Prediction, Analysis And Relief Measure Reports For Disaster Crisis Management Using Regression, Artificial Neural Network And RFC

Ramya Prakash[1], S Mathangi[2], Utkarshini Acharya[3], Zainab Noorain[4], Kavita Horadi[5]

[1,2,3,4]*Students, Department of Computer Science* (B.E)*, BNM Institute of Technology (VTU), Bangalore, Karnataka -560078, India*
[5]*Assistant Professor, Department of Computer Science, BNM Institute of Technology (VTU), Bangalore, Karnataka -560078, India*

ramyakondabagil@gmail.com[1] *mathangi.s98@gmail.com*[2]
utkarshiniacharya@gmail.com[3]*zainabn98@gmail.com*[4]*kavitahoradi@gmail.com*[5]
ramyakondabagil@gmail.com

Abstract

The Disaster Crisis Management System aims to provide and facilitate model tha treads, interprets and entirely comprehends the parameters and factors involved in a natural disaster like Floods and Earthquake, and the subsequent outputs and relief measures that need to be deployed within hours for faster and improved public relief. The parameters being considered for the disaster management and relief include data from various intensity units, Relief funds, Extent of Disaster, casualties, mode of Rescue operations, etc. The system makes use of Regression and Perceptron Model for the prediction of disaster and the Random Forest Classifier algorithm for Post Disaster analysis.

Keywords: Disaster Management, Crisis Management, Earthquake Detection, Flood Detection, relief measures.

1. INTRODUCTION

Natural disasters, caused due to changing climatic conditions, extreme weather and environmental degradation, have been causing tremendous loss of lives and property. Rebuilding not only the geographical area but also the lives of the people involves a detailed risk analysis, disaster management and post disaster relief. In the structural framework of response and recovery, disaster is the impact which exhausts the capacity of local responders and increases demands on resources which are unavailable locally. During the event of a natural disaster, the Response phase, includes activities like search, rescue, damage and needs assessments, and the Recovery phase includes reconstruction efforts that are performed to assess the damage and destruction of infrastructure.

Measuring risk of physical damage, victims and economic equivalent loss are required for Disaster crisis management. Risk assessment is associated with the conditions that facilitate disasters. Vulnerability can be defined in terms of the fragility, susceptibility or lack of resilience of the population in the event of the disaster.
Disaster Crisis Management involves assessment of disaster prone geographical areas for the extent of coverage of the disaster, potential damage, loss of life, funds required on the basis of

the intensity or predicted intensity of the ongoing disaster. Taking into account the area of occurrence, the history of previous disasters occurred, the environmental conditions governing the previous disasters and the current environmental conditions for comparative analysis, the Intensity of the Disaster can be determined. The previous techniques used for this prediction involved manually looking at the conditions and determining if they can be related to any disaster. Crisis management included techniques like manual calculation of the funds required for recovery depending on the Sensex of the area and the infrastructure development, along with the evacuation system during the wake of the disaster.

A model has been presented in this paper to use advanced techniques of Machine Learning to perform prediction and relief assessment using learning models. The most commonly occurring disasters in India like Floods are predicted using the Time Series Regression model while the Perceptron Model in Artificial Neural Networks is used for predicting Earthquake. The Random Forest Classifier has been used to design The Relief assessment model.

2. Related Work

There are various machine learning (ML) algorithms that have been previously used for the prediction model for Floods and Earthquake.

M Khalaf et al. [1] discussed the use of various ML algorithms for prediction of a flood's severity and classification of the floods into three classes, normal, high-risk and abnormal floods. It produced enhanced results for pre- processing of flood dataset based on time series. The models used for comparative accuracy prediction included Random Forest Classifier (RFC), Support Vector Machines (SVM), Levenberg-Marquartdt training algorithm (LEVNN), Linear Neural Network (LNN) and RFC performed better using the performance measures examined.

J Opella et al. [2] used the data available on the Geographic Information System (GIS) to generate reliable flood disaster susceptibility and probability maps. Fusion Convolutional Network was used along with Support Vector Machine for a better image map result. Flood mapping system is used to calculate the range and approximate depth of water in flood affected areas. In this paper, we intend on using Artificial Neural Network and not Convolutional Neural Network as the data is not in the form of visual imagery.

S Saravi et al. [3] discussed the use of Artificial Intelligence on the big data, that was collected from previous other flood disaster events to train the algorithm about past events and also extracted information and patterns, and understanding flood's behavior to improve the degree of preparedness and prevent damage in the events of disaster. Random Forest technique is being used as it guarantees the highest rate of accuracyforclassification.J48 decision tree and Artificial Neural Networks (ANN) are next in line for predicting flash flood and Lazy methods. Disaster-monitoring methods used in the paper are based on detection algorithm based on change, where the area affected can be recognized using a complex study on pre disaster and post disaster event data. This paper helped us understand the working of ANN and Time Series Algorithm to predict upcoming disasters, which is being implemented in the pre disaster analysis phase of this paper.

F Ahmad et al. [4] used Machine Learning models for early identification and future earthquake prediction by analyzing continuous time series data. Seismic stations continuously gather data which can be used to identify earthquake prone regions. The first phase uses the K-means algorithm for clustering applications to gives the result as clusters for different earthquake locations.

A Ranit et al. [5] developed a model for a reliable flood forecasting system where the reliability is based on the ability of the system to provide advance warning. The modelis designed based on the scale, types of flood, flooding behavior, types of landscape. The various approaches used are statistical, ANN and clustering approaches. This paper helps us understand the working of application software for the deployment of a disaster prediction, warning system and post disaster report.

