Journals Pub

International Journal of
Water Resources Engineering

ISSN: 2456-1606
Volume 11, Issue 1, 2025
January–June
DOI (Journal): 10.37628/IJWRE

https://journalspub.com/journal/ijwre/

Research IJWRE

Flood Susceptibility Modeling Using Logistic Regression in QGIS Environment

Sandaru G.M.C.[1,*], Illeperuma I.A.K.S.[2]

Abstract

Flood is the most common natural disaster of Sri Lanka. Sri Lanka is suffering from extensive flood events more frequently than previously due to climatic changes. Due to rapid urbanization, there are limited land resources to live in Kalutara district which is not prone to flood events. This research aims to model the flood susceptibility in Kalutara district Sri Lanka which is having frequent flood events using Logistic Regression Machine Learning model. Rainfall, Elevation, Slope, Aspect, Stream Power Index (SPI), Topographic Wetness Index (TWI), Distance to Rivers, River Water Level were selected as flood contribution factors. Flood extent was extracted using SAR images captured from Sentinel 1 using Google Earth Engine (GEE). Elevation data was downloaded from ALOS PALSAR dataset. Logistic Regression model fitting was done using python language with some open-source libraries. Fitted Logistic Regression model was able to achieve 0.89 Area Under the Curve (AUC) value in the Receiver Operating Characteristic (ROC) curve. The results show that flood susceptibility is very high around the major water features which are Kalu Ganga and Kuda Ganga. Among the other factors, River Water Level is having the highest contribution to the flood susceptibility in the area. Rainfall does not make much contribution to the flood susceptibility over the area.

Keywords: Flood, flood susceptibility, Google Earth engine, logistic regression, machine learning, SAR, sentinel

INTRODUCTION

A balance of natural ecosystem is needed for the well-being of all living creatures of the world. But due to the changes happen in many parts of the world day by day ecosystem is changing and environmental balance is losing. This is caused to increase the frequency and intensity of natural disasters happened in the world. The frequency of intense natural disasters has been on the rise over the past 40 years [1]. The changes happen in a one place creates unbalance in the eco system and the reaction will be happened in the different place of the world. The climatic change in the world has led to an exponential increasement of flood occurrences and their severity [2].

***Author for Correspondence**
Sandaru G.M.C.
E-mail: chamindusandaru1@gmail.com

[1]Student, Department of Remote Sensing and GIS, Faculty of Geomatics, Sabaragamuwa University of Sri Lanka, Belihuloya, Sri Lanka
[2]Senior Lecturer, Department of Remote Sensing and GIS, Faculty of Geomatics, Sabaragamuwa University of Sri Lanka, Belihuloya, Sri Lanka

Received Date: November 07, 2024
Accepted Date: November 23, 2024
Published Date: January 02, 2025

Citation: Sandaru G.M.C., Illeperuma I.A.K.S. Flood Susceptibility Modeling Using Logistic Regression in QGIS Environment. International Journal of Water Resources Engineering. 2025; 11(1): 1–11p.

Floods are among the major natural extreme and dangerous events that cause loss of life and property [3]. Flood is defined as water that temporarily submerges land. It is a consequence of migration of the boundary between land and water bodies, reflecting the normal interaction of the atmosphere, hydrosphere, and lithosphere [4]. From the past humans suffered from flood events as they established their homes and agricultural lands in the river valleys.

Floods are one of most common natural disaster for the people live in Sri Lanka. Sri Lanka is suspect to Indian Ocean (IO) Monsoon system which gives systematic variation of intense rainfall over the year

in a particular area [5]. Due to climatic changes happening in the world Sri Lanka has been having unusual rainfall patterns and more frequent rainy and dry seasons. Also, annual rainfall is higher than what we had few years back [6]. Sri Lanka suffered from severe flood events in 2016, 2017, 2019, 2021 and around 14 million people were affected by floods between 2010 and 2018. Also, there are recorded death events as well due to the flood events happened in these years.

Other than that flood events are more frequent in western part of the country where the population density is higher in comparison with other parts of the country [7]. This region is having limited places to live, and it is hard to find new places for homes which aren't prone for flood events. While working on the development of the country, managing the limited space available to live should be used in an optimum way. With a flood risk map, the identification of areas that can be developed without having the flood risk and minimal disturbance to the environment is possible. Also, necessary actions for highest risk areas can be implemented by respective authorities. For these purposes, we must model the existing situation and provide appropriate solutions.

