

A Generalized Machine Learning Model to Predict Significant Wave Height from Wind Parameters

Submitted By

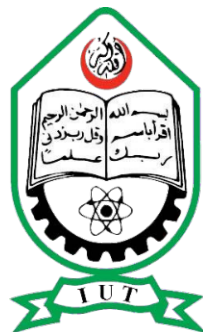
Imrul Kayes	ID: 180011101
Abid Hasan	ID: 180011203
Minhazul Alam	ID: 180011217

Supervised By

Prof. Dr. Mohammad Ahsan Habib

Co-Supervised By

Mr. Tanvir Shahriar, Assistant Professor, MPE



Department of Mechanical and Production Engineering (MPE)

Islamic University of Technology (IUT)

May 2023

A Generalized Machine Learning Model to Predict Significant Wave Height from Wind Parameters

Submitted By

Imrul Kayes	ID: 180011101
Abid Hasan	ID: 180011203
Minhazul Alam	ID: 180011217

Supervised By

Prof. Dr. Mohammad Ahsan Habib

Co-Supervised By

Mr. Tanvir Shahriar, Assistant Professor, MPE



**A Thesis submitted in partial fulfillment of the requirement for the degree
of Bachelor of Science in Mechanical Engineering**

Department of Mechanical and Production Engineering (MPE)

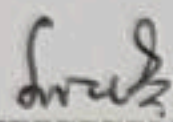
Islamic University of Technology (IUT)

May 2023

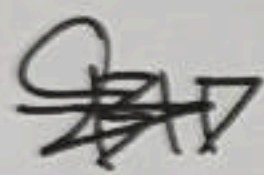
Candidate's Declaration

This is to certify that the work presented in this thesis, titled, "A Generalized Machine Learning Model to Predict Significant Wave Height from Wind parameters", is the outcome of the investigation and research carried out by us under the supervision of Dr. Mohammad Ahsan Habib, Professor, Department of Mechanical and Production Engineering.

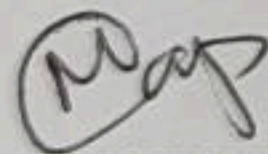
It is also declared that neither this thesis nor any part of it has been submitted elsewhere for the award of any degree or diploma.



Name of the Student: Imrul Kayes Emon
Student No: 180011101



Name of the Student: Abid Hasan
Student No: 180011203



Name of the Student: Minhazul Alam
Student No: 180011217

RECOMMENDATION OF THE BOARD OF SUPERVISORS

The thesis titled "A Generalized Machine Learning Model to Predict Significant Wave Height from Wind Parameters" submitted by Imrul Kayes Emon, Student No: 180011101, Abid Hasan, Student No: 180011203 and Minhazul Alam, Student No:180011217 has been accepted as satisfactory in partial fulfillment of the requirements for the degree of B Sc. in Mechanical Engineering on 19th May, 2023.

BOARD OF EXAMINERS

1. *Ahsan*
19/05/2023

Dr. Mohammad Ahsan Habib

(Supervisor)

Professor

MPE Dept., IUT, Board Bazar, Gazipur-1704, Bangladesh.

2. *t@mmmm*
19/05/2023

Mr. Tanvir Shahriar

(Co-Supervisor)

Assistant Professor

MPE Dept., IUT, Board Bazar, Gazipur-1704, Bangladesh.

Acknowledgement

In the Name of Allah, the Most Beneficent, the Most Merciful. We would like to begin by expressing our gratitude to the Almighty Allah, who enabled us by bestowing upon us the forbearance, strength, ability, and knowledge to undertake this project and finish it satisfactorily on schedule.

We would like to express our deepest gratitude to our supervisor Dr. Mohammad Ahsan Habib, Professor, Department of Mechanical and Production Engineering, IUT for sharing his knowledge and ideas regarding this project, providing us with invaluable guidance and the necessary resources to execute this task, patiently checking this thesis and papers, caring and inspiring us throughout each phase of this project. We would like to express our thanks to our co-supervisor Mr. Tanvir Shahriar, Assistant Professor, Department of Mechanical and Production Engineering, IUT, for sharing his technical expertise pertaining to this project, for guiding us from the very beginning of this work, helping in the coding sections, and patiently explaining the difficult aspects of this project. Without the significant time and effort of our supervisor and co-supervisor, it would not have been possible for us to complete this project on time.

We would also like to express our gratitude to Department of Research, Extension, Advisory Services and Publications (REASP), IUT for providing us with the necessary funding (154 USD) for this project and our friend, Hasan Saju from the CSE Department for helping us make the payment for Google Colab Pro which we used for running the scripts. We would like to thank our examiners for the insightful suggestions and feedback about our project and double checking our work.

We would like to say thanks to our parents, family members and friends for providing us with mental support and inspiration throughout this journey.

Abstract

Due to recent energy crises in the post-pandemic world, exploration of marine renewable energy sources is more crucial than ever. Significant wave height is a key parameter for wave energy extraction, it also has a wide range of applications, including ship navigation, oil and gas extraction, and the construction of coastal structures. Among the existing methods of measuring significant wave height, direct measurements using buoys are very expensive and limited in number, moreover these provide data with low time and spatial resolutions whereas numerical models are based on mathematical equations, assumptions and becomes complex when they are applied for generalization purpose. With a view to facilitate the utilization of wave energy and foster research activities by providing cheap dataset of wave properties with high spatial and time resolution, this work focuses on developing a generalized machine learning model that is able to predict significant wave height from wind parameters on a huge area around the coastlines of USA and Canada. Four machine learning models have been used in this work; 2 deep learning models (Artificial Neural Network (ANN) and Self Normalizing Neural Network (SNN)) and 2 gradient boosting tree-based models (XGBoost and LightGBM) and performance of these models have been evaluated on test data, distinct from the one used for training. The deep learning models have showed greater fitting capacity compared to tree based model on the training data, achieving the lowest Mean Squared Error(MSE) (0.047 for ANN, 0.063 for SNN, 0.226 for XGBoost, 0.108 for LightGBM) and highest R^2 score (0.953 for ANN & 0.937 for SNN, 0.894 for XGBoost, 0.892 for LightGBM) whereas the gradient boosting models demonstrate better generalizing capacity compared to the deep learning models on both the known (data from these buoys are included in the training set) and unknown (data from these buoys are not included in the training set) buoys. Furthermore, impact of outlier detection and removal using Tukey's Fence method on the performance of ANN & SNN has been evaluated and found to be insignificant.

Table of Contents

Acknowledgement	v
Abstract	vi
Table of Contents	vii
List of Figures	x
List of Tables	xiii
Chapter 1 Introduction	1
1.1 Background of this study	2
1.2 Objectives with Specific Aims.....	5
1.3 Organization of this thesis	6
Chapter 2 Literature Review	8
2.1 Ocean Energy	8
2.2 Different types of ORE and their Utilization.....	9
2.2.1 Energy Harvesting Technology from the Wave	9
2.2.2 Tidal & Current Energy Extraction Technology	15
2.2.3 Ocean Thermal Energy Extraction Technology (OTEC)	21
2.2.4 Salinity Gradient Energy Extraction Technology.....	24
2.3 Relationship between Wave Energy Flux and Significant Wave Height:	25
2.4 Machine Learning Models to measure significant wave height:	26
2.4.1 Prediction Models	26
2.4.2 Forecasting Models.....	28

2.5 Summary:.....	31
Chapter 3 Materials and Methods	33
3.1 Introduction.....	33
3.2 Study Area and Data Collection	34
3.3 Data Sets Pre-Processing	42
3.4 Outlier Removal with Box-Cox method:.....	50
3.5 Machine Learning Models to Predict Significant Wave Height.....	55
3.5.1 XGBoost	55
3.5.2 LightGBM.....	61
3.5.3 Artificial Neural Network (ANN).....	63
3.5.4 Self-Normalizing Neural Network:.....	70
3.6 Modelling Procedure.....	71
3.6.1 XGBoost	72
3.6.2 LightGBM.....	73
3.6.3 Artificial Neural Network (ANN).....	74
3.6.4 Self-Normalizing Neural Networks (SNN)	75
3.7 Model Performance and Validation Criteria.....	76
3.7.1 Evaluation Metrics	76
Chapter 4 Results and Discussion.....	78
4.1 XGBoost	78
4.2 LightGBM.....	84

4.3 Artificial Neural Network (ANN).....	89
4.4 Self-Normalizing Neural Network (SNN).....	94
4.5 Model Comparison:	99
4.6 Outlier Evaluation.....	109
Chapter 5 Conclusions and Future Plan.....	114
5.1 Conclusion	114
5.2 Future Plan.....	115
References.....	117

List of Figures

Figure 2-1 Different Types of Wave Energy Converters [40].....	11
Figure 2-2 Different Type of Tidal Stream Generator [69]	17
Figure 2-3 Schematic of DTP System and Traditional Tidal Barrage System [86]	19
Figure 2-4 Schematic of a basic OTEC system [95].....	22
Figure 3-1 Flowchart for Overall Modelling Procedure	34
Figure 3-2 (a-d) Wind Rose Plots for Different Stations.....	37
Figure 3-3 Locations & Bathymetry of the Buoys.....	42
Figure 3-4 Work flow of Data processing	43
Figure 3-5 (a-f) Distribution of features before and after standardization.....	47
Figure 3-6 Correlation Matrix for the selection of input features.....	48
Figure 3-7 (a-f) Histogram, box plot and probability plot.....	49
Figure 3-8 (a-f) Distribution of variables before and after transformation.....	53
Figure 3-9 (a-f) Distribution of variables before and after outlier removal.....	54
Figure 3-10 XGBoost Tree Building Process	58
Figure 3-11 ReLU activation function.....	66
Figure 3-12 SELU activation function.....	71
Figure 4-1 Scatter plot of Wave Height against XGBoost prediction on test and validation data	79
Figure 4-2 Line plot of actual values and XGBoost prediction on known buoys.....	82
Figure 4-3 Line plot of actual values and XGBoost prediction on unknown buoys.....	82
Figure 4-4 Scatter plot of Wave Height against XGBoost prediction on known buoys.....	83
Figure 4-5 Scatter plot of Wave Height against XGBoost prediction on unknown buoys.....	84
Figure 4-6 Scatter plot of Wave Height against LightGBM prediction on test and validation data.....	85

Figure 4-7 Line plot of actual values and LightGBM prediction on known buoys.....	87
Figure 4-8 Line plot of actual values and LightGBM prediction on unknown buoys.....	88
Figure 4-9 Scatter plot of Wave Height against LightGBM prediction on known buoys	88
Figure 4-10 Scatter plot of Wave Height against LightGBM prediction on unknown buoys .	89
Figure 4-11 Scatter plot of Wave Height against ANN prediction on test and validation data	90
Figure 4-12 Line plot of actual values and ANN prediction on known buoys	92
Figure 4-13 Line plot of actual values and ANN prediction on unknown buoys	92
Figure 4-14 Scatter plot of Wave Height against ANN prediction on known buoys	93
Figure 4-15 Scatter plot of Wave Height against ANN prediction on unknown buoys	93
Figure 4-16 Scatter plot of Wave Height against SNN prediction on test and validation data	95
Figure 4-17 Line plot of actual values and SNN prediction on known buoys.....	96
Figure 4-18 Line plot of actual values and SNN prediction on unknown buoys.....	96
Figure 4-19 Scatter plot of Wave Height against SNN prediction on known buoys.....	98
Figure 4-20 Scatter plot of Wave Height against SNN prediction on unknown buoys.....	98
Figure 4-21 Bar plot of error metrics for all the models on different stations.....	100
Figure 4-22 Bar plot of error metrics on testing and validation data for all four models.....	101
Figure 4-23 Line plot of actual values and prediction of all four models on known buoys ..	102
Figure 4-24 Line plot of actual values and prediction of all four models on unknown buoys	102
Figure 4-25 Scatter plot of actual values and predicted values of all four models on known buoys.....	104
Figure 4-26 Scatter plot of actual values and predicted values of all four models on unknown buoys.....	105
Figure 4-27 Box plot of actual wave height and predicted wave height on known buoys	106

Figure 4-28 Box plot of actual wave height and predicted wave height on unknown buoys	106
Figure 4-29 Box plot of actual wave height and predicted wave height on test and validation data.....	107
Figure 4-30 Violin plot of residual error of all four models on known buoys.....	108
Figure 4-31 Violin plot of residual error of all four models on unknown buoys.....	108
Figure 4-32 Violin plot of residual error of all four models on test and validation data.....	109
Figure 4-33 Line plot of actual wave height and ANN, SNN model outputs on known buoys after outlier removal.....	111
Figure 4-34 Line plot of actual wave height and ANN, SNN model outputs on unknown buoys after outlier removal.....	111
Figure 4-35 Line plot of actual wave height and ANN model outputs on unknown buoys after outlier removal.....	112
Figure 4-36 Line plot of actual wave height and SNN model outputs on unknown buoys after outlier removal.....	112

List of Tables

Table 1-1 Input variables & Target variable used in ML models.....	5
Table 3-1 Buoy specifications & description of measurements	38
Table 3-2 Buoy Location & Serial Number.....	38
Table 3-3 Station wise mean properties for the parameters.....	39
Table 3-4 Station wise deviation for the parameters	40
Table 3-5 Train Test and Validation Data Distribution among stations.....	44
Table 3-6 Hyperparameter Optimization Search Space for XGBoost.....	73
Table 3-7 Hyperparameter Optimization Search Space for LightGBM	74
Table 3-8 Hyperparameter Optimization Search Space for ANN	75
Table 3-9 Hyperparameter Optimization Search Space for SNN.....	76
Table 4-1 Model Evaluation of XGBoost on test and Validation data.....	79
Table 4-2 Comparison of Means and Standard Deviation of Predicted Value with Actual Value	79
Table 4-3 Model Evaluation of XGBoost on unknown buoys	81
Table 4-4 Model Evaluation of XGBoost on known buoys	81
Table 4-5 Model Evaluation of LightGBM on test and Validation data	84
Table 4-6 Comparison of Means and Standard Deviation of LightGBM Predicted Value with Actual Value	85
Table 4-7 Model Evaluation of LightGBM on unknown buoys.....	86
Table 4-8 Model Evaluation of LightGBM on known buoys.....	86
Table 4-9 Model Evaluation of ANN on test and Validation data	89
Table 4-10 Comparison of Means and Standard Deviation of ANN Predicted Value with Actual Value.....	90
Table 4-11 Model Evaluation of ANN on unknown buoys.....	90

Table 4-12 Model Evaluation of ANN on known buoys.....	91
Table 4-13 Model Evaluation of SNN on test and Validation data	94
Table 4-14 Comparison of Means and Standard Deviation of SNN Predicted Value with Actual Value	94
Table 4-15 Model Evaluation of SNN on unknown buoys.....	95
Table 4-16 Model Evaluation of SNN on known buoys.....	95
Table 4-17 Error evaluation of all four models on validation and test data	99
Table 4-18 Error evaluation of all four models on known and unknown buoy data	99
Table 4-19 Error evaluation of ANN and SNN models on validation and test data after outlier removal	110
Table 4-20 Error evaluation of ANN and SNN models on known and unknown buoy data after outlier removal	110

Chapter 1 Introduction

Significant wave height is instrumental to several coastal activities like ship navigation, marine traffic control, Port locating, environmental monitoring, deploying coastal protection structures, wave energy converters, building platform for the oil & gas extraction, dredging activities [1]–[4]. Hazardous occurrences or disasters brought on by extreme weather events can be predicted with the use of high-seas estimations of significant wave height and it aides in planning coastal risk management tasks [5]–[9]. Prediction of significant wave height is also vital for assessing the wave resistance which causes significant fuel consumption on the ship hull and offers safety and improved performance for ship navigation[10]. Prior knowledge of significant wave height is critical for the continually evolving aquaculture business, numerous military exercises along the coasts & deep seas and the efficient usage of ocean energy potential [11]. In recent decades, there has been a widespread adoption of the wave-based power industry, which has predominantly stemmed from the imperative of global leaders to transition to renewable energy sectors [12]–[14]. The precise history of significant wave height permits forecasting of future weather trends and patterns, which is paramount for maximization of the potential of these power industries [15]. The assessment of wave parameters in seaports is crucial for safety reasons, as it enables the shipment and unloading of freight in safe conditions. As such, numerous research activities have been conducted all over the world to improve accuracy and reduce the computational demands of predicting and forecasting wave height.

In this study, a Machine Learning strategy has been proposed to deliver reliable datasets while making efficient use of scarce resources and minimizing the necessary amount of computing effort. The aim of this research is to illustrate the reliability of ML models as an alternative to numerical models for generalization purposes and to show that high-cost rate associated with direct buoys implementation can be circumvented by ML models. In this work, a generalized

machine learning model is trained & tested based on standard datasets provided by NDBC (National Data Buoy Centre) & necessary datasets were downloaded through ERRDAP. Location of the study area is selected mostly along the coastlines of USA, Canada and several European countries and finally the model is validated based on some unseen datasets. When fed wind data & atmospheric data, this generalized model can provide significant wave height. This will significantly reduce the cost of adopting direct buoy measurement, increase both spatial and time resolution of the datasets and facilitate academics with precise datasets of significant wave height which will ultimately boost the economic prosperity of least developed and emerging nations by exploiting their wave energy potentials. After analyzing several machine learning models, tree-based models (XGBoost, LightGBM) are selected as these models perform well in structured & tabular dataset, also performances of these models have been compared with two deep learning models Artificial Neural Network (ANN) and Self-Normalizing Neural Network (SNN).

1.1 Background of this study

Significant wave height can be influenced by a variety of meteorological variables, including water depth, tides, air temperature, atmospheric pressure, and wind related parameters such as wind direction, gust, speed, fetch [16], [17]. The buoys are reliable sources of measuring these meteorological data in near and deep sea which can be categorized into two main categories [18]:

1. Moored buoys
2. Drifting buoys

Moored buoys are fixed at ocean beds even in adverse weather conditions to routinely gather precise data, monitor weather condition, but relatively expensive to deploy & maintain whereas Drifting buoys are typically connected to a drogue or marine anchor, inexpensive to

deploy, float on the ocean's surface, they are typically carried by ocean eddies, and provide less precise data than moored buoys[19], [20]. NDBC has installed roughly 50 C-MAN stations and over 100 Moored buoys primarily along the coastlines of the United States which can measure air temperature, seawater temperature, water level, humidity levels, sea surface pressure, wind related parameters, wave properties and significant wave height continuously & transmit data in real time [21]. Although data collected by Moored buoys are highly quality controlled, but these have limited spatial coverage, limited time resolution, expensive to maintain and repair, susceptible to sensor drifting, biofouling which might affect the accuracy level. To increase the spatial coverage and to provide inexpensive datasets, NDBC has also deployed numerous drifting buoys as part of various research programs, including GARP, FGGE, and TOGA which can measure longitude and latitude, barometric pressure, sea surface temperature, wind properties[18]. The less precise data collected by drifting buoys is transmitted to the ground station via NOAA satellite, which takes a considerable amount of time [22]. In addition, there can be data loss because of difficulties in transmission and marine life interference. Among the various buoys that have been deployed by the NDBC, only wave rider buoys can estimate wave parameters directly whereas most of the buoys can measure sea surface temperature, wind parameters, air temperature, longitude and latitude. Due to their high initial cost and ongoing maintenance requirements, Wave rider buoys have seen relatively limited deployment outside of the coasts of the United States, Canada, and Europe. Unfortunately, poor nations lack access to these buoys and research operations in these regions are being severely hampered and their wave energy harnessing potential is being underutilized.

To compensate for the drawbacks of direct measurement mentioned above, scientists from all over the world have employed a variety of approaches and created a multitude of wave models that can be categorized into three distinct groups according to Hu et al. [23].

1. Numerical Model (rely heavily on mathematical equations and diverse boundary conditions)
2. Machine Learning and statistical models (utilize the correlations between parameters to estimate output without taking ocean physics into account)
3. Hybrid models (integration of numerical and ML models)

Several numerical models like JONSWAP [24], WAM [25], SWAN [26], WAVEWATCH III [27] have been invented over the years to replicate wave properties. The implementation of numerical models is very challenging due to the fact that a change in location affects the physics and boundary conditions of the mathematical equations which augments computational burden to a great extent [28]. Moreover, due to the necessity for accurate local bathymetric measurements, generalization across a vast region also results in considerable inaccuracies for numerical models [29], [30]. Additionally, these models perform badly during drastic occurrences such as sudden storm peaks [31]. Large amounts of high-quality data, which are quite difficult and expensive to gather, may be needed to train and validate numerical models.

So, the feasibility of utilizing machine learning model as an alternative to numerical model for wave height prediction has been examined comprehensively in this work. After an extensive literature review, two tree based models (XGBoost and LightGBM) & two deep learning models (ANN, SNN) has been selected as they perform well in large structured and tabular dataset. Input parameters have been selected based on the literature review, also to justify the parameters, Pearson Correlation matrix has been determined to find the correlation. Based on the result of correlation matrix, input parameters that have been selected for our model are listed in Table 1-1.

Table 1-1: Input variables & target variable used in ML models

	Variables		
	Acronym	Description	Unit
Input	longitude	Longitude	°E
	latitude	Latitude	°N
	atmp	Air Temperature	°C
	bar	Air Pressure	hPa
	mwd	Wind Propagation Direction	° true
	gst	Wind Gust Speed	ms ⁻¹
	wspd	Wind speed	ms ⁻¹
	wspu	Wind speed, Zonal	ms ⁻¹
	wspv	Wind speed, Meridional	ms ⁻¹
Output	wvht	Significant wave height	m

1.2 Objectives with Specific Aims

Present work aims to build a generalized machine learning model as an alternative to direct measurement & numerical model to predict significant wave height for a humongous area around the coastlines of USA and Canada. It will be a prediction-based model which will take wind parameters, location, temperature, pressure of a specific buoy (which lies within the coverage of the trained model) as input and provide significant wave height as output for that specific buoy. The computation time will be minimal for this model and both the time resolution and spatial resolution would be higher than conventional methods of predicting significant wave height. This model aims to provide cheap dataset with satisfactory precision and accuracy to the nations that lack access to wave height measuring buoy to foster research activities and unleash their wave energy harnessing potential.

This work has following objectives:

- To analyze the feasibility of extracting wave energy and the potential of wave energy to mitigate energy crisis around the world through extensive literature review
- To comprehend the importance of significant wave height prediction and the feasibility of machine learning model as an alternative to numerical models for wave height prediction through extensive literature review
- Selection of Input parameters for the proposed Machine learning model and implementing Pearson Correlation matrix to justify the selection of input parameters
- Selection of suitable Machine Learning models capable of handling this huge dataset (XGBoost, LightGBM, ANN, SNN)
- To Train this model based on train and validation data and later the performance of the models will be evaluated on the unknown stations.
- To contrast the tree-based model with the deep learning models based on their generalization capability in wave height prediction.

1.3 Organization of this thesis

This thesis is structured into five chapters. Chapter 1 starts with a brief introduction about the importance of significant wave height, its applications in different fields, different methods of measuring significant wave height and finally the proposition of implementing machine learning methods to predict significant wave height from wind parameters. In the next subsections, background of this work has been discussed where problems related to the two conventional approaches of significant wave height prediction (Direct Measurement & Numerical Models) have been identified and in the next subsection, ultimate aim and the objectives of this work have been discussed and this chapter ends by exhibiting the organization of this dissertation.

In chapter 2, an extensive literature review is given which is divided into 5 sections. The first section gives a brief introduction on Ocean energy, its harnessing potentials across different parts of the world. In the second section, different types of ocean renewable energy (ORE) along with their utilization have been discussed which can be subdivided into 4 sub-sections. For each type of energy extraction processes, insights of relevant energy extraction devices have been given. In the third section, significant wave height has been established as a vital parameter by exhibiting its relationship with wave energy flux and for the final section extensive literature review has been given for current progress of utilizing machine learning techniques to predict significant wave height. The overall discussion in this sub-section have been based on two different perspectives: prediction-based works and forecasting based works. The final section for this chapter gives an overall summary of this chapter.

Chapter 3 of the thesis starts with a introduction of the overall modelling procedure of this work and the subsequent sections detail the different methods used for the study along with how the data selection and processing has been done. The different comparison metrics for result justification are also described in this section.

In chapter 4, the results found from the study are presented and discussed thoroughly.

The thesis ends with the conclusion and future scope in chapter 5. Lastly, references are added.

Chapter 2 Literature Review

2.1 Ocean Energy

Many studies on renewable energies have been conducted globally in recent years as a result of the global energy crisis and issues related to the usage of fossil fuels, such as greenhouse gas emissions. The energy potential of the seas and oceans has received a lot of attention among these resources since they have a large quantity of energy stored in many forms, such as wave, tidal, thermal, and current energy, which can more than supply the whole global need for power [32]. Compared to other renewable sources, ocean energy has a minimal environmental impact, a high-power density, and a high utilization rate. At a level 19.5 m above the sea surface, wave power is approximately five times more than wind power. Ocean energy can be harnessed in different forms, including tidal and current energy, waves, salinity gradients, and thermal gradients. Electrical energy can be generated from the motion of tides through the harnessing of tidal energy. The amount of electricity produced is dependent on the intensity of water level height ranges or tidal current velocities, with higher electricity generation resulting from greater intensity. Onshore or offshore sites can use different wave energy converter technologies to harness the kinetic and potential energy associated with ocean waves. Pressure-retarded reverse osmosis processes and associated conversion technologies can exploit the energy resulting from the mixing of fresh water and seawater at the mouth of rivers, known as salinity gradient energy. Additionally, ocean thermal energy conversion processes can harness the thermal energy generated by the temperature gradient between the sea surface and deep water. The capacity of the world's oceans to store 93,100 TWh of electrical energy annually is estimated by the International Energy Agency (IEA). Tidal wave and thermal power potentials caused by gravity, wind, and sunlight were previously estimated to be 22,000, 2000, and 87600 TWh/y, respectively, but more recent calculations indicate that wave power alone has a

potential of 2985 GW. From 24 to 28 °C at the surface to 4-6 °C in 1-kilometer depths, the ocean's temperature changes. The ocean thermoelectric generator might have its roots in the temperature difference. In the arid and temperate regions between latitudes 30°S and 30°N, temperature fluctuations of 20 °C are frequently encountered. Worldwide, tidal potential is evenly distributed, equatorial regions have high OTEC potential, although the tropics have higher wave energy potential [33]–[35].

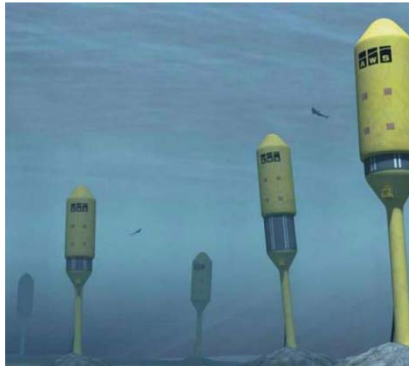
In different parts of the world, ocean renewable sources are already producing good quantities of energy. Countries like the UK, the USA, South Korea, China and the Netherlands are leaders in this sector being surrounded by ocean all around [36].

2.2 Different types of ORE and their Utilization

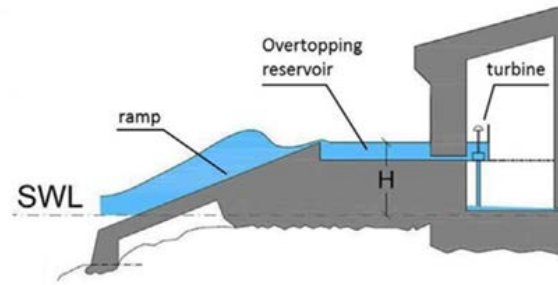
2.2.1 Energy Harvesting Technology from the Wave

The most well-known representation of ocean energy is energy from the wave. Waves are a result of the effects of the wind on the water's surface. Because of this, wave power is tangentially seen as a form of solar energy. It has long been proposed to use wave energy for energy purposes. The exploration of wave energy involved a number of experiments and programs, but these efforts eventually came to an end due to the drop in the price of oil and the underwhelming results of the experimental units [37]. Due to the volatility of the oil market, worries about global warming, and restrictions on greenhouse gas emissions over the past decades, interest in WECs has increased [32]. Wave energy in particular and renewable energy in general attracted a lot of attention in the 1970s. This desire, however, did not lead to the development of any system with respectable technological and financial qualities. Having high goals for the programs was one of the primary causes. In the UK, for instance, the goal was to create 2 GW of wave power producing capacity by 1983. As a result, there were exceedingly expensive, oversized designs. Recent systems have been developed as small individual units,

typically less than 500 kW, which may be deployed in huge numbers to generate enormous amounts of electricity, based on the previous experiences [38]. Shoreline, near shore, and offshore systems are the three basic divisions of these devices. The complicated mooring systems, undersea electrical cables, or high-pressure pipes are not needed by the shoreline systems to transport generated electricity to the shore. They are also simpler to install and maintain. However, there aren't many places that would be suitable for such plants, and the wave energy there isn't as high as it is offshore. Additionally, because each coastline has specific characteristics, it's possible that mass manufacture and economies of scale cannot be used to shoreline systems, which could result in higher installation costs. Since the waves are calmer and less damaging to the coastal zone once the wave energy has been absorbed, the near shore systems can be useful for coastal protection. Before commercialization can be achieved for the offshore ones, the accessibility and dependability of the plants must be greatly increased. The way the WECs transform wave energy into the required product can also be used to identify them. Wave energy is extracted and has great force but low velocity. Before it can be used in a generator to generate electricity, this energy must be converted into a high-speed spinning motion. This necessitates a complex mechanical system, which raises the plant's construction and maintenance costs and decreases system reliability due to the large number of moving parts, which lowers the system's ability to endure powerful storms [39].



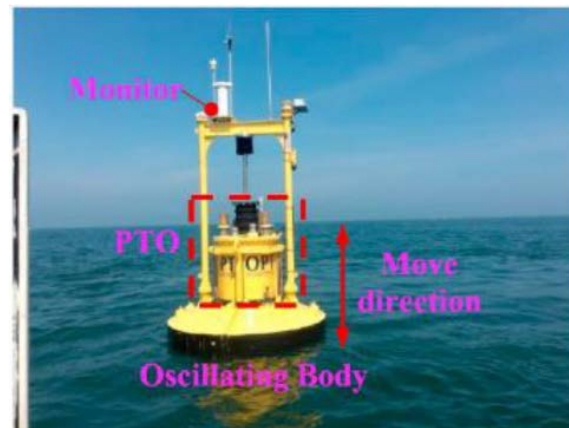
Submerged Pressure Differential WEC



Overtopping Device



Attenuator and Terminator Device



Point Absorber Device



Oscillating Wave Surge Converter

Figure 2-1: Different types of wave energy converters [40]

2.2.1.1 Point Absorbers

As indicated in Figure 2-1, point absorbers are a form of floating oscillating body that consists of a heaving buoy that uses a Power Take-Off (PTO) to extract wave energy from the motion between the buoy and a fixed reference or between a submerged body and the sea bed [41].

Point absorbers were first theorized in the 1900s but the first small scale studies took place in Japan in the 1980s [42] and later in Norway [43].

The generated energy may take the form of shaft power, air or liquid pressure, which can be converted to electrical energy in coastal or offshore facilities, or direct electricity, depending on the mechanism utilized. Additionally, there are proposals to operate reverse osmosis water desalination facilities with generated high pressure sea water [44], [45].

2.2.1.2 Oscillating Water Column (OWC)

One of the most common types of wave energy converters are the oscillating water column types though it spent a long time in the development phase. The system utilizes a gradually narrowing water column and water trapping. The top of the water column is narrowed at the inlet of the turbine. This design increases the trapped air flow in the designed column cell. The pressurized air flows through the turbine which ultimately produces electricity. The system does not directly use any wave but uses the air trapping and pressurizing capacity of the waves to generate power by rotating the turbine using air flow. Air in these systems can flow in both directions. This flaw is rectified by using Wells Turbine which spins unidirectionally regardless of air direction [32].

2.2.1.3 Attenuators and Terminators

Attenuators and terminators technology of wave energy harvesting utilizes the relative motion between multiple floating structures. Double-action pumps are moved in these systems by the pitch and yaw motions between the structures or sections riding parallelly to the wave direction. Lower force is induced the area being lower due to its relative orientation with the wave. Several attenuator and terminator type wave converters have been designed and deployed. The most prominent was the Scottish Pelamis wave converter. The Pelamis converter, cylindrical in shape is segmented in nature. It can convert the vertical and horizontal induced wave motion

at the hinged joints to pressurized fluid which can be used to run electric generators through hydraulic motors. The final transfer of electrical power occurs by means of cables to the shore. There are several instances of Pelamis being used – one full-fledged system in the UK in 2004, a 2.25 MW WEC attenuator plant in Portugal, a 3 MW 4unit plant in Scotland and several others [32]. After several attempts at making their product commercially viable, the company had to declare bankruptcy finally in 2014 after failing to secure funds needed [46]. Other wave energy converters based on the attenuator and terminator concept are the wave star energy converted which was registered in 1898 [32]. The model has been mostly experimental except for one installed in the North Sea which was capable of developing 500 kW of power. Another similar technology is the McCabe wave pump which is close to the pelamis in the concept of power generation but yet to be commercially realized [47].

2.2.1.4 Overtopping Devices

Most ocean wave energy converters are complicated in nature and consists of many parts. The abundance of parts possess complication in maintenance and repairing. To eliminate these issues overtopping devices were introduced. These devices have very small number of movable parts making them very suitable and easy to implement. The concept of these devices is very simple where some type of a reservoir or barrage is created that can hold the water coming in with the waves. The water is then let to flow back to the sea through a low-head turbine which produces the wanted electric power [32]. One of the devices of this type is the Waveplane A/S which is capable of storing incoming wave waters at different heights. The water is then allowed to pass through the turbine to produce electricity. The product initially showed good promise but later went under the radar [48]. Another overtopping device is the famous Wave Dragon, credited to be one of the most successful devices of the type. The system utilizes the concept of hydro power stations and can be added with wings to increase wave deposition in the reservoir [49]. Aalborg University conducted a laboratory test worth 1:50 scale model in

the 1990s by developing a software for simulation. The study showed the proposal to be usable and insurable. Another 1:4 turbine test conducted at TU Munich showed that the model was as efficient as existing hydro turbine-based power plants. Based on these experiments, Soerensen et al. deployed a 1:50 prototype in the Danish Wave Energy Test Station at Nissum Bredning in 2003 and found the survivability and power production to be satisfactory [50]. The Tapered Channel Wave Energy Device, shortened as Tapchan is another kind of overtopping devices. This one has guided channels through which the incoming waves are amplified and then discharged into an above the water line reservoir. The process of electricity generation is similar to the others using a low head turbine. The first of the kind was introduced in Norway but yet to see practical implementations [38].

2.2.1.5 Oscillating Wave Surge Converters

Oscillating Wave Surge Converter has a design that uses a buoyant flap anchored to the sea bottom. The oncoming waves hit the flap and the movement of the flap at the hinge is converted to electrical energy. These type of converters usually operate at an intermediate depth of water [51]. Several projects were undertaken as the CETO OSWC device to be installed in Australia capable of generating about 5 MW of power. The project had several prototypes installed in 2003, 2006 and 2008 with a final commercial scale set up in 2009. Another system based on the CETO model named bioWAVE was proposed to be setup in Australia with a capacity of 250 kW [32]. No subsequent information about the plans were found. WaveRoller and Pendolor are two similar projects. Several proposals were developed about these designs but other than a prototype, no practical implementation was seen [52]. Salter Cam is a patented device that can extract energy from the waves. The device promises to convert almost 90% of the impacted wave energy to electrical, chemical or mechanical energy with minimal loss [32], [53].

2.2.1.6 Submerged Pressure Differential Converters

Commonly installed near the shore on the seabed, Submerged Pressure Differential Converters are another type of Ocean wave converters that can generate electricity by using the created pressure differential due to the rise and fall of the level of water due to waves. The pressure gradient allows pumping the fluid through an arrangement that generates the electricity [54]. Archimedes Wave Swing by AWS Ocean Energy is a submerged pressure differential converter. Several models have been tested to study the prospects as – a 16 kW swing in Glasgow and an 80-kW unit in Orkney, both of which has exceeded expectations. The company plans to develop 500 kW units soon that can be integrated to develop about 10 MW of power in each platform [55]–[57].

2.2.2 Tidal & Current Energy Extraction Technology

Most renewable energy sources are in one way or other linked to solar energy and derived from. But the tides of the ocean are created due to the effect of the existing gravitational forces between the Moon and the Sun with the water bodies on the Earth and the rotation of the earth [58]. The global reserve of Tidal energy is about 100 GW but only some of it can be extracted due to geographical limitations [59].

2.2.2.1 Tidal Barrage

Tidal energy extraction can be done in a variety of ways but the primary idea is very similar to the hydroelectricity power production technologies. But in tidal energy extraction, the flow of the tides is not unidirectional in all cases requiring several arrangements. Primarily a barrage is created that traps the oncoming tidal water and separates them on either side of the barrage erected. The height differential allows the water to pass from one side to another through a turbine installed in the barrage wall. This generates the required electricity [32].

There are many successful implementations of the Tidal Barrage extraction system. The largest tidal power plant in the world follows this principle to generate up to 254 MW of power. The Sihwa Lake Tidal Power Station generates this power on an unpumped flood generation scheme where only the tidal inflow generates power. The plant has 10 25.4 MW Bulb type turbines installed with an annual generation of 552 GWh of electricity [60], [61]. South Korea has another Tidal Barrage Power plant in Uldolmok that powers 430 households. The Tidal Barrage station has a barrage width of 3260 meters and channel speed exceeding 6.5 meters per second with a capacity of 1.5 MW peak and an annual output of 2.4 GWh [62], [63]. The world's first Tidal Barrage system set up in France on the Rance River in France is still working producing peak 240 MW of power with a capacity factor of 28% and an annual generation of 500 GWh. The plant has a tidal range of 8 meters and spans 750 m in length accommodating 24 Turbines. The electricity generation cost is cheaper in comparison to other sources and the plant attracts quite a lot of visitors also [64]. China's Jiangxia Tidal power station is another working tidal barrage power station with a peak of 3.2 MW. The plant has three turbines of 500, 600 and 700 kW respectively. The maximum tidal range stands at 8.39 meters and the annual generation is 6500 MWh [65].

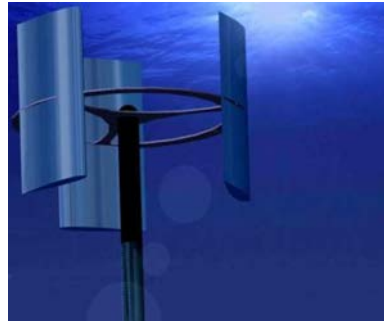
2.2.2.2 Tidal Stream Generator

Tidal Stream Generators can be said to be the different types of underwater devices capable of extracting energy from the flowing water. These are similar to the wind turbines and known as tidal turbine. Classified into six types, the tidal stream generators are horizontal and vertical axis turbines, oscillating hydrofoils, venturi devices, Archimedes screws and tidal kites [66]. Three commonly tidal turbines are shown in Figure 2-2. One of the most notable tidal turbines developed by Orbital Marine Power was set up at the Fall of Warness Tidal Test Site, Orkney owned by the European Marine Energy Centre (EMEC) back in 2016. The model known as SR2000 was capable of generating peak 2 MW of power and had two turbines with rotor blades

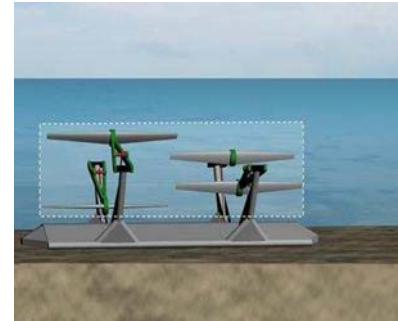
spanning 16m in length. The blades were at a water depth of 33 meters. The turbine tested for a 12 month continuously generated 3200 MWhs of electricity. Though it had a lifetime of 20 years, it was taken off in 2018 from the site and the later model Orbital O2 has been deployed in the location since 2021 [66]–[68].