S Abdullahi et al. [6] designed a Flood monitoring system which combines the uses of water level sensors and flow sensors. It uses neural network and Microsoft's Azure Machine Learning. The updated data from sensors is made available using ANN, weather radar images and hydrological flood mappings. Azure Web Services is used to predict with the Common Information Space (CIS) along with neural state cloud. Flow rate monitoring and water level monitoring are evaluated based on accuracy, recall and precision and ROC curve for true positive versus false positive. This paper is used to understand the deployment of Azure services for post disaster analysis.

J Caldera et al. [7] conducted a study on identifying parameters which reflected the severity of the eruptions, intensity, impacted population, fatalities, impact area, damage funds and GDP per capita that were required to determine the severity level. The paper suggests a formula based on the Link Function for logit model, for evaluating the damage due to the disaster. We make use of this formula to perform a fund estimation based on the above parameters. This equation has been tested to produce 90% confidence level.

J de Boer et al. [8] worked to give a meaningful definition to the term disaster and classify the severity scale of a disaster. The paper suggests the use of the algorithm of conceptions to define the severity of the disaster post a destructive event. This paper helps us define the post disaster severity model using the Disaster Severity Scale (DSS) it suggests.

R Below et al. [9] highlights impacts of disaster, and draws attentions to its problems and areas of management of disaster preparedness. This paper helped us understand the grouping of the meteorological, hydrological, climatological and biological disasters.

S.C. Wirasinghe et al. [10] performs a comparative analysis of probability of occurrence of disaster, its intensity, region of impact, death and major injuries and to categorize the extremity of the disaster into Emergency, Disaster, Catastrophe, Calamity and a Cataclysm. This paper helped us determine the fatality range based on the category of disaster in the Disaster Scale.

V Hristidis et al. [11] proposed different paths to analyze and manage information produced during calamities by surveying and organizing current knowledge for the analysis, while considering the present challenges and future research prospects. It makes use of a Business Continuity Information Network (BCIN) to organize the findings over data mining, information acquisition, processing, retrieval and recovery of insights. This paper helped us devise the various visualizations associated with the data organized and manipulated for post disaster management.

3. METHODS AND TECHNIQUES

A. Linear Regression:

Linear regression can be used to develop a binary relation for different variable by adjusting the linear equation between the variables against data taken. First variable is the explanatory variable and another variable is dependent on the first. It gives the equation in the form $Y = bX + a$, wherein X can be considered an explanatory variable and Y can be considered a dependent variable. The slope line is set to b, and the intercept determined is a (Consider y, in which $x = 0$).

B. Perceptron:

Perceptron is an ANN algorithm which can be used for the production of a binary classifier. The algorithm takes the input data that is binary classified, with the class membership and gives the output as a line that seems to separate the data of one class from data of the other class.

C. RFC:

Random forest is an algorithm for supervised learning of machines which is used to apply and classify problems that develops decision tree on the given data and then determines prediction from each and finally uses voting to select the best solution. It is the set method that is better than a single decision tree as it averages the result to reduce over-fitting.

4. Proposed Scheme

Pre-Disaster analysis consists of prediction of the disaster on a timely basis. Prediction becomes a vital aspect of managing a calamity as it minimizes loss damage that can occur to the species and resources around the are prone to disaster. The pre disaster analysis of datasets makes use of the timely data and processes it in order to train other models for prediction of disasters like flood and earthquakes using a vector xt of observations made at different time instants. The Perceptron model is used for the prediction of earth quakes. The Pre Disaster Analysis Model is reflected in Fig. 1.

The Post-Disaster Analysis makes use of Random Forest Classifier Model to predict, compare and classify data sets from an on-going disaster and satellite images of the disaster affected areas to detect anomalies to analyses the extent of damage caused due to the disaster. Ituses RFC to extent of the damage occurred in the event of the disaster by comparing with previous disasters. It uses Classification Model and combines the results of the Classification Model estimate the funds with better accuracy. Post Disaster Analysis is reflected in Fig. 2. The block diagram representations of the pre disaster and post disaster models have been illustrated in details in the figures, namely Fig. 1 and Fig. 2.

5. Comparative Study

The comparative analysis of each of the research techniques reference in this paper is given in the Table 1. This helps us to understand and justify the techniques used in this paper. To generate relief and alert measure reports for the duration of the disaster using Support Vector Machine considering factors affecting and leading to the calamity. The preparedness for a disaster can be defined by having an accurate system that predicts the occurrence of a disaster and manages the events and accurate reports before, during and after the occurrence of the disaster. This can be done by retrieving geographical, meteorological and geological data from reliable sources for the prediction and preparedness of the disaster and thereby, alerting and evacuating the public and deploying precautionary measures to reduce the extent of the damage caused due to the disaster.

6. Implementation

Introducing the disaster crisis management, a model is built to predict and analyze natural calamities based on previous occurrences and current circumstances. The management portal accepts user input and draws parallel conclusions with the help of supervised learning models trained using authenticated datasets. To predict a flood, linear regression is used where rainfall is plotted against the water level on a hyper plane. Rainfall here is considered as the independent variable and water level is considered as the dependent variable. The dataset will be organized with a time series stamp, thus allowing the maintaining of a chronological sequence of numbers to improve accuracy. The regression model helps in determining the strength of the predictors, since the model chosen is a simple linear algorithm for regression, the basic and simple idea is to allow determination of rise/ maintenance of water level based on the same, that is the changing value of the water level based on the rainfall, forecasting and effect, predicting the possible occurrence of a flood and trend forecasting.. Fig. 3 represents a prototype of the hyper plane that is created for the purpose of prediction.