When we model the flood events, we need to have a large and accurate dataset. Until now most of the research are done as a qualitative approach to the flood risk assessment due to the limited data availability. Practically, it is hard to collect field data in a disaster situation real time to create a large data set but collecting data remotely and verifying it with field observations will be a better approach. Remote sensing techniques are very useful to create flood inventory datasets, such as by using multispectral, radar, and LIDAR satellite [8]. Satellite data is a good source of data but most of the time cloud cover will be high since most flood events are triggered by heavy rainfall. That barrier can be overrun by using SAR images because they are insensitive to cloud cover and darkness, all weather and high resolution. By using appropriate extraction methods, the flood extent can be derived [9].

High resolution data processing is requiring a powerful computer with higher processing power. Processing a set of SAR images will be a tuff situation to manage by a normal computer. In case of that as a modern solution Google Earth Engine (GEE) provide cloud base computer system with access to a massive library of satellite images and data processing using programming languages JavaScript and Python. It is used in wide range of applications including environmental protection, disaster management, and food security [10]. Flood inventory dataset can be derived from SAR images in particular time in which flood happened in the GEE environment and export it in a usable manner.

Qualitative and quantitative are two different approaches to gathering and analyzing information, each with its own strengths and weaknesses. Flood risk analysis using quantitative methods includes probabilistic approaches, deterministic analysis, and statistical procedures that primarily rely on mathematical models. Machine learning models have a distinct set of mathematical algorithms which can handle vast volumes of data, spot trends, and forecast future flood events, which are essential for flood monitoring and analysis. All dependent and explanatory variables are defined to model and by using given data model fit to the given data, compute a contribution value for each explanatory variable. Fitted model can be used to make predictions at unsampled locations or for scenario testing. Support vector machines (SVMs), Random forests (RF), K-nearest neighbours (KNN), Artificial neural networks (ANN) and Deep Learning models can be identified as most used machine learning models for flood analysis and prediction. But machine learning models are very data intensive that means it need a much bigger dataset to derive more accurate result by model fitting.

Regression models are a specific type of machine learning model used for predicting values. They learn the relationship between input features and a target variable. Regression models can be used with a smaller number of data compared with other machine learning models. Linear Regression, Polynomial Regression, Support Vector Regression (SVR), Decision Tree Regression are several most used regression models for distinct types of work.

Logistic regression is a powerful and popular statistical method that's particularly using for binary classification problems. It models the relationship between one or more independent variables with the

dependent variable that has two categories, such as 0 or 1, True or False. Result shows the probability of an observation belonging to a specific category based on the derived values according to the fitted model. It differs from the linear regression since it provides only two classified results other than predicting continuous values. In logistic regression, the sigmoid function plays a crucial role in transforming the model's output into a probability between 0 and 1. It ensures gradual changes in predicted probabilities as the input features change and enables gradient-based optimization algorithms to efficiently train the logistic regression model. Since logistic regression is a binary classifier, it can be used to predict flood occurrences. The probability of each result to be in each class out of True or False can be used as a measure of probability of flood event at a specific location. That result can be mapped accordingly to get a better visualization.

METHODOLOGY

Data Collection

Kalutara district was selected as the study area for this research since it is frequently flooded as compared to other parts of Sri Lanka.

Elevation data was collected from ALOS PALSAR dataset and after applying corrections, all the derivatives were derived.

Rainfall data was collected from the Meteorological department of Sri Lanka for both Ratnapura and Kalutara districts.

River Network data was obtained from the OSM and required corrections were done.

As software resources, Google Earth Engine (GEE), Quantum GIS (QGIS), SPYDER IDE were used.

Methodology Used

Figure 1 shows the flowchart of the methodology used.

