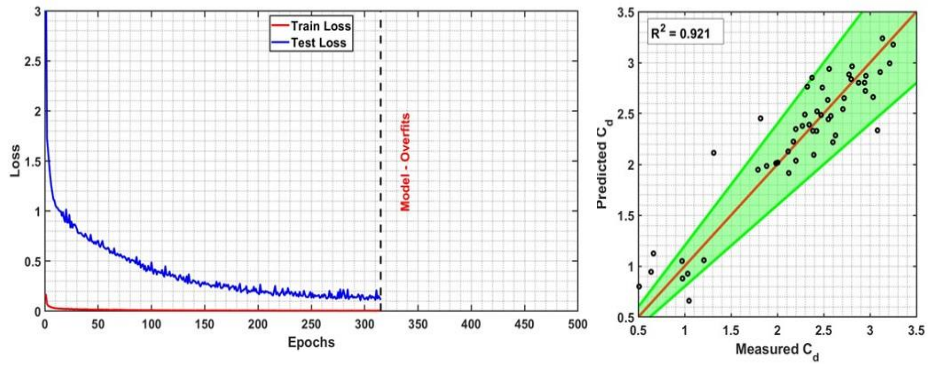



**Fig. 5**. Loss history (left) and prediction deviation (right) band plot for model with the proposed submerged ratio

Most prior studies (He et al., 2019; Mendez & Losada, 2004; H. Wang et al., 2022) have reported the significance of parameters mentioned in reference studies for wave vegetation interaction, except $C_m^*$ and $C_d^*$. This study further incorporates these two additional parameters, $C_m^*$ and $C_d^*$, to examine their effect in predicting the dissipative coefficient $C_D$. These parameters are added to the same set of 11 input features, along with the proposed submergence ratio, for model prediction. Fig 6 illustrates the loss curve evaluation for model training and testing, along with the prediction plot. The performance metric of $R^2$ value 0.892 was achieved during the test case. This shows that the model achieves reasonable prediction compared to the prediction achieved in the inputs of the reference study. To analyze the effect of these two additional parameters individually, another two sets of input features, including only one of the parameters at a time, were considered. The performance metrics of these two sets, along with all the prior sets, are listed in Table 4. It can be noted that compared to $C_d^*$ the inclusion of $C_m^*$ has a more positive effect on the prediction of the dissipative coefficient, and this result is supported by the correlation analysis presented in section 2.1, where $C_m^*$ was shown to have a stronger correlation than $C_d^*$.

In summary, this study provides a data-driven analysis of the inclusion of the two additional parameters $C_m^*$ and $C_d^*$ and also examines their potential impacts on the model performance.

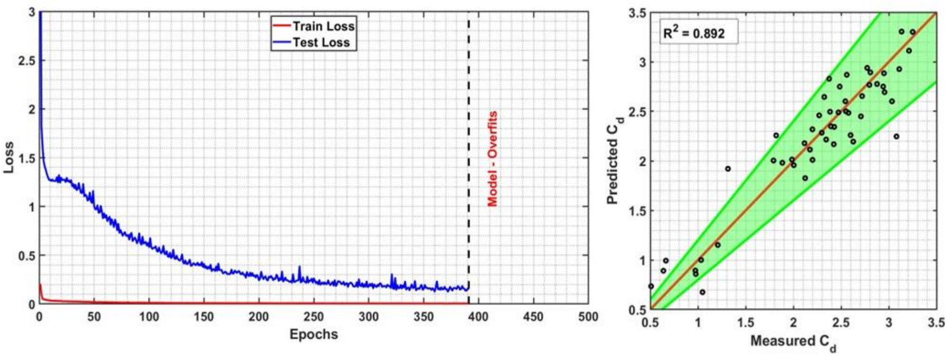



**Fig. 6**. Loss history (left) and prediction deviation (right) band plot for model with proposed submerged ratio along with the additional parameters of $C_m^*$ and $C_d^*$