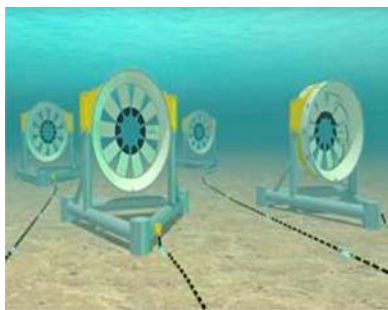
Horizontal-Axis Turbine



Vertical-Axis Turbine



Oscillating Hydrofoil



Ducted Turbine



Archimedes Screw



Tidal Kite

Figure 2-2: Different types of Tidal Stream Generator [69]

The Orbital O2 was successfully deployed on 22nd of April 2021 amid the coronavirus pandemic from the Port of Dundee. The largest existing tidal turbine has the capacity of meeting the demand of 2000 homes and can offset near to 2200 tons of produced Carbon-dioxide a year. The O2 is part of the FORWARD 2030 project by the EU’s Horizon 2020 research plan where it is planned to deploy 2030 MW of tidal energy before 2030. The integrated plan for a greener future will see multiple O2 turbines connected with a host of other renewable energy devices with the EMEC systems for massive decarbonization efforts [70]–[72]. Other than the Orbital O2, several other small-scale prototypes have been developed such as the HS300 Turbine in Kvalsund which is a 0.3 MW grid connected turbine set at a depth of

50m within a 700 m wide channel in Norway. The turbine was developed for a period of 3 years and after testing was taken off in 2007. This model led to the development of the HS1000 Turbine which helped develop the 1.5 MW MK 1 turbine which was installed in the MeyGen project [73]. The ongoing project has a plan of installing capacity to 86 MW from installed 6 MW by the installation of 49 more along with its installed 4 turbines. The project lease can have up to 398 MW capacity installed and divided into 3 phases with each increasing the capacity gradually [74], [75]. Atlantic Resources Corporation, responsible for the AR 1500 model of the MeyGen project also deployed a 1 MW turbine before at the EMEC facility during 2011 [76]. Other smaller implementations include the Seaflow, a 300 kW propeller type prototype turbine installed in Devon in 2003 [77]; a prototype project in New York City between Queen and the Rossevelt Island in the East River; SeaGen, a 1.2 MW turbine installed in Norther Ireland in 2008; Evopod, a 1/10 scale turbine set in Northern Ireland and the Triton 3, a 3 MW turbine set up in the Thames [66].

Crossflow turbines based on the Darreius design was deployed as a prototype in the South Korea with plans to be scaled up to 90 MW but no further information could be retrieved [78]. Flow Augmented Turbines were also proposed in Australia but the project was not commercially deployed after the initial trials [66]. Oscillating devices has been developed by Australian company BioPower System based on the Biomimicry of swimming species to develop the bioSTREAM tidal power converter system. Several power rated version of the device are still in R&D phase [79]. Swedish company Minesto proposed a tidal kite device capable of developing between 150 and 800 kW of power [80].

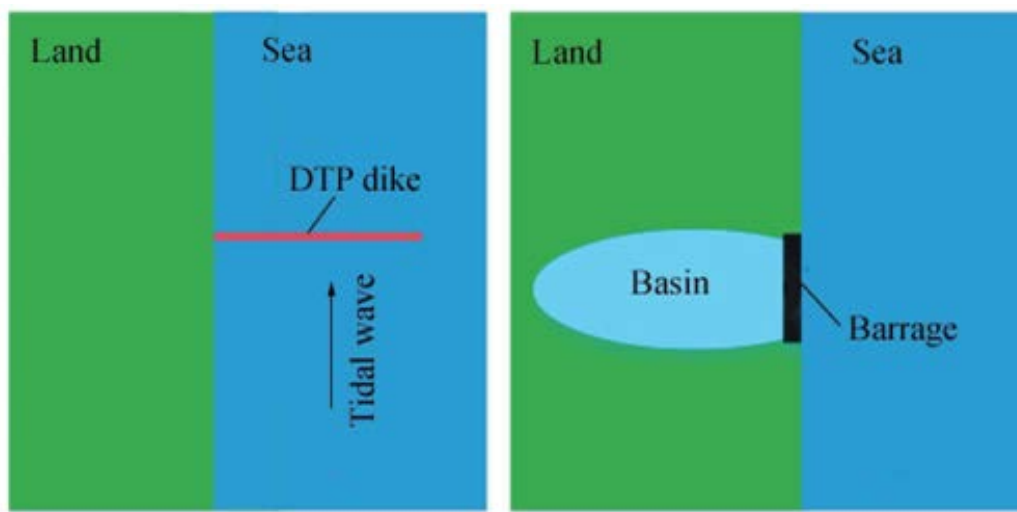
2.2.2.3 Tidal Lagoon

Tidal lagoons are by principle of energy generation similar to the tidal barrage system of energy generation but they differ from the tidal barrages in the sense that tidal barrages are natural systems whereas tidal lagoons are created artificially. Turbines are incorporated into the created

walls that has the ability to capture the tidal potential energy effectively [81]. Tidal lagoon method of tidal energy harnessing is still a very new concept and has not been explored much. The very first Tidal Lagoon was planned to be deployed in Swansea Bay in the UK but the project was rejected after initial go ahead. The project funded by the government was eventually dropped and a new Blue Eden plan was conceptualized which will be completely privately funded [82]–[84]. A comparative study by D. Vandercruyssen et al. found that tidal barrages and lagoons both have their pros and cons and depending on the location being considered, the schemes should be choose [85].

2.2.2.4 Dynamic Tidal Power

Dynamic tidal power is a novel idea for tidal energy harnessing. Commonly known as DTP, the system can be imagined to be a barrier 30 or more kilometers long ending up in the sea. It requires a structure perpendicular to the coast and a parallel barrier at the end. The two barriers will form a ‘T’ shape (shown in Figure 2-3). The structure inhibits the flow of the tidal wave and a heigh difference in the level of water on either side of the barrier is created. The water is allowed to flow through Bi-Directional Turbines installed in the dam.



DTP System

Traditional Barrage System

Figure 2-3: Schematic of DTP system and traditional tidal barrage system [86]

The positive side of the DTP includes a higher power output, stable power, availability and combined implementation. But the DTP implementation requires at least 30 kilometers length to be viable. This being a considerable length makes it difficult to implement. Currently no power plant following the concept Kees Hulsbergen and Rob Steijn, the inventors of DTP exists but feasible locations are proposed along the coast of China, South Korea and the UK [87]. Y.H. Park from the Korea Institute of Ocean Science & Technology ran numerical simulations of DTP of 10 km, 30 km and 50 km in the Yellow Sea for a period of 30 days case by case basis. Though the produced power might be an over estimation but it was observed that the peak power had the potential to reach up to 23,383 MW which would have a net total annual generation of 61 TW h equaling 68% of the installed tidal barrage capacity in 2014. Similar studies in Netherlands showed a potential of annual generation of 40 TW h by a 35 km T-shaped DTP [88]. Dai et al. predicted DTP in the Taiwan Strait of China and made a comparative analysis of the power production with the open ratio. The open ratio can be said to be the ratio of the open area of the turbines to that of the lateral area of the dike. The numerical study found the mean power increase from 0.43 GW to 0.89 GW as the open ratio was changed in between 2% and 8% but the power decreased with open ratios greater than 8% [89]. Shao et al. investigated DTP systems numerically in the Yellow Sea and the Bohai Sea and concluded that in place of a single larger dam, the combination of more than one dam has the potential to exceed the efficiency of a single working dam due to less interference with surrounding hydrodynamic conditions [90]. Y.H. Park from the KIOST carried out numerical simulations with different shapes of DTP but it did not show massive improvement in the results. Park concluded that the traditional DTP might be the most economical choice though future studies and practical experimentations can give more accurate results. DTP shapes should be optimized based on the tidal condition of the location chosen and the best output can be extracted from a 70 kilometers long simple DTP [91]. Armoudli et al. simulated a 30 km

long DTP near the Faror Island in the Persian Gulf. The 3D numerical model found up to 760 MW of power can be generated by using direct-drive permanent magnet synchronous generators with a water level difference of 1.2 meters between the two sides of the dam [92]. Dynamic Tidal Power though conceptualized and patented in 1997 but yet lacks any practical development or experimentation on actual setups due to being quite costly as the size of the barrier should be at least 30 kilometers long to be considered viable. But numerical studies show good promise and this should be explored extensively.

2.2.3 Ocean Thermal Energy Extraction Technology (OTEC)

The potential of ocean thermal energy conversion (OTEC), which is one of the permanent and readily available renewable energy sources that could help with base-load power supply, is significantly bigger than that of other ocean energy kinds. OTEC systems are practically infinite sources of energy since they rely exclusively on the sun and ocean currents, which cover more than 70% of the earth's surface and absorb solar radiation. They are essentially the most efficient energy storage systems in existence because of this. OTEC merely stores solar thermal energy as a result of the ocean's temperature gradient between the surface-warm water and the deep-sea cold water [93]. OTEC is suitable for a variety of tasks, including cooling and power production. The core idea behind OTEC technology is to use surface-warm ocean water (25–30 °C) to heat and evaporate a working liquid, which creates steam that drives a turbine to spin and produce power. When the fluid is once more condensed to a liquid state at a deep cold water source's temperature (4-6 °C), the thermodynamic cycle is complete [94]. The schematic of a basic OTEC system is shown in Figure 2-4. The change in water temperature has a significant impact on the cycle's efficiency. Higher efficiencies are produced by larger temperature differences.

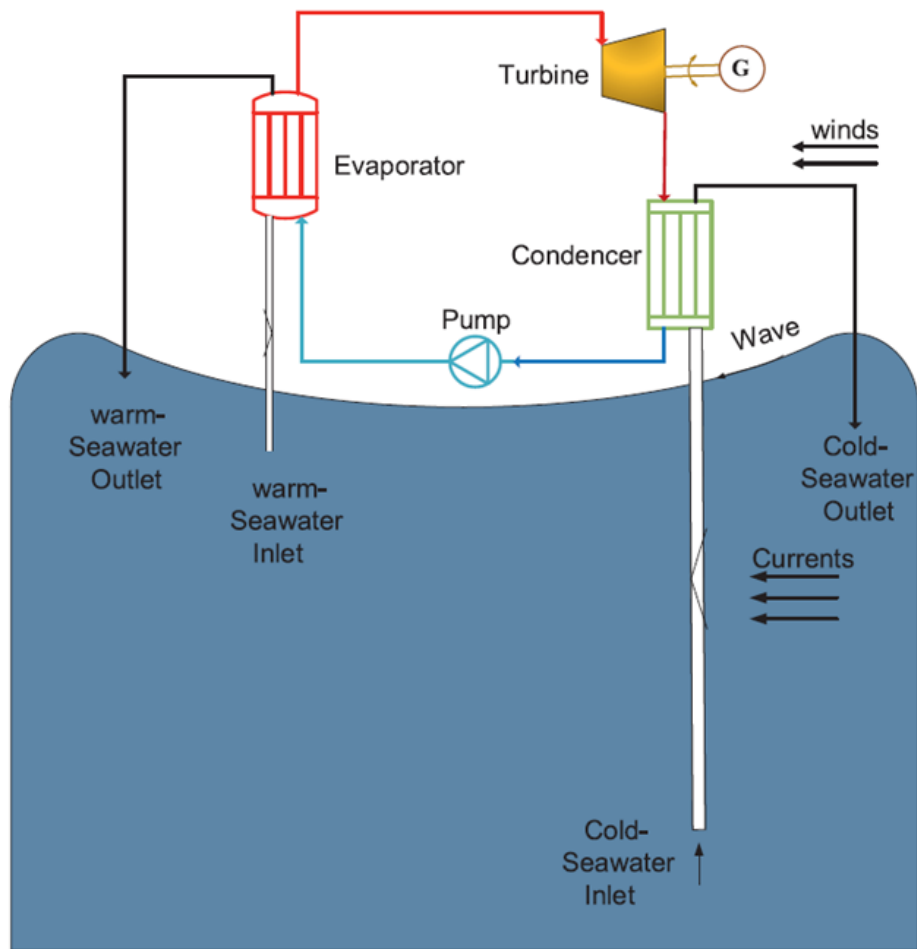


Figure 2-4: Schematic of a basic OTEC system [95]

This happens as a result of the water's modest temperature difference (of roughly 20 °C) and the thermodynamic cycle's generally low efficiency. When the temperature difference is 20 °C, the Carnot cycle efficiency—the ideal thermodynamic cycle in theory—is approximately 6.7%. As a result, choosing an appropriate working fluid and thermodynamic cycle are crucial factors in boosting OTEC systems' efficiency [96], [97]. D. Vera et al. has examined a wide range of OTEC system types and given an outline of the major challenges facing OTEC technology. The steam turbine needs to be optimized if the OTEC system is to operate more efficiently. Due to the restrictions put forth by the minimal water temperature difference, the OTEC power generation system's efficiency is frequently subpar [95]. J. Langer et al. reported a case of upscaling of OTEC plants as they are mostly available in small scale which is not really viable. They concluded that up to 45 GW of power can be harnessed in Indonesia by taking proper

steps in regions with higher electricity demands [98]. In a different work, J. Langer et al. ran a study for studying the feasibility of a 135 MW gross plant in Ende and found that the project is feasible. They also suggested that similar power plants could be developed in at least 11 locations as Hawaii, the USA; Lamu, Kenya; Kumejima, Japan; Tarawa, Kiribati; Manay, the Philippines; Rainbow Beach, Australia; Lagos, Nigeria; Fernando de Noronha, Brazil; Ofu, Samoa; Montego, Jamaica and Port Gentil in Gabon. The temperature profile and electricity demand of these places should vary from the one in Indonesia but the locations might be suitable given the ocean condition [99]. OTEC plants were first conceptualized in the 1800s with the first of a kind 22 kW plant being set up in 1930. The plant could not stand and a natural calamity and got destroyed. Not much practical progress was made regarding commercial deployment. Several plants were designed and numerically studied in the USA, India, Indonesia and Cuba but commercial deployment is still not there [32]. In the 1970s, a 100 kW closed cycle OTEC planned was built by Tokyo Electric Power Company on Nauru Island which produced about 30 kW of useful electricity [100]. During the Arab-Israeli War of 1973, President Carter planned to produce 10 GW of electricity from OTEC systems by 1999 but that was never realized [101]. Several privately funded research beds were developed in Europe without any actual output. A 1 MW floating OTEC test bed was deployed in India near Tamil Nadu which eventually failed due to a faulty cold water pipe but researches are still being funded by the Government of India [102]. A plant capable of generating up to 30 MW of power was planned by Ocean Thermal Energy company on St. Croix costing a maximum of 300 million dollars but the project stopped due to the associated resort's bankruptcy [103]. Makai Ocean Engineering connected the very first OTEC plant to the grid in Hawaii, the USA for the first time in 2015. The plant is capable of generating 105 kW at the moment with plants to install up to 100 MW of power generating capacity to provide electricity to 120,000 Hawaiian homes [104], [105]. Another demonstrator plant is set up in Okinawa prefecture of

Japan in collaboration with Saga University. The 100 kW class plant works as a test bed for future OTEC integration with the Japanese grid and is also open to visitors [106]. Commercial application of OTEC though remains unrealized but the researches in this direction show its viability.

2.2.4 Salinity Gradient Energy Extraction Technology

Salinity gradient technologies use the chemical pressure difference between freshwater and seawater's differing ionic concentrations to produce power. Due to the high concentration of salt in seawater, it has a higher osmotic pressure than freshwater. Reverse electrodialysis (RED) and pressure-retarded osmosis (PRO), two major technological kinds, use semi-permeable membranes to create an osmotic potential that can be used to drive turbines in deltas or fjords to produce power. By converting the osmotic pressure of saline solutions to hydraulic pressure through a process known as pressure retarded osmosis, or PRO, electricity can be produced. PRO technologies produce energy from the differential in salt content between saltwater and freshwater, much like reverse electrodialysis systems do. Reverse electrodialysis, or RED, is a process that uses the carefully regulated mixing of two pools of water with various salinities to produce electricity. Several cation and anion exchange membranes are commonly constructed into high- and low-salinity compartments in RED technology. These membranes form charged poles like batteries when freshwater and saltwater are passed through them in opposition to each other [107].

The total estimated energy generation capacity of Salinity Gradient Extraction Technology is around 2.6 TW. These systems are very expensive to implement and as such makes them difficult to deploy. The advent of newer technologies and the development of cheaper membranes provide hope for such systems [32]. Statkraft Osmotic Power Prototype at Tofte in Hurum, Norway is the world's pioneering plant that runs on the principal of Salinity Gradient Energy Extraction Technology. Researchers at the plant calculated that up to 2.85 GW of power

can be extracted through the methods implemented [108], [109]. Apart from this, full-fledged practical operations based on this technology is yet to be realized but researchers are continuously working to make it viable.

2.3 Relationship between Wave Energy Flux and Significant Wave

Height:

In the preceding sections, the immense potential offered by wave energy to alleviate the global energy crisis has been exhaustively discussed, and various devices to extract wave energy, their applications, and limitations have been outlined. Prediction of significant wave height is crucial as wave energy flux is directly related to the significant wave height and can be described by the following equation [110]:

$$P = \frac{\rho g^2}{64\pi} H_s^2 T_e \quad (2.1)$$

Where,

P= Energy Flux

T_e= Wave Period

H_s=Significant Wave Height

So, this work aims to predict significant wave height from wind parameters and later this significant wave height can be related to harness wave energy flux for a specific location and based on these wave energy flux, specific wave energy extracting devices can be employed at that location.

2.4 Machine Learning Models to measure significant wave height:

Machine learning models may be a great substitute to mitigate the shortcomings imposed above as they work based on the correlation between input & target parameters which reduces the computational burden in generalization approach and provide precise values of target variable in a short amount of time, also both spatial & time resolution is high for these models. Based on their objective and data source, machine learning models can be divided into various categories. The two major categories are:

2.4.1 Prediction Models

These types of models take input data from multiple buoys to train a regression model, which then predicts the desired output of a nearby buoy. Mahjoobi et al. [111] evaluated the performance of CART, ANN, and C5 algorithms for predicting wave height in Lake Michigan and concluded that regression models are superior to classification models and akin to ANN models in terms of their effectiveness in prediction purpose. Elbisy et al. [112] has compared the performance of MART model and multiple ANN models in terms of significant wave height prediction at Abu Qir Bay coastal zone, Egypt. Various wind parameters have been used as input features after analyzing their significance in terms of output prediction, and it has been determined that MART model is superior to all other models, while RBFNN is the best among ANN models. Gracia et al. [113] has proposed an artificial neural network (MLP) model and a decision tree model, and then combined their performance via ensemble learning to enhance the wave parameter predictions of numerical models. In this study, 12 years of data from 4 buoys along the coast of Spain have been used, and the authors have concluded that in coming years, a machine learning model could even completely supplant the numerical model. Kumar et al. [114] has implemented two sequential learning algorithms, MRAN and GAP-RBF to predict the significant wave height and the prediction performance is compared with SVR and ELM. The author has collected data from 14 stations in the Gulf of Mexico, the United

Kingdom, and the Korean region for approximately five years. It has been determined that MRAN outperforms all preceding models. The integration of meteorological and structural data has been proposed in the training of supervised machine learning models (ANN, Tree based models) to enhance wave height forecasting ability by Demetriou et al. [11]. Gunaydin et al. [115] compared ANN models and regression models for predicting the monthly average wave height from mean wind speed, air temperature, and sea level pressure for three distinct buoys in the Atlantic. After analyzing seven various configurations of input features, this study justifies the aforementioned three input features as the optimal combination. The ANN model trained with all input features has been identified as the superior one and finally the pi theorem has been implemented in regression models to generate prediction-capable formulas. Shamshirband et al. [116] compared the efficacy of three machine learning models (SVR, ANN, ELM) to the SWAN simulation model of two distinct stations situated in Persian Gulf. It has been determined that ELM outperforms other models, but that all models fail to accurately predict extreme weather events. In this study, numerical models demonstrated identical efficacy to ML models. As an alternative to numerical models for simulating wave properties, James et al. [117] proposed two machine learning (ML) models (MLP as a wave height regression model and SVM as a wave period classification model). The wind parameters were derived from the SWAN model, validated against buoy measurements, and used as input features for the surrogate model. The bathymetry datasets were obtained from the National Oceanic and Atmospheric Administration (NOAA). When steady-state SWAN model conditions are examined, the author has demonstrated that ML can replace numerical models admirably in predicting wave height and period. Hu et al. [23] have compared the efficacy of two machine learning (ML) models (XGBoost, LSTM) with the numerical model WAVEWATCH3 in predicting wave parameters in Lake Erie based on wind parameters. This study utilized National Data Buoy Center (NDBC) standard wind data from 1994-2013 for

model training, 2014-2015 for model validation, and unobserved data from 2016-2017 for model performance evaluation on wave height and period prediction. This work's results justify the use of ML models due to the dramatic decrease in computation time compared to numerical models; while XGBoost performed best among the individual models, the ensemble implementation of all the models showed much greater promise than the separate implementations. Etemad-Shahidi et al. [118] has suggested a tree-based model as an alternative to ANN models for predicting wave height in Lake Superior using wind parameters. For the study, NDBC standard datasets of wind and wave parameters were collected from 2000 to 2001. The subsequent training and testing of the models revealed that the M5 algorithm based on Tress outperformed ANN models. Also, the use of tree-based models has been validated because they do not require the time-consuming topology determination phase of ANN models, their results are readily interpretable, and their computational unit and time requirements are less. Malekmohamadi et al. [16] have used multiple soft computing methods (ANN, SVM, BN, and ANFIS) to predict wave height in Lake Superior based on wind parameters. In this work, the dataset preprocessing phase includes sensitivity analysis of the input parameters, and the optimal input parameters were chosen based on the performance of several different configurations on an MLP model. Except for the BNs, all models predict wave height exceedingly well and can be relied upon in this sector, according to the findings of this study.

2.4.2 Forecasting Models

These types of models take historical data from a specific buoy to train a model capable of predicting the future measurement of the objective variable for that buoy. Deo et al. [119] proposed an ANN model to forecast wave period and wave height from wind parameters for three locations in Karwar, India, at a very nascent stage in the application of machine learning models to replicate wave parameters. The author has performed sensitivity analysis to

determine the optimal combination of input parameters, and fifty distinct input combinations have been supplied for this purpose. Although it is a very rudimentary model with limited capacity, the author has demonstrated the viability of ML models for predicting wave parameters, paving the way for future research. Makarynsky et al. [120] implemented ANN models to forecast wave height and wave period with a lead time of one to twenty-four hours in an earlier work employing ML models in wave height forecasting. The model was applied in two distinct locations, the Atlantic and the Irish Coast, and was based on some initial predictions, which were later corrected and then merged with measurements and previous day forecasting to predict wave parameters for the following day. The forecasting and subsequent correction and integration of multiple ANNs validated the feasibility of ML models to forecast wave parameters, but it was discovered that this simple model cannot capture a broad range of variations in the target parameters for the oceanic site. Agrawal et al. [121] have compared the performance of ANN models to that of two stochastic models (ARMA and ARIMA) for predicting wave height in the Indian Ocean with a lead time of 3, 6, 12 and 24 hours. This study demonstrates that ANN outperforms stochastic models for shorter lead time forecasting, whereas all models produce comparable results for long range forecasting. Zamani et al. [122] implemented ANN models with time-lag and other models with IBL technique, the author compared the performance of wave height forecasting for two distinct locations in the Caspian Sea, concluding that ANN with time-lag outperforms other models. Nitsure et al. [17] proposed using genetic programming as an alternative to numerical models to forecast the peak wave height for four buoys located along the United States and Indian coasts. For the proposed model, wind parameters were used as input variables, the lead time was 12h and 24h and satisfactory results were obtained using this technique. Ali et al. [123] has proposed an MLR-CWLS model to forecast wave height 30 minutes in advance. Australia's coastal zone has been chosen as the study area and eight years of data have been selected for this study. Finally, the

efficacy of this model was contrasted to that of conventional MLR, MARS, M5 Model Tree, and it was concluded that MLR-CWLS model is superior to other models. Berbic et al. [124] has proposed ANN and SVM models to forecast wave height with a 0.5-5.5-hour lead time. Initially, only wave height with time latency was used for forecasting; subsequently, wind parameters with time lag have been added to the model, which enhanced its accuracy. For greater accuracy, it has been determined that SVM and ANN should be used in shorter and longer lead times, respectively. To facilitate the navigation of a self-governing ship, Domala et al. [125] has contrasted the efficacy of Deep learning models (FBProphet) and tree-based models (XGBoost and Random Forest) in forecasting wave height for six distinct stations (three from Hawaii and three from Puerto Rico). Wind-related parameters, Pressure, and Temperature have been utilised to train the model, and wave properties have been chosen as the dependent variable for this study. XgBoost has been deemed the finest ensemble model in terms of both accuracy and prediction time, and FBProphet has demonstrated comparable performance without the need for hyper-parameter tuning. Gomez-Orellana et al. [126] have proposed a novel MTEANN model to forecast wave height and energy flux simultaneously with a 6-hour and 12-hour lead time for six distinct stations in the United States. This paper's primary objective is to validate the effectiveness of a novel strategy known as zonal strategy, in which a single model can predict wave parameters for each station located within an identical zone. MTEANN with zonal strategy outperforms all other methods when compared to MTEANN trained for individual stations first and SVM, Lasso regression with zonal strategy subsequently. Jorges et al. [127] have proposed an innovative method of forecasting and reconstructing wave height using LSTM neural networks and bathymetry data. In comparison to SVR, MLR, FFNN, and RF, Parallel LSTM has been proven to be the preferable model. Incorporating bathymetry data as an input feature enhanced the forecasting performance of these models, justifying the incorporation of this feature in future works. Mahjoobi et al. [128]

have used SVM with both an RBF and a Polynomial Kernel to predict wave height and compared the results to those of MLP and ANN models. Wind characteristics collected with a 6-hour lag were utilized as input features, and data from Lake Michigan was collected over the course of 2 years. The author has demonstrated that SVM with RBF outperforms all other models, and the implementation of SVM is validated due to its capacity for generalization, fewer input parameters, and shorter computational time. Using meteorological data as input characteristics generated from numerical models, Fernández et al. [129] have constructed eight nominal & ordinal classifiers for wave energy & height forecasting in two buoys (4 points around each buoy). The author concluded that, among ordinal classifiers, SVORIM and SVOREX performed exceptionally well in wave height prediction but not satisfactory in wave energy forecasting. Guijo-Rubio et al. [130] has proposed an MTEANN model for forecasting wave energy flux with multiple time horizons (both short and long). Reanalysis data from three buoys in the Gulf of Alaska were used as input in this study. Output for four various time horizons was retrieved from a single model, and the results were compared to those of the SVR and ELM models. According to the author, MTEANN with Su-Li outperforms all other models.

2.5 Summary:

Ocean Energy has tremendous potential as a renewable energy to alleviate the worldwide energy crisis. However, to utilize this source, prior knowledge of significant wave height can be effective as it is directly related to the wave energy flux which will ultimately aid in choosing the appropriate locations to deploy wave energy converters. Although direct measurement is the most precise and accurate method, it is extremely expensive and has seen limited deployment, also both the spatial & time resolution is low. Most underdeveloped nations lack access to these buoys that can predict significant wave height directly and for this research activities are being seriously hampered and wave energy harnessing potential is underutilized. Numerical models could have been an alternative, but they are based on assumptions and

equations, if location changes, subsequently all the assumptions have to be altered for that specific location which is computationally burdensome, accuracy is also less. Finally, after extensive literature review, machine learning models have been found to be the most appropriate method to predict significant wave height from the wind parameters. Two Tree based models and two deep learning models have been selected for this work, and the study area is around the coastlines of USA and Canada. To the best of our knowledge, no other work has incorporated data from more than 14 stations to train models, whereas in this study, data from 47 stations have been taken which covers a huge area, also the efficacy of trained models was evaluated on unknown stations for that model, a novel concept in and of itself.

Chapter 3 Materials and Methods

3.1 Introduction

Before training a machine learning model, a number of factors with significant impact on the model must be considered. The entire modelling procedure of this work is illustrated in Figure 3-1. The first step is data collection, as the model is entirely based on the training data, its performance will have a significant impact if the data is not quality controlled. For this work, standard dataset from 47 distinct stations were collected using ERRDAP and the dataset is provided by NDBC. The second step is selection of the input parameters for the training purpose of the model and this entire work is based on extensive literature review and also justified by employing Pearson Correlation Matrix in Figure 3-6. As the raw dataset contains many outliers and null values which drastically deteriorate models' performance if not removed. So, in the next step, all the null values, missing values were completely removed from the dataset. After that the entire dataset was standardized to transform the input feature range to identical values and thus the bias for a particular input feature is avoided while training the deep learning-based models. The feature distribution before and after standardization is illustrated in Figure 3-5. The impact of outlier removal on model performance was also evaluated in this work. For the outlier removal, Box-Cox transformation technique has been employed and the distribution of the features after the transformation is illustrated in Figure 3-8 and the distribution of the features after outlier removal is illustrated in Figure 3-9. In the next step, entire data set was divided into training data and test data, the training is again divided into training data and validation data. Train and validation data have been used to train the model while test data have been kept aside to measure the performance of the model on a completely unseen data. Hyper-parameter tuning was carried out to find the optimized model and finally the performance for each model is evaluated and contrasted based on three error matrices (MAE, MSE, R^2).

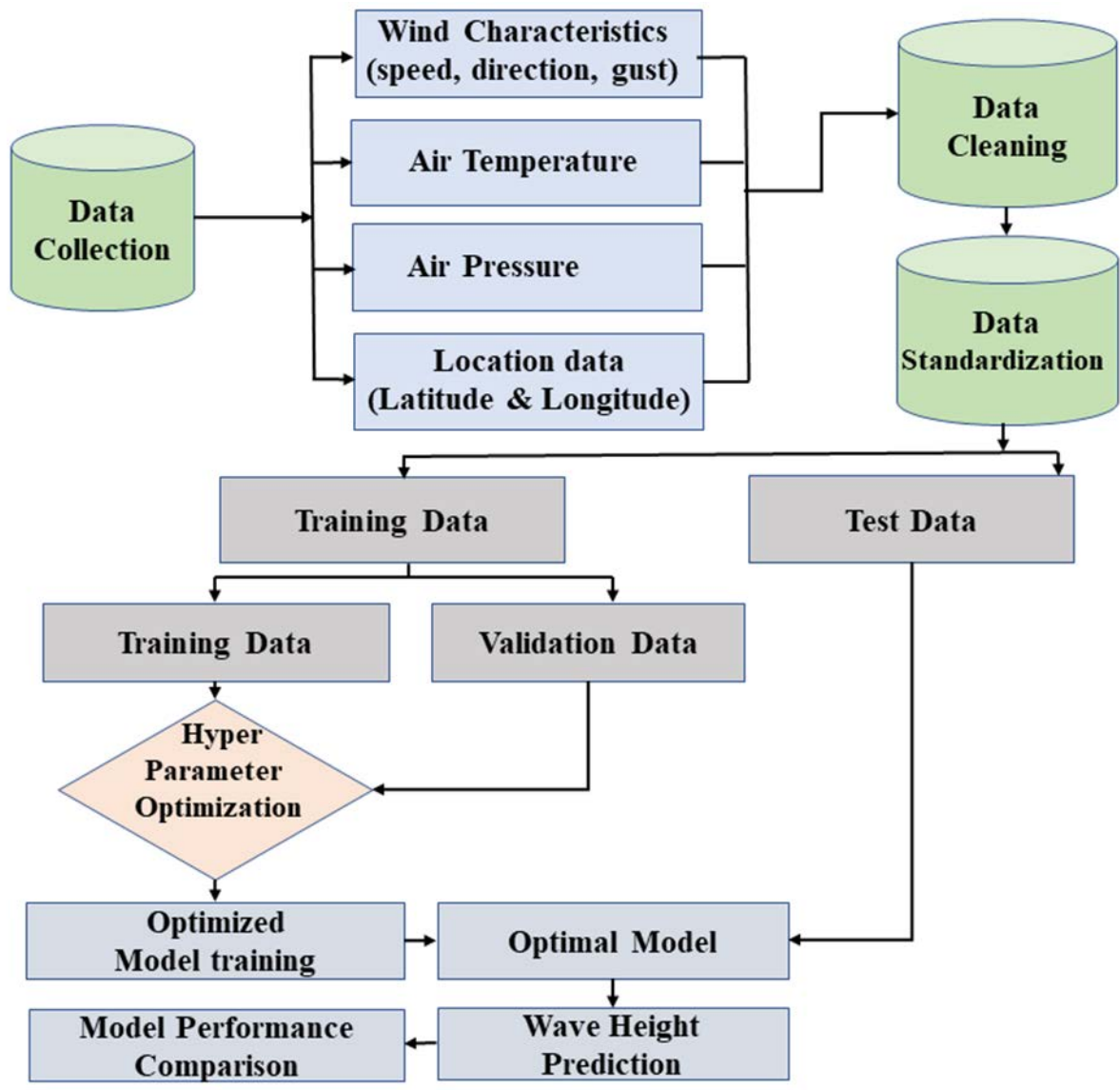


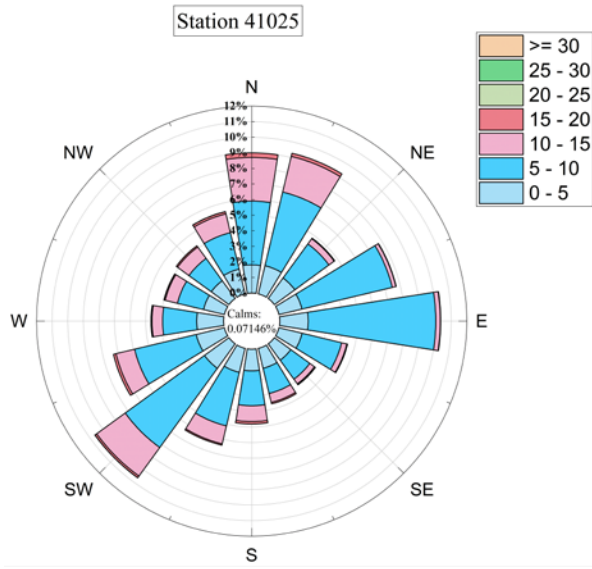
Figure 3-1: Flowchart for Overall Modelling Procedure

3.2 Study Area and Data Collection

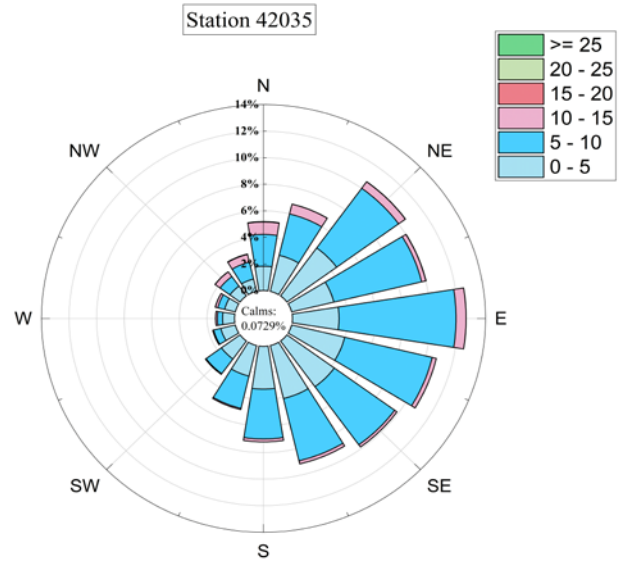
This research examines a vast region mostly along the coastlines of the United States and Canada. For the research, 47 stations are selected, including 1 station from the Bering Sea, 8 stations from the Gulf of Alaska, 2 stations from the North Pacific Ocean (near USA coast), 6 stations from the Caribbean Sea, 9 stations from the Gulf of Mexico, and the remaining 21 stations from the North Atlantic Ocean. The predominant reasons for selecting these regions include the availability and ease of access to datasets in these regions, as well as the heightened research interest in these regions because of the enormous wave & wind energy harnessing

potential these places offer, their effect on the marine transportation & global food chain sector and the recent population, industrial, and economic growth adjacent to these regions [131]. The datasets are provided by the National Data Buoy Center (NDBC) which is governed by the National Oceanic and Atmospheric Administration (NOAA) as a part of their National Weather Service (NWS) program. The National Data Buoy Center (NDBC) operates a global network of data collecting moored buoys extended from the western Atlantic to Pacific, Bering Sea to South Pacific & the majority of these provide a wide variety of wind-related characteristics, including wind direction, speed, gust, barometric pressure, wind, and sea temperature, from which wave parameters may be derived, which is essential for scientific research, meteorological observation, and forecasting. Standardized and slightly modified datasets titled 'NDBC Meteorological Buoy Data' provided by NOAA NMFS SWFSC ERD were obtained through ERDDAP, a site specialized for accumulating scientific data [132]. The input features were selected on basis of previous work after extensive literature review, and also by evaluating their correlation to the target variables. The selected features are listed in Table 1-1. The datasets used in this study were primarily collected using three types of buoys: 3-meter discus buoy, moored buoy, and 3-meter foam buoy which are maintained by NDBC. Specifications of the buoys and description of the measurements have been listed in Table 3-1. In Figure 3-2, both the wind propagation direction and velocity is illustrated by Rose plot for 4 distinct stations denoted by a, b, c, d .From the distribution, it is clear that wind velocity is predominantly between 5 and 10 ms^{-1} and that it blows from distinct directions at various stations. The 'Matplotlib Basemap Toolkit' (a dedicated library of python to generate 2D maps from coordinates of the location) has been employed to visually illustrate the chosen buoys' positions in this study. The Basemap module is provided with input that includes the precise latitude and longitude coordinates of each individual buoy, as listed in Table 3-2. The longitude range provided for the input data spans from 58 degrees West to 178 degrees West, while the

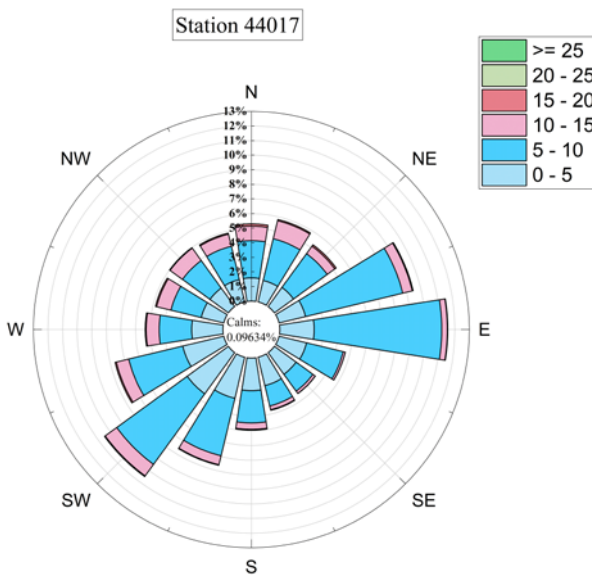
latitude range covers from 15 degrees North to 61 degrees North. Additionally, a standardized bathymetry dataset sourced from ERDDAP is also included as input. This bathymetry dataset furnishes essential information regarding the ocean depth at various points on the map. Consequently, Basemap utilizes the buoy coordinates and bathymetry data to generate a comprehensive map in Figure 3-3 illustrating the precise locations of each buoy in conjunction with the surrounding oceanic topography. From January 2010 to January 2022, over twelve years of data were collected for this study. The dataset has a temporal resolution of 1 hour, with some buoys having a resolution of 10 minutes. Station wise mean properties for the entire dataset have been listed in Table 3-3 and subsequently the deviation of the input and output features are listed in Table 3-4 which shows moderate deviation for each feature in the dataset except Wind direction. The dataset exhibits a substantial presence of null values across different categories. Since the buoy data are real-time weather data, noise and other types of disturbances are also present in the measured dataset. Additionally, the buoy data has a very high degree of variability, as observed from the further investigation: the standard deviation of the data is very high. For wave height, often the standard deviation is half or almost equal to that of the mean. This, along with the presence of outlier, sensor misreading and other forms of indistinguishable contamination renders the dataset quite challenging for our purpose: predicting the significant wave height.



