Journal of
Information Systems Engineering
and Business Intelligence

Vol.10, No.3, October 2024
Available online at: http://e-journal.unair.ac.id/index.php/JISEBI

Classification of Non-Seismic Tsunami Early Warning Level Using Decision Tree Algorithm

Elmo Juanara [1)*], Chi Yung Lam [2)]

[1)2)]*Japan Advanced Institute of Science and Technology, Nomi, Japan*
[1)]elmo.juanara@jaist.ac.jp, [2)]cylam@jaist.ac.jp

Abstract

Background: Tsunami caused by volcanic collapse are categorized as non-seismic uncommon events, unlike tsunamis caused by earthquakes, which are common events. The traditional tsunami early warning based on the seismic sensor (e.g. earthquake detectors) may not be applicable to volcanic tsunamis because they do not generate seismic waves. Consequently, these tsunamis cannot be detected in advance, and warnings cannot be issued. New methods should be explored to address these non-seismic tsunamis caused by volcanic collapse.
Objective: This study explored the potential of machine learning algorithms in supporting early warning level issuing for non-seismic tsunamis, specifically volcanic tsunamis. The Anak Krakatau volcano event in Indonesia was used as a case study.
Methods: This study generated a database of 160 collapse scenarios using numerical simulation as input sequences. A classification model was constructed by defining the worst tsunami elevation and its arrival time at the coast. The database was supervised by labeling the warning levels as targets. Subsequently, a decision tree algorithm was employed to classify the warning levels.
Results: The results demonstrated that the classification model performs very well for the Major Tsunami, Minor Tsunami, and Tsunami classes, achieving high precision, recall, and F1-Score with very high accuracy of 98%. However, the macro average indicates uneven performance across classes, as there are instances of 'No Warning' in some coastal gauges.
Conclusion: To improve the model performance in the 'No Warning' class, it is necessary to balance the dataset by including more 'No Warning' scenarios, which can be achieved by simulating additional scenarios involving very small-volume collapse. Additionally, exploring additional collapse parameters such as dip angle and outlier volume could contribute to developing a more robust classification model.

Keywords: Machine Learning, Classification, Volcanic Tsunamis, Early Warning, Decision Tree

Article history: Received 12 July 2024, first decision 17 September 2024, accepted 10 October 2024, available online 28 October 2024

I. INTRODUCTION

High-consequence, low-probability (HCLP) events present significant challenges. 'High-consequence' refers to events with severe impacts or outcomes resulting in major damages. Conversely, 'low probability' indicates events that are unlikely to occur frequently. The consequences of HCLP events extend rapidly across various sectors, often causing second or third-order impacts. Recent trends show that HCLP tragedies are increasing in severity and frequency [1], yet historical data are scarce, making prediction difficult [2]. One example of a high-consequence, low-probability (HCLP) event is tsunamis [3]. Tsunamis are commonly generated by seabed displacements caused by earthquakes. However, they can also result from other mechanisms, such as volcanic activity, which is categorized as non-seismic tsunamis [4]. Although non-seismic tsunamis are rare events (low probability), there have been approximately 100 notable such as volcanic tsunamis recorded in the past three centuries, yet they pose significant hazards due to their high consequences within the tsunami domain [4]. Furthermore, as highlighted by [5], one significant source of tsunamis from volcanic activity is flank or sector collapse.

Anak Krakatau, part of the renowned Krakatau volcanic complex located in the Sunda Strait between Java and Sumatra in Indonesia, has a storied history of volcanic activity. In December 2018, Anak Krakatau (AKV) experienced a major lateral collapse during a period of eruptive activity [6]. While flank instability and sector collapse are common occurrences on volcanic islands, the collapse of the AKV posed a particularly severe threat, triggering a deadly tsunami [7]. This catastrophic tsunami struck the Sunda Strait region, resulting in hundreds of fatalities, and thousands of injuries [6], [8], [9], [10]. The volcano is mainly located together with the caldera wall [4], [11] which will be contributing to the flank collapse of AKV and the triggering tsunamis. As happened in the event of December 2018,

* Corresponding author

ISSN 2443-2555 (online) 2598-6333 (print)
doi: http://dx.doi.org/10.20473/jisebi.10.3.378-391

the thickness, volume, and velocity of collapse become important points in tsunami generation from volcano collapse [4]. In order to save people's lives from volcanic tsunamis, especially for coastal society, an early warning system is significant for taking decisions and action [12]. However, currently, there is no standardized approach for early warning systems particularly for non-seismic tsunamis [13]. Most tsunami warning systems are developed for seismogenic tsunamis (e.g., earthquake-generated) and not for non-seismic tsunamis [14]. Therefore, the traditional tsunami early warning method based on the seismic sensor may not be appropriate for detecting volcanic tsunamis, as they do not produce seismic waves. Hence, detecting these kinds of non-seismic tsunamis and issuing early warnings become very challenging tasks in the tsunami domain. Previous studies have only focused on numerical modeling to understand the mechanisms of volcanic tsunamis [15], [16], [17], [18], [19], [20], yet still less emphasis on developing effective early warning systems.

Moreover, other past research has studied various Machine Learning (ML) algorithms for disaster prediction. For instance, Support Vector Machines (SVM) are widely used in classification tasks since this algorithm handles high dimensional data and is able to make an optimal decision [21], [22]. Even though the accuracy of SVM is high, in most cases, SVM is considered as black-box model. The reason because the model cannot give the appropriate insight about how the decision occurs. Similar to SVM, k-Nearest Neighbors (kNN) is also usually for classification tasks in many research and application fields, including disaster prediction [23], [24]. kNN has a strength point such as being easy to interpretable and simple to implement. However, the computational cost is very high if the study uses a large dataset since the kNN will be storing all training data dan have to calculate the distance for each new prediction. This is incompatible with the purpose of tsunami early warning systems that require a real-time application. Alternatively, the Decision Tree method is one of the ML algorithms that has a simple and interpretable model for dealing with the decision process [25], [26]. The interpretability makes the Decision Tree suitable for the aim of immediate response such as tsunami early warning. By splitting the dataset using some features and determined thresholds, this model forms a structure like a tree structure where each branch stands for a decision rule, and each leaf illustrates a classification score or output. This model gives two main advantages, namely transparency and efficiency. Transparency is very important in the process of decision-making. Decision Tree can visualize the simple process and make it easier to understand if-then rules. Regarding efficiency, Decision Trees can deal with large datasets and are capable of reducing the computational cost. This is very important in the case of early warning issuing. Moreover, Decision Trees are normally appropriate for handling imbalanced datasets and non-linear relationships between given features, which usually appear in the processing of tsunami datasets.

In this research, a Decision Tree model and a tsunami numerical model were combined to optimize the classification of early warning levels for non-seismic tsunamis. By integrating these two approaches, the numerical model plays an important role in simulating tsunami events and creating a dataset while the Decision Tree algorithm classifies the outcome data for improving the decision-making capabilities of tsunami early warning systems. This combination allows for both accurate simulation of complex, non-seismic tsunami events, such as the Anak Krakatau collapse, and efficient classification for real-time decision-making. By focusing on interpretability and operational efficiency, this integrated approach ensures that early warning provides actionable insights promptly.

II. METHODS

The methodology of this study employs a comprehensive warning decision model that integrates tsunami numerical simulations, input sequence consistency, and model construction for supervising the database within a decision-making framework (Fig. 1). Tsunami numerical simulation utilizes COMCOT (Cornell Multi-grid Coupled Tsunami Model) software [27], which simulates tsunami scenarios based on various volcanic collapse parameters.

