IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)
An International Open Access, Peer-reviewed, Refereed Journal

"Proactive Disaster Detection"

Authors:

Shashank J K Parth Sagar M Teja Reddy M

Dr. Manjunath K V

School of Computer Science and Engineering, Presidency University, Bangalore, India

Guide:
Dr. Manjunath K V, Assistant Professor, School of Computer Science and Engineering(data science), Presidency University

A R T I C L E I N F O

Keywords:

Machine learning Algorithms, Rainfall induced disasters, Hyperparameter Tuning, Time series forecastin.

A B S T R A C T

The central objective of this project is to employ machine learning techniques to forecast natural disasters proactively. Utilizing statistical models applied to historical data, the system aims to predict future occurrences, incorporating Geographic Information System (GIS) data on tectonic plates and past floods. Through this approach, the model can be trained to anticipate floods and tsunamis, enabling early predictions that play a pivotal role in facilitating timely evacuation and disaster preparedness.Throughout the course of history, natural disasters have consistently posed a global-level threat, unleashing catastrophic events that cause significant harm to both lives and property. These disasters have not only eradicated entire civilizations but have also steered the course of human history. The toll exacted by natural disasters is immense, resulting in the loss of hundreds of thousands of lives and billions in property damage. Despite advancements in physical, mental, and technological aspects, there are challenges that humans cannot fully overcome, with natural disasters standing out as one such formidable force. While humanity may not possess the ability to prevent natural disasters entirely, adequate preparation lies within our control. Natural disasters can be broadly classified into geological and meteorological types. Despite ongoing efforts to predict earthquakes and other natural disasters, achieving precise predictions remains an enduring challenge.

1. Introduction

Over the last decade, over 2.6 billion people have been affected by disaster events that included tsunamis, flooding, earthquakes among many other hazards. Such disasters, be it natural or artificial ones, have already caused hundreds of deaths. These put a human's life, the ecology, as well as a nation's economy in danger. Computer science has made great advancement with time and the study of disaster management also got a huge amount of data from this field. Nevertheless, much of this data is unstructured and thus a challenge to clean and process at this high volume.

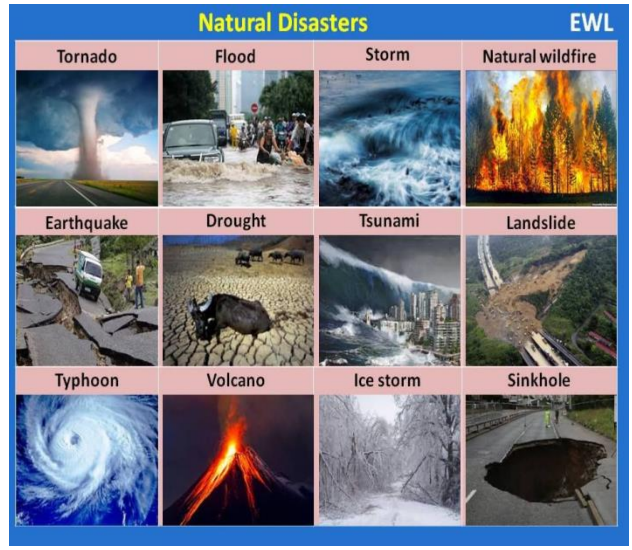
The project addresses the issues with a comprehensive review of the existing procedures and techniques applicable during both pre- and post-disaster periods. This will reduce the losses to the minimum level and maximize preparedness against disasters.

The occurrence or incidents of natural calamities can, therefore, be described as forceful or violent happenings with tremendous devastations on humanity and the existence of living organisms. Within the past decades, natural disasters have hugely increased. Common reported occurrences have comprised floods, droughts, heavy rain, wild bushfires, wind, and unprecedented cases of high temperatures accompanied by volcanic ash, earthquakes, landslides, tsunamis avalanches, and eventually drought (Barrett & Curtis, 1999). Human-induced climate change brings many more disasters via events: "wildfire, heat-waves, floods.".

Natural disasters have the potential to cause devastation in human populations, including loss of life, property damage, and disruption of economic activity. Sometimes, they also cause environmental damage, such as deforestation, soil erosion, and water pollution.

