Predicting the Characteristics of Tsunamis Using Machine Learning

Shuchen Lu
Montverde Academy, 17235 7th St, Montverde, FL 34756, U.S.A.

Keywords: Machine Learning, Tsunami Prediction, Data Visualization

Abstract: This study examines the pressing need to advance tsunami prediction methods, emphasizing the drawbacks of existing strategies and the potential of machine learning. Accurate forecasting is crucial for risk management because tsunamis are a global threat with brief warning times. Since existing methods—like analytical and empirical modelling—have shortcomings, new approaches need to be looked into. This study uses a large historical dataset spanning 1800 to 2024 and focuses on seismic characteristics and maximum water height. The methodology makes use of a random forest regression model that integrates machine learning, data exploration, and visualization. Among the results are informative bar charts, heat maps, interactive maps, and dependable machine-learning models with low mean square error. The discussion emphasizes the importance of specific tsunami incidents, the impact of geo-visualization on vulnerability assessment, and the efficacy of machine learning models. While acknowledging the limitations of the models, the paper emphasizes the interdisciplinary nature of the results and their practical significance for disaster management. The conclusions highlight the study's combined contributions to academic understanding and practical application, and they also project future developments in predictive models and their continuous improvement to improve response to disasters globally.

1 INTRODUCTION

Tsunamis are massive natural hazards that are primarily caused by underwater seismic activity and pose a global threat to coastal areas. Because these catastrophic events occur with little or no warning, accurate forecasting of their characteristics, particularly the maximum height of water they are likely to reach, is critical for effective risk assessment, mitigation, and preparation. The significance of this prediction stems from its ability to save lives, protect infrastructure, and facilitate timely evacuation measures.

Existing tsunami prediction methods typically rely on historical data and mathematical models that incorporate seismic parameters. Analytical models like linear wave theory provide a fundamental understanding of wave propagation, but they may oversimplify the complexities of tsunami dynamics. Empirical methods use historical data to establish relationships between seismic parameters and tsunami characteristics, but their predictive power is frequently limited by the nonlinear nature of tsunami dynamics.

Due to existing methods having limitations, innovative approaches are required to create robust predictive models using a large amount of available historical data. Machine learning, a field that excels at detecting complex patterns in large datasets, presents a promising avenue for improving tsunami prediction capabilities.

The motivation for this research stems from the need to improve current tsunami prediction methods. While there are several approaches available, including analytical models and empirical methods, their limitations highlight the critical need for more sophisticated and accurate prediction tools. This study seeks to fill a gap in existing research by investigating the use of machine learning techniques (specifically Random Forest Regression) to improve the prediction accuracy of tsunami-related maximum water height.

This paper addresses these research gaps by taking a two-pronged approach: first, by investigating historical tsunami data to gain insights into the patterns and characteristics of previous tsunami events; and second, by developing machine learning models to predict maximum water heights using seismic characteristics. This paper is divided into

Lu, S.
Predicting the Characteristics of Tsunamis Using Machine Learning.
DOI: 10.5220/0012827900004547

In *Proceedings of the 1st International Conference on Data Science and Engineering (ICDSE 2024)*, pages 293-298
ISBN: 978-989-758-690-3

several sections, including data exploration and visualization, machine learning model development, results and discussion, and conclusions. This structured approach enables a thorough examination of the historical context and machine learning's potential in improving tsunami prediction methods.

The methodology used in this study is based on a comprehensive historical dataset derived from tsunami records. The dataset was loaded into Python programming and data analysis libraries (such as Pandas, Folium, and Plotly Express), relevant features were extracted, and visual images were created to identify patterns and trends. A random forest regression model was then created, trained, and tested to predict the maximum water level height using seismic parameters.

The primary goal of this research is to contribute to the advancement of tsunami prediction methods by incorporating machine learning techniques. This study aims to provide a more accurate and adaptable tool for predicting maximum water level heights associated with tsunamis by examining historical data critically and developing predictive models. The findings of this study have the potential to significantly improve the existing understanding of tsunami dynamics, improve the effectiveness of early warning systems, and, ultimately, help mitigate tsunamis' devastating effects on coastal communities.