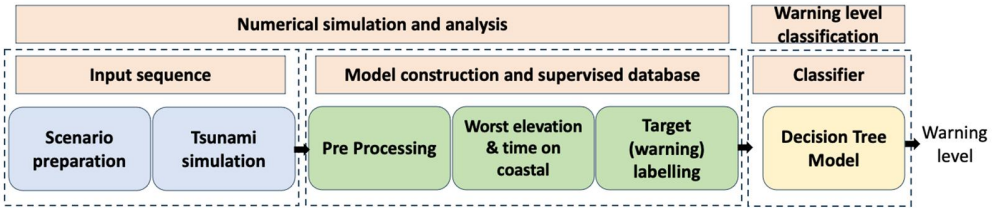

Fig. 1 Flowchart of the warning decision model

A. Tsunami simulation numerical model

When a tsunami happens, the dispersion effect is minimal because the tsunami's wavelength usually exceeds the depth of the water. Shallow water equations are sufficient to generate tsunami numerical simulation as used by previous studies [16], [28], [29], [30]. Numerical models based on shallow water equations are very effective for

simulating transoceanic tsunamis because they use explicit numerical methods and do not require higher-order derivatives related to non-linearity and frequency dispersion [27]. This study's main software tool for simulating tsunami events caused by landslides is the COMCOT (Cornell Multi-grid Coupled Tsunami Model), created by [27]. This model uses linear shallow water equations in spherical coordinates and is based on shallow water equations (equations 1-4). In order to accurately capture the dynamics of tsunami propagation and inundation, the model ensures fine spatial resolution by using a grid size of 0.2 minutes, which is equivalent to 12 arc seconds or roughly 370 meters. The following is a written representation of the nonlinear shallow water equations governing equations.

$$\frac{\partial \eta}{\partial t} + \frac{1}{R\cos\varphi}\left\{\frac{\partial P}{\partial \psi} + \frac{\partial}{\partial \varphi}(\cos\varphi Q\right\} = -\frac{\partial h}{\partial t} \tag{1}$$

$$\frac{\partial \mathrm{P}}{\partial t} + \frac{1}{R\cos\varphi}\frac{\partial}{\partial \psi}\left\{\frac{P^2}{H}\right\} + \frac{1}{R}\frac{\partial}{\partial \psi}\left\{\frac{PQ}{H}\right\} + \frac{gH}{R\cos\varphi}\frac{\partial \eta}{\partial \psi} - fQ + F_x = 0 \tag{2}$$

$$\frac{\partial Q}{\partial t} + \frac{1}{R\cos\varphi}\frac{\partial}{\partial \psi}\left\{\frac{PQ}{H}\right\} + \frac{1}{R}\frac{\partial}{\partial \psi}\left\{\frac{Q^2}{H}\right\} + \frac{gH}{R}\frac{\partial \eta}{\partial \varphi} + fP + F_y = 0 \tag{3}$$

in which H is the total water depth and $H = \eta + h$; F^x and F_y represents the bottom friction in X and Y directions, respectively. And these two terms are evaluated via Manning's formula, where n is the Manning's roughness coefficient.

$$F_x = \frac{gn^2}{H^{\frac{7}{3}}} P(P^2 + Q^2)^{1/2} \tag{4}$$

$$F_x = \frac{gn^2}{H^{\frac{7}{3}}} Q(P^2 + Q^2)^{1/2} \tag{5}$$

There are a number of consecutive steps in the COMCOT model's flow chart. The first step in creating realistic and precisely located tsunami scenarios is for COMCOT to read the setup parameters. This is followed by the crucial input of bathymetric data. The simulation bathymetric data was derived from the official Indonesian source (https://tanahair.Indonesia.go.id/demnas/#/), which also included the topography data. The model then goes on to create the computational grid that will be used during the simulation procedure. The Landslide simulation model is selected concurrently, and its parameters are included in the framework. When the model has finished initializing, it configures the initial conditions (IC), which include the initial displacement of the water surface. Boundary conditions (BC) are defined to finish the model configuration after the IC setup. The simulation then moves on to the computation of the solver for the Shallow Water Equations (SWEs), which runs numerical techniques to solve the fluid dynamics in the specified region. The model proceeds to handle floor deformation phenomena after the SWE solver calculation, reflecting changes in the underwater landscape brought about by dynamic forces. At the end of the simulation, the tsunami wavefield is generated, containing parameters like wave amplitude and temporal evolution, as well as time-series records that are derived and provide comprehensive outputs that show the spatiotemporal evolution of the event.

1) Database generation

This study uses 160 volcanic collapse scenarios as a dataset which are generated from numerical modeling explained previously. These scenarios are referred to the case study of the December 2018 Anak Krakatau Volcano tsunami to define the worst-case wave elevation (see Table 1).

TABLE I
EXAMPLE OF SIMULATION SCENARIO FOR VOLCANIC TSUNAMI DATABASE GENERATION

Scenario	Collapse Direction	Angle Degree	Thickness (m)	Length (m)	Width (m)	Volume (km^3)
1	East	9°	176	2906	78.2	0.04
22	north	9°	62	1250	903.2	0.07
43	northeast	9°	127	1395	846.7	0.15
66	northwest	9°	130	3288	444.5	0.19
91	south	9°	127	3490	541.5	0.24
114	southeast	9°	63	1499	2859.0	0.27
137	southwest	9°	43	1439	4848.3	0.30
160	west	9°	106	3727	835.3	0.33

The dataset included the all of potential collapse possibilities by diversifying the collapse parameters such as collapse direction, volume, dimensions (length, thickness, width), and angle-degree collapse. The collapse volumes vary starting start from 0.04 to 0.33 cubic kilometers and the collapse directions covered all directions such as east, north, northeast, northwest, south, southeast, southwest, and west referring to previous research [6], [8], [31], [32], [33]. The initial landslide lengths and thicknesses also range from 1.0 km to 4.0 km as proposed in a previous study [8]. Each scenario set in the simulation is expected to reflect the real event of a volcanic tsunami so that the resulting database is a database that is close to real conditions so that it is accurate for model training.

B. Observation stations

We utilized four observation stations located on Java Island (Marina Jambu and Ciwandan) and Sumatera Island (Panjang and Kota Agung) to record the time series of tsunami waveforms. These stations were strategically selected to provide detailed record of tsunami waveforms and validate the accuracy of our simulation model. Initially, we validated our simulation results against the actual 2018 Anak Krakatau tsunami event to ensure the model accurately replicates the event for validating the simulation approach. This validation step was crucial to confirm the model's reliability for future predictions and to generate training and testing databases. The simulated waveforms were compared with the tsunami waveforms recorded at existing four tide gauge stations. All observed waveforms at each gauge showed a reasonable match with the simulation waveforms (Fig. 2), using one scenario with an estimated collapse volume of 0.24 km^3 [32].

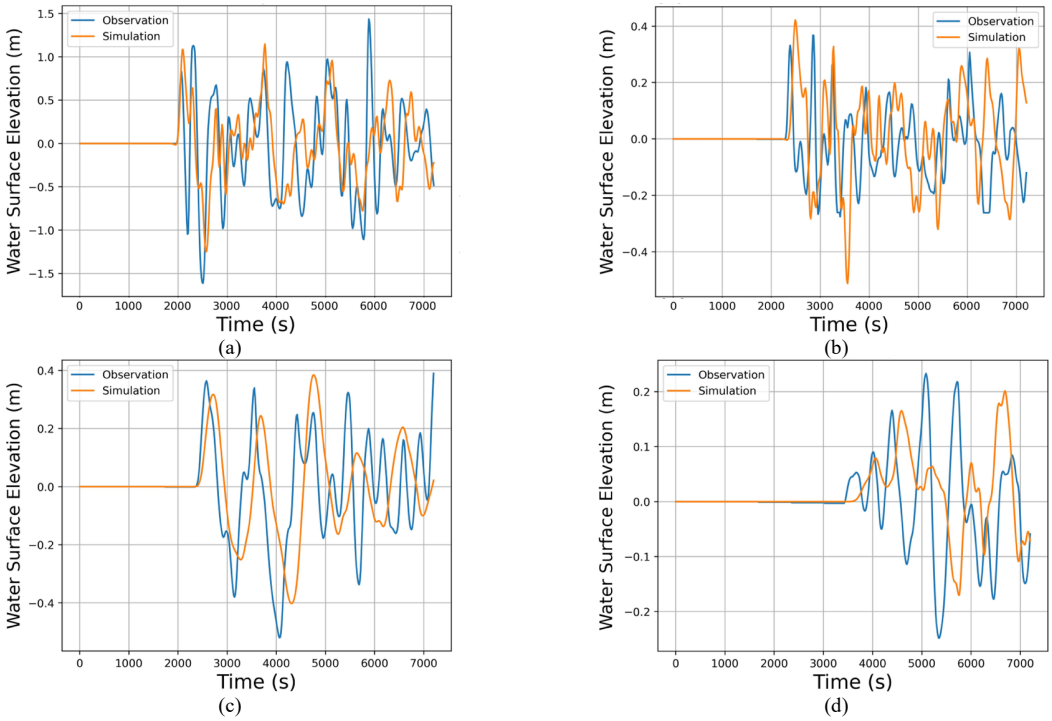

Fig. 2 Simulation test with a real event of the 2018 Anak Krakatau tsunami comparison in each station. (a) Marina Jambu (b) Ciwandan (c) Kota Agung (d) Panjang