It is necessary to be well informed about the risk of natural disasters and to adopt measures to mitigate their impact. This includes constructing earthquake-resistant structures, adopting flood control measures, and preparing for hurricanes and tornadoes

Fig 1.



Some of the most common kinds of natural calamities include the following:

Fig 1.1 Earthquakes: These are the phenomena of tectonic plate movement, which causes great shaking, ground displacement, and the occurrence of landslides.

Fig 1.2 Floods: It is an incident that takes place when water in a general way covers normally dry land, and its reasons include heavy rainfall and melting snow as well as breakage of the dam.

Fig 1.3 Floods Hurricanes: These are tropical cyclones generated over a warm ocean characterized by strong winds, heavy rainfall, and storm surge

Fig 1.4 Tornadoes: Turbulent columns of air, rotating and touching the ground, which can be very destructive to buildings and structures.

Fig 1.5 Wildfires: These are wildfires that can flare up in moments and cover wide areas of ground.

The field of artificial intelligence, particularly machine learning, deals with the design of statistical models and algorithms that allow computers to learn from data, identify patterns, and make decisions on their own.

The last few years have seen the intensive involvement of machine learning in the forecast area that basically endeavors through identification to avoid natural disasters. Of late, machine learning algorithms have been devised which compute the possibility of natural disasters based upon past history.

These algorithms analyze various sources of data that include satellite data, atmospheric data, and historical data on weather and catastrophes to come up with some pretty accurate forecasting.

To make an effective prediction of the onset of a natural disaster, we have used one neural network and two separate machine learning strategies. Long Short Term Memory(LSTM), Linear Regression, ARIMA is the name of the different types of algorithms.

The models that performed best at predicting when a natural disaster would first manifest itself have been chosen after we assessed how well each of them performed.

Machine Learning:

Artificial intelligence is expressed through machine learning, which is defined as the process by which computer systems are capable of improving their performances or skills on any particular task independently over time. Put in simple words, machine learning gives computers the ability to learn from data without explicit programming. Among many others, this has posed significant methodological challenges for the present work.

Those will be further categorized into several types of machine learning: supervised, unsupervised, semi-supervised, and reinforcement learning. In supervised learning, an algorithm learns from data labeled-that is, the input data is previously classified. Unsupervised learning handles unstructured data to find a pattern or a relation. Semi-supervised learning contains part of supervised and part of unsupervised, while reinforcement learning is about agents that learn through interaction with an environment toward some goal in mind. Machine learning finds a number of applications in computer vision, natural language processing, speech recognition, recommendation systems, and predictive analytics, among other fields. It proves particularly valuable in solving complex problems that are difficult or impossible to solve with traditional programming techniques.

The advantages of machine learning are numerous and include:

Automation: Machine learning can also automate many tasks that would require human intervention; this frees up time and resources. For example, it could classify images, recognize speech, and make predictions.

Accuracy: Machine learning algorithms excel in tasks that require both precision and consistency, such as medical diagnosis and financial forecasting. Their ability to learn from large data sets fuels their accuracy, making them indispensable tools in these vital fields. Efficiency: Swiftness through oceans of data, churning out crucial decisions in milliseconds-that is the power of machine learning, which makes it the go-to solution for tasks that demand both speed and mastery of data. Adaptability: Unlike static algorithms, machine learning models are perpetual learners, sharpening their insight with every data bite. Personalization: Be it recommending the perfect pair of shoes or predicting individual health risks, machine learning weaves its magic across diverse fields in creating highly individualized experiences.

Scalability: Where data explodes, machine learning thrives. This insatiable algorithm, ever-evolving and endlessly scalable, holds the key to unlocking a Pandora's box of efficiency and accuracy across industries.

2.Related work and Literature Surve

(a)Author: Amezquita-Sanchez J., Valtierra,Rodriguez M., Adeli H **Methodology:** Harness the combined capabilities of signal processing, image processing, and statistical techniques to drive
Outcomes: Enhance the precision in predicting natural disasters. **Drawbacks**: Restricted statisticalparameters available for forecasting.

