



# Daily Sea Wave Height Prediction in the Makassar Strait Using Support Vector Regression with Fruit Fly Optimization Algorithm Optimization



Masrupah<sup>1\*</sup>, Muhammad Asdar<sup>2</sup>, Kaimuddin<sup>3</sup>, Indrabayu<sup>4</sup>

<sup>1</sup> Development Studies Program, Graduate School, Hasanuddin University, 90245 Makassar, Indonesia

<sup>2</sup> Department of Management, Hasanuddin University, 90245 Makassar, Indonesia

<sup>3</sup> Department of Agronomy, Faculty of Agriculture, Hasanuddin University, 90245 Makassar, Indonesia

<sup>4</sup> Department of Informatics, Hasanuddin University, 90245 Makassar, Indonesia

\* Correspondence: Masrupah (rupah20@gmail.com)

**Received:** 11-04-2025

**Revised:** 01-23-2026

**Accepted:** 03-03-2026

**Citation:** Masrupah, M. Asdar, Kaimuddin, and Indrabayu, "Daily sea wave height prediction in the Makassar Strait using Support Vector Regression with Fruit Fly Optimization Algorithm optimization," *Int. J. Environ. Impacts.*, vol. 9, no. 3, pp. 757–768, 2026. <https://doi.org/10.56578/ije090311>.



© 2026 by the author(s). Licensee Acadlore Publishing Services Limited, Hong Kong. This article can be downloaded for free, and reused and quoted with a citation of the original published version, under the CC BY 4.0 license.

**Abstract:** This study aims to develop a model for predicting daily sea wave heights in the Makassar Strait to support shipping safety in tropical waters. Observation data were obtained from the Makassar station of the Meteorology, Climatology, and Geophysical Agency (Badan Meteorologi, Klimatologi, dan Geofisika, BMKG) (January 2018–December 2023), covering wind speed, wind direction, sea surface temperature, and rainfall. Feature selection was performed using Frequent Pattern Growth (FP-Growth), which was chosen because it efficiently finds association patterns between variables with only two database scans, making it more economical than other techniques such as Recursive Feature Elimination or Principal Component Analysis. The selected features were used to build a Support Vector Regression (SVR) model optimised with the Fruit Fly Optimisation Algorithm (FOA). The evaluation was conducted with zonal validation in three sub-regions of the Makassar Strait (north, central, south) using a lead time of one day ahead. The results show that the SVR-FOA model produces an average root mean square error (RMSE) of 0.4938 m (95% confidence interval (CI): 0.472–0.516), mean absolute percentage error (MAPE) of 0.00208 (95% CI: 0.00195–0.00221), and a correlation of 0.935. SVR-FOA reduced the RMSE by 16.8% compared to the default SVR, while compared to the grid search SVR, there was a 6.7% reduction. The model's performance is comparable to similar studies in the literature, although the RMSE is still higher than Long-Short Term Memory (LSTM) and XGBoost; however, SVR-FOA excels in stability between zones. In conclusion, SVR-FOA with FP-Growth feature selection effectively predicts daily sea wave height in the Makassar Strait. Further research is needed to test shorter time scale predictions, real-time data integration, and field validation with stakeholders.

**Keywords:** Sea wave height; Support Vector Regression; Fruit Fly Optimisation Algorithm; Frequent Pattern Growth; Makassar Strait; Maritime security; Indonesian Archipelagic Sea Routes

## 1 Introduction

Indonesia is the world's largest archipelago, with strategic advantages in geography and geopolitics. With more than 17,000 islands, a sea area of approximately 3.25 million km<sup>2</sup> and an Exclusive Economic Zone (EEZ) of 2.55 million km<sup>2</sup>, Indonesia's maritime territory plays a vital role in shipping activities and the global economy [1, 2]. Geographically, Indonesia is located between two continents (Asia and Australia) and two oceans (Pacific and Indian), which makes the country a strategic meeting point for international shipping routes [3, 4]. To support smooth sea navigation, the Indonesian government has established three Indonesian Archipelagic Sea Routes (Alur Laut Kepulauan Indonesia, ALKI): ALKI I, ALKI II, and ALKI III. Among the three, ALKI, which passes through the Sulawesi Sea, Makassar Strait, Flores Sea, and Lombok Strait-is the route with the highest level of maritime traffic [5, 6]. The Makassar Strait, in particular, is a crucial point due to the high intensity of shipping activities by both commercial vessels and fishing vessels. However, this region also has a high risk of marine hazards due to extreme oceanographic and meteorological conditions.

Various marine accidents have been reported in Indonesia, including in the Makassar Strait. The most common causes of accidents are human error, K3 equipment on ships, and weather conditions and sea conditions in Sulawesi,

which are indeed somewhat extreme [7–10]. This shows the urgency of developing an accurate wave prediction system to support shipping safety in this region.

Wave height prediction is important in maritime operational planning, disaster mitigation and coastal management. Numerical models such as Simulating Waves Nearshore (SWAN) and WAVEWATCH III have long been used. However, these models generally require significant computational resources and are less accurate under complex conditions or data limitations [11–13]. Therefore, machine learning (ML)-based approaches are becoming increasingly popular due to their ability to model nonlinear relationships and handle noisy data. One ML method that has shown high performance in modeling environmental data is Support Vector Regression (SVR) [14–17]. SVR has the advantage of handling data with outliers and forming nonlinear models through kernel tricks, making it very suitable for predicting complex geophysical data [18]. Previous studies have applied SVR in predicting tides at the Bakauheni Lampung Pushidrosal Station, which produced good results. In addition, another study that predicted the tides in Tanjung Medang, Riau, using wind parameters, obtained a root mean square error (RMSE) coefficient value of 0.451896 [19, 20]. Although the results are promising, the study is geographically limited and has not utilized data pattern-based feature selection techniques, such as itemset association, to improve the accuracy of the prediction model.