C. Supervised database and classifier model

After confirming that the simulation results reasonably match with the observed waveforms, we stored them in the database as the simulated results. These simulation results were used to construct a supervised database, forming the foundation for the decision support model. In this study, a supervised database refers to a dataset specifically structured for use in supervised learning. It comprises input features derived from simulation results, such as maximum wave elevation and collapse thickness, paired with labeled outputs representing tsunami warning levels. These labels enable a classifier model to learn the relationship between the input attributes and the correct warning levels. The classifier was trained using this supervised database, utilizing the learned relationship to predict warning levels from new simulation outputs by determining the appropriate warning level based on the key attributes (Table 2). The definition

of warning levels was based on the guidelines from the Indonesian Meteorology, Climatology, and Geophysical Agency [34] as shown in Table 3: No Warning if wave height < 0.1 meter, Minor Tsunami issued if wave height range between 0.1 and < 0.5 meter, Tsunami if wave height between 0.5 and < 3.0 meter, and Major Tsunami if wave height range ≥ 3.0 meter.

TABLE 2
SUPERVISED DATABASE WITH WARNING LEVEL FOR CLASSIFIER MODEL

Scenario	Collapse Direction	Angle Degree	Thickness (m)	Length (m)	Width (m)	Volume (km^3)	Worst Elevation (WE)	Time at WE (min)	Warning Level
1	East	9°	176	2906	78.2	0.04	0.27	49.72	Minor Tsunami
22	north	9°	62	1250	903.2	0.07	0.18	41.00	Minor Tsunami
43	northeast	9°	127	1395	846.7	0.15	0.76	41.13	Tsunami
66	northwest	9°	130	3288	444.5	0.19	1.30	32.94	Tsunami
91	south	9°	127	3490	541.5	0.24	1.14	41.32	Tsunami
114	southeast	9°	63	1499	2859.0	0.27	1.24	32.85	Tsunami
137	southwest	9°	43	1439	4848.3	0.30	0.76	41.62	Tsunami
160	west	9°	106	3727	835.3	0.33	0.83	42.72	Tsunami

TABLE 3
DEFINE WARNING LEVELS REFER TO INDONESIA METEOROLOGY, CLIMATOLOGY, AND GEOPHYSICAL AGENCY [34]

Warning Level	Wave Height (WH) Range [m]
No Warning	0.0 = WH < 0.1
Minor Tsunami	0.1 = WH < 0.5
Tsunami	0.5 = WH <3.0
Major Tsunami	WH ≥ 3.0

D. Decision tree algorithm

The Decision Tree is a supervised learning method applicable to both classification and regression tasks, primarily used for classification problems. This method employs a tree-structured model where internal nodes signify dataset features, branches indicate decision rule, and leaf nodes signify outcomes. Within a Decision Tree, there are two types of nodes: Decision Nodes facilitate decision-making and have multiple branches, while Leaf Nodes represent final outcomes without further branching. Decisions within the tree are based on dataset features, presenting a graphical representation of all potential solutions or decisions given specific conditions. The tree starts with a root node and branches out, forming a hierarchical structure, hence the name 'Decision Tree'.

The tree is constructed using the Classification and Regression Tree (CART) algorithm, which iteratively splits the dataset based on specific criteria. The Decision Tree algorithm operates by asking a series of yes/no questions and splitting the tree into sub-trees based on the answers (Fig. 3). Decision Trees are preferred for several reasons. Firstly, they mimic human decision-making processes, making them easy to understand [35]. Secondly, their logic is transparent and straightforward to interpret due to their tree-like structure [36]. To predict the class of a given dataset, the Decision Tree algorithm starts at the root node and compares the root attribute values with the actual dataset attributes. Based on this comparison, the algorithm follows the corresponding branch to the next node. This process continues recursively until it reaches a leaf node.

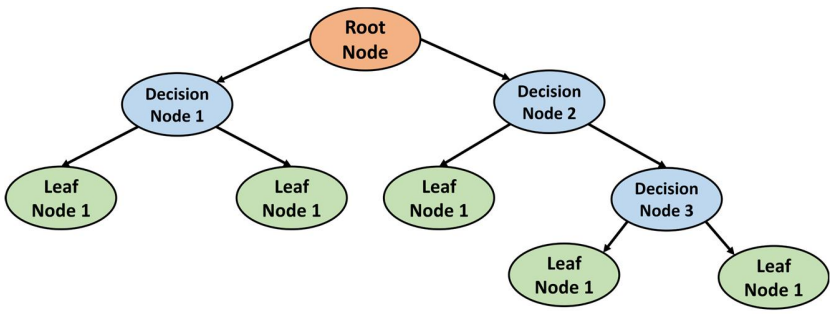

Fig. 3 Decision tree model

Attribute Selection Measures (ASMs) are crucial for determining the best attribute for the root and sub-nodes. Two popular ASMs are Information Gain and the Gini Index. Information Gain helps estimate the change in entropy following dividing the data set based on the attributes present. It examines the sensitivity of information given to a feature present about a class. The splitting in the Decision Tree algorithm starts with the highest gain first in order to reach the maximum Information Gain. The calculation of how Information Gain and entropy occurred in the Decision Tree are shown in equation 5-6.

$$Information\ Gain = Entropy\ (S) - (Weighted\ Avg * Entropy\ of\ each\ feature) \quad (5)$$

$$Entropy\ (S) = -P(yes) \log_2 P\ (yes) - P(no) \log_2 P(no) \quad (6)$$

On the other hand, the Gini Index is used to measure how good and optimal a node is in selecting features and determining the separation point based on its target class. Attributes with a small Gini Index indicate something better (purer) and, it is calculated as shown in equation 7 below.

$$Gini\ Index = 1 - \sum jP_j^2 \quad (7)$$

E. Evaluation method

In using machine learning algorithms, evaluation methods are very important to measure the success of a model, in this case the ability to classify. In the case of Decision Tree, confusion matrices, F1 score, macro, and weighted average accuracies are used as evaluation methods. Confusion metrics will evaluate the comparison of actual versus predicted tsunami warning levels at four coastal stations. While F1 is used to assess the accuracy of the model for each class, especially when imbalanced datasets occur. To treat all classes equally, macro accuracy is used, and to find out the frequency level in each class, weighted average accuracy is applied. The metrics mentioned above are expected to be able and well proper to evaluate the performance of the classification model for volcanic tsunami early warning level.

III. RESULTS

This result session will display tsunami waveform timeseries at each coastal tide gauge station along with information related to the specific time the tsunami will arrive at the coast and its worst elevation information. In addition, the calculation results of the decision tree algorithm and evaluation of the model classification performance will be visualized.

A. Tsunami waveform on each gauge or station

Tsunami waves resulting from 160 scenario simulation datasets were recorded at four coastal tide gauges, namely Marina Jambu, Ciwandan, Kota Agung, and Panjang. These four gauges are located on the islands of Java and Sumatra in the Sunda Strait. Marina Jambu and Ciwandan are located on the coast of Java (Banten) and Kota Agung and Panjang are located on the southern coast of Sumatra. The recorded tsunami wave recordings are important as input information called water surface elevation (WSE). The waveform is then visualized into a plot that depicts the waveform over time indicates tsunami behavior and its potential impact on coastal regions (Fig. 4).