(b)Author: Zhang X.Y., Li X.,Lin X.
Methodology: Particle swarm optimization
Outcomes:Forecast the seismic magnitude of an earthquake. **Drawbacks**: Focus solely on forecasting within the seismic dataset.

(c)Author: Adeli H., Panakkat A.
Methodology: Artificial neural network.
Outcomes: Anticipate the seismic magnitude of an Earthquake **Drawbacks**: A constrained set of parameters employed for forecasting.

(d)Author: Kradolfer U
Methodology:Improve customer satisfaction by mining text from feedback channels
Outcomes:indicate earthquakes with both swiftness and precision using seismological data.
Drawbacks:Rely on public feedback as a factor in earthquake detection.

(e)Author: Merz B., Kreibich H., Lall U.
Methodology:Tree-based decision model.
Outcomes: Evaluating how well our model detects both the onset **Drawbacks**: Constraints on parameters for identifying flood-affected areas.

(f)Author: Sahay R.R., Srivastava A
Methodology:Unlocking the synergy of Artificial Neural Networks (ANNs), Genetic Algorithms (GAs)
Outcomes: Provide a concise overview of positive outcomes in comparison to existing methodologies in Southeast Asia.
Drawbacks: Conduct research focused on time series data pertaining to monsoon floods occurring in specific regions of India during June and September.

(g)Author: Venkatesan M., Thangavelu A., Prabhavathy P.
Methodology:Support vector machine, Bayesian Naïve classifier **Outcomes**:Categorize natural disasters based on diverse parameters.
Drawbacks: Restricted to the initial phases of natural disasters.

(h)Author: Korup O., Stolle A
Methodology:Technique based on machine learning.
Outcomes:Forecast landslides with an accuracy ranging from 75 to 95 percent.
Drawbacks:Provide additional guidelines for selecting models to predict large-scale landslides.

(i)Author: Di Salvo R., Montalto P., Nunnari G., Neri M., Puglisi G
Methodology: Artificial neural network with backpropagation **Outcomes**: Forecasting based on historical datasets.
DrawbacksDynamic prediction is indispensable for the effectiveness of this system

(j)Author: Das H.S., Jung H
Methodology:Utilize clustering methods for analyzing multivariate time series data.
Outcomes:Proposed a dynamic clustering methodology for time series analysis, integrating a self-optimizing organizing mapping technique.
Drawbacks: The clustering process necessitates dynamic time series data.

The research gap in natural disasters lies in the inadequate understanding of the long-term socio-economic impacts on communities that are susceptible to these calamities. Whereas various studies focus on immediate response and recovery, there is a shortage of research examining the long-term consequences on societies, particularly in developing regions. The complex interaction of pre-existing vulnerabilities, post-disaster adaptive capacities, and the chronic challenges faced by affected populations in their struggle to rebuild their lives has not received sufficient attention. Furthermore, there is a serious lack of an integrated approach, drawing on environmental science, social dynamics, and policy analysis in putting together comprehensive strategies for mitigating the long-term effects of natural disasters. Contributing to this prevailing research gap allows enhanced effectiveness of preparatory, responsive, and recovery strategies-thereby enhanced levels of resilience when facing an increased climatic uncertain future.

2. Methodology

Data Collection:

First come the tasks of data collection that are essential for the model. The dataset needs to be developed with information such as temperature, average rainfall in millimeters, and monthly distribution of rainfall; it can be sourced from a database online or by visiting government websites.

Data Preprocessing:

After data acquisition, the steps involve the pre-processing of data to make it ready for analysis. This process entails refining and changing the data to make it most effective to be analyzed. In cleaning the data, duplicate values are removed, knowledge gaps are filled, and outliers are excluded that might have been introduced.

Data transformation involves scaling features, encoding categorical variables, and generating new features when needed. For a deeper understanding of data transformation, you may want to read this article.

Feature Selection:

The next in the line of procedures is Feature Selection, which comes in where identifying the key attributes that will feature within the model scope takes centre stage. While at this step, we would want to detect the pertinent characteristic variables that keep a great stake in the variations of the target variable under inquiry. The selection involves judgment about the relevant traits.