**Table 4.** Summary of Performance metrics achieved for each set of input features

| Input features | Remarks | MSE | | $R^2$ | |
|---|---|---|---|---|---|
| | | Average | Best | Average | Best |
| $Kc, \frac{H}{d}, \frac{d}{H_v}, \lambda_f, \lambda_p, \frac{H}{L}, \frac{d}{L}, \frac{L}{L_v}, Fr, Ur, Re$ | Reference (Wang et al. 2023) | 0.205 | 0.112 | 0.850 | 0.918 |
| $Kc, \frac{H}{d}, \frac{(a+d)}{H_v}, \lambda_f, \lambda_p, \frac{H}{L}, \frac{d}{L}, \frac{L}{L_v}, Fr, Ur, Re$ | Replacing with the proposed submergence ratio | 0.206 | 0.108 | 0.849 | 0.921 |
| $Kc, \frac{H}{d}, \frac{(a+d)}{H_v}, \lambda_f, \lambda_p, \frac{H}{L}, \frac{d}{L}, \frac{L}{L_v}, Fr, Ur, Re, C_d^*, C_m^*$ | Inclusion of both $C_d^*$, $C_m^*$ | 0.246 | 0.144 | 0.820 | 0.894 |
| $Kc, \frac{H}{d}, \frac{(a+d)}{H_v}, \lambda_f, \lambda_p, \frac{H}{L}, \frac{d}{L}, \frac{L}{L_v}, Fr, Ur, Re, C_d^*$ | Inclusion of only $C_d^*$ | 0.259 | 0.137 | 0.810 | 0.900 |
| $Kc, \frac{H}{d}, \frac{(a+d)}{H_v}, \lambda_f, \lambda_p, \frac{H}{L}, \frac{d}{L}, \frac{L}{L_v}, Fr, Ur, Re, C_m^*$ | Inclusion of only $C_m^*$ | 0.223 | 0.124 | 0.836 | 0.909 |

## 5. Conclusions

The present study investigated the dependence of several non-dimensional parameters involved in wave vegetation interaction. Through systematic analysis and correlation studies, key parameters – Froude number, Reynolds number, relative length of vegetation patch, relative water depth, inertial and drag co of a single cylinder, and Ursell number are involved in predicting dissipative coefficient associated with wave vegetation interaction process. Similarly, the monotonic relationship trend between the modified submergence ratio and the dissipative coefficient was highlighted, which aligned with the experimental observation in the literature. Furthermore, five ANN models were investigated with a combination of the modified submergence ratio single cylinder inertial coefficient ($C_m^*$) and drag coefficient ($C_d^*$), which showed improved $C_D$ prediction compared to the existing empirical model. The present study has achieved the prediction of $C_D$ close to a maximum MSE of 0.259 out of 5 ANN models considered. The ANN model can be a potential tool supporting the macroscopic numerical studies related to wave vegetation interaction studies. This study proved an improved ML-based solution for solving the present gap in the microscopic numerical investigation of wave vegetation research. Further, the predicted from the ANN model will be tested with macroscopic numerical models in the future and fine-tuned.

**Acknowledgments.** This work is supported by the Department of Science & Technology, India Grant No: DST/CCP/CoE/141/2018C under SPLICE – Climate Change Programme and Scheme for Promotion of Academic and Research Collaboration (SPARC). We also gratefully acknowledge the Prime Minister's Research Fellowship (PMRF), India and HPCE, Indian Institute of Technology Madras, for providing us the computational resources for the completion of this study.

**Disclosure of Interests.** The authors have no competing interests to declare relevant to this article's content

## References


1. Augustin, L. N., Irish, J. L., & Lynett, P. (2009). Laboratory and numerical studies of wave damping by emergent and near-emergent wetland vegetation. *Coastal Engineering*, *56*(3), 332–340. https://doi.org/10.1016/j.coastaleng.2008.09.004
2. Busari, A. O., & Li, C. W. (2015). A hydraulic roughness model for submerged flexible vegetation with uncertainty estimation. *Journal of Hydro-Environment Research*, *9*(2), 268–280. https://doi.org/10.1016/j.jher.2014.06.005
3. Chen, H., & Zou, Q. P. (2019). Eulerian–Lagrangian flow-vegetation interaction model using immersed boundary method and OpenFOAM. *Advances in Water Resources*, *126*, 176–192. https://doi.org/10.1016/j.advwatres.2019.02.006
4. Chen, J., Pillai, A. C., Johanning, L., & Ashton, I. (2021). Using machine learning to derive spatial wave data: A case study for a marine energy site. *Environmental Modelling and Software*, *142*. https://doi.org/10.1016/j.envsoft.2021.105066
5. Dalrymple, R. A., Kirby, J. T., & Hwang, P. A. (1984). Wave diffraction due to areas of energy dissipation. *Journal of Waterway, Port, Coastal, and Ocean Engineering*, *110*(1), 67–79.
6. Danielsen, F., Sørensen, M. K., Olwig, M. F., Selvam, V., Parish, F., Burgess, N. D., Hiraishi, T., Karunagaran, V. M., Rasmussen, M. S., Hansen, L. B., Quarto, A., & Suryadiputra, N. (2005). The Asian tsunami: A protective role for coastal vegetation. *Science*, *310*(5748), 643. https://doi.org/10.1126/science.1118387
7. Deogaonkar, V., Jadhav, A., & Somayajula, A. (2023). *DATA DRIVEN IDENTIFICATION OF SHIP MANEUVERING COEFFICIENTS*. http://asmedigitalcollection.asme.org/OMAE/proceedings pdf/OMAE2023/86878/V005T06A050/7041174/v005t06a050-omae2023-104644.pdf
8. Divya, R., Sriram, V., & Murali, K. (2020). Wave-vegetation interaction using Improved Meshless Local Petrov Galerkin method. *Applied Ocean Research*, *101*. https://doi.org/10.1016/j.apor.2020.102116
9. Elmar Balas, C., Koç, L., & Balas, L. (2004). Predictions of Missing Wave Data by Recurrent Neuronets. *Journal of Waterway, Port, Coastal, and Ocean Engineering*, *130*(5), 256–265. https://doi.org/10.1061/ASCE0733-950X2004130:5256
10. Fan, S., Xiao, N., & Dong, S. (2020). A novel model to predict significant wave height based on long short-term memory network. *Ocean Engineering*, *205*. https://doi.org/10.1016/j.oceaneng.2020.107298
11. Figueroa-Alfaro, R. W., van Rooijen, A., Garzon, J. L., Evans, M., & Harris, A. (2022). Modelling wave attenuation by saltmarsh using satellite-derived vegetation properties. *Ecological Engineering*, *176*. https://doi.org/10.1016/j.ecoleng.2021.106528
12. Hari Ram, N., Saincher, S., & Sriram, V. (2022). Numerical investigation into wave attenuation characteristics of vegetation belt in terms of vortex shedding due to different arrangement configuration. *In 23rd Congress of the International Association for Hydro Environmental Engineering and Research Asia Pacific Division. IAHR-APD 2022*.

