International Journal of Engineering & Scientific Research
Vol.12 Issue 10, November 2024
ISSN: 2347-6532 Impact Factor: 6.660
Journal Homepage: http://www.ijmra.us, Email: editorijmie@gmail.com
Double-Blind Peer Reviewed Refereed Open Access International Journal - Included in the International Serial Directories Indexed & Listed at: Ulrich's Periodicals Directory ©, U.S.A., Open J-Gage as well as in Cabell's Directories of Publishing Opportunities, U.S.A

Enhanced Tsunami Prediction Through Real-Time Data Integration and Hybrid Machine Learning Models

Kishor Yadav Kommanaboina**

Abstract

Tsunamis are the most catastrophic natural hazards and are very dangerous to every coastal community in the world. Very accurate and timely tsunami forecasting is critically important in reducing these impacts and enhancing the resilience of society. This paper presents a comprehensive system for tsunami prediction using real-time data from various sources such as seismic, oceanographic, GNSS, satellite imagery, and IoT sensors. Advanced hybrid machine learning models combined with deep learning methods involving Convolutional Neural Networks (CNNs) and Long Short-Term Memory networks (LSTMs) make the proposed system quite powerful in predicting tsunami incidents. This covers the application of stream processing frameworks in managing real-time data, and scalable storage options that allow for timely and reliable prediction. Above all, model interpretability shall be the prime factor in stakeholder buy-in that helps derive an actionable insight into disaster preparedness and response. This is a method aimed at empowering tsunami early warning systems and hence reducing the effects of tsunamis in highly susceptible coastal areas.

Keywords:

Tsunami Prediction;
Real-Time Data;
Machine Learning;
Hybrid Models;
CNN;
LSTM;
Disaster Management;
IoT Sensors;
GNSS;
Data Integration.

Author correspondence:

Kishor Yadav Kommanaboina,
Independent researcher, The Ohio State University Alumni
kkishore.iiith@gmail.com

1. Introduction

Tsunamis rank among the most destructive of all natural hazards in terms of lives and property taken and destroyed. Recent events, especially the destructive Indian Ocean tsunami of 2004 and the devastating Tōhoku earthquake and subsequent tsunami in 2011 bring up very strongly the dire need for reliable and timely tsunami forecasting models. Such tragedies also point to a continuing quest for improvements in the forecast to mitigate these tragic events.

Machine learning and data pipelines now play a vital role in enhancing the accuracy of tsunami predictions. The integration of diverse data streams- seismic data, oceanographic measurements, and satellite scans- enables better monitoring and analysis. For example, Astri et al. [1] have shown that seismic attributes, when used with artificial neural networks, can provide early detection of tsunami occurrences and greatly enhance prediction accuracy by analyzing various parameters of the earthquake, including magnitude and focal depth. The current simulation systems developed by Gailler et al. [2] predict tsunami wave propagation by using numerical models in conjunction with real-time seismic data, thus providing fast and accurate wave predictions.

Other ideas involve the use of IoT sensors and deep learning techniques. For example, Gokulnath et al. [3] integrated IoT sensors with deep neural networks, demonstrating for the first time the scalability and cost-effectiveness of IoT technology in real-time tsunami prediction. Rim et al. [4] instead proposed the use of CNNs for analyzing GNSS data, coming up with a new way of detecting ground displacements caused by seismic activity-a precursor to tsunamis that provides an early warning. Kishor Yadav [5] used the same techniques to predict the floods instead of Tsunami

However, despite these advances, many gaps still exist in incorporating heterogeneous data sources, processing data in real-time, and incorporating advanced hybrid models. This research paper tries to address these gaps by developing an integrated tsunami prediction system using real-time data originating from

** Independent researcher, The Ohio State University Alumni disparate sources, using state-of-the-art hybrid machine learning models, and ensuring the models are scalable and interpretable. It will be developed using an integrated system of the latest technologies and methodologies to enhance the accuracy, reliability, and usefulness of tsunami forecasting for better disaster preparedness and response.