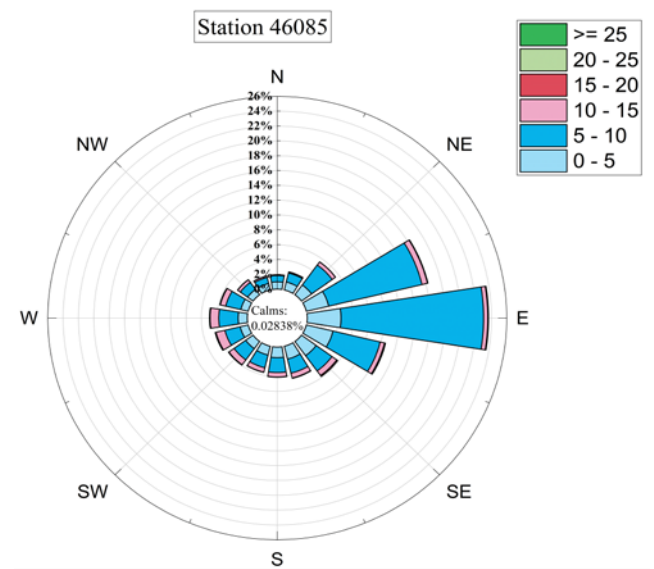
(a)



(b)



(c)



(d)

Figure 3-2 (a-d): Wind Rose plots for different stations

Table 3-1: Buoy specifications and description of measurements

Measuring Height of Air Temperature	Approximately 3.7m above elevation of the site
Height of Anemometer	Approximately 4.1 m above elevation of the site
Measuring depth of Sea Temperature	1.5 m below water line
Radius of Watch Circle	2000 yards avg.
Depth of Water	1900 m avg.
Elevation of the Site	Sea level
Elevation of the Barometer	2.7 m above sea level

Table 3-2: Buoy location and serial number

Buoy Number/ID	Location	Longitudes	Latitudes
41013	33.44 N 77.76 W	-77.76	33.44
41025	35.01 N 75.45 W	-75.45	35.01
41043	21.03 N 64.79 W	-64.79	21.03
41044	21.58 N 58.63 W	-58.63	21.58
41046	23.82 N 68.38 W	-68.38	23.82
41047	27.47 N 71.45 W	-71.45	27.47
41048	31.83 N 69.57 W	-69.57	31.83
41049	27.49 N 62.94 W	-62.94	27.49
41053	18.47 N 66.10 W	-66.10	18.47
41056	18.26 N 65.46 W	-65.46	18.26
42002	26.06 N 93.65 W	-93.65	26.06
42003	25.93 N 85.62 W	-85.62	25.93
42012	30.06 N 87.55 W	-87.55	30.06
42019	27.91 N 95.35 W	-95.35	27.91
42035	29.23 N 94.41 W	-94.41	29.23
42036	28.50 N 84.51 W	-84.51	28.50
42039	28.79 N 86.01 W	-86.01	28.79
42040	29.21 N 88.24 W	-88.24	29.21
42055	22.12 N 93.94 W	-93.94	22.12
42056	19.82 N 84.95 W	-84.95	19.82
42059	15.30 N 67.48 W	-67.48	15.30
42060	16.43 N 63.33 W	-63.33	16.43
42085	17.87 N 66.53 W	-66.53	17.87
44005	43.20 N 69.13 W	-69.13	43.20
44007	43.53 N 70.14 W	-70.14	43.53
44009	38.46 N 74.69 W	-74.69	38.46
44013	42.35 N 70.65 W	-70.65	42.35
44014	36.61 N 74.84 W	-74.84	36.61
44017	40.69 N 72.05 W	-72.05	40.69
44018	42.20 N 70.15 W	-70.15	42.20
44029	42.52 N 70.57 W	-70.57	42.52
44030	43.18 N 70.43 W	-70.43	43.18

44032	43.72 N 69.36 W	-69.36	43.72
44033	44.06 N 69 W	-69.00	44.06
44034	43.50 N 67.88 W	-67.88	43.50
44037	37.00 N 76.09 W	-76.09	37.00
46001	56.30 N 148.02 W	-148.02	56.30
46022	40.75 N 124.53 W	-124.53	40.75
46027	41.84 N 124.38 W	-124.38	41.84
46060	60.59 N 146.79 W	-146.79	60.59
46061	60.24 N 146.83 W	-146.83	60.24
46073	55.01 N 172.01 W	-172.01	55.01
46076	59.47 N 148.01 W	-148.01	59.47
46078	55.56 N 152.64 W	-152.64	55.56
46080	57.95 N 150.04 W	-150.04	57.95
46082	59.67 N 143.35 W	-143.40	59.69
46085	55.88 N 142.88 W	-142.88	55.88

Table 3-3: Station wise mean properties for the parameters

Station Number	Wind Direction (°)	Wind Speed (m/s)	Wind Gust (m/s)	Air Pressure (hPa)	Air Temperature (° C)	Wind Speed Zonal (m/s)	Wind Speed Meridional (m/s)	Significant Wave Height (m)
41013	171.45	6.78	8.31	1017.52	19.77	-0.20	0.61	1.32
41025	187.50	7.32	9.02	1016.78	20.70	-0.05	1.75	1.48
41043	103.33	6.19	7.58	1016.58	26.50	-0.04	-4.85	1.80
41044	108.59	6.04	7.41	1017.41	26.07	0.12	-4.57	1.87
41046	120.15	5.94	7.30	1017.23	26.05	0.46	-3.73	1.74
41047	150.41	6.04	7.41	1018.20	24.46	0.42	-1.64	1.71
41048	186.28	6.74	8.31	1018.13	22.60	0.94	1.00	1.91
41049	142.52	5.67	7.05	1019.04	24.76	0.29	-1.90	1.84
41053	100.74	5.76	7.26	1015.62	28.14	0.18	-5.28	0.99
41056	93.55	5.89	7.40	1015.35	27.83	0.19	-5.37	0.96
42002	130.62	6.05	7.39	1016.17	25.45	1.16	-3.08	1.16
42003	133.45	5.80	7.22	1016.88	24.80	-0.46	-2.49	1.07
42012	167.47	5.55	6.77	1017.47	21.59	-0.21	-0.69	0.76
42019	137.07	6.41	7.85	1016.27	23.52	1.03	-2.54	1.21
42035	150.89	5.78	6.99	1016.67	21.36	0.70	-1.73	0.85
42036	158.81	5.34	6.55	1017.39	23.02	-0.31	-1.24	0.86
42039	165.99	5.55	6.85	1017.44	23.25	0.03	-1.09	0.96
42040	160.36	6.01	7.36	1017.30	22.92	0.00	-1.21	0.94
42055	107.46	6.29	7.58	1014.26	26.30	-0.40	-4.33	1.24
42056	100.07	6.40	7.72	1014.19	27.23	-0.54	-4.79	1.22
42059	92.20	7.39	8.87	1013.54	27.75	-0.04	-6.93	1.57
42060	87.35	6.77	8.12	1014.49	27.36	-0.87	-6.07	1.35
42085	79.22	6.44	7.86	1014.90	28.54	-1.21	-5.71	1.09
44005	198.29	6.81	8.34	1015.36	10.03	0.06	1.46	1.41

44007	199.44	5.62	6.89	1015.02	8.82	0.01	1.35	0.96
44009	195.38	6.35	7.69	1016.49	14.37	-0.20	1.20	1.19
44013	194.28	6.10	7.41	1015.40	10.30	-0.21	1.17	0.91
44014	182.60	6.38	7.77	1017.39	16.41	-0.74	0.81	1.48
44017	203.90	6.74	8.23	1016.21	11.72	0.04	1.92	1.32
44018	192.04	6.30	7.64	1015.57	11.20	0.39	1.17	1.24
44029	195.83	5.83	7.13	1014.82	10.72	-0.08	1.19	0.94
44030	193.19	5.56	6.82	1014.84	9.93	0.12	1.20	1.00
44032	198.52	5.90	7.27	1014.78	8.93	-0.09	1.34	1.15
44033	191.88	5.32	6.66	1014.67	7.76	-0.54	1.20	0.65
44034	198.87	5.88	7.23	1014.94	8.19	-0.23	1.40	1.19
44037	201.15	6.46	7.96	1015.18	10.05	0.22	1.52	1.47
46001	198.90	7.63	9.46	1006.78	7.57	1.59	1.42	2.62
46022	208.99	6.01	7.48	1017.70	11.67	-1.13	0.30	2.43
46027	212.64	5.50	6.92	1017.22	10.69	-0.64	0.19	2.19
46060	168.14	5.25	6.52	1009.29	7.09	0.10	-2.12	0.67
46061	156.55	6.28	7.82	1008.33	7.07	0.28	-2.78	1.45
46073	191.49	8.66	10.60	1005.18	5.46	-0.14	0.43	2.51
46076	161.50	6.35	7.91	1006.89	7.23	-0.41	-1.48	2.00
46078	206.46	8.05	9.98	1007.37	7.46	0.48	2.10	2.58
46080	188.31	7.07	8.74	1006.48	7.87	0.68	0.47	2.25
46082	162.81	6.73	8.44	1008.00	8.01	1.36	-3.04	2.31
46085	194.66	7.38	9.15	1008.22	8.52	1.77	0.84	2.62

Table 3-4: Station wise deviation for the parameters

Station Number	Wind Direction (°)	Wind Speed (m/s)	Wind Gust (m/s)	Air Pressure (hPa)	Air Temperature (° C)	Wind Speed Zonal (m/s)	Wind Speed Meridional (m/s)	Significant Wave Height (m)
41043	57.24	2.24	2.61	2.82	1.56	3.12	3.18	0.62
41044	61.74	2.31	2.68	2.95	1.63	3.10	3.37	0.64
41046	70.65	2.42	2.82	3.34	2.01	3.66	3.69	0.71
41047	90.21	2.98	3.50	4.17	2.95	4.47	4.74	0.86
41048	94.07	3.40	4.07	5.60	3.85	5.33	5.16	1.08
41049	88.85	2.76	3.24	4.21	2.69	4.18	4.31	0.86
41053	33.03	2.03	2.41	2.01	0.90	1.86	2.44	0.23
41056	29.42	1.52	1.92	2.45	0.79	2.31	1.67	0.25
42002	73.80	2.72	3.21	4.37	3.67	4.77	3.22	0.66
42003	94.23	2.72	3.24	3.89	3.56	4.30	4.02	0.68
42012	100.82	2.74	3.27	4.89	6.40	4.43	4.27	0.47
42019	82.31	2.82	3.36	5.33	4.87	5.47	3.42	0.63
42035	87.77	2.71	3.24	5.49	6.63	4.93	3.60	0.41
42036	102.19	2.88	3.42	4.42	5.02	4.22	4.17	0.61
42039	98.88	2.87	3.42	4.42	5.04	4.28	4.41	0.64
42040	100.23	3.02	3.58	4.78	5.34	4.86	4.49	0.60

42055	74.37	2.59	3.03	4.26	2.59	4.25	3.05	0.70
42056	64.03	2.39	2.79	2.78	1.70	3.84	2.95	0.54
42059	27.73	1.87	2.21	2.04	1.07	2.29	2.22	0.49
42060	35.25	2.21	2.58	2.09	1.16	2.56	2.57	0.42
42085	41.63	1.99	2.39	1.97	1.36	2.11	2.65	0.31
44005	100.12	3.60	4.38	8.79	6.80	5.68	5.00	0.91
44007	100.74	3.18	3.84	9.00	8.02	4.87	4.02	0.64
44009	101.16	3.31	4.01	7.46	7.98	5.50	4.42	0.66
44013	97.82	3.31	4.08	8.72	7.98	4.79	4.88	0.73
44014	108.00	3.40	4.14	7.02	7.40	5.70	4.30	0.86
44017	93.71	3.49	4.27	8.12	7.98	4.84	5.53	0.77
44018	95.38	3.45	4.18	8.38	6.80	5.23	4.77	0.85
44029	98.82	3.05	3.76	8.41	7.90	4.68	4.47	0.69
44030	99.21	3.02	3.70	8.39	7.70	4.58	4.21	0.69
44032	99.96	3.31	4.09	8.76	7.25	4.92	4.45	0.73
44033	106.32	3.00	3.76	9.00	7.46	4.57	3.83	0.39
44034	100.07	3.47	4.28	8.83	6.28	4.90	4.54	0.77
44037	96.30	3.38	4.17	8.67	6.91	5.21	4.87	0.95
46001	86.75	3.67	4.49	13.71	4.13	5.29	6.26	1.40
46022	125.73	3.59	4.29	5.98	2.06	6.51	2.30	1.10
46027	107.60	3.96	4.68	6.07	2.05	6.07	2.94	0.98
46060	104.59	3.62	4.36	12.26	4.88	3.30	5.03	0.44
46061	100.07	4.07	4.98	12.25	4.66	4.46	5.33	0.93
46073	103.70	4.04	4.94	14.13	3.84	6.63	6.87	1.47
46076	101.16	3.83	4.61	12.83	4.24	4.53	5.67	1.22
46078	90.71	3.83	4.68	13.23	4.38	5.45	6.72	1.38
46080	92.30	3.83	4.68	12.84	4.00	5.04	6.21	1.29
46082	88.46	4.39	5.31	12.38	4.21	3.50	6.42	1.38
46085	87.48	3.47	4.26	13.39	3.65	5.15	6.01	1.37

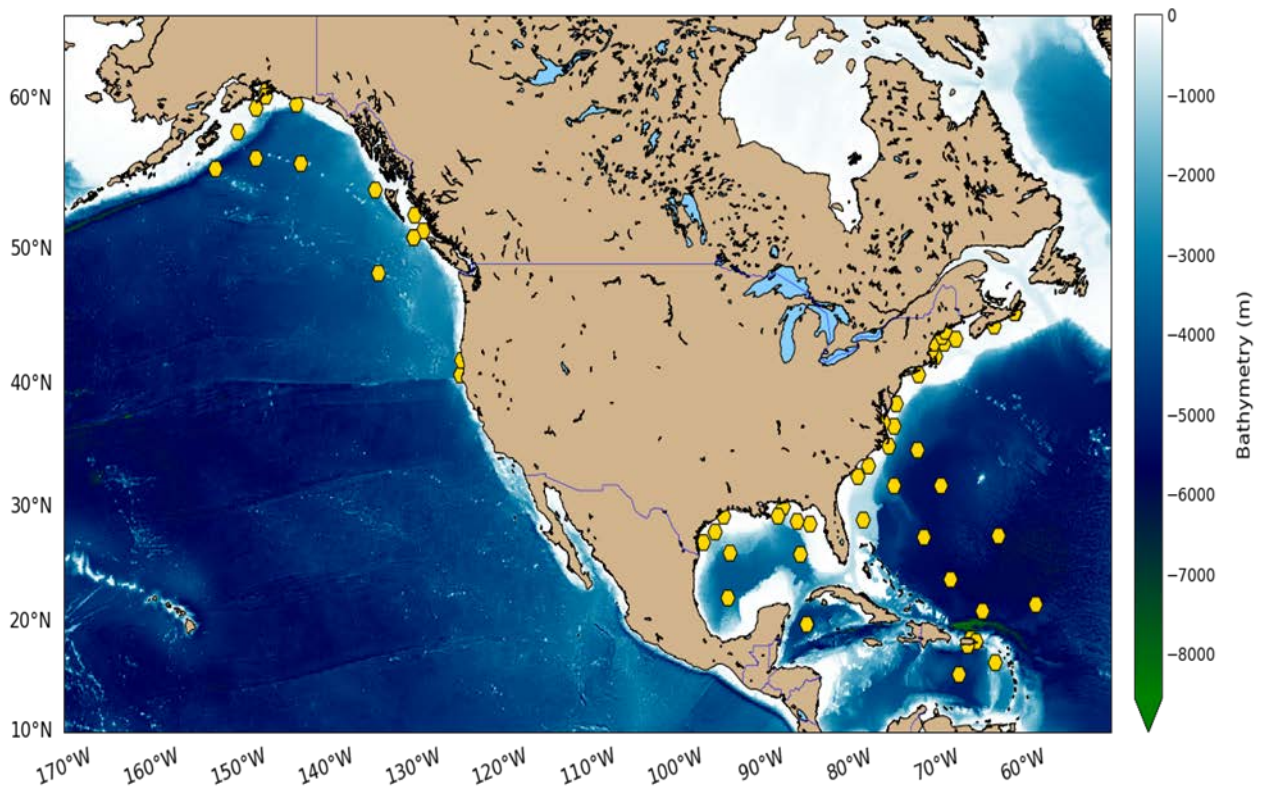


Figure 3-3: Locations & Bathymetry of the buoys

3.3 Data Sets Pre-Processing

The entire preprocessing steps are divide into 5 steps as illustrated in Figure 3-4 which are described in the following sections. The data retrieved from the buoys is formatted in comma-separated values (.csv) and is stored as individual files for each buoy. This format is not suitable for the purpose of training and testing machine learning and deep learning models, as it necessitates a single compiled file. The obtained dataset exhibits a significant quantity of missing values and instances of unsuitable modeling, necessitating the need for data cleaning. On the basis of earlier studies, samples with wave heights less than 0.05 meters were eliminated [23]. Additionally, based on the findings of the literature review, it was noted that incorporating time lags can enhance the efficacy of the model. Therefore, a time lag of 8 hours was implemented. Despite the lack of a substantial correlation between time lag and model performance, subsequent testing revealed that a model trained on a maximum of 8 hours of

time lag exhibits superior performance compared to a model trained on fewer hours of time lag or no time lag whatsoever.

The dataset also contained features of varying scale, i.e., features with a non-uniform range; this makes gradient descent challenging and deep learning models inapplicable. Thus, the data were normalized. The data were then divided into training, test, and validation sets. For modeling and hyperparameter optimization, training and validation data were used. The testing data were only used for final evaluations and were not utilized in any capacity during model training or optimization. The distribution of data from different stations in the training, Testing and Validation Set are listed in Table 3-5 which exhibits quite an equal contribution from each of the station and thus the bias of a particular station is avoided in the modelling procedure. In addition, an exploratory data analysis was conducted.

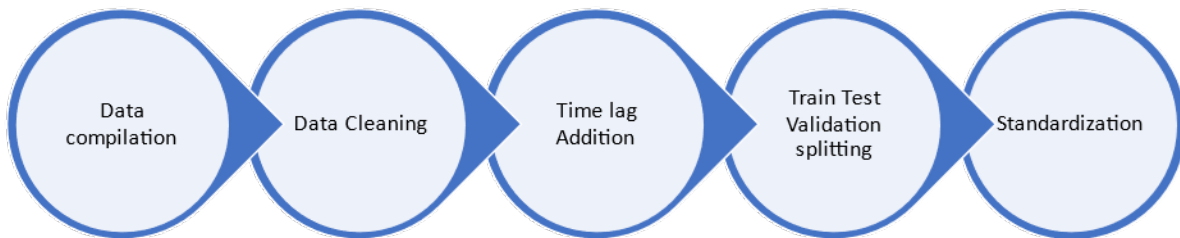


Figure 3-4: Work flow of data processing

Incorporating time latency necessitates the utilization of atmospheric data characteristics from earlier samples. The dataset is sampled every hour; the use of an 8-hour time lapse indicates that the characteristics of the previous 8 samples were considered. Previously, time lags were incorporated into wave height forecasting [16], [119], and the technique was found to also improve the model's performance in the regression task.

The dataset was split into train test and validation parts using the scikit learn's [133] train test split module. The data was first split into training and test set, and then the training set was further split into two datasets: training and validation dataset. The train, test and split dataset

was done in such a way that the amounts of samples from each station on the datasets are proportional. The validation and testing sets contained 75000 samples each and the training data contained 2291184 samples, and the numbers of features were 65 after incorporation of time lag. In addition, for model evaluation, one-year data (ranging from 2022 to 2023) from some buoys as well as data from buoys not included in the training set were considered. Wind speed, wind direction, wind velocity, atmospheric pressure, atmospheric temperature, and the vertical and horizontal components of the wind were taken into account because they are major wave generation elements [134]. Latitude and longitude were also taken into account to add geographical context to the model.

Table 3-5: Train test and validation data distribution among stations

Station	Number of data		
	Train Set	Validation Set	Test Set
41013	40454	1288	1358
41025	29877	1030	949
41043	42881	1392	1346
41044	47682	1526	1547
41046	47601	1518	1567
41047	49559	1628	1633
41048	53239	1755	1741
41049	59562	1880	1903
41053	2027	57	57
41056	1808	64	62
42002	47854	1571	1549
42003	44912	1472	1465
42012	58024	1949	1991
42019	46452	1539	1566
42035	62366	2054	2067
42036	61796	1988	2062
42039	55151	1803	1756
42040	32365	1056	999
42055	59950	1912	2015
42056	43644	1411	1421
42059	43553	1450	1424
42060	48647	1635	1593
42085	1472	53	43
44005	34091	1120	1077

44007	71669	2371	2420
44009	40748	1276	1343
44013	78470	2497	2583
44014	52388	1750	1718
44017	38467	1239	1262
44018	45453	1541	1456
44029	58634	1954	1875
44030	68074	2212	2208
44032	77389	2590	2573
44033	67463	2179	2169
44034	72423	2379	2365
44037	44421	1443	1476
46001	62510	2138	2034
46022	41561	1355	1430
46027	60637	1988	2013
46060	63486	2059	2013
46061	67913	2235	2234
46073	28699	933	924
46076	61177	2011	2000
46078	40129	1371	1354
46080	42479	1363	1340
46082	48398	1576	1566
46085	43629	1389	1453

The dataset was scaled by using standardization process, standardization includes computing the mean and standard deviation of the data, subtracting the mean and then dividing by standard deviation of the dataset. The standardization process can be expressed by equation (3.1).

$$z = \frac{x - \mu}{\sigma} \quad (3.1)$$

where,

x = actual value of the data point

μ = mean of the dataset

σ = standard deviation of the dataset

x = calculated value of the data point

Standardization scales all data points in a dataset to the same scale, ensures that all data points have the same scale, and can be used to detect outliers. It aids significantly in gradient descent by preventing the optimizer from being affected by variables with larger scales. Moreover, it prevents the gradient from exploding or vanishing. Standardization is performed under the assumption of a Gaussian distribution, but it can be applied to datasets with other distributions. In Figure 3-5 (a-f), the distribution of input features and target variable is illustrated both before and after standardization. Although the range of the different features have become identical after standardization, the shape of the distribution remains unchanged for all the features.

As a component of the exploratory data analysis, a correlation matrix was generated in conjunction with a histogram, box plot, and probability plot. The correlation matrix illustrated in Figure 3-6 indicates a strong positive correlation between wave heights and various wind features, including wind speed, wind gust, and the vertical and horizontal components of wind. Moderate correlations have been observed between wave height and the remaining parameters.

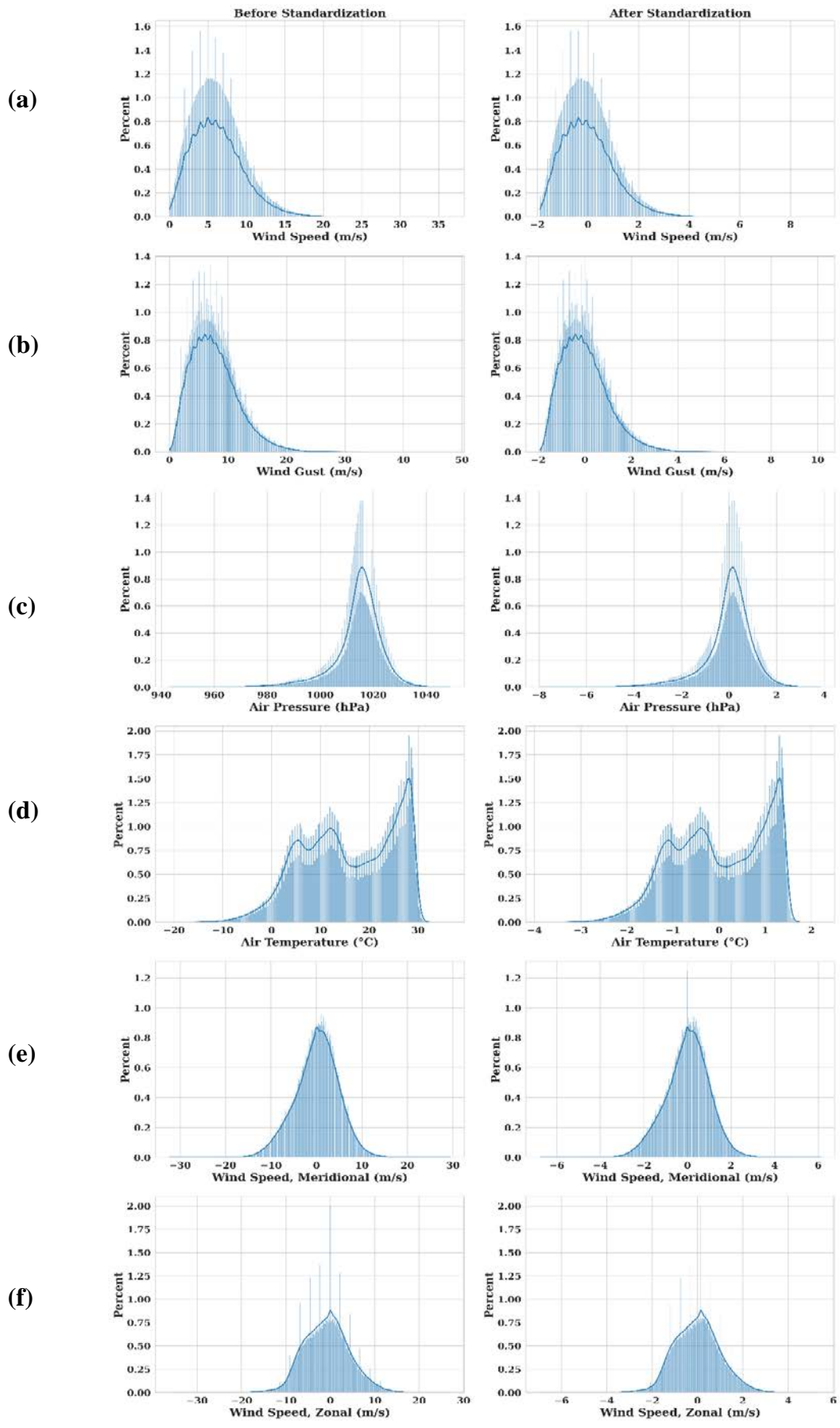


Figure 3-5 (a-f): Distribution of features before and after standardization

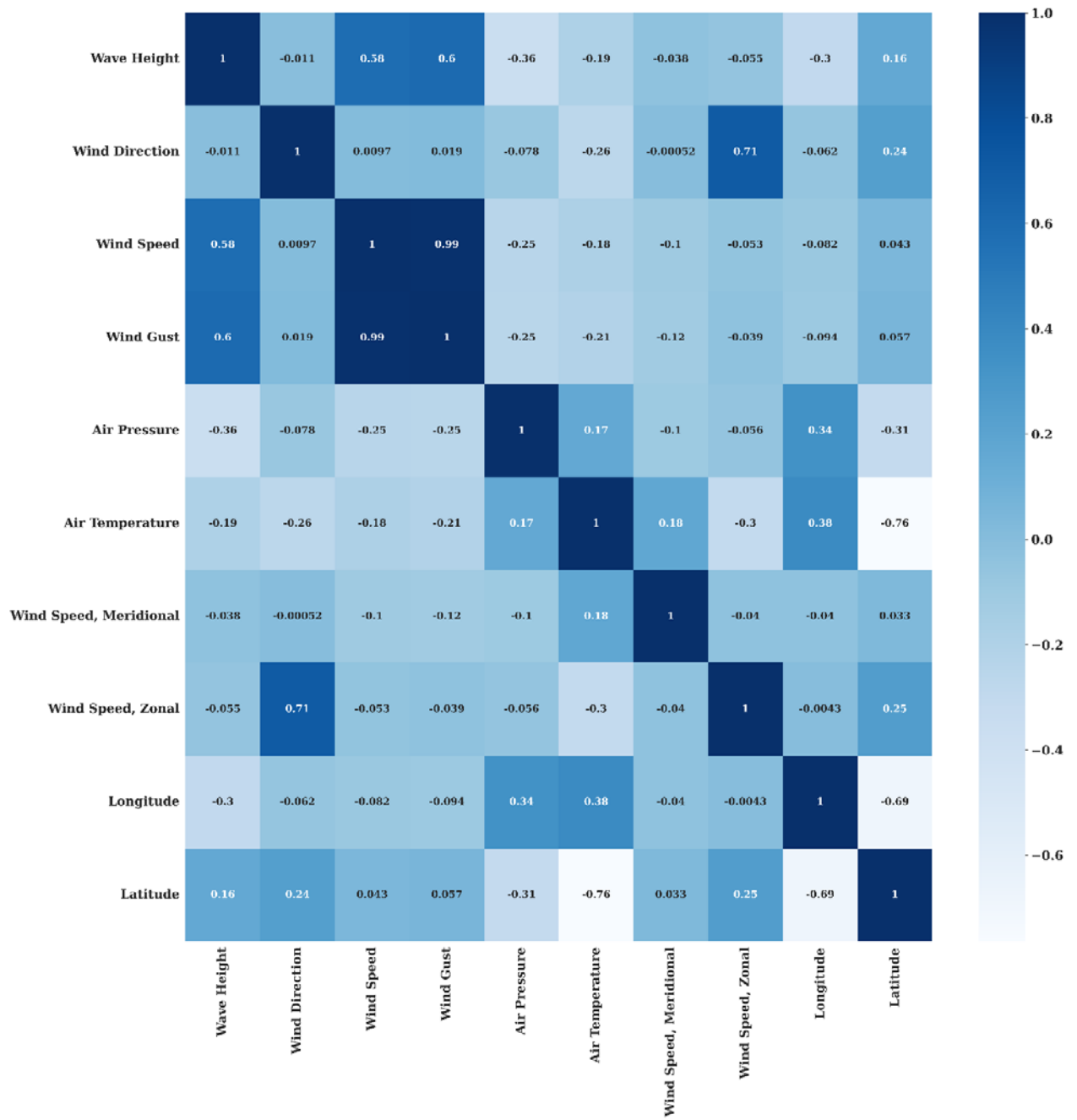


Figure 3-6: Correlation Matrix for the selection of input features

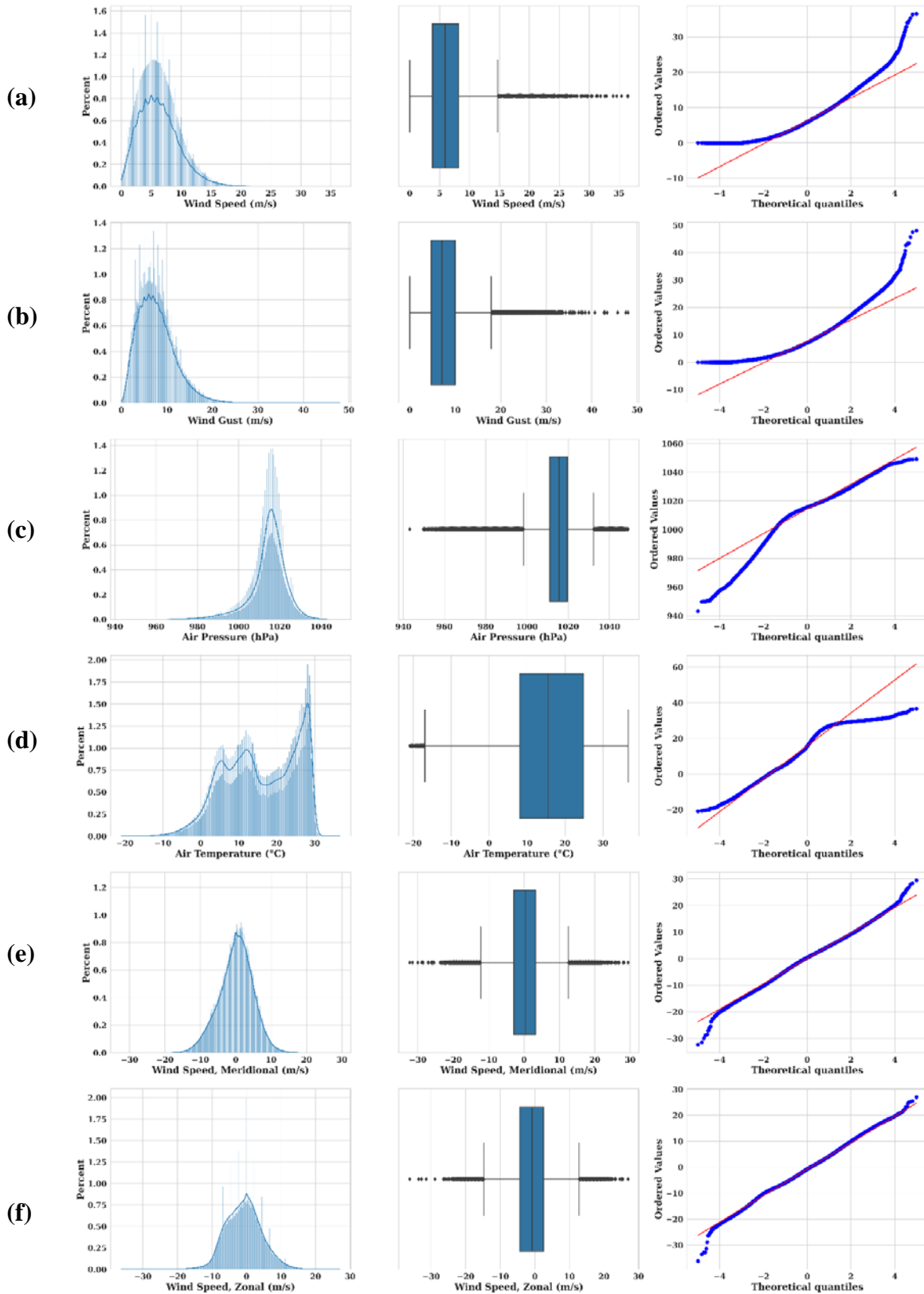


Figure 3-7 (a-f): Histogram, box plot and probability plot

The histogram in Figure 3-7 for 6 variables denoted by (a-f) illustrates the distribution of various features and indicates that, with the exception of the Meridional and Zonal components

of winds, all features exhibit a skewed distribution to either the left or right. The box plot reveals that a substantial quantity of data lies beyond the whiskers, suggesting the existence of outlier values in all features, except for air temperature. The presence of curved lines in a probability plot or quantile-quantile plot indicates nonconformity of the data to a normal distribution. The q-q plot indicates that wind gust and wind speed have a higher tail, suggesting that the outliers are larger than the normal values. Conversely, the lower tail for the remaining variables indicates that the outer value is lower than the normal distribution.

3.4 Outlier Removal with Box-Cox method:

In statistics and data analysis, an outlier is a data point that substantially deviates from the expected or normal distribution of the dataset. It is an observation that deviates from other values in a random sample from a distribution by an abnormal distance. Outliers can be caused by a number of factors, including measurement errors, environmental anomalies, sensor limitations, data entry errors, natural variations, and rare events.

Outliers can have a significant impact on statistical analyses and models due to their ability to bias results and distort data interpretation. They can have an impact on the estimation of parameters, calculation of summary statistics, and performance of machine learning algorithms.

In this study the outliers were detected using the Tukey's Fence method. The Tukey's Fence method involves the following steps:

1. Calculation of the first quartile (Q1) and the third quartile (Q3) of the dataset.
2. Calculation of the interquartile range (IQR) by subtracting Q1 from Q3,

$$\text{IQR} = \text{Q3} - \text{Q1}.$$

3. Determining the lower fence (LF) by subtracting 1.5 times the IQR from Q1,

$$LF = Q1 - 1.5 \times IQR.$$

4. Determining the upper fence (UF) by adding 1.5 times the IQR to Q3,

$$UF = Q3 + 1.5 \times IQR.$$

5. Identifying any data points that fall below the lower fence or above the upper fence.
6. Classifying the outliers as mild outliers or extreme outliers based on their distance from the fences:

- Mild outliers: Data points that lie between 1.5 times and 3 times the IQR beyond the fences.
- Extreme outliers: Data points that lie beyond 3 times the IQR.

Tukey's Fence method requires a normal distribution of the data, in our study the data didn't have normal distribution. The data was converted to normal distribution using scikit learn's quantile transformer. In Figure 3-8 (a-f), distribution of the variables before and after transformation is depicted and it is evident from the illustration that all the input features are transformed into a normalized shape in this step. Then outlier analysis was done for 6 variables:

- 1) Wind Speed
- 2) Wind Gust
- 3) Air Temperature
- 4) Air Pressure
- 5) Wind Speed, Meridional
- 6) Wind Speed, Zonal

Following Outlier analysis, a total of 7257 samples were found to be extreme outliers and a further 71528 samples were found to be mild outliers. Both kinds of outliers were removed. 71528 samples were removed from the dataset. In Figure 3-9 , distribution of the variables after outlier removal using Tukey's Fence method have been depicted. For wind speed, gust and zonal component of wind speed, more normalized distributions after outlier removal have been observed and the shapes of the remaining variables remain almost same before and after outlier removal.

Following outlier removal, two models: Artificial Neural Networks (ANN) and Self-Normalizing Neural Networks (SNN) were trained and compared with similar models trained on non-outlier removed data to evaluate the outlier removal process's impact.