Some of the approaches toward important feature selection objectives include various methods: incorporating metrics like mean absolute error, mean squared error, and root mean square error; correlation analysis; and recursive feature elimination.

Model Selection:

With important features for the situation identified, there needs to come the choice of the model serving the purpose more efficiently. The selection of needed features was realized, then assessed by various machine learning algorithms performance in view: a linear regression one, XGBoost, and ARIMA. It is a thoughtful comparison of advantages and disadvantages aimed at finding such an option of models that balances perfectly between the achieved precision and functions' efficiency.

Model Evaluation:

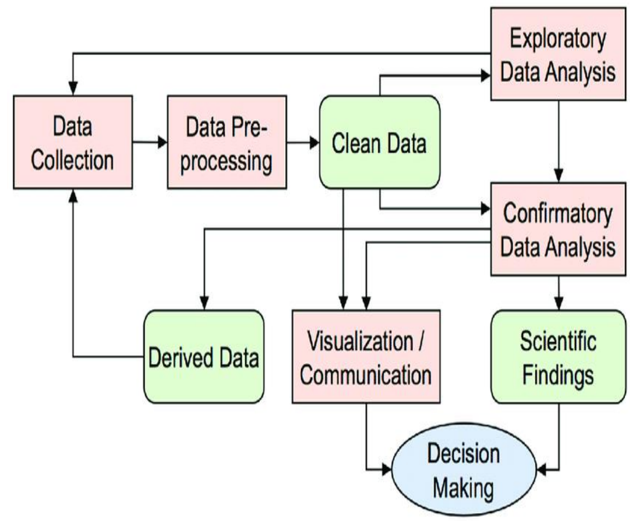
After model selection comes model performance evaluation, which includes computation of the various metrics: RMSE, MAE, and MSE. Some measurements that were considered in this process were accuracy and precision, because this helps find a place it could be improved upon.

Hyper-parameter Tuning:

Once the model analysis is done, the second step would be revising the hyperparameters of the model. This method is more general and is addressed as hyperparameter tuning. The most formal name of the above technique is "hyperparameter tuning," a process that simply includes choosing such hyperparameters that this model can do its best toward meeting its objectives in terms of performance. Therefore, this optimum may be executed via different tactics; for example, grid search and random search are among these tactics.

Deployment:

The last stage, after the model is trained and tested, involves its deployment into the real world for practical applications. Deployment is done after all the previous steps are complete. In integrating the model into the real world, there is the possibility of doing it either through a web application or incorporating it into an existing system, making it flexible in terms of implementation. For the model to retain its credibility over time as a representation



of reality, it must be continually analyzed and refined. The key components of the proposed methodology for disaster detection include data collection, preprocessing, feature selection, model selection, model evaluation, hyperparameter tuning, and deployment. Empowering communities with knowledge and tools to stay ahead of the curve. Early detection of disasters saves lives, saves infrastructure, and builds resilience against the fury of nature.

