Retraction

Retraction: Advanced flood severity detection using ensemble learning models (J. Phys.: Conf. Ser. 1916 012048)

Published 23 February 2022

This article (and all articles in the proceedings volume relating to the same conference) has been retracted by IOP Publishing following an extensive investigation in line with the COPE guidelines. This investigation has uncovered evidence of systematic manipulation of the publication process and considerable citation manipulation.

IOP Publishing respectfully requests that readers consider all work within this volume potentially unreliable, as the volume has not been through a credible peer review process.

IOP Publishing regrets that our usual quality checks did not identify these issues before publication, and have since put additional measures in place to try to prevent these issues from reoccurring. IOP Publishing wishes to credit anonymous whistleblowers and the Problematic Paper Screener [1] for bringing some of the above issues to our attention, prompting us to investigate further.

[1] Cabanac G, Labbé C and Magazinov A 2021 arXiv:2107.06751v1

Retraction published: 23 February 2022
Advanced flood severity detection using ensemble learning models

A. Sunitha Nandhini¹, J Balakrishna¹, R Bala Manikandan¹, S Bharath Kumar⁴
¹Department of Computer Science and Engineering, Sri Krishna College of Technology, Coimbatore, Tamil Nadu, India
Email: jonnalakrish@gmail.com, ¹tucs018@skct.edu.in, ¹sunithanandhini.a@skct.edu.in, ¹tucs016@skct.edu.in

Abstract. On the life of human beings and financial development of the nation, the phenomenon of River flooding has its catastrophic effects. There are various approaches in finding out watercourse flooding but depleted understanding and restricted information regarding flooding conditions hinder the management estimates this particular phenomenon. The ensemble model approach has been used in this paper, i.e., the combination of Multilayer layer perceptron model (MLP) + K-Means Clustering (KMC) for flood severity prediction. Our ensemble way can support the modern and recent growth and development inside the IoT (IoT stands for Internet of things), with the help of some tools such as smart sensors, RFID and learning based on machine for the prediction of flood severity and its automatic analysis and it is expected to help human beings and can be a useful rescue from such kind of natural disasters. Analysis outcome indicate that ensemble model is more reliable to predict flood severity. The experimental output shows that by the usage of ensemble learning along with Multilayer Perceptron (MLP) model, Particle Swarm Optimization (PSO), K-Means Clustering, Long-Short Term Memory and Random Forest Classifier will produce an optimized result and also with greater accuracy.

1. Introduction

River flooding is a natural disaster which causes destruction to human life, goods, properties etc. Providing an early and quick warning for such a deadly disaster will be very helpful for saving lives and properties [1]. With the help of current technology trends (ML, IOT, AI) flooding of river cannot be avoided completely but an alert can be given, so everyone can take care of their life and property. A recent research study by Gilad David Maayan and Gartner shows that for every second nearly 127 IoT devices are connected through web, and there were nearly 31 billion IoT devices which were connected through web in 2020. It has been estimated by the experts that by the end of year 2021, on the web there would be as much as 35 billion IoT connected devices, by 2025 75 billion IoT devices will be connected over web. As a part of this different varieties of sensors will be used for collection of data with high accuracy. [2] IoT is rapidly growing day by day which can provide access to physical devices remotely. By using Machine Learning (ML) useful data can be obtained from the huge volumes of data’s and predicting can be done more accurately. ML techniques provide reliable models for extraction of sensor data and also provide optimized solution [3]. Majority of Environmental centers are using IoT assisted with ML for flood prediction purposes since prediction can be done more accurately.
The most important thing in flood prediction is gathering most important, reliable and meaningful data. By using machine learning one can obtain such data. Researchers say that the learning based on machines has tremendously enhanced the flood detection and for this purpose the algorithms for deep learning are used [4]. The models such as supporting vector machine and random forest are the models of machine learning and they will be very useful for data analysis and clustering model like K-Means clustering will be very useful clustering of data, which are the most important stages while doing machine learning project [5].