At Marina Jambu tide gauge, the tsunami waveform is the most important to be noticed. By searching in the database, scenario 109 becomes the very first significant elevation that surpasses the 0.01-meter threshold occurring at around 28.72 minutes. This timing shows at what minute in the worst-case scenario from the database, the tsunami starts to reach the coastal area and causes coastal inundation (indicated by the red dashed vertical line). This early detection is critical, given that Marina Jambu is one of the closest points on the coast of Java to the tsunami source. The relatively short time frame of approximately 27-28 minutes before the WSE exceeds the threshold underscores the urgent need for rapid evacuation and early warning systems in this area. The Ciwandan gauge shows a similar pattern, with the first elevation surpassing the 0.01-meter threshold at 35.15 minutes in scenario 109. The data indicates that while Ciwandan has a slightly longer lead time compared to Marina Jambu, the difference is not substantial, highlighting the need for swift action in both locations.

In Kota Agung, the first significant elevation above the 0.01-meter threshold occurs later, at 40.04 minutes in scenario 140, as marked by the red dashed vertical line in the plot. This delay compared to the Java coastal gauges suggests that communities in Sumatra, such as Kota Agung, have a slightly longer window to respond to tsunami warnings. However, this extended time frame should not lead to complacency, as the waveform data still indicates a pressing need for efficient and effective evacuation procedures. For Panjang gauge, the first elevation surpassing the 0.01-meter threshold occurs at 59.82 minutes in scenario 38. The waveform data here shows that Panjang, while

experiencing the longest lead time among the four gauges, still faces significant tsunami risks that require prompt and coordinated response efforts.

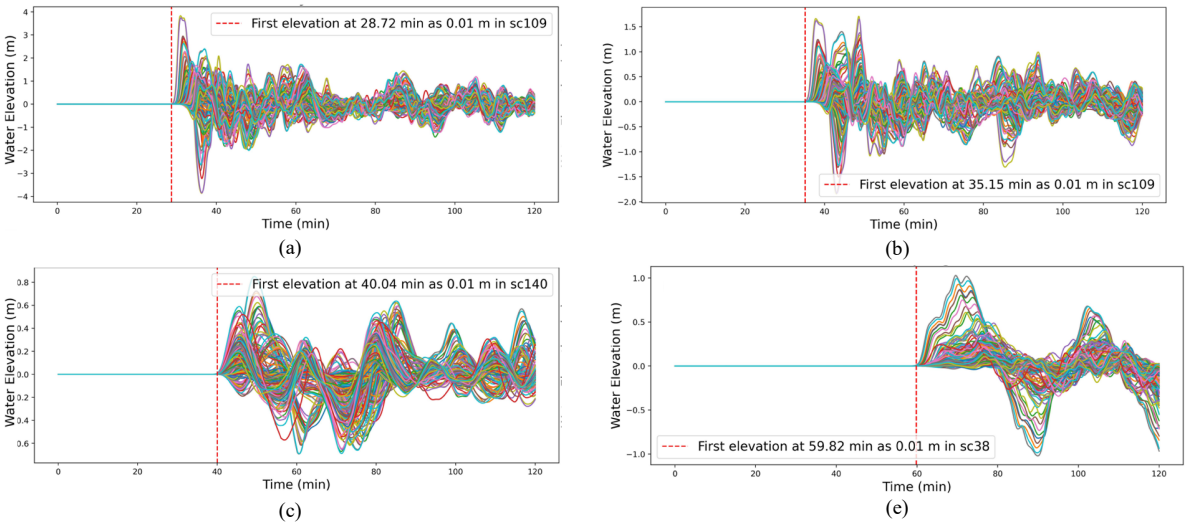

Fig. 4 Waveforms of water elevation at four tide gauges during multiple scenarios. (a) Marina Jambu gauge (b) Ciwandan gauge (c) Kota Agung gauge (d) Panjang gauge

Each subplot in the visualization not only highlights the initial exceedance of the 0.01-meter threshold but also depicts the overall dynamics of water surface elevation changes across different coastal locations and scenarios. These detailed waveforms offer a clear and comprehensive picture of how tsunamis evolve and impact various regions, providing essential data for improving tsunami early warning systems and emergency response strategies.

B. Specific time for worst elevation of tsunami waveform on each gauge

The visualization of tsunami waveform data from 160 scenarios at the four coastal gauges—Marina Jambu, Ciwandan, Kota Agung, and Panjang—provides a detailed look at the specific time intervals and maximum elevations recorded at these locations. Each subplot in the image represents one gauge and plots the Water Surface Elevation (WSE) in meters against time in minutes (Fig. 5). This detailed representation is crucial for decision-making when determining warning levels, as it helps pinpoint the periods of highest tsunami elevation.

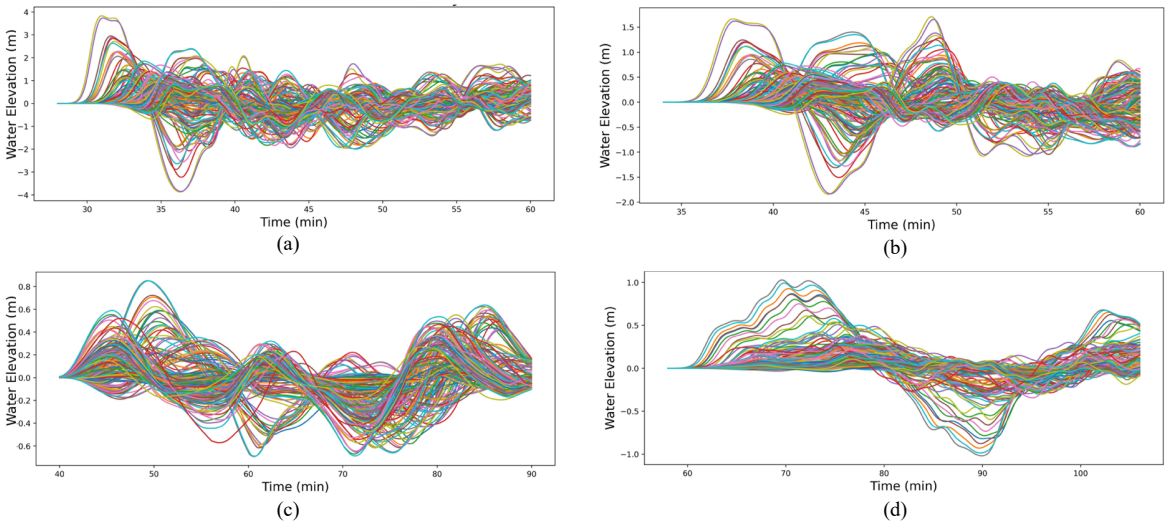

Fig. 5 Specific time and worst elevation when reached coastal gauges (a) Marina Jambu record in 26-60 min (b) Ciwandan record in 34-60 min (c) Kota Agung record in 50-90 min (d) Panjang record in 58-105 min

For Marina Jambu, the worst elevation occurs between 26 and 60 minutes. This is the fastest time span for a tsunami to reach the coast compared to the other three coastal gauges because it arrived in less than 30 minutes. Similarly, the Ciwandan gauge records the worst elevation between 34 and 60 minutes. The waveform data within this timeframe emphasizes significant peaks in water surface elevation, underscoring the importance of this interval as tsunami worst elevation input into the database. This targeted approach to warning and evacuation can significantly enhance the effectiveness of tsunami response efforts. For Kota Agung, the worst elevation occurs between 40 and 90 minutes. The wave shape recorded in Kota Agung tends to be similar to Marina Jambu with dynamic up and down fluctuations even though the height is below one meter. For Panjang gauge, records the worst elevation between 58 and 105 minutes with the average waveform in a flat form.

The plot visualization of waveform data on each coastal gauge above revealed the specific time intervals for deciding the tsunami worst elevations. By identifying the periods when the WSE is highest, the early warning model can be developed as one of the significant parameters of warning level classification. This targeted approach ensures that warning issue strategies are timely and effective, reducing the potential impact of tsunamis on coastal communities.