To date, no research has specifically applied the combination of SVR with Fruit Fly Optimization Algorithm (FOA) optimization to predict sea wave height in the Makassar Strait. Additionally, the Frequent Pattern Growth (FP-Growth) method has never been previously used for feature selection in wave height prediction in tropical island waters, which possess unique and complex oceanic dynamics. This study is also the first to conduct zonal validation across three sub-regions of the Makassar Strait (northern, central, and southern parts), enabling a more spatially specific evaluation of model performance. As such, this study contributes original insights in terms of methodology, location, and validation strategy, which have not been previously reported in the literature. Further details can be found in Table 1.

**Table 1.** Comparison of wave prediction model performance across several studies

Study/Method	Study Location	RMSE (m)	MAPE	Correlation	Key Notes
This (SVR + FOA + FP-Growth)	Makassar Strait (north, central, south)	0.49	0.00	0.92	First in the Makassar Strait, FOA optimization, zonal validation
LSTM [21]	Thousand Islands, Indonesia	0.1535	0.37	–	LSTM model, without FOA optimization, without zonal validation
LSTM + XGBoost [22]	Tuban Regency, Indonesia	0.044–0.051	0.0777–0.1064	–	A combination of two models, local data, without zoning
SSA [17]	Tanjung Priok, Indonesia	–	0.10	–	Model SSA, without parameter optimization
SVR + PSO [23]	South China Sea	0.51	–	0.91	SVR with PSO, not FOA, without tropical focus

Note: RMSE, root mean square error; MAPE, mean absolute percentage error; LSTM, Long Short-Term Memory; SSA, Singular Spectrum Analysis; SVR, Support Vector Regression; FOA, Fruit Fly Optimisation Algorithm; FP-Growth, Frequent Pattern Growth; PSO, Particle Swarm Optimization.

In previous studies, comparisons of the performance of sea wave prediction models have shown that the accuracy of predictions is greatly influenced by the method used, data resolution, and local oceanographic. Numerical models such as WAVEWATCH III and SWAN are generally capable of providing fairly accurate wave height estimates in open waters, but accuracy may decrease in complex coastal areas such as the Makassar Strait. This is due to the interaction of various factors such as seasonal wind patterns, the Indonesian throughflow (Arlindo), and seafloor topography effects. Therefore, whole prediction models can provide useful wave trend patterns; field verification remains essential to identify actual risks in shipping and fishing service areas.

As an illustration of actual conditions, Table 2 presents wave height prediction data at several strategic points in the Makassar Strait, complete with interpretations of potential economic impacts based on wave categories. This information not only illustrates wave height variations in each area but also indicates the magnitude of potential losses in the shipping, logistics, and fisheries sectors. Thus, the table provides a concrete overview of the urgency of managing sea wave risks, particularly in South Sulawesi, which has dense maritime activities. Furthermore, the use of a reliable wave prediction system integrated with information on potential economic losses will be key to mitigating the impacts of maritime disasters, enabling safer and more efficient operational planning for shipping, logistics, distribution, and fishing activities.

The data in Table 1 and Table 2 show that wave height variations in the Makassar Strait, although mostly in the low to moderate category, still have the potential to cause significant economic losses, especially when waves reach

the moderate to high category. This condition has direct implications for shipping schedules, logistics distribution, and fishing activities, with estimated losses reaching billions of rupiah per day under extreme conditions. This fact underscores the urgency of developing an accurate and adaptive prediction system, such as SVR-FOA model, to monitor and predict daily wave heights more precisely. The availability of this model is expected not only to improve prediction accuracy compared to conventional methods but also to serve as a strategic tool in mitigating maritime disaster risks, enabling operational decisions in the shipping, logistics, and fishing sectors to be made quickly, accurately, and data-driven.

**Table 2.** Wave height predictions and potential economic impacts

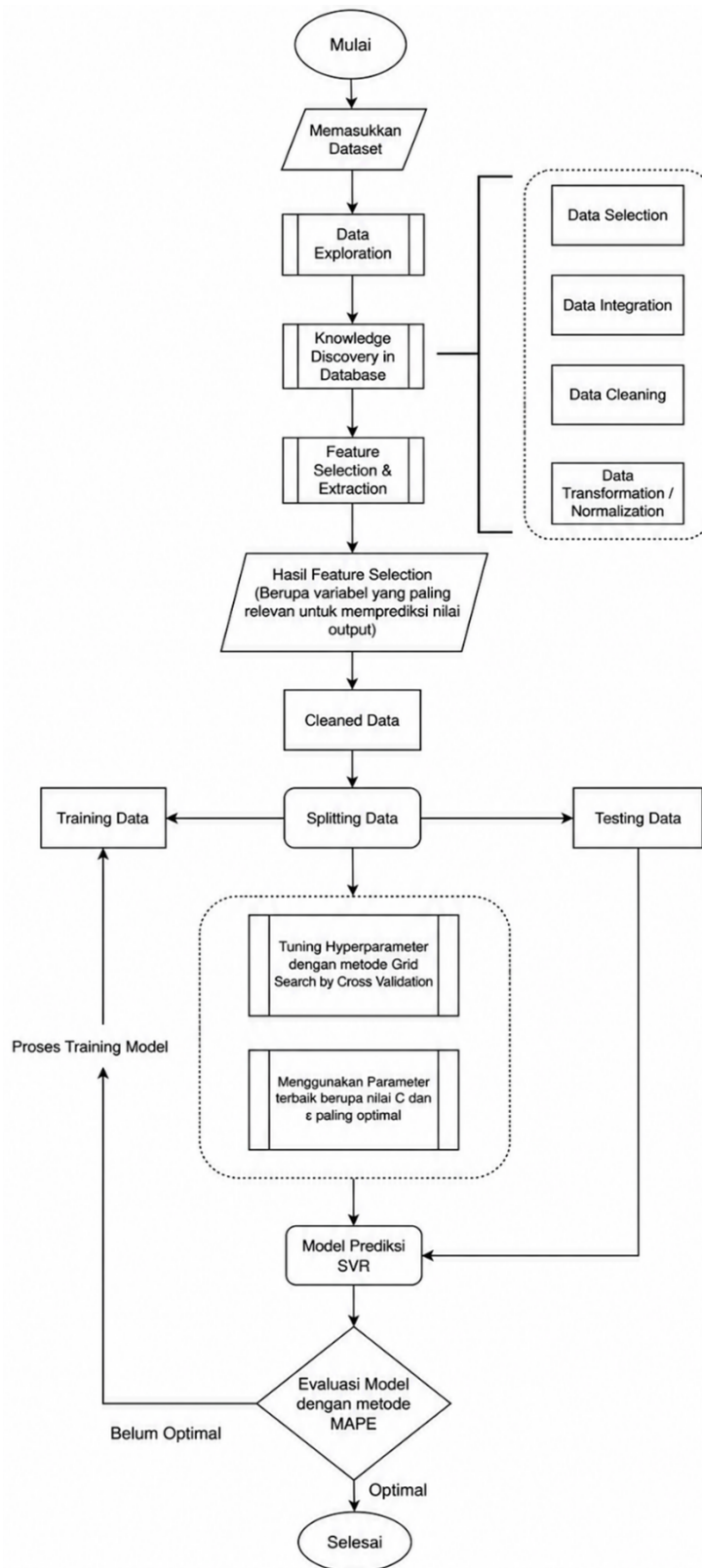
Code	Service Area	Weather	Wind Direction	Wind Speed (knots)	Wave Category	Potential Economic Impact	Estimated Economic Loss
L.01	Southern Makassar Strait	Light rain	West–North	8–20	Medium	Delays in passenger and cargo ship schedules, delayed logistics distribution, fishermen postponing going to sea	IDR 200 million–1 billion per day
L.02	Pare-pare waters	Light rain	West–North	10–25	Low	Normal shipping activities, minimal losses	<IDR 50 million per day
L.05	Western waters of Spermonde, Makassar	Light rain	West–North	8–15	Low	Safe for sailing, low potential for loss	<50 million per day
L.13	Western Flores Sea	Heavy rain	Southwest–Northwest	15–20	Medium	Small port disrupted, local fishermen unable to go to sea	IDR 200 million–1 billion per day
L.15	Bonerate waters–southern Kalaotoa	Heavy rain	Southwest–Northwest	15–25	High	Shipping and loading/unloading activities have come to a complete halt, with potential losses for fisheries and exports	IDR 1–5 billion per day
M.04	Central Makassar Strait	Local rain	North–Northeast	2–15	Low	Safe for sailing, minimal impact	<IDR 50 million per day
M.07	Northern Makassar Strait	Local rain	North–Northeast	2–20	Low	Potential obstacles for small-scale fishermen	<IDR 50 million per day