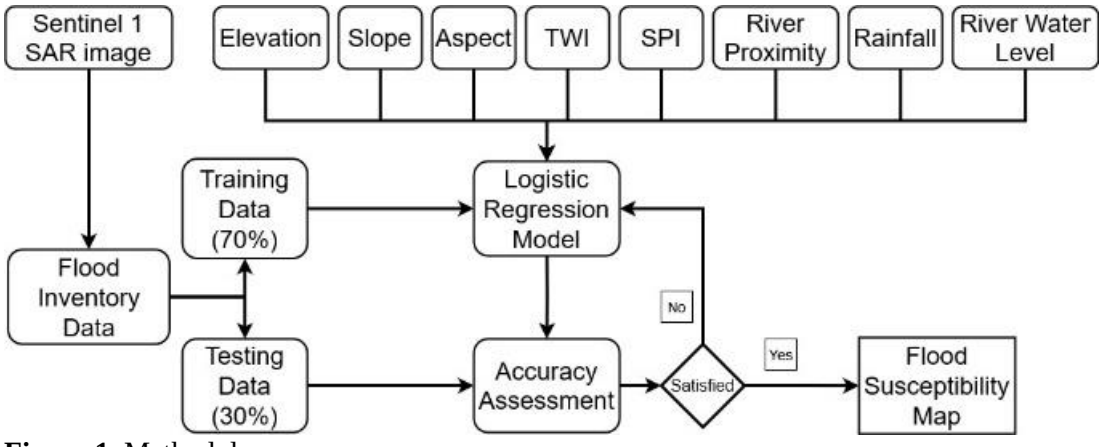

Figure 1. Methodology.

Sentinel 1 SAR images were processed with GEE. River water level was derived from the rainfall values in the Ratnapura district to evaluate the commitment of rainfall in Ratnapura to flood susceptibility in Kalutara district. Since the rainwater in Ratnapura district flows to Kalutara by two main water features, their relation was created with inverse distance to the river and mean the rainfall value of each river catchment area in Ratnapura. These maps were generated through QGIS software. Then Python with Sklearn library was used to create the Logistic Regression model. All the inventory dataset was passed as a .csv file to the model and Confusion matrix and ROC curves were generated through the model for verification. For the mapping, model coefficient for each factor was exported and mapping was done using raster calculator in QGIS software.

RESULTS ANALYSIS AND DISCUSSION

Flood Inventory Dataset

The raw data was processed for creating the flood inventory dataset. Later, the data can be inserted to the model and fit the model for accurate detection over the area. The derived data from the raw data before adding to the logistic regression model is shown in Figures 2–10.

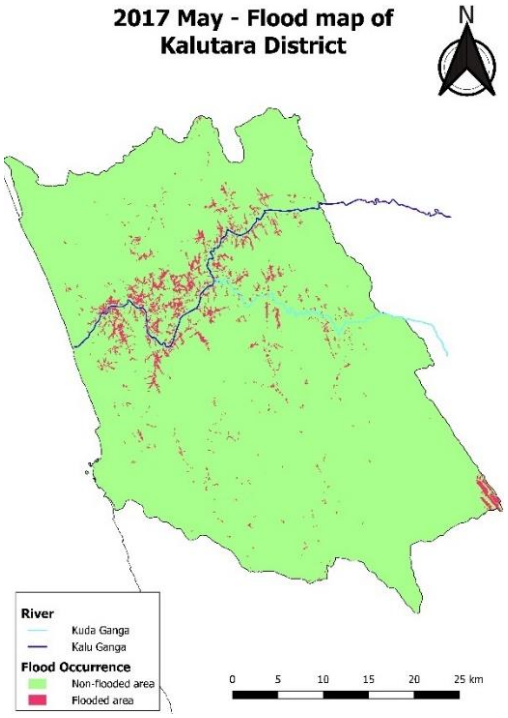

Figure 2. 2017 May flood extent map derived from SAR.

Most of the flooded areas are closer to the water features in the study area.

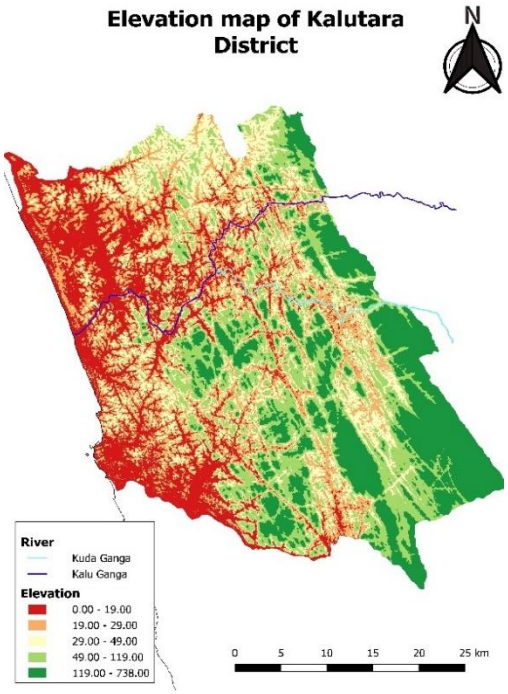

Figure 3. Elevation map of Kalutara district.

