Flood Predicition Using AI Model

T.Dharani, T.KumudbenJoshi, K.Sai Himaja ,Mrs.Vinmathi

Student, Dept of Computer Science Engineering, PEC College, Tamil Nadu, India
Student, Dept of Computer Science Engineering, PEC College, Tamil Nadu, India
Student, Dept of Computer Science Engineering, PEC College, Tamil Nadu, India
Associate Professor, Dept of Computer Science Engineering, PEC College, Tamil Nadu, India

***

Abstract - Floods are one of the most catastrophic natural disasters, and modelling them is extremely difficult. The progress of flood prediction models has helped to risk reduction, policy recommendations, the reduction of human life loss, and the reduction of flood-related property damage. To replicate the complicated mathematical expressions of physical flood processes, neural network approaches have made significant contributions to the evolution of prediction systems during the last two decades, enabling higher performance and cost-effective solutions. The full dataset will be analysed using a Multi-Layer Perceptron Classifier (MLP) to collect various information such as variable of finding identification, missing value treatments, data validation, and data cleaning/preparing.The performance of the algorithms will be shown in the form of graph and to understand it in better way confusion matrix graphs will be represented

Keywords—Dataset, python, preprocessing, Algorithm Calculation

I. INTRODUCTION

Flood prediction necessitates quantitative knowledge of infiltration and runoff dynamics, which is often acquired at a local level. When scaling up local measurements to watershed scale, the catchment's organisation must be taken into account (connectivity and patchiness).

To this end, we devised a new method for mapping soil flood susceptibility based on two steps:
(1) identifying flow mechanisms at the plot scale, and
(2) scaling this knowledge up to the catchment scale. The marked differences in textural and structural porosities between Forest and grassland plots appear to control runoff processes.

2. LITERATURE SURVEY

1. Hybrid Machine Learning Approach for Classifying Aerial Images of Flood-Hit Areas

The classification is performed on an image dataset that is divided into two subgroups: aerial photos with flood-affected areas and aerial images without any flooded areas. The downloaded dataset is randomly partitioned before image categorization. The SVM classifier is combined with k-means clustering in this classification. The SVM classifier is used to train the dataset's numerous classifications. Comparative studies reveal that the SVM classifier is one of the most preferred among all the others that are commonly employed in many remote sensing applications, and that it gives much superior accuracy than other algorithms such as decision trees. It was then utilised to solve difficulties involving multi-class classification.

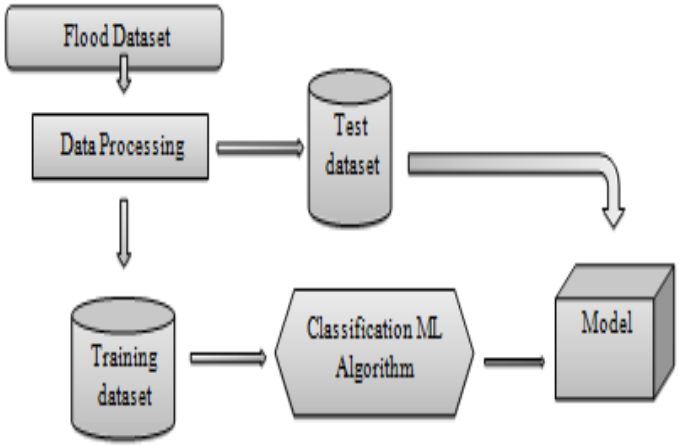
2. Flood Prediction Using Multi-Layer Artificial Neural Neural Network in Monitoring System with Rain Gauage, Water Level, Soil Moisture Sensors.

This study was about to implement a real-time of monitoring system capable of measuring parameters such as rainfall intensity, soil moisture, water level and rate of increase in water level. The different sensor was integrated into a system, wherein the data was logged and stored. A prediction model based on multi-layered artificial neural network was developed and was tested using an actual setup. This study validated the response of the multi-layered network model. The flood the prediction model had a small deviation with reference to the actual water level, with RMSD of 2.2648.In the Philippines that was a major issue as it leads to damage of properties, damage to infrastructures or even loss of Lives.

3. PROPOSED SYSTEM

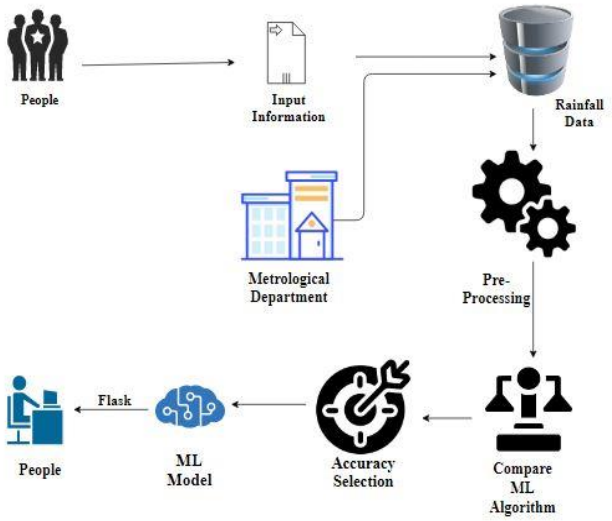
There are Different types of floods like river flood, and flash flood have been observed over the years. A given dataset from several sources would be combined to generate a generalised dataset, which would then be used to extract patterns and obtain accurate findings. The training and test sets of data collected for predict the given data are separated. In most cases, 7:3 ratios are used to divide the Training and Test sets. The Multi-Layer Perceptron Classifier Data Model will be