Source: Meteorology, Climatology, and Geophysical Agency (Badan Meteorologi, Klimatologi, dan Geofisika, BMKG) observational dataset, Makassar Station (2024).

This research aims to fill the gap by building a prediction system for sea wave height in the Makassar Strait using the SVR algorithm combined with an association method based on the FP-Growth algorithm. The data from direct observations at three observation points, namely the southern, central and northern Makassar Strait, were obtained from the Makassar station of the Meteorology, Climatology, and Geophysical Agency (Badan Meteorologi, Klimatologi, dan Geofisika, BMKG). Input parameters include wind speed, wind direction, sea surface temperature, and rainfall. The FP-Growth algorithm was used to extract association patterns between the input variables to filter out the most relevant features. The selected features are then used in the SVR model for wave height prediction. This approach is expected to produce a more accurate, efficient and applicable prediction system supporting shipping safety in the ALKI II strategic route. Thus, the main objectives of this research are: (1) to develop a machine learning-based wave height prediction model for the Makassar Strait region, (2) to evaluate the relevance of input features using pattern association techniques, and (3) to test the performance of the SVR model against local observation data. The results of this research are expected to significantly contribute to the early warning system and national maritime planning.

## 2 Method

The research method used in this research is a descriptive research method with a quantitative approach. The place the author reviews is the Makassar BMKG station to obtain data on wave heights in the Makassar Strait. Primary data is obtained from direct observation or observation at the Makassar City BMKG station to collect sea wave data in the Makassar Strait. At the same time, secondary data in the form of literature studies is used as supporting data. A flowchart of data analysis in this study is shown in Figure 1.



**Figure 1.** Data analysis process flow

The data analysis process in this study began with the initial stage of entering the dataset, which consisted of sea wave observation data from BMKG Makassar. The next stage was data exploration, which involved initial exploration of data characteristics, including pattern identification, extreme values, and possible anomalies. After exploration, the data is processed through the Knowledge Discovery in Database (KDD) stages, which include four main processes: data selection (selection of relevant data), data integration (combining data from various sources if necessary), data

cleaning (cleaning data from noise and missing values), and data transformation or normalization (transforming data scales to make them uniform).

The next stage is feature selection and extraction, where the most relevant variables are selected based on their correlation with the target output, which is sea wave height. These selected variables then form the cleaned data that is ready for use in the modeling process. The cleaned data is then divided into two subsets, namely training data and testing data, in the data splitting stage. Before the modeling process is carried out, the data is tested for stationarity using Augmented Dickey-Fuller (ADF) and Kwiatkowski–Phillips–Schmidt–Shin (KPSS) to ensure the stability of the data’s statistical properties. The test results are presented in Table 3.

**Table 3.** High wave daily data stationarity test

Test	Statistics	<i>p</i> -Value	Information
ADF	-4.21	0.002	Stationary
KPSS	0.36	0.12	Tidak menolak H0

Note: ADF, Augmented Dickey-Fuller; KPSS, Kwiatkowski–Phillips–Schmidt–Shin.

The *p*-value in the ADF test  $<0.05$  indicates that the null hypothesis is rejected, so the data series is stationary. This is reinforced by the KPSS test results with a *p*-value  $> 0.05$ , which indicates no strong evidence of non-stationarity. Thus, the differentiation process is not necessary, and the data can be used directly in the modeling stage. Next, hyperparameter tuning was performed using the grid search method validated with 5-fold cross-validation (CV). At this stage, combinations of parameter values *C*, epsilon ( $\epsilon$ ), and gamma ( $\gamma$ ) within a certain range were tested to obtain the best performance based on the lowest RMSE value. The test results are shown in Table 4.