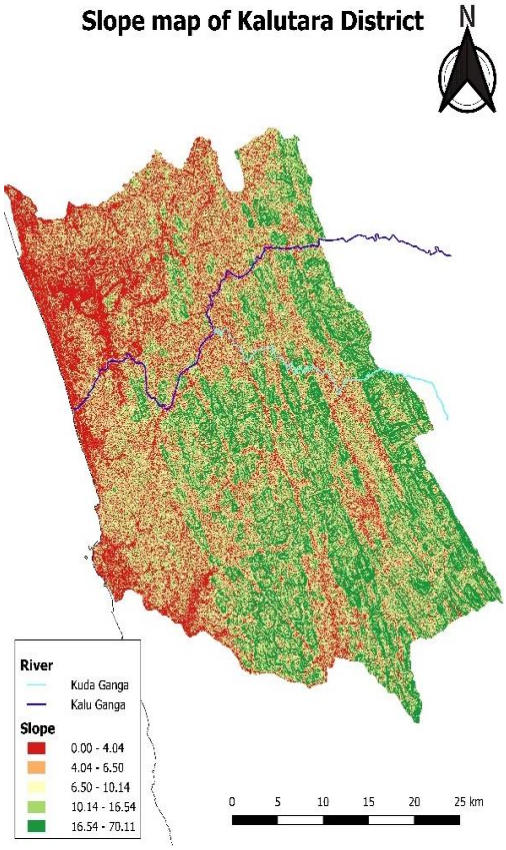

Figure 4. Slope map of Kalutara district.

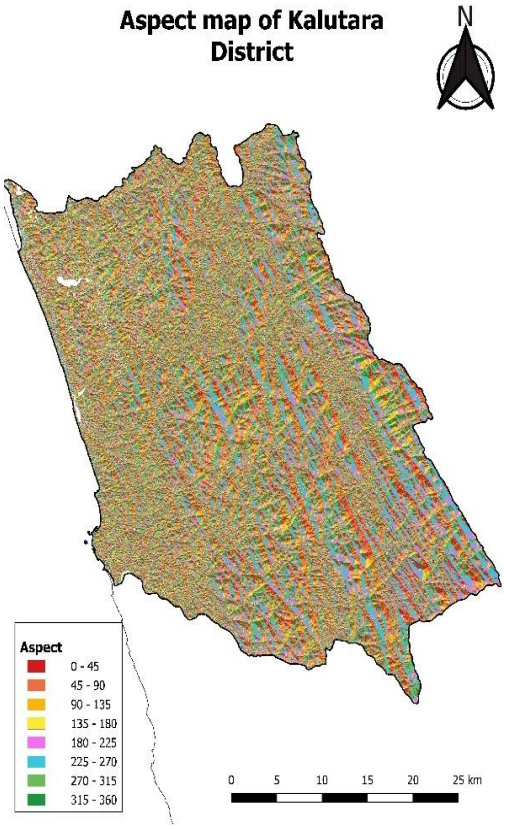

Figure 5. Aspect map of Kalutara district.