2. Methodology

This research will develop a coherent tsunami prediction framework that merges a wide suite of real-time data with state-of-the-art machine learning techniques, thereby deploying scalable architectures to process data voluminous in scale. Our approach involves data acquisition and preprocessing from seismic, oceanographic, GNSS, satellite, and historical records, followed by the use of hybrid machine learning architectures such as convolutional neural networks and long short-term memory networks. It is also intended that the system provides real-time forecasting with automated notification to reduce the impacts of tsunamis through timely notifications.

2.1. Data Acquisition and Integration:

2.1.1 Data Sources:

- Seismic Data: We are constantly gathering near real-time earthquake notifications from global monitoring networks containing magnitude, depth, and location. This dynamic data captures recent seismicity that is possibly tsunami-triggering.
- Oceanographic Observations: Real-time data on sea level variations, wave amplitudes, and ocean currents are obtained from DART buoys, tide gauges, and altimeters in orbit around the Earth. Monitoring oceanic dynamics following seismic events is essential for evaluating the potential risk of tsunamis.
- Satellite Imagery: High-resolution imagery from Sentinel and MODIS provides near-continuous updates of the evolving ocean surface changes and wave propagation that are needed for near real-time monitoring.
- GNSS displacements: It detects ground displacements from earthquakes using real-time data from global satellite positioning networks. These provide indications of seismic events capable of generating tsunamis.
- Historical Database: Past tsunami records including travel durations, heights, and impacted areas provide context and boost model training—vital for developing and validating forecasts.

2.1.2 Data Collection:

- Streaming Data Hub: Develop a near real-time data integration topology, similar to Apache Kafka, to acquire variable inputs from multiple, unpredictable sources in a well-managed manner for current and future processing.
- Adaptive Data Warehouse: This can store all kinds of data-static, historical, and real-time streamed-in an elastic data lake such as Hadoop HDFS, AWS S3, or Google Cloud Storage while supporting efficient archiving, access, and processing of large volumes.

2.2. Information Preparation and Analysis:

2.2.1 Data Cleaning:

- Addressing Absent Information: Complete the latter by interpolation, imputation, and deletion. Anomalies should be identified and corrected since information integrity can be guaranteed only that way.
- Normalization and Standardization: Homogenize information from the various sources standardized and normalized in different scales for their due analyses and modeling.

2.2.2 Feature Derivation:

- Seismic Attributes: The salient attributes shall be derived from seismic information, P-wave and S-wave arrival period, energy discharge, and fault system parameters.
- Physical Oceanographic Characteristics: Establish attributes that represent sea level rise, wave transmission, and current patterns.
- Geospatial Features: The GIS will help in inferring the spatial attributes, such as underwater topography and coastal characteristics.

2.3. Predictive Modelling:

2.3.1 Machine Learning Techniques:

● Support Vector Machines (SVMs): SVMs classify these seismic characteristics to estimate the possibility of a tsunami. Given its efficiency in high-dimensional space, the support vector machine is very suitable for recognizing complicated patterns in seismic data.

● Extreme Learning Machines (ELMs): ELMs provide for rapid wave count preparation and precise forecasting of the arrival time. Their foremost superiority lies in the swiftness of learning and feature mapping associated with big data sets.

2.3.2 Deep Learning Models:

● Convolutional Neural Networks (CNNs): These have been modified to analyze seismic waveforms and undersea topographical data in search of patterns indicative of a possible tsunami. They can understand this spatial hierarchy, along with other complicated patterns present in big datasets, very efficiently.

● Recurrent Neural Networks (RNNs), particularly Long Short-Term Memory (LSTM) networks: The Long Short-Term Memory LSTM networks perform amazingly well for complex sequential patterns in ocean and seismic datasets. The LSTMs do a great job of keeping the critical information for a very long time, hence making highly accurate forecasts.

