Journal of Applied Data Sciences
Vol. 5, No. 1, January 2024, pp. 84-99


# Comparative Analysis of SVM and RF Algorithms for Tsunami Prediction: A Performance Evaluation Study

Husni Teja Sukmana[1,*], Yusuf Durachman[2], Amri[3], Supardi[4]

[1,2] *State Islamic University Syarif Hidayatullah, Jakarta, Indonesia*

[3,4] *Atma Luhur Institute of Science and Business, Pangkal Pinang, Indonesia*

(Received: November 23, 2023; Revised: December 8, 2023; Accepted: January 14, 2024; Available online: January 29, 2024)

**Abstract**

This study explores the use of machine learning algorithms, specifically SVM and RF, for predicting tsunamis, a crucial aspect of disaster management. The research utilized earthquake data from 2001 to 2023, evaluating these models based on accuracy, precision, recall, F1-score, and ROC AUC, emphasizing features like magnitude, depth, and alert levels. The SVM model demonstrated an accuracy of 65.61%, precision of 70.59%, recall of 19.67%, F1-score of 30.77%, and ROC AUC of 62.15%. In comparison, the RF model showed an accuracy of 61.15%, precision of 50.00%, higher recall of 36.07%, F1-score of 41.90%, and ROC AUC of 63.84%. These results highlight the distinct strengths of each model: SVM's precision makes it suitable for minimizing false positives, while RF's higher recall indicates its effectiveness in detecting actual tsunamis. The findings underscore the significance of selecting the appropriate model for tsunami prediction based on specific disaster management needs and the inherent trade-offs in model selection. The research concludes that SVM and RF models provide valuable yet distinct contributions to tsunami prediction. Their application should be customized to disaster management requirements, balancing accuracy and operational efficiency. This study contributes to disaster risk management and opens avenues for further research in enhancing the accuracy and reliability of machine learning in natural disaster prediction and response systems.

*Keywords:* Machine Learning, Tsunami Prediction, Support Vector Machine, Random Forest, Disaster Management

## 1. Introduction

Tsunamis, while infrequent, are among the most catastrophic natural disasters, often striking with little warning and leaving a trail of devastation. The 2011 Tohoku earthquake and subsequent tsunami, which resulted in over 15,000 deaths and significant nuclear incidents, serves as a stark reminder of the urgent need for improved predictive models. In a world where technology and data science have advanced rapidly, the potential to enhance tsunami prediction through sophisticated algorithms presents a critical opportunity in disaster preparedness. This paper explores the application of two such algorithms, SVM and RF, in tsunami prediction. By leveraging the strengths of these machine learning techniques, we aim to contribute to the development of more accurate and reliable models, potentially saving lives and minimizing the impact of future tsunamis. In this context, the comparative analysis of SVM and RF is a scientific endeavor and a step towards more effective disaster risk management.

Predictive modeling in natural disasters is crucial for effective disaster management and risk reduction. Various studies have emphasized the significance of spatial modeling and data-based methods for predicting natural and engineering disasters [1]. These methods are essential for understanding the influences of disasters on the environment and for establishing effective predictive models.

Furthermore, the limitations of physics-based models in predicting natural disasters have been acknowledged, as these models may have prediction errors due to the complexity and dynamic nature of natural disaster behavior [2]. This underscores the importance of integrating data-driven approaches in predictive modeling. Research and development efforts are ongoing to develop and improve simulation and prediction modeling systems in the field of natural disasters, indicating the continuous pursuit of more effective predictive models [3].

***Corresponding author: Husni Teja Sukmana (husniteja@uinjkt.ac.id)**
**DOI: https://doi.org/10.47738/jads.v5i1.159**

A tsunami is a series of ocean waves with very long wavelengths (typically hundreds of kilometers) caused by large-scale ocean disturbances, such as earthquakes, volcanic eruptions, and meteorite impacts [4]. These disturbances displace a substantial volume of water, generating waves that can travel across entire ocean basins at high speeds. When these waves approach shallow coastal areas, their energy becomes compressed, dramatically increasing wave height and causing devastating flooding and destruction. Tsunamis can have catastrophic effects on coastal communities, resulting in the loss of human lives and properties.

Tsunami prediction and disaster preparedness are critical components of mitigating the impact of tsunamis. Predictive modeling plays a crucial role in anticipating and preparing for such disasters. [5] conducted numerical experiments on tsunami flow depth prediction for clustered areas using regression and machine learning models, emphasizing the importance of prediction methods with robust and light computational loads for preparing for unforeseen situations during large-scale earthquakes and tsunami disasters. [6] demonstrated the potential of machine learning-based models as surrogates for conventional physics-based models to predict near-field tsunami inundations in real-time, highlighting the advancement in predictive modeling techniques for tsunami events. Collectively, these studies underscore the significance of predictive modeling in tsunami prediction in reducing the vulnerability of communities to tsunami disasters.

Applying machine learning algorithms in disaster prediction has gained significant attention in recent research. [7] highlighted the potential of predictive data analytics, enabled by machine learning algorithms, for assessing the long-term impacts of disasters, emphasizing the role of technological advancements in predicting the effects of disasters over time. Furthermore, [8] compared the performance of machine learning algorithms, including Artificial Neural Network (ANN), SVM, and Decision Tree (DT), in flood prediction models, indicating the diverse applications of machine learning techniques in disaster prediction.

Accurately predicting tsunamis is crucial for mitigating the devastating impact of these natural disasters. Tsunami prediction models, particularly those utilizing machine learning algorithms, are crucial in providing timely warnings and enabling effective disaster preparedness and response efforts. This highlights the practical significance of efficient prediction models in ensuring timely and effective disaster management.

After discussing the broader context of predictive modeling in natural disasters and the specific challenges associated with tsunami prediction, it becomes evident that innovative approaches are required to enhance prediction accuracy and reliability. In this vein, machine learning algorithms have emerged as powerful tools, offering significant potential in predicting such catastrophic events. Among these algorithms, SVM and RF have shown considerable promise in various domains of environmental prediction and disaster management. However, their specific utility and comparative effectiveness in tsunami prediction still need to be explored. This gap in research forms the crux of our study. Our research specifically focuses on these two algorithms, delving into a comparative analysis of SVM and RF for predicting tsunamis. This comparison is crucial for understanding each algorithm's strengths and weaknesses within the unique challenges posed by tsunami prediction and is pivotal in advancing the field of disaster predictive modeling. Thus, we aim to bridge the gap in current research by thoroughly evaluating these models, focusing on their predictive performance in the context of tsunamis, a critical aspect of disaster preparedness and risk reduction.

Machine learning models have been increasingly utilized for their predictive capabilities in natural disaster prediction. The comparison between SVM and RF models is fascinating in tsunami prediction. Several studies have explored the performance of these models in various domains, shedding light on their strengths and weaknesses.

[9] demonstrated the effectiveness of a stacking algorithm, which included SVM and RF, in predicting forest canopy height. This suggests that SVM and RF have been successfully applied in environmental prediction tasks, showcasing their potential for natural disaster prediction.