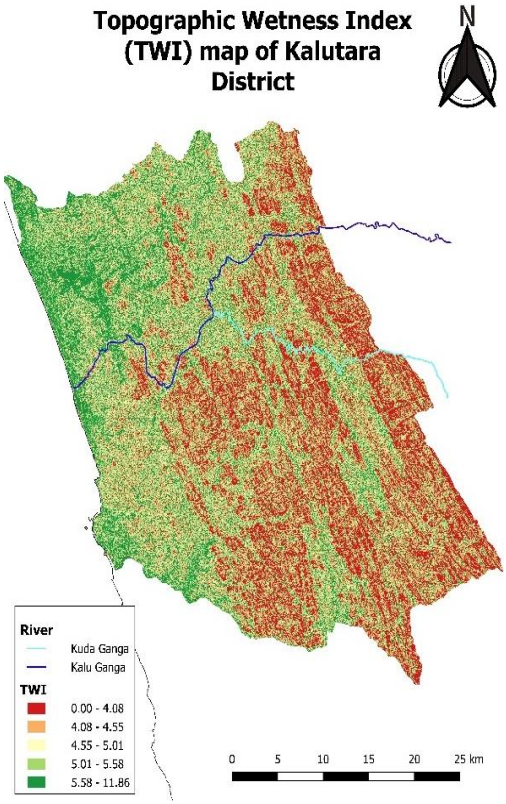

Figure 6. Topographic Wetness Index (TWI) map of Kalutara district.

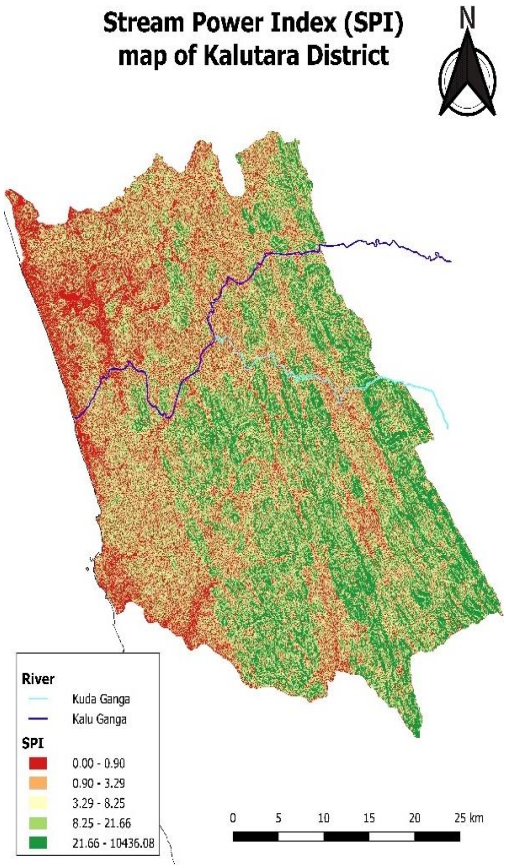

Figure 7. Stream Power Index (SPI) map of Kalutara district.

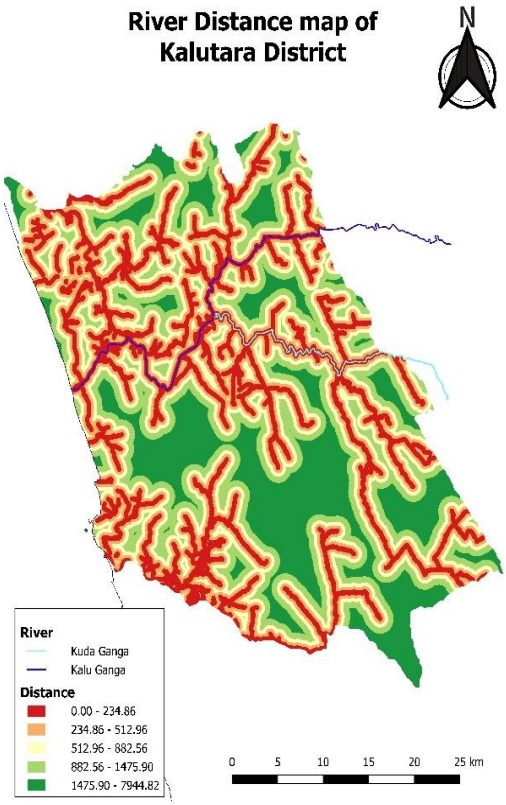

Figure 8. River Distance map of Kalutara district.