● Hybrid Model: spatial features captured by CNN feed into LSTM for temporal dependencies. The hybrid construction takes advantage of the relative strengths of both: spatial analysis by the CNNs and the LSTMs for learning temporal sequences.

2.3.3 Model Development and Optimization:

● Training and Cross-validation: The models must be trained based on a comprehensive historical and real-time data bank by applying k-fold cross-validation, where iteratively trained, robustness is increased, and overfitting is inhibited by continually assessing the model against different data cuts.

● Hyperparameter Tuning: The optimization of parameters through grid searches and random sampling methodically examines a spectrum of values to determine the most effective combinations. The meticulous adjustment of these hyperparameters is crucial for enhancing both reliability and accuracy.

● Evaluation Metrics: Mean Absolute Error, Root Mean Squared Error, precision, recall, and F1 scores are used for the evaluation of performance. These evaluative metrics will extend thorough analysis and further substantiate that the model meets the anticipated standards of accuracy and reliability.

2.4. Real-Time Data Processing and Prediction:

2.4.1 Stream Data Handling:

● Apache Spark Streaming: Apache Spark Streaming efficiently handles the continuous flow of real-time data, hence allowing continuous retraining and forecasting. It ensures, in real-time, that the processing of the data is very near the input.

● Reduction of Latency: Apply techniques to reduce latency not only while processing data but also in generating predictions. Design suitable data pipelines that allow low latency in prediction due to in-memory processing and data flow.

2.4.2 Real-time Prediction and Warning System:

● Routine Forecast Generation: The continuous generation of tsunami forecasts based on current updates translates into timely forecasts, which are produced at regular intervals with updated information.

● Automation of Warning Dissemination: Automating systems instantly send alerts regarding tsunami occurrences to concerned stakeholders through SMS, e-mails, and Public Warning Systems for immediate preparedness measures against such a situation to mitigate tsunami devastation

2.5. Visualization and Decision-Making Support:

2.5.1 Interactive Dashboard Development:

● Geospatial Visualization: Geospatial visualization presents an interactive visualization dashboard of tsunami forecasts, historical data, and real-time sensor data visualizations on GIS displays for fast representation of the spatial distributions and possible impacts to the relevant stakeholders.

● User-friendly interfaces: User-friendly interfaces allow intuitive, available interfaces to a wide range of stakeholders for exploring various scenarios, getting detailed reports, and interacting with data.

The users of the dashboard include emergency services, municipal governments, policymakers, researchers, and the public. Emergency services benefit through real-time warnings, and imagery showing areas of heightened danger, allowing them to respond with urgency and distribute resources effectively. In addition, municipal governments and policymakers benefit from information on tsunami behavior to build infrastructure, create policies on zoning, or prepare for disaster. Researchers and scientists will have historical and current data to help them with ongoing research and improving models. Serving public safety and awareness, timely warnings, and preparedness are concerns provided to community leaders and citizens.

2.5.2 Enhanced Visualization Techniques:

● Augmented Reality (AR): Leverage the AR tools that apply tsunami predictions and real-world data on physical landscapes, fully immersing virtual components into real settings. In such an intuitive manner, the first respondents and policymakers can get an experiential view of how tsunamis will look concerning specific areas in detail. It will help in better comprehension and planning as such visualization gives full insight into the potential impacts.

● Scenario Simulation: The capability for simulation is also to be provided in the dashboard to visualize or hypothetically simulate different tsunami scenarios based on various conditions and inputs. Such simulations allow the stakeholder to prepare for a range of possible events in building upon an overall preparedness and response strategy against any disaster situation.

Such an approach would enable the proposed tsunami forecasting system to provide real-time precision forecasts using such an inclusive methodology that will integrate both historical archives and dynamic real-time data, backed by advanced hybrid modeling. Enhancements in this direction would thus lead to increased preparedness and response in handling such calamities. State-of-the-art advances in data processing and machine learning ensure scalability, reliability, and meaningful insights into mitigating tsunami impacts.