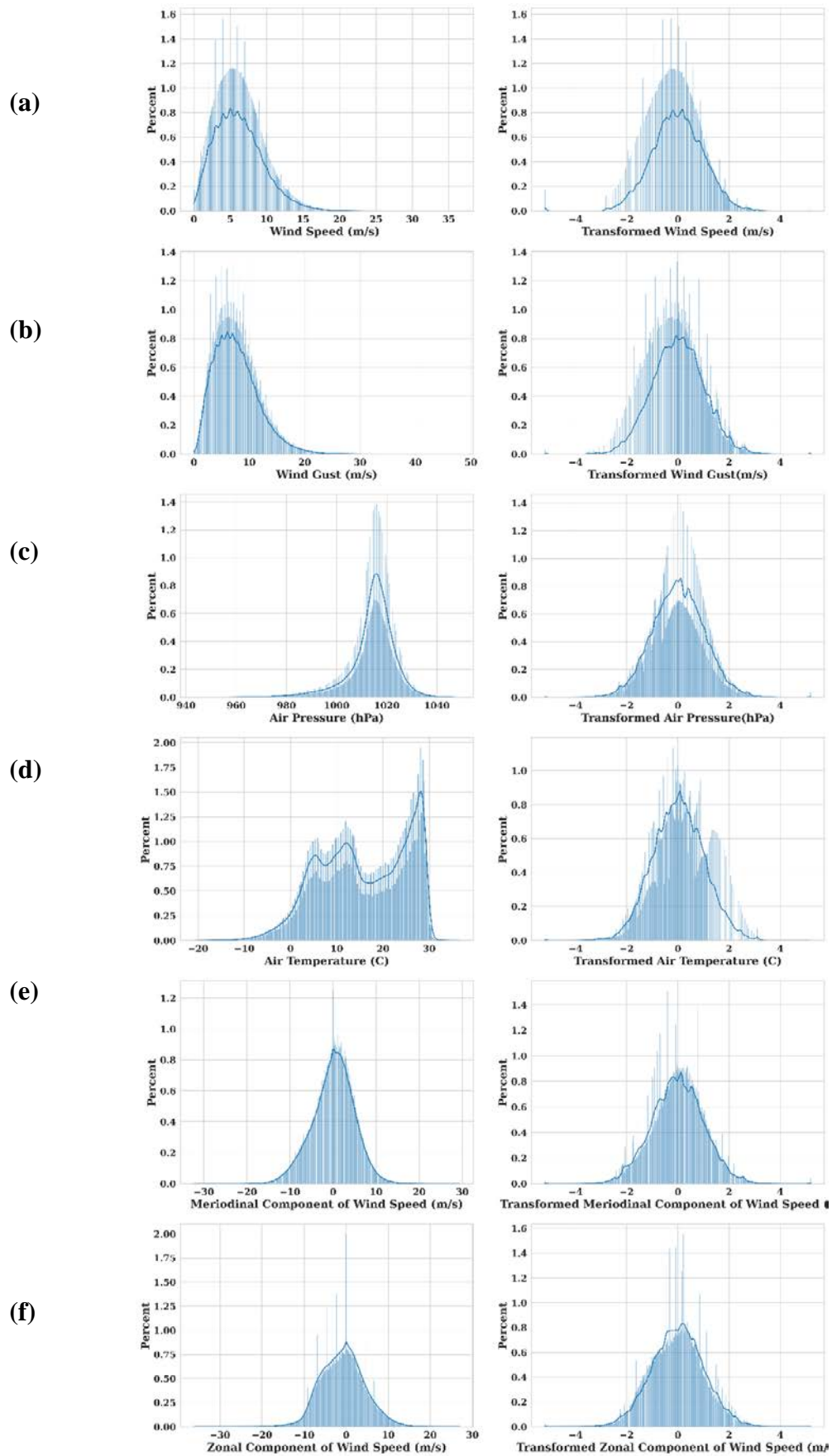


Figure 3-8 (a-f): Distribution of variables before and after transformation

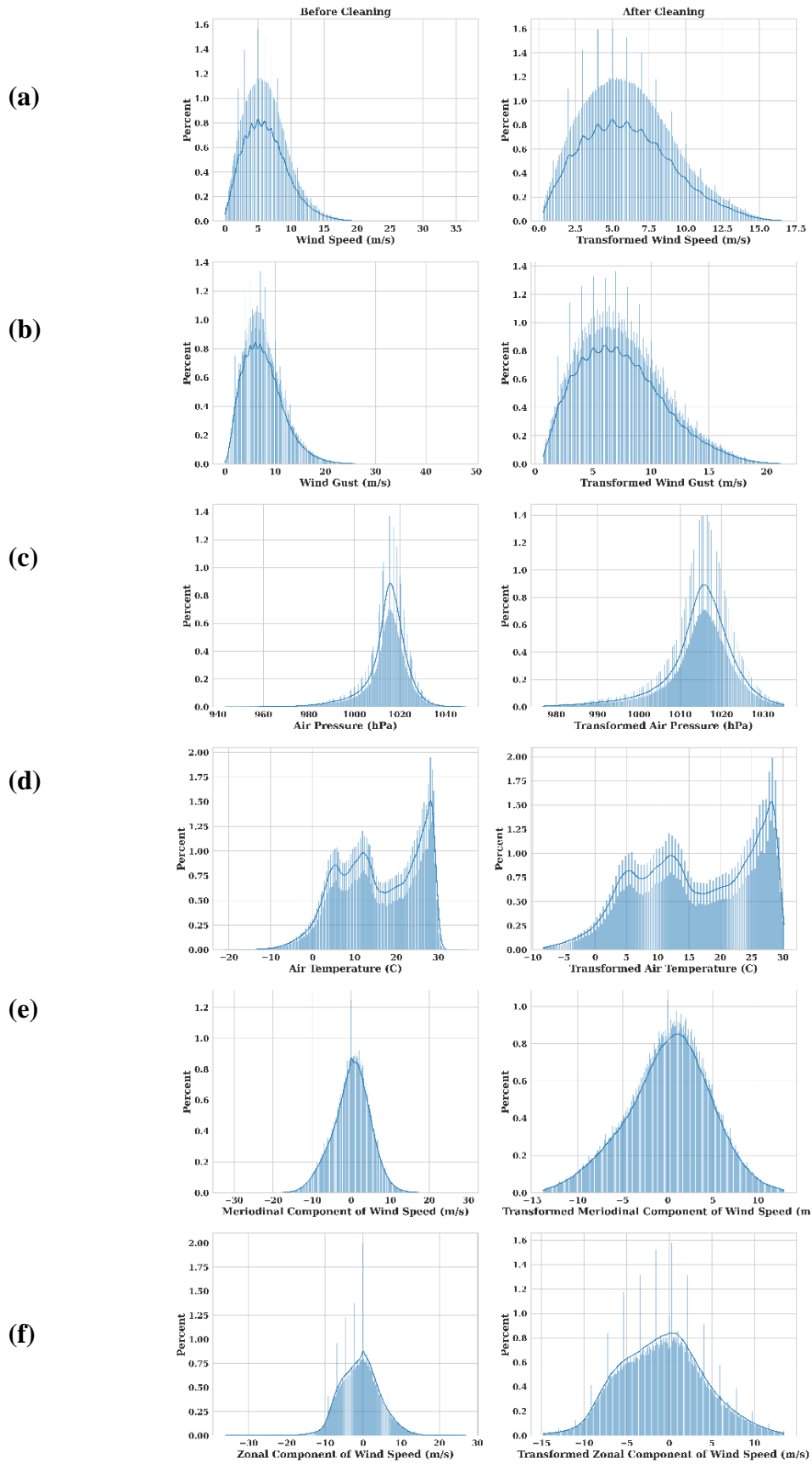


Figure 3-9 (a-f): Distribution of variables before and after outlier removal

3.5 Machine Learning Models to Predict Significant Wave Height

3.5.1 XGBoost

It is a ML model extensively used for classification and regression related problems in supervised learning tasks. XGBoost [135] is an improved version of Gradient Boosting[136], although both implements gradient descent to reduce the error in ensemble tree architecture, a strong regularization (L1 & L2) substantially reduces the overfitting problem in XGBoost. It is a widely used algorithm by data scientists because of its minimal time of training, high performance capabilities in large tabular & structured dataset. XGBoost uses a sequential ensemble model known as boosting instead of Random Forest's parallel ensemble model, which is frequently referred to as bagging.

Both bagging and boosting are components of an ensemble model in which, rather than counting on a single predictive model, a collection of weak models are trained to make distinct choices, and then the choices of multiple weak models are incorporated to generate a robust model. Bagging employs a popular technique known as Bootstrapping, in which random subsets of the provided dataset are generated to train each weak learner independently and same data can be used in the sample multiple times. Then, all the weak models are independently trained in parallel on this bootstrapped dataset to make predictions and finally a robust model is built by aggregating these predictions considering all the models are of equal weight. Although the Bagging method can perform well for weak learners with high variance and low bias, it typically performs poorly on stable datasets with low variance, computationally expensive for larger datasets as well. Boosting on the other hand transforms a set of weak learners into strong learners in a sequential training method by reducing error into each iteration. At first, the data is used to fit a simple base classifier considering all the data of equal weight. To enhance the present model's performance, it fits a new weak classifier without

altering the prior one. Each new classifier has to take into account the data points in which its predecessors performed poorly and higher weightages are assigned to those points for to correct those in the next iteration. Each model learns from its predecessor, assigned a weight based on its performance and finally weighted average of all the models are incorporated to make a robust model. Boosting is typically implemented in stable models with high bias and low variance. There are several boosting algorithms such as Adaptive boosting which works on updating weights on the misclassified data in each iteration until a strong prediction is done (simple boosting), Gradient boosting which changes the hyper parameters to improve the loss function in each iteration, XGBoost which is an optimized version of gradient boosting technique. For the regression problem using XGBoost, an initial prediction (average value of the target variable) is made by the base model. After calculating all residuals from predicted and observed values, an initial decision tree is constructed and a binary split of the tree is applied. To identify the optimal split of a leaf, a similarity score is assigned to each of the right, left, and upper leaf based on residuals. Using this similarity score, the gain value for various splitting conditions is determined, and the split with the highest gain is chosen. Using the residuals, an outcome value for every individual leaf is also computed. The Tree building process for XGBoost is depicted in Figure 3-10. For the classification problem, output of each leaf will be the similarity score whereas for regression type model it will be average of residuals for that leaf. If this is a classification problem, log of odds method is further used to calculate base model output. Then the output for this tree is again compared to the observed values to evaluate the residuals and subsequently using these residuals, a new tree is constructed in a similar manner and this iteration continues until the residuals cease decreasing or a predetermined number of iterations have been completed. Finally, from the whole set of decision trees, the final prediction is the summation of base model output with the product of

learning rate and each of decision tree output. Here are some of XGBoost's distinctive features that make it an optimized model.

- **Regularization:** Because of being an ensemble decision tree, sometimes trees can get complex, XGBoost offers the capability to penalize intricate models via L1 and L2 regularization and circumvent overfitting as well.
- **Sparsity Awareness:** Utilizing a distinct split algorithm that can detect various sparse patterns in the dataset, it manages sparse values to that are frequently the result of several preprocessing steps, such as removing null values and one-hot encoding and ensures optimum split of the node.
- **Blocks for Out-of-core Computation:** When XGBoost is provided with a dataset that is too large for main memory, the dataset is compressed and stored in separate blocks. In lieu of accessing them from the hard drive, which is quite sluggish, these small dataset units are decompressed in the main memory only when needed, resulting in a faster response.
- **Weighted quantile sketch:** When the data points are equally weighted, the majority of tree-based algorithms can identify the splitting points using quantile sketch method, but they cannot deal with weighted data. XGBoost provides a distributed weighted quantile sketch algorithm that is capable of dealing with weighted data.
- **Cache awareness:** This algorithm has been designed to maximize the use of hardware resources by designating internal buffers for gradient statistics in each thread.
- **Parallel Learning:** This divides the data into multiple segments that can be utilized concurrently for tree generation.

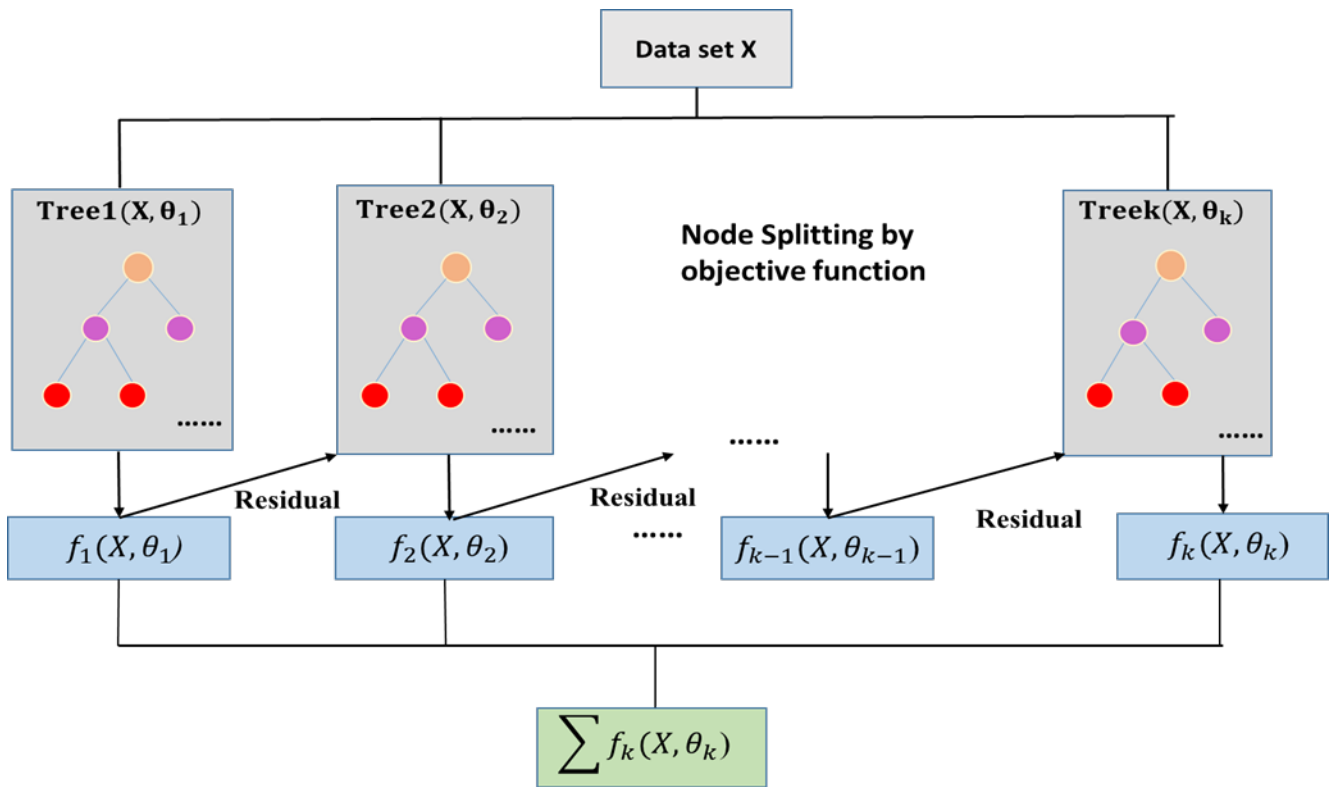


Figure 3-10: XGBoost tree building process

Assuming D as a given data set.

$$D = x; y, |D| = n, x \in \mathbb{R}^m, y \in \mathbb{R} \quad (3.2)$$

n = no of examples in D ,

m = no of features,

y =target variable

x = features

In our dataset, $n= 3354160$ observations

$m= 12$ features (6 actual features and 6 synthetic features generated by time lag addition)

A tree ensemble model predicts output by combining results from all the trees. Assuming

K = no of additive functions

$$\hat{y}_i = \sum_{k=1}^k f_k(x_i), f_k \in F \quad (3.3)$$

Where \hat{y}_i = prediction of the target variable at i-th instance, k-th boost

x_i = i-th sample of the training dataset

$f_k(x_i)$ = output of the k-th tree

F= all the output values of the decision tree

Loss Function,

$$L_k = \sum_{i=1}^n L(\hat{y}_i, y_i) \quad (3.4)$$

To increase the effectiveness & overall model performance, XGBoost employs several hyper-parameters.

The Objective Function is defined as,

$$\text{Obj} = \sum_{i=1}^n L(\hat{y}_i, y_i) + \sum_{i=1}^k R(f_i) \quad (3.5)$$

Where,

L = Loss function which assesses the performance of the model

\hat{y}_i = Predicted Value

y_i = Observed Value

$R(f_i)$ = penalizes the complexity of the model and reduces overfitting

Now, Function of a tree is defined as, $f(x)$ to define the complexity

$$f(x) = w_{q(x)}, w \in \mathbb{R}^T, q: \mathbb{R}^m \rightarrow \{1, 2, \dots, T\} \quad (3.6)$$

Here,

$w_{q(x)}$ = Leaves scores vector

q = mapping functions that map input data to the leaves

T = number of leaves

Formula for penalizing models' complexity is given below,

$$R(f) = \gamma T + \alpha (\|w\|) + \frac{1}{2} \lambda (\|w\|^2) \quad (3.7)$$

Here,

γ, λ = Hyper-parameters or constant coefficient

$\|w\|^2$ = L2-norm of the weight of the leaf which is controlled by λ

$\|w\|$ = L1-norm of the weight of the leaf which is controlled by α

T = total number of leaves

Combining Equation (3.5), Equation (3.6), Equation (3.7) we get,

$$\text{Obj}^{(t)} = \sum_{i=1}^n [L(y_i, \hat{y}_i^{(t-1)}) + f_t(x_i)] + R(f_t) + \text{constant} \quad (3.8)$$

The second-order Taylor approximation is computed as we do not know a derivative for every objective function.

$$\text{Obj}^{(t)} = \sum_{i=1}^n [L(y_i, \hat{y}_i^{(t-1)}) + g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i)] + R(f_t) + \text{constant} \quad (3.9)$$

Here,

$$g_i = \partial_{\hat{y}_i^{(t-1)}} L(y_i, \hat{y}_i^{(t-1)}) \quad (3.10)$$

$$h_i = \partial^2_{\hat{y}_i^{(t-1)}} L(y_i, \hat{y}_i^{(t-1)}) \quad (3.11)$$

After the removal of constant term and subsequently addition of the regularization factor gives the final form of the objective function which is for t^{th} step.

$$\text{Obj}^{(t)} = \sum_{i=1}^n [g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i)] + \gamma T + \alpha \sum_{j=1}^T \omega_j + \frac{1}{2} \lambda \sum_{j=1}^T \omega_j^2 \quad (3.12)$$

3.5.2 LightGBM

LightGBM is a widely used tree-based algorithm that relies on the gradient-boosting sequential ensemble approach & performs well both in classification and regression problems [137]. Gradient boosting is a widely used approach in which each successor in the tree is built and adjusted based on the residual errors of its predecessors, resulting in an algorithm that is both efficient and accurate. In comparison to XGBoost, LightGBM decreases calculation time by a substantial margin due to its utilization of unique characteristics. LightGBM expands leaf-wise meaning that, given a condition, just a single leaf is divided based on the gain whereas XGBoost expands level-wise which costs a lot of computation time. Also, depth is restricted for the leaf-wise growth to control the overfitting which makes it more efficient. LightGBM employs a histogram-based technique in which data is bucketed into bins using a distribution's histogram, as opposed to XGBoost's pre-sorted algorithm, which is highly inefficient in terms of complexity and CPU demand. Additional feature of LightGBM is exclusive feature bundling, whereby the algorithm bundles exclusive characteristics to minimize complexity, hence making it more rapid and effective. LightGBM uses Gradient-based One Side Sampling (GOSS) to sample the dataset, wherein instead of putting all the data points, data points are arranged first best on their gradient value in a descending order, from the first 80%, which has a higher error percentage and should be given most priority, 20% data points are taken

randomly, and from the remaining 20%, only 10% is fed. This strikes a balance between precision and processing time. Light GBM is an efficient approach for processing large-scale data and features because, unlike previous GBDT-based methods like XGBoost and GBDT, it would build the tree vertically rather than horizontally.

Provided that, a supervised training set, $X = \{(x_i, y_i)\}_{i=1}^n$ Suppose a certain function $f^*(x)$, for this function LightGBM strives to find an appropriate approximation function $\hat{f}(x)$ that will minimize the lost function $L(y, f(x))$ as described below:

$$\hat{f} = \arg \min_f E_{y,x} L(y, f(x)) \quad (3.13)$$

After integrating various T regression trees, LightGBM reaches to the final model,

$$f_T(X) = \sum_{t=1}^T f_t(X) \quad (3.14)$$

$$w_{q(x)}, q \in \{1, 2, \dots, J\} \quad (3.15)$$

This represents the regression trees. Here,

J = Number of leaves

q = Decision rules of the tree

w = Vector that represents leaf node's sample weight

At step t, an additive form training which is described below would be done to LightGBM,

$$\Gamma_t = \sum_{t=1}^T L(y_i, F_{t-1}(x_i) + f_t(x_i)) \quad (3.16)$$

LightGBM uses a fast approximation of the objective function based on Newton's technique.

Removing the constant term in Equation 3.16 gives a simplified form:

$$\Gamma_t \cong \sum_{i=1}^n (g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i)) \quad (3.17)$$

Where

g_i = First order gradient statistics of loss function

h_i = Second order gradient statistics of loss function

Considering I_j = Sample set of leaf j and transferring Equation 3.17 as follows:

$$\Gamma_t = \sum_{j=1}^j ((\sum_{i \in I_j} g_i) w_j + \frac{1}{2} (\sum_{i \in I_j} h_i + \lambda) w_j^2)) \quad (3.18)$$

The extreme value of Γ_k and the optimum leaf weight scores of each leaf node w_j^* for a given tree structure $q(x)$ can be found in this way:

$$w_j^* = -\frac{\sum_{i \in I_j} g_i}{\sum_{i \in I_j} h_i + \lambda} \quad (3.19)$$

$$\Gamma_T^* = -\frac{1}{2} \sum_{j=1}^j \frac{(\sum_{i \in I_j} g_i)^2}{\sum_{i \in I_j} h_i + \lambda} \quad (3.20)$$

Here,

Γ_T^* =Scoring function for measuring tree structure quality. Adding the split gives the final form of the objective function:

$$G = \frac{1}{2} \left(\frac{(\sum_{i \in I_L} g_i)^2}{\sum_{i \in I_L} h_i + \lambda} + \frac{(\sum_{i \in I_R} g_i)^2}{\sum_{i \in I_R} h_i + \lambda} - \frac{(\sum_{i \in I} g_i)^2}{\sum_{i \in I} h_i + \lambda} \right) \quad (3.21)$$

Here the left and right branches' samples are denoted by I_L and I_R

3.5.3 Artificial Neural Network (ANN)

The artificial neural network is a type of soft computation that takes inspiration from biological neurons and can approximate any function. The multilayer perceptron model, which was

introduced in 1957[138], is widely regarded as the earliest iteration of artificial neural network. In 1980[139], the introduction of the backpropagation algorithm revived interest in artificial neural networks. Diverse types of neural networks have since been developed for an extensive range of applications, including regression and classification tasks, as well as more complex endeavors such as question answering and the generation of artificial images from text prompt, among many others.

The artificial neural networks are universal approximators, which means they are capable of approximating any continuous function of arbitrary precision. In this study, a complex multilayer perceptron network with backpropagation was used as an artificial neural network.

Overfitting, exploding and vanishing gradients, sparse gradients, and slow convergence are just some of the challenges that can arise during the training of deep neural networks. Various strategies have been implemented to address such problems.

Typically, a neural network consists of an input layer, an output layer, and several concealed layers in between. Determining the appropriate number of layers and nodes per layer is a hyperparameter selection process that typically involves experimentation, practical application, or hyperparameter optimization. The nodes are assigned weights and biases, which are then fed into an activation function. The resultant output is then fed to the subsequent nodes. The process being referred to is commonly known as forward propagation. Mathematically,

$$Z = (W \times X) + b \quad (3.22)$$

$$A = f(Z) \quad (3.23)$$

Where,

- Z is the weighted sum of the inputs, including the bias term.
- W represents the weight matrix connecting the inputs to the neuron.

- X is the input vector or matrix.
- b is the bias vector.
- A is the output or activation of the neuron.
- f() is the activation function applied to the weighted sum

The output then computed from the neural network is then compared with the actual target values, a loss function is used to compute the loss. Then the gradients are calculated and the weights and biases are changed using backpropagation algorithm. The backpropagation algorithm employs some optimization algorithm to optimize the cost. The most common optimization algorithms are: Stochastic Gradient Descent, RMSprop, Adam, Nadam etc.

The back propagation algorithm starts with the calculation of the loss using some form of loss function. The loss function is a function that quantifies the quantity of prediction error. It measures the deviation between the predicted value and the actual value. For the particular regression task, the loss function is given in Equation 3.24.

$$L = \frac{1}{2n} \sum_x \|y(x) - a^L(x)\|^2 \quad (3.24)$$

Where,

n = total number of training data points,

y(x)= desired output

L= number of layers

$a^L(x)$ is the output of the network when x is input

Then the gradient of the loss with respect to each weight $\frac{\delta L}{\delta w}$ and bias $\frac{\delta L}{\delta b}$ are calculated. Then the weights and biases are updated.

$$W_{\text{new}} = W - l \frac{\delta L}{\delta W} \quad (3.25)$$

$$b_{\text{new}} = b - l \frac{\delta L}{\delta b} \quad (3.26)$$

Where,

W_{new} is the updated weight

W is the old weight

l is the learning rate

b_{new} is the updated bias

b is the old bias

3.5.3.1 Activation:

Activation function is a mathematical function applied to the output a neural network node, before passing it to the next layer. In the experiment Rectified Linear Unit or ReLU [140] is used as activation function.

The function is,

$$f(X) = \max(0, X) \quad (3.27)$$

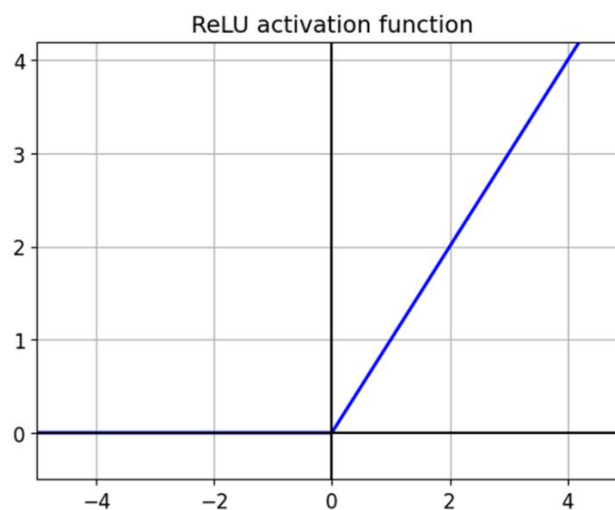


Figure 3-11: ReLU activation function

Without non-linearity, deep neural networks behave as a single-layer network. The ReLU activation function is depicted in Figure 3-11 which mitigates the problem of vanishing gradients and prevents weight saturation.

3.5.3.2 Weight Initialization:

The behavior of a neural network is determined by its initial weights. If all weights are initialized with the same value, either zero or one, symmetry will result, meaning the weights of all layers will be identical. Gradients will be identical for each weight, so the updated weights will remain unchanged. Therefore, weight initialization is required.

Modern initialization techniques offer superior performance to random initialization and accelerate model convergence.

In this study, He initialization [141] was used for the weight initialization at the beginning of the training. He initialization is suitable initializer for the deep neural networks with ReLU activation function. It also addresses the problem of vanishing or exploding gradient often encountered while training very deep neural networks.

3.5.3.3 Optimizer:

Optimizers are used to change and update the weights on every iteration of a deep neural network training. The optimizers are used to calculate the gradients of loss with respect to weights and biases, which are then used to update the model weights and biases. Many optimizer models are proposed and widely used; most optimizers use gradients for optimization while some optimizers employ momentum as well.

Adam optimizer[142] is a stochastic gradient based adaptive learning optimizing algorithm that is based on adaptive implementation of first order and second order moments. Adam optimizers are computationally efficient, has low memory requirements and are suitable for problems with large data or parameters.

The adaptation of adaptive learning rate as well as momentum makes Adam optimizer a powerful optimizer algorithm capable of fast convergence and address issues like noisy gradient, flat region, sparse data. Adam optimizers are also less sensitive to learning rate changes, and are used in a wide variety of applications

3.5.3.4 Dropout:

Very large neural networks are prone to overfitting, where the model performs too well on the training dataset, but does not generalize well outside of the training data. Regularizing is a method to address the problem of overfitting, where the model is penalized for overfitting or fitting the training data too well. Dropout[143] is a regularizing method. Dropout is performed to counteract the effect of overfitting and to strengthen the model. During training, random weights are deleted from a neural network by randomly setting weights of neuron activations to zero, making the model more robust and less susceptible to outliers. The intensity of dropout is controlled by a hyper parameter called dropout rate, which is the possibility of neuron's being dropped out.

Recent studies have demonstrated that dropout can assist with underfitting as well as controlling overfitting.

3.5.3.5 Batch Normalization:

In the very deep neural networks, the gradients of the weights in the deep layers can be very large or very low. This can lead to problems as the very large gradient would result in very large amount of weight update, while very low gradient would hardly change the weights. This problem slows down and destabilizes the model training process, resulting in the failure of the model's learning. This problem is called vanishing or exploding gradient problem.

Batch Normalization[144] is done in order to address the problem. In Batch Normalization, the input to each layer is normalized and the output is shifted and scaled. The shifting and scaling are controlled by two hyper parameters.

The Batch Normalization algorithm consists of the following equations:

$$\mu_B = \frac{1}{m_B} \sum_{i=1}^{m_B} x^{(i)} \quad (3.28)$$

$$\sigma_B^2 = \frac{1}{m_B} \sum_{i=1}^{m_B} (x^{(i)} - \mu_B)^2 \quad (3.29)$$

$$\hat{x}^{(i)} = \frac{x^{(i)} - \mu_B}{\sqrt{\sigma_B^2 + \epsilon}} \quad (3.30)$$

$$z^{(i)} = \gamma \hat{x}^{(i)} + \beta \quad (3.31)$$

Where,

- μ_B is the input mean, evaluated over the whole mini-batch B
- σ_B is the input standard deviations, also evaluated over the whole minibatch B
- m_B is the number of samples in minibatch B
- $\hat{x}^{(i)}$ is the vector of zero-centered and normalized inputs for instance i .
- γ is the output scale parameter vector for the
- β is the output shift (offset) parameter vector for the layer
- ϵ is a tiny number that avoids division by zero (typically 10^{-5}). This is called a *smoothing term*.
- $z^{(i)}$ is the output of the BN operation. It is a rescaled and shifted version of the inputs.

Apart from addressing the exploding and vanishing gradient problem, the Batch Normalization method provides some unique advantages. It speeds up the training and help with convergence, it also mitigates the covariance shift of the activation. Batch Normalization imparts slight regularization effect as well.

3.5.4 Self-Normalizing Neural Network:

Self-Normalizing Neural Networks[145] are a special kind of Neural Networks that uses the Scaled Exponential Linear Unit (SeLU) activation function.

The SELU activation function is given by

$$\begin{aligned} f(x) &= \lambda x \text{ if } x \geq 0 \\ f(x) &= \lambda \alpha (\exp(x) - 1) \text{ if } x < 0 \end{aligned} \quad (3.32)$$

with $\alpha \approx 1.6733$ and $\lambda \approx 1.0507$.

The use of SeLU activation function which is depicted in Figure 3-12, eliminates the need of using normalization techniques as the weights naturally normalizes themselves. These models outperform similar models trained with other activation functions and allows for training of very deep neural networks. The SNN model weights need to be initialized with the LeCun Normal initialization technique. Furthermore, for regularization, regular dropout is not used, as it will break normalization. A special kind of dropout named Alpha Dropout is used. The normalization of weights results in stable activation during train which helps with convergence as well as improving the training speed. SNN models are impervious to the vanishing or exploding gradient problem, and provides better generalization.

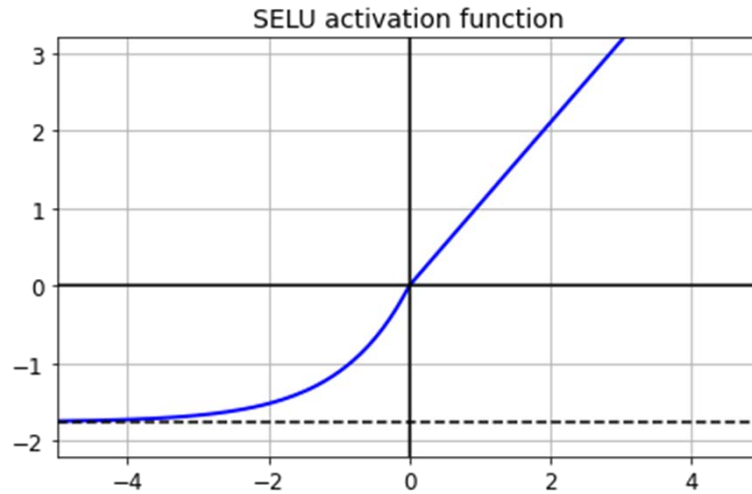


Figure 3-12: SELU activation function

3.6 Modelling Procedure

In this study we trained 2 kinds of machine learning models: tree-based models and deep learning models. XGBOOST and LIGHTGBM are tree-based model while ANN and SNN are deep learning models.

The modelling of all four models followed identical process. At first hyper parameter tuning was done using the Optuna [146] framework. A model was then trained on the optimized set of hyperparameter. The optimized model was then used for prediction and further evaluation on test data and data from buoys not included in the training set.

A hyperparameter is a parameter that is set prior to the learning phase of a machine learning algorithm. Hyperparameters serve as configuration settings for the learning algorithm, governing various aspects including model complexity, learning rate, regularization strength, number of iterations or epochs, and choice of optimization algorithm. They are typically determined through prior knowledge, domain expertise, or trial and error.

The selection of appropriate hyperparameters is crucial, as they can have a substantial effect on the model's ability to learn and generalize from the training data. Different hyperparameter

values can result in varying model performance, accuracy, convergence speed, and ability to prevent overfitting and underfitting.

Hyperparameter optimization, also known as hyperparameter tuning, is the process of determining the optimal hyperparameter values for a machine learning model. It seeks to search for and select hyperparameters that optimize the model's performance and generalizability.

Hyperparameter optimization is crucial because the efficacy of a machine learning model is highly dependent on the values of its hyperparameters. By selecting suitable hyperparameter values, the model's precision, convergence speed, and ability to generalize well to unobserved data can be enhanced.

3.6.1 XGBoost

For hyper parameter optimization, 50 XGBoost models were trained for 1000 iterations each. Then from the search space, the optimal set of hyper parameters was selected and was used for model training. The search space for XGBoost hyper parameters is listed in Table 3-6

Following parameters were optimized in hyper parameter optimization process:

- 1) Learning rate
- 2) Lambda
- 3) Alpha
- 4) Gamma
- 5) Growth Policy
- 6) Subsample
- 7) Col sample
- 8) Maximum Depth

9) Minimum Child Weight

Table 3-6: Hyperparameter optimization search space for XGBoost

Parameter	Search Space	Optimal Value
Learning rate	1^{-3} to 1^{-1}	0.095
Lambda	1^{-8} to 1	2.46^{-7}
Alpha	1^{-8} to 1	0.47
Gamma	1^{-8} to 1	$1,25^{-5}$
Growth Policy	Depthwise, Lossguide	Lossguide
Sub Sample	0.6 to 0.1	0.97
Col Sample	0.6 to 0.1	0.66
Maximum Depth	10 to 12	12
Minimum Child Weight	2 to 50	25

The XGBoost model on optimal set of hyper parameters were trained for 72,787 iterations before converging, early stopping was used to prevent overfitting. The trained model was then subjected to evaluation.

3.6.2 LightGBM

LightGBM models are more sensitive to hyper parameters compared with XGBoost. 30 LightGBM models were trained for 100 iterations as part of hyper parameter tuning using optuna and the search space for the hyperparameter tuning is listed in Table 3-7.

The following parameters were considered for hyper parameter tuning:

- 1) Learning Rate

- 2) Lambda L1
- 3) Lambda L2
- 4) Number of Leaves
- 5) Feature Fractions
- 6) Bagging Fractions
- 7) Bagging Frequency

Table 3-7: Hyperparameter optimization search space for LightGBM

Parameter	Search Space	Optimal Value
Learning Rate	1^{-3} to 1^{-1}	0.065
Lambda L1	1^{-8} to 10	9.849
Lambda L2	1^{-8} to 10	0.091
Number of Leaves	1000 to 50000	44000
Feature Fractions	0.40 to 1.0	0.509
Bagging Fractions	0.40 to 1.0	0.862
Bagging Frequency	1 to 7	2

A lightGBM model was trained on the optimal sets of parameters. The model was trained for 100 iterations.

3.6.3 Artificial Neural Network (ANN)

For ANN model, the whole model was split into blocks. Each block contained 3 fully connected layer followed by a Batch Normalization and Dropout layer. The neural network was uniform in shape: the number of neurons were uniform throughout the layers.

Following parameters were subjected to hyper parameter optimization:

- 1) Number of Blocks
- 2) Number of Neurons
- 3) Learning Rate
- 4) Dropout Rate

50 models were created and trained for 50 iterations each, then the best performing set of parameters was selected. The search space and selected values for this model are given in Table 3-8.

Table 3-8: Hyperparameter optimization search space for ANN

Parameter	Search Space	Optimal Values
Number of Blocks	1 to 7	3
Number of Neurons	64 to 512	512
Learning Rate	1^{-5} to 1^{-3}	0.000957
Dropout Rate	0.05 to 0.25	0.05

The model trained on the optimal parameter converged after 167 iterations.

3.6.4 Self-Normalizing Neural Networks (SNN)

The SNN model had a triangular shape, this too was split into blocks. Each blocks contained 2 fully connected layer followed by an Alpha Dropout layer. The number of neurons were 1024 for the widest layer and 32 for the last layer. The number of neurons in a block was half of that of previous block.

The following parameters were used for hyper parameter tuning:

1) Dropout Rate

2) Learning Rate

For hyper parameter tuning 25 models were trained for 30 iterations each and the search space for dropout rate and learning rate is given in Table 3-9.

Table 3-9: Hyperparameter optimization search space for SNN

Parameter	Search Space	Optimal Values
Dropout Rate	0.005 to 0.15	005
Learning Rate	1^{-5} to 1^{-3}	0.00075

The optimal model converged in only 150 iterations.

3.7 Model Performance and Validation Criteria

3.7.1 Evaluation Metrics

3.7.1.1 R-square (R^2)

R-square indicates the degree to which the data match the regression model. The upper limit for the R-square is 1 which indicates both the dependent & independent variables are perfectly correlated (perfect fit), however it is not bounded by any lower limit, 0 often indicates a trivial fit & negative value indicate fit below average level. R-square is scale-free value, therefore whether the values are tiny or high, the R square value will be lower than 1. R-square equation is:

$$R^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y}_i)^2} \quad (3.33)$$

Where

y_i = measured value, \hat{y}_i = predicted value, \bar{y}_i = average value

3.7.1.2 Mean Squared Value (MSE)

In ML applications, it is a common error measure to calculate the mean squared deviation between the actual and estimated values. The average squared residual is represented by the mean squared error in regression models.

$$\text{MSE} = \frac{\sum (y_i - \hat{y}_i)^2}{n} \quad (3.34)$$

Where n= number of observed values

3.7.1.3 Mean Absolute Error (MAE)

The mean absolute error is calculated by averaging all the absolute errors from each prediction and comparing them to the real value and can be estimated by the following equation.

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (3.35)$$

Chapter 4 Results and Discussion

In this section the visual and numerical evaluation of the four models were done as well as a comparative analysis of the four algorithm's performance at predicting the wave heights of varying range.

The numerical evaluation was conducted using the validation data, testing data, as well as data gathered from buoys that were not part of the training set and gathered after the training period. As part of exploratory analysis, a wide range of graphs have been employed, including bar plots of actual and predicted values, to compare the output of multiple models with the actual ones and determine, among other factors, if they follow the general trend.

A violin plot was generated to examine the distribution and density of the residual error across various models. The residual error was computed as,

$$\text{error} = \|y_{\text{actual}} - y_{\text{prediction}}\| \quad (4.1)$$

Moreover, a scatter plot was generated for each of the four models in order to assess the quality of the output model's fit. The predicted values of models are depicted against the actual wave height in a scatter plot.

A line plot was utilized to visually assess the accuracy of the fit and the extent of deviation between the predicted and actual values, as both sets of data were plotted on the same axis.

Ultimately, a bar chart was generated to represent the error metrics and facilitate a comparison of the error magnitudes across the four models.

4.1 XGBoost

The XGBoost models provide a satisfactory approximation to the actual values, with the predicted output not deviating significantly from the actual values. The performance of the model based on three evaluation metrics is listed in Table 4-1 which exhibits that the errors of

the test set were nearly identical to those of the valid set, indicating no overfitting. The Mean Absolute Error (MAE) of the model is comparatively lower than the Mean Squared Error (MSE), implying that a significant proportion of the error values are close to zero. The R^2 value on the validation set and Test set is 0.896 and 0.894 respectively which implies a satisfactory fit of the trained model to the actual value. The predicted value also followed the mean of actual data closely as listed in Table 4-2, while deviating away from standard deviation.

Table 4-1: Model evaluation of XGBoost on test and validation data

	MSE	MAE	R^2
Valid Set	0.227	0.106	0.896
Test Set	0.226	0.106	0.894

Table 4-2: Comparison of means and standard deviation of predicted value with actual value

	Actual Mean	Predicted Mean	Actual Standard Deviation	Predicted Standard Deviation
Valid Set	1.441	1.442	1.01	0.944
Test Set	1.434	1.436	1.003	0.937

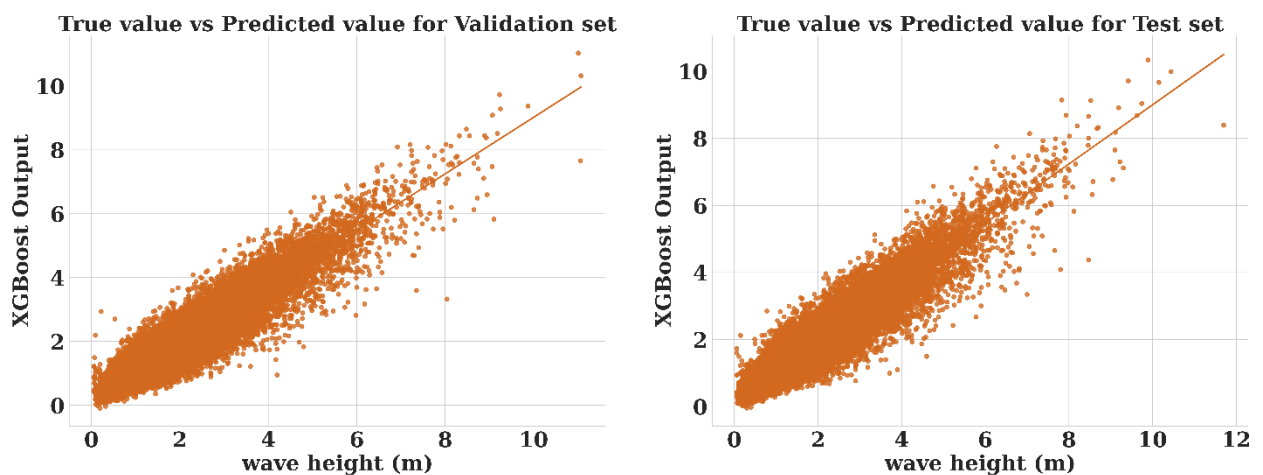


Figure 4-1: Scatter plot of wave height against XGBoost prediction on test and validation data

The efficacy of the XGBoost model is evident in the scatterplot illustrated in Figure 4-1 which exhibits minimal dispersion around the mean, with only a handful of outliers located at the

extremes of the distribution. The model exhibits satisfactory accuracy when evaluated against the test and validation data.