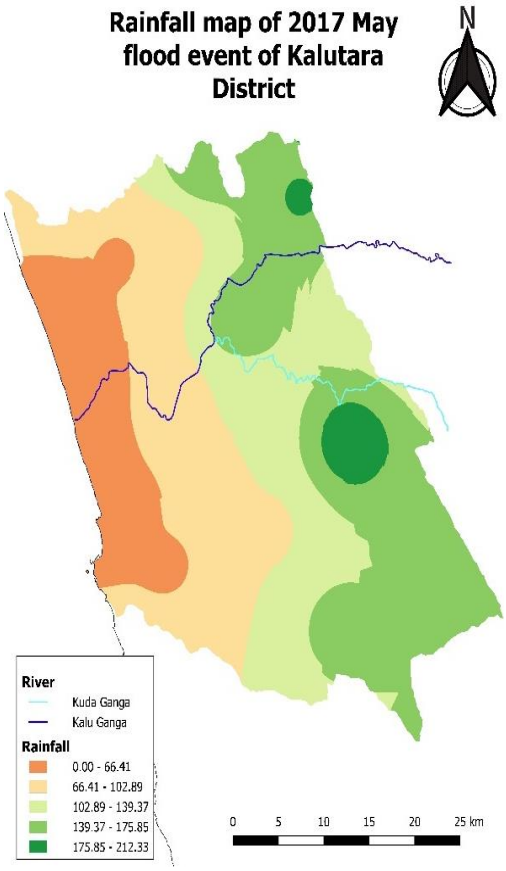

Figure 9. Rainfall map of Kalutara district.

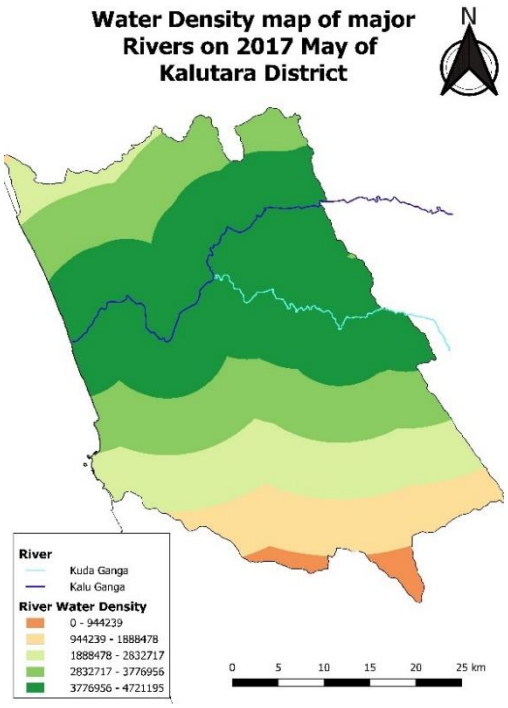

Figure 10. River water density map of Kalutara district.

Model Accuracy

Model fitting was done using different sizes of sample point 1000, 3000, 6000. Their accuracies are evaluated using confusion matrices and ROC curves shown in Figures 11–13.

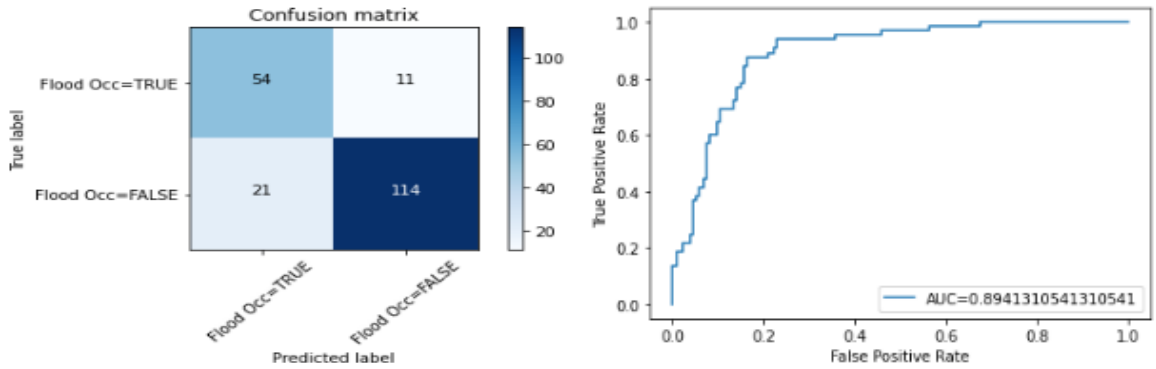

Figure 11. Confusion matrices (left) and ROC curves (right) of fitted model with 1000 sample points.