This research includes the usage of multi-sensor data and also includes data collected from smart sensors which is collected from various flood centers across the world to determine the river water levels. Advanced models like (Particle Swarm Optimization (PSO), Multilayer Perceptron (MLP), ANN which is an abbreviation of Artificial Neural Network, SVSM (which stands for Support Vector Machine), LSTM which is an abbreviation of long short-term memory etc. are used for prediction purposes. The major use machine learning and deep learning algorithms is to analyze and predict flood sensor data [6-10].

Collection of raw data is a simple process, but the quality of data is biggest factor for prediction purpose. Collected data get affected by, missing values in it, noise in data and so on. In order to deal with this data science approach has been used. Different machine learning approaches have been used for prediction purpose. Ensemble approach has been used, since single machine learning model has not provided the desired outcome. Visualization of the outcome has been also done for clear understanding purposes [11-16].

The major contribution of the project/paper as follows
• For the prediction of severity of flood with the help of collected data at various sensors ensemble of machine learning algorithms.
• Multi-level classification improvement and classifier’s accuracy with the help of MLP, LSTM and SVM which are deep learning algorithms.

2. Literature Review

Studies using ensemble learning models and usage of deep learning algorithms for flood severity prediction are in still developing stage. There are many approaches for the prediction of flood occurrences, but there are only very few approaches for prediction of flood severity and the major disadvantage in these approaches is lack of accuracy and performance. Above discussed approaches use only one machine learning model which is major disadvantage. In order to overcome this our project uses ensemble learning (i.e. Multiple machine learning and deep learning models). Our approach also uses current trends in IoT sensors which include smart sensors and RFID. The usage of smart sensors reduces human involvement with sensors and data can be obtained with high quality. With stress on sanctionative technologies, protocols, and application problems the IoT is enabled by the most recent developments in RFID, good sensors, communication technologies, and web protocols.

The basic premise is to own good sensors collaborate directly while not human involvement to deliver a replacement category of applications. It has been forecasted in the last few years to bridge modernized innovations so that new applications can be altered by the connection of physical items in addition to supporting the smart devices. The initial technical detailed summary can also be offered as we have this tendency so that sanctionative technologies, protocols and applications can be pertained. Internet of things is a need of this era and in this modern time, sensing devices and their colossal quantity is likely to generate sensory data with the passage of time for the applications in various fields. Over such kind of information when analytics is applied for the generation of modern and latest data, future insight
prediction and creating choices for management, all these are essentially crucial methods to create IoT which can ideally be used in the businesses paradigm and technologies for enhancing the quality of life.

On the employment of category of modern techniques regarding machine learning process we offered a radical summary especially DL which stands for deep learning and this is done to facilitate the analytics and internet of things related domain. Hardware development, protocols and various algorithms and collective knowledge is used to sense various networking elements, information of this kind is essential in making of the smart systems.

Vertical system integration is explained in detail for the nodes of sensing elements along with a machine learning algorithm toolkit. The sensible knowledge is mixed by supporting a knowledge set and data is further introduced and in this way we can forecast the people and their number in any zone. Dataset analysis and performance can be assessed by machine learning algorithms of two kinds on this data set which are described generally as classification and regression. Sensing element network development especially in the past few years as ended up in extending of some more relevant domains which are so many in numbers such as preservation of heritage, driving the environment and recognition act. Most of the systems which work on the bases of sensor are used in the applications. Highest layers are conjointly included in it which are supposed to deal with the data in associate degree economical and helpful manner.

3. Proposed System

With the help of generation of early warning this assessment can be beneficial in predicting the situation of floods and thereby the situation can be handled, and instant help can be offered. The region of flood can be provided with rescue operation as well. In the work suggested, the use of deep learning models can be used as well for the accurate predictions regarding flood severity and it is captured in numerical data in the regions of flood by the habitats. The flooding data is used by the proposed method for the determination of flood severity level and event is recorded as an input. A flood dataset has also been created by us on the basis of unavailability of dataset of this special kind. On this dataset the proposed model is evaluated and then it is contrasted with the baseline adasvm classification to predict better flood severity results. This work includes the comparison of machine learning (ML) algorithms and proposes an ensemble approach for the severity of flood prediction. The dataset is divided into three categories namely normal abnormal and highly risky for flooding.