[10] compared machine learning models for predicting forest fires and found that RF exhibited high accuracy and precision. This highlights the robustness of RF in environmental risk assessment, which could be relevant to the predictive modeling of tsunamis.

In conclusion, the comparison of SVM and RF in various domains demonstrates the versatility and effectiveness of both models in predictive tasks. While RF has shown promise in environmental prediction and risk assessment, SVM has also been successful in diverse applications. The findings from these studies provide valuable insights for the comparative analysis of SVM and RF in the context of tsunami prediction, offering a foundation for further exploration and experimentation.

The application of SVM and RF algorithms in disaster prediction has been investigated in various studies. [11] demonstrated the prediction effectiveness of the SVM algorithm in assessing sensitivity to coal ash blasts for different degrees of deterioration, highlighting the potential of SVM in predicting and assessing disaster-related events. Additionally, [12] showcased the use of RF in flood risk assessment, emphasizing its relevance in predicting natural disasters such as floods.

In pursuit of advancing the field of tsunami prediction, this research sets forth two interconnected objectives. Firstly, it seeks to rigorously evaluate the predictive performance of SVM and RF algorithms within the intricate context of tsunami forecasting. This evaluation entails a comprehensive analysis of both algorithms' capacity to accurately predict tsunami events, considering their proficiency in recognizing seismic data patterns and their ability to deliver timely and reliable predictions. Furthermore, the research endeavors to compare the strengths and weaknesses of SVM and RF algorithms in tsunami prediction. This comparison focuses on three critical dimensions: accuracy, computational efficiency, and scalability. By scrutinizing these aspects, the study aims to provide valuable insights into which algorithm exhibits superior predictive performance while shedding light on the specific attributes that make one algorithm more adept than the other in addressing the unique challenges posed by tsunami prediction.

As we stand on the cusp of significant advancements in predictive modeling, this study aims to contribute to a pivotal shift in disaster management strategies, particularly in the context of tsunami prediction. The comparative analysis of SVM and RF algorithms in this research is not just a theoretical exercise; it has profound practical implications. Our findings could enhance the accuracy and reliability of tsunami prediction models, which are crucial for early warning systems. Improved prediction models can lead to better preparedness and quicker response times, significantly reducing the loss of life and property damage in coastal communities vulnerable to tsunamis.

Furthermore, the insights gained from this study could inform the development of more sophisticated, data-driven approaches in the broader field of disaster management. By understanding the strengths and weaknesses of SVM and RF algorithms in predicting natural disasters, we can better tailor predictive models to suit these events' complex and dynamic nature. This could lead to more efficient allocation of resources and more effective evacuation plans, ultimately contributing to the resilience and safety of at-risk populations.

## 2. Literature Review

In selecting appropriate machine learning algorithms for tsunami prediction, it is essential to consider the characteristics of the data typically encountered in this domain and the specific requirements of disaster prediction models. SVM and RF algorithms were chosen for this comparative analysis due to their distinct yet complementary strengths in handling such challenges.

SVM is renowned for its effectiveness in classification tasks, particularly in high-dimensional data situations. Its capability to model complex, non-linear relationships makes it a strong candidate for tsunami prediction, where the data often involve intricate patterns and many influencing factors. The robustness of SVM in dealing with overfitting, even in cases of limited training data, is particularly advantageous, considering the relative rarity of tsunami events. Additionally, SVM's flexibility in kernel choice allows for fine-tuning the model to accommodate the specific nature of seismic data associated with tsunamis.

On the other hand, RF is chosen for its proficiency in handling large datasets and its inherent feature selection capability, which is crucial in analyzing the vast and diverse datasets typical in natural disaster prediction. RF's ensemble approach, aggregating multiple decision trees, provides high accuracy and stability, reducing the risk of overfitting. This is particularly beneficial in tsunami prediction, where the model must be robust against the variability and noise inherent in environmental data. Moreover, RF's ability to provide insights into feature importance is invaluable in understanding which factors most significantly influence tsunami genesis and propagation.

 

Together, these algorithms offer a comprehensive approach to the complex task of tsunami prediction. By comparing their performance, we aim to determine which algorithm is more effective and gain insights into their strengths and limitations in disaster predictive modeling. This comparison will contribute to developing more accurate and reliable tsunami prediction models, ultimately aiding in better disaster preparedness and risk reduction efforts.

### 2.1. Description of SVM Algorithm

SVM is a widely used algorithm in various fields due to its effectiveness in classification, regression, and prediction tasks. It has been applied in real-time crash risk evaluation [13], power prediction for PV power smoothing [14], and fire detection and recognition optimization [15]. Additionally, it has been found to outperform other algorithms in terms of accuracy and prediction error rates in various applications [16].

SVM operates by searching for optimal classification hyperplanes in both linearly separable and inseparable cases, utilizing structured risk minimization to construct the optimal hyperplane in the attribute space, ensuring the classifier achieves the global optimum and meets the expected risk at a certain upper bound [15].

Kernel functions are crucial in the effectiveness of SVM algorithms. These functions enable SVM to handle non-linearly separable data by mapping it to a higher-dimensional space, where linear separation becomes feasible [17].

The application of SVM in tsunami prediction has been a subject of interest in recent research. [18] developed a real-time tsunami prediction system utilizing an ocean floor network for local regions, providing outputs such as tsunami arrival time, height, inundation area, and depth. Furthermore, [6] focused on machine learning-based tsunami inundation prediction derived from offshore observations, highlighting the potential of machine learning in enhancing tsunami prediction capabilities.

### 2.2. Description of RF Algorithm

RF is a popular machine learning algorithm widely used in various domains due to its robustness and effectiveness. It is an ensemble learning method based on decision trees, where multiple decision trees are trained independently and combined to make predictions [19]. The algorithm has been applied in diverse areas such as security risk assessment [20], cyber-attack prediction [21], and geoscience data analysis [22].

Moreover, various studies have compared the RF algorithm with other machine learning algorithms, demonstrating its superior performance in different applications. For example, in the context of predicting the yield of non-breakeven financial products, the practical results of five machine learning algorithms were compared, and the RF was found to have the best prediction effect [23].

The algorithm has several hyperparameters that need to be set by the user, such as the number of observations drawn randomly for each tree, the number of variables drawn randomly for each split, and the minimum number of samples that a node must contain [24].

The application of RF in tsunami prediction has gained attention due to its potential to improve the accuracy and reliability of forecasting models and develop a practical evaluation method for tsunami debris and accumulation, demonstrating the use of prediction analysis to reveal hazards undetected by conventional tsunami inundation analysis [18]. Collectively, these studies underscore the growing interest in leveraging machine learning, including RF, to enhance the prediction and forecasting of tsunamis.

### 2.3. Relevance of Algorithms to Tsunami Prediction

The selection of the SVM algorithm for this study is grounded in several compelling reasons:

1) SVM has demonstrated robustness when dealing with high-dimensional data, making it well-suited for handling tsunami prediction dataset’s complex and multi-dimensional nature. Tsunami prediction often relies on a wide range of variables, including seismic data, oceanographic parameters, and historical tsunami records, all of which contribute to a high-dimensional feature space.
2) SVM's structural risk minimization principle enables it to find an optimal classification hyperplane that generalizes well to unseen data. This ensures the model does not overfit the training data, which is essential for reliable tsunami prediction.