13. Hari Ram, N., & Sriram, V. (2024). Energy Dissipation Characteristics in Vegetation Belt due to mono-chromatic and bi-chromatic waves. *Ocean Engineering*, *312*, 119062.
14. Hari Ram, N., Sriram, V., & Murali, K. (2022). Experimental investigation on the characteristics of solitary and elongated solitary waves passing over vegetation belt. *Journal of Ocean Engineering and Marine Energy*, *8*(3), 305–318. https://doi.org/10.1007/s40722-022-00233-2
15. Hari Ram, N., & Sundar, V. (2024). A Comprehensive Review on Wave-Vegetation Interaction. In *Hydrodynamics of Wave-Vegetation Interactions* (pp. 203–238). World Scientific.
16. He, F., Chen, J., & Jiang, C. (2019). Surface wave attenuation by vegetation with the stem, root and canopy. *Coastal Engineering*, *152*. https://doi.org/10.1016/j.coastaleng.2019.103509
17. Hu, K., Chen, Q., & Wang, H. (2015). A numerical study of vegetation impact on reducing storm surge by wetlands in a semi-enclosed estuary. *Coastal Engineering*, *95*, 66–76. https://doi.org/10.1016/j.coastaleng.2014.09.008
18. Hu, Z., Lian, S., Wei, H., Li, Y., Stive, M., & Suzuki, T. (2021). Laboratory data on wave propagation through vegetation with following and opposing currents. *Earth System Science Data*, *13*(10), 4987–4999. https://doi.org/10.5194/essd-13-4987-2021
19. Kim, T., & Lee, W.-D. (2022). Review on Applications of Machine Learning in Coastal and Ocean Engineering. *Journal of Ocean Engineering and Technology*, *36*(3), 194–210. https://doi.org/10.26748/ksoe.2022.007
20. Li, Y.-H., Xie, L., & Su, T. (2018). Resistance of Open-Channel Flow under the Effect of Bending Deformation of Submerged Flexible Vegetation. *Journal of Hydraulic Engineering*, *144*(3). https://doi.org/10.1061/(asce)hy.1943-7900.0001419
21. Liu, H., Zhang, K., Li, Y., & Xie, L. (2013). Numerical study of the sensitivity of mangroves in reducing storm surge and flooding to hurricane characteristics in southern Florida. *Continental Shelf Research*, *64*, 51–65. https://doi.org/10.1016/j.csr.2013.05.015
22. Maza, M., Lara, J. L., & Losada, I. J. (2015). Tsunami wave interaction with mangrove forests: A 3-D numerical approach. *Coastal Engineering*, *98*, 33–54. https://doi.org/10.1016/j.coastaleng.2015.01.002
23. Maza, M., Lara, J. L., & Losada, I. J. (2016). Solitary wave attenuation by vegetation patches. *Advances in Water Resources*, *98*, 159–172. https://doi.org/10.1016/j.advwatres.2016.10.021
24. Mendez, F. J., & Losada, I. J. (2004). An empirical model to estimate the propagation of random breaking and nonbreaking waves over vegetation fields. *Coastal Engineering*, *51*(2), 103–118. https://doi.org/10.1016/j.coastaleng.2003.11.003
25. Nepf, H. M. (1999). Drag, turbulence, and diffusion in flow through emergent vegetation. *Water Resources Research*, *35*(2), 479–489. https://doi.org/10.1029/1998WR900069
26. Noarayanan, L., Murali, K., & Sundar, V. (2012). Role of vegetation on beach run-up due to regular and cnoidal waves. *Journal of Coastal Research*, *28*(1A), 123–130.