C. Decision tree model visualization

Figure 6 and 7 illustrate the results of decision tree classification applied to data from four coastal tide gauges: Kota Agung, Panjang, Ciwandan, and Marina Jambu. Each decision tree represents a series of splits based on different features related to tidal measurements, ultimately categorizing the potential tsunami warning levels at each gauge.

As shown in Fig. 6, for the Kota Agung gauge (a), the decision tree classifier begins by splitting based on the feature highest_kota_agung. If this value is less than or equal to 0.099, the event is classified as a "Minor Tsunami" with a Gini impurity of 0.428, indicating some uncertainty in the classification. This branch further subdivides into pure nodes (Gini = 0) that classify as "No Warning" if the specified criterion is met. On the other hand, if highest_kota_agung is greater than 0.099, the model results in further subdivisions, eventually classifying the observations as either "Minor Tsunami" or "Tsunami." For the Panjang gauge (b), the initial split is also based on the feature highest_panjang. If then rule also applied to the value is less than or equal to 0.099, then the event is classified as a "Minor Tsunami". The next splitting on highest_panjang with different thresholds and samples leads to three classifications namely "No Warning," "Minor Tsunami," and "Tsunami". Naturally, the nodes will become purer as well as the Gini values decrease, demonstrating to become more confidence in classifications at the leaves of the tree such as "Minor Tsunami" and "Tsunami".

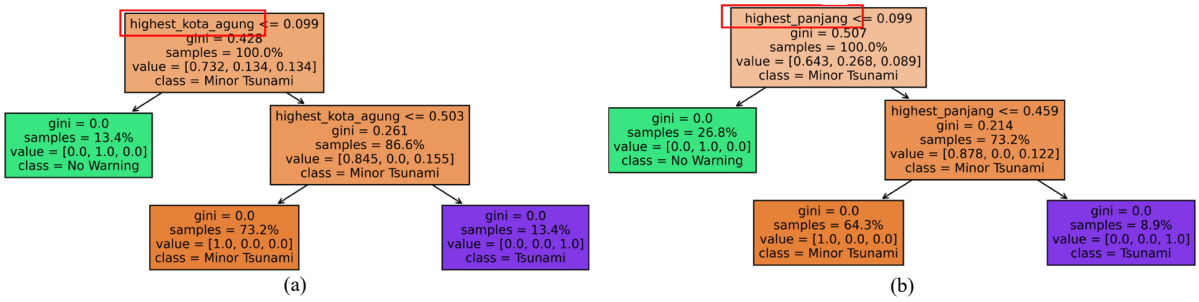

Fig. 6 Decision tree classification results (a) Kota Agung gauge; (b) Panjang gauge

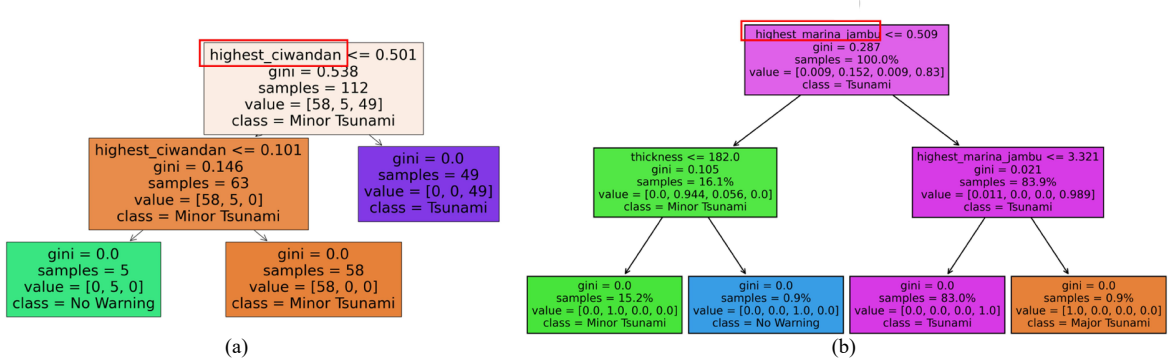

Fig. 7 Decision tree classification results. (a) Ciwandan gauge (b) Marina Jambu gauge

Based on Fig. 7, at Ciwandan gauge (a), the decision tree classifier starts the process with highest_ciwandan as a feature. Then splitting at 0.501 is categorized as "Minor Tsunami". More splitting based on the different thresholds classify into either "No Warning" or "Minor Tsunami". For Marina Jambu gauge (b), similar to Ciwandan, the decision tree begins with the feature highest_marina_jambu. An initial split at 0.509 "Tsunami" is the most important feature. The next splitting leads to a "Minor Tsunami" or "Major Tsunami" based on the threshold.

The decision tree classifiers for each gauge demonstrate a logical approach to categorizing tsunami early warning levels. In addition, from all four gauges visualized above, the decision trees revealed that the highest of waveform becomes a primary feature over the other features such as thickness. In order to improve the classification, trees regularly split on important features at different thresholds to find the lowest gini impurity. As a result, a consistent reduction of impurity confirms that the model predictions become increasingly accurate and reliable. The classifications range from "No Warning" to "Minor Tsunami," "Tsunami," and "Major Tsunami," demonstrating how various feature values determine these results. This structured approach allows the model to handle the complexity of varying tsunami threats, enhancing its reliability and effectiveness for early warning.

D. Model evaluation: confusion matrix, f1, macro, and weighted average accuracies score

Figure 8 below shows the evaluation results of decision tree classification using confusion matrices at four coastal stations. As widely used for assessing classification performance, confusion matrices provide detailed insights into the number of true and false predictions produced by the model, categorized by the actual and predicted classes.

Based on Fig. 8, the confusion matrix on Marina Jambu gauge (a) shows the predictions result by the model. It correctly predicted 1 "Major Tsunami", 6 "Minor Tsunami", and 40 "Tsunami" cases. Only 1 misclassification in this gauge that predicted "Minor Tsunami" that should be "No Warning". Similarly, the Ciwandan gauge (b) revealed 24 correct predictions for "Minor Tsunami", 23 correct predictions for "Tsunami" and 1 correct for "No Warning". There are no misclassifications in this gauge. Overall, confusion metrics for Marina Jambu and Ciwandan demonstrated the model's proficiency in classifying between "Minor Tsunami" and "Tsunami" with high accuracy, despite a single misclassification suggesting some room for improvement.

The confusion matrix for the Kota Agung gauge (c) shows high accuracy similar to the Ciwandan gauge previously, with 36 correct predictions for "Minor Tsunami" 6 predictions accurate for "Tsunami", 6 correct predictions for "No Warning". Importantly, there are no misclassifications in this gauge. For the Panjang gauge (d), the matrix shows 28 correct predictions for "Minor Tsunami," 15 correct for "No Warning," and 4 for "Tsunami". However, there is 1 misclassification where a "Minor Tsunami" is predicted as "Tsunami".

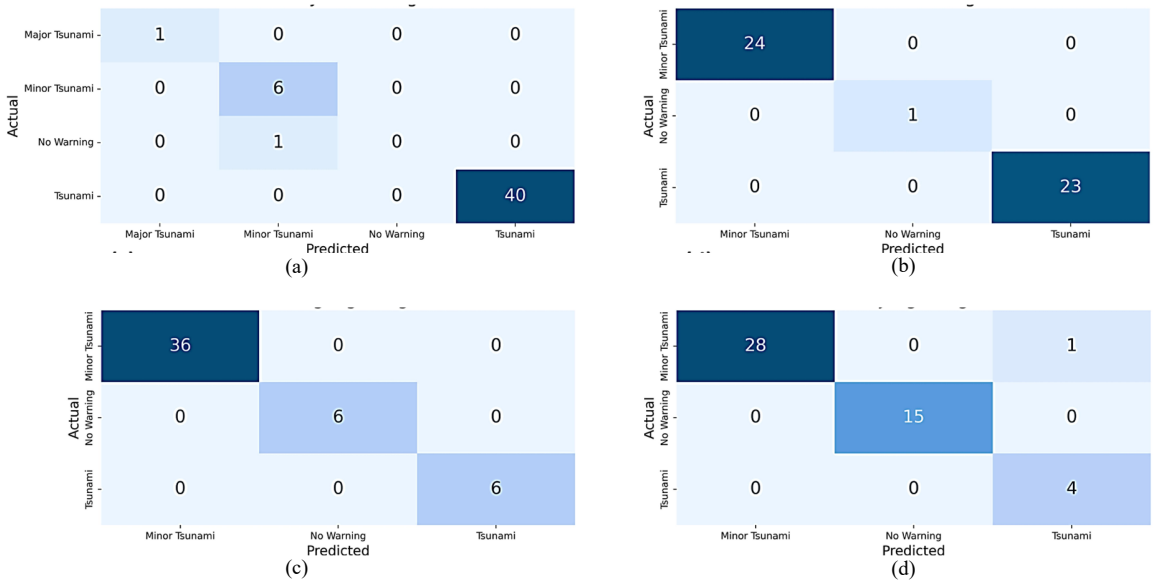

Fig. 8 Model evaluation on each gauge. (a) confusion matrix on Marina Jambu (b) confusion matrix on Ciwandan (c) confusion matrix on Kota Agung (d) confusion matrix on Panjang