 ISSN 2723-6471 

The RF algorithm complements SVM in several critical aspects, justifying its inclusion in this study:

1) RF's ensemble learning approach, which combines multiple decision trees, enhances prediction accuracy and robustness. This is essential in tsunami prediction, where the combination of various predictive factors and data sources can significantly improve forecasting.
2) The ensemble nature of RF allows for model interpretability, which can provide insights into the relative importance of features in tsunami prediction. Understanding the importance of features aids in identifying key indicators of tsunami events.

In summary, the selection of both SVM and RF is well-justified due to their complementary strengths and suitability for addressing the challenges posed by tsunami prediction. Combining SVM's high-dimensional data handling and non-linearity handling capabilities with RF's ensemble learning and resilience to noisy data forms a robust framework for accurate and reliable tsunami forecasting.

**3. Method**

## 3.1. Data Collection

The dataset is sourced from Kaggle, a comprehensive collection of earthquake records from January 1, 2001, to January 1, 2023. It consists of 782 detailed earthquake entries. Each record in the dataset includes various attributes crucial for understanding and analyzing seismic events. These attributes encompass the title of the earthquake, its magnitude, date and time, various intensity measures such as CDI (Community et al.) and MMI (Modified Mercalli Intensity), and the alert level which ranges from green to red, indicating the severity of the earthquake.

Significantly, the dataset includes a 'tsunami' field, marked "1" for oceanic events and "0" otherwise, which is vital for tsunami prediction studies. This field, along with others such as magnitude, depth, and location coordinates (latitude and longitude), plays a crucial role in assessing the potential for a tsunami following an earthquake. The 'sig' attribute quantifies the significance of each event, factoring in magnitude, MMI, felt reports, and estimated impact, providing a comprehensive overview of the earthquake's overall impact.

The dataset also details the source of the data ('net'), the number of seismic stations involved in the data collection ('nst'), and the distance from the epicenter to the nearest station ('dmin'). The 'gap' attribute describes the azimuthal gap between stations, essential for determining the reliability of the earthquake's location data. The 'magType' field indicates the method used to calculate the earthquake's magnitude, and the 'depth' field specifies the depth at which the earthquake began. Additionally, the dataset includes the specific location within the country, the affected continent, and the country, offering a complete geographical context of each seismic event.

## 3.2. Data Preprocessing

In the data preprocessing stage, meticulous attention was paid to preparing the dataset for practical machine learning analysis. This process was essential to ensure the reliability and validity of the subsequent model predictions.

The initial task involved a thorough examination of the dataset to identify and address missing values. Several columns, such as 'location,' 'continent,' and 'country,' exhibited many missing entries. Decisions regarding the imputation or removal of these values were made based on their criticality to the study and the extent of the missing data. Additionally, the dataset was scrutinized for errors, focusing on rows containing incomplete information in the 'title' or 'location' fields. Appropriate measures were taken to correct or exclude these inaccuracies, thereby preserving the dataset's overall integrity.

Feature selection and extraction constituted another vital component of the preprocessing phase. The study specifically targeted features that were deemed pivotal for tsunami prediction. These included 'magnitude,' 'depth,' 'latitude,' 'longitude,' and the target variable 'tsunami.' While the dataset's manageable size rendered dimensionality reduction unnecessary, feature engineering was employed to augment its predictive power. This involved innovative approaches like amalgamating latitude and longitude data into a singular feature indicative of the distance from the nearest coast.

Normalizing and scaling the data were critical steps in the preprocessing strategy, especially for numerical features such as 'magnitude,' 'depth,' and 'dmin.' The application of standardization ensured that these features were transformed to a standard scale without distortion. This step is crucial for models sensitive to the scale of input features, as it enhances their performance by providing a level playing field for all input variables.

Finally, the data imbalance issue, particularly in the 'tsunami' field, was rigorously addressed. Given its potential impact on model performance, strategies like resampling were considered to ensure a balanced representation of both classes in the model training process. This approach is critical in machine learning contexts to mitigate biases towards the majority class.

The dataset was primed for further analytical procedures after completing these preprocessing steps. It was now suitably structured and standardized for in-depth exploratory data analysis, advanced feature engineering, and the development of robust predictive models tailored for tsunami prediction.

## 3.3. Data Splitting into Training and Testing Datasets

In machine learning, partitioning the dataset into training and testing subsets is critical, serving as the cornerstone for model evaluation. This process is instrumental in determining the efficacy of the models developed for tsunami prediction.

The methodology for splitting the data adhered to a well-established rationale. The fundamental purpose of dividing the dataset was to create two sets: one for training the model, encompassing features and the target variable, and another for testing to evaluate the model's accuracy and generalization capability. The training set enables the model to learn and adapt to the patterns within the data. In contrast, the testing set offers a means to assess the model's predictive performance against unseen data rigorously.

In this study, the earthquake dataset underwent a strategic division, adhering to standard practices in the field. An 80-20 split was employed, allocating 80% of the data (comprising 625 records) to the training set and the remaining 20% (157 records) to the testing set. This allocation was based on the dataset's size and characteristics, aiming to provide a comprehensive learning base for the models while reserving a substantial portion for their evaluation.

The process of splitting the data also encompassed elements of randomization and stratification. The randomization ensured that both the training and testing sets were representative of the dataset's overall distribution, thus eliminating potential biases associated with ordered or non-random datasets. Additionally, stratification was employed given the slight imbalance in the 'tsunami' variable within the dataset. This approach ensured that both subsets maintained a similar proportion of tsunami and non-tsunami events, approximately 38.8%, thereby addressing any potential skewness in class distribution.

The culmination of this data-splitting process was pivotal in the study. It guaranteed that the machine learning models were trained on a dataset that was not only representative of the diverse range of scenarios but also unbiased. Simultaneously, it provided a distinct and equally representative subset of data for validating the models' predictive prowess. This meticulous division of data into training and testing sets thus laid the foundation for a fair and comprehensive assessment of the models' performance in tsunami prediction.

## 3.4. SVM Model Training

The training of the SVM model was a pivotal component of our research, aimed at harnessing its capabilities for tsunami prediction. This section encapsulates the comprehensive methodology undertaken for training the SVM model, including its theoretical framework, kernel choice, training process, and hyperparameter optimization.

The SVM algorithm, renowned for its robustness in supervised learning tasks, plays a crucial role in classification and regression scenarios. In tsunami prediction, a classification challenge, the SVM algorithm's objective is to delineate an optimal hyperplane that distinctly segregates different classes within the feature space. The essence of the SVM methodology lies in its use of support vectors, the closest data points to the hyperplane, which are pivotal in defining the hyperplane that maximizes the margin between the classes. This characteristic renders the SVM algorithm particularly effective in high-dimensional spaces and highlights its computational efficiency, making it an apt choice for complex classification tasks such as tsunami prediction.

