ISSN No:-2456-2165

Detection of Flood Images using Different Classifiers

Roshni R Menon
Asst. Professor
Dept of Computer Science and Engineering
SCET, Thrissur

Sandhra Simon, Maria A S, Rosemaria Shaju, Sachin
UG Students
Dept of Computer Science and Engineering
SCET, Thrissur

Abstract:- Flood is not a new disaster that we face nowadays in every part of the world. It is sudden, fast, and the impact is beyond the imagination. Its frequency is increasing day by day. Although we can't avoid this natural disaster, We should manage it properly. For that, image detection has a great role and should find the best classifier to detect it. The classifiers we use are k-nearest neighbors, Logistic Regression, Support Vector Classifier, Decision Tree, and Random Forest machine learning algorithms. By learning through each algorithm we found the best among them. The accuracy obtained by learning each algorithm on our trained model is quite different and we found out the best. First, we prepared the image dataset which includes remote dataset and satellite images. Second, we passed the dataset to each classifier and obtained the variant accuracies. Best results are produced in each method. The classifier which gives the best can be taken for the early prediction of the flood. By using new technologies to manage the flood will help us with evacuation faster and take care of people who are affected. Flood prediction has done here using history rainfall data so that we just predicted the chance. Detection is done mainly with high accuracy and the accuracy of each classifier is shown. Also the image tested result shown.

Keywords:- *Flood Detection ; Accuracy ; Training ; Convolutional Neural Network ;Logistic Regression ; K-Nearest Neighbor ; Naive Bayes Classifier ; Support Vector Machine ; Synthetic Minority Oversampling Technique.*

I. INTRODUCTION

Flood become a frequent disaster that occurs in every part of the world. In India, almost every year in new places the floods are born and which is almost uncontrollable. The government is hardly fighting for the management before, during, and after occurred. Satellite images have increased the ability to predict the flood and thereby preventing the loss of lives, property, and homes. It can help in the early evacuation of nearby places. Since the number of satellites and sensors have increased their prediction of a flood. Significant improvements in the near real-time assessments of floods have been made due to increased data acquisition rates, higher sensor resolution, the improvement of change detection algorithms, and the integration of remote sensing systems. We cannot predict these floods at their full accuracy just can make sure of the lives as possible as earliest.

We need to know about the flood is happening nearby or should take preventive measures at the earliest. For, this we are proposing a method known as Detection of Flood using different classifiers. Here we use different satellite images to detect the occurrence of flood. These images are fed to different machine learning algorithms and predicting the accuracy of each classifier.

This paper deals with the 2018 Kerala flood. The heavy rainfall and bad weather conditions have taken Kerala in its most worse situation. Every dam was overflowed in such a short time. The network had become down fast and communication was not that much easy for the rescue team as well as the people around there. In such cases, rescue teams need to know the current situation in the field. This information can be best mediated if the affected area is mapped in real-time. Our procedure assures rapid and accurate mapping using machine learning techniques. The best dataset have taken for the training the model. Tested using different classifiers and better classifier with best accuracy is taken.

A. PROBLEM DOMAIN

Flooding detection has a very important role in managing the flood in the affected area. Existing systems use fewer data and thereby getting less accuracy for the detection. The dataset including satellite images is used for detection. Moreover, the use of images taken from remote areas should also consider for the exact detection. Those images gave more accurate segments for the image segmentation. Today there are a variety of techniques that we can choose from. The important role of images and classifiers is crucial in each technology or whatever the techniques are using detection of image using better classifier is the best choice. The main drawback of the existing system is not knowing the better classifier. Some of the systems use Twitter images combining with Artificial Intelligence to detect which is more time-consuming. There is also a technique that uses hardware at each flood-prone area, detecting the factors of a flood using sensors and predicting the occurrence of flood and all. The main demerit of these things is network gets down in those areas when a flood occurs, thereby no further information about that area can be received. In that cases, detecting the flood using machine learning holds a huge role and thereby finding the best algorithm to train the model. Thus it has to be detected in a very accurate way. We use machine learning algorithms that give perfect results. The dataset taken should be accurate for classifying. Images act as a weapon for detecting the disaster at the most valuable time. Using recent floods in the last years can be used for high accuracy in the

detection. The suggested classification system for flood detection looks promising and addresses the issues raised by the previous systems. The data is collected in real-time with high-resolution cameras which eliminates the need for a poor classifier to detect the images. The suggested classifier will be better.