The model was also evaluated on the data: unknown and known buoys and the performances have been enlisted in Table 4-3 and in Table 4-4 respectively.

Compared to known buoys, the model performs marginally worse when predicting unknown buoys; however, the magnitude of the difference is significantly smaller than the performance decline when compared to test and validation data. The performance was quite satisfactory on unknown buoys except Station 46035 and 46066 considering the humongous area this model is covering, also there might be some localized factors that deteriorate the performance of the model as these haven't been considered while training the model. For Station 42057, best performance is observed (6.7% MSE and 20% MAE), while the MSE is within 14 % and MAE is within 30% for other stations except the first two mentioned earlier which is quite impressive. Therefore, it can be concluded that although the model overfits the training data marginally, it shows satisfactory performance on most of the unknown stations.

For the known buoys, the performance is much better than the unknown buoys as dataset from these stations were utilized while training the model. The model is fitted quite satisfactorily as all of the known stations have R^2 value close to 0.80. Both the MAE and MSE is impressive for all of the known stations except station 46082 which gives 47.5 % MAE and 38 % MSE. The errors might be emanated from some local factors for this station which were not considered during the training process, also the model might be slightly over fitted to the remaining known stations.

Table 4-3: Model evaluation of XGBoost on unknown buoys

Buoy number	XGBOOST MAE	XGBOOST MSE	XGBOOST R²
46035	0.464086	0.420102	0.836135
46066	0.663947	0.82167	0.591437
42058	0.296003	0.136799	0.737573
42057	0.201888	0.067683	0.741534
42001	0.23583	0.094794	0.781998
44025	0.274198	0.134604	0.74491

Table 4-4: Model evaluation of XGBoost on known buoys

Buoy number	XGBOOST MAE	XGBOOST MSE	XGBOOST R²
41013	0.212389	0.074468	0.812042
42002	0.247476	0.112031	0.810281
42040	0.19589	0.078396	0.806552
46080	0.32323	0.181435	0.855107
46082	0.47586	0.380273	0.776286
44034	0.267377	0.126234	0.754938

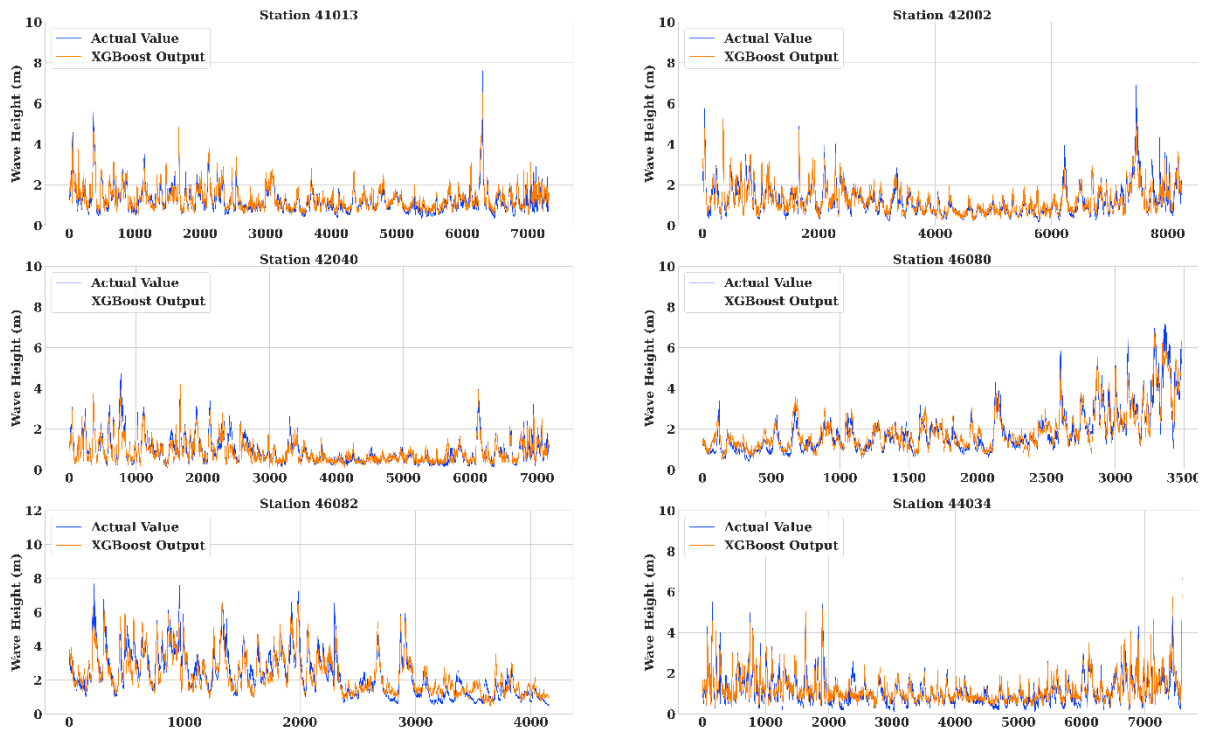


Figure 4-2: Line plot of actual values and XGBoost prediction on known buoys

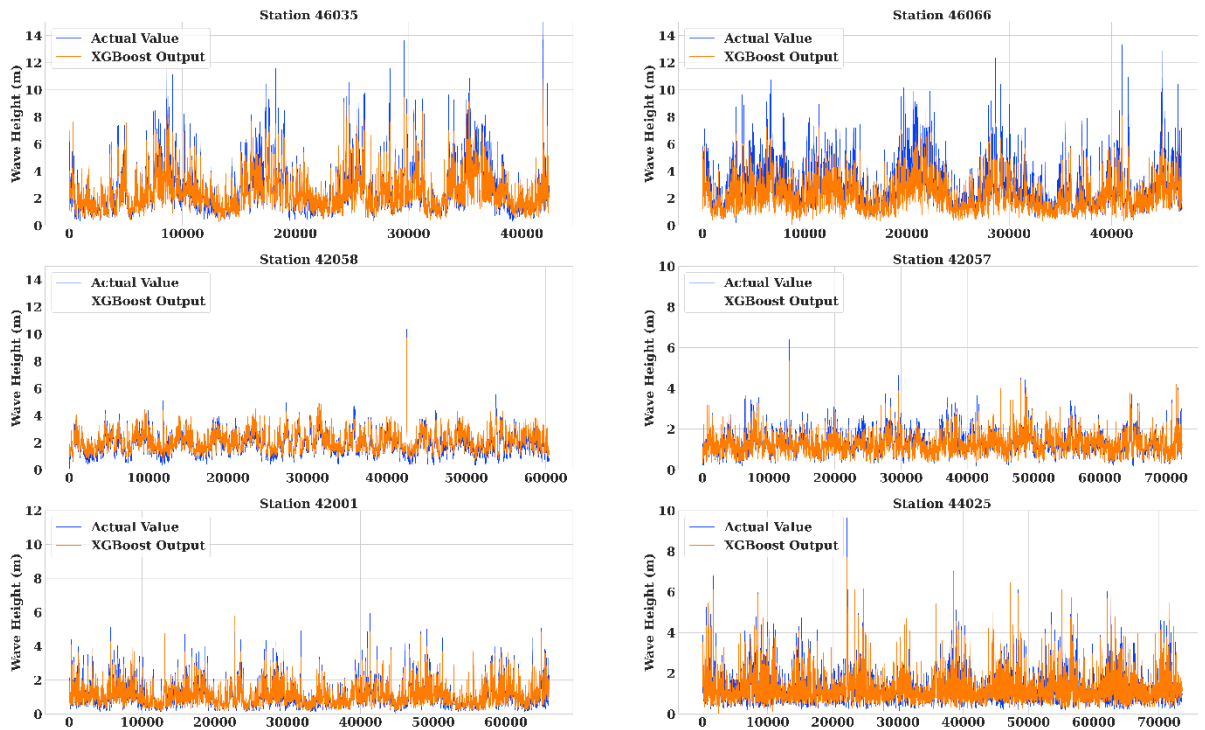


Figure 4-3: Line plot of actual values and XGBoost prediction on unknown buoys

The line graphs for the known buoys in Figure 4-2 and for the unknown buoys in Figure 4-3 are illustrating actual data and predicted output on the same axis. These plots indicate that the output of the models closely follows the actual data and their trends for all of the stations for

the known buoys. However, it is incapable of adapting to spikes, extreme changes, and outlier values. For the unknown buoys, although the performance is comparable with the known buoys for the station 42058, 42057, 42001, 44025, but for station 46035, 46066, XGBoost couldn't follow the trend of the actual wave height properly. However, it should be acceptable for a model that is covering such a huge area and many of the localized factors were not considered while training the model as well for the simplification of the modelling.

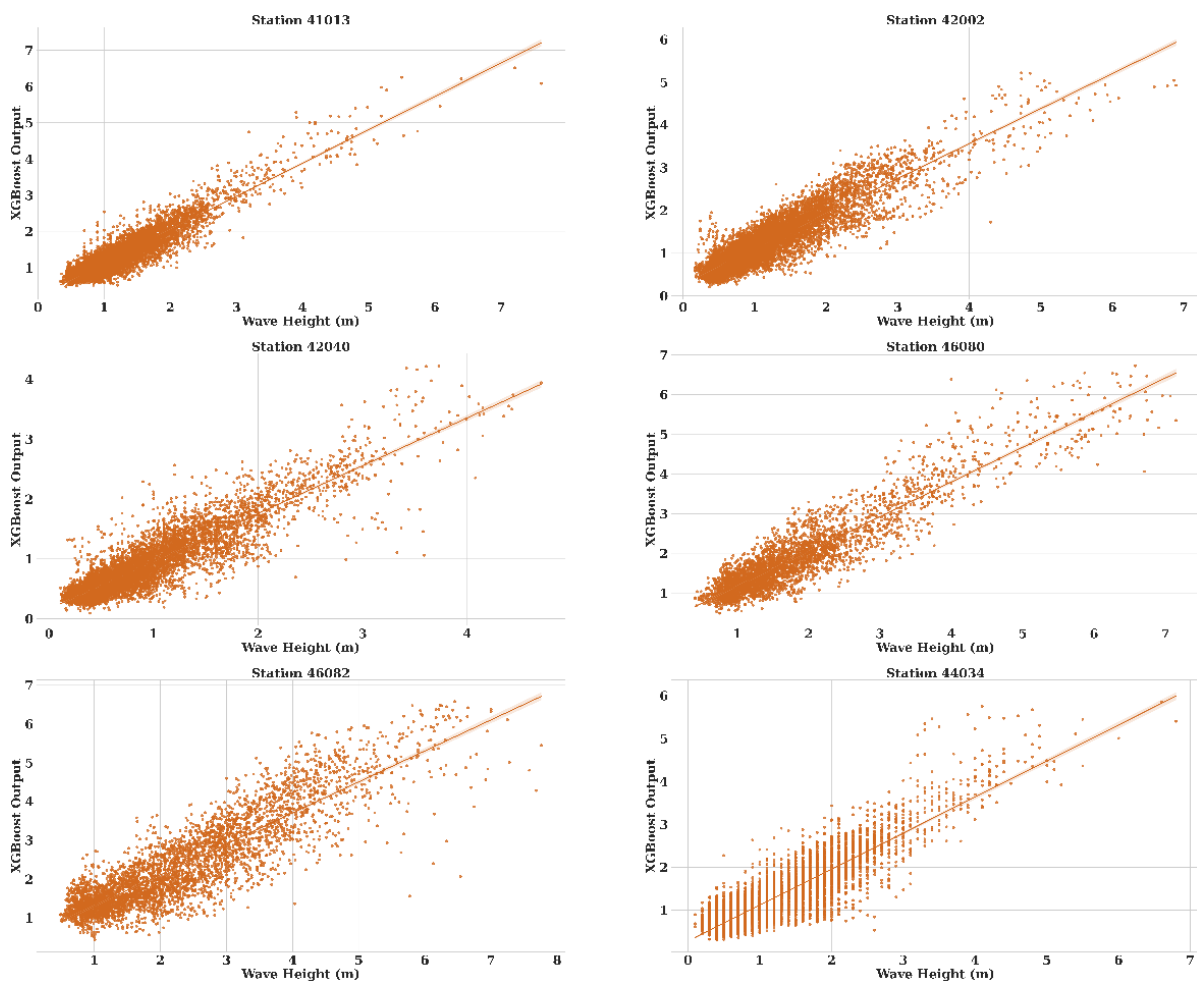


Figure 4-4: Scatter plot of Wave Height against XGBoost prediction on known buoys

For both known and unknown buoy data, the scatter plot in Figure 4-4 and Figure 4-5 respectively display a greater degree of dispersion than that of the training buoys. Furthermore, the error increased for all indicators as the wave height increased.

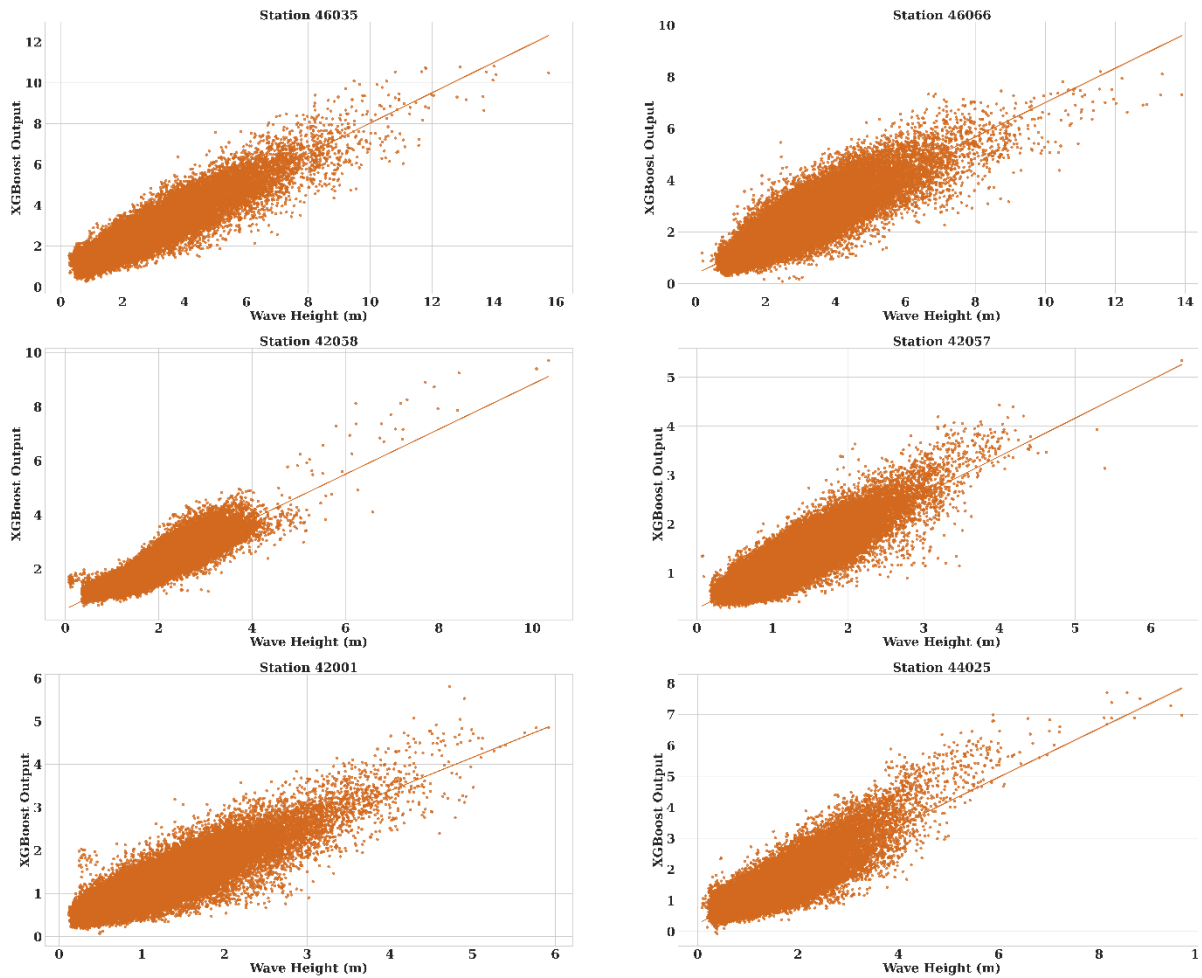


Figure 4-5: Scatter plot of Wave Height against XGBoost prediction on unknown buoys

4.2 LightGBM

In addition, the LightGBM model performed similarly on both the valid and training sets. The LightGBM model provided an outstanding fit without overfitting, with a regression coefficient greater than 0.89 on both the testing and validation set as enlisted in Table 4-5. The mean squared error was only 10.8% for both the Test Set and Validation Set and significantly less than the mean absolute error, indicating that the majority of the error lies close to zero.

Table 4-5: Model evaluation of LightGBM on test and validation data

	MSE	MAE	R ²
Valid Set	0.108	0.225	0.893
Test Set	0.108	0.224	0.892

Table 4-6: Comparison of means and standard deviation of LightGBM predicted value with actual value

	Actual Mean	Predicted Mean	Actual Standard Deviation	Predicted Standard Deviation
Valid Set	1.441	1.442	1.010	0.922
Test Set	1.434	1.436	1.003	0.913

The Mean and Standard deviation of predicted values and actual values are listed in Table 4-6 which clearly implies that LightGBM model's mean closely tracked the actual mean, slightly exceeding it in some portions. However, the discrepancy between the actual and predicted values demonstrate that the model failed to account for the data's variability and dispersion. The model provided a decent fit overall.

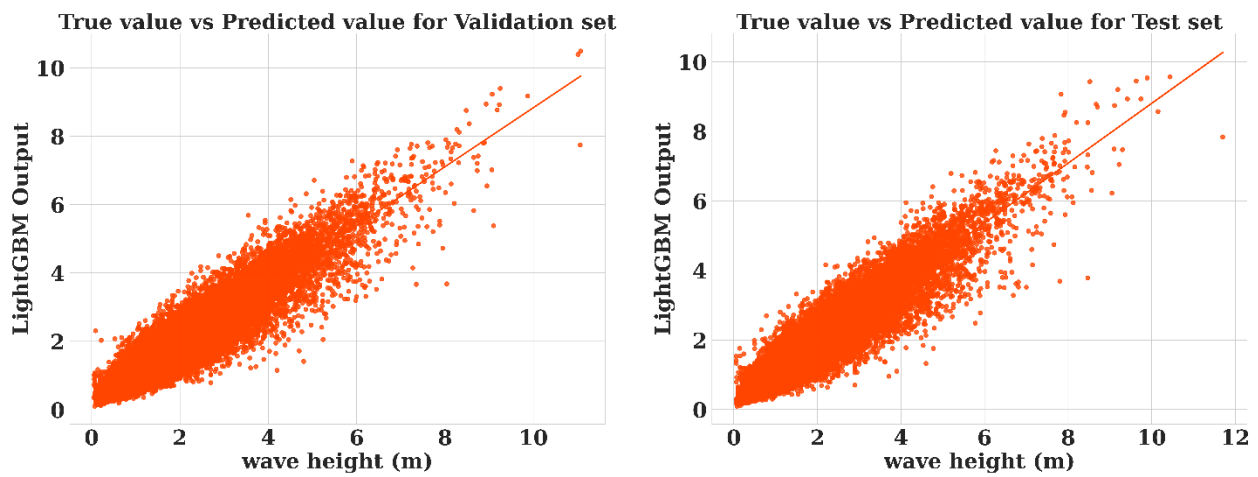


Figure 4-6: Scatter plot of Wave Height against LightGBM prediction on test and validation data

The scatterplot in Figure 4-6 is depicting minimum dispersion around the mean when wave height is less and for wave height greater than 8 m, performance of the model is not that good. So, it is evident that, the performance of the model is impressive for wave height less than 8 m on both Test and Validation dataset, but the model can't perform well on the outliers or extreme peak values of the wave height.

Table 4-7: Model evaluation of LightGBM on unknown buoys

Buoy number	LGBM MAE	LGBM MSE	LGBM R²
46035	0.486137	0.452828	0.82337
46066	0.527157	0.550724	0.726161
42058	0.29417	0.133242	0.744396
42057	0.205622	0.072871	0.721722
42001	0.235217	0.096707	0.777597
44025	0.250827	0.112994	0.785863

Table 4-8: Model evaluation of LightGBM on known buoys

Buoy number	LGBM MAE	LGBM MSE	LGBM R²
41013	0.209483	0.071336	0.819946
42002	0.243407	0.11211	0.810146
42040	0.19366	0.07635	0.811602
46080	0.328074	0.186386	0.851154
46082	0.509492	0.435545	0.74377
44034	0.263408	0.121411	0.764302

The LightGBM model was further evaluated on known and unknown buoy data and the results are enlisted in Table 4-7 and Table 4-8. The model exhibited comparable performance on both familiar and unfamiliar buoys, although with a marginal decline in performance between the validation and test sets and the known unknown buoys. This drop in performance may be an indication of overfitting. To evaluate the performance of LightGBM on known and unknown stations, line plots have been depicted in Figure 4-7 and Figure 4-8. The output of the model exhibits a high degree of conformity with the means and trends; however, it falls short in accurately representing the occurrence of spikes or extreme values. The model output appears to align with the actual values, even for the outliers. For the known stations, the performance of the model is satisfactory as it replicates the trend both on mean values and spikes quite correctly, however for the unknown stations, the performance on station 46035 and 46066 could have been improved for the sudden spike values. It should be still acceptable considering the huge range that is being covered with a single generalized model.

For further evaluation of the model, scatter plot on known and unknown buoys are illustrated in Figure 4-9 and Figure 4-10 respectively to measure the extent to which the model fit to the actual value. From the figures, it is evident that The LightGBM model exhibits an increase in error magnitude at the extreme values near the tails and the small values, with the exception of those cases the model demonstrates a good fit. For the known buoys, the performance is quite impressive for all of the stations for wave height below 3m, however for the extreme values of significant wave height, there is a room for improvement both for the known and unknown buoys. There could have been several reasons behind these errors on unknown stations like overfitting to the known stations while training the model, some localized factors that significantly affect the overall performance of the model, removal of null values which caused some valuable information loss, accuracy of the measurement of actual dataset and also the complexity of the model. As the complexity of the model is avoided to provide dataset with a least computational cost and time, these limitations should be acceptable considering the extent to which this single model operates and provides both high spatial resolution and high time resolution data.

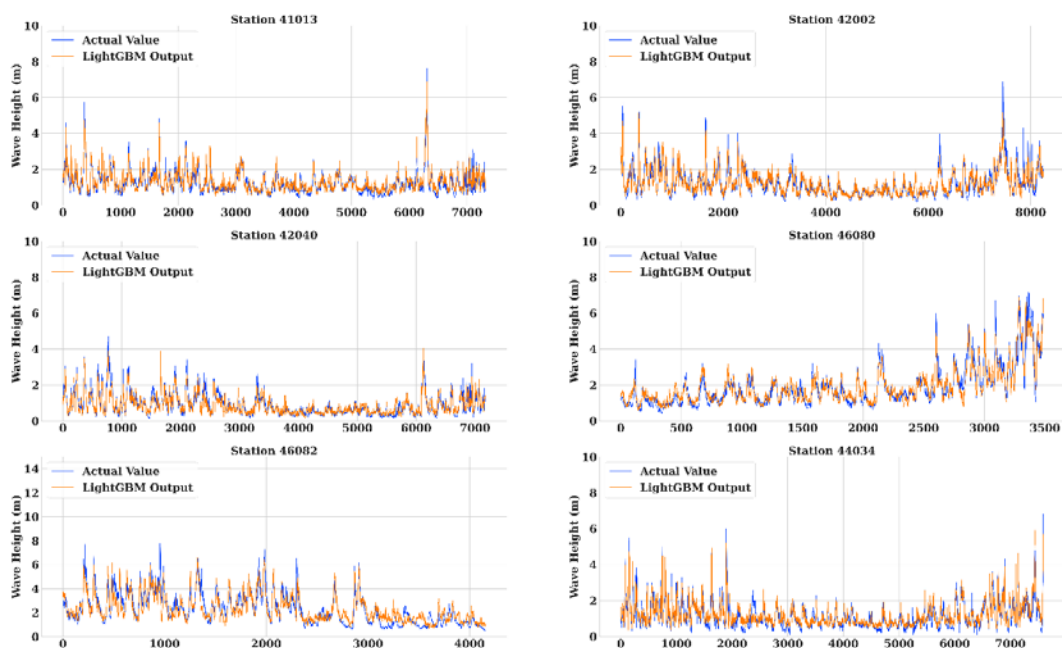


Figure 4-7: Line plot of actual values and LightGBM prediction on known buoys

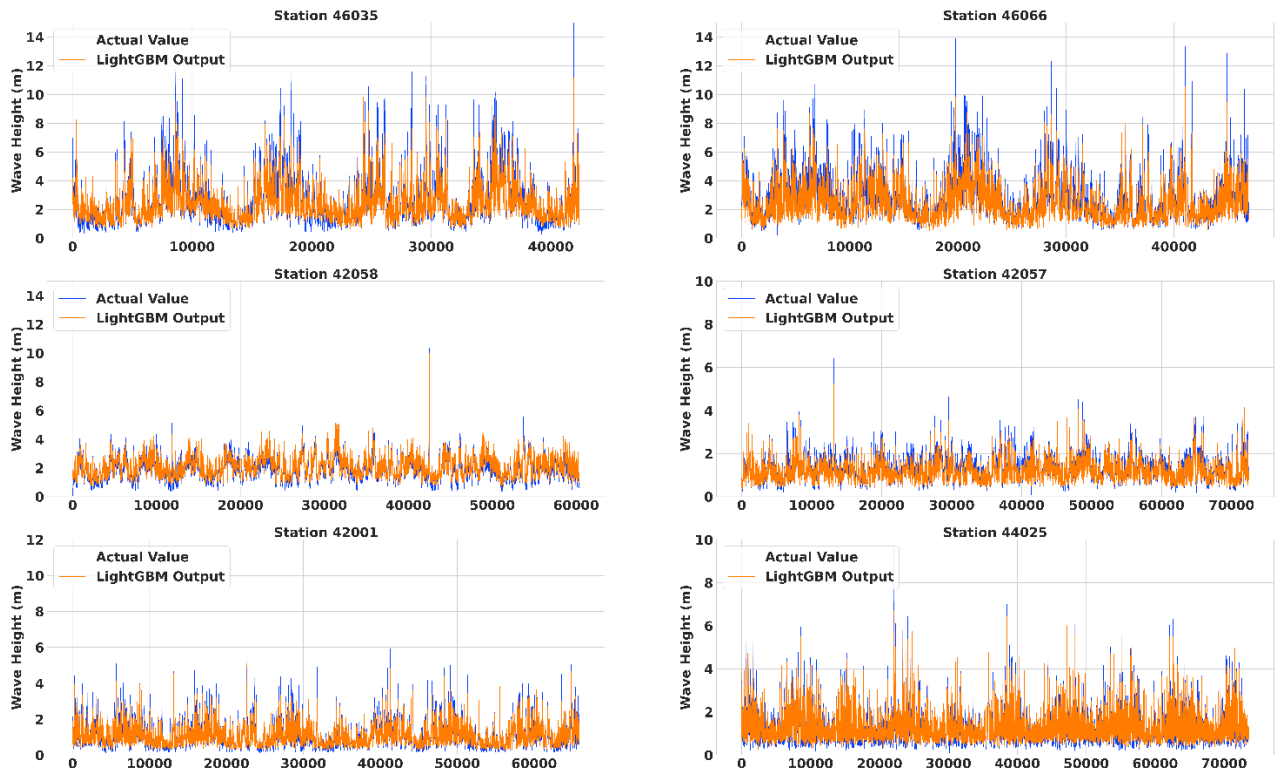


Figure 4-8: Line plot of actual values and LightGBM prediction on unknown buoys

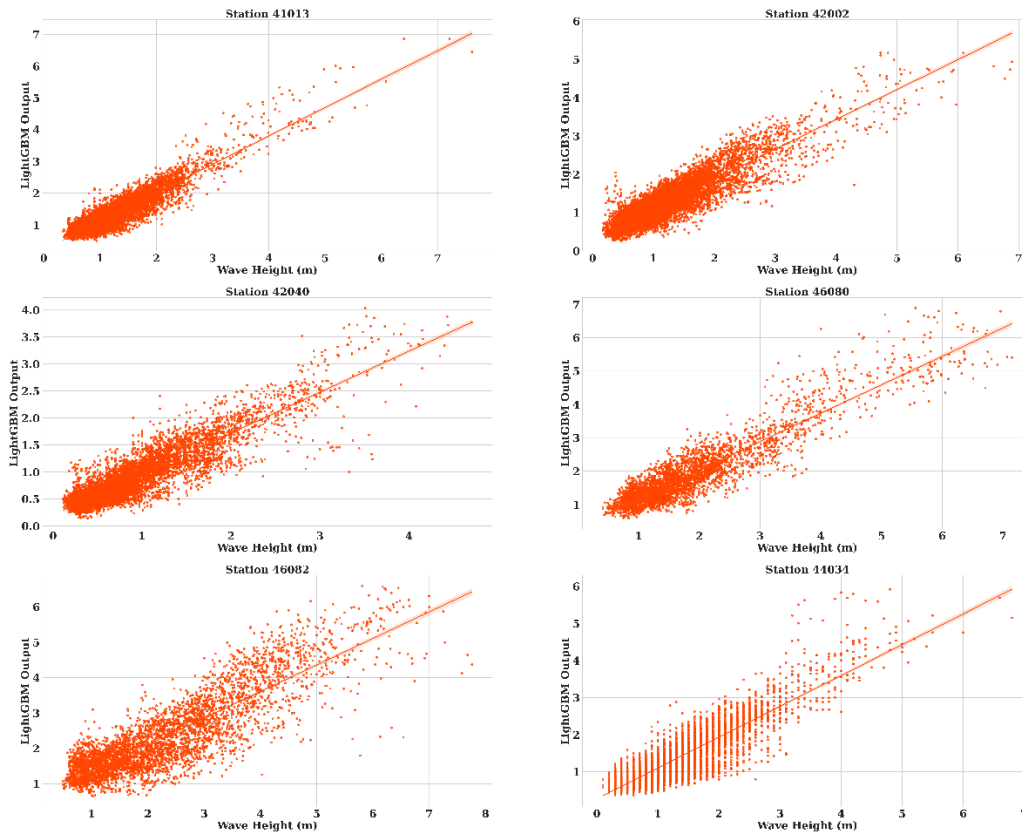


Figure 4-9: Scatter plot of Wave Height against LightGBM prediction on known buoys

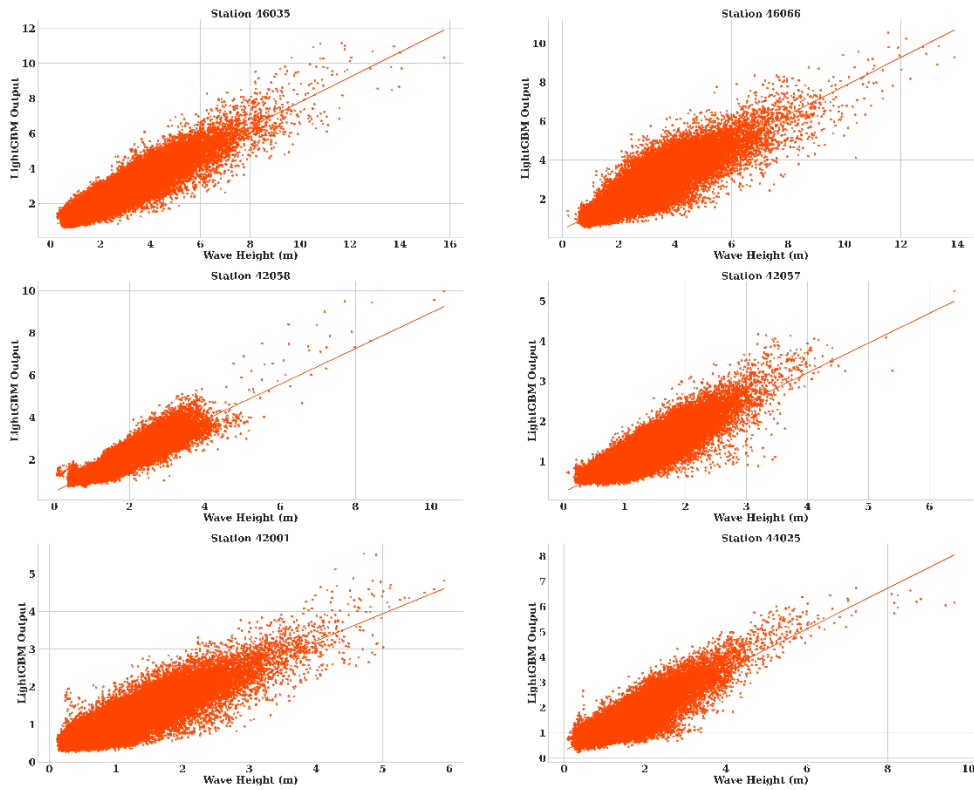


Figure 4-10: Scatter plot of Wave Height against LightGBM prediction on unknown buoys

4.3 Artificial Neural Network (ANN)

The performance of the Artificial Neural Network model is satisfactory as it demonstrates a good fit with both the validation and testing datasets as listed in Table 4-9. The value of the mean squared error is negligible, while the mean absolute error exhibits a relatively low magnitude. The model's coefficient of determination (R-squared) exceeds 0.95, indicating that the model explains over 95% of the variance in the output. The comparison of the actual and anticipated means and deviation in Table 4-10 further demonstrates this. The predicted mean and predicted standard deviation nearly match the actual mean and deviation, respectively.

Table 4-9: Model evaluation of ANN on test and validation data

	MSE	MAE	R²
Valid Set	0.046	0.153	0.954
Test Set	0.047	0.153	0.953

Table 4-10: Comparison of means and standard deviation of ANN predicted value with actual value

	Actual Mean	Predicted Mean	Actual Standard Deviation	Predicted Standard Deviation
Valid Set	1.434	1.438	1.003	0.998
Test Set	1.441	1.445	1.010	1.006

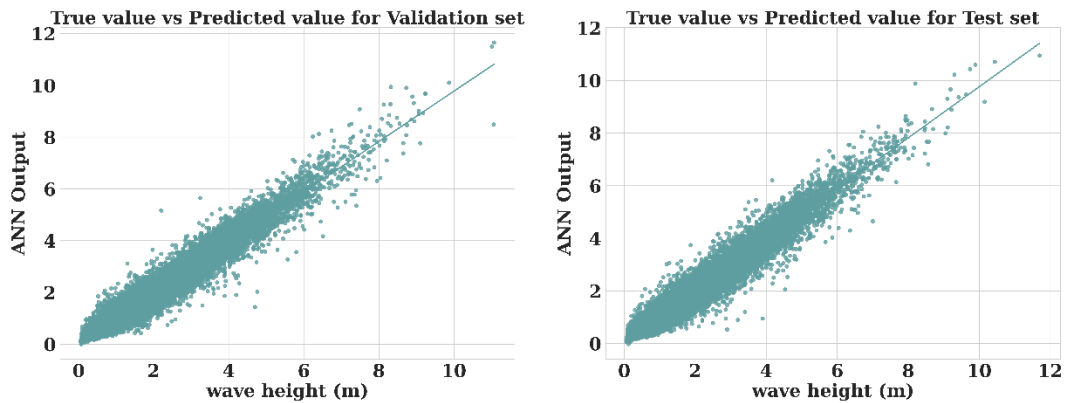


Figure 4-11: Scatter plot of Wave Height against ANN prediction on test and validation data. The scatterplot illustrated in Figure 4-11 exhibits a high degree of consistency between the predicted and actual values, indicating a high degree of accuracy and precision in the model.

Nevertheless, the model exhibits a significant decline in its performance when applied to both unknown and known buoy data which is enlisted in Table 4-11 and Table 4-12 respectively, with the model demonstrating superior performance when applied only to the known buoy data. This observation suggests the presence of an overfitting issue in the model.

Table 4-11: Model evaluation of ANN on unknown buoys

Buoy number	ANN MAE	ANN MSE	ANN R ²
46035	0.540105	0.578674	0.774283
46066	0.549995	0.56601	0.718561
42058	0.311971	0.16119	0.690781
42057	0.231311	0.08992	0.656617
42001	0.245369	0.11166	0.743209
44025	0.246579	0.118135	0.77612

Table 4-12: Model evaluation of ANN on known buoys

Buoy number	ANN MAE	ANN MSE	ANN R²
41013	0.235422	0.096828	0.755602
42002	0.288798	0.157617	0.733082
42040	0.211426	0.099419	0.754677
46080	0.357746	0.233546	0.813492
46082	0.519712	0.507427	0.701482
44034	0.288438	0.154875	0.699337

From the observation of the line plots depicted in Figure 4-12 and Figure 4-13, it is evident that the ANN model conforms to the mean and trend satisfactorily, but it is inadequate in capturing the variability of the observed values, particularly for the unknown buoys. Consequently, the model exhibits a higher degree of inaccuracy in predicting those particular data points both for known and unknown buoys.

The degree of dispersion in predicted values for both known and unknown buoys is observed in Figure 4-14 and Figure 4-15 are considerably higher when compared with the results of the validation and testing data. This implies inferior predictive capabilities, inadequate generalization, and overfitting.