The limitations of current tsunami prediction methods are numerous. Analytical models may oversimplify the complexities of tsunami dynamics, producing inaccurate predictions. While empirical methods can be used to analyze historical data, they may struggle to adapt to the complex relationships between different factors. The machine learning model proposed in this study aims to address these limitations by leveraging the power of data-driven prediction to capture complex patterns that traditional methods may miss.

Machine learning, as a data-driven approach, has proven successful in a wide range of fields, including image recognition and natural language processing. Machine learning's ability to recognize complex patterns and relationships in data makes it a useful tool for improving tsunami prediction accuracy. Machine learning models can use historical data to learn subtle relationships between seismic parameters and maximum water heights, allowing them to make more accurate predictions about future events.

In conclusion, this paper introduces a novel approach to tsunami prediction that incorporates machine learning techniques. The use of data exploration, visualization, and the creation of a random forest regression model provides a comprehensive approach to understanding historical tsunami patterns and predicting maximum water heights. This study is significant because it has the potential to improve current tsunami forecasting methodologies by providing more accurate and adaptable tools. As the research respond to complex natural hazards, such innovations are critical for protecting coastal communities and mitigating the effects of these catastrophic events.

2 METHODOLOGIES

The foundation of this research lies in the careful collation of a rich dataset obtained from the National Center for Environmental Information (NCEI), a division of the National Oceanic and Atmospheric Administration (NOAA) (2023). This dataset covers a wide historical period, from 1800 to 2024, and includes a plethora of parameters critical to understanding the complexity of tsunamis. These parameters include earthquake magnitude, tsunami causes, geographic coordinates, maximum water levels, and impact statistics. The information's authenticity and reliability were ensured by using NCEI's Tsunami Event Data Portal.

The first step was to load the dataset into a Pandas DataFrame using Python code. The initial investigation involved selecting relevant columns and applying filters to include instances where the maximum water height exceeded the significant threshold of 10 meters (Parwanto 2014). As a result, a new data frame was created for visualization, focusing on key columns such as year, country, magnitude, and maximum water height.

A notable aspect of the data exploration phase was the creation of a bar chart. The visualization was created with Plotly Express and shows a ranking of the top 30 historical maximum water levels. The chart cleverly displays the relevant countries and years to show the effect of tsunamis on water table height.

To improve geographic understanding of a significant event, specifically the "1958 Lituya Bay Earthquake and Mega-tsunami" (NOAA 2024), shown as “USA_1958” in a bar chart, an interactive map was created with Folium. The map shows the location of the epicenter (marked as the center) and Lituya Bay. The interactive map depicts the spatial context, which aids in providing a nuanced visualization of the affected area.

During the exploration phase, an innovative effort was made to create thermal maps of areas with high water levels (>10 meters). Distinct map was created using latitude and longitude data to show the areas most affected by the tsunami and with the highest water table.

The transition into the machine learning domain began with the preparation of data specifically designed for predicting maximum water heights in Japanese tsunamis. Critical features such as latitude, longitude, and earthquake magnitude were retrieved from the dataset's Japanese subset (Alhamid 2022). To guarantee uniformity, these attributes underwent standard scaling, which normalized their contribution on the resulting model.

A key component of the methodology is the use of a random forest regression model to forecast the maximum water height of the Japanese tsunami (Li 2023). Using the Scikit-Learn library, the model is made up of 200 decision trees, each with its own depth and leaf node parameters. The model was meticulously trained on a subset of the data, and its performance was critically assessed on a separate test set.

The model's prediction accuracy was measured quantitatively using Mean Square Error (MSE). The MSE assessment classified the model's performance as "High MSE" "Low MSE" or "Medium MSE" based on predefined thresholds (Allen 2013). As a supplement to the quantitative assessment, a scatter plot was created to visually compare the actual and predicted maximum water level elevations.

Ensuring model robustness is critical. The methodology includes a systematic check to ensure the sample size is adequate. If the sample size for the Japanese tsunami falls below a specified threshold (10 in this case) (Abdullah 2022), a message appears indicating that the parameters must be adjusted, or the data should be combined with data from other regions.

This research is built on the confluence of data exploration and machine learning technologies (Mulia 2022). Geo-visualization techniques were seamlessly combined with statistical analysis and predictive modeling to provide an integrated framework for assessing and predicting maximum water heights connected with tsunamis.