B.PROPOSED METHOD

In this current scenario, natural disasters are more common compared to previous environmental situations. Flooding due to heavy rainfall is the major issue that we face in our daily life. Our project is focusing on the early flood detection that can take place by analyzing different satellite images. Floodnet dataset is used to train with different machine learning classifiers. The main four classifiers that we applying are SVM, KNN, Logistic Regression, Decision Tree, etc. we also input the testing images additionally for checking the result of each algorithm used. It shows the result as this is Flooded or Not Flooded. After the execution there we also compare the accuracy obtained by each algorithm and will take the best classifier leading with a higher accuracy rate.

C. DATASET

The dataset that we are implementing for our project is mainly from the Floodnet Dataset. This data set from the Microsoft teams are basically collected by the International Conference combined with UMBC for the Satellite image data learning. This study is utilizing for research on different satellite images related to weather, natural disasters like floods, cyclones, and to find different hydrological parameter ,etc. Here, this dataset is classified as about 500-1000 flooded and non-flooded images. Besides, this information is more useful for exploring the climatic changes happening nowadays. Heavy rainfall is occurring due to the high rate of precipitation and humidity. Which leads to the occurrence of the life-threatening disaster known as Flooding. These satellite images are then trained with different machine learning algorithms.

The algorithm that is used for further training is CNN (Convolutional Neural Network). The system will try to learn the features, attributes of the particular image. For example, if the image containing buildings, short-term streams, playgrounds, etc may be used to be classified accordingly. The image is labeled as using the Longitudinal and Latitude values. The data divided into two processes, tests and the validation part. We use 60 to 70percentage of data to train the machine. Balance 30 percentage data is given into the test folder for validation if it works or not. Here we use the mobile net model which preprocesses the input on your inputs before passing them to the model. Dividing each into three batches given ten as a size. Image data generator is the function used. using this function we converting into a machine-understandable form. Thus actually machine studies this and show the output as how many images belong to each class. The library called Keras is used in the model. And find the above-mentioned features and divides them into different classes. After the learning process, it will be tested using an image whether it is flooded or not. For example, If we are giving an image of agriculture field a testing image it will give the result as Not Flooded. The confusion matrix representing the prediction label is shown in the Figure 1.

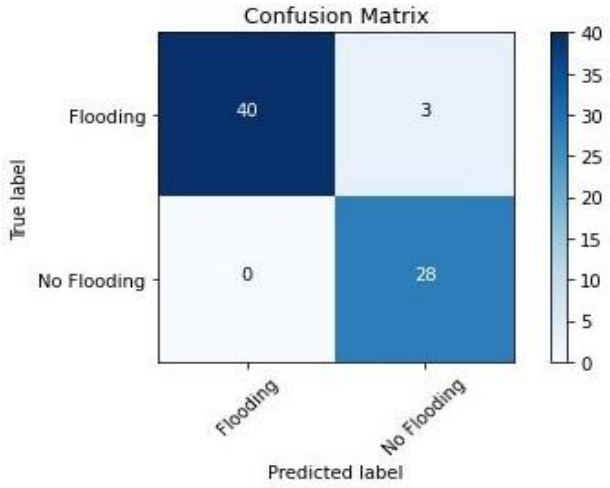


Fig 1.Confusion Matrix of Flood detection.

The models are used for the comparison between the accuracy of detecting flood. Each classifier will be trained with the given dataset and predict the output. Flooded and No-flooded images are shown in the Fig1.1 & Fig 1.2.

Fig 1.1 Flooded Images

Fig 1.2 Non-Flooded Images

Training network used here is CNN(Convolutional Neural Networks). It is also a Multilayered neural network representing different classes. CNN models can detect complex features in data. In this, we used satellite images for the classification for detecting whether it is flooded or Not flooded. CNN is also used for image classification and recognition. The weight of each data is shared between the inputs. Thus, weight sharing is an advantage in this system.