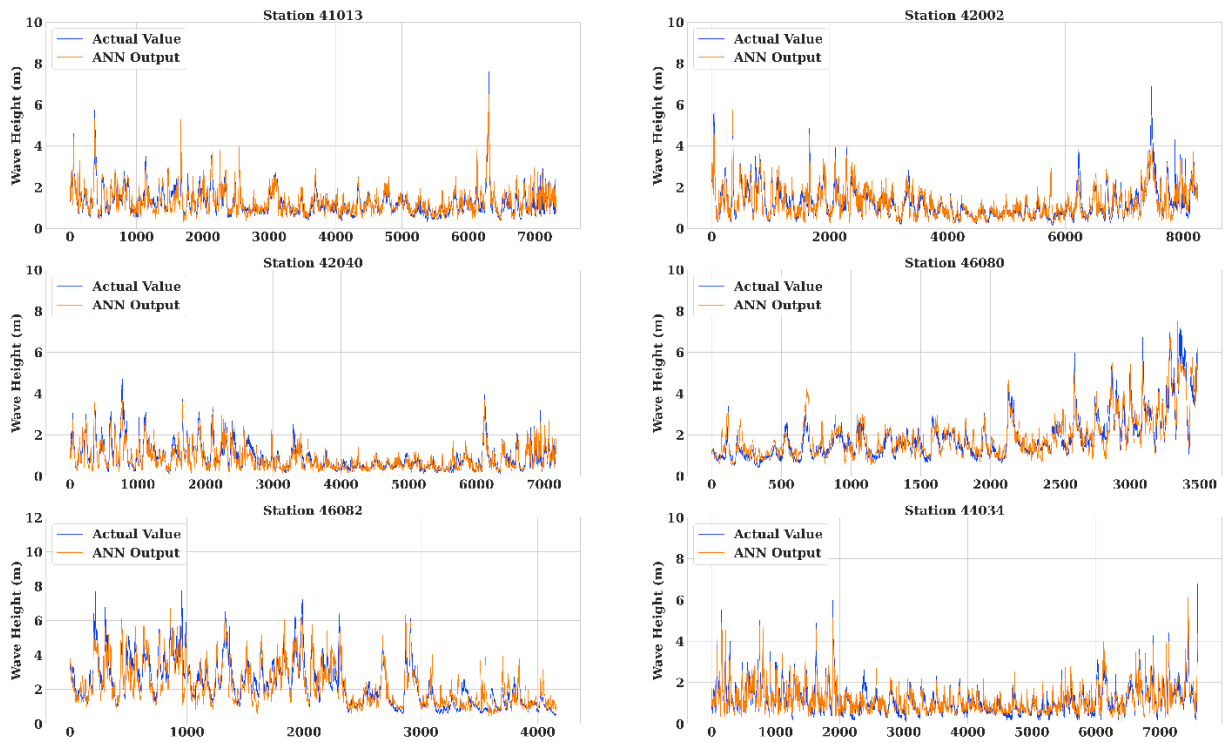


Figure 4-12: Line plot of actual values and ANN prediction on known buoys

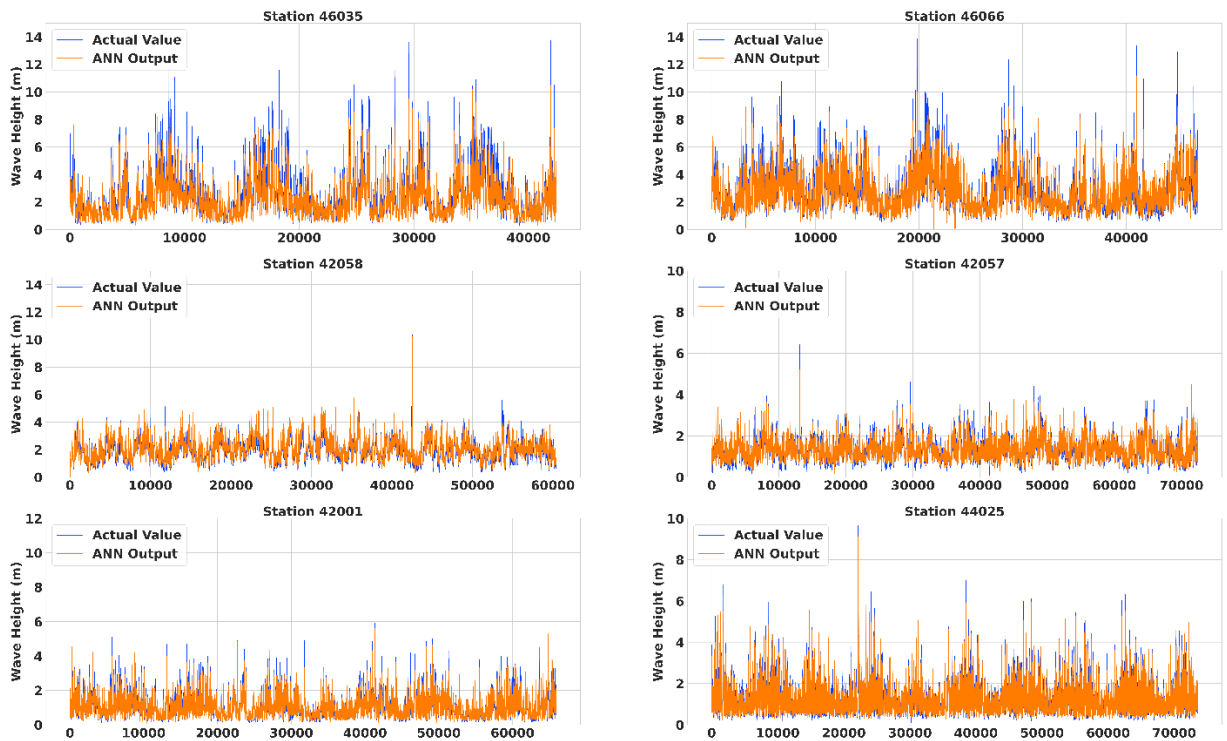


Figure 4-13: Line plot of actual values and ANN prediction on unknown buoys

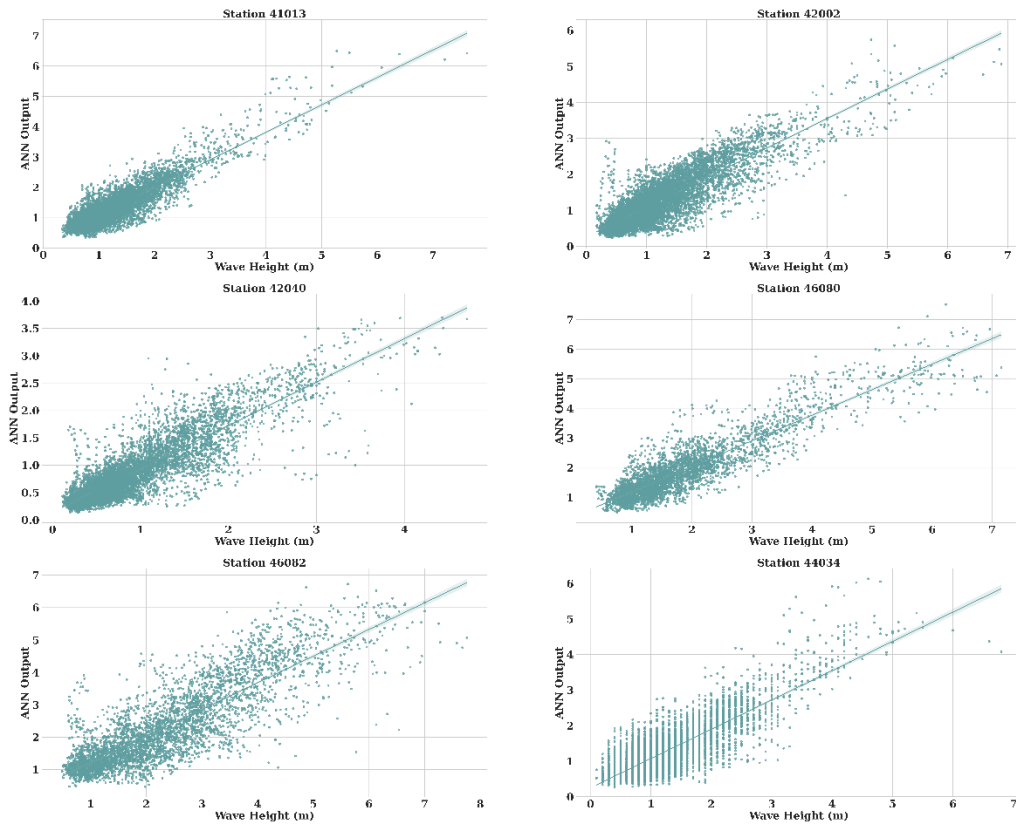


Figure 4-14: Scatter plot of Wave Height against ANN prediction on known buoys

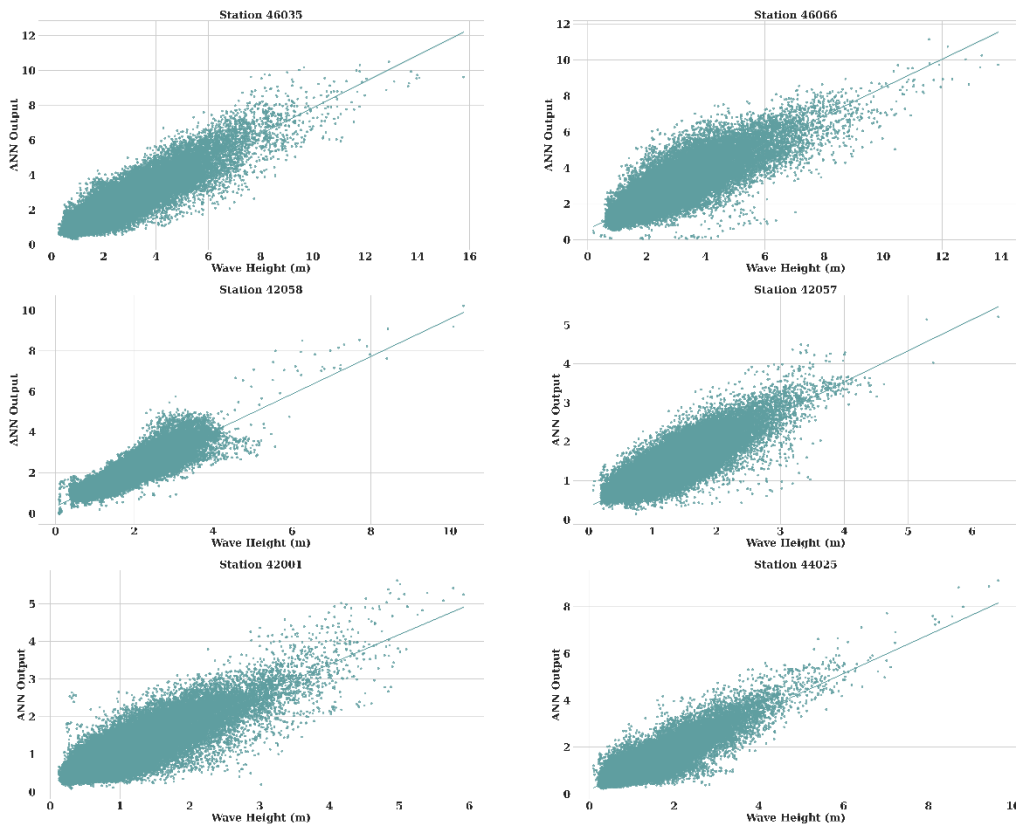


Figure 4-15: Scatter plot of Wave Height against ANN prediction on unknown buoys

4.4 Self-Normalizing Neural Network (SNN)

The Self Normalizing Neural Network exhibits a remarkable level of accuracy on both the testing and validation datasets enlisted in Table 4-13. The regression coefficient exceeds 0.90, indicating that 90% of the variance has been accounted for. Moreover, the relative insignificance of the mean squared error and mean average error should be noted. The mean squared error exhibits a significantly lower value in comparison to the mean absolute error.

Table 4-13: Model evaluation of SNN on test and validation data

	MSE	MAE	R²
Valid Set	0.063	0.182	0.938
Test Set	0.063	0.180	0.937

Table 4-14: Comparison of means and standard deviation of SNN predicted value with actual value

	Actual Mean	Predicted Mean	Actual Standard Deviation	Predicted Standard Deviation
Valid Set	1.441	1.429	1.010	0.981
Test Set	1.434	1.423	1.003	0.937

The statistical analysis enlisted in Table 4-14 indicates that the mean of the predicted value is marginally lower; however, it is extremely similar and closely follows to the actual mean. The model's predicted values exhibit a slightly lower deviation than the actual deviation, with the magnitude of the difference increasing for the test set in comparison to the validation set.

The regression model exhibits a low degree of dispersion between the predicted and actual values, resulting in a narrow and tidy scatter plot for Test and Validation dataset which is depicted in Figure 4-16. Furthermore, the residual error appears to remain fairly constant despite the increase in wave height.

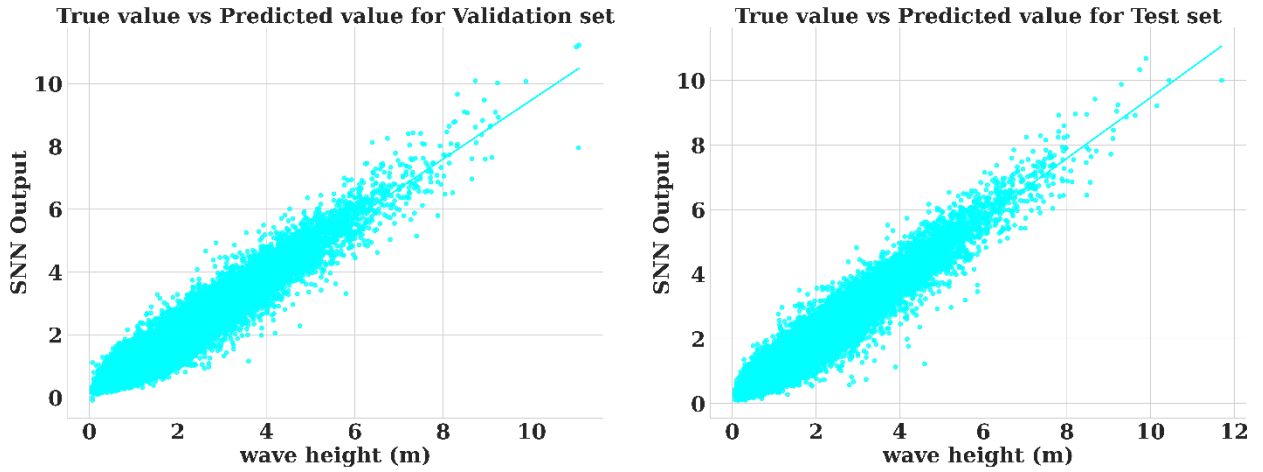


Figure 4-16: Scatter plot of Wave Height against SNN prediction on test and validation data

Table 4-15: Model evaluation of SNN on unknown buoys

Buoy number	SNN MAE	SNN MSE	SNN R ²
46035	0.538552	0.57411	0.776062
46066	0.56178	0.605564	0.698893
42058	0.307767	0.156215	0.700326
42057	0.223573	0.083056	0.682827
42001	0.237079	0.105712	0.756888
44025	0.254127	0.124277	0.764479

Table 4-16: Model evaluation of SNN on known buoys

Buoy number	SNN MAE	SNN MSE	SNN R ²
41013	0.239485	0.107099	0.729678
42002	0.283563	0.165317	0.720043
42040	0.213917	0.112283	0.722933
46080	0.345	0.219518	0.824695
46082	0.546505	0.569457	0.66499
44034	0.287519	0.152976	0.703024

The model exhibits similar performance on both unknown and known buoys as enlisted in Table 4-15 and Table 4-16 respectively. There is a marginal improvement in the prediction performance of known buoys. Nonetheless, the model's efficacy experiences a substantial drop in comparison to that of the testing and validation data.

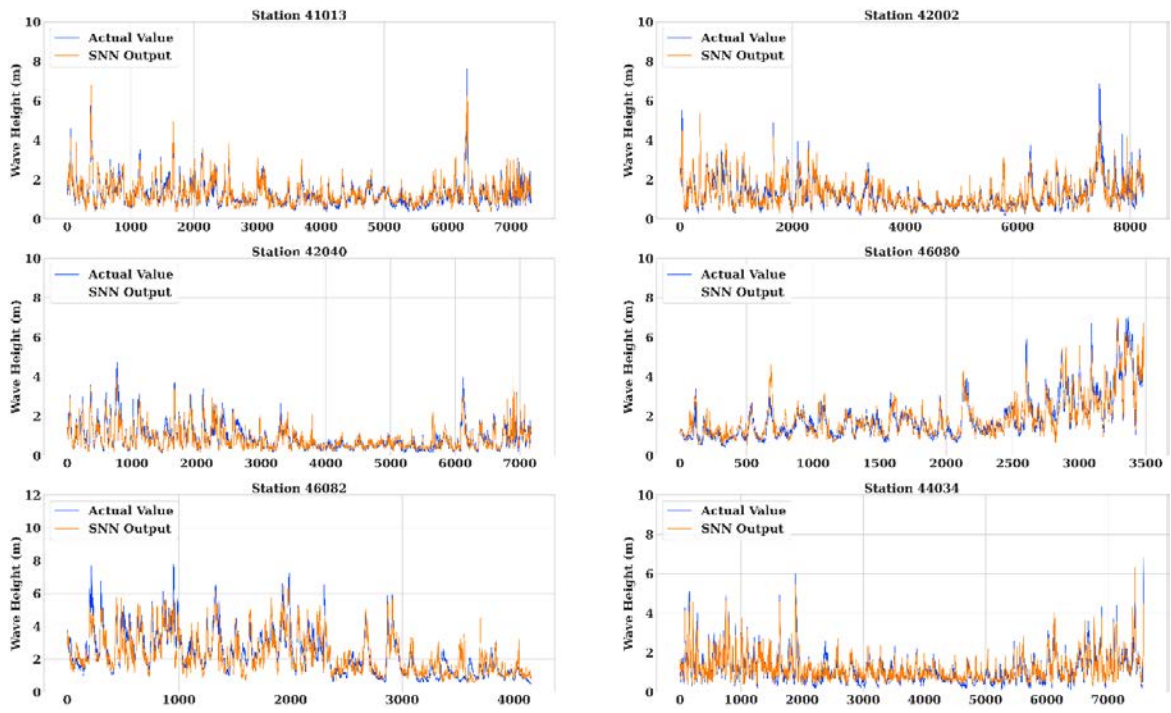


Figure 4-17: Line plot of actual values and SNN prediction on known buoys

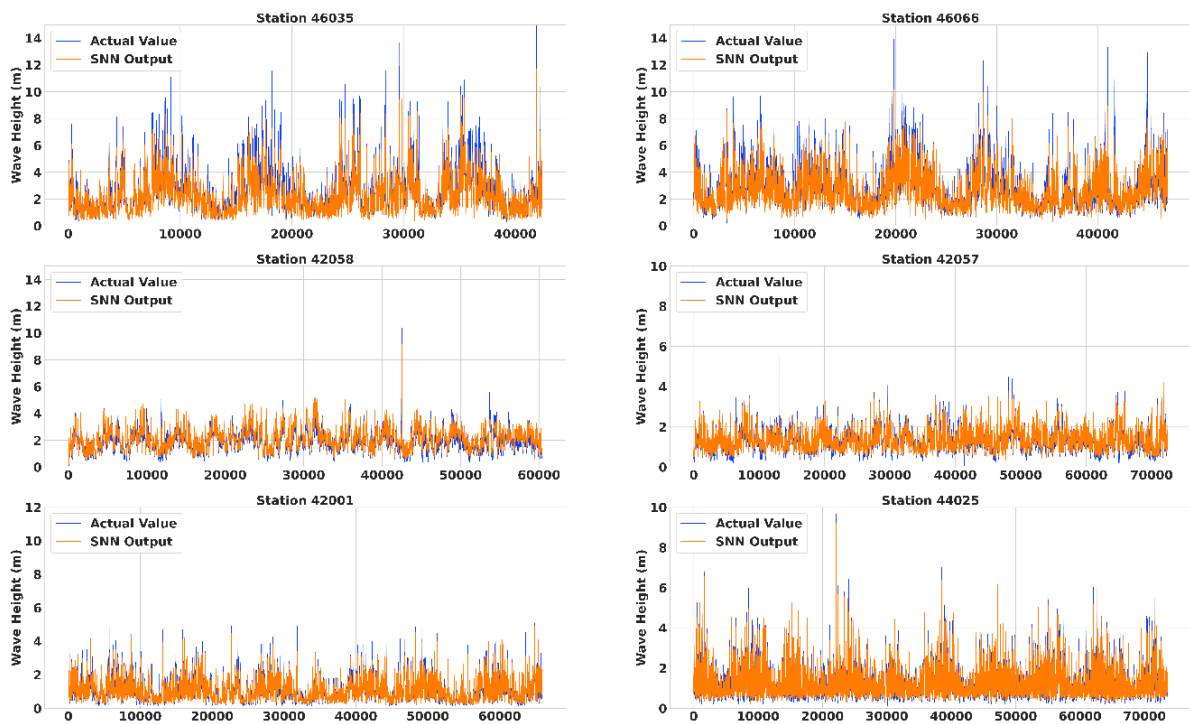


Figure 4-18: Line plot of actual values and SNN prediction on unknown buoys

The line plot presented in Figure 4-17 on known buoys and in Figure 4-18 on unknown buoys exhibits that The SNN model is capable of capturing both the primary trend and the patterns exhibited in the data. But it does not fluctuate or increase like the actual wave height; this issue is more apparent in unknown buoys compared to known buoys. For the six distinct known buoys, the performance of the model is satisfactory, replicating both the mean values and sudden spikes with insignificant error. The performance deteriorates slightly as expected for the unknown buoys particularly for Station 46035 and Station 46066. However, the performance is excellent in the other four unknown buoys both in prediction of mean properties and sudden spikes of wave height.

To evaluate the extent to which the SNN model output is correlated to the actual value, scatter plot for six known buoys and six unknown buoys are illustrated in Figure 4-19 and Figure 4-20. The models' predicted values exhibit a deviation from the actual values at both the lower and upper ends of the distribution, resulting in an increase in error for data points of both lower and higher magnitudes. The dispersion of data points in the scatter plot appears to be significantly larger than that of the validation and testing datasets, suggesting insufficient generalization and overfitting. Although there remains a significant room for improvement for this model, considering the minimum computational requirements while providing a huge operational range, high spatial and high time resolution dataset, the performance of this model should be acceptable for the generalization purpose. The performance can be improved further by increasing the complexity of the model, considering the localized factors that affect the target variable for each of the station, hyper parameter tuning with a large search space.

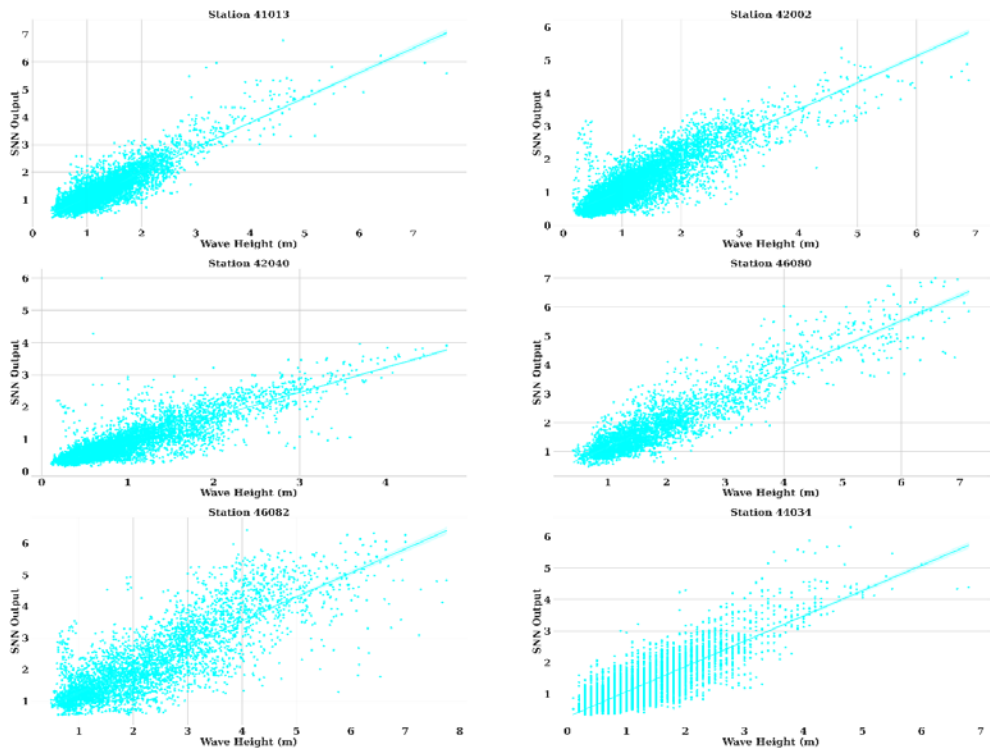


Figure 4-19: Scatter plot of Wave Height against SNN prediction on known buoys

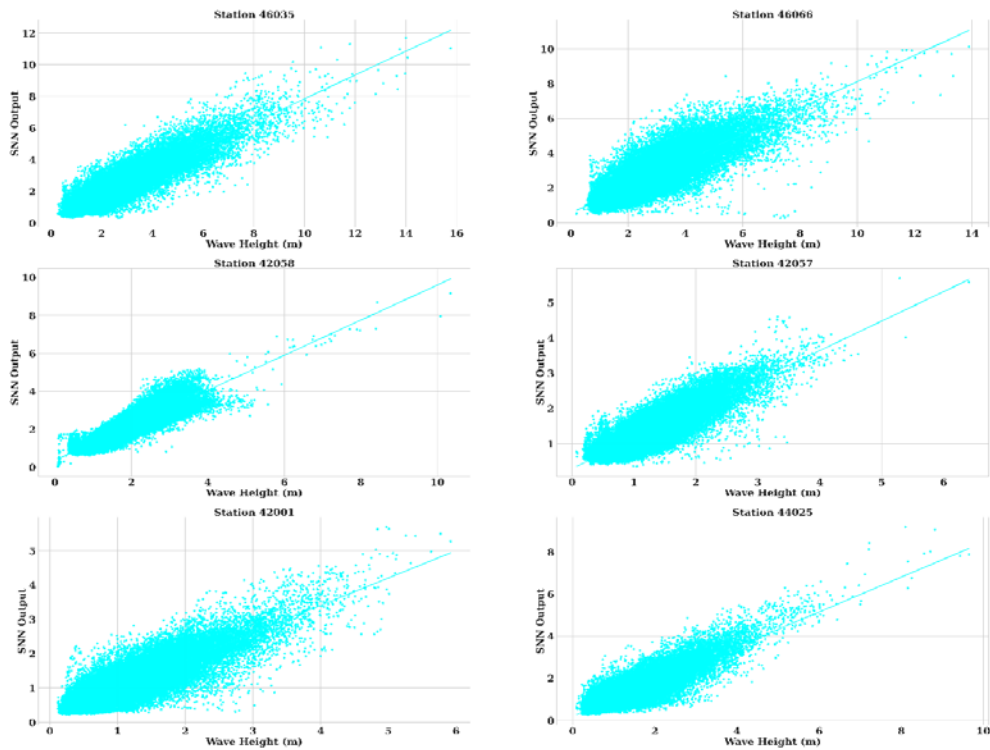


Figure 4-20: Scatter plot of Wave Height against SNN prediction on unknown buoys

4.5 Model Comparison:

This section presents a comparative analysis of the performance of all four models, in regard to both adequateness of fit and generalization. The evaluation of performance was based on both testing, validation datasets and known, unknown buoys which are enlisted in Table 4-17 and Table 4-18 respectively.

It is evident from the illustration in Figure 4-21 for known-unknown buoys and in Figure 4-22 for Test-Validation dataset that the performance of tree models and deep learning models is comparable across all datasets. Although all models perform above an acceptable level of accuracy, the deep learning models do not generalize well on external data (known and unknown buoy data) in comparison to the gradient boosting models. The deep learning models, on the other hand, provide a superior fit on validation and testing data and converge well.

Table 4-17: Error evaluation of all four models on validation and test data

	XGBoost			LightGBM			ANN			SNN		
	MSE	MAE	R ²	MSE	MAE	R ²	MSE	MAE	R ²	MSE	MAE	R ²
Valid Set	0.227	0.106	0.896	0.108	0.225	0.893	0.046	0.153	0.954	0.063	0.182	0.938
Test Set	0.226	0.106	0.894	0.108	0.224	0.892	0.047	0.153	0.953	0.063	0.180	0.937

Table 4-18: Error evaluation of all four models on known and unknown buoy data

	Buoy Number	XGBoost			LightGBM			ANN			SNN		
		MSE	MAE	R ²	MSE	MAE	R ²	MSE	MAE	R ²	MSE	MAE	R ²
Known Buoys	41013	0.07	0.21	0.81	0.07	0.21	0.82	0.10	0.24	0.76	0.11	0.24	0.73
	42002	0.11	0.25	0.81	0.11	0.24	0.81	0.16	0.29	0.73	0.17	0.28	0.72
	42040	0.08	0.20	0.81	0.08	0.19	0.81	0.10	0.21	0.75	0.11	0.21	0.72
	46080	0.18	0.32	0.86	0.19	0.33	0.85	0.23	0.36	0.81	0.22	0.35	0.82
	46082	0.38	0.48	0.78	0.44	0.51	0.74	0.51	0.52	0.70	0.57	0.55	0.66
	44034	0.13	0.27	0.75	0.12	0.26	0.76	0.15	0.29	0.70	0.15	0.29	0.70
Unknown Buoys	46035	0.42	0.46	0.84	0.45	0.49	0.82	0.58	0.54	0.77	0.57	0.54	0.78
	46066	0.82	0.66	0.53	0.55	0.53	0.73	0.57	0.55	0.72	0.61	0.56	0.70
	42058	0.14	0.30	0.29	0.13	0.29	0.74	0.16	0.31	0.69	0.16	0.31	0.70
	42057	0.07	0.20	0.21	0.07	0.21	0.72	0.09	0.23	0.66	0.08	0.22	0.68
	42001	0.09	0.24	0.24	0.10	0.24	0.78	0.11	0.25	0.74	0.11	0.24	0.76
	44025	0.13	0.27	0.25	0.11	0.25	0.79	0.12	0.25	0.78	0.12	0.25	0.76

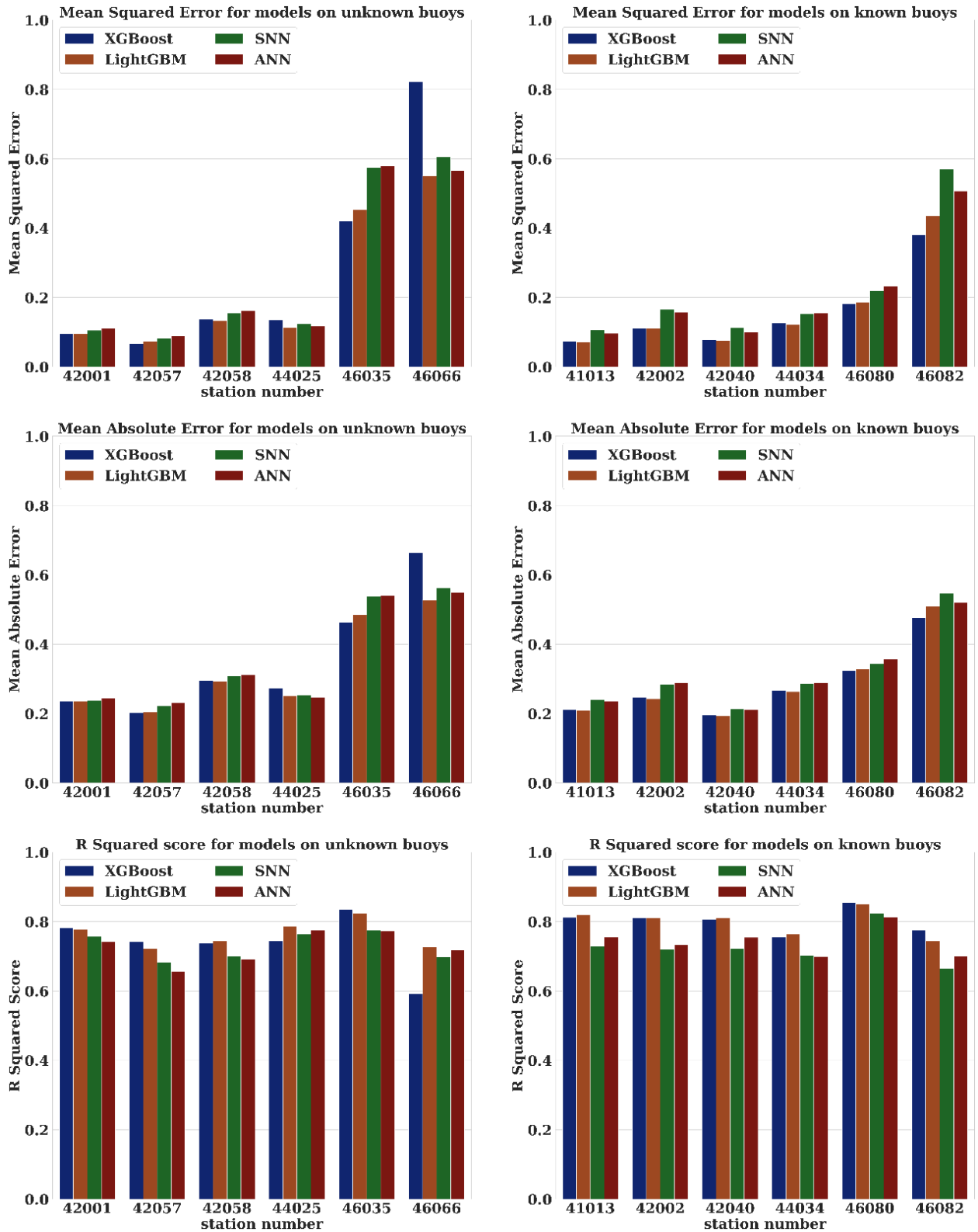


Figure 4-21: Bar plot of error metrics for all the models on different stations

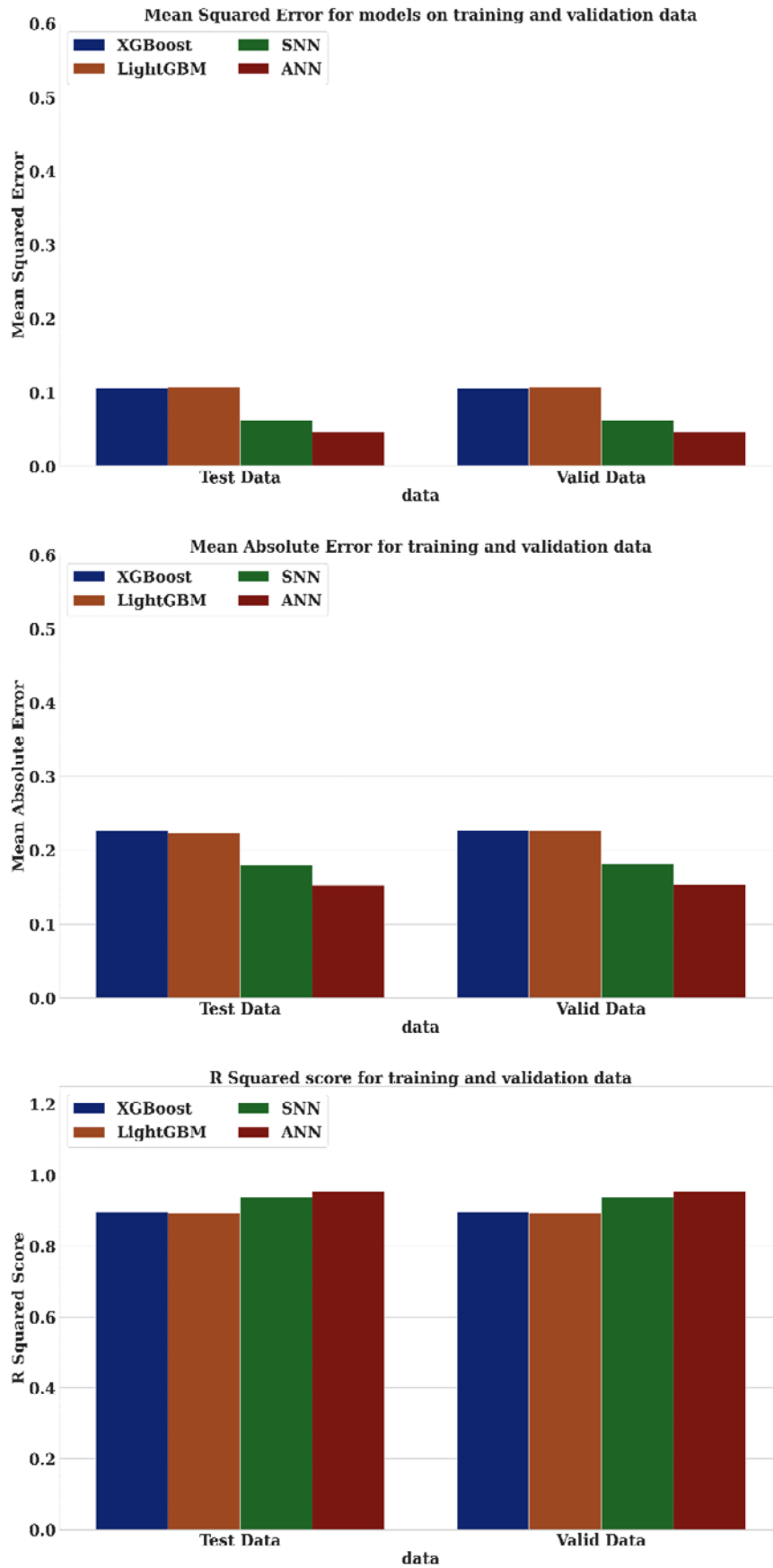


Figure 4-22: Bar plot of error metrics on testing and validation data for all four models

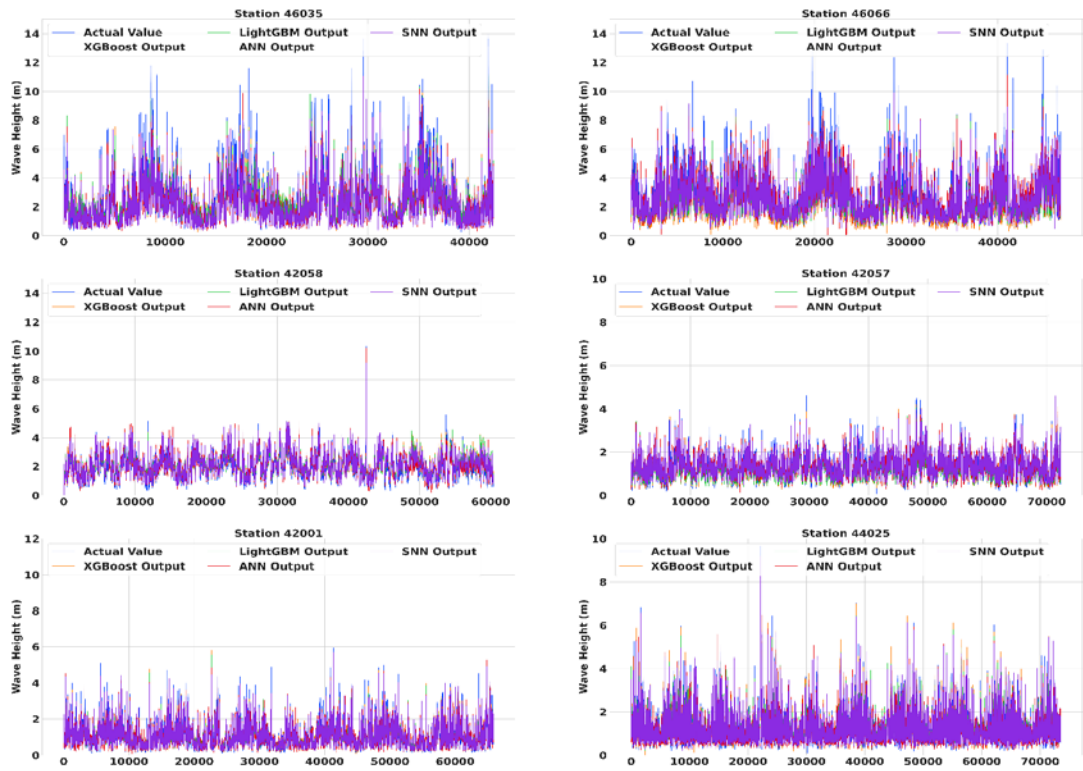


Figure 4-23: Line plot of actual values and prediction of all four models on known buoys

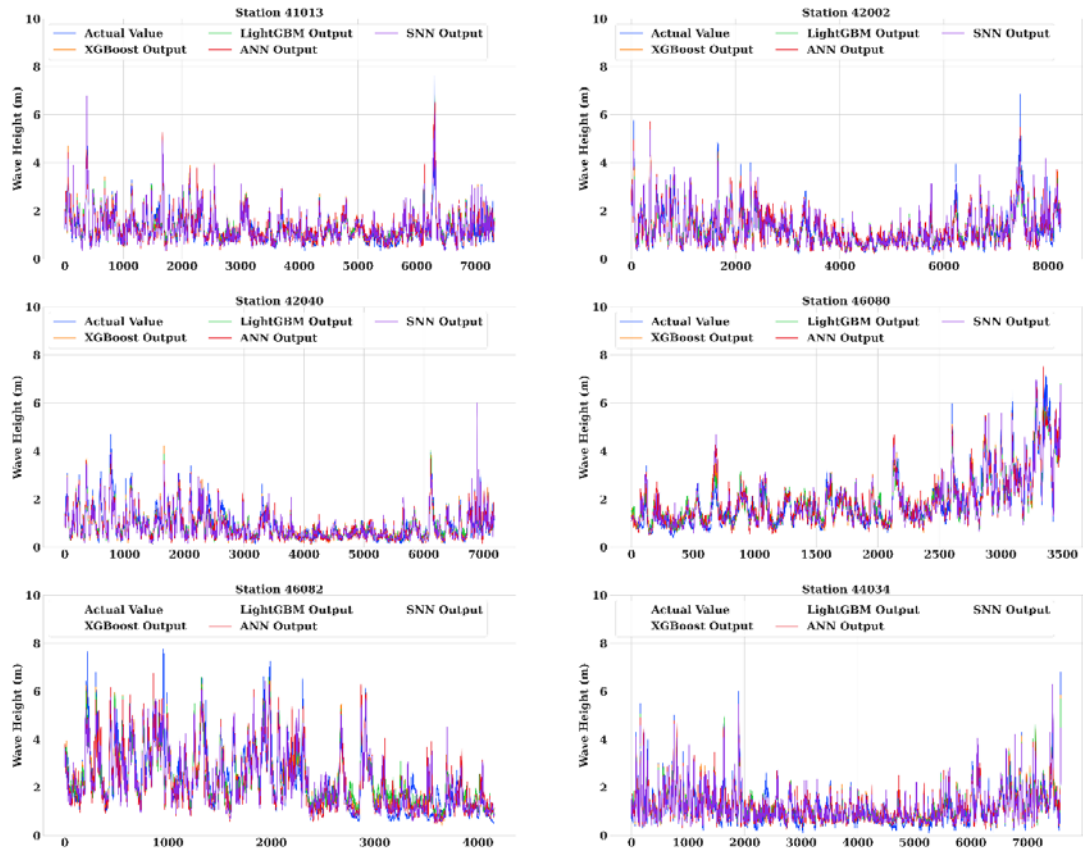


Figure 4-24: Line plot of actual values and prediction of all four models on unknown buoys

The overall performance for the four models is illustrated for known and unknown stations in Figure 4-23 and Figure 4-24 respectively. The four models exhibit comparable trends and means in their predictions. All the models give superior performance on known buoys compared to unknown buoys as expected. Although the mean values can be predicted quite accurately by all of the models, prediction of sudden spike of significant wave height become difficult for all of the models and there are significant errors for these types of values. Both of the tree-based models have shown better performance on known, unknown buoys compared to deep learning models while deep learning models perform better on test and validation dataset. Collectively, they account for a substantial portion of the wave height range. Under these circumstances, it is expected that an ensemble model would exhibit outstanding results, with the individual models demonstrating similar behavior.