3. Conclusion

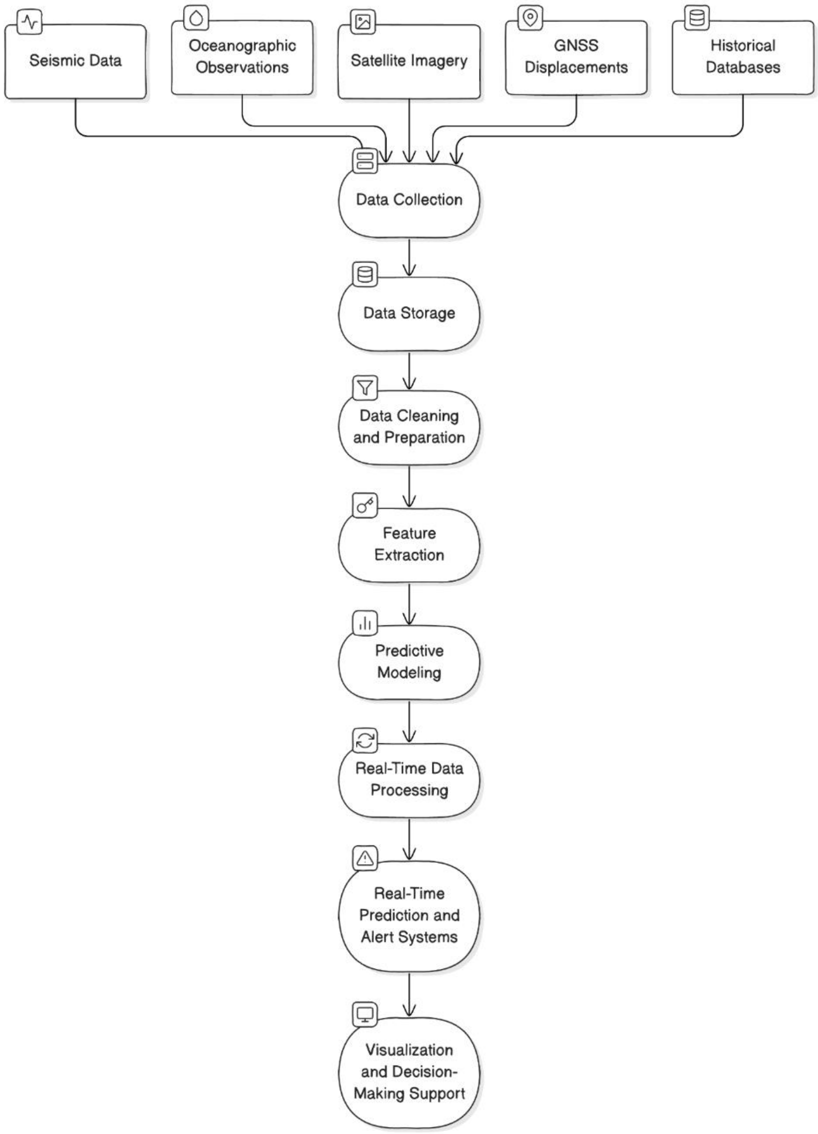

Figure SEQ Figure * ARABIC 1. Tsunami Prediction Data

The proposed architecture of the Tsunami Early Warning System presents an Elaborate pipeline, containing all the nuances of Big Data. Aggregate data from various sources includes seismic information, oceanographic data from buoys and sensors, satellite images depicting tidal changes, GNSS-based land measurements to high accuracy, and historic tsunami data from several centuries ago, all in real-time are processed using the streaming platform Apache Kafka. After summarization, data is stored in highly scalable storage facilities like Hadoop HDFS, AWS S3, and Google Cloud Storage for computationally intensive analysis over exabytes of data.