3 RESULTS

This work uses geo-visualization and machine-learning techniques to conduct a thorough examination of historical tsunami data. Complex analysis was carried out employing two different code segments: one for exposing historical patterns of maximum water heights and the other for forecasting maximum water heights for the Japanese tsunami.

As illustrated in Figure 1, the initial data exploration resulted in a fascinating graphic. The bar chart methodically lists the top 30 tsunamis with the highest water heights in history, visually representing the tsunamis with the biggest water height impacts.

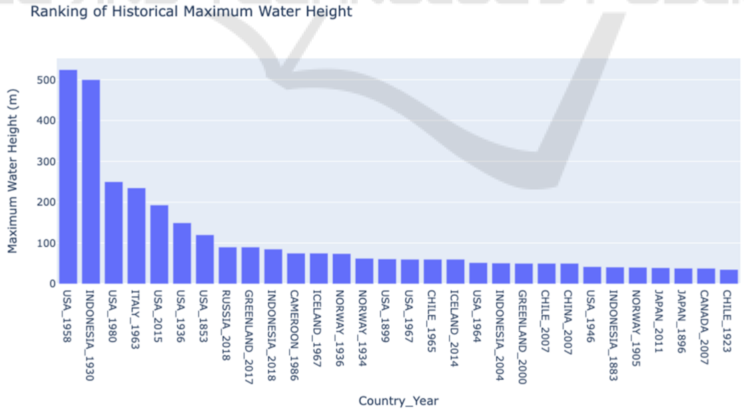

Figure 1. Bar Chart illustrating the ranking of the top 30 historical maximum water heights with associated countries and years (Photo/Picture credit: Original).

Figure 2 depicts an interactive map that shows the location and epicenter of the 1958 Lituya Bay earthquake and mega-tsunami, the event with the highest water table height (OpenStreetMap), which had the highest water level height. This spatial contextualization enables a more sophisticated view of the impacted area.

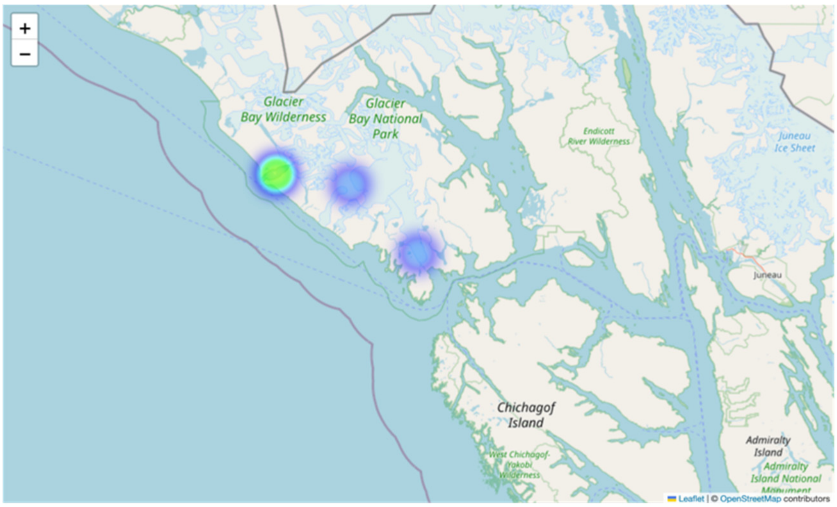
+
−
Glacier Bay Wilderness
Glacier Bay National Park
Endicott River Wilderness
Juneau Ice Sheet
Juneau
Chichagof Island
West Chichagof-Yakobi Wilderness
Admiralty Island
Admiralty Island National Monument
Leaflet | © OpenStreetMap contributors

Figure 2. Heat Map illustrating the location of 1958 Lituya Bay earthquake and mega-tsunami, the one with the greatest maximum water height, and epicenter with the use of OpenStreetMap (Photo/Picture credit: Original).

Figure 3 broadens geographic inquiry by providing heat maps that vividly depict regions with maximum water levels greater than 10 meters high. This map is an effective tool for determining which areas were most affected by the tsunami.