To be enable earthquake prediction, a multilayer perceptron model is used, with the layers helping in prediction. The input layer accepts the required parameters for analysis, that is- latitude, longitude, depth and time. The hidden layers use these parameter values. The output layer predicts the possible magnitude of the earthquake generated by the model and its learning. Fig. 4 is a representation of a multilayer perceptron demonstrating the three layers used by the algorithm.

Post disaster effects may be catastrophic. Relief and rescue measures demand high speed and efficiency to help the affected area and avoid further damages due to negligence. The main idea having to develop a model to estimate funds for, and based on, these measures allow a better rescue protocol. The Random Forest Classifier model helps categorize a calamity based on the scale of damages it has caused. This severity is a numerical value that helps us further determines the cost required for the same. Fig. 5 is a pictorial representation of a random forest classifier. The three classes A, B and C are the categories under which a particular category will fall. Finally the intensity is determined and a corresponding fund, derived mathematically for the same is generated.

The data flow diagrams for the proposed application are as follows:

- Level 0: Fig. 6 represents the level 0 of the data-flow diagram (DFD or commonly called context diagram which represents the data base system as a complete and focuses on data and its relationship with outside entities.
- Level 1: The level 1 of the data flow diagram (DFD) shown in Fig. 7 gives more details than a level 0 DFD but less than a level 2 DFD. The level 1 data flow diagram helps break crucial process into mini process which is later analyzed and bettered on a more thorough and minute basis.
- Level 2: Fig. 8 shows the level 2 of the data flow diagram (DFD) which gives thorough representation of the process which constitute a knowledge system versus level 1 DFD. Level 2 DFD is essential to schedule or report the model of a system.

Prediction / Probability of a flood:
Flood is the overflow of water above ground-level on a usually dry ground. Floods may occur due to snowmelt, a dam break, overflowing river, or heavy rainfall. Most floods in the Indian subcontinent occur due to excess rainfall ultimately leading to rise in water level of the river. The idea behind prediction of floods is to obtain a line that helps best fit the previous data. The best fit line is true when the total prediction error(for every data point) are minimum. Error is the distance between the points upto the regression line. A regression line can be determined using the training data,which will give theminimum error. The linear equation can then be used for any new data as shown in Equation (5.1).

Y (predicted) = b0 + b1*x -------------------------------Eq (5.1)

The values b0 and b1 has to selected such that the error is minimized. If the sum of the squared error is considered to perform evaluation of the model, then the aim is to form a line that best reduces the error as shown in Equation (5.2).

Error = Σ (actual_output – predicted output)**2 -----Eq (5.2)

If the error is not squared, then it will cancel out the positive and negative points. The value of b0 for a model with only one predictor is given as:
b0=y‘-b1x---Eq (5.3)

$b1 = [\Sigma (y_i-y') (x_i-x')/\Sigma (x_i-x')^2]$

Exploring value of 'b1'

- If in case b1 value > 0, the predictors x and the goal y have a positive relation. This means rise in x directly adds to y.
- If in case b1 value < 0, predictors x and the goal y have a negative relation. This means rise in x decline in y.

Exploring value of 'b0'

- The prediction gets useless with just b0, when the system does not include x=0. For example, consider a data point that gives the relation between height x and weight y. When x=0 i.e, height is 0, the equation will have only b0 which is completely unrealistic since the real-time value of height and weight will never be zero. This happens when the model considers values beyond its scope.
- 'b0' is an average value of predictions where x=0, if the model includes the value 0. Making zero value for all the predictions is not possible in real-time.
- b0 guarantees the remaining values will give a mean value zero. The regression line can be compelled to cross the origin if there is no 'b0' term. Both the prediction and regression co-efficient will be biased.

The datasets used were taken from the Government of India's official website (data.gov.in). The model is tested on two parameters namely rainfall and water level. In the regression model, rainfall (predictor) in mm is independent variable at x axis and the dependent variable at y axis is water - level (target) in the river.

The data is stored in a timely manner. This time series methodology allows a better understanding of data collected for rainfall. The first step is to read the csv file containing cumulative versus current rainfall against each time-stamp, mainly done to remodel and understand the structure of a dataset in terms of an array. The same is repeated for the hourly water level dataset. This dataset too contains and holds data in a timely manner. Both these datasets are shaped in python. To better understand the data from both, graphs are plotted as displayed in Fig. 9 and Fig. 10.

The goal is perform prediction of the water level which depends on the current rainfall, thus the two are combined and plotted against one another, the predictor variable rainfall is plotted against the target variable water level from the previous datasets as shown in Fig. 11 and Fig. 12.

The graph is plotted and thus the hyper plane is developed. This allows the model to run training and test data. The best points are selected to plot the line to help predict the water level, ultimately letting us know the possibility of a flood.