Furthermore, F1, micro, and weighted average accuracy scores were used for evaluation (Table 4). These metrics provided an understanding of model performance by combining precision and recall into a single score for each class. For the Marina Jambu gauge, the F1 scores are as follows: 0.00 for "No Warning," 0.92 for "Minor Tsunami," 1.00

for "Tsunami," and 1.00 for "Major Tsunami." The macro average of 0.73 and the weighted average of 0.97. These high scores indicate the model's strong performance in predicting "Minor Tsunami," "Tsunami," and "Major Tsunami" events. However, the score of 0.00 for "No Warning" reflects significant challenges in predicting this class, consistent with the previously noted misclassifications. For the Ciwandan gauge, the F1 scores are 1.00 for both "No Warning" and "Minor Tsunami," and 1.00 for "Tsunami." Since the gauge does not have data for "Major Tsunami", no score is provided for that class. Both macro and weighted averages are also at 1.00. This perfect score demonstrates the model's exceptional capability in predicting all relevant classes for this station, with no errors recorded in the confusion matrix.

For the Kota Agung gauge, the F1 scores are 0.00 for "No Warning," 0.92 for "Minor Tsunami," 1.00 for "Tsunami," and 1.00 for "Major Tsunami." Similar to Marina Jambu, the macro average of 0.73 and a weighted average of 0.97. While the model effectively predicts "Minor Tsunami," "Tsunami," and "Major Tsunami" events, it struggles significantly with the "No Warning" class, as indicated by the score of 0.00. Moving to the Panjang gauge, the F1 scores are 1.00 for "No Warning", 0.98 for "Minor Tsunami", and 0.89 for "Tsunami". Since there is no data for "Major Tsunami", no score is provided for that class. The macro average of 0.96 and the weighted average of 0.96. These scores indicate a generally strong performance, with the model performing slightly less effectively for the "Tsunami" class compared to the other gauges.

TABLE 4
F1-SCORE, MACRO AND WEIGHTED AVERAGE ACCURACIES OF MODEL EVALUATION METRICS

	Marina Jambu	Ciwandan	Kota Agung	Panjang
F1-score				
No warning	0.00	1.00	0.00	1.00
Minor tsunami	0.92	1.00	0.92	0.98
Tsunami	1.00	1.00	1.00	0.89
Major tsunami	1.00	-	1.00	-
Accuracy				
Macro average	0.73	1.00	0.73	0.96
Weighted average	0.97	1.00	0.97	0.96

IV. DISCUSSION

This study simulating a 160-scenario dataset encompassing all possible collapse directions and varying volumes has proven to be a suitable approach in anticipating future events similar to the 2018 Anak Krakatau Volcanic tsunami. Based on this database, the selected decision tree algorithm is able to classify the warning level of volcanic tsunamis, enhancing its predictive capabilities and solving challenges for early warning systems as will be discussed below.

A. Analysis of significant features and influences from geographical context

The results of the study show that the integration of numerical simulations and Decision Tree classification is effective in predicting tsunami warning levels. The input features in the dataset included collapse thickness, maximum amplitude on the coast, collapse volume, direction, dip, width, length, and time to maximum wave elevation. The study finding is the identification of maximum wave amplitude (highest waveform) as the most significant feature followed by collapse thickness by referring to information gain and gini index in the Decision Tree model. Information gain measures the reduction in uncertainty (entropy) when the dataset is split while gini index measures the purity of the splits generated by an attribute. Information gain and Gini index procedurally evaluated each feature contribution to improving the predictive accuracy. Information gain was calculated by assessing the entropy before and after the split, with attributes like maximum amplitude and collapse thickness. This result indicates that these two features were the most critical in distinguishing between tsunami events and non-tsunami or minor tsunamis or even no warning.

Regarding Gini index, in this model, maximum amplitude achieved the lowest gini values at most of the tide gauges showing they produced the purest splits. While other attributes, such as collapse volume, are also relevant, they contributed less significantly to only on Marina Jambu gauge. The reason because that Marina Jambu has many input features since this gauge is the closest to the volcano. By focusing on these significant attributes, the Decision Tree model was able to effectively predict whether a tsunami would occur and classify the warning level. For example, the high predictive accuracy at Marina Jambu and Ciwandan reflects the shorter tsunami arrival times in these two locations, due to their closeness to the volcanic collapse. This also discovered the new insight into how the geographical context plays a role in the model predictions, providing information for future evacuation planning.

B. Model performance across different locations

The model performance differed among various coastal gauges, mainly due to the geographical context of the collapse site as previously mentioned. Marina Jambu and Ciwandan on Java Island revealed faster tsunami arrival times, with Marina Jambu having a critical time of less than 30 minutes. The reason because these gauges are closer to the Anak Krakatau volcanic collapse site. On the other hand, gauges on Sumatra Island, such as Kota Agung, had longer lead times of about 40 minutes due to their greater distance from the collapsed volcano. It emphasizes the importance of geographic context when issuing tsunami warnings such as distance from the collapse site. Moreover, the simulation not only identifies the arrival time but also the worst-case wave elevations. For example, Ciwandan had a time for the worst wave heights between 34 and 60 minutes, while Panjang in Sumatra saw these between 60 and 75 minutes. This information can help local decision makers with evacuation planning based on when the highest risk and time of arrival is expected.

C. Class imbalance during the model learning

Another challenge of the model is the imbalance class which had limited representation between the different categories in the dataset, especially "No Warning" category. The class imbalance problem appears when tsunamis events dominate the dataset while no warning occusr less frequently. This imbalance can lead to bias toward predicting the more common class such as "Tsunami", "Major Tsunami", and "Minor Tsunami" while struggling to predict the less frequent "No Warning" class. Even though the dataset used in this study included a variety of scenarios, there was less observance of the "No Warning" category. This made it difficult for the model to learn from these examples and resulted in weaker performance in predicting when no tsunami would occur. Oversampling or undersampling techniques can be used to address this issue. Alternatively, cost-sensitive learning can be implemented, where the model is penalized more heavily for misclassifying the minority class. This forces the model to pay closer attention to the underrepresented class, improving its ability to correctly predict less frequent events like "No Warning."

D. Comparative analysis with other studies

To point out the strength of this model, comparison to other studies using machine learning models such as Regression Trees and integrated networks in tsunami early warning systems are conducted. For example, [37] demonstrated the use of regression trees for tsunami wave forecasting, emphasizing the interpretability and computational efficiency of decision-tree-based methods. This aligns with our findings, where the Decision Tree algorithm enhances interpretability and provides real-time classifications, offering advantages over other approaches that struggle with high training times and limited transparency, such as neural networks. Additionally, [38] proposed a real-time tsunami monitoring system using submerged mooring systems, which we have referenced to highlight the value of combining real-time data collection with machine learning models like ours to improve tsunami early warning systems.