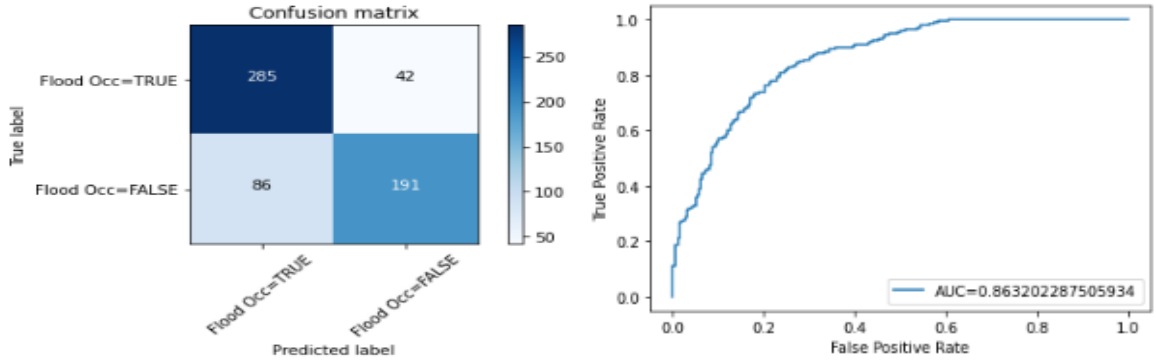

Figure 12. Confusion matrices (left) and ROC curves (right) of fitted model with 3000 sample points.

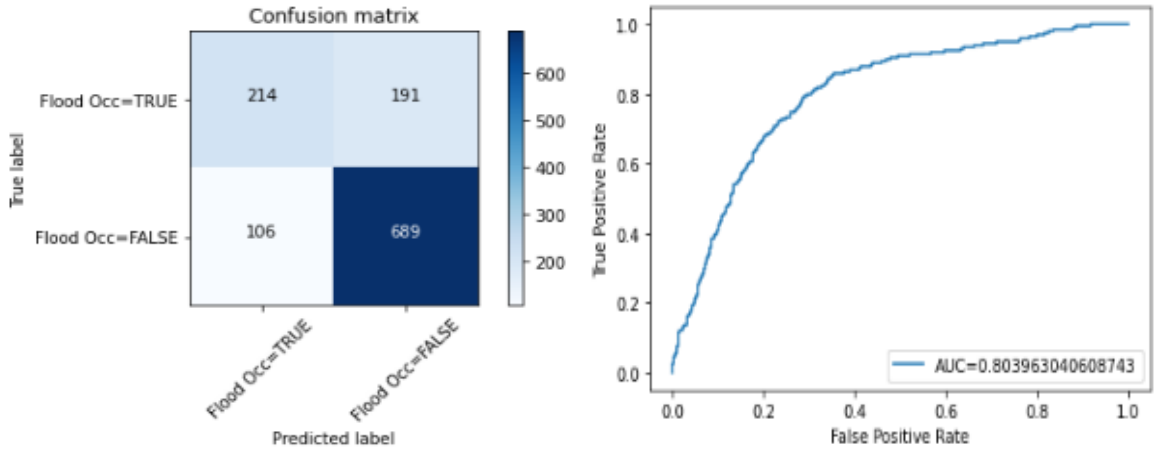

Figure 13. Confusion matrices (left) and ROC curves (right) of fitted model with 6000 sample points.

When comparing the results of the confusion matrices, the values archived from the model of 1000 sample values have better model accuracy. This result also can be identified from the ROC curves and AUC values of the models. Then for the flood susceptibility modeling, the coefficients derived from the fitted model with 1000 sample points were selected.

Model Coefficients

The river water level is having the highest values from other factors, and we can say that it is the most contributing factor for the flood susceptibility over the study area. Aspect has the lowest model coefficient, and so it is a factor that isn't contributing to flood over the study area (Table 1).

Table 1. Derived model coefficients for each factor.

Factor	Coefficient
Rainfall	0.06
Elevation	–0.40
Slope	–0.20
Aspect	0.01
SPI	–0.11
TWI	0.16
Distance to Rivers	–0.17
River Water Level	0.57

Especially, rainfall of the study area has a lower coefficient value when compared with the other factors. That indicates the rainfall of the study area is not much affecting to flood susceptibility over the study area.