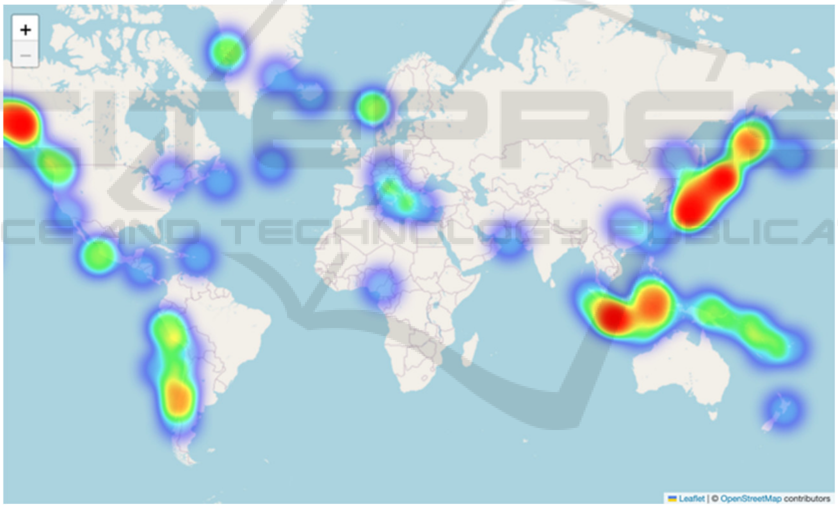

Figure 3. Heat Map highlighting locations with maximum water heights exceeding 10 meters (Photo/Picture credit: Original).

The machine learning endeavor, illustrated in Figure 4, entailed creating a Random Forest Regression model to forecast the highest water levels associated with tsunamis in Japan. The scatter plot revealed a close alignment of actual and projected values, with a commendably low MSE of 27.04. This low MSE classifies the model's performance as "Low MSE," implying strong predictive capabilities.

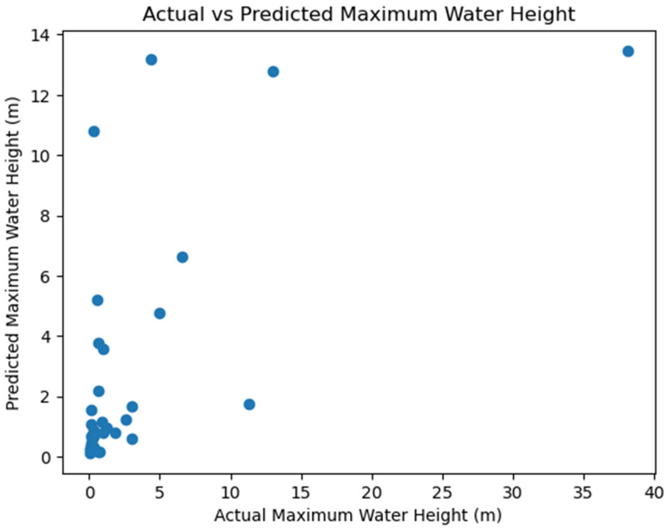

Figure 4. Scatter Plot comparing actual and predicted maximum water heights from the Random Forest Regression model (Photo/Picture credit: Original).

4 DISCUSSIONS

The bar charts illustrating the ordering of the highest water heights throughout history provide a thorough picture of key tsunami incidents. Notably, several incidents stand out for their extraordinary magnitude, such as the "1958 Gulf of Lituya earthquake and mega-tsunami." In-depth historical analysis demonstrates the significance of analyzing specific catastrophes, demonstrating distinct trends that contribute to the existing knowledge of tsunami variety.

The intricate details of tsunami incidents may only be fully comprehended with the use of geo-visualization. Spatial vulnerability was identified by mapping epicenters and creating heat map, which highlighted locations vulnerable to high water levels. The preparation and reaction plan for disasters will be directly impacted by these revelations. Geographic knowledge can be used by decision-makers to better allocate resources, bolster early warning systems (Jin 2011), and put focused mitigation measures in place in high-risk areas.

The incorporation of machine learning, specifically random forest regression models, represents a noteworthy progression in the forecasting of maximum water levels. The model's efficacy is demonstrated by the low mean square error (MSE) in the evaluation. The model's ability to anticipate tsunami impacts with reasonable accuracy is promising for enhancing current capacity for prediction, especially about the Japanese tsunami catastrophe. But it's equally critical to acknowledge the model's shortcomings and possible areas for development.