E. Model improvement strategies

To further improve the performance of the Decision Tree model in handling class imbalance, several strategies can be implemented. Ensemble methods such as boosting and bagging are highly effective techniques for improving model robustness. Boosting trains several weak classifiers, each one correcting the mistakes of the previous, then improves the models ability to handle hard-to-predict cases like the "No Warning" category. Furthermore, in the case of a tsunami dataset, bagging can be used by training multiple models on random portions of the data, such as wave height or time intervals to reduce errors and avoid overfitting to specific tsunami scenarios. By integrating the predictions of multiple classifiers, ensemble methods can handle the class imbalance of challenge. Besides to ensemble methods, adding another significant feature dataset in the tsunami domain could enhance its predictive capabilities. For example, seismic activity, atmospheric pressure, or collapse precursor signs may provide the model with a wider perspective for predicting non-tsunami events.

V. CONCLUSIONS

The decision tree classifier model used in this study presents a solution to dealing with early warning decisions, especially for non-seismic tsunamis such as volcanic tsunamis. Volcanic tsunamis, caused by events like submarine and subaerial collapse, pose a distinct challenge for early warning systems. Unlike seismic tsunamis that can be identified through seismic activity monitoring, volcanic tsunamis demand a different early warning strategy because of their unique detection causes. Given the current lack of early warning for such events, this study tried to fill a critical gap in preliminary non-seismic early warning system research by providing a classification warning model for anticipating and responding to decision-making. The model addresses this challenge by simulating different collapse

scenarios and storing the resulting data in a comprehensive database for use in a model classifier. As a result, the model accuracy performance shows the reliability and effectiveness of the classification.

Moreover, to enhance the robustness and reliability of the decision model, it is important to explore additional collapse parameters, such as seismic activity and collapse precursor signs. Including these parameters can lead to a more detailed and accurate model that can handle a wider range of cases and give better predictions. A stronger grasp of these factors will help build a full and effective early warning system. For example, the seismic activity affects the collapse timing, and knowing this link can sharpen the model ability to predict. Furthermore, it is also important to continue developing models and frameworks that integrate technical early warning systems with robust decision-making processes. The current model is capable of classifying the warning level with high accuracy. This ability could be integrated into the decision-making process to provide a more effective and timely early warning system for volcanic tsunamis.

Additionally, improving the performance in predicting the "No Warning" class is also an important area for model development. To cope with this situation, it is essential to balance the dataset by adding more scenarios that simulate very small volume collapses, which may trigger the "No Warning" classification in gauge recordings. Training the model with a more balanced dataset, can strengthen the ability to accurately "No Warning" class. Balancing the dataset prevents bias toward frequent tsunami events, enabling the model to effectively identify all potential scenarios. Moreover, training the model capabilities to handle outlier volumes, an exceptionally, large or small collapse, is very important. Outliers can significantly impact the model predictive accuracy. Ensuring the model can effectively account for these extremes in all collapse possibilities.

In conclusion, this study highlights the contribution of machine learning algorithm application, specifically the Decision Tree, in classifying tsunami early warning levels based on non-seismic tsunami or volcanic tsunamis. Furthermore, it also advances by integrating numerical simulations with machine learning to improve real-time prediction and classification of tsunami warning levels. Although this research primarily focuses on the case study of a tsunami triggered by the Anak Krakatau volcano, its implications extend beyond this specific event. Anak Krakatau is only one of several key volcanoes in Southeast Asia identified as potential sources of subaerial volcanic tsunami hazards. Other significant volcanoes include Ritter Island in Papua New Guinea, Lliwerung in Indonesia, and Didicas in the Philippines. The methodology can be adapted to other volcanic settings, providing a foundation for enhancing tsunami early warning systems in regions with similar risks of collapse-induced tsunamis.

Author Contributions: *Elmo Juanara*: Conceptualization, Methodology, Software, Data Curation, Writing - Original Draft. *Chi Yung Lam*: Conceptualization, Writing - Original Draft – Supervision.

All authors have read and agreed to the published version of the manuscript.

Funding: This research received no specific grant from any funding agency.

Conflicts of Interest: The authors declare no conflict of interest.

Data Availability: The software used in this paper are available at [27]. The code presented in this study are available on github online repository https://github.com/elmojuanara/warning_level_classification.

Institutional Review Board Statement: Not applicable.

Informed Consent: There were no human subjects.

Animal Subjects: There were no animal subjects.

ORCID:
Elmo Juanara: https://orcid.org/0000-0003-1224-2405
Chi Yung Lam: https://orcid.org/0000-0003-2124-2386

REFERENCES

[1] B. Lee, F. Preston, and G. Green, "Preparing for high-impact, low-probability events," Chatham House, London, 2012.
[2] L. Lakshita and N. K. C. Nair, "High impact low probability weatherization impact analysis for electricity infrastructure," in *IEEE Region 10 Annual International Conference, Proceedings/TENCON*, Institute of Electrical and Electronics Engineers Inc., 2021, pp. 958–963. doi:

10.1109/TENCON54134.2021.9707369.
[3] J. Madeira, R. S. Ramalho, D. L. Hoffmann, J. Mata, and M. Moreira, "A geological record of multiple Pleistocene tsunami inundations in an oceanic island: The case of Maio, Cape Verde," *Sedimentology*, vol. 67, no. 3, pp. 1529–1552, Apr. 2020, doi: 10.1111/sed.12612.
[4] R. Paris *et al.*, "Volcanic tsunami: A review of source mechanisms, past events and hazards in Southeast Asia (Indonesia, Philippines, Papua New Guinea)," *Nat. Hazards*, vol. 70, no. 1, pp. 447–470, Jan. 2014, doi: 10.1007/s11069-013-0822-8.
[5] S. J. Day, "Volcanic tsunamis," in *The Encyclopedia of Volcanoes*, Second Edi., London: Elsevier, 2015, pp. 993–1009. doi: 10.1016/B978-0-12-385938-9.00058-4.
[6] S. T. Grilli *et al.*, "Modelling of the tsunami from the December 22, 2018 lateral collapse of Anak Krakatau Volcano in the Sunda Straits, Indonesia," *Sci. Rep.*, vol. 9, no. 1, p. 11946, Dec. 2019, doi: 10.1038/s41598-019-48327-6.
[7] T. R. Walter *et al.*, "Complex hazard cascade culminating in the Anak Krakatau sector collapse," *Nat. Commun.*, vol. 10, no. 1, p. 4339, Dec. 2019, doi: 10.1038/s41467-019-12284-5.
[8] M. Heidarzadeh, T. Ishibe, O. Sandanbata, A. Muhari, and A. B. Wijanarto, "Numerical modeling of the subaerial landslide source of the 22 December 2018 Anak Krakatoa volcanic tsunami, Indonesia," *Ocean Eng.*, vol. 195, p. 106733, Jan. 2020, doi: 10.1016/J.OCEANENG.2019.106733.
[9] M. Heidarzadeh, P. S. Putra, S. H. Nugroho, and D. B. Z. Rashid, "Field survey of tsunami heights and runups following the 22 December 2018 Anak Krakatau volcano tsunami, Indonesia," *Pure Appl. Geophys.*, vol. 177, no. 10, pp. 4577–4595, Oct. 2020, doi: 10.1007/s00024-020-02587-w.
[10] A. Muhari *et al.*, "The December 2018 Anak Krakatau volcano tsunami as inferred from post-tsunami field surveys and spectral analysis," *Pure Appl. Geophys.*, vol. 176, no. 12, pp. 5219–5233, Dec. 2019, doi: 10.1007/s00024-019-02358-2.
[11] C. Deplus, S. Bonvalot, D. Dahrin, M. Diament, H. Harjono, and J. Dubois, "Inner structure of the Krakatau Volcanic complex (Indonesia) from gravity and bathymetry data," *J. Volcanol. Geotherm. Res.*, vol. 64, no. 1–2, pp. 23–52, Feb. 1995, doi: 10.1016/0377-0273(94)00038-I.
[12] United Nations Office for Disaster Risk Reduction (UNDRR), "Words into action: A Guide to multi-hazard early warning systems," *United Nations Office for Disaster Risk Reduction (UNDRR)*, Geneva, Switzerland, 2023. Accessed: May 30, 2024. [Online]. Available: https://www.undrr.org/words-into-action/guide-multi-hazard-early-warning?sfnsn=scwspmo
[13] J. Selva *et al.*, "Tsunami risk management for crustal earthquakes and non-seismic sources in Italy," *La Riv. del Nuovo Cim.*, vol. 44, no. 2, pp. 69–144, Feb. 2021, doi: 10.1007/s40766-021-00016-9.
[14] V. V Titov *et al.*, "Real-time tsunami forecasting: challenges and solutions," *Nat. Hazards*, vol. 35, no. 1, pp. 35–41, May 2005, doi: 10.1007/s11069-004-2403-3.
[15] A. Perttu *et al.*, "Reconstruction of the 2018 tsunamigenic flank collapse and eruptive activity at Anak Krakatau based on eyewitness reports, seismo-acoustic and satellite observations," *Earth Planet. Sci. Lett.*, vol. 541, p. 116268, Jul. 2020, doi: 10.1016/J.EPSL.2020.116268.
[16] R. Sabeti and M. Heidarzadeh, "Estimating maximum initial wave amplitude of subaerial landslide tsunamis: A three-dimensional modelling approach," *Ocean Model.*, vol. 189, p. 102360, Jun. 2024, doi: 10.1016/j.ocemod.2024.102360.
[17] R. Omira and I. Ramalho, "Evidence-calibrated numerical model of December 22, 2018, Anak Krakatau flank collapse and tsunami," *Pure Appl. Geophys.*, vol. 177, no. 7, pp. 3059–3071, Jul. 2020, doi: 10.1007/s00024-020-02532-x.
[18] J. P. Terry, J. Goff, N. Winspear, V. P. Bongolan, and S. Fisher, "Tonga volcanic eruption and tsunami, January 2022: globally the most significant opportunity to observe an explosive and tsunamigenic submarine eruption since AD 1883 Krakatau," *Geosci. Lett.*, vol. 9, no. 1, Dec. 2022, doi: 10.1186/s40562-022-00232-z.
[19] S. T. Grilli *et al.*, "Modeling of the Dec. 22nd 2018 Anak Krakatau Volcano lateral collapse and tsunami based on recent field surveys: Comparison with observed tsunami impact," *Mar. Geol.*, vol. 440, p. 106566, Oct. 2021, doi: 10.1016/J.MARGEO.2021.106566.
[20] K. S. Cutler *et al.*, "Downward-propagating eruption following vent unloading implies no direct magmatic trigger for the 2018 lateral collapse of Anak Krakatau," *Earth Planet. Sci. Lett.*, vol. 578, p. 117332, Jan. 2022, doi: 10.1016/J.EPSL.2021.117332.
[21] K. Sun, Z. Li, S. Wang, and R. Hu, "A support vector machine model of landslide susceptibility mapping based on hyperparameter optimization using the bayesian algorithm: a case study of the highways in the southern Qinghai–Tibet Plateau," *Nat. Hazards*, vol. 120, no. 12, pp. 11377–11398, Sep. 2024, doi: 10.1007/s11069-024-06665-3.
[22] V. Linardos, M. Drakaki, P. Tzionas, and Y. Karnavas, "Machine learning in disaster management: Recent developments in methods and applications," *Mach. Learn. Knowl. Extr.*, vol. 4, no. 2, pp. 446–473, May 2022, doi: 10.3390/make4020020.
[23] A. Gulghane, R. L. Sharma, and P. Borkar, "A formal evaluation of KNN and decision tree algorithms for waste generation prediction in residential projects: a comparative approach," *Asian J. Civ. Eng.*, vol. 25, no. 1, pp. 265–280, Jan. 2024, doi: 10.1007/s42107-023-00772-5.
[24] K. Abraham, M. Abdelwahab, and M. Abo-Zahhad, "Classification and detection of natural disasters using machine learning and deep learning techniques: A review," *Earth Sci. Informatics*, vol. 17, no. 2, pp. 869–891, Apr. 2024, doi: 10.1007/s12145-023-01205-2.
[25] M. Pota, G. Pecoraro, G. Rianna, A. Reder, M. Calvello, and M. Esposito, "Machine learning for the definition of landslide alert models: a case study in Campania region, Italy," *Discov. Artif. Intell.*, vol. 2, no. 15, Dec. 2022, doi: 10.1007/s44163-022-00033-5.
[26] M. Kappi and B. Mallikarjuna, "Artificial intelligence and machine learning for disaster prediction: A scientometric analysis of highly cited papers," *Nat. Hazards*, vol. 120, no. 12, pp. 10443–10463, Sep. 2024, doi: 10.1007/s11069-024-06616-y.
[27] P. L. F. Liu, S.-B. Woo, and Y.-S. Cho, "Computer programs for tsunami propagation and inundation," Seoul, Korea, 1998. [Online]. Available: https://tsunamiportal.nacse.org/documentation/COMCOT_tech.pdf
[28] S. M. Abadie, J. C. Harris, S. T. Grilli, and R. Fabre, "Numerical modeling of tsunami waves generated by the flank collapse of the Cumbre Vieja Volcano (La Palma, Canary Islands): Tsunami source and near field effects," *J. Geophys. Res. Ocean.*, vol. 117, no. 5, pp. 1–26, 2012, doi: 10.1029/2011JC007646.
[29] T. Esposti Ongaro *et al.*, "Modeling tsunamis generated by submarine landslides at Stromboli Volcano (Aeolian Islands, Italy): A numerical benchmark study," *Front. Earth Sci.*, vol. 9, May 2021, doi: 10.3389/feart.2021.628652.
[30] R. N. Ratnasari, Y. Tanioka, Y. Yamanaka, and I. E. Mulia, "Development of early warning system for tsunamis accompanied by collapse of Anak Krakatau volcano, Indonesia," *Front. Earth Sci.*, vol. 11, Sep. 2023, doi: 10.3389/feart.2023.1213493.
[31] T. Giachetti, R. Paris, K. Kelfoun, and B. Ontowirjo, "Tsunami hazard related to a flank collapse of Anak Krakatau Volcano, Sunda Strait, Indonesia," *Geol. Soc. Spec. Publ.*, vol. 361, no. 1, pp. 79–90, 2012, doi: 10.1144/SP361.7.
[32] I. E. Mulia, S. Watada, T. Ho, K. Satake, Y. Wang, and A. Aditiya, "Simulation of the 2018 tsunami due to the flank failure of Anak Krakatau Volcano and implication for future observing systems," *Geophys. Res. Lett.*, vol. 47, no. 14, Jul. 2020, doi: 10.1029/2020GL087334.
[33] A. Paris, P. Heinrich, R. Paris, and S. Abadie, "The December 22, 2018 Anak Krakatau, Indonesia, landslide and tsunami: Preliminary modeling results," *Pure Appl. Geophys.*, vol. 177, no. 2, pp. 571–590, Feb. 2020, doi: 10.1007/s00024-019-02394-y.

[34] N. Baumert, M. Mück, and A. Muhari, "Guideline for tsunami risk assessment in Indonesia scientific Indonesian-German working group on tsunami risk assessment," 2018. doi: 10.13140/RG.2.2.20247.11685.
[35] J. R. Quinlan, "Induction of decision trees," *Mach. Learn.*, vol. 1, no. 1, pp. 81–106, Mar. 1986, doi: 10.1007/BF00116251.
[36] B. Letham, C. Rudin, T. H. McCormick, and D. Madigan, "Interpretable classifiers using rules and bayesian analysis: Building a better stroke prediction model," *Ann. Appl. Stat.*, vol. 9, no. 3, pp. 1350–1371, Sep. 2015, doi: 10.1214/15-AOAS848.
[37] E. Cesario, S. Giampá, E. Baglione, L. Cordrie, J. Selva, and D. Talia, "Machine learning for tsunami waves forecasting using regression trees," *Big Data Res.*, vol. 36, p. 100452, May 2024, doi: 10.1016/j.bdr.2024.100452.
[38] B. Zhou *et al.*, "Development and application of a novel tsunami monitoring system based on submerged mooring," *Sensors*, vol. 24, no. 18, p. 6048, Sep. 2024, doi: 10.3390/s24186048.