Flood Susceptibility Map

As the river water level is one major contributing factor in the fitted model, the final flood susceptibility map shows much higher susceptibility around the main water features over the study area. Also, lower elevated areas are having much susceptibility than higher elevated areas. The spreading of higher susceptibility is increased in the lower elevated areas which are closer to the sea (Figure 14).

CONCLUSIONS

Floods are the most common natural disaster for the living beings of Sri Lanka. But there is still no proper way of managing the flood risk of the flood prone areas. Lack of flood extent data is one problem faced in this matter. But with the Remote Sensing techniques, this flood extent data can be derived with some limitations.

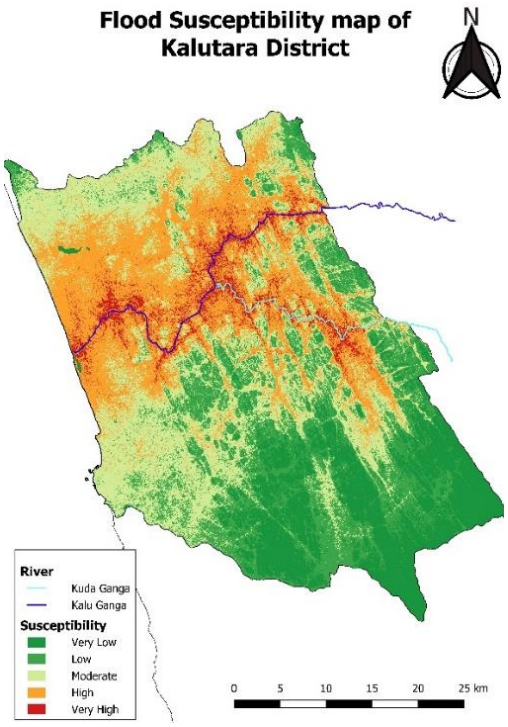

Figure 14. Flood susceptibility map of Kalutara district.

Machine Learning algorithms can provide good results in predicting different types of situations. Here the accuracy of the Logistic Regression model for the flood susceptibility modeling is in the acceptable range and its performance is very good. From the given flood contribution factors, water level in the rivers shows the highest coefficient value among others. Therefore, it can be called the most intensive factor for the flood susceptibility in the study area. Following elevation, slope, distance to rivers, TWI show considerable coefficient values when compared to rest. The results show that rainfall of the study area does not make big impact on the flood susceptibility.

When comparing the model accuracies to different sets of sample points, the number of sample points increases the model accuracy which does not seem to be going upwards. It shows good results in the average size of samples.

In the results, the flood susceptibility is much higher around the main water features in the study area.

REFERENCES

1. Thomas V, Albert JRG, Hepburn C. Contributors to the frequency of intense climate disasters in Asia-Pacific countries. Clim Change. 2014;126(3-4):381–98.
2. Costache R, Arabameri A, Elkharachy I, Ghorbanzadeh O, Pham QB. Detection of areas prone to flood risk using state-of-the-art machine learning models. Geomatics, Nat Hazards Risk. 2021;12(5):1488–507.
3. Şen Z. Flood modeling, prediction, and mitigation. Cham: Springer; 2018.
4. Mandych AF. Classification of floods. Proc 2011 Int Symp ELMAR. 2011:5–10.
5. Burt TP, Weerasinghe KDN. Rainfall distributions in Sri Lanka in time and space: An analysis based on daily rainfall data. Climate. 2014;2(3):242–63.
6. Alahacoon N, Edirisinghe M. Spatial variability of rainfall trends in Sri Lanka from 1989 to 2019 as an indication of climate change. ISPRS Int J Geo-Inf. 2021;10(84):1–15.
7. Weerasinghe KM, Gehrels H, Arambepola N, Vajja HP, Herath J, Atapattu K. Qualitative flood risk assessment for the Western Province of Sri Lanka. Procedia Eng. 2018;212:503–10.
8. Munawar HS, Hammad AWA, Waller ST. Remote sensing methods for flood prediction: A review.

Sensors. 2022;22(3):960.
9. Planinšič P, Gleich D. SAR-images categorization and applications. Proc 2017 Int Symp ELMAR. 2017:5–10.
10. Gorelick N, Hancher M, Dixon M, Ilyushchenko S, Thau D, Moore R. Google Earth Engine: Planetary-scale geospatial analysis for everyone. Remote Sens Environ. 2017;202:18–27.