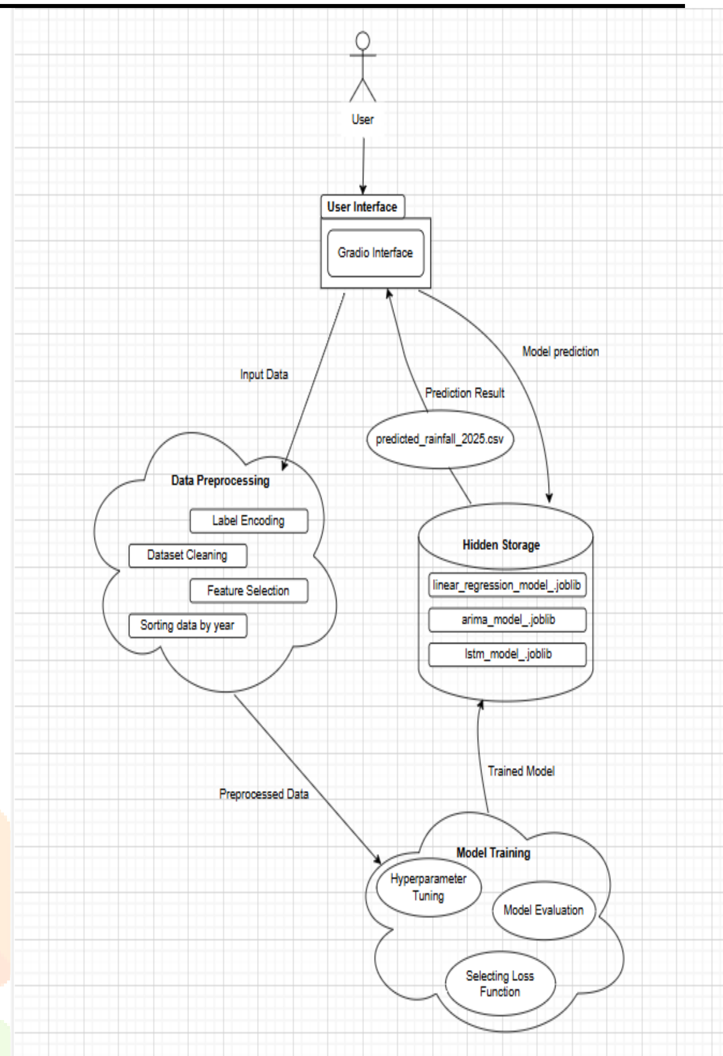
The recommended approach for early disaster detection is to develop a machine learning model that predicts the risk of diabetes, considering a variety of risk factors. The model will be trained and tested on a dataset representing the monthly rainfall distribution from 2000 to 2024. The machine learning model will be trained using various methods, including random linear regression, LSTM, and ARIMA, on the provided dataset. The deployed model will be accessible through a web application where users can input their information and get a corresponding disaster risk score. This solution aims to facilitate early disaster detection, empowering individuals to take preventive measures.

3. **Discussions**

4.1 Objectives

The core purpose of the research study will be the use of machine learning in developing a model which can predict disasters at an early stage. The objective is to set up a base system that may use satellite data and various sources of data in predicting the probabilities and times of natural disasters. The enhanced base system can provide the possibility to extend its functions toward an early warning and decision support tool for disaster management agencies and other actors to take proper actions with a minimum level of losses in human lives in the case of a natural disaster.

4.2 System Architecture



4.3 Algorithms:

1.Linear Regression:

In statistics, a linear regression is a statistical model that defines the linear relationship between a scalar response and one or more explanatory variables, which may be usually referred to as dependent and independent variables. If one explanatory variable is considered, the procedure is termed simple linear regression, and if more than one, then the process is regarded as multiple linear regression.

Note that this differs from multivariate linear regression, where the prediction has multiple correlated dependent variables, instead of a single scalar variable. If some of the explanatory variables have measurement errors, then what is called the errors-in-variables model or measurement error model needs to be used.

2. Lstm:

Long Short-Term Memory Long Short-Term Memory, or LSTM, is a type of RNN architecture that was designed to model long-range dependencies in sequential data. An LSTM network is an improved form of traditional RNNs that have the ability to overcome the problem of vanishing gradients through a gating mechanism that allows controlled flow of information through the network.

This architecture is basically made up of three different gates, which model whether to remember or forget data in a sequence. The LSTM finds its application mainly in time-series forecasting and in other areas with sequences, including NLP tasks. Stacked or bidirectional LSTM architectures might be employed where multiple correlated input sequences need consideration. Besides, for a sequence with noised or lacking data, implementations of hybrid models, which introduce the integration between LSTM and typical preprocessing techniques-data imputation and filtering-can act to enhance its performance.

3. Arima-:

Auto-Regressive Integrated Moving Average In time series analysis, one of the more common statistical approaches to forecasting and understanding data involves the use of an Autoregressive Integrated Moving Average, or ARIMA for short. It brings three aspects together in ARIMA: an autoregressive component showing the relationship of an observation and previous observations, a differencing component to get the series into stationarity for removing trends, and finally the moving average one concerning the relationship that an observation may take with respect to residual errors at earlier time points. So, in some cases where there are periods in a time series, this gives us a name called SARIMA. Though ARIMA models are pretty good at working out univariate time series, if there is more than one operational time series that might influence each other, then a choice must be made between a vector autoregression model and an ARIMA-based hybrid approach. If there is noise or incompleteness, methods of data preprocessing could involve smoothing or the detection of outliers in order to clean the input for the model and raise its accuracy.