In selecting the appropriate kernel for the SVM model, the study gravitated toward the Radial Basis Function (RBF) kernel, recognized for its proficiency in managing non-linear relationships. This choice was informed by the nature of the dataset and the specific requirements of tsunami prediction. The kernel selection process was complemented by careful parameter tuning, focusing on the regularization parameter (C) and the kernel coefficient (gamma), both crucial in enhancing the model's performance.

The training of the SVM model was grounded in a robust mathematical framework centered around solving an optimization problem aimed at maximizing the margin between the classes. This process involved intricate quadratic programming to ascertain the most effective hyperplane. The training steps included:

1) Select the RBF kernel and initialize the parameters.
2) Transforming the training data using the selected kernel.
3) Training the SVM model to find the hyperplane that maximizes the margin between classes.
4) Utilizing the support vectors and the hyperplane to classify new data points.

Hyperparameter tuning emerged as an integral aspect of the training process, encompassing the exploration of various parameter combinations through grid search and cross-validation techniques. This meticulous approach aimed to identify the optimal set of parameters that would culminate in the highest level of predictive accuracy.

The outcomes of the SVM model training were promising, identifying the best-performing model characterized by a regularization parameter (C) of 10 and a kernel coefficient (gamma) of 'auto.' The model exhibited a commendable accuracy of approximately 63.2%, as determined through a 5-fold cross-validation approach. While this accuracy serves as a positive initial indicator of the model's potential, it also opens avenues for further enhancements. Continuous improvements, including more refined feature engineering and the exploration of alternative kernels or other advanced machine learning models, could further elevate the model's efficacy in accurately predicting tsunami events.

### 3.5. RF Model Training

The training of the RF model, aimed at leveraging its ensemble learning capabilities for tsunami prediction, was a critical aspect of this study. This section details the comprehensive approach adopted for the RF model training, including its conceptual framework, ensemble learning principles, decision tree construction, and hyperparameter optimization.

The RF algorithm is a sophisticated ensemble learning method widely utilized in classification and regression tasks. It creates an array of decision trees during the training phase. It outputs the mode of the classes (for classification) or the mean prediction (for regression) derived from these individual trees. The RF model operates by constructing multiple decision trees, each developed from a bootstrap sample (a sample drawn with replacement) from the training set. These decision trees are then aggregated to produce a more accurate and stable prediction.

Ensemble learning, the underlying principle of the RF algorithm, is a technique that employs multiple learning algorithms to achieve better predictive performance than what might be obtained independently from any single learning algorithm. In the context of the RF model, it amalgamates predictions from several decision trees to enhance accuracy over any individual model. This approach effectively reduces overfitting, a common pitfall in machine learning, by averaging the results.

The construction of decision trees within the RF model incorporates an element of randomness, with each tree being built on a different subset of the data. This methodology introduces substantial diversity into the model, contributing to a more robust and accurate ensemble than a solitary decision tree. The RF model's decision trees are built on different samples and involve selecting the best split among a random subset of features at each node. This strategy further augments the model's robustness.

Optimizing the performance of the RF model involved tuning key parameters, including the number of trees (n_estimators), maximum depth (max_depth), minimum samples required to split an internal node (min_samples_split), and minimum samples required at a leaf node (min_samples_leaf). The tuning process, akin to that used for the SVM model, involved a combination of grid search and cross-validation to identify the most effective set of parameters.

The target variable for the model was 'tsunami,' with predictors including 'magnitude,' 'alert,' previous tsunami occurrence, 'depth,' and 'location/continent.' These features were carefully selected and prepared for the training process. However, the training of the RF model encountered challenges due to its time-intensive nature, particularly given the extensive hyperparameter grid and the use of cross-validation. This highlighted one of the potential limitations of RFs: their computational intensity during hyperparameter tuning.

A more streamlined approach was adopted for the RF model training. This involved simplifying the hyperparameter grid, reducing the cross-validation folds, and implementing incremental training. These modifications allowed for a more efficient training process while still capturing the essential characteristics of the model. The outcomes of this streamlined training indicated that the best-performing model configuration had no limit on tree depth, required a minimum of 2 samples at each leaf node, 5 samples to split an internal node, and consisted of 200 trees. The model achieved an accuracy of approximately 57.1% using 3-fold cross-validation.

This initial performance provided insights into the RF model's capability to predict tsunamis. While the model showed promise, the accuracy indicated potential areas for improvement. Future enhancements include further hyperparameter tuning, advanced feature engineering, or exploring alternative modeling techniques. Notably, the balance between model complexity and its ability to generalize is critical in achieving optimal performance in ensemble methods like RF.

### 3.6. Evaluation Metrics

In evaluating the performance of machine learning models, particularly for classification tasks like tsunami prediction, various evaluation metrics are employed, each offering unique insights into the model's performance. These metrics are crucial in understanding how effectively a model can predict and classify data, and they play a pivotal role in guiding the model selection process.

Accuracy is the most straightforward and intuitive metric, representing the proportion of correct predictions made by the model out of the total number of predictions. While it provides a general overview of the model's overall effectiveness, accuracy may only sometimes be the most reliable metric, especially in cases where the classes in the dataset are imbalanced.

Precision, another key metric, measures the model's accuracy in predicting positive observations. It is calculated as the ratio of true positives (correctly predicted positive observations) to the sum of true positives and false positives (incorrectly predicted positive observations). Precision is essential in scenarios where avoiding false positives is more critical than detecting all positive instances.

Recall, also known as sensitivity or the true positive rate, quantifies the model's ability to correctly identify actual positives. It is calculated as the ratio of true positives to the sum of true positives and false negatives (missed positive observations). Recall becomes a critical metric in situations where the cost of missing a true positive is significant.

The F1-Score, which is the harmonic mean of precision and recall, serves as a more balanced metric, especially useful in datasets with uneven class distributions. It effectively combines the aspects of both precision and recall into a single measure, providing a more comprehensive view of the model's performance.

The Receiver Operating Characteristic (ROC) curve and the Area Under the Curve (AUC) are used to assess a model's ability to discriminate between classes. The ROC curve plots the true positive rate against the false positive rate at various threshold settings, and the AUC provides an aggregate measure of the model's performance across all possible classification thresholds. This metric is particularly valuable for evaluating models in scenarios with imbalanced class distributions.

This study evaluated both the SVM and RF models using these metrics. The SVM model exhibited higher accuracy (65.61%) and precision (70.59%), suggesting its overall effectiveness and reliability in predicting tsunamis. However, its relatively low recall (19.67%) indicated a potential limitation in identifying many actual tsunami events. In contrast, the RF model demonstrated slightly lower accuracy (61.15%) and precision (50.00%) but achieved a higher recall (36.07%) and F1 score (41.90%). This suggested the RF model's enhanced ability to identify actual tsunamis, albeit at the expense of a higher rate of false positives. The ROC AUC scores for both models were relatively close, with the RF model slightly outperforming the SVM, indicating its marginally better capability in distinguishing between tsunami and non-tsunami events.