Prediction of earthquake:

Earthquake happens when two squares of the earth out of nowhere slip past each other. The surface where it slips is known as the fault or fault plane. The area underneath the earth's surface where the quake starts is known as the hypocentre, and the area straightforwardly above it on the surface layer of the earth is known as the epicentre. Feed-forward mechanism is a artificial neural network algorithm which aims for approximation of a function given f. Consider the example, given a classification equation $\mathbf{y = f *(x)}$ which helps map and obtain value of x to a class y, the problem find the most accurate estimation for classification which is done using map, $y = f(x ; \theta)$ and determining the most relevant value θ for it. The Learning problems consists of various function which have been tied to each other. The network of 3 layer gives the Equation (5.4):

f(x) = ((f (1)(x))f (2)) f (3)--------------------------------Eq (5.4)

Every layer contains measures which give an affine kind of transformation for linear sum of values. Every layer given by the Equation (5.5):

y = f (b_fc + W_fcxT)-----------------------------------Eq (5.5)

where. activation function is given as f, W_fc is a collection of parameters, in every layer, x is the vector of inputs, or possibly results for pre layers, and the vector of bias is b_fc. The best practice for training a Neural Network is to try and normalize the data in order to obtain a mean that is close to 0. Normalizing the data leads to faster convergence. The network accepts a total of four inputs, namely the latitude, longitude, depth and time. Since the data has varying scales, normalization is performed on the input layer. The dataset is a universal collection of earthquakes that have occurred in the past. It includes various parameters that include the input requirements along with seismic errors and magnitude to help train the model better. When a neural network is trained on the training set, it is initialized with a set of weights. These weights are then optimized during the training period and the optimum weights are produced. The strategy involves initializing the weight to some arbitrary value and refining repetitively obtains lesser loss. Refining can be done with motion towards the intend which is given by the degree of function in gradient loss. It is significant to initially assume rate of learning that defines the rate for the model moving in all repetitions. This function can be called at the beginning of the program before tensors or graphs and other structures have been created, and before devices have been initialized. It switches all global behaviors that are different between TensorFlow 1.x and 2.x to behave as intended for 1.x.
Activation functions as shown in Fig. 13, describes the data interaction relationships using non-linear manner. The function helps give the algorithm the ability to be open to describe the random relation. The rectified linear unit of activation (ReLU) model is defined as a piecewise-linear function which gives the result as the input, if is positive, or, it outputs zero. It is considered as the initial activation function for various kinds of NN as it is a system which is simple for handling to get better performance.

The loss function has been used to estimate the performances of classifier. The loss function is predicted greater when the estimated class cannot be corresponding to real class, or else it will be less. In some cases, the issues with over-fitting and under-fitting may occur while preparing the system. The system’s performance is high while preparing datasets except the tests data. Therefore, to prepare the system, the optimizing technique performed which requires an optimizer and loss function. The optimization procedure helps find data using weights, W which helps minimize the function of loss. The loss function is shown in Fig. 14.

Post Analysis Model:
Post Disaster Analysis model uses the Random Forest Classifier model to analyze the damage caused due to the disaster and to estimate funds for relief measures. RFC Classifier is particularly used as it generates a decision tree whose prediction by committee is more accurate than any individual tree. This model uses the dataset which consists of the various damage grades of buildings in an area, the materials used, the vulnerability to damage score, the no. of floors, the height and area per square feet of the building, the ward and location of the building, the roof type, how old the building is, amount of repair required, etc. Based on this dataset, the RFC model creates a correlation matrix to classify the various categories of damage caused due to a disaster into features and categorizes all the dataset values to a relevant cell. The difference in the fund estimates proves the efficiency. Post classification, the model implements a simple python code that includes another range within itself. This range allows estimation of data on a decimal basis, thus improving accuracy. Thus in this case, the 4 inputs from the user may differ on a 0.3

or 0.5, which only leads to more clarity and better results.

7. Performance Analysis

Training loss, for each example in the training sets, is defined as the sum of the errors made. In Fig. 15, the graphs represent the loss of the two different models, where the graph in the left shows greater loss and the graph on the right shows lesser loss. Arrows in the graphs are a representation of the loss and the blue lines are a representation of the predictions. Confusion Matrix describe the performance of the models. For a binary classification problem, samples generally belong to two different classes: YES / NO. Also, there is one more classifier that estimates classes for an obtained samples. While test are done for the system for 166 different sample, the result obtained is given as shown in Fig. 16.

MSE can be considered as mean square error diff among an estimate and the datasets. This is an identical calculation to the calculation of the variance of a statistic, where the estimate is the mean. Smaller MSE generally indicates a better estimate, at the data points in question. The RMSE has the same unit as the dependent variable (DV) which is the water-level in this case. There is no absolute good or bad threshold, however you can define it on the basis of your Dependent Variable. For a data which normally ranges between 0 and 1000, an RMSE value of 0.7 is considered small, but if the ranges between 0 and1, it is not considered that small. The smaller the RMSE value, the better is the theoretical claims on the levels of the RMSE when we know what is expected from the water-level.

If loss in preparation is set to very low values when compared to validating losses, the system may be over-fitting. Resolution for such issue is for reducing the size of the networks, or upscaling the value for dropout. If the loss of training loss or the loss of validation is almost as the system is under-fitting. Increasing the size of the system is the only resolution. The Earthquake prediction model which uses the Neural Network Model produces a loss in Validation of 0.38 and Training Loss of 037 for iteration varying from 0 to 800 for an epoch value of date time set to 01/01/1970. The lower the value of the validation loss, the more accurate is the prediction. The accuracy can therefore be further increased by providing more data for training and expanding the epoch value.

Once the Random Forest model has analyzed the data, correlated it and classified the data in damage grades using decision trees, it uses the decision tree to generate a cumulative damage caused to property and life. The model re-samples the data with 3-class (Low, Medium and High Grades) to obtain a normalized confusion matrix with highest precision, accuracy and recall score, as shown in Fig. 17.