4.4 Evaluation:

1. MAE

The Mean Absolute Error (MAE) is calculated by averaging all absolute errors, and its formula is:

$$\text{MAE}=\frac{1}{N}\sum_{i=1}^{N} \quad |x-x|$$

Where:
n denotes the total number of errors,
Σ is the symbol, Depicting the process of adding all the values together,
|xi – x| denotes the absolute errors.

2. MSE

The fundamental aim of this research is to formula for Mean Squared Error (MSE) is as follows:

$$\text{MSE}=\sum \quad \left(\frac{(yi-yi\prime)\text{^}2}{n}\right)$$

In the context of this formula:
- "yi" stands for the ith observed value.
- "ŷi" represents the corresponding predicted value.
- "n" indicates the total number of observations

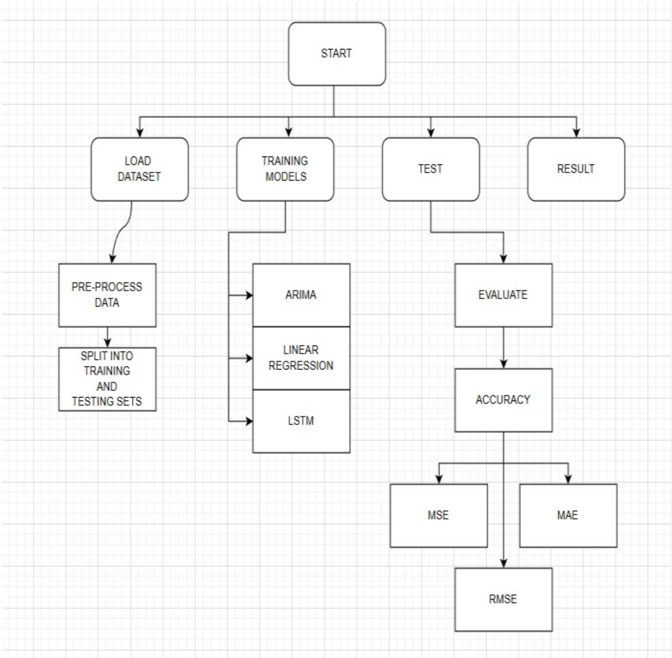
3. RMSE

The formula for computing the Root Mean Square Error (RMSE), often abbreviated as RMSE, is articulated as follows:

$$\text{RMSE} = \sqrt{}\Sigma(P_i - O_i)^2 / n$$

In this context:
- Σ is a mathematical symbol denoting summation,
- Pi stands for the predicted value corresponding to the ith observation in the dataset,
- Oi represents the observed value corresponding to the ith observation in the dataset,
- n indicates the sample size.

4.5 Flowchart:



5.Results

The project undertaken is about the development of a rainfall prediction system using different machine learning algorithms, namely Linear Regression, Auto Regression Integration Moving Average (ARIMA), and Long Short Term Memory (LSTM). The work started with the loading of the labeled dataset that contained relevant features and color-coded labels signifying different intensities of rainfall. It was followed by the splitting of the dataset into training and testing sets to allow both model training and testing.

Each of the machine learning algorithms was thereafter trained independently using their different strengths in pattern understanding of the data. Further, the test set was used for predictions, and then the performance for each model was checked using the MSE, RMSE, and MAE metrics from the sklearn-metrics library.

The basic objective of this research is to assess the results, showing different performances for the different models. The ARIMA algorithm had an impressive MSE of 61.00, RMSE of 7.81, and MAE of 5.25, followed closely by LSTM. Though Linear Regression performed well, it showed relatively low accuracy in the context of the current problem.