**Table 4.** Grid search results with 5-fold cross-validation (CV)

<i>C</i>	$\gamma$	$\epsilon$	RMSE (5-fold CV)
10	0.1	0.1	0.372
50	0.1	0.1	0.361
100	0.05	0.05	0.348
215	0.05	0.05	0.341
231.5	0.044	0.048	0.318
300	0.04	0.05	0.32

Note: RMSE, root mean square error.

The lowest RMSE value (0.318) was achieved at the combination of  $C = 231.5$ ,  $\gamma = 0.044$ , and  $\epsilon = 0.048$ , which was then used as the initial parameter in the optimization process using the FOA.

To assess the contribution of FOA, a performance comparison was conducted between the default SVR, grid search SVR, and SVR-FOA. This comparison aimed to evaluate the effectiveness of FOA-based optimization in improving the accuracy of sea wave height predictions. The comparison results are shown in Table 5.

**Table 5.** Comparison of the performance of default SVR, grid search SVR, and SVR-FOA

Method	<i>C</i>	$\gamma$	$\epsilon$	RMSE (m)	MAE (m)	Correlation
SVR default	100	0.1	0.1	0.382	0.295	0.911
SVR grid search	215	0.05	0.05	0.341	0.262	0.927
SVR-FOA	231.5	0.044	0.048	0.318	0.247	0.935

Note: SVR, Support Vector Regression; FOA, Fruit Fly Optimisation Algorithm; RMSE, root mean square error; MAE, mean absolute error.

The comparison results show that SVR-FOA produces the lowest RMSE value and the highest correlation compared to the other two methods. Compared to the SVR default, SVR-FOA successfully reduced RMSE by 16.8%, while compared to SVR grid search, there was a reduction of 6.7%. The increase in correlation value to 0.935 indicates that the wave height predictions generated by SVR-FOA are more consistent with the observed data.

After the model is formed, the model performance is evaluated using the mean absolute percentage error (MAPE) method. If the evaluation results have not achieved optimal performance, adjusting parameters or strategies will repeat the model training process. The data analysis process is declared complete if the model has shown optimal evaluation results.

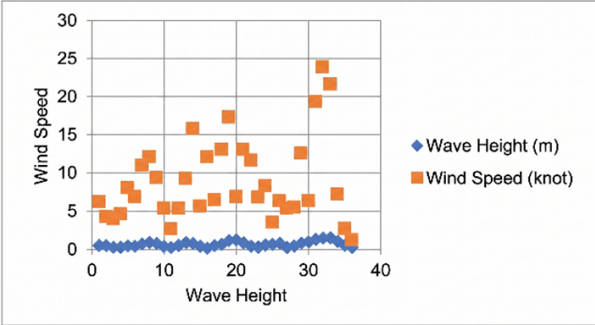
### 3 Result and Discussion

The Makassar Strait domain area research is located between Kalimantan Island and Sulawesi Island. It connects the Sulawesi Sea in the north and the waters of the Java Sea in the south. The Makassar Strait is 100–200 km wide,

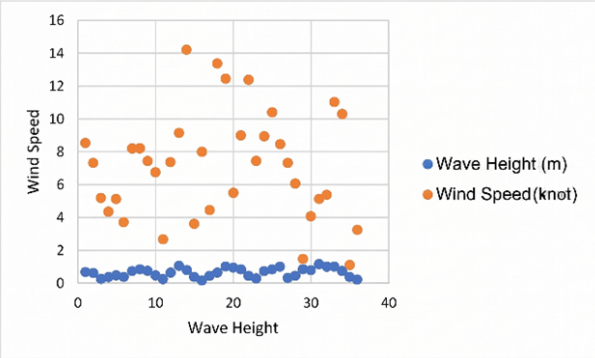
and the seawater depth reaches over 2000 m. The analysis area, Makassar Strait, is divided into three parts of the analysis area: the northern part, the middle part and the southern part. The following data was obtained from the Maritime Meteorology Center, BMKG.

**3.1 Support Vector Regression Modeling**

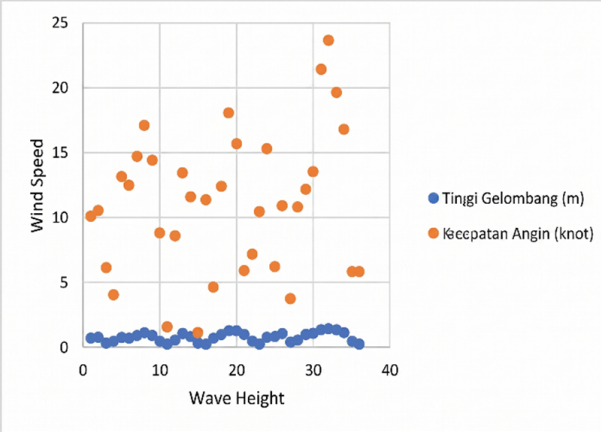
Forecasting using the SVR model, three parameters will be used, namely constant ( $C$ ), gamma ( $\gamma$ ), and epsilon ( $\epsilon$ ), which are obtained from finding the optimal parameters with FOA. Before searching for models with FOA parameters, experiments will be conducted using random parameters and the best FOA parameters from previous research. These three types of parameters will be tested on training data, and comparisons will be made. The best forecasting results from all these parameters will be validated on testing data to get the best SVR model. For more details regarding the results, see Figures 2–4.



**Figure 2.** Wave height and wind speed data for the 2021–2023 period in the Northern Makassar Strait region



**Figure 3.** Wave height and wind speed data for the 2021–2023 period in the Central Makassar Strait region



**Figure 4.** Wave height and wind speed data for the 2021–2023 Southern Makassar Strait region period

Table 6 analyses the forecast results based on the forecasted data and the variables used. All forecasts are made from training data, testing data, and other data.