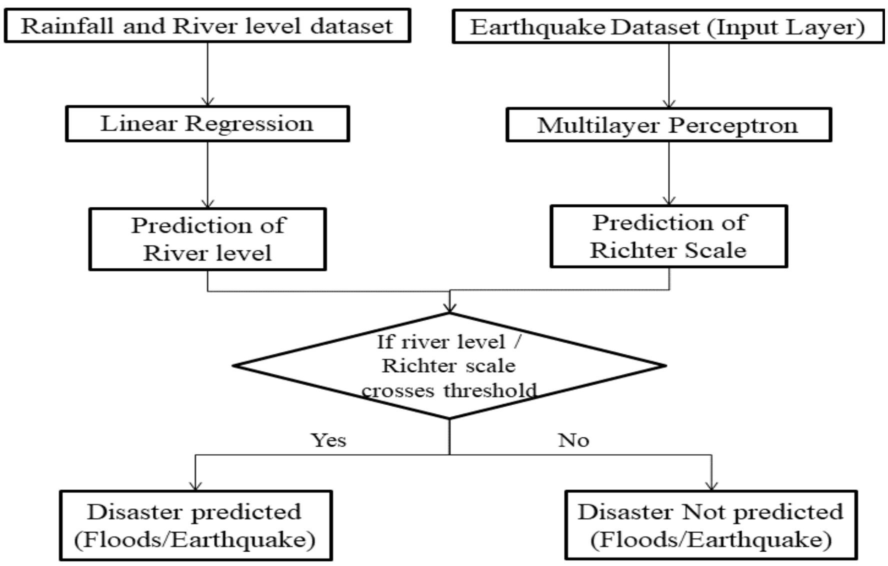

Figure 1.– Pre Disaster Analysis

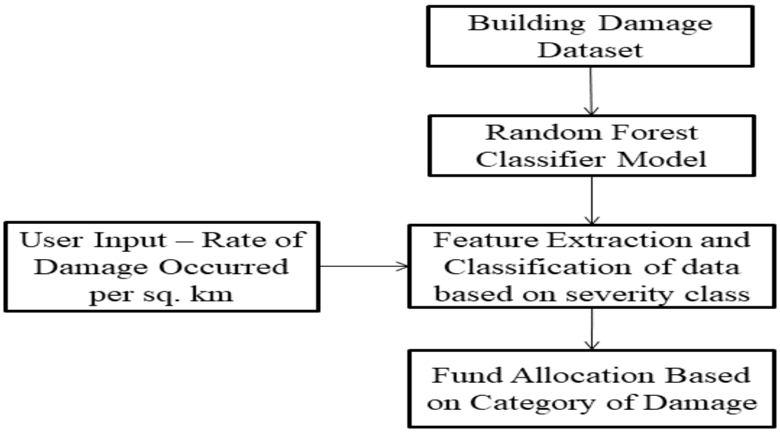

Figure 2. Post Disaster Analysis

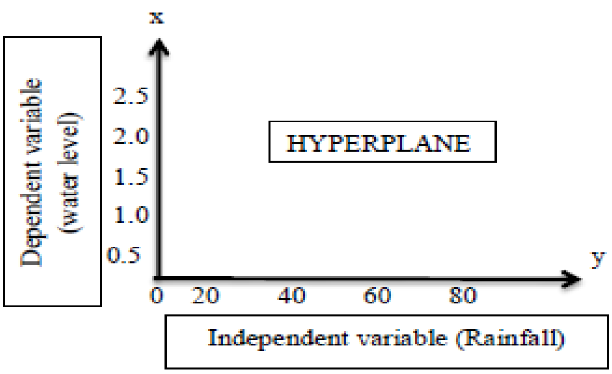

Figure 3. Logistic Regression

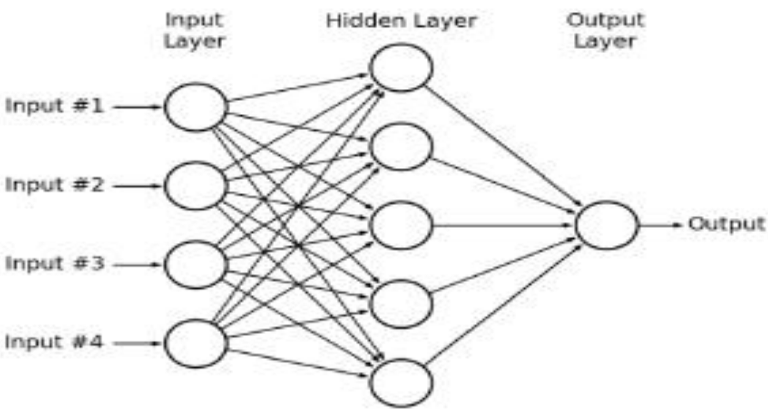

Figure 5. Random Forest

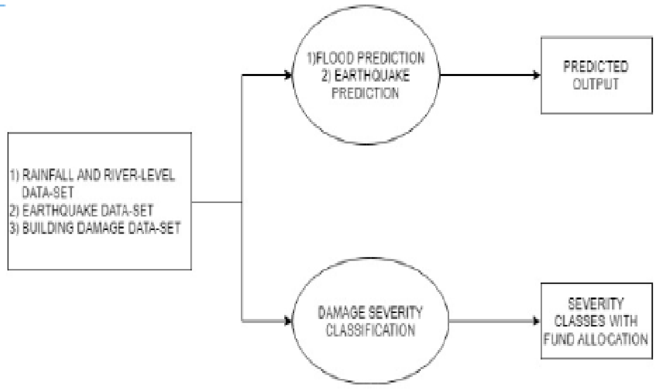

Figure 6. DFD Level 0

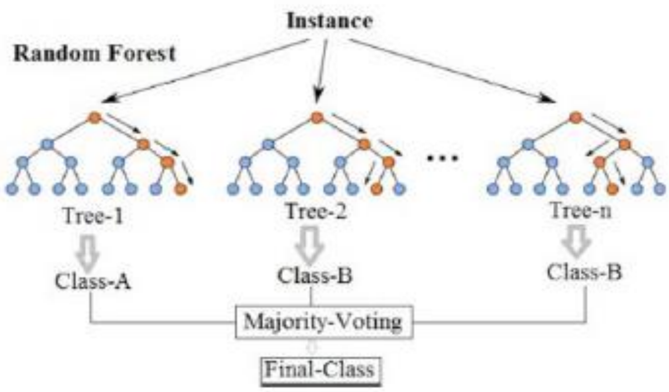

Figure 7. DFD Level 1

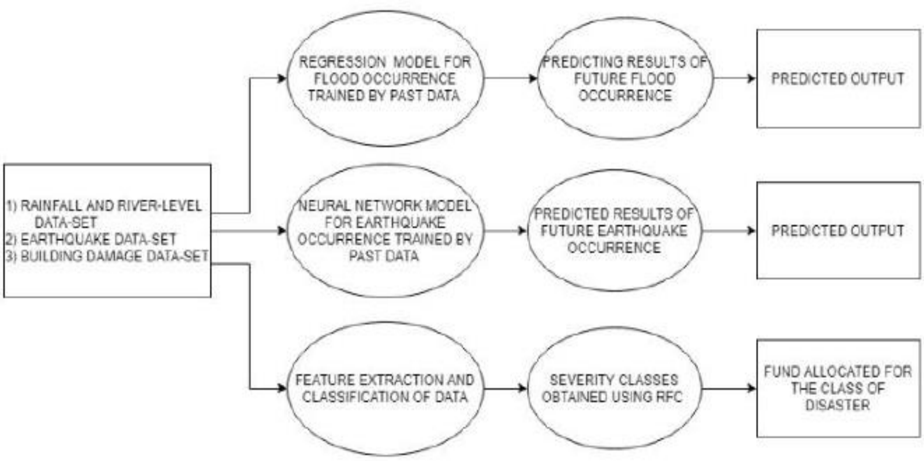

Figure 8. DFD Level 2