These evaluation metrics collectively provided a nuanced understanding of the strengths and weaknesses of each model. While the SVM model was more precise in its predictions, the RF model's higher recall and F1 score suggested its potential suitability for situations where the cost of missing an actual tsunami is more critical than generating false alarms. The comparative analysis of these metrics was instrumental in assessing the models' capabilities and guiding the selection process based on the specific requirements of tsunami prediction.

### 3.7. Tsunami Classification and Prediction Process

The tsunami classification and prediction process using machine learning models encompasses several critical stages, from data preparation to practical decision-making. This section outlines the systematic approach adopted for tsunami prediction using the SVM and RF models.

1) Data Preparation
   The initial step in the prediction process involved carefully selecting relevant features that significantly influence tsunami prediction. This study identified crucial features such as 'magnitude,' 'depth,' and 'alert.' Additionally, data encoding was performed, particularly for categorical variables like 'alert,' converting them into a numerical format compatible with the machine learning models. Furthermore, scaling and normalization were applied, especially for the SVM model, to adjust the features to a uniform scale, as this model is sensitive to the scale of the data.
2) Model Prediction
   The prepared input data was then formatted appropriately, typically structured into a format such as a pandas DataFrame or a NumPy array, to align with the requirements of the models. The prediction phase involved the application of the prediction function of the trained models. For the SVM model, this was executed through svm_model.predict(X), and for the RF model, through rf_model.predict(X).
3) Interpretation of Outputs
   The interpretation of the models' outputs is crucial for understanding their predictions. Both models produce binary outputs (0s and 1s), with '1' indicating a tsunami prediction and '0' indicating no tsunami. The SVM model, known for its higher precision, is more likely to predict tsunamis accurately, but it may miss several actual events due to its lower recall. In contrast, with its higher recall, the RF model is better at identifying actual tsunamis, though it may result in more false positives.
4) Post-Processing and Additional Analyses
   Post-processing involves evaluating the confidence measures provided by the RF model, which offers probabilities of each class, which is essential in high-stakes decision-making. Error analysis was also conducted to identify and understand the models' errors, such as false positives and false negatives, providing insights for future improvements. Additionally, combining outputs from both models could enhance the reliability of the predictions, especially when both models concur.
5) Practical Decision-Making
   The final stage involved integrating the models' predictions into practical decision-making processes. This included using the predictions in conjunction with historical data, geographical information, and seismic activity patterns for comprehensive risk assessment. Balancing the metrics of precision and recall was crucial, especially in scenarios where the implications of false positives and negatives vary significantly. Moreover, integrating model outputs into broader tsunami alert systems involving advanced GIS mapping, public warning systems, and coordination with emergency response teams was considered. Responsive action plans were developed based on the level of risk indicated by the models, ensuring preparedness and prompt response in the event of a tsunami.

## 4. Result and Discussion

### 4.1. Analysis of SVM Model Results

In the evaluation of the SVM model for tsunami prediction, several vital metrics were employed to assess its performance. The model demonstrated an accuracy of 65.61%, precision of 70.59%, recall of 19.67%, F1-score of 30.77%, and a Receiver Operating Characteristic (ROC) Area Under the Curve (AUC) of 62.15%. These results illuminated the SVM model's strengths and limitations in its application to tsunami prediction.

ISSN 2723-6471


The model's relatively high precision indicated its effectiveness in correctly identifying tsunami events when it predicted them, which is crucial in scenarios where the goal is to minimize false alarms. Additionally, SVM's known proficiency in handling high-dimensional datasets was evident, showcasing its adaptability and strength in managing complex classification tasks.

However, the model exhibited a low recall rate, a significant shortcoming, as it needed to identify many actual tsunami events. This limitation is particularly critical in real-world tsunami prediction scenarios, where missing an event could lead to catastrophic consequences. The ROC AUC score, while moderate, suggested potential room for improvement in the model's discriminatory power between tsunami and non-tsunami events.

A visualization approach was undertaken to understand the SVM model's performance better. The initial attempt to visualize the model's decision boundaries faced challenges due to discrepancies in the number of features used in training versus those represented in the visualizations. Addressing this, the SVM model was retrained using only 'magnitude' and 'depth', facilitating a more accurate visual representation of the model's decision boundaries and predictions.

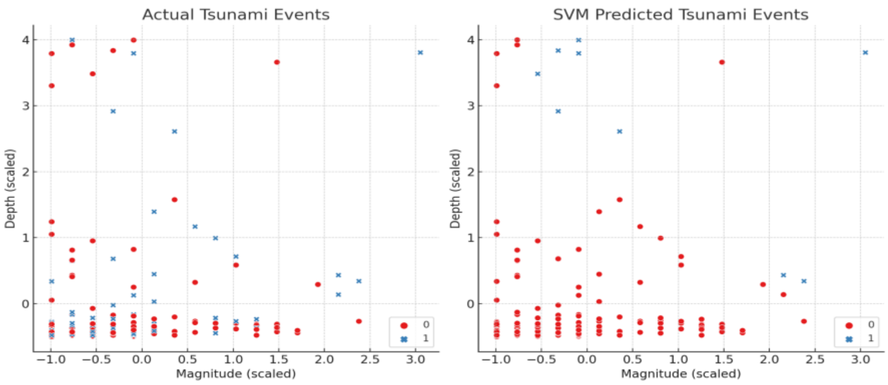


**Figure 1.** Comparison of the actual tsunami events and the model's predictions

The visualizations comprised a side-by-side comparison of the actual tsunami events and the model's predictions, focusing on magnitude and depth. These visualizations were crucial in assessing the alignment of the model's predictions with actual events.

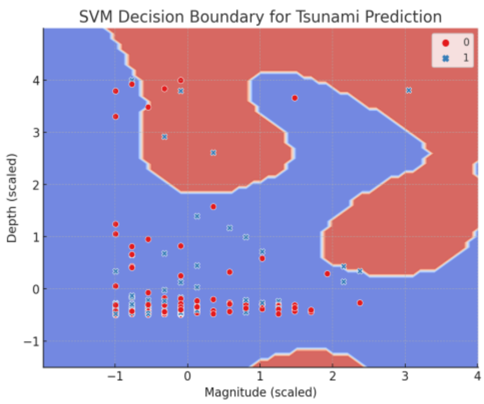


**Figure 2.** The decision boundary plots

The decision boundary plots further revealed the intricacies of the classification task, showing how the SVM model distinguished between tsunami and non-tsunami events based on these critical features. The predictive visualization highlighted the model's conservative nature in forecasting tsunamis, evidenced by its higher precision but lower recall. The decision boundary complexity depicted in the visualization underscored the model's capability to differentiate between the classes, albeit with limitations. Moreover, the visualizations emphasized the significant role of magnitude and depth as determinants in tsunami prediction, underscoring the model's reliance on these features.

In conclusion, the analysis and visualization of the SVM model provided a comprehensive view of its capabilities and limitations in tsunami prediction. While the model showed a degree of accuracy, the insights from the visualizations pointed to the need for further enhancement, particularly in improving its recall rate and ability to differentiate between tsunami and non-tsunami events more effectively. This highlights an area for future research and development to bolster the model's performance in accurately predicting tsunami occurrences.