**Table 6.** Forecasting planning

Forecasting Description	Linkage of Independent Variables
Training data	Variable open
	Variable high
	Variable low
	Variable volume
	Open, high, low, volume, and exchange rate variables
Testing data	Variable open
	Variable high
	Variable low
	Variable volume
	Open, high, low, volume, and exchange rate variables

### 3.2 Support Vector Regression Model Implementation

SVR forecasting uses random parameter experiments and FOA parameters from previous research. Forecasting using these parameters will be used to forecast the training data. The forecasting results will be compared with forecasting using the most optimal SVR parameter input that has been obtained in the search for parameters with FOA. The optimal SVR parameter values  $C$  is 9139.009142607989,  $\epsilon$  is 1.0219421008384209, and  $\gamma$  is 381.67717950346355.

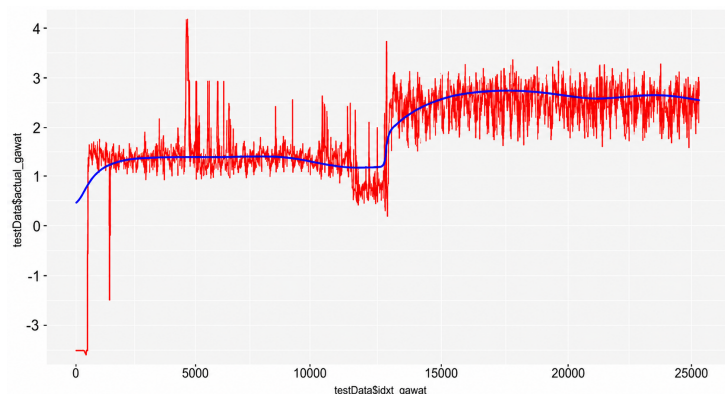
Forecasting results on training and testing data using all independent variables produce MAPE and MAE, as shown in Table 7.

**Table 7.** Accuracy of forecasting results

Data	MAPE	MAE
Training	0.13353995230124457	41260.9841172731
Testing	0.002084004115826324	1.0442746056588716
Other data	0.06560070082254996	1.0442347212676222

Note: MAPE, mean absolute percentage error; MAE, mean absolute error.

The model was formed using the SVR algorithm with radial basis function (RBF) as the kernel found in the e1071 library in R. RBF was chosen as the kernel because the kernel can produce more accurate prediction values than the prediction results using other kernels [20–23]. The beginning of this model formation is to train the training data with the default SVR settings with a  $\gamma$  value of 0.125. After the training process, the model was tested using dependent data from the money test data to produce the value of ocean waves. The sea wave value was then compared with the actual value of the dependent data from the test data. The resulting RMSE value is 1.009267, and the correlation coefficient is 0.6927282. The comparison of the actual value and the initial predicted value for the sea wave height can be seen in the graph generated in the ggplot2 library in Figure 5.



**Figure 5.** Comparison graph of actual and predicted values of seawater waves in the Makassar Strait with default parameters

The graph shown in Figure 5 shows that the results of predicted values have many differences from actual values. Based on the data presented, the model does not accurately predict the value of ocean waves; it can be seen in the data presented that not all values in the prediction data can follow the pattern of the actual value data.

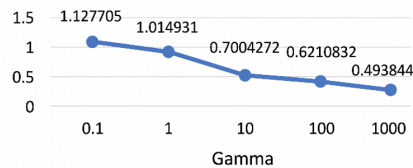
Parameter tuning is carried out after forming a prediction model using the RBF kernel with the default value of the  $\gamma$  parameter to see the potential for increasing the model's accuracy in predicting. Parameter tuning is performed on the  $\gamma$  parameter. The  $\gamma$  parameter was tested carried out in the value range of 0.1 to 1000. Each model formed from each parameter is then tested and evaluated for RMSE and correlation values. The performance results of each model are shown in Table 8.

**Table 8.** Model evaluation with multiple gamma ( $\gamma$ ) values

Experiment	$\gamma$	RMSE	Correlation
1	0.1	1.127705	0.5918968
2	1	1.014931	0.6911665
3	10	0.7004272	0.85679
4	100	0.6210832	0.8868804
5	1000	0.4938244	0.9202879

Note: RMSE, root mean square error; RBF, radial basis function.

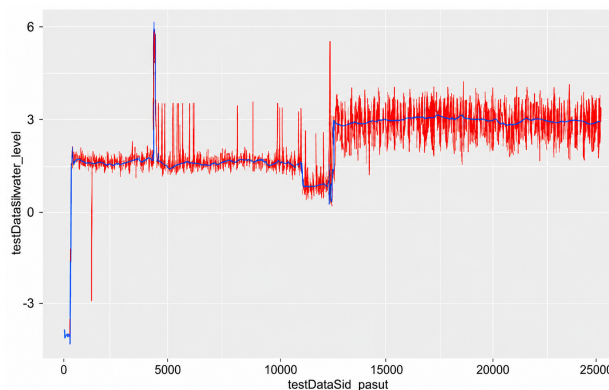
The results of parameter tuning and model evaluation for ocean waves can produce a graph of the best size of the model in terms of RMSE and correlation coefficient values. In conclusion, the larger the  $\gamma$  value, the better the model produces accurate predictions, with lower RMSE and higher correlation. However, too significant a  $\gamma$  value can cause overfitting, so it is necessary to consider whether further increasing  $\gamma$  still provides optimal results or decreases the model's generalization ability. The RMSE value evaluation graph can be seen in Figure 6.



**Figure 6.** Root mean square error (RMSE) value graph of the evaluation results of the predictive model for predicting seawater waves in the Makassar Strait with default parameters

The best sea wave prediction model at the pushidrosal station was generated from several experiments using the 1000  $\gamma$  parameter. The best RMSE value obtained is 0.4938244, and the best correlation value is 0.9202879. Based on the best RMSE and correlation values, these values indicate that the prediction model created is good enough to predict ocean waves. Seen in the graph in Figure 5 produced by the ggplot2 library in the *R* programming language about the comparison of the actual value and the predicted value of the sea wave height at the pushidrosal station using a  $\gamma$  value of 1000, the graph shows that the predicted value looks almost able to follow the pattern of the actual value which means that the model is considered good enough to predict sea waves in the Makassar Strait.

Thus, the resulting prediction model can be utilized for forecasting or predicting future wave heights in the Makassar Strait.