Despite the remarkable success of the Random Forest regression model, it is crucial to acknowledge its inherent limits. The caliber and volume of the supplied data determine the model's accuracy. Issues including a lack of data, particularly in regions with a low number of past tsunami occurrences, could compromise the model's generalizability. Furthermore, under rapidly changing climatic conditions, the models' assumption of a certain degree of constancy in the underlying patterns of tsunami recurrence may not hold.

In the future, there are many interesting prospects to explore this field of study. Future research should focus on improving predictive models by adding new features, investigating other machine-learning techniques, and incorporating real-time data. With the help of developing technology capabilities, model refinement can continue to strengthen early warning systems and enhance extent ability to lessen the effects of tsunamis.

This study is at the nexus of academia and industry, with an emphasis on both theoretical knowledge and practical applications. Predictive modeling and geographic insights not only add to scientific debates but also give policymakers, first responders, and communities at risk useful tools. The significance of this research in supporting an all-encompassing strategy for tsunami research is shown by this interdisciplinary intersection.

5 CONCLUSIONS

The thorough study's findings provide important insight into the complicated subject of tsunami dynamics. First, historical research using histograms exposes the major patterns of tsunami occurrence and emphasizes specific catastrophes that have left an indelible stamp on the tsunami record. In addition to expanding the knowledge of space, geo-visualization techniques like mapping epicenters and heat map also highlight areas that are susceptible to flooding, adding to the difficulties involved in mitigating and preventing natural disasters.

The combination of geo-visualization with machine learning proved to be an effective method, which was the high point of this study project. The scatterplot shows the effective construction and evaluation of the Random Forest regression model, which demonstrates predictive modeling's potential in understanding and predicting tsunami consequences. The low Mean Squared Error (MSE) illustrates the model's ability to anticipate the maximum water height of the tsunami in Japan, confirming the robustness of the methodology used.

In addition to academic research, the findings of this study have direct relevance to real-world applications. The geographical insights gained from the visualizations can be useful for disaster management strategies, such as directing resources to vulnerable areas. The study's predictive models have the potential to contribute to early warning systems, allowing decision-makers to prepare and respond more quickly. This research's dual applicability, which bridges the gap between theoretical knowledge and practical application, is an invaluable asset in disaster risk reduction.

This area of research is predicted to make further advances in the future. Further improvement of prediction models, such as the incorporation of new features and the exploration of advanced machine learning methods, can increase tsunami predictions' accuracy and reliability. Future research will focus on integrating real-time data and continuously refining models, allowing for the development of more dynamic and responsive warning systems. Technological developments present an opportunity to strengthen defenses against the catastrophic effects of tsunamis, creating a more secure and resilient global community.

Overall, this research project deepens the study of tsunamis and establishes the foundation for real-world disaster management applications. The integration of machine learning, geo-visualization, and historical analysis offers a holistic approach to understanding and reducing tsunami damage. The research provides guidance toward a future where geographical insights and predictive capabilities may help to protect vulnerable areas when dealing with complex natural disasters.

REFERENCES

Wikipedia, 1958 Lituya Bay Earthquake and Mega tsunami, 2023, available at https://en.wikipedia.org/wiki/1958_Lituya_Bay_earthquake_and_megatsunami.

N. B. Parwanto, International Journal of Disaster Risk Reduction, 7: 122-141, (2014).

National Centers for Environmental Information, National Oceanic and Atmospheric Administration (NOAA) - Tsunami Event Data, 2024, available at https://www.ngdc.noaa.gov/hazel/view/hazards/tsunami/event-data?maxYear=2024&minYear=1800.

A. K. Alhamid, Structural Safety, 99, (2022).

Y. Li, Computers & Geosciences, 179, (2023).

S. C. R. Allen, Pure and Applied Geophysics, 170: 1601-1620, (2013).

F. A. R. Abdullah, Journal of Ocean Engineering and Marine Energy, 8, 183-192, (2022).

I. E. Mulia, Nature Communications, 13, 5489, (2022).

OpenStreetMap. (n.d.), OpenStreetMap, available at https://www.openstreetmap.org/#map=5/38.007/-95.844.

D. Jin, Ocean & Coastal Management, 54, 189-199, (2011).