applied to the Training set, and Test set prediction will be made based on the test result accuracy.



4. SYSTEM ARCHITECTURE

Collect the information needed for a rainfall forecast From people or ancient records So, once the data has obtained, use certain pre-processing approaches before testing and training them with the machine learning algorithms.



After running the dataset through the ML algorithms (Logistic Regression, Decision Tree, and the Random Forest) and comparing their accuracy, we can determine which approach is best for flood prediction. Finally,we using that approach, we can develop a machine learning model and make predictions.

5.Variable Identification Process / data validation process:

Importing library packages and loading the specified dataset. Identifying variables based on data shape, data type, and the evaluating missing values and duplicate values. A validation dataset is a sample of data kept back from training your model that is used to measure model skill when tweaking models and techniques for making the most of validation and test datasets while evaluating your models. To examine the uni-variate, bi-variate, and multi-variate processes, rename the given dataset and eliminate the columns, among other things. Data cleaning methods and techniques will be differ from dataset to dataset.The primary goal of the data cleaning is to detect and the remove errors and anomalies to increase the value of data in analytics and decision making.

6. COMPARING ALGORITHM WITH PREDICTION IN THE FORM OF BEST ACCURACY RESULT:

It is important to compare the performance of multiple different machine learning algorithms consistently and it will discover to create a test harness to compare the multiple different machine learning algorithms in Python with scikit-learn. It can use this test harness as a template on your own machine learning problems and add more and different algorithms to compare. Each model will have different performance characteristics.

Using resampling methods like cross validation, you can get an estimate for how accurate each model may be on unseen data. It needs to be able to use these estimates to choose one or two best models from the suite of models that you have created. When have a new dataset, it is a good idea to visualize the data using different techniques in order to look at the data from the different perspectives. The same idea applies to model selection.

You should use a number of different ways of looking at the estimated accuracy of your machine learning algorithms in order to choose the one or two to finalize. A way to do this is to use different visualization methods to show the average accuracy, variance and other properties of the distribution of model accuracies.

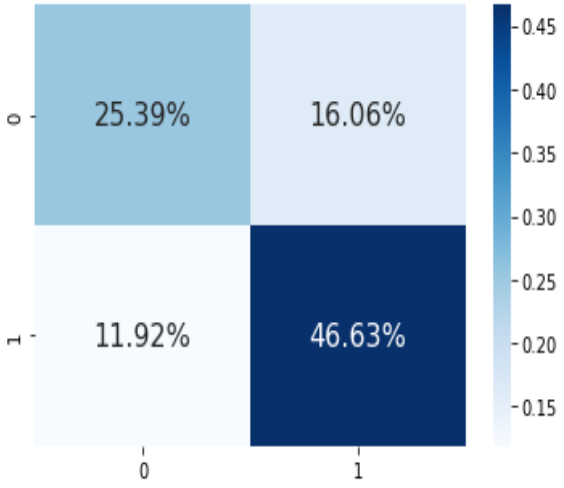
7. ACCURACY & COMPARISON OF RESULT:

MODEL	ACCURACY
Logistic Regression	71.3
Decision Tree	71.6
Random Forest	81.1
Support Vector Machine	60.5

8. SYSTEM ANALYSIS:

Performances metrics for all algorithms:

LOGISTIC REGRESSION



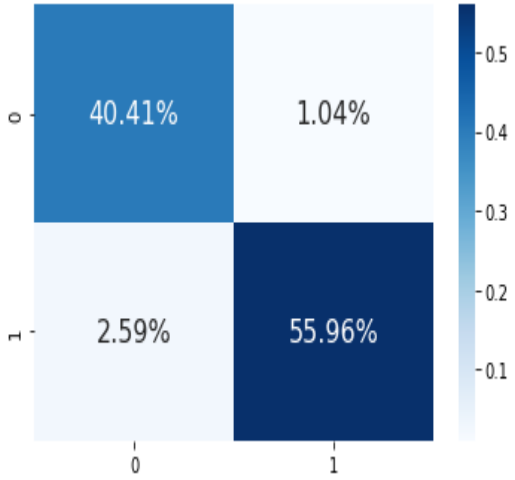
True positive (TP): 25.39%

False positive (FP): 11.92%

True negative (TN): 46.63%

False negative (FN):16.06%

DECISION TREE CLASSIFIER



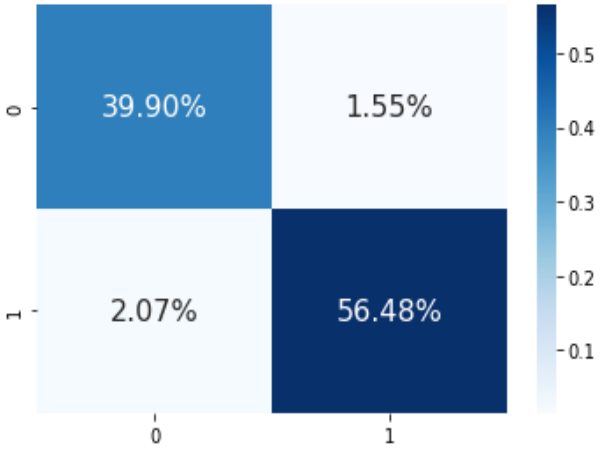
True positive (TP): 40.41%

False positive (FP): 2.59%

True negative (TN): 1.04%

False negative (FN):55.96%

RANDOM FOREST



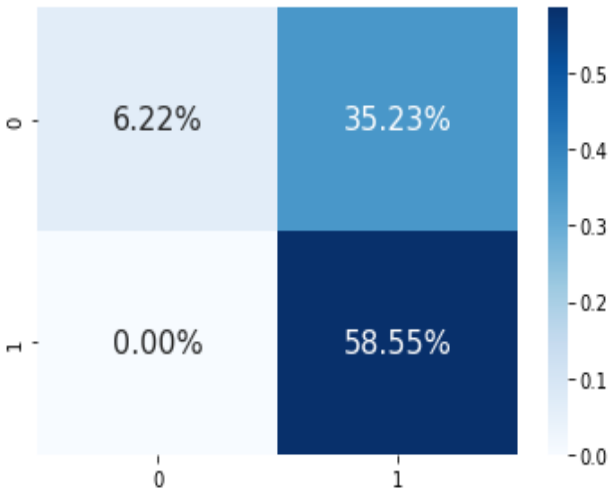
True positive (TP): 39.90%

False positive (FP): 56.48%

True negative (TN): 1.55%

False negative (FN):56.48%

SUPPORT VECTOR MACHINE



True positive (TP): 6.22%

False positive (FP): 0.00%

True negative (TN):35.23 %

False negative (FN):58.55%

9. CONCLUSION

The analytical process started from data cleaning and processing, missing value, exploratory analysis and finally model building and evaluation. Finally we predict the flash flood using the machine learning algorithm with different results. This brings some of the following insights about flood prediction.

10. FUTURE ENHANCEMENT

Disaster management wants to automate the detecting the flood happened or not from eligibility process (real time).

To automate this process by show the prediction result in web application or desktop application.

To optimize the work to implement in Artificial Intelligence environment.

11. REFERANCES

1 K. Smith, Environmental hazards: assessing risk and reducing disaster. Routledge, 2003.

2 Z. Sen, Flood Modeling, Prediction and Mitigation, 1st ed. Springer International Publishing, 2018.

3 A. Y. Sun and B. R. Scanlon, ‘‘How can big data and machine learning benefit the environment and water management: A survey of methods, applications, and future directions,’’ Environ. Res. Lett., vol. 14, no. 7, p. 28, 2019.

4 Cortes and V. Vapnik, ‘‘Support-vector networks,’’ Mach. Learn., vol. 20, no. 3, pp. 273–297, 1995.

5 K. Kravari and N. Bassiliades, “A survey of agent platforms,” Journal of Artificial Societies and Social Simulation, vol. 18, no. 1, p. 11, 2015.

6 M. Westergaard and H. Verbeek, “Cpn tools.” [Online]. Available: http://cpntools.org/

7 Ö. Kişi, ‘‘Streamflow forecasting using different artificial neural network algorithms,’’ J. Hydrol. Eng., vol. 12, no. 5, pp. 532–539, Sep. 2007

8 Glossary of Meteorology (June 2000) Flood Archived 2007-08-24 at the Wayback Machine, Retrieved on 2009-01-09

9 Brown, Richard; Chanson, Hubert; McIntosh, Dave; Madhani, Jay (2011). Turbulent Velocity and Suspended Sediment Concentration Measurements in an Urban Environment of the Brisbane River Flood Plain at Gardens Point on 12–13 January 2011. Hydraulic Model Report No. CH83/11. p. 120. ISBN 978-1- 74272-027-2.

10 Siegel, Eric (2013). Predictive Analysis: The Power to Predict Who Will Click, Buy, Lie, or Die. Hoboken, NJ: John Wiley & Sons. ISBN 978-1-118-35685-2.