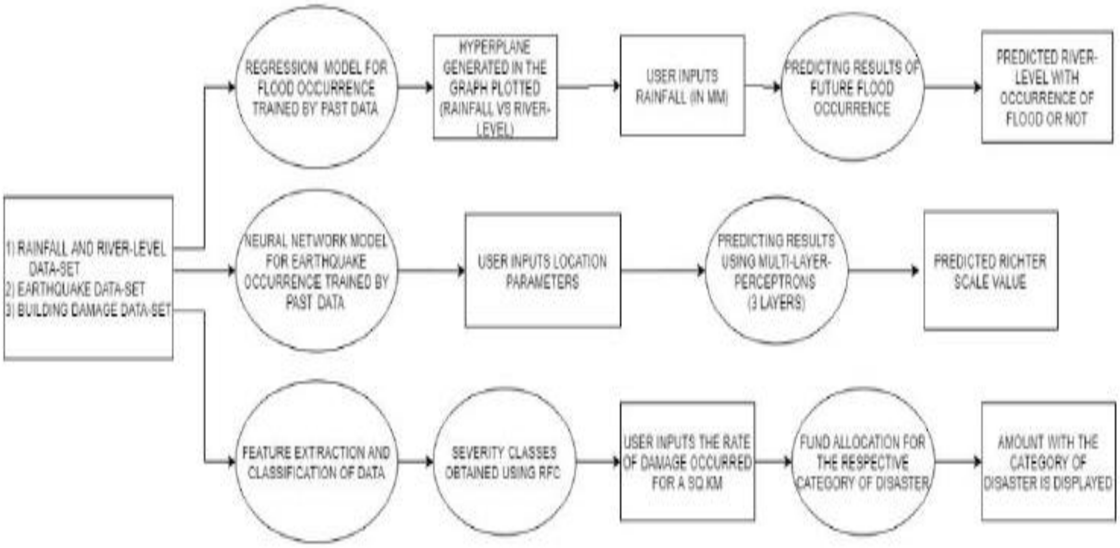

Figure 9. Time Versus Rainfall

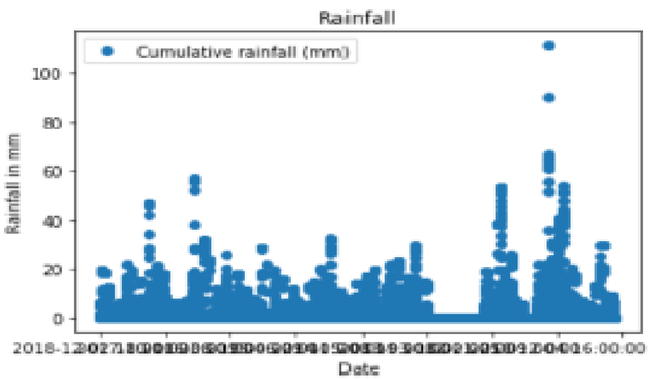

Figure 9. Time Versus Rainfall

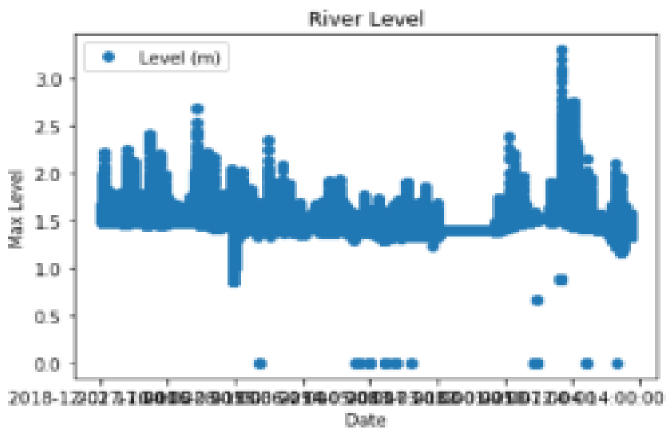

Figure 10. Time Versus Max River Level

	Date/Time	Current rainfall (mm)	Cumulative rainfall (mm)	Level (m)
0	2018-12-02 18:00:00	0.0	0.0	NaN
1	2018-12-02 17:00:00	0.0	0.0	1.61
2	2018-12-02 16:00:00	0.0	0.0	1.61
3	2018-12-02 15:00:00	0.0	0.0	1.62
4	2018-12-02 14:00:00	0.0	0.0	1.63