II. EVALUATION ALGORITHMS

- **KNN(K-Nearest Neighbours)**

It is a super simple way of classifying data and supervised machine learning models. It classifies the new data point into the target class, depending on the features of its neighboring data points. Features of KNN model, it is a classification algorithm and a Non-Parametric, which means it does not assume any particular kind of distribution or functional relationship of the data that we are trying to predict or learn the pattern. It is also used to transform underlying data the features into a higher dimension after that parametric approach is applied.

The training process includes different variables to store the elements like function variables to represent each section like which classifier, accuracy rate, etc. For example (knn_clf) represents the knn classifiers used, the accuracy score is calculated as such using the variable, (metrics.accuracy_score). Training elements used are(knn_clf.fit(X_train, y_train). Then testing the image is then fed to the algorithm to check whether it is flooding or not. For example, we used the (filename='evaluate/flood4.jpg') and the result was it belongs to the flooded class. The drawbacks of the KNN algorithm are outlier sensitivity and missing value treatments.

- **SVM (Support Vector Machine)**

It is a discriminative classifier that is formally designed by a separative hyperplane. It is a representation of examples as points in space that are mapped so that the points of different categories are separated by a gap as wide as possible. It is also a supervised learning algorithm that can be used for classification or regression problems. It works well even with unstructured and semi-structured data like text images and trees.it is maximum margin hyperplane, finds the linear model with maximum margin unlike the linear classifiers, the objective is not minimizing the sum of squared errors but finding a line/plane that separates two or more groups with maximum margins.

Support Vector Machine used in this project for the classification of different satellite images and detection of flooding process. The classifier used here is the (svm_clf). Train the model using the training sets using the function variable (svm_clf.fit(X_train, y_train)). The training elements used are (X_train, y_train).Predicting the dataset using the variable using (y_pred = svm_clf.predict(X_test)). The demerits of the SVM it is not suitable for large datasets.

- **Logistic Regression**

Regression analysis typically is represented as the predictive modeling technique.it estimates the relationship between a dependent (target) and an independent variable(predictor).It is also a supervised learning algorithm. Logistic Regression produces results in a binary format which is used to predict the outcome of a categorical dependent variable. So the outcome should be discrete/categorical such as:

- 0 or 1
- YES or NO
- True or False
- High or Low

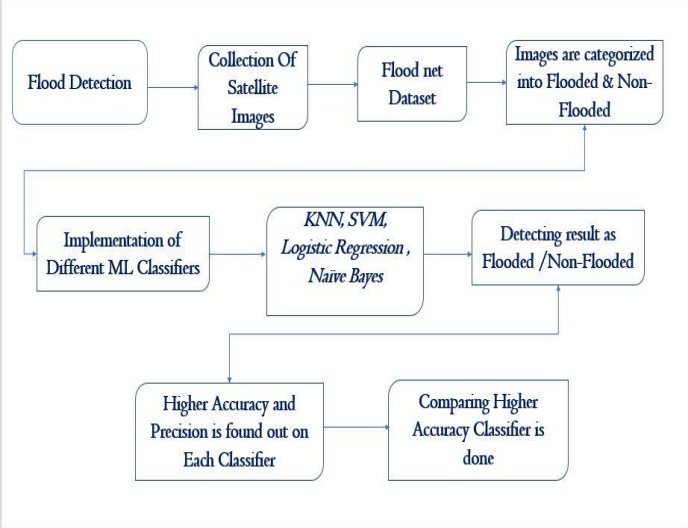
Logistic regression used in our project is used to classify different satellite images to detect flood. Variable used is (logis_clf = LogisticRegression()).The variable used for training the elements are (logis_clf.fit(X_train, y_train))the prediction variable used here is(y_pred = logis_clf.predict(X_test)). The main draw backs of logistic regression is mainly Overfitting.

- **Decision Tree**

A Decision Tree is a tree-shaped diagram used to determine a course of action. Each branch of the tree represents a possible decision, occurrence, or reaction.it is a graphical representation of all the possible solutions to decisions. These decisions are based on some conditions. It is mainly classified as the root node, leaf node, splitting, branch or subtree, parent/child node, pruning.

The implementation of the Decision Tree in our project is consists of different elements to train the particular images etc. The decision tree classifier is represented as dec_clf. Training the model using the training sets (dec_clf.fit(X_train, y_train)).Predicting the response for test dataset (y_pred = dec_clf.predict(X_test)).