Afterward, raw data are cleaned and formatted; examples of techniques used may include interpolation to fill gaps, imputation to replace missing values, and normalization to make diverse measurements standard. Key features are then extracted by performing elaborate statistical analyses on seismic, oceanographic, geophysical, and geographical characteristics. The features extracted using the above procedure are used as input for advanced tsunami forecast approaches using techniques like support vector machines, extreme learning machines, deep convolutional neural networks, and long short-term memory networks, which are capable of capturing the fine-grained pattern.

Apache Spark Streaming updates these models in real-time, with continuous coming-in data. The resultant updated models generate accurate short-term predictions of tsunamis and effectively disseminate them to emergency management personnel, government organizations, researchers, and local populations through customized dashboards. Stakeholders can visualize and analyze the forecast while issuing early warnings using advanced interactive visualization technologies.

4. Interpretation and Implication

The integration of diversified real-time data streams and state-of-the-art machine learning methodologies within one framework significantly improves the accuracy and speed of tsunami forecasting. Using huge volumes of data from various sources in concert provides the ability to detect subtle signs that are of utmost importance for gaining precious minutes of early warnings. Continuous real-time reevaluation ensures the system adapts promptly to changed conditions, hence retaining a very responsive decision support system. Combined, these enable better preparedness for tsunamis and reduce loss of life by early action.

It is the visualization and decision-making tools in this system that shall help the stakeholders analyze predictions and act accordingly. Visualization through interactive dashboards, and advanced visualization through augmented reality and scenario modeling, provide a readable and rich view of what tsunami effects could be, thus driving preparedness planning and response efforts.

5. Limitations and Future Scope

While this architecture advances tsunami forecasting significantly, limitations remain. Chiefly, reliability depends on consistent, accurate real-time data from diverse sources. Any data interruption or inaccuracy impacts performance. Moreover, the complexity of collecting, integrating, and processing varied data streams necessitates robust infrastructure and efficient data administration.

Future work can strengthen the robustness and scalability of the data pipeline. Integrating extra sources of information, such as social media and crowdsourced reports, comprehensively extends situational awareness. Interpreting machine learning models more understandably and embedding explainable AI techniques further advances the adoption and stakeholder trust.

In turn, the continuous single platform for all hazard management is expanded toward earthquake and hurricane forecasting-and-response amongst others. Ongoing technological and algorithmic advances in data handling support the continuous refinement and effectiveness of the system.

The proposed tsunami prediction data pipeline is very apt in harnessing real-time data and advanced analytics for preparedness and response, hence an important landmark. The limitations found within could be addressed, and enhancements explored to turn it into an indispensable tool in mitigating tsunami and disaster impacts.

References

[1] Novianty, A., Machbub, C., Widiyantoro, S., Meilano, I., & Daryono. (2022). Tsunami Potential Prediction using Seismic Features and Artificial Neural Network for Tsunami Early Warning System. International Journal on Advanced Science, Engineering and Information Technology, 12(1), 16. https://doi.org/10.18517/ijaseit.12.1.14237

[2] Berto, M.C. and Vruce K., "Implementation of a Fuzzy PI Controller for Speed Control of Induction Motors Using FPGA," *Journal of Power Electronics*, vol. 10, pp. 65-71, 2010.

[3] S. Gokulnath, R. Sumathi, J. Yuvaraj, C. Poorani, & K.Leena Priya. (2022). Analysis of Early Warning Tsunami Prediction System using the Arduino Uno R3 and DNN. 2022 6th International Conference on Electronics, Communication and Aerospace Technology, 369–374. https://doi.org/10.1109/iceca55336.2022.10009246.

[4] Rim, D., Baraldi, R., Liu, C. M., LeVeque, R. J., & Terada, K. (2022). Tsunami Early Warning From Global Navigation Satellite System Data Using Convolutional Neural Networks. Geophysical Research Letters, 49(20). https://doi.org/10.1029/2022gl099511

[5] Kishor Yadav Kommanaboina. (2024). Real-time Flood Prediction using a Big Data Approach. International Journal for Multidisciplinary Research, 6(5). https://doi.org/10.36948/ijfmr.2024.v06i05.28244