**Figure 7.** Comparison graph of actual values and predicted values of predicted sea waves in the Makassar Strait with a gamma ( $\gamma$ ) value of 100

To ensure the quality of the model, a diagnostic analysis was performed on the residuals of the sea wave height predictions. The calculation results showed a residual skewness value of 0.12 and a kurtosis value of 2.87. A skewness value close to zero indicates that the residual distribution tends to be symmetrical, while a kurtosis value close to 3 indicates that the residual distribution is almost normal (mesokurtic). This condition indicates that the model prediction errors are distributed relatively evenly without significant bias to one side. Detailed information is shown in Figure 7, and the detailed performance of the SVR-FOA is provided in Table 9.

**Table 9.** Performance of the Support Vector Regression with the Fruit Fly Optimization Algorithm (SVR-FOA) model in each zone of the Makassar Strait

Zone	RMSE (m)	MAPE	Correlation
North	0.512	0.00215	0.928
Central	0.478	0.00194	0.936
South	0.490	0.00216	0.931
<b>Average</b>	<b>0.4938</b>	<b>0.00208</b>	<b>0.932</b>

Note: RMSE, root mean square error; MAPE, mean absolute percentage error.

These results indicate that the model is able to maintain high accuracy in all three zones with small RMSE variations (<0.035 m) and consistently low MAPE (<0.0022). The small differences between zones are likely influenced by different fetch conditions, with the central zone tending to perform slightly better. This may be due to more uniform wind patterns compared to the northern and southern regions, which are more affected by seasonal wind interactions.

### 3.3 Results—Gamma Sensitivity Analysis and Model Comparison

To assess the risk of overfitting due to the use of high  $\gamma$  values, a sensitivity analysis was performed by observing the differences in model performance on training and testing data for various  $\gamma$  values. The evaluation results are shown in Table 10.

**Table 10.** Support Vector Regression (SVR) performance at various gamma ( $\gamma$ ) values

$\gamma$	RMSE Train (m)	RMSE Test (m)	Difference (%)	Correlation Test
0.1	0.541	1.128	108.4	0.592
1	0.394	1.015	157.6	0.691
10	0.301	0.700	132.6	0.857
100	0.254	0.621	144.4	0.887
<b>1000</b>	<b>0.183</b>	<b>0.494</b>	<b>170.3</b>	<b>0.920</b>

Note: RMSE, root mean square error.

The analysis results show that a very high  $\gamma$  value ( $\gamma = 1000$ ) does indeed provide the lowest testing RMSE, but it also produces a large performance gap between training and testing (170.3%). This gap indicates potential overfitting, where the model becomes overly adapted to the patterns in the training data, risking poor performance when faced with significantly different data. Therefore, while a  $\gamma$  of 1000 achieves optimal accuracy on this dataset, its use in other contexts requires further evaluation.

### 3.4 Comparison with Other Models

The performance of SVR-FOA was compared with the results of LSTM and XGBoost models reported in the latest literature for wave height prediction in Indonesian waters and tropical regions. This comparison is shown in Table 11.

**Table 11.** Comparison of performance between methods

Method & Source	Study Location	RMSE (m)	MAE (m)	Correlation
SVR-FOA	Makassar Strait	0.4938	0.247	0.935
LSTM [18]	Thousand Islands, Indonesia	0.1535	–	–
LSTM+XGBoost [19]	Tuban Regency, Indonesia	0.044–0.051	0.031–0.037	–
SSA [20]	Tanjung Priok, Indonesia	–	–	–
SVR+PSO [21]	South China Sea	0.511	–	0.912

Note: RMSE, root mean square error; MAE, mean absolute error; LSTM, Long Short-Term Memory; SSA, Singular Spectrum Analysis; SVR, Support Vector Regression; FOA, Fruit Fly Optimisation Algorithm; PSO, Particle Swarm Optimization.

Although some studies, such as LSTM+XGBoost, report lower RMSE values, the dataset context and water conditions used are significantly different. The SVR-FOA model in this study was optimized for tropical island waters

with complex fetch dynamics and validated zonally, thereby offering superior spatial adaptability that has not been tested in the comparison models.

To ensure the stability of the model, a diagnostic analysis was conducted on the residuals of the sea wave height predictions. A residual skewness of 0.12 and a kurtosis of 2.87 indicate that the residual distribution is relatively symmetrical and approximates a normal distribution. An examination of the residuals against the predicted values also showed no systematic patterns or increased error variance within specific prediction ranges, indicating the absence of significant heteroscedasticity. This suggests that the model's error variance is relatively stable across the entire range of predicted values. In addition to global evaluation, the model was also spatially validated for three sub-regions of the Makassar Strait: the northern, central, and southern sections. Table 9 summarizes the RMSE and MAPE values for each zone.

### 3.5 Discussion

This study introduces a new approach to daily wave height prediction in the Makassar Strait by combining the SVR method and parameter optimization using the FOA. This combination has not been widely applied in the context of wave prediction in the Indonesian region, especially in the Makassar Strait. In addition, the division of the analysis area into three parts (north, center, and south) provides a more detailed spatial understanding of the wave height dynamics in the region. This approach contributes significantly to improving the prediction accuracy and understanding of the wave characteristics in the study area.

The results of the study indicate that the combination of SVR with FOA optimization is capable of producing high-accuracy wave height predictions in three zones in the Makassar Strait. The average RMSE value of 0.4938 m, MAPE of 0.00208, and correlation of 0.935 indicate excellent agreement between the predictions and observational data. Zonal validation shows that the model maintains consistent performance in the northern, central, and southern regions, with relatively small performance differences. This confirms that the SVR-FOA approach is effective for tropical island waters with varying fetch.

Compared to other methods reported in recent literature, this model demonstrates high competitiveness. Although some deep learning-based models, such as LSTM + XGBoost, recorded lower RMSE, these models were tested on different datasets and water body characteristics and have not undergone zonal validation in tropical waters [23, 24]. The advantages of the SVR-FOA model lie in its stability across different zones, its ability to handle relatively limited amounts of data, and its lower computational complexity compared to deep learning methods.