III. SYSTEM ARCHITECTURE



IV. WORKING EXPLANATION

The problem domain that we are facing floods due to unconditional climatic changes and heavy rainfall. So to detect the flood we should be alert on upcoming flooding. Thus, We propose the method of flood detection using image processing with different training machine learning classifiers. Different types of steps are taking place in this detection phase of flooding.

A.PSEUDOCODE

1. Importing necessary packages
2.Collection of Satellite Images
3. Flood net Dataset preparation.

4. Image Classification labeled as Flooded & Non - Flooded.
5. Implementation of Different ML Classifiers.
6. Different training method algorithms like KNN, SVM, Logistic Regression, Decision Tree
7. Output is represented as Flooded or Non-Flooded
8. Higher Accuracy and Precision are found out on each Classifier.
9. Comparing Higher Accuracy Classifiers is done.

V. HARDWARE SPECIFICATIONS

- Processor : i5
- RAM : 4 GB
- Hard Disk : 500 GB

VI. SOFTWARE SPECIFICATIONS

- Python

Python is most simplest language that can be used for various tasks such as classification, Software Development and even many more. Python 3.7 is used here.

- IDE : SPYDER

For the programming in python language we use open source cross platform called Spyder here.Spyder is best used with jupyter notebooks or other scientific computing tool such as Anaconda which provides a general python development environment.

- Scikit-learn (Sklearn)

Package used for accuracy calculation,modelling in machine learning which helps us also to draw confusion matrix.

- Keras

The library used for python interfacing for neural networks.

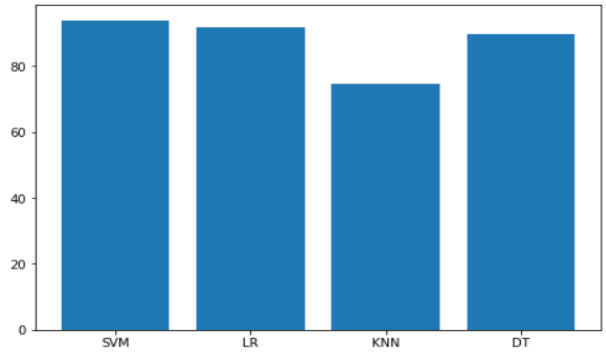
VII. PERFORMANCE EVALUATION

Classifier Algorithms gave different accuracy levels using the same dataset are shown in the table 1.1.

Logistic Regression	91.79978700745474
Support Vector Machine	93.9297124600639.
K-Nearest Neighbour	74.54739084132055
Decision Tree	87.43343982960596.

Table 1.1

The training success with each machine learning method is presented with the percentage of correctly classified training points.



VIII. CONCLUSION

Flood Detection using different algorithms was successfully classified using well balanced dataset. For balancing dataset we used SMOT analysis in which used for accurate classification. For better classification the use of a dataset and the most accurate classifier is essential. Mobilenet data model is used for converting the data into batches with 10 as the size for each. To compare accuracies we used the four best among them and from that, SVM is the best to detect the flood dataset. This is the optimal classification algorithm in which we tested for binary classification labeled as flooded or no-flooded.

Among the classifiers we used Logistic Regression, Support Vector Machine crossed the real accuracy level that exists. After all, more input data and machine learning algorithms are needed to be tested if we wish to find the best combination for detecting flooded areas. The accuracy has to be increased shortly for better real-time flood management.

ACKNOWLEDGMENT

We offer our heartfelt gratitude to everyone ho helped us to make this effort a grant success. We thank the almighty God for all the benefits he has bestowed upon us. Our guide, Prof. Roshni R Menon deserves a lot. Our principal, DR. Nixon Kuruvila, and our Department Head Dr. M Rajeswari, for unwavering support throughout the project.