## 4.2. Analysis of RF Model Results

The performance of the RF model in tsunami prediction was rigorously evaluated through a comprehensive set of metrics. The model demonstrated an accuracy of 61.15%, precision of 50.00%, recall of 36.07%, F1-score of 41.90%, and a Receiver Operating Characteristic (ROC) Area Under the Curve (AUC) of 63.84%. These metrics provide a detailed assessment of the model's capabilities in accurately predicting tsunami events.

In assessing the RF model's predictive accuracy and reliability, it was observed that the model correctly predicts tsunamis approximately 61% of the time, indicating a moderate level of accuracy. The ROC AUC score, reflecting the model's ability to differentiate between tsunami and non-tsunami events, also suggested moderate discrimination. The RF model's balance between precision and recall indicates a cautious approach in prediction, not significantly biased toward avoiding false positives or capturing all positive tsunami cases.

The model's strengths in tsunami prediction were highlighted by its comparatively higher recall than the SVM model, indicating an enhanced capability in identifying actual tsunami events. This is a critical feature in tsunami prediction, where missing a real event can have dire consequences. Additionally, the RF model exhibited an effective handling of non-linear relationships between features, a significant advantage in dealing with complex patterns inherent in natural disaster data. However, the model encountered challenges, including a lower precision, which implied a higher incidence of false positives. There was also an indication that the model might struggle to interpret complex underlying patterns specific to tsunami occurrences, a potential limitation in its predictive power.

A series of visualizations were created to gain a deeper understanding of the RF model's performance. These visualizations aimed to compare the actual tsunami events with the model's predictions, provide insight into the diversity of decision-making processes within the RF ensemble, and illustrate the importance of various features in the model's predictions. The visualization process, however, faced challenges similar to those encountered with the SVM model due to the initial training of the RF model on a different set of features than those represented in the visualizations. To address this, the model was retrained using only 'magnitude' and 'depth', enabling more accurate visual representations of its performance.

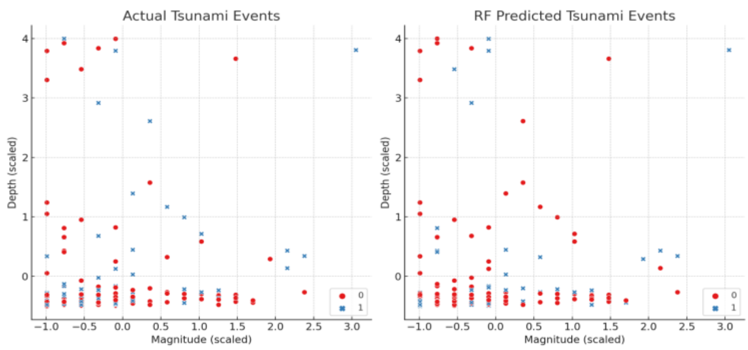


**Figure 3.** The comparative plots of actual tsunami events and the RF model's predictions

The resulting visualizations offered valuable insights into the model's performance. The comparative plots of actual tsunami events and the RF model's predictions highlighted how closely the model's predictions aligned with actual events, particularly in magnitude and depth.

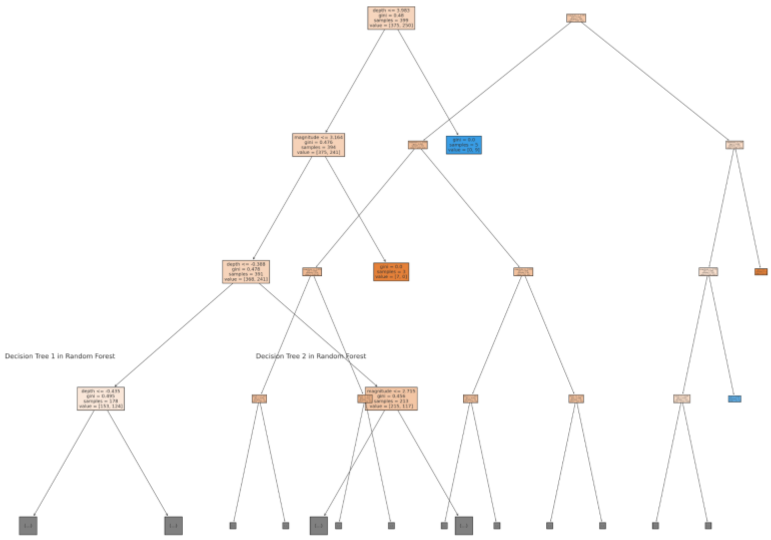


**Figure 4.** Individual Decision Tree

Visualizing individual decision trees within the RF ensemble provided a window into the diverse decision-making processes within the model, with each tree contributing uniquely to the final prediction. This diversity in approach underscored the model's robustness and adaptability in predicting tsunamis.

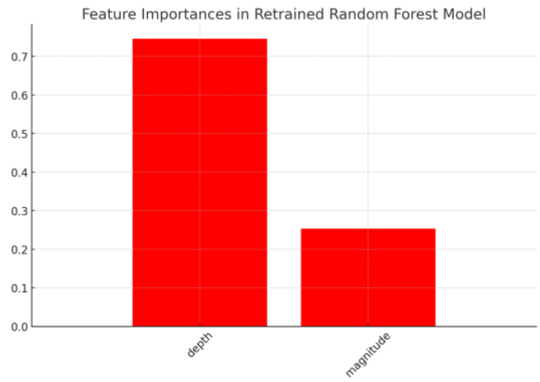


**Figure 5.** Feature Importance Chart

Additionally, the feature importance chart elucidated the relative significance of 'magnitude' and 'depth' in the model's predictions, offering clarity on the critical drivers of tsunami prediction within the model. In conclusion, the visualizations provided a comprehensive understanding of the RF model's strengths and limitations in predicting tsunamis. The model demonstrated a reasonable ability to discern between tsunami and non-tsunami events, with noted areas for improvement in accuracy. The insights gained from the visualizations, particularly regarding feature importance and the diversity of decision trees within the RF ensemble, are invaluable for refining the model. These insights will guide future enhancements, aiming to bolster the model's predictive accuracy and reliability in the critical context of tsunami prediction.

### 4.3. Performance Comparison Between SVM and RF

This study conducted a detailed comparative analysis between the SVM and RF models, focusing on their application in tsunami prediction. This comparison was rooted in evaluating various key performance metrics, providing insights into each model's strengths and weaknesses and their practical implications.

The accuracy of the SVM model stood at 65.61%, surpassing the RF model, which recorded an accuracy of 61.15%. This higher accuracy of the SVM model indicates its superior capability in correctly identifying both tsunami and non-tsunami events. Regarding precision and recall, the SVM model exhibited a precision of 70.59% and a recall of 19.67%, reflecting its higher propensity to avoid false positives but at the expense of missing a significant number of actual tsunamis. Conversely, the RF model demonstrated a lower precision of 50.00% but a notably higher recall of 36.07%, suggesting its enhanced ability to identify actual tsunamis, albeit with an increased likelihood of false positives.