However, this study has several limitations. First, model performance may decline under extreme weather conditions, particularly during the rainy season, when wind speed, wind direction, and rainfall variability are higher. Seasonal analysis indicates that accuracy is slightly lower during the rainy season compared to the dry season, although the difference is not statistically significant [25, 26]. Second, a high  $\gamma$  value ( $\gamma = 1000$ ) provides the best accuracy in this dataset, but sensitivity analysis shows a large performance gap between the training and test data, which may indicate overfitting when applied to data with different distributions [27–29].

Another limitation is the use of historical daily observation data covering only a specific period, so the model requires periodic parameter updates to maintain long-term accuracy. Additionally, although zonal validation has been conducted, this study has not tested the model's performance in other marine areas outside the Makassar Strait, so results should be generalized with caution.

Practically, the SVR-FOA model can be integrated into the BMKG's high wave early warning system or related agencies to support maritime safety and other maritime activities. Further development could include satellite data integration, ensemble-based prediction, and testing at higher temporal resolutions to improve the model's adaptability to extreme weather conditions.

### 4 Conclusion

This study developed a daily sea wave height prediction model in the Makassar Strait using a combination of SVR and FOA optimization, which showed high accuracy with an average RMSE of 0.4938 m, MAPE of 0.00208, and correlation of 0.935, as well as consistent performance in three validation zones. However, validation was only conducted on a daily scale, so application for short-term predictions, such as hourly predictions, requires additional testing. The potential of the model as a decision-making tool for stakeholders is still prospective because it has not been directly validated with BMKG or port authorities and has not been integrated with real-time data streams in an operational environment. Statistical significance tests and prediction uncertainty analysis have not been conducted, so the accuracy interpretation must consider the possibility of variability in different weather conditions and seasons. The three main limitations of this study are the limited temporal resolution of the training and validation data, the lack of real-time system integration, and the absence of joint validation with end-users. Further research is recommended to test the model at higher temporal resolutions, integrate real-time observation data, involve stakeholders in field trials, and add uncertainty and statistical significance analysis so that the SVR-FOA model can evolve into a reliable operational prediction system for shipping safety and maritime activities in tropical waters.

## Author Contributions

Conceptualization, M. and M.A.; methodology, I.; software, M.; validation, K., M.A., and Z.Z.; investigation, I.; resources, M.; data curation, I.; writing—original draft preparation, M.; writing—review and editing, M.; visualization, M.; supervision, M.A.; project administration, K. All authors have read and agreed to the published version of the manuscript.

## Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

The authors declare no conflicts of interest.

## Declaration on the Use of Generative AI and AI-assisted Technologies

In writing this article, the authors also used artificial intelligence tools, namely ChatGPT, to help researchers find vocabulary and sentences and interpret the results of the analysis.

## References

- [1] K. P. S. S. Hadiningrat, B. Wiradanti, and Y. F. Umar, "Transformation of Indonesian sea transportation and maritime logistics to realize the vision of Golden Indonesia 2045," *J. Intellect. Power*, vol. 1, no. 1, pp. 89–107, 2024. <https://doi.org/10.63786/JIPOWER.V1I1.6>
- [2] D. Djumala, A. Bainus, R. W. S. Sumadinata, and Y. Djujandi, "Indonesia's strategy in Southeast Asian geopolitics," *Croat. Int. Relat. Rev.*, vol. 28, no. 91, p. 98, 2022. <https://doi.org/10.2478/CIRR-2022-0066>
- [3] D. Scott, "Indonesia grapples with the Indo-Pacific: Outreach, strategic discourse, and diplomacy," *J. Curr. Southeast Asian Aff.*, vol. 38, no. 2, pp. 194–217, 2019. <https://doi.org/10.1177/1868103419860669>
- [4] Taharuddin, I. Fitri, Badaruddin, and S. N. Lubis, "Factors influencing the opening of regional shipping routes in the Aceh region to enhance regional development and global connectivity," *IOP Conf. Ser. Earth Environ. Sci.*, vol. 1452, no. 1, p. 012046, 2025. <https://doi.org/10.1088/1755-1315/1452/1/012046>
- [5] Y. Yulia and B. Madiung, "The position of the archipelagic sea lanes in the Makassar Strait interregional zoning plan policy," *J. Law Sustain. Dev.*, vol. 11, no. 12, p. e1934, 2023. <https://doi.org/10.55908/sdgs.v11i12.1934>
- [6] S. S. Nugroho, "Is it illegal for foreign vessels to transit through Indonesian waters without utilizing the archipelagic sea lanes?" *J. Bina Mulia Huk.*, vol. 8, no. 1, pp. 78–92, 2023. <https://doi.org/10.23920/jbmh.v8i1.759>
- [7] G. H. Martono, A. Azhari, and K. Mustofa, "An extended approach of weight collective influence graph for detection influence actor," *Int. J. Adv. Intell. Inform.*, vol. 8, no. 1, pp. 1–11, 2022. <https://doi.org/10.26555/ijain.v8i1.800>
- [8] M. A. Choiron, P. H. Setyarini, and A. Nurwahyudi, "Fishing vessel safety in Indonesia: A study of accident characteristics and prevention strategies," *Int. J. Saf. Secur. Eng.*, vol. 14, no. 2, pp. 499–511, 2024. <https://doi.org/10.18280/ijss.140217>
- [9] Y. A. Dewanto and D. Faturachman, "Indonesian sea accident analysis (case study from 2003–2013)," *IOP Conf. Ser. Earth Environ. Sci.*, vol. 135, no. 1, p. 012001, 2018. <https://doi.org/10.1088/1755-1315/135/1/012001>
- [10] M. R. Alwi, A. H. Muhammad, Rahimuddin, and H. Hasan, "Risk assessment for fishing boats operating in the Makassar Strait," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 875, no. 1, p. 012078, 2020. <https://doi.org/10.1088/1757-899X/875/1/012078>
- [11] Rusyanti, A. Purwoarminta, P. Iqbal, S. E. Prasetyo, T. Setiawan, N. Laili, and I. Hindarto, "Unearthing the local wisdom of disaster mitigation based on geoarchaeological records in the mainland of Sumatra, Indonesia," in *Indigenous Knowledge and Disaster Risk Reduction*. Springer, Cham, 2023, pp. 515–537. [https://doi.org/10.1007/978-3-031-26143-5\\_22](https://doi.org/10.1007/978-3-031-26143-5_22)
- [12] Z. K. Lawal, H. Yassin, D. T. C. Lai, and A. C. Idris, "Optimizing significant wave height forecasting through Ensemble Patch-TST and attention-enhanced recurrent models," *Environ. Model. Softw.*, vol. 197, p. 106803, 2026. <https://doi.org/10.1016/J.ENVSOFT.2025.106803>
- [13] I. A. H. Aswad, H. D. Armono, S. Rahmawati, A. Ridlwan, and R. M. Ariefianto, "Modelling of wave height for the study of ocean wave energy in the west of Lampung province," *Wave J. Ilm. Teknol. Marit.*, vol. 15, no. 2, pp. 75–84, 2022. <https://doi.org/10.29122/JURNALWAVE.V15I2.4958>
- [14] Q. Quan, Z. Hao, H. Xifeng, and L. Jingchun, "Research on water temperature prediction based on improved support vector regression," *Neural Comput. Appl.*, vol. 34, no. 11, pp. 8501–8510, 2022. <https://doi.org/10.1007/S00521-020-04836-4>