In Figure 4-25 and Figure 4-26, scatter plot for all the four models have been illustrated to compare the performances of each model. Both the trend and distribution of the model dispersion are comparable for all known and unknown buoys. The difference in model performance is determined by both the dispersion of the scatter and the behavior of the outlier. In addition, the dispersion of the two gradient boosting models and the two deep learning models are comparable. All the models perform poorly on larger significant wave height values both on known buoys and unknown buoys.

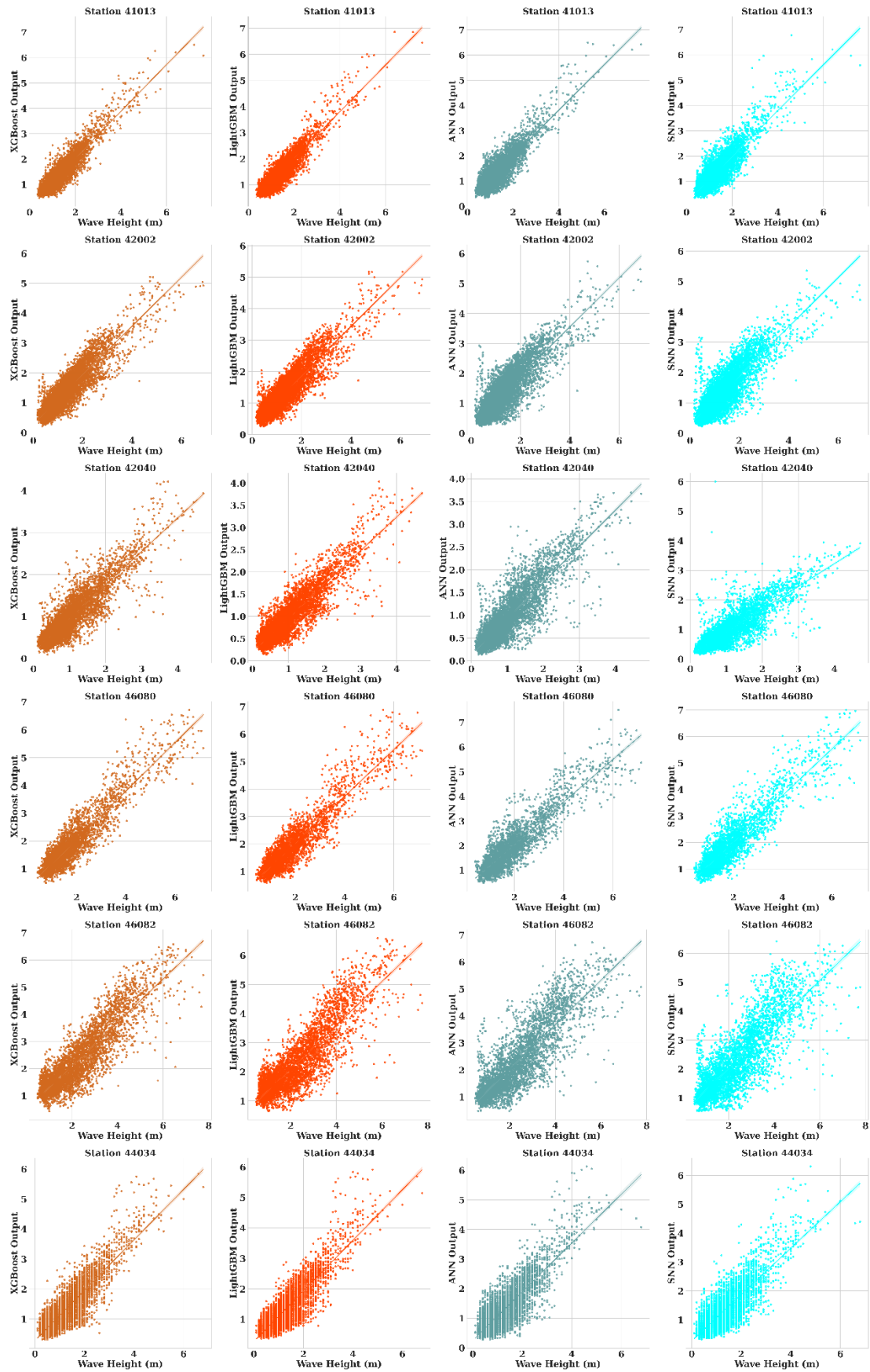


Figure 4-25: Scatter plot of actual values and predicted values of all four models on known buoys

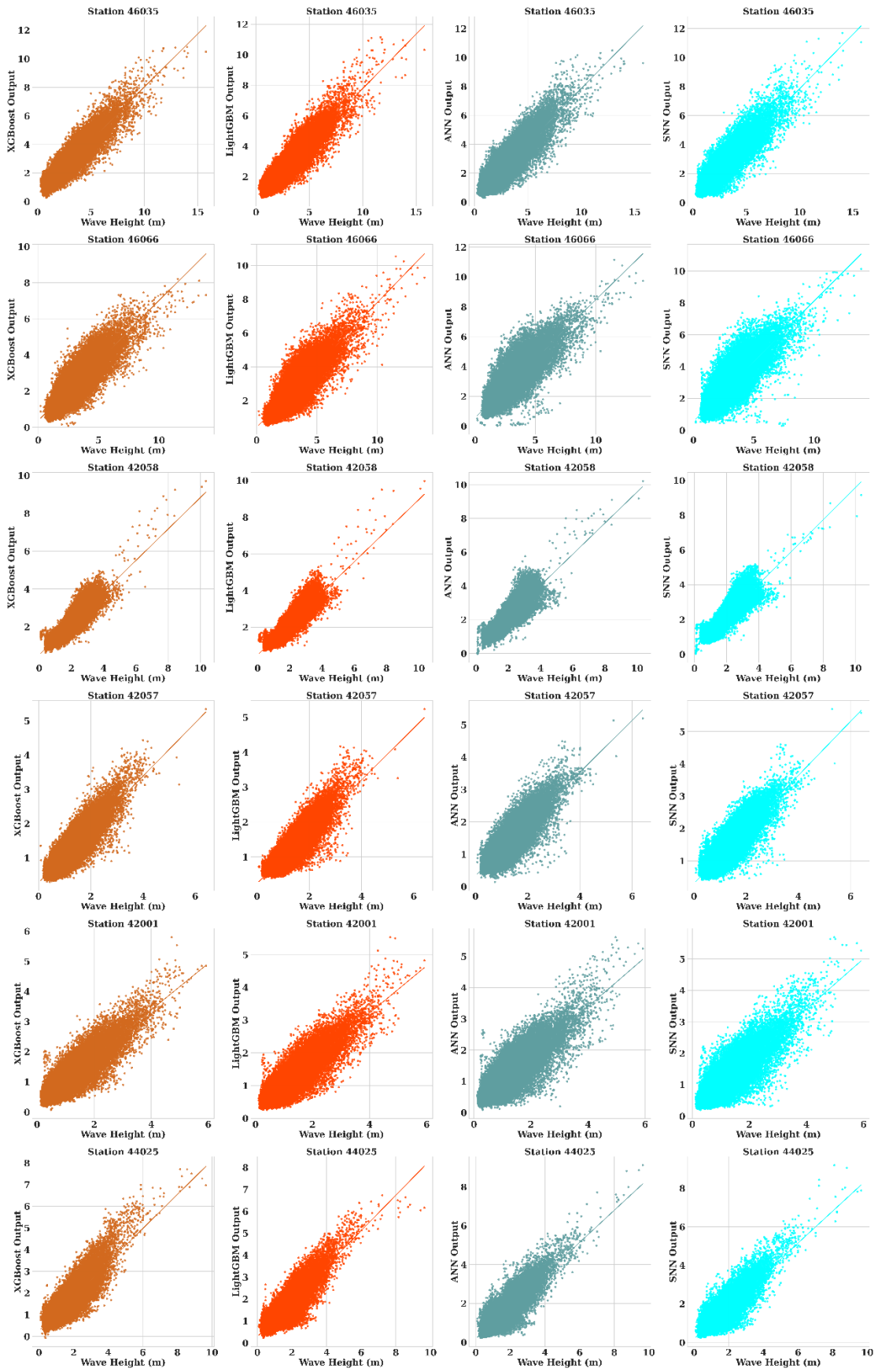


Figure 4-26: Scatter plot of actual values and predicted values of all four models on unknown buoys

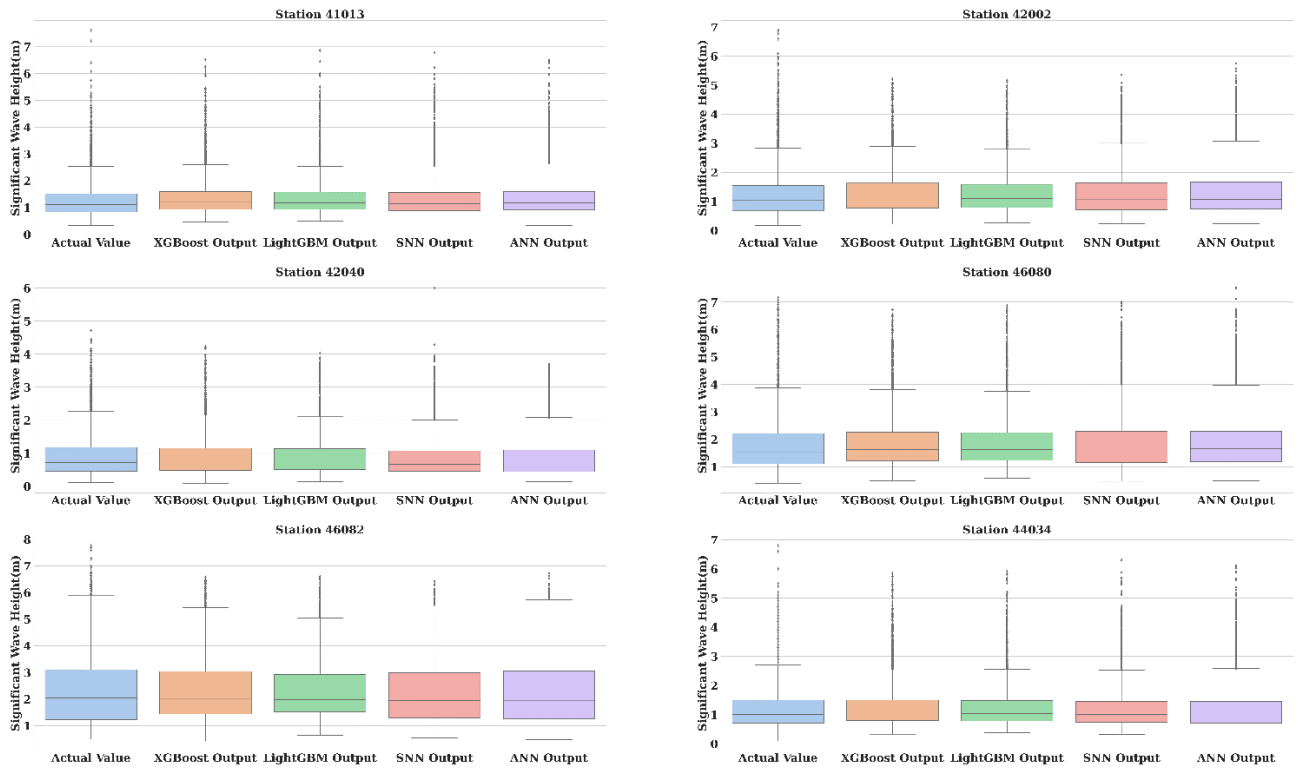


Figure 4-27: Box plot of actual wave height and predicted wave height on known buoys

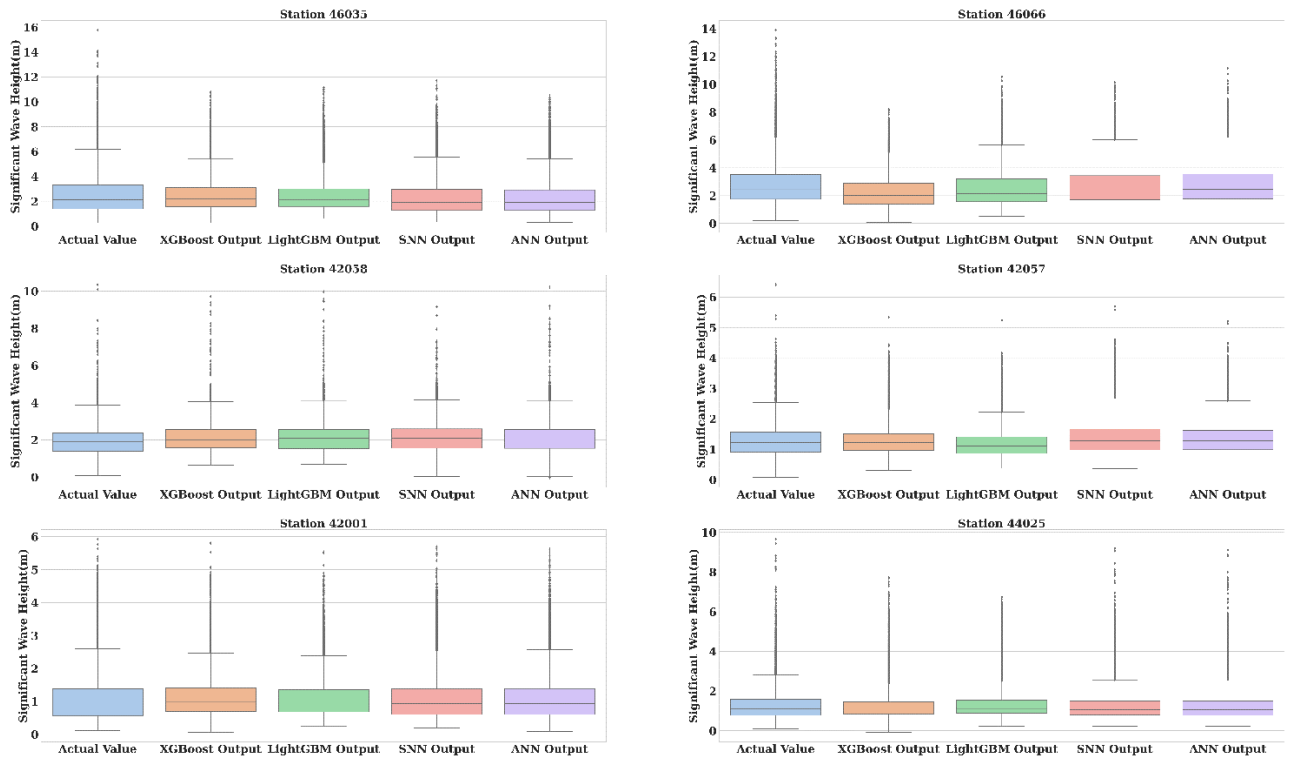


Figure 4-28: Box plot of actual wave height and predicted wave height on unknown buoys

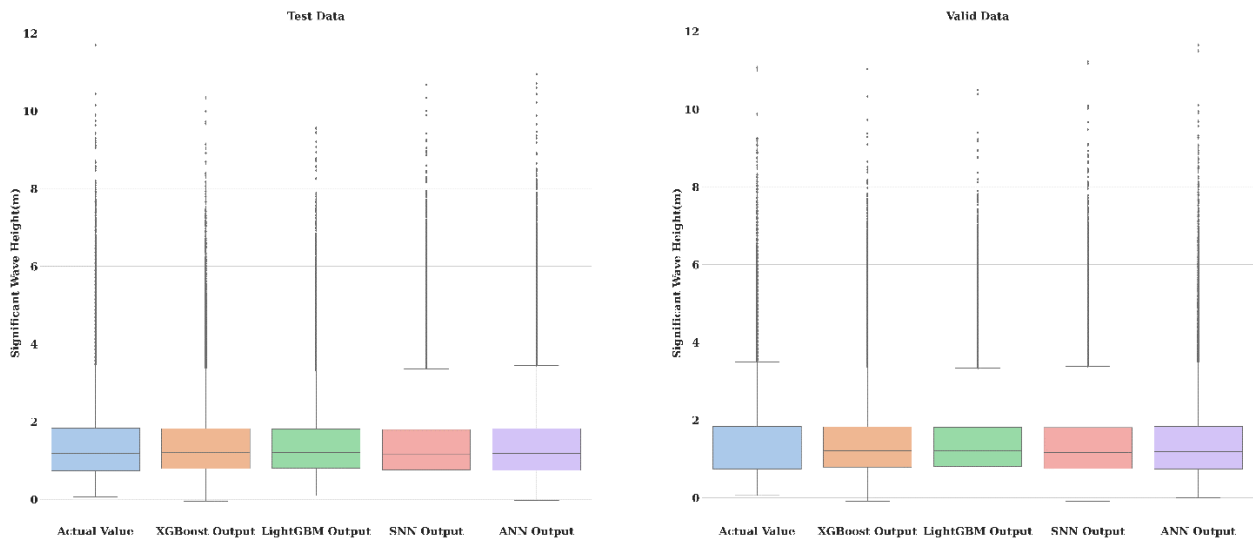


Figure 4-29: Box plot of actual wave height and predicted wave height on test and validation data

The box plot serves as a visual representation of the data distribution, where the box enclosed within the plot corresponds to the Interquartile range (IQR). The two whiskers in a box plot represent a distance of 1.5 times the interquartile range (IQR), and any data points that fall beyond this range are typically classified as outliers.

The distribution of significant wave height both for actual value and predicted values are illustrated in Figure 4-27 for known buoys, Figure 4-28 for unknown buoys and Figure 4-29 for test and validation set. The models exhibit a tendency to be fairly close to the interquartile range (IQR), albeit with a marginal deviation or contraction in the overall span. The models, nevertheless, encounter difficulties in following to the outlier values. The box exhibits a consistent shape across various models, with variations in tail length observed.

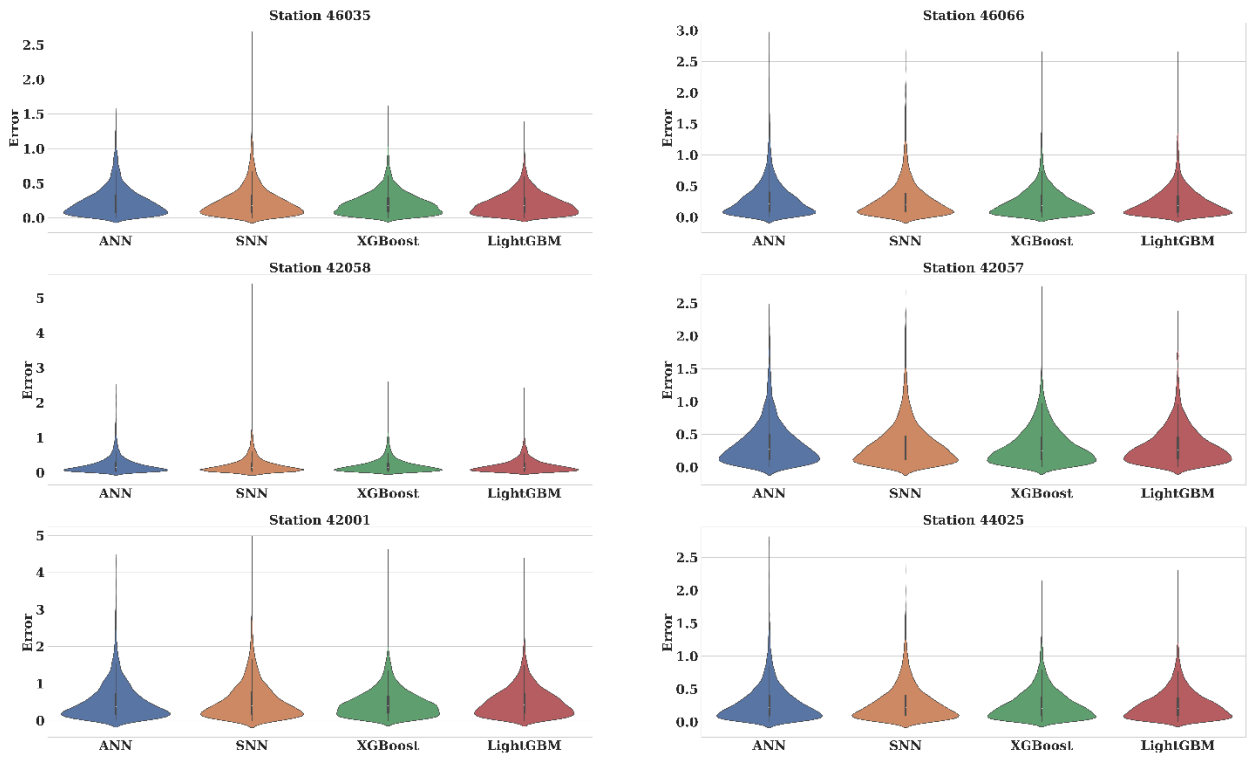


Figure 4-30: Violin plot of residual error of all four models on known buoys

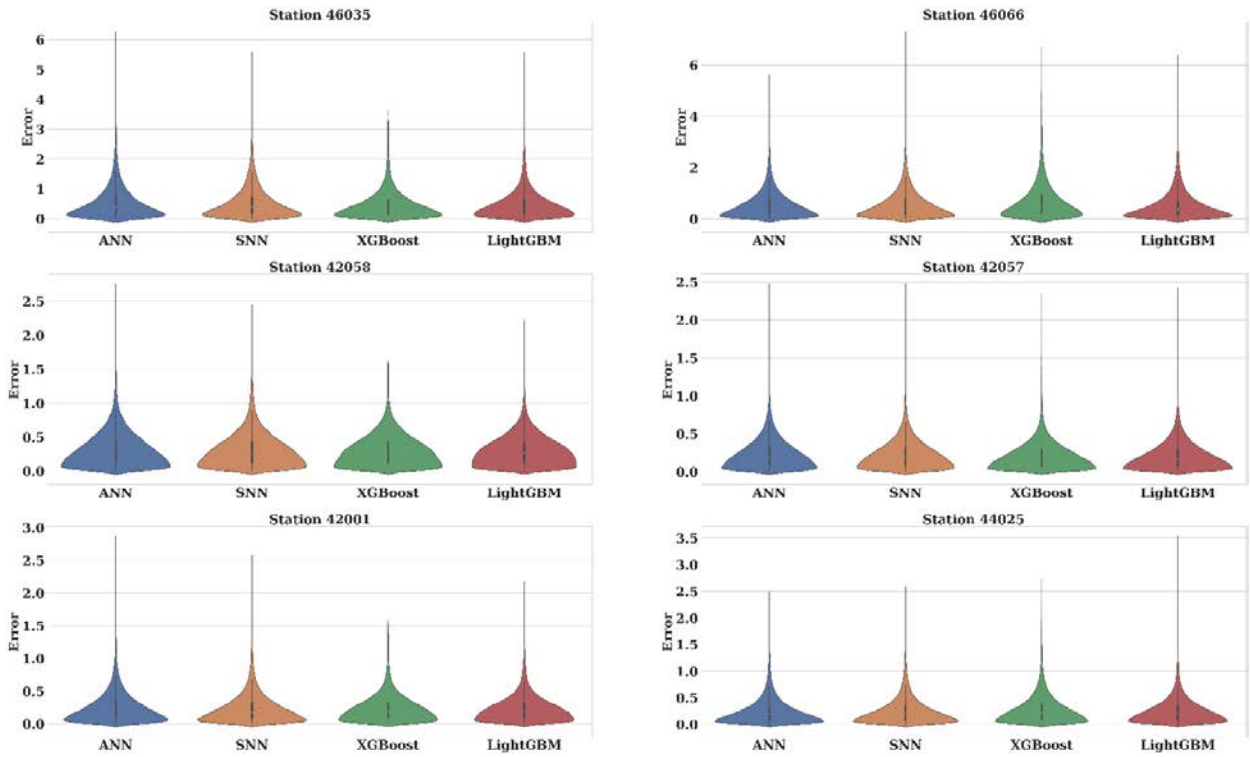


Figure 4-31: Violin plot of residual error of all four models on unknown buoys

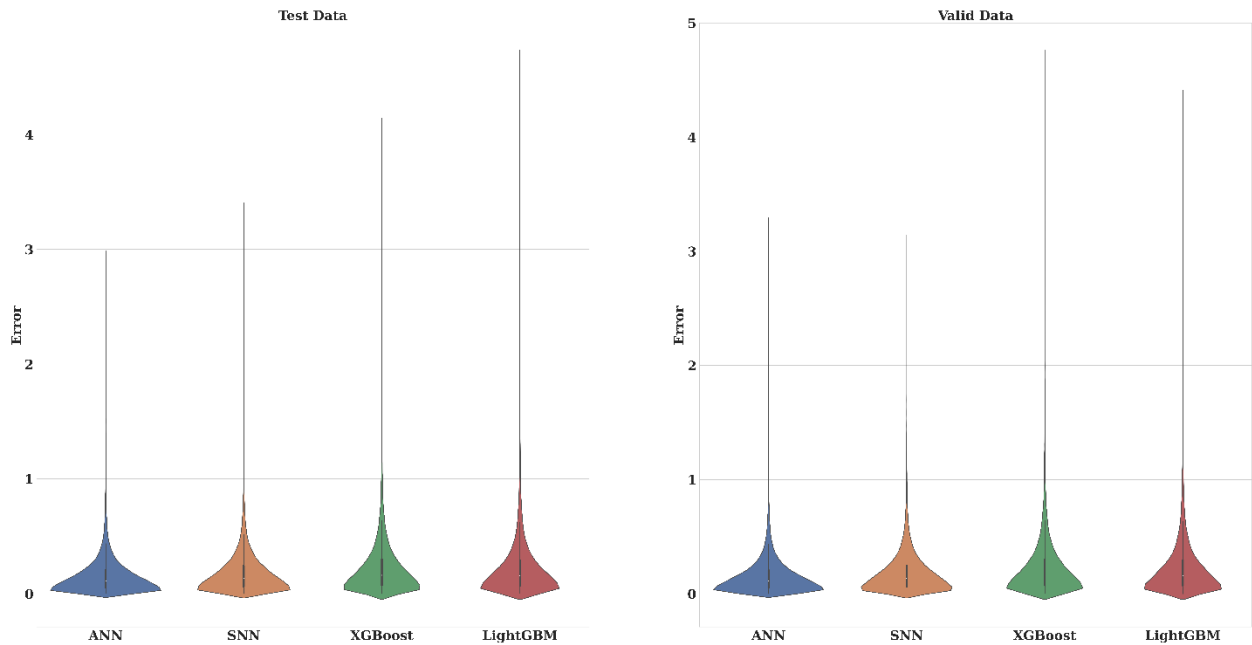


Figure 4-32: Violin plot of residual error of all four models on test and validation data

The error distribution illustrated by violin plot in Figure 4-30, Figure 4-31, Figure 4-32 which exhibit a concentration of error values around zero, with a minor proportion of values increasing towards larger magnitudes. This phenomenon can be ascribed to the influence of outliers and other extraneous factors, and may be considered anomalous and therefore negligible. Moreover, the violin plots exhibit a flat distribution, which suggests a comparatively smaller interquartile range and, thus, a reduced range of values.

4.6 Outlier Evaluation

To compare the effectiveness of the outlier removal algorithm, two models were trained on a dataset with outliers removed and then compared with models trained on conventional data. The identical models were hyper-parameter tuned and trained until convergence. Early stopping was used to prevent overfitting.

The error results for ANN & SNN model after outlier removal are enlisted in Table 4-19 and Table 4-20, it is observed that removing outliers marginally enhances the performance of the ANN model as evidenced by a slight decrease in the Mean Absolute Error. The Mean Squared Error and R-Squared error, however, increase. This may be due to the fact that Mean Squared

Error penalizes outliers severely. The model trained on outlier-removed data augmented its performance on non-outlier data, but performed poorly on outliers in the test set. The outlier removal technique didn't cause any substantial improvement in the SNN model.

Table 4-19: Error evaluation of ANN and SNN models on validation and test data after outlier removal

		ANN			SNN		
		MSE	MAE	R ²	MSE	MAE	R ²
Valid Data	Before Outlier removal	0.046	0.153	0.954	0.063	0.182	0.938
	After Outlier removal	0.049	0.148	0.951	0.072	0.188	0.929
Test Data	Before Outlier removal	0.047	0.153	0.953	0.063	0.180	0.937
	After Outlier removal	0.048	0.146	0.951	0.072	0.187	0.928

Table 4-20: Error evaluation of ANN and SNN models on known and unknown buoy data after outlier removal

	Buoy Number	ANN			SNN		
		MSE	MAE	R ²	MSE	MAE	R ²
Known Buoys	41013	0.235	0.097	0.756	0.239	0.107	0.730
	42002	0.289	0.158	0.733	0.284	0.165	0.720
	42040	0.211	0.099	0.755	0.214	0.112	0.723
	46080	0.358	0.234	0.813	0.345	0.220	0.825
	46082	0.520	0.507	0.701	0.547	0.569	0.665
	44034	0.288	0.155	0.699	0.288	0.153	0.703
Unknown Buoys	46035	0.540	0.579	0.774	0.539	0.574	0.776
	46066	0.550	0.566	0.719	0.562	0.606	0.699
	42058	0.312	0.161	0.691	0.308	0.156	0.700
	42057	0.231	0.090	0.657	0.224	0.083	0.683
	42001	0.245	0.112	0.743	0.237	0.106	0.757
	44025	0.247	0.118	0.776	0.254	0.124	0.764

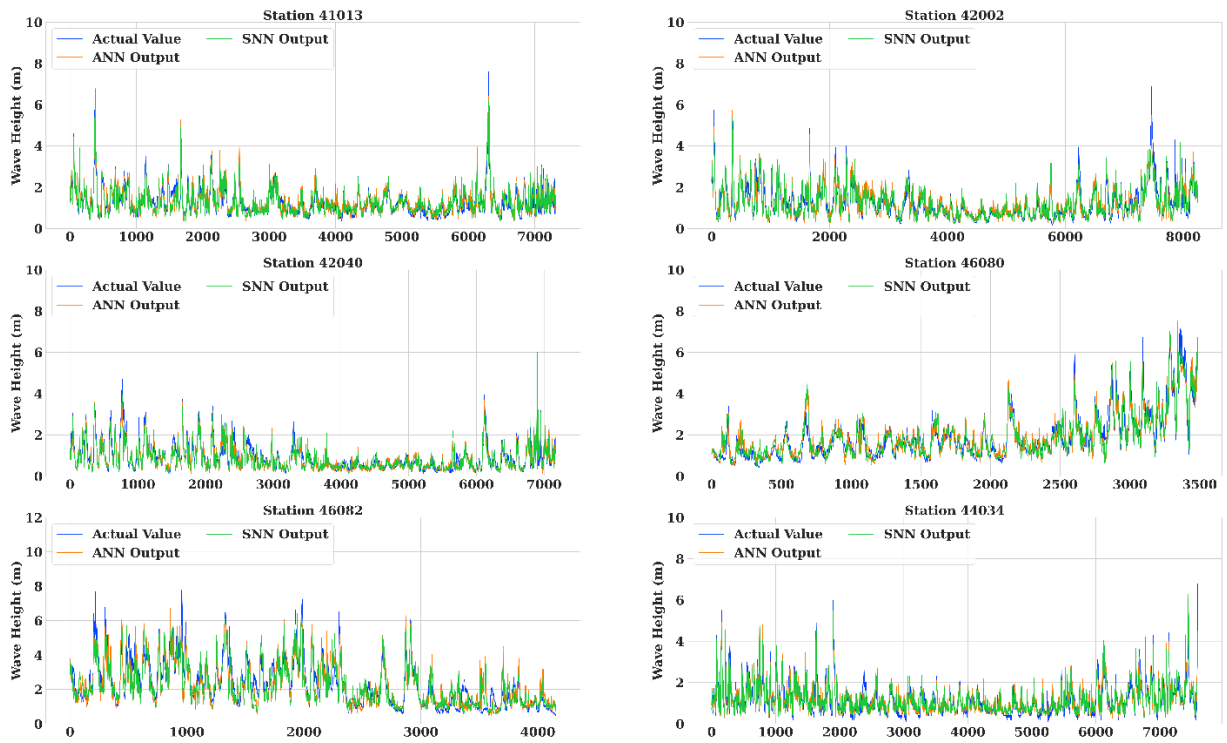


Figure 4-33: Line plot of actual wave height and ANN, SNN model outputs on known buoys after outlier removal

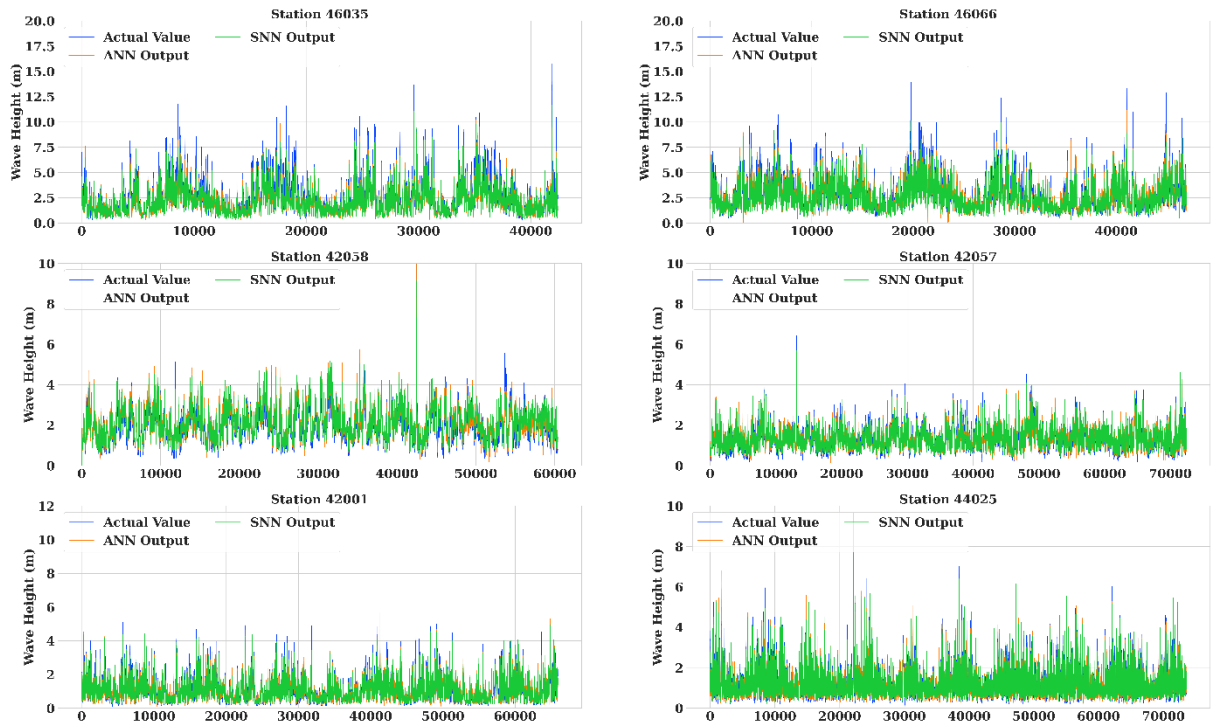


Figure 4-34: Line plot of actual wave height and ANN, SNN model outputs on unknown buoys after outlier removal

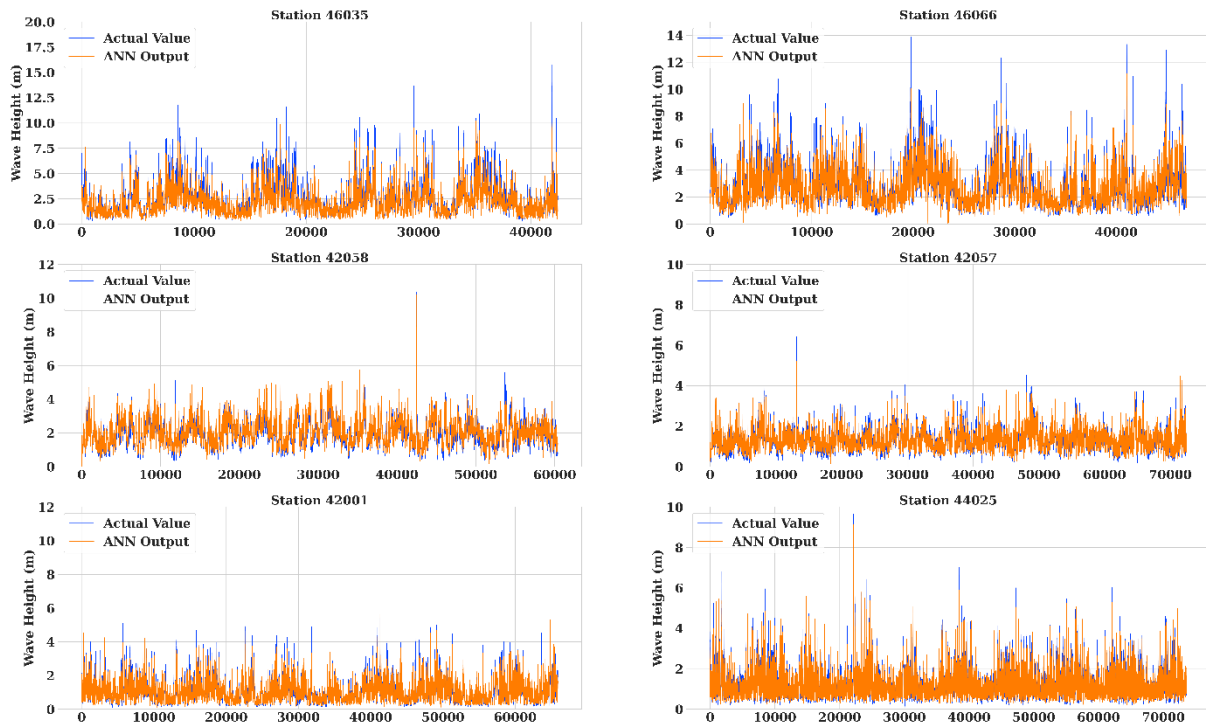


Figure 4-35: Line plot of actual wave height and ANN model outputs on unknown buoys after outlier removal

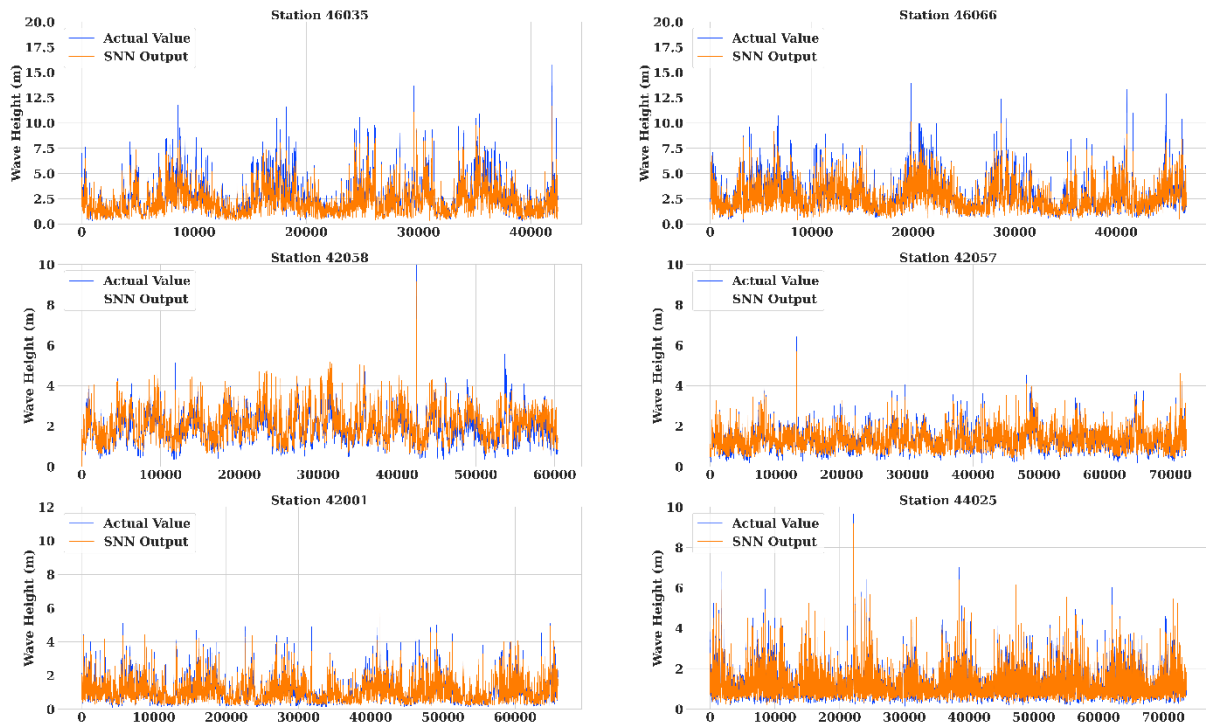


Figure 4-36: Line plot of actual wave height and SNN model outputs on unknown buoys after outlier removal

To assess the impact of outlier removal on model performance for both ANN & SNN model , line plots have been illustrated in Figure 4-33 and Figure 4-34 for known and unknown buoys respectively. If compared with the performance of these two models (ANN & SNN) before outlier removal and after outlier removal, it is evident that the models provide a reasonable fit for the median data but don't improve model performance overall. The model's performance is nearly identical to that of the model trained on the unfiltered dataset, albeit marginally worse, which may be attributable to outlier values or noise in the actual data.