The F1-Score, which balances precision and recall, was higher in the RF model at 41.90%, compared to 30.77% in the SVM model. This suggests that the RF model achieves a better equilibrium between precision and recall, a crucial factor in scenarios with imbalanced classes. Additionally, the RF model slightly outperformed the SVM in the ROC AUC metric, with scores of 63.84% and 62.15%, respectively. This marginal superiority indicates the RF model's better ability to distinguish between tsunami and non-tsunami classes.

The efficiency and scalability of the two models also presented a contrast. The SVM model tends to be more computationally demanding, particularly with larger datasets and requires meticulous hyperparameter tuning. This characteristic may limit its scalability in handling huge datasets. In contrast, the RF model demonstrates robustness against overfitting and efficiently manages larger feature sets, making it more scalable and suitable for increasing data sizes and complex feature spaces.

The selection between the SVM and RF models for tsunami prediction is contingent on the specific requirements of the task. The SVM model, with its higher precision and overall accuracy, is preferable in situations where the primary concern is to minimize false positives, and the dataset is manageable. In contrast, the RF model, characterized by its higher recall and F1-Score, is more suitable for scenarios where failing to detect an actual tsunami is more critical than avoiding false alarms. This model is also favorable for large datasets or those involving numerous features, particularly when handling non-linear relationships is vital.

In conclusion, the choice of the SVM or RF model in tsunami prediction should be informed by carefully considering factors such as dataset size, feature complexity, and computational resources. This decision is crucial in determining the effectiveness and reliability of the tsunami prediction task, with each model offering distinct advantages that must be weighed against the specific needs and constraints of the scenario.

### 4.4. Discussion of Findings and Their Implications

The comparative analysis conducted between the SVM and RF models in the context of tsunami prediction offers insightful revelations, each with significant implications for the application of machine learning in disaster management and preparedness.

The analysis highlights a crucial trade-off between precision and recall, with the SVM model demonstrating higher precision, thereby making it more suited for scenarios where reducing false positives is a priority. In contrast, with its higher recall, the RF model proves more effective in correctly identifying actual tsunami events. This attribute is particularly critical in disaster scenarios, where the consequences of missing an event can be severe. Despite the SVM's superior accuracy, indicating its effectiveness in general classification tasks within tsunami prediction, the often higher cost of false negatives in disaster management could necessitate a preference for the RF model.

The practical implications of this comparative analysis are profound, particularly concerning disaster response and alert systems. The selection of either model significantly impacts the strategy for disaster response. For instance, while an RF model may trigger more alerts, potentially including false alarms, it offers a greater assurance of capturing actual tsunami events. On the other hand, due to its high precision, the SVM model might reduce public panic caused by false alarms but at the risk of overlooking critical events. This consideration is pivotal in enhancing public safety, as models like the RF could provide more reliable alerts for potential tsunamis with their comprehensive coverage potential.

In the realm of decision-making for disaster management and preparedness, both models can be integrated into broader risk assessment frameworks. This integration provides valuable insights for decision-makers, assisting in informed policy formulation and emergency planning. It ensures that resources are allocated effectively and that response mechanisms are robust and well-prepared.

Future research and improvement in this area could focus on several key aspects. Advancements in feature engineering and selection, particularly incorporating temporal and geographical data, could enhance the models' predictive performance. Exploring ensemble techniques that combine the strengths of various models might offer a more nuanced approach to tsunami prediction. Integrating real-time seismic and oceanographic data can lead to more dynamic and

ISSN 2723-6471


responsive prediction models. Furthermore, aligning these models with the needs and expectations of end-users, including emergency responders and the public, is essential. This alignment involves ensuring the interpretability and actionability of model outputs, which is crucial in human-centered design.

In conclusion, the findings from this comparative analysis emphasize the complexities and considerations involved in employing machine learning for tsunami prediction. The selection between the SVM and RF models is influenced by specific demands and constraints inherent in disaster management scenarios. Continuous refinement and adaptation of these models, guided by ongoing research and feedback from real-world applications, are essential for enhancing disaster preparedness and strengthening response efforts. This ongoing development is critical in ensuring these technological advancements effectively contribute to public safety and disaster management initiatives.

## 5. Conclusion

The analysis conducted on the SVM and RF models for tsunami prediction has elucidated their respective performances and characteristics. The SVM model exhibited a notable strength in precision, effectively minimizing false positives, a crucial aspect in situations where prediction accuracy is paramount. Conversely, the RF model demonstrated a superior recall ability, indicating its effectiveness in identifying actual tsunami events. This characteristic is precious in disaster scenarios where the failure to detect an event can have grave consequences. The evaluation of both models encompassed a range of metrics, including accuracy, precision, recall, F1-score, and ROC AUC. The SVM model excelled in accuracy and precision, suggesting its effectiveness in classifying tsunami and non-tsunami events. On the other hand, the RF model achieved a more balanced performance between precision and recall and marginally outperformed the SVM in ROC AUC. Additionally, the visualizations created for both models provided more profound insights into their classification behaviors and highlighted the importance of critical features in tsunami prediction.

The decision to select between the SVM and RF models for tsunami prediction should be carefully considered, considering the task's specific requirements. The trade-offs between the potential consequences of false negatives and false positives are pivotal in this decision-making process. In disaster management, where the costs associated with missed tsunamis (false negatives) are often significantly high, models with higher recall, such as RF, may be more appropriate.

The SVM model is characterized by its high precision and effectiveness in high-dimensional spaces, making it a suitable choice for certain classification tasks. However, it is limited by a lower recall and potential computational challenges when handling large datasets. The RF model, in contrast, is proficient in managing non-linear data, shows robustness against overfitting, and offers scalability for large datasets. Its primary limitations include a tendency to produce more false positives and potential difficulties in interpretability due to its ensemble nature.

Future research in this area could explore developing hybrid or ensemble models that leverage the strengths of both SVM and RF to enhance tsunami prediction capabilities. Future feature engineering advancements, incorporating data such as real-time seismic information, geographical factors, and historical records, could improve predictive accuracy. Developing real-time prediction systems that integrate current data for dynamic and responsive tsunami prediction is another promising area of research. Additionally, focusing on user-centric design to ensure prediction systems are accessible and actionable for disaster management authorities and the public is crucial. Enhancing the interpretability of these models is also essential for fostering trust and reliability among users.

This study underscores the potential and inherent challenges of using machine learning models like SVM and RF in the context of tsunami prediction. The effectiveness of these models in real-world scenarios hinges on various factors, including the nature of the data, the specific requirements of the prediction task, and the overall disaster management framework. Continued research and development, particularly in integrating these models into comprehensive disaster response systems, are essential for maximizing their practical utility. The ongoing advancement in this field is crucial for harnessing technological innovations to improve disaster preparedness and response efforts.