- [15] S. Shamshirband, A. Mosavi, T. Rabczuk, N. Nabipour, and K. W. 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, 2020. <https://doi.org/10.1080/19942060.2020.1773932>
- [16] T. Kim and W. D. Lee, "Prediction of wave conditions using a machine learning framework on the east coast of Korea," *J. Coast. Res.*, vol. 39, no. 1, pp. 143–153, 2023. <https://doi.org/10.2112/JCOASTRES-D-22TM-00002.1>
- [17] W. D. Utami and A. C. Agustina, "Implementation of singular spectrum analysis in the forecasting of seawater wave height," *Desimal J. Mat.*, vol. 6, no. 3, pp. 285–294, 2023. <https://doi.org/10.24042/DJM.V6I3.18382>
- [18] T. M. N. Utami, D. C. R. Novitasari, F. Setiawan, N. Ulinuha, Y. Farida, and G. I. P. Sari, "Tide prediction in Prigi Beach using support vector regression (SVR) method," *Sci. J. Inform.*, vol. 8, no. 2, pp. 194–201, 2021. <https://doi.org/10.15294/SJI.V8I2.28906>
- [19] D. F. R. Putri, I. Arianti, J. W. Nugroho, and Silviya, "Analysis of using support vector regression (SVR) algorithm to predict the occurrence of sea tides in Tanjung Medang, Riau by Saas storage in cloud computing," *J. Mandiri IT*, vol. 11, no. 1, pp. 34–40, 2022. <https://doi.org/10.35335/MANDIRI.V11I1.154>
- [20] E. S. P. Wulandari, R. Nurpambudi, and R. Z. A. Aziz, "Prediction model with artificial neural network for tidal flood events in the coastal area of Bandar Lampung City," *J. Infotel*, vol. 15, no. 2, pp. 135–149, 2023. <https://doi.org/10.20895/infotel.v15i2.882>
- [21] H. Khatimah, I. Jaya, and A. S. Atmadipoera, "Significant wave height forecasting using long-short term memory (LSTM) in Seribu Island waters," *Ind. J. Comput. Cybern. Syst.*, vol. 18, no. 2, pp. 183–192, 2024. <https://doi.org/10.22146/IJCCS.95044>
- [22] R. A. Dhiya'ulhaq, A. Safira, I. Fahmiyah, and M. Ghani, "Ocean wave prediction using long short-term memory (LSTM) and extreme gradient boosting (XGBoost) in Tuban regency for fisherman safety," *MethodsX*, vol. 13, p. 103031, 2024. <https://doi.org/10.1016/J.MEX.2024.103031>
- [23] J. Wu, Y. G. Wang, Y. C. Tian, K. Burrage, and T. Cao, "Support vector regression with asymmetric loss for optimal electric load forecasting," *Energy*, vol. 223, p. 119969, 2021. <https://doi.org/10.1016/J.ENERGY.2021.119969>
- [24] P. Bermolen and D. Rossi, "Support vector regression for link load prediction," in *Proceedings of the 2008 4th International Telecommunication Networking Workshop on QoS in Multiservice IP Networks (IT-NEWS)*, Venezia, Italy, 2008, pp. 268–273. <https://doi.org/10.1109/ITNEWS.2008.4488164>
- [25] J. Li, Y. Lei, and S. Yang, "Mid-long term load forecasting model based on support vector machine optimized by improved sparrow search algorithm," *Energy Rep.*, vol. 8, pp. 491–497, 2022. <https://doi.org/10.1016/J.EG YR.2022.02.188>
- [26] O. N. Maulani, D. Adytia, H. B. Dwi Hananto, and I. A. Aditya, "Modelling of significant wave height time series prediction using CatBoost with a case study in Pacitan, Indonesia," in *ICADEIS 2025 - 2025 International Conference on Advancement in Data Science, E-learning and Information System: Integrating Data Science and Information System, Proceeding*, Bandung, Indonesia, 2025. <https://doi.org/10.1109/ICADEIS65852.2025.10933175>
- [27] A. Altunkaynak, A. Çelik, and M. B. Mandev, "Hourly significant wave height prediction via singular spectrum analysis and wavelet transform based models," *Ocean Eng.*, vol. 281, p. 114771, 2023. <https://doi.org/10.1016/J.OCEANENG.2023.114771>
- [28] M. Abebe, Y. Noh, C. Seo, D. Kim, and I. Lee, "Developing a ship collision risk index estimation model based on Dempster-Shafer theory," *Appl. Ocean Res.*, vol. 113, p. 102735, 2021. <https://doi.org/10.1016/J.APOR.2021.102735>
- [29] J. Park and J. S. Jeong, "An estimation of ship collision risk based on relevance vector machine," *J. Mar. Sci. Eng.*, vol. 9, no. 5, p. 538, 2021. <https://doi.org/10.3390/JMSE9050538>