Figure 11. Time Series for chronological readings

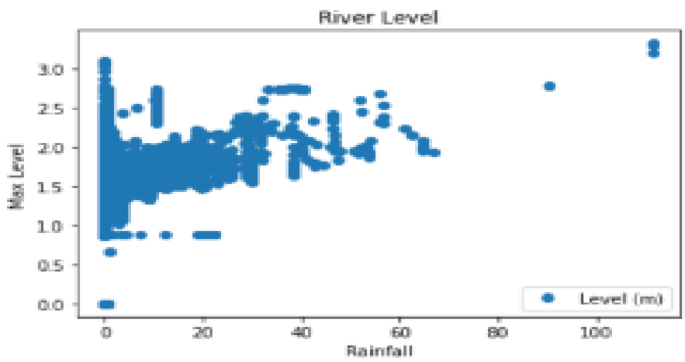

Figure 12. Rainfall Versus River Level

```
#Layer 1
matmul_fc1=tf.matmul(X, W_fc1) + b_fc1
h_fc1 = tf.nn.relu(matmul_fc1)   #ReLU activation
#Layer 2
matmul_fc2=tf.matmul(h_fc1, W_fc2) + b_fc2
h_fc2 = tf.nn.relu(matmul_fc2)   #ReLU activation
#Layer 3
matmul_fc3=tf.matmul(h_fc2, W_fc3) + b_fc3
h_fc3 = tf.nn.relu(matmul_fc3)   #ReLU activation
#Output Layer
matmul_fc4=tf.matmul(h_fc3, W_fO) + b_fO
output_layer = matmul_fc4  #Linear activation
```

Figure 13. Activation Function

```
#Loss function
mean_square =  tf.reduce_mean(tf.square(Y-output_layer))
train_step = tf.train.AdamOptimizer(learning_rate).minimize(mean_square)
```