## 6. Declarations

### 6.1. Author Contributions

Conceptualization: H.T.S. and Y.D.; Methodology: Y.D.; Software: H.T.S.; Validation: H.T.S. and Y.D.; Formal Analysis: H.T.S. and Y.D.; Investigation: A.; Resources: S.; Data Curation: A.; Writing Original Draft Preparation: A. and H.T.S.; Writing Review and Editing: A. and S.; Visualization: H.T.S.; All authors have read and agreed to the published version of the manuscript.

### 6.2. Data Availability Statement

The data presented in this study are available on request from the corresponding author.

### 6.3. Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

### 6.4. Institutional Review Board Statement

Not applicable.

### 6.5. Informed Consent Statement

Not applicable.

### 6.6. Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## References

[1] J. Liu et al., "Editorial: Spatial modelling and failure analysis of natural and engineering disasters through data-based methods," *Frontiers in Earth Science*, vol. 10, no. 1, pp. 1–13, Aug. 2022. doi:10.3389/feart.2022.1000540.

[2] S.-J. Yun, J.-W. Kwon, and W.-T. Kim, "A Novel Digital Twin Architecture with similarity-based hybrid modeling for supporting dependable disaster management systems," *Sensors*, vol. 22, no. 13, pp. 4774–4786, Jun. 2022. doi:10.3390/s22134774

[3] S. Paul and H. Sharif, "Analysis of damage caused by hydrometeorological disasters in Texas, 1960–2016," *Geosciences*, vol. 8, no. 10, pp. 384–396, Oct. 2018. doi:10.3390/geosciences8100384

[4] T. Ulrich et al., "Coupled, physics-based modeling reveals earthquake displacements are critical to the 2018 Palu, Sulawesi Tsunami," *Pure and Applied Geophysics*, vol. 176, no. 10, pp. 4069–4109, Aug. 2019. doi:10.1007/s00024-019-02290-5.

[5] M. Kamiya, Y. Igarashi, M. Okada, and T. Baba, "Numerical experiments on tsunami flow depth prediction for clustered areas using regression and machine learning models," *Earth, Planets and Space*, vol. 74, no. 1, pp. 1–12, Aug. 2022. doi:10.1186/s40623-022-01680-9.

[6] I. E. Mulia, N. Ueda, T. Miyoshi, A. R. Gusman, and K. Satake, "Machine learning-based tsunami inundation prediction derived from offshore observations," *Nature Communications*, vol. 13, no. 1, p. `-12, Sep. 2022. doi:10.1038/s41467-022-33253-5.

[7] S. K. Mishra, "Assessing long-term impacts of disaster using predictive data ANALYTICSFOR effective decision support," *International Journal of Advanced Research in Computer Science*, vol. 14, no. 02, pp. 17–25, Apr. 2023. doi:10.26483/ijarcs.v14i2.6956.

[8] F. H. Hassan, "Comparing performance of machine learning algorithms in a flood prediction model with real data sets," *International Journal of Advanced Trends in Computer Science and Engineering*, vol. 1, no. 1, pp. 152–157, Sep. 2019. doi:10.30534/ijatcse/2019/2381.42019.

[9] F. Jiang, F. Zhao, K. Ma, D. Li, and H. Sun, "Mapping the forest canopy height in northern China by synergizing icesat-2 with sentinel-2 using a stacking algorithm," *Remote Sensing*, vol. 13, no. 8, pp. 1535–1543, Apr. 2021. doi:10.3390/rs13081535.

[10] Y. Shao et al., “Mapping China’s forest fire risks with machine learning,” *Forests,* vol. 13, no. 6, pp. 856–869, May 2022. doi:10.3390/f13060856

[11] J. Zhang, Q. Wang, W. Guo, and L. Li, “A support vector machine based prediction on sensitivity to coal ash blast for different degrees of deterioration,” *Journal of Sensors,* vol. 2022, no. 1, pp. 1–12, Aug. 2022. doi:10.1155/2022/7604338.

[12] G. Li, G. Chen, P. Li, and H. Jing, “Efficient and accurate 3-D numerical modelling of landslide tsunami,” *Water,* vol. 11, no. 10, pp. 2033–2046, Sep. 2019. doi:10.3390/w11102033.

[13] R. Yu and M. Abdel-Aty, “Utilizing support vector machine in real-time crash risk evaluation,” *Accident Analysis & Prevention,* vol. 51, no. 1, pp. 252–259, Mar. 2013. doi:10.1016/j.aap.2012.11.027.

[14] Z. Zhao et al., “An ultra-fast power prediction method based on simplified LSSVM hyperparameters optimization for PV Power Smoothing,” *Energies,* vol. 14, no. 18, pp. 5752–5764, Sep. 2021. doi:10.3390/en14185752.

[15] X. Huang and L. Du, “Fire detection and recognition optimization based on Virtual Reality Video Image,” *IEEE Access,* vol. 8, no. 1, pp. 77951–77961, 2020. doi:10.1109/access.2020.2990224.

[16] J. Gu, “Risk prediction of enterprise credit financing using machine learning,” *Informatica,* vol. 46, no. 7, pp. 1–13, Nov. 2022. doi:10.31449/inf.v46i7.4247.

[17] G. I. Sayed, A. Tharwat, and A. E. Hassanien, “Chaotic dragonfly algorithm: An improved metaheuristic algorithm for feature selection,” *Applied Intelligence*, vol. 49, no. 1, pp. 188–205, Aug. 2018. doi:10.1007/s10489-018-1261-8.

[18] K. Imai et al., “Development of a practical evaluation method for tsunami debris and its accumulation,” *Applied Sciences*, vol. 12, no. 2, pp. 858–870, Jan. 2022. doi:10.3390/app12020858.

[19] H. Cao and C. Dong, “An English pronunciation error detection system based on improved random forest,” *Mobile Information Systems,* vol. 2022, no. 1, pp. 1–10, Apr. 2022. doi:10.1155/2022/6457286.

[20] Y. Cheng, M. Guo, and L. Guo, “Radar signal recognition exploiting information geometry and Support Vector Machine,” *IET Signal Processing*, vol. 17, no. 1, pp. 1–13, Oct. 2022. doi:10.1049/sil2.12167.

[21] S. Madhavi, D. Divyadharshini, and S. Divya, “Machine learning model to predict cyber attack,” *International journal of health sciences*, vol. 1, no. 1, pp. 1341–1349, Jun. 2022. doi:10.53730/ijhs.v6ns6.9745.

[22] H. Talebi, L. J. Peeters, A. Otto, and R. Tolosana-Delgado, “A truly spatial random forests algorithm for geoscience data analysis and modelling,” *Mathematical Geosciences,* vol. 54, no. 1, pp. 1–22, Jul. 2021. doi:10.1007/s11004-021-09946-w.

[23] X. Tong and J. Duan, “Research on the prediction of Nonbreakeven financial products’ yield of commercial banks based on machine learning,” *Mobile Information Systems,* vol. 2022, no. 1, pp. 1–7, Oct. 2022. doi:10.1155/2022/8731261.

[24] M. A. Bayir, K. Shamsi, H. Kaynak, and C. G. Akcora, “Topological Forest,” *IEEE Access,* vol. 10, no. 1, pp. 131711–131721, 2022. doi:10.1109/access.2022.3229008.