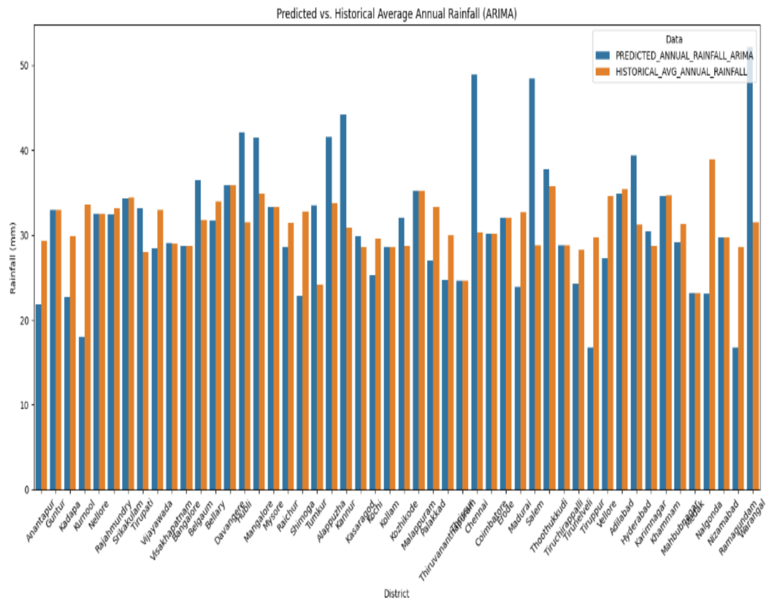


Figure 5.1 depicts the final result of the prediction for monitoring climate change using Machine Learning.

SlN O.	ALGORITHMS	MAE	MSE	RMSE
1.	LINEAR REGRESSION	398.52	420349.41	648.34
2.	LSTM	26.24	698.13	26.42
3.	ARIMA	5.25	61.00	7.81

The ARIMA model, which was used in the system, beat other models and resulted in an notable MSE (61.00), RMSE (7.81) and MAE (5.25).

4. Conclusion

In brief, the performance of the results of the proposed machine learning early disaster detection approach proved promising. The performance of the ARIMA algorithm applied to the system outperformed other models, with an impressive MSE of 61.00, RMSE of 7.81, and MAE of 5.25. Presently, real-time updates on rainfall are easily accessed by the public.

A system that could detect high probabilities of large-scale natural disasters and dynamically model post-disaster spread would have strong potential to reduce disaster risks and improve resilience. In this regard, early warning and decision support could be instrumental in minimizing damage and saving lives during any natural catastrophe.

It is a possible development based on processing satellite data and machine learning connected with simulations modeling. Every design of such a system shall be focused on issues such as scaling up and adaptability, to various forms of natural disaster occurrences and, accordingly, to different geographical areas of Earth.

Key advantages of implementing such a system include:

• Early warning capabilities

• Improved disaster management • Reduced damage and loss of life.

7. References

1. Amezquita-Sanchez J., Valtierra-Rodriguez M., Adeli H. Current efforts for prediction and assessment of natural disasters: Earthquakes, tsunamis, volcanic eruptions, hurricanes, tornados, and floods. Sci. Iran. 2017;24:2645–2664. doi: 10.24200/sci.2017.4589. [CrossRef] [Google Scholar]

2. Zhang X.Y., Li X., Lin X. The data mining technology of particle swarm optimization algorithm in earthquake prediction. Adv. Mater. Res. 2014;989–994:1570–1573. doi: 10.4028/www.scientific.net/AMR.989-994.1570. [CrossRef] [Google Scholar]

3. Adeli H., Panakkat A. A probabilistic neural network for earthquake magnitude prediction. Neural Netw. 2009;22:1018–1024. doi: 10.1016/j.neunet.2009.05.003. [PubMed] [CrossRef] [Google Scholar]

4. Kradolfer U. SalanderMaps: A rapid overview about felt earthquakes through data mining of web-accesses; Proceedings of the EGU General Assembly Conference; Vienna, Austria. 7–12 April 2013; pp. EGU2013–6400. [Google Scholar]

5. Merz B., Kreibich H., Lall U. Multi-variate flood damage assessment: A tree-based data-mining approach. Nat. Hazards Earth Syst. Sci. 2013;13:53–64. doi: 10.5194/nhess-13-53-2013. [CrossRef] [Google Scholar]