27. Ozeren, Y. (2011). Phase I Report for SERRI Project No. 80037: Investigation of surge and wave reduction by vegetation. *Laboratory Publication*. https://www.researchgate.net/publication/308546946
28. Panda, J. P. (2023). Machine learning for naval architecture, ocean and marine engineering. In *Journal of Marine Science and Technology (Japan)* (Vol. 28, Issue 1, pp. 1–26). Springer. https://doi.org/10.1007/s00773-022-00914-5
29. Phan, K. L., Stive, M. J. F., Zijlema, M., Truong, H. S., & Aarninkhof, S. G. J. (2019). The effects of wave non-linearity on wave attenuation by vegetation. *Coastal Engineering*, *147*, 63–74. https://doi.org/10.1016/j.coastaleng.2019.01.004
30. Sangeetha, S., Ram, H., Somayajula, A., & Sriram, V. (2024). *Prediction of nonlinear wave elevation using deep learning approach*. http://onepetro.org/ISOPEPACOMS/proceedings pdf/PACOMS24/PACOMS24/ISOPE-P-24-160/4177214/isope-p-24-160.pdf/1
31. Sriram, V., Stoesser, T., Yan, S., & Murali, K. (2023). Hydrodynamics of Wave-vegetation Interactions. *World Scientific*, *14*.
32. Struve, J., Falconer, R. A., & Wu, Y. (2003). Influence of model mangrove trees on the hydrodynamics in a flume. *Estuarine, Coastal and Shelf Science*, *58*(1), 163–171. https://doi.org/10.1016/S0272-7714(03)00072-6
33. Sundar, V., Sannasiraj, ; S A, Murali, ; K, & Sundaravadivelu, R. (2007). Runup and Inundation along the Indian Peninsula, Including the Andaman Islands, due to Great Indian Ocean Tsunami. *Journal of Waterway, Port, Coastal, and Ocean Engineering*, *133*(6), 401–413. https://doi.org/10.1061/ASCE0733-950X2007133:6401
34. Tanaka, N. (2009). Vegetation bioshields for tsunami mitigation: Review of effectiveness, limitations, construction, and sustainable management. In *Landscape and Ecological Engineering* (Vol. 5, Issue 1, pp. 71–79). https://doi.org/10.1007/s11355-008-0058-z
35. Wang, H., Yin, Z., Luan, Y., Wang, Y., & Liu, D. (2022). Hydrodynamic Characteristics of Idealized Flexible Vegetation under Regular Waves. *Source: Journal of Coastal Research*, *38*(3), 673–680. https://doi.org/10.2307/48668658
36. Wang, Y., Liu, Y., Yin, Z., Jiang, X., & Yang, G. (2023). Numerical simulation of wave propagation through rigid vegetation and a predictive model of drag coefficient using an artificial neural network. *Ocean Engineering*, *281*. https://doi.org/10.1016/j.oceaneng.2023.114792
37. Wu, C., Wu, S., Wu, X., Zhang, Y., Feng, K., Zhang, W., & Zhao, Y. (2023). Hydrodynamics affected by submerged vegetation with different flexibility under unidirectional flow. *Frontiers in Marine Science*, *9*. https://doi.org/10.3389/fmars.2022.1041351
38. Wu, W. C., & Cox, D. T. (2015). Effects of wave steepness and relative water depth on wave attenuation by emergent vegetation. *Estuarine, Coastal and Shelf Science*, *164*, 443–450. https://doi.org/10.1016/j.ecss.2015.08.009
39. Yin, Z., Yang, G., Wang, Y., Qiu, Q., & Jiang, X. (2024). Numerical simulations of breaking wave propagation through the vegetation on a slope based on a drag coefficient prediction model. *Ocean Engineering*, *291*. https://doi.org/10.1016/j.oceaneng.2023.116440

40. Zhang, C., & Zhang, M. (2023). Numerical investigation of solitary wave attenuation and mitigation caused by vegetation using OpenFOAM. *Coastal Engineering Journal*, *65*(2), 198–216. https://doi.org/10.1080/21664250.2022.2163844