REFERENCES

[1]. Lorenzo Alfieri, Sagy Cohen, John Galantowicz, Guy JP Schumann, Mark A Trigg, Ervin Zsoter, Christel Prud-homme, Andrew Kruczkiewicz, Erin Coughlan de Perez, Zachary Flamig, et al, (2018) , "A global network for operational flood risk reduction" . Environmental science & policy, 84:149–158.
[2]. Serena Ceola, Francesco Laio, and Alberto Montanari, (2014) "Satellite nighttime lights reveal increasing human exposure to floods worldwide" . Geophysical Research Letters, 41(20):7184–7190,
[3]. Ben DeVries, Chengquan Huang, Megan Lang, John Jones, Wenli Huang, Irena Creed, and Mark Carroll, (2017) "Automated Quantification of Surface Water

Inundation in Wetlands Using Optical Satellite Imagery“, Remote Sensing, 9(8):807.
[4]. Laura Giustarini, Renaud Hostache, Patrick Matgen, Guy J.P. Schumann, Paul D. Bates, and David C. Mason, (2013) "A Change Detection Approach to Flood Mapping in Urban Areas Using TerraSAR-X". IEEE Transactions on Geoscience and Remote Sensing, 51(4):2417–2430.
[5]. Fabio Cian, Mattia Marconcini, and Pietro Ceccato, (2018) "Normalized Difference Flood Index for rapid flood mapping: Taking advantage of EO big data". Remote Sensing of Environment, 209:712–730.
[6]. Hanqiu Xu. (2006) "Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery" . International Journal of Remote Sensing, 27(14):3025–3033.
[7]. Jian Li and David Roy. (2017) "A Global Analysis of Sentinel 2A, Sentinel-2B and Landsat-8 Data Revisit Intervals and Implications for Terrestrial Monitoring". Remote Sensing, 9(9):902.
[8]. Igor Klein, Andreas Dietz, Ursula Gessner, Stefan Dech, and Claudia Kuenzer, (2015). "Results of the Global WaterPack: a novel product to assess inland water body dynamics on a daily basis". Remote Sensing Letters, 6(1):78–87.
[9]. Sandro Martinis, Andre Twele, and Stefan Voigt. (2009) " To wards operational near real-time flood detection using a split based automatic thresholding procedure on high resolution TerraSAR-X data" Natural Hazards and Earth System Sciences, 9(2):303–314 ,
[10]. Perry C Oddo and John D Bolten. (2019) "The value of near real time earth observations for improved flood disaster response". Frontiers in Environmental Science, 7:127,
[11]. Jean-Franc ̧ois Pekel, Andrew Cottam, Noel Gorelick, and Alan S Belward. (2016) " High-resolution mapping of global surface water and its long-term changes". Nature 2.
[12]. Stefan Schlaffer, Patrick Matgen, Markus Hollaus, and Wolf gang Wagner. (2015) "Flood detection from multi-temporal SAR data using harmonic analysis and change detection". International journal of Applied Earth Observation and Geo information, 38:15–24.
[13]. Yuxin Wu and Kaiming He. (2018) "Group normalization, In Proceedings of the European Conference on Computer Vision" (ECCV), pages 3–19, 5.
[14]. A. Al-Fuqaha, M. Guizani, M. Mohammadi, M. Aledhari, and M. Ayyash, (2015) ‘‘Internet of Things: A survey on enabling technologies, protocols, and applications,’’ IEEE Commun. Surveys Tuts., vol. 17, no. 4, pp. 2347–2376, 4th Quart.
[15]. A. Y. Sun and B. R. Scanlon, (2019). ‘‘How can big data and machine learning benefit environment and water management: A survey of methods, applications, and future directions,’’ Environ. Res. Lett., vol. 14, no. 7, p. 28.
[16]. K. Panchal, H. Ajmani, and S. Y. Sait,(2019) ‘‘Flooding level classification by gait analysis of smartphone sensor data,’’ IEEE Access, vol. 7, pp. 181678–181687.
[17]. K. Tuberson and K. Tremaine, (2017) ‘‘How the Internet of Things can prepare cities for natural disasters,’’ Harvard Bussniss Review.
[18]. D. B. Soubhagyalaxmi, L. S. Bhosle, T. Swetha, P. D. Veena, and B. V. Chetan, (2018) ‘‘Disaster management system using IoT,’’ Int. J. Res. Trends Innov., vol. 3, no. 6, pp. 148–154.
[19]. X. Chen, X. Zhu, and D. Zhang, (2010) ‘‘A discriminant bispectrum feature for surface electromyogram signal classification,’’ Med. Eng. Phys., vol. 32, no. 2, pp. 126–135.
[20]. T. K. Ho, (2002)‘‘A data complexity analysis of comparative advantages of decision forest constructors,’’ Pattern Anal. Appl., vol. 5, no. 2, pp. 102–112.