6. Sahay R.R., Srivastava A. Predicting monsoon floods in rivers embedding wavelet transform, genetic algorithm and neural network. Water Resour. Manag. 2014;28:301–317. doi: 10.1007/s11269-013-0446-5. [CrossRef] [Google Scholar]

7. Venkatesan M., Thangavelu A., Prabhavathy P. An improved Bayesian classification data mining method for early warning landslide susceptibility model using GIS; Proceedings of the Seventh International Conference on Bio-Inspired Computing: Theories and Applications (BIC-TA 2012); Springer, Gwalior, India. 12–16 December 2012; pp. 277–288. [Google Scholar]

8. Korup O., Stolle A. Landslide prediction from machine learning. Geol. Today. 2014;30:26–33. doi: 10.1111/gto.12034. [CrossRef] [Google Scholar]

9. Di Salvo R., Montalto P., Nunnari G., Neri M., Puglisi G. Multivariate time series clustering on geophysical data recorded at Mt. Etna from 1996 to 2003. J. Volcanol. Geotherm. Res. 2013;251:65–74. doi: 10.1016/j.jvolgeores.2012.02.007. [CrossRef] [Google Scholar]

10. Das H.S., Jung H. An efficient tool to assess risk of storm surges using data mining. Coast. Hazards. 2013:80–91. doi: 10.1061/9780784412664.008. [CrossRef] [Google Scholar]

11. Chatfield A.T., Brajawidagda U. Twitter early tsunami warning system: A case study in Indonesia's natural disaster management; Proceedings of the 2013 46th Hawaii International Conference on System Sciences; Maui, HI, USA. 7–10 January 2013; pp. 2050–2060. [Google Scholar]

12. Khalaf M., Alaskar H., Hussain A.J., Baker T., Maamar Z., Buyya R., Liatsis P., Khan W., Tawfik H., Al-Jumeily D. IoT-enabled flood severity prediction via ensemble machine learning models. IEEE Access. 2020;8:70375–70386. doi: 10.1109/ACCESS.2020.2986090. [CrossRef] [Google Scholar]

13. Zhai C., Zhang S., Cao Z., Wang X. Learning-based prediction of wildfire spread with real-time rate of spread measurement. Combust. Flame. 2020;215:333–341. doi: 10.1016/j.combustflame.2020.02.007. [CrossRef]

[Google Scholar]

14. Tan J., Chen S., Wang J. Western North Pacific tropical cyclone track forecasts by a machine learning model. Stoch. Environ. Res. Risk Assess. 2020:1–14. doi: 10.1007/s00477-020-01930-w. [CrossRef] [Google Scholar]

15. Liu Y.Y., Li L., Liu Y.S., Chan P.W., Zhang W.H., Zhang L. Estimation of precipitation induced by tropical cyclones based on machine-learning-enhanced analogue identification of numerical prediction. Meteorol. Appl. 2021;28:e1978. doi: 10.1002/met.1978. [CrossRef] [Google Scholar]

16. Meadows M., Wilson M. A Comparison of Machine Learning Approaches to Improve Free Topography Data for Flood Modelling. Remote Sens. 2021;13:275. doi: 10.3390/rs13020275. [CrossRef] [Google Scholar]

17. Nisa A.K., Irawan M.I., Pratomo D.G. Identification of Potential Landslide Disaster in East Java Using Neural Network Model (Case Study: District of Ponogoro) J. Phys. Conf. Ser. 2019;1366:012095. doi: 10.1088/1742-6596/1366/1/012095. [CrossRef] [Google Scholar]

18. Syahaneim Marzukhi, Mohd Amirul Shafiq Mohamad, Sidik, Haidawati Mohd Nasir, Zuraini Zainol and Mohd Nazri Ismail. Flood Detection and Warning System (FLoWS) doi:. 10.1145/3164541 [CrossRef] [Google Scholar]

19. S. N. K. B. Amit and Y. Aoki, "Disaster detection from aerial imagery with convolutional neural network," 2017 International Electronics Symposium on Knowledge Creation and Intelligent Computing (IES-KCIC), Surabaya, Indonesia, 2017, pp. 239-245, doi: 10.1109/KCIC.2017.8228593.