Figure. 14 – Loss Function

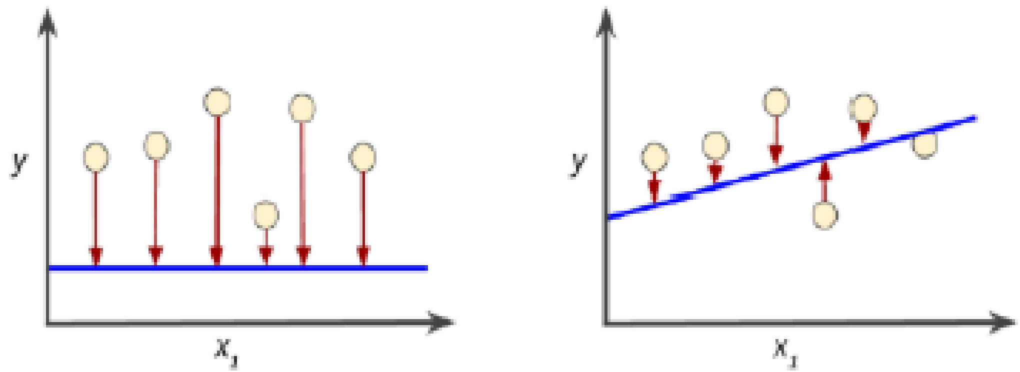

Figure 15 . Result analysis for linear regression

n=165	Predicted: NO	Predicted: YES
Actual: NO	50	10
Actual: YES	5	100

Figure 16. Confusion Matrix

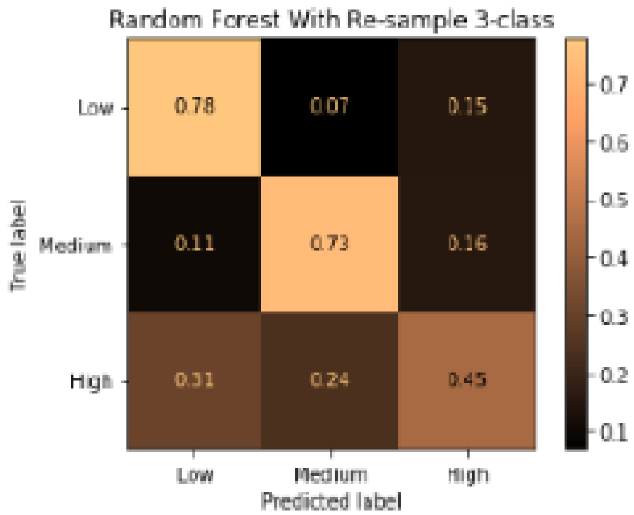

Figure 17. Confusion Matrix of RFC (re-sampled

Table 1. Research Techniques comparisons

Author & Year	Dataset	Techniques	Results/Accuracy

Mohammed Khalaf et al. (2018) [1]	"Flood Data and Resources" dataset from environment agency website.	Support Vector Machine, Random Forest Classifier, Random Oracles Model and Linear Neural Network.	The accuracy measured on the basis of AUCs with 3 classes of 0.994 for RFC and 0.858 for LEVNN
Joe Marlou A. Opella et al. (2019) [2]	Topographic map from NAMRIA, Rainfall dataset of the Boholfrom 1961 to2017.	Support Vector Machine, Convolutional Neural Networks	CNN and SVM produce a more accurate flood map when put together in comparison to when deployed individually.
Amitkumar B. Ranit et al. (2018)[3]	-	Artificial Neural Networks, K-Means Algorithm	Flood Forecasting and Warning System is reliable
Salami Ifedapo Abdullahi et al. (2018)[4]	-	Azure Web Services	98.9% accuracy and 100% precision using 3 hidden layers
C.P.Shabariram, Dr. K.E.Kannammal (2017) [5]	USGS Datset	Map Reduce Model	Generateed graph useful in identifying shaky places and fault Lines
Sara Saravi et al. (2019)[6]	Dataset taken from National Climatic Data Centre(NCDC), National Oceanic and Atmospheric Administration (NOAA) and Federal Emergency Management Agency (FEMA)	Random Forest Classifier, J48 Decision Tree, Artificial Neural Networks, Lazymethods	RFC has an accuracy rateof 80.49%, ANN has an accuracy rate of 77.44%
Faraz Ahmad et al. (2019)[7]	USGS Dataset	K-Means Clustering, Hierarchical clustering	Hierarchical clustering gives better efficiency with respect to entropy, But has lower co-

			variance measure than K-Means
Qing-Quan Tan et al. (2017)[8]	-	Geographic Information System	Datsets collected using GIS can be used for research purposes

8. CONCLUSION

A comparative study of these papers reveals that proper planning is required to determine the usefulness of data. Even beyond that the datasets require sufficient count of records for our proposed model to deliver adequate results. In order to make the application more reliable, a time series algorithm must be incorporated along with Artificial Neural Networks as the proposed methodology contains modules for both prediction and post analysis of disasters. This is because data updated more regularly will be more advantageous to prepare for a disaster and analyze the amount of resources required to recover from one.

REFERENCES

[1] Mohammed Khalaf, Abir Jaafar Hussain, Dhiya Al-Jumeily, Thar Baker, Robert Keight, Paulo Lisboa, Paul Fergus, Ala S. Al Kafri, "A Data Science Methodology Based on Machine Learning Algorithms for Flood Severity Prediction", 2018 IEEE Congress on Evolutionary Computation (CEC), pp. 230-237, July 9, Liverpool john moors university, United Kingdom.
[2] Joe Marlou A. Opella, Alexander A. Hernandez, "Developing a Flood Risk Assessment Using Support Vector Machine and Convolutional Neural Network: A Conceptual Framework", 2019 IEEE 15th International Colloquium on Signal Processing & its Applications (CSPA 2019), pp. 260-265, 8 -9 March 2019, Penang, Malaysia.
[3] Amitkumar B. Ranit, Dr. P.V. Durge, "Different Techniques of Flood Forecasting and their Applications", 2018 3rd IEEE International Conference on Research in Intelligent and Computing in Engineering , August 22nd, Universidad Don Bosco (UDB), El Salvador.
[4] Salami Ifedapo Abdullahi, Mohamed HadiHabaebi, and Noreha Abdul Malik, "Flood Disaster Warning System on the go", 2018 IEEE 7th International Conference on Computer and Communication Engineering (ICCCE), 19th September, Kuala Lumpur, Malaysia.
[5] Sara Saravi, Roy Kalawsky, Demetrios Joannou, Monica Rivas Casado, Guangtao Fu and Fanlin Meng, "Use of Artificial Intelligence to Improve Resilience and Preparedness Against Adverse Flood Events", Volume 11, Issue 5, 2019, Article No. 973, 9 May 2019, Molecular Diversity Preservation International, Switzerland.
[6] Faraz Ahmad, Om Ashish Mishra, ShivamBhagwani, Jabanjalin Hilda J, "A Model for Empirical Earthquake Prediction and Analysis in a Data Intensive Environment", International Journal of Innovative Technology and Exploring Engineering, ISSN: 2278-3075, Volume-8, Issue-6S3, April 2019.
[7] Caldera, H. J. &Wirasinghe, S.C., 2014. "Analysis and Classification of Volcanic Eruptions" In R. R. Rapp & W. Harland, eds. The 10th International Conference of the International Institute for Infrastructure Resilience and Reconstruction (I3R2). West Lafayette, Indiana: Purdue University., pp. 20–22, May 2014.
[8] de Boer, J., "Definition and classification of disasters: Introduction of a disaster severity scale." The Journal of Emergency Medicine, 8(5):591–595, 1990.
[9] Below, R., Wirtz, A. & GUHA-SAPIR, D., 2009. "Disaster Category Classification and peril Terminology for Operational Purposes." Centre for Research on the Epidemiology of Disasters (CRED) and Munich Reinsurance Company (Munich RE), October 2009.
[10] Wirasinghe, S.C., Caldera, H. J., Durage, S. W., &Ruwanpura, J. Y. "Preliminary Analysis and Classification of Natural Disasters." In P. H. Barnes & A. Goonetilleke, eds. The 9th Annual International Conference of the International Institute for Infrastructure Renewal and

Reconstruction. Brisbane, Queensland, Australia: Queensland University of Technology, p. 11, 2003.

[11] Hristidis, V., Chen, S.-C., Li, T., Luis, S., & Deng, Y. "Survey of data management and analysis in disaster situations." Journal of Systems and Software, 83(10):1701–1714, 2010