Impact of outlier removal on the performance of ANN model is illustrated in Figure 4-35 for six unknown buoys and in Figure 4-36 for six known buoys. The model performance is quite similar to that of before outlier removal. The model is performing quite well on known stations for the mean values, but poor performance is observed for some specific unknown stations and sudden peak values of significant wave height for each of the station. Outlier removal process couldn't provide significant improvement to the overall performance of the generalized model and so it can be avoided to reduce to complexity of the modelling procedure if the model is further employed on a large scale.

Chapter 5 Conclusions and Future Plan

5.1 Conclusion

The study demonstrated that all four models exhibited acceptable proficiency in predicting significant wave height. Although the models may exhibit inaccuracies in predicting extreme values, it is important to note that these values are considered anomalies and, as such, any associated errors should be deemed negligible. The models exhibit a reasonable level of accuracy in forecasting wave heights at both familiar and unfamiliar buoys, thereby demonstrating their ability to generalize. In this study, gradient boosting models performed better than deep learning models, but deep learning models were simpler to train. Additionally, these models are lightweight and appropriate for IoT applications.

One of the distinctive difficulties in forecasting weather or atmospheric information is the presence of noise or uncertainty that is inherently associated with it. In the context of our study, the training data underwent quality control procedures. However, there was a lack of information pertaining to the collection method, sensor noise, weather anomalies, and other factors that could potentially impact the data collection process. In addition, relevant data regarding the vertical placement and alignment of the temperature, air speed, and air pressure sensors was not available. The presence of these factors introduced noise into the data we collected, resulting in the emergence of outliers and increased complexity in the modeling process.

While the atmospheric data utilized in this investigation hold considerable influence over wave height, certain factors were not taken into account due to limitations in scope, unavailability of data, and implementation complexity. Tidal data, bathymetric data, and water level data are of paramount importance.

The present investigation involved the exclusion of data points with missing values, without any imputation technique applied. Significant amounts of information were lost through this method. The utilization of contemporary imputation techniques has the potential to substantially enhance the quality of the model. Additionally, a simple approach for eliminating outliers was employed. The implementation of complex and effective techniques for eliminating outliers has the potential to enhance the performance of the model.

Ultimately, the enhancement of a model's performance can be achieved through the implementation of appropriate data acquisition and processing techniques. This includes the standardization of data to eliminate any additional noise that may have been introduced during the procurement, processing, imputation, and outlier removal stages. The investigation of complex deep learning, tree models, and hybrid models can serve as a means of enhancing performance. However, it is crucial to exercise caution against overfitting during such studies. The primary objective of a general model is to exhibit robustness in its ability to perform well on novel data, rather than simply memorizing the patterns inherent in the training data.

5.2 Future Plan

- In this work, for the outlier's detection and removal task, a very primitive technique called Tukey's Fence method have been implemented which couldn't cause significant improvement in overall performance of the models both on known and unknown buoys. The next phase of our work will be to implement other modern methods of outlier removal techniques like Winsorization, Hampel Filter, One-Class Support Vector Machines (SVM).
- In this work, all the null values have been excluded from the dataset prior to the training of the model which have caused significant information loss and ultimately deteriorated the overall performance. Instead of removing the null values, in the next phase of our

work, we will use imputation techniques like Iterative Imputation (Multiple Imputation), K-Nearest Neighbors (KNN) Imputation, Random Forest Imputation, Expectation-Maximization (EM) Algorithm, Gaussian Mixture Models (GMM) Imputation to fill out the missing values.

- In this work, standard dataset from NDBC have been utilized without considering the errors of the sensors during measurement, any localized factor that has significant impact on a particular station. In our future work, these factors will be considered for individual stations to make the model a robust one.

References

- [1] A. Agarwal, V. Venugopal, and G. P. Harrison, “The assessment of extreme wave analysis methods applied to potential marine energy sites using numerical model data,” *Renew. Sustain. Energy Rev.*, vol. 27, pp. 244–257, Nov. 2013, doi: 10.1016/j.rser.2013.06.049.
- [2] M. Zhu, “Soil erosion risk assessment with CORINE model: case study in the Danjiangkou Reservoir region, China,” *Stoch. Environ. Res. Risk Assess.*, vol. 26, no. 6, pp. 813–822, Aug. 2012, doi: 10.1007/s00477-011-0511-7.
- [3] D. D. Wu, “Introduction to the special SERRA issue on ‘Risks, Uncertainties and the Environment,’” *Stoch. Environ. Res. Risk Assess.*, vol. 25, no. 3, pp. 301–304, Mar. 2011, doi: 10.1007/s00477-010-0368-1.
- [4] M. Nazir, F. Khan, and P. Amyotte, “Fatigue reliability analysis of deep water rigid marine risers associated with Morison-type wave loading,” *Stoch. Environ. Res. Risk Assess.*, vol. 22, no. 3, pp. 379–390, Apr. 2008, doi: 10.1007/s00477-007-0125-2.
- [5] P. Dixit and S. Londhe, “Prediction of extreme wave heights using neuro wavelet technique,” *Appl. Ocean Res.*, vol. 58, pp. 241–252, Jun. 2016, doi: 10.1016/j.apor.2016.04.011.
- [6] R. Mínguez, A. Tomás, F. J. Méndez, and R. Medina, “Mixed extreme wave climate model for reanalysis databases,” *Stoch. Environ. Res. Risk Assess.*, vol. 27, no. 4, pp. 757–768, May 2013, doi: 10.1007/s00477-012-0604-y.
- [7] M. A. Benitz, M. A. Lackner, and D. P. Schmidt, “Hydrodynamics of offshore structures with specific focus on wind energy applications,” *Renew. Sustain. Energy Rev.*, vol. 44, pp. 692–716, Apr. 2015, doi: 10.1016/j.rser.2015.01.021.
- [8] C. W. Zheng and C. Y. Li, “Variation of the wave energy and significant wave height in the China Sea and adjacent waters,” *Renew. Sustain. Energy Rev.*, vol. 43, pp. 381–387, Mar. 2015, doi: 10.1016/j.rser.2014.11.001.
- [9] E. Vanem, “Long-term time-dependent stochastic modelling of extreme waves,” *Stoch. Environ. Res. Risk Assess.*, vol. 25, no. 2, pp. 185–209, Feb. 2011, doi: 10.1007/s00477-010-0431-y.

- [10] L. Liu, D. Wang, and Z. Peng, "Path following of marine surface vehicles with dynamical uncertainty and time-varying ocean disturbances," *Neurocomputing*, vol. 173, pp. 799–808, Jan. 2016, doi: 10.1016/j.neucom.2015.08.033.
- [11] D. Demetriou, C. Michailides, G. Papanastasiou, and T. Onoufriou, "Coastal zone significant wave height prediction by supervised machine learning classification algorithms," *Ocean Eng.*, vol. 221, p. 108592, Feb. 2021, doi: 10.1016/j.oceaneng.2021.108592.
- [12] S. Foteinis, "Wave energy converters in low energy seas: Current state and opportunities," *Renew. Sustain. Energy Rev.*, vol. 162, no. September 2021, p. 112448, Jul. 2022, doi: 10.1016/j.rser.2022.112448.
- [13] A. F. d. O. Falcão, "Wave energy utilization: A review of the technologies," *Renew. Sustain. Energy Rev.*, vol. 14, no. 3, pp. 899–918, 2010, doi: 10.1016/j.rser.2009.11.003.
- [14] D. Qiao, R. Haider, J. Yan, D. Ning, and B. Li, "Review of wave energy converter and design of mooring system," *Sustain.*, vol. 12, no. 19, pp. 1–31, 2020, doi: 10.3390/su12198251.
- [15] L. Cuadra, S. Salcedo-Sanz, J. C. Nieto-Borge, E. Alexandre, and G. Rodríguez, "Computational intelligence in wave energy: Comprehensive review and case study," *Renew. Sustain. Energy Rev.*, vol. 58, pp. 1223–1246, May 2016, doi: 10.1016/j.rser.2015.12.253.
- [16] I. Malekmohamadi, M. R. Bazargan-Lari, R. Kerachian, M. R. Nikoo, and M. Fallahnia, "Evaluating the efficacy of SVMs, BNs, ANNs and ANFIS in wave height prediction," *Ocean Eng.*, vol. 38, no. 2–3, pp. 487–497, Feb. 2011, doi: 10.1016/j.oceaneng.2010.11.020.
- [17] S. P. Nitsure, S. N. Londhe, and K. C. Khare, "Wave forecasts using wind information and genetic programming," *Ocean Eng.*, vol. 54, pp. 61–69, Nov. 2012, doi: 10.1016/j.oceaneng.2012.07.017.
- [18] S. Elipot, A. Sykulski, R. Lumpkin, L. Centurioni, and M. Pazos, "A dataset of hourly sea surface temperature from drifting buoys," *Sci. Data*, vol. 9, no. 1, pp. 1–27, 2022, doi: 10.1038/s41597-022-01670-2.

- [19] “The Use of Buoys In The Ocean And What Weather Conditions They Measure.” .
- [20] “NDBC - NDBC’s Drifting Buoy Program.” .
- [21] “NDBC - Moored Buoy Program.” .
- [22] S. Elipot, R. Lumpkin, R. C. Perez, J. M. Lilly, J. J. Early, and A. M. Sykulski, “A global surface drifter data set at hourly resolution,” *J. Geophys. Res. Ocean.*, vol. 121, no. 5, pp. 2937–2966, 2016, doi: 10.1002/2016JC011716.
- [23] H. Hu, A. J. van der Westhuysen, P. Chu, and A. Fujisaki-Manome, “Predicting Lake Erie wave heights and periods using XGBoost and LSTM,” *Ocean Model.*, vol. 164, Aug. 2021, doi: 10.1016/j.ocemod.2021.101832.
- [24] K. Hasselmann *et al.*, “Measurements of wind-wave growth and swell decay during the joint North Sea wave project (JONSWAP).,” no. July 2015, 1973.
- [25] T. W. Group, “The WAM Model—A Third Generation Ocean Wave Prediction Model,” *J. Phys. Oceanogr.*, vol. 18, no. 12, pp. 1775–1810, Dec. 1988, doi: 10.1175/1520-0485(1988)018<1775:TWMTGO>2.0.CO;2.
- [26] N. Booij, R. C. Ris, and L. H. Holthuijsen, “A third-generation wave model for coastal regions: 1. Model description and validation,” *J. Geophys. Res. Ocean.*, vol. 104, no. C4, pp. 7649–7666, Apr. 1999, doi: 10.1029/98JC02622.
- [27] H. L. Tolman *et al.*, “Development and Implementation of Wind-Generated Ocean Surface Wave Models at NCEP*,” *Weather Forecast.*, vol. 17, no. 2, pp. 311–333, Apr. 2002, doi: 10.1175/1520-0434(2002)017<0311:DAIOWG>2.0.CO;2.
- [28] M. Zijlema, “Computation of wind-wave spectra in coastal waters with SWAN on unstructured grids,” *Coast. Eng.*, vol. 57, no. 3, pp. 267–277, Mar. 2010, doi: 10.1016/j.coastaleng.2009.10.011.
- [29] N. K. Kumar, R. Savitha, and A. Al Mamun, “Ocean wave height prediction using ensemble of Extreme Learning Machine,” *Neurocomputing*, vol. 277, pp. 12–20, Feb. 2018, doi: 10.1016/j.neucom.2017.03.092.
- [30] M. Browne, B. Castelle, D. Strauss, R. Tomlinson, M. Blumenstein, and C. Lane, “Near-shore swell estimation from a global wind-wave model: Spectral process, linear, and artificial neural network models,” *Coast. Eng.*, vol. 54, no. 5, pp. 445–460, May

- 2007, doi: 10.1016/j.coastaleng.2006.11.007.
- [31] J.-H. G. M. Alves, A. Chawla, H. L. Tolman, D. Schwab, G. Lang, and G. Mann, “The Operational Implementation of a Great Lakes Wave Forecasting System at NOAA/NCEP*,” *Weather Forecast.*, vol. 29, no. 6, pp. 1473–1497, Dec. 2014, doi: 10.1175/WAF-D-12-00049.1.
- [32] F. Zabihian and A. S. Fung, “Review of marine renewable energies: Case study of Iran,” *Renew. Sustain. Energy Rev.*, vol. 15, no. 5, pp. 2461–2474, 2011, doi: 10.1016/j.rser.2011.02.006.
- [33] M. Fadaeenejad, R. Shamsipour, S. D. Rokni, and C. Gomes, “New approaches in harnessing wave energy: With special attention to small islands,” *Renew. Sustain. Energy Rev.*, vol. 29, pp. 345–354, 2014, doi: 10.1016/j.rser.2013.08.077.
- [34] M. A. J. R. Quirapas, H. Lin, M. L. S. Abundo, S. Brahim, and D. Santos, “Ocean renewable energy in Southeast Asia: A review,” *Renew. Sustain. Energy Rev.*, vol. 41, pp. 799–817, 2015, doi: 10.1016/j.rser.2014.08.016.
- [35] N. Khan, A. Kalair, N. Abas, and A. Haider, “Review of ocean tidal, wave and thermal energy technologies,” *Renew. Sustain. Energy Rev.*, vol. 72, no. January, pp. 590–604, 2017, doi: 10.1016/j.rser.2017.01.079.
- [36] D. Khojasteh, S. M. Mousavi, W. Glamore, and G. Iglesias, “Wave energy status in Asia,” *Ocean Eng.*, vol. 169, no. April, pp. 344–358, 2018, doi: 10.1016/j.oceaneng.2018.09.034.
- [37] R. Pelc and R. M. Fujita, “Renewable energy from the ocean,” *Mar. Policy*, vol. 26, no. 6, pp. 471–479, Nov. 2002, doi: 10.1016/S0308-597X(02)00045-3.
- [38] T. . Thorpe, “An overview of wave energy technologies: Status, performance and costs,” *Wave Power Mov. Towar. Commer. viability*, no. November, p. not paginated, 1999.
- [39] M. L. Ae *et al.*, “Wave Energy from the North Sea: Experiences from the Lysekil Research Site,” vol. 29, pp. 221–240, 2008, doi: 10.1007/s10712-008-9047-x.
- [40] Y. P. Wimalaratna *et al.*, “Comprehensive review on the feasibility of developing wave energy as a renewable energy resource in Australia,” *Clean. Energy Syst.*, vol. 3, no. April, p. 100021, 2022, doi: 10.1016/j.cles.2022.100021.

- [41] E. Al Shami, R. Zhang, and X. Wang, “Point absorber wave energy harvesters: A review of recent developments,” *Energies*, vol. 12, no. 1, 2019, doi: 10.3390/en12010047.
- [42] T. Hirohisa, “Sea trial of a heaving buoy wave power absorber,” in *Proceedings of 2nd international symposium on wave energy utilization, Trondheim, Norway*, 1982, pp. 403–417.
- [43] K. Budal *et al.*, “The Norwegian wave-power buoy project,” 1982.
- [44] M. E. McCormick, “Wave-powered reverse-osmosis desalination,” *Sea Technol.*, vol. 42, pp. 37–39, 2001.
- [45] J. Ringwood, “The dynamics of wave energy,” pp. 23–34, Jan. 2007, doi: 10.1049/CP:20060408.
- [46] “Pelamis Wave Power - Wikipedia.” https://en.wikipedia.org/wiki/Pelamis_Wave_Power (accessed Jan. 25, 2023).
- [47] A. H. Clément *et al.*, “Wave energy in Europe: current status and perspectives,” *Renew. & Sustain. Energy Rev.*, vol. 6, pp. 405–431, 2002.
- [48] “WPP A/S Products- Waveplane.” <http://www.waveplane.com/products.html> (accessed Jan. 25, 2023).
- [49] “Wave Dragon.” wavedragon.net.
- [50] H. C. Soerensen, W. Panhauser, D. Duncce, J. Nedkvintne, P. Frigaard, and J. P. Kofoed, “Development of Wave Dragon from Scale 1 : 50 to Prototype,” *Coast. Eng.*, pp. 5–11, 2003.
- [51] L. Cameron *et al.*, “Design of the Next Generation of the Oyster Wave Energy Converter,” 2010.
- [52] “Home - AW-Energy.” <https://aw-energy.com/> (accessed Jan. 25, 2023).
- [53] S. Salter, “Apparatus and method for extracting wave energy,” *United States Pat.*, vol. 3,928,967, no. 521,385, p. 8, 1975.
- [54] N. Tom, Y. Yu, A. Wright, N. Tom, Y. Yu, and A. Wright, “Submerged Pressure Differential Plate Wave Energy Converter with Variable Geometry Preprint Submerged Pressure Differential Plate Wave Energy Converter with Variable

- Geometry Preprint,” no. October, 2019.
- [55] “News - AWS Ocean Energy.” <https://awsocan.com/news/> (accessed Jan. 28, 2023).
- [56] “AWS wave energy converter takes shape in Glasgow ahead of testing - AWS Ocean Energy.” <https://awsocan.com/2021/06/waveswing-takes-shape-in-glasgow/> (accessed Jan. 28, 2023).
- [57] “AWS Waveswing trials exceed expectations - AWS Ocean Energy.” <https://awsocan.com/2022/11/aws-waveswing-trials-exceed-expectations/> (accessed Jan. 28, 2023).
- [58] “Tide - Wikipedia.” <https://en.wikipedia.org/wiki/Tide> (accessed Jan. 28, 2023).
- [59] R. Pelc and R. M. Fujita, “Renewable energy from the ocean,” *Mar. Policy*, vol. 26, no. 6, pp. 471–479, 2002, doi: [https://doi.org/10.1016/S0308-597X\(02\)00045-3](https://doi.org/10.1016/S0308-597X(02)00045-3).
- [60] “Sihwa Lake Tidal Power Station - Wikipedia.” https://en.wikipedia.org/wiki/Sihwa_Lake_Tidal_Power_Station (accessed Jan. 28, 2023).
- [61] “Sihwa Tidal Power Plant | Tethys,” *tethys.pnnl.gov*, Accessed: Jan. 28, 2023. [Online]. Available: <https://tethys.pnnl.gov/project-sites/sihwa-tidal-power-plant#:~:text=To date, the Sihwa TBPP is the largest and most expensive tidal installation in the world.>
- [62] *Korea’s Uldolmok tidal power project completed.* .
- [63] “Uldolmok Tidal Power Station - Wikipedia.” https://en.wikipedia.org/wiki/Uldolmok_Tidal_Power_Station (accessed Jan. 29, 2023).
- [64] “Rance Tidal Power Station - Wikipedia.” https://en.wikipedia.org/wiki/Rance_Tidal_Power_Station (accessed Jan. 28, 2023).
- [65] “Jiangxia Tidal Power Station - Wikipedia.” https://en.wikipedia.org/wiki/Jiangxia_Tidal_Power_Station (accessed Jan. 29, 2023).
- [66] “Tidal stream generator - Wikipedia.” https://en.wikipedia.org/wiki/Tidal_stream_generator (accessed Jan. 29, 2023).
- [67] “ScotRenewables SR2000 at EMEC | Tethys.” <https://tethys.pnnl.gov/project->

- sites/scotrenewables-sr2000-emec (accessed Jan. 31, 2023).
- [68] M. Dickie, “Scotland unveils world’s largest tidal stream power project,” Sep. 2016, Accessed: Jan. 29, 2023. [Online]. Available: <https://webcache.googleusercontent.com/search?q=cache:NLYheToLCeAJ:https://www.ft.com/content/d197308a-7826-11e6-97ae-647294649b28+&cd=1&hl=en&ct=clnk&gl=us>.
- [69] M. Nachtane, M. Tarfaoui, I. Goda, and M. Rouway, “A review on the technologies, design considerations and numerical models of tidal current turbines,” *Renew. Energy*, vol. 157, pp. 1274–1288, 2020, doi: 10.1016/j.renene.2020.04.155.
- [70] “Project overview - FORWARD2030.” <https://forward2030.tech/project-overview> (accessed Jan. 31, 2023).
- [71] “Orbital Marine Power Launches O2: World’s Most Powerful Tidal Turbine - Orbital Marine.” <https://orbitalmarine.com/orbital-marine-power-launches-o2/#> (accessed Jan. 31, 2023).
- [72] “Innovation - Orbital Marine.” <https://orbitalmarine.com/innovation/#> (accessed Jan. 31, 2023).
- [73] “Kvalsund Tidal Turbine Prototype | Tethys.” <https://tethys.pnnl.gov/project-sites/kvalsund-tidal-turbine-prototype> (accessed Jan. 31, 2023).
- [74] “MeyGen - Wikipedia.” <https://en.wikipedia.org/wiki/MeyGen> (accessed Jan. 29, 2023).
- [75] “MeyGen Tidal Energy Project - Phase I | Tethys.” <https://tethys.pnnl.gov/project-sites/meygen-tidal-energy-project-phase-i> (accessed Jan. 31, 2023).
- [76] H. Chen, T. Tang, N. Ait-Ahmed, M. E. H. Benbouzid, M. MacHmoum, and M. E. H. Zaim, “Attraction, Challenge and Current Status of Marine Current Energy,” *IEEE Access*, vol. 6, pp. 12665–12685, Jan. 2018, doi: 10.1109/ACCESS.2018.2795708.
- [77] “World’s First Open Sea Tidal Turbine | REUK.co.uk.” <http://www.reuk.co.uk/wordpress/tidal/worlds-first-open-sea-tidal-turbine/> (accessed Jan. 31, 2023).
- [78] “Worldchanging | Evaluation + Tools + Best Practices: Gorlov’s Helical Turbine.”

- <https://web.archive.org/web/20130511010214/http://www.worldchanging.com/archives/002383.html> (accessed Jan. 31, 2023).
- [79] “Shark biomimicry produces renewable energy system.”
<https://news.mongabay.com/2006/11/shark-biomimicry-produces-renewable-energy-system/> (accessed Jan. 31, 2023).
- [80] “Underwater kite-turbine may turn tides into green electricity | Renewable energy | The Guardian.” <https://www.theguardian.com/environment/damian-carrington-blog/2011/mar/02/underwater-kite-turbine-green-electricity> (accessed Jan. 31, 2023).
- [81] “Tidal power - Wikipedia.” https://en.wikipedia.org/wiki/Tidal_power (accessed Jan. 28, 2023).
- [82] S. Waters and G. Aggidis, “A world first: Swansea Bay tidal lagoon in review,” *Renew. Sustain. Energy Rev.*, vol. 56, pp. 916–921, Apr. 2016, doi: 10.1016/J.RSER.2015.12.011.
- [83] “Green light expected for multi-billion-pound tidal lagoon project,” *Nation.Cymru*, Jan. 2023, Accessed: Feb. 01, 2023. [Online]. Available: <https://nation.cymru/news/green-light-expected-for-multi-billion-pound-tidal-lagoon-project/>.
- [84] “Tidal Lagoon Swansea Bay - Wikipedia.”
https://en.wikipedia.org/wiki/Tidal_Lagoon_Swansea_Bay (accessed Feb. 01, 2023).
- [85] D. Vandercruyssen, S. Baker, D. Howard, and G. Aggidis, “Tidal range electricity generation: A comparison between estuarine barrages and coastal lagoons,” *Heliyon*, vol. 8, no. 11, p. e11381, 2022, doi: 10.1016/j.heliyon.2022.e11381.
- [86] P. Dai *et al.*, “Numerical study of hydrodynamic mechanism of dynamic tidal power,” *Water Sci. Eng.*, vol. 11, no. 3, pp. 220–228, 2018, doi: 10.1016/j.wse.2018.09.004.
- [87] “Dynamic tidal power - Wikipedia.”
https://en.wikipedia.org/wiki/Dynamic_tidal_power (accessed Feb. 01, 2023).
- [88] Y. H. Park, “Analysis of characteristics of Dynamic Tidal Power on the west coast of Korea,” *Renew. Sustain. Energy Rev.*, vol. 68, no. October 2016, pp. 461–474, 2017, doi: 10.1016/j.rser.2016.10.008.

- [89] P. Dai, J. Zhang, and J. Zheng, “Predictions for Dynamic Tidal Power and Associated Local Hydrodynamic Impact in the Taiwan Strait, China,” *J. Coast. Res.*, vol. 33, no. 1, pp. 149–157, 2017, doi: 10.2112/JCOASTRES-D-16-00068.1.
- [90] D. Shao, W. Feng, X. Feng, and Y. Xu, “Reinvestigation of the Dynamic Tidal Power Dams and their Influences on Hydrodynamic Environment,” *IOP Conf. Ser. Earth Environ. Sci.*, vol. 63, no. 1, 2017, doi: 10.1088/1755-1315/63/1/012048.
- [91] Y. H. Park, “The Application of Dynamic Tidal Power in Korea,” *J. Coast. Res.*, vol. 85, pp. 1306–1310, 2018, doi: 10.2112/SI85-262.1.
- [92] E. Armoudli, A. Mohseni, A. Sheykhi, and K. Javaherdeh, “A Case Study of Energy Harvesting by Dynamic Tidal Power in the Persian Gulf,” *2019 Iran. Conf. Renew. Energy Distrib. Gener. ICREDG 2019*, pp. 11–12, 2019, doi: 10.1109/ICREDG47187.2019.194178.
- [93] J. G. McGowan, “Ocean thermal energy conversion—A significant solar resource,” *Sol. Energy*, vol. 18, no. 2, pp. 81–92, Jan. 1976, doi: 10.1016/0038-092X(76)90042-6.
- [94] W. Liu *et al.*, “A review of research on the closed thermodynamic cycles of ocean thermal energy conversion,” *Renew. Sustain. Energy Rev.*, vol. 119, p. 109581, Mar. 2020, doi: 10.1016/J.RSER.2019.109581.
- [95] S. M. Abbas, H. D. S. Alhassany, D. Vera, and F. Jurado, “Review of enhancement for ocean thermal energy conversion system,” *J. Ocean Eng. Sci.*, no. xxxx, 2022, doi: 10.1016/j.joes.2022.03.008.
- [96] F. Chen, L. Zhang, W. Liu, L. Liu, and J. Peng, “Thermodynamic analysis of rankine cycle in ocean thermal energy conversion,” *Int. J. Simul. Syst. Sci. Technol.*, vol. 17, no. 13, pp. 7.1-7.4, 2016, doi: 10.5013/IJSSST.A.17.13.07.
- [97] K. Rajagopalan and G. C. Nihous, “Estimates of global Ocean Thermal Energy Conversion (OTEC) resources using an ocean general circulation model,” *Renew. Energy*, vol. 50, pp. 532–540, Feb. 2013, doi: 10.1016/J.RENENE.2012.07.014.
- [98] J. Langer, J. Quist, and K. Blok, “Upscaling scenarios for ocean thermal energy conversion with technological learning in Indonesia and their global relevance,” *Renew. Sustain. Energy Rev.*, vol. 158, no. January, p. 112086, 2022, doi: 10.1016/j.rser.2022.112086.

- [99] J. Langer, C. Infante Ferreira, and J. Quist, “Is bigger always better? Designing economically feasible ocean thermal energy conversion systems using spatiotemporal resource data,” *Appl. Energy*, vol. 309, no. July 2021, p. 118414, 2022, doi: 10.1016/j.apenergy.2021.118414.
- [100] T. Mitsui, F. Ito, Y. Seya, and Y. Nakamoto, “Outline of the 100 kW OTEC Pilot Plant in the Republic of Nauru,” *IEEE Trans. Power Appar. Syst.*, vol. PAS-102, no. 9, pp. 3167–3171, 1983, doi: 10.1109/TPAS.1983.318124.
- [101] “Hawaii About to Crack Ocean Thermal Energy Conversion Roadblocks? | OilPrice.com.” <https://oilprice.com/Alternative-Energy/Renewable-Energy/Hawaii-About-To-Crack-Ocean-Thermal-Energy-Conversion-Roadblocks.html> (accessed Feb. 04, 2023).
- [102] “Ocean thermal energy conversion - Wikipedia.” https://en.wikipedia.org/wiki/Ocean_thermal_energy_conversion (accessed Feb. 04, 2023).
- [103] “Ocean Thermal to begin talks for renewable energy plants in St. Croix, St. Thomas | Local Business | lancasteronline.com.” https://lancasteronline.com/business/local_business/ocean-thermal-to-begin-talks-for-renewable-energy-plants-in/article_e68b41f4-4da4-11e6-8d72-1352558baa6f.html (accessed Feb. 04, 2023).
- [104] “Celebrating Hawaii ocean thermal energy conversion power plant.” <https://techxplore.com/news/2015-08-celebrating-hawaii-ocean-thermal-energy.html> (accessed Feb. 04, 2023).
- [105] “OTEC – Ocean Thermal Energy Conversion | Makai Ocean Engineering.” <https://www.makai.com/renewable-energy/otec/> (accessed Feb. 04, 2023).
- [106] “Okinawa OTEC Demonstration Project.” <http://otecokinawa.com/en/Project/index.html> (accessed Feb. 04, 2023).
- [107] “Salinity Gradient | Tethys.” <https://tethys.pnnl.gov/technology/salinity-gradient> (accessed Feb. 04, 2023).
- [108] “Statkraft osmotic power prototype in Hurum - Wikipedia.” https://en.wikipedia.org/wiki/Statkraft_osmotic_power_prototype_in_Hurum

- (accessed Feb. 04, 2023).
- [109] “Osmotic power - Wikipedia.” https://en.wikipedia.org/wiki/Osmotic_power (accessed Feb. 04, 2023).
- [110] N. Guillou, “Estimating wave energy flux from significant wave height and peak period,” *Renew. Energy*, vol. 155, pp. 1383–1393, 2020, doi: 10.1016/j.renene.2020.03.124.
- [111] J. Mahjoobi and A. Etemad-Shahidi, “An alternative approach for the prediction of significant wave heights based on classification and regression trees,” *Appl. Ocean Res.*, vol. 30, no. 3, pp. 172–177, 2008, doi: 10.1016/j.apor.2008.11.001.
- [112] M. S. Elbisy and A. M. S. Elbisy, “Prediction of significant wave height by artificial neural networks and multiple additive regression trees,” *Ocean Eng.*, vol. 230, p. 109077, Jun. 2021, doi: 10.1016/j.oceaneng.2021.109077.
- [113] S. Gracia, J. Olivito, J. Resano, B. Martin-del-Brio, M. de Alfonso, and E. Álvarez, “Improving accuracy on wave height estimation through machine learning techniques,” *Ocean Eng.*, vol. 236, p. 108699, Sep. 2021, doi: 10.1016/j.oceaneng.2021.108699.
- [114] N. K. kumar, R. Savitha, and A. Al Mamun, “Regional ocean wave height prediction using sequential learning neural networks,” *Ocean Eng.*, vol. 129, pp. 605–612, Jan. 2017, doi: 10.1016/j.oceaneng.2016.10.033.
- [115] K. Günaydin, “The estimation of monthly mean significant wave heights by using artificial neural network and regression methods,” *Ocean Eng.*, vol. 35, no. 14–15, pp. 1406–1415, Oct. 2008, doi: 10.1016/j.oceaneng.2008.07.008.
- [116] S. Shamshirband, A. Mosavi, T. Rabczuk, N. Nabipour, and K. Chau, “Prediction of significant wave height; comparison between nested grid numerical model, and machine learning models of artificial neural networks, extreme learning and support vector machines,” *Eng. Appl. Comput. Fluid Mech.*, vol. 14, no. 1, pp. 805–817, Jan. 2020, doi: 10.1080/19942060.2020.1773932.
- [117] S. C. James, Y. Zhang, and F. O’Donncha, “A machine learning framework to forecast wave conditions,” *Coast. Eng.*, vol. 137, pp. 1–10, Jul. 2018, doi: 10.1016/j.coastaleng.2018.03.004.

- [118] A. Etemad-Shahidi and J. Mahjoobi, "Comparison between M5' model tree and neural networks for prediction of significant wave height in Lake Superior," *Ocean Eng.*, vol. 36, no. 15–16, pp. 1175–1181, Nov. 2009, doi: 10.1016/j.oceaneng.2009.08.008.
- [119] M. C. Deo, A. Jha, A. S. Chaphekar, and K. Ravikant, "Neural networks for wave forecasting," *Ocean Eng.*, vol. 28, no. 7, pp. 889–898, Jul. 2001, doi: 10.1016/S0029-8018(00)00027-5.
- [120] O. Makarynskyy, "Improving wave predictions with artificial neural networks," *Ocean Eng.*, vol. 31, no. 5–6, pp. 709–724, Apr. 2004, doi: 10.1016/j.oceaneng.2003.05.003.
- [121] J. D. Agrawal and M. C. Deo, "On-line wave prediction," *Mar. Struct.*, vol. 15, no. 1, pp. 57–74, Jan. 2002, doi: 10.1016/S0951-8339(01)00014-4.
- [122] A. Zamani, D. Solomatine, A. Azimian, and A. Heemink, "Learning from data for wind-wave forecasting," *Ocean Eng.*, vol. 35, no. 10, pp. 953–962, 2008, doi: 10.1016/j.oceaneng.2008.03.007.
- [123] M. Ali and R. Prasad, "Significant wave height forecasting via an extreme learning machine model integrated with improved complete ensemble empirical mode decomposition," *Renew. Sustain. Energy Rev.*, vol. 104, pp. 281–295, 2019, doi: <https://doi.org/10.1016/j.rser.2019.01.014>.
- [124] J. Berbić, E. Ocvirk, D. Carević, and G. Lončar, "Application of neural networks and support vector machine for significant wave height prediction," *Oceanologia*, vol. 59, no. 3, pp. 331–349, 2017, doi: 10.1016/j.oceano.2017.03.007.
- [125] V. Domala, W. Lee, and T. Kim, "Wave data prediction with optimized machine learning and deep learning techniques," *J. Comput. Des. Eng.*, May 2022, doi: 10.1093/jcde/qwac048.
- [126] A. M. Gómez-Orellana, D. Guijo-Rubio, P. A. Gutiérrez, and C. Hervás-Martínez, "Simultaneous short-term significant wave height and energy flux prediction using zonal multi-task evolutionary artificial neural networks," *Renew. Energy*, vol. 184, pp. 975–989, Jan. 2022, doi: 10.1016/j.renene.2021.11.122.
- [127] C. Jörges, C. Berkenbrink, and B. Stumpe, "Prediction and reconstruction of ocean wave heights based on bathymetric data using LSTM neural networks," *Ocean Eng.*, vol. 232, no. March, 2021, doi: 10.1016/j.oceaneng.2021.109046.

- [128] J. Mahjoobi and E. Adeli Mosabbebi, “Prediction of significant wave height using regressive support vector machines,” *Ocean Eng.*, vol. 36, no. 5, pp. 339–347, Apr. 2009, doi: 10.1016/j.oceaneng.2009.01.001.
- [129] J. C. Fernández, S. Salcedo-Sanz, P. A. Gutiérrez, E. Alexandre, and C. Hervás-Martínez, “Significant wave height and energy flux range forecast with machine learning classifiers,” *Eng. Appl. Artif. Intell.*, vol. 43, pp. 44–53, 2015, doi: 10.1016/j.engappai.2015.03.012.
- [130] D. Guijo-Rubio, A. M. Gómez-Orellana, P. A. Gutiérrez, and C. Hervás-Martínez, “Short- and long-term energy flux prediction using Multi-Task Evolutionary Artificial Neural Networks,” *Ocean Eng.*, vol. 216, no. July, p. 108089, 2020, doi: 10.1016/j.oceaneng.2020.108089.
- [131] “Wave power - U.S. Energy Information Administration (EIA).” .
- [132] “ERDDAP - NDBC Standard Meteorological Buoy Data, 1970-present - Subset.” .
- [133] F. Pedregosa FABIANPEDREGOSA *et al.*, “Scikit-learn: Machine Learning in Python,” *J. Mach. Learn. Res.*, vol. 12, no. 85, pp. 2825–2830, 2011.
- [134] S. R. Massel, “Ocean Surface Waves: Their Physics and Prediction,” vol. 11, Feb. 1996, doi: 10.1142/2285.
- [135] T. Chen and C. Guestrin, “XGBoost: A scalable tree boosting system,” *Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, vol. 13-17-Aug, pp. 785–794, 2016, doi: 10.1145/2939672.2939785.
- [136] B. J. H. Friedman, “1999 REITZ LECTURE,” vol. 29, no. 5, pp. 1189–1232, 2001.
- [137] G. Ke *et al.*, “LightGBM: A highly efficient gradient boosting decision tree,” *Adv. Neural Inf. Process. Syst.*, vol. 2017-Decem, no. Nips, pp. 3147–3155, 2017.
- [138] F. Rosenblatt, “The perceptron: A probabilistic model for information storage and organization in the brain,” *Psychol. Rev.*, vol. 65, no. 6, pp. 386–408, Nov. 1958, doi: 10.1037/H0042519.
- [139] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, “Learning representations by back-propagating errors,” *Nat. 1986 3236088*, vol. 323, no. 6088, pp. 533–536, 1986, doi: 10.1038/323533a0.

- [140] A. M. Fred Agarap, “Deep Learning using Rectified Linear Units (ReLU),” Mar. 2018.
- [141] H. Kaiming, Z. Xiangyu, R. Shaoqing, and S. Jian, “Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification Kaiming,” *Biochem. Biophys. Res. Commun.*, vol. 498, no. 1, pp. 254–261, 2018.
- [142] D. P. Kingma and J. L. Ba, “Adam: A Method for Stochastic Optimization,” *3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc.*, Dec. 2014.
- [143] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, “Dropout: A Simple Way to Prevent Neural Networks from Overfitting,” *J. Mach. Learn. Res.*, vol. 15, no. 56, pp. 1929–1958, 2014.
- [144] S. Ioffe and C. Szegedy, “Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift,” *32nd Int. Conf. Mach. Learn. ICML 2015*, vol. 1, pp. 448–456, Feb. 2015.
- [145] G. Klambauer, T. Unterthiner, A. Mayr, and S. Hochreiter, “Self-Normalizing Neural Networks,” *Adv. Neural Inf. Process. Syst.*, vol. 2017-Decem, pp. 972–981, Jun. 2017.
- [146] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, “Optuna: A Next-generation Hyperparameter Optimization Framework,” *Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, pp. 2623–2631, Jul. 2019, doi: 10.1145/3292500.3330701.