The system suggested comprises of some special components such as collection of data, the pre data processing, data representation for the feature space and training of classification model and configuration of data and analysis using the testing data. The dataset has been splitted into 70:30 ratio for training and testing purpose, so the dataset will undergo many training stages and then it is tested so the accuracy can be improved which is the major factor. The proposed system consists of data collection stage where the flood sensory data is collected and pushed into pre-processing stage where the null values are removed, noise in data is removed and the data is normalized. The normalized data is pushed into training stage and then machine learning and deep learning algorithms are applied on the data for prediction. The overview of the proposed method is represented in Figure 1. The classifiers and algorithms discussed in the paper is represented in table 1.
3.1 Data Collection

The data collected from sensors is always stored in environmental agencies the data from environmental agencies which contains collection of different datasets of various cities all around the world is collected. Flood sensor data used in this research consists of 4117 flood samples. This collected data is categorized into three major parts named as normal (with 1181 data points), abnormal (with 306 data points) and highly risky (with 456 data points). Aggregate function is used for statistical procedures for combining of data from several measurements.

3.2 Data Pre-Processing

To get accurate and more optimized outcome with the help of machine learning models the collected data must be pre-processed. Pre-processing of data involves cleaning of data, removal of null values from dataset, removal noise from data, and normalizing the data. The most important stage in data pre-processing while dealing with machine learning algorithm is data normalization and noise reduction. By making the data to undergo all these processes more accuracy can be obtained, and overall performance of the outcome can be improved. Finally, data pre-processing step is the most essential step in machine learning.

3.3 Exploratory Analysis

It is used for identifying outliers in the pre-processed data. It is most important step to improve the efficiency of training models. To illustrate this visualization is shown in Figure 2.
3.4 Feature Selection

It is the most important and essential step for prediction and analysis of data. The older approaches use only one machine learning model for prediction, so the outcome was not much accurate. In this approach multiple machine learning models are used, called as ensemble models. This approach also uses deep learning models for analysis and prediction purpose. Machine Learning (ML) models like (Random Forest Classifier, Artificial Neural Network, Multilayer Perceptron, Support Vector Machine, Long-Short Term Memory) are used. Given below is the brief description of these models.

3.5 K- Means Clustering

The simpler and one of the most popular algorithms for machine learning are known as K-means clustering. For the portioning of many observations it is a vector quantization method. With the nearest mean each partitioned observation belongs to the cluster. It reduces the within cluster variance but does not reduce Euclidian distance. The purpose of using the K-Means clustering in this paper is to optimize squared errors.

3.6 Artificial Neural Network

Derived from the connectionist model the neural network is an approach of solving problems. ANN is inspired by biological neural network founded in mammal brain. The input given is feed into network layer which is then passed into some hidden layers and the input will be processed and passed to output layer. Backpropagation algorithm is used during training stage to updates weights and errors. ANN is being used in many applications like text classification, paraphrase detection etc.

3.7 Multilayer Perceptron

MLP which stands for multilater perceptron is actually the distribution of ANN feed forward; ANN stands for the artificial neural network. Supervised learning technique is used by it and it is also called backpropagation of training. Non linear activation and multiple layers are there inside it and they are supposed to identify it from the linear perceptron. The non-linear advisable memory and knowledge can be distinguished with the help of it.

| Architecture | Training Algorithm |
|--------------|--------------------|
| Inputs: 11   | ANN (Multilayer Perceptron) Damped Least Squares |
| Hidden Layers: 2 | |
| Inputs: 11 | K - Means Clustering Clustering is done by measuring distance |
| Classes: 3   | |
| Layers: 2    | LSTM (Long Short-Term Memory) Error Backpropagation algorithm |

Table 1. Algorithms and classifiers used.
4. Experimental Results

To verify the performance of the proposed method monthly rainfall dataset for 115 years (1901 – 2015) is used. This dataset consists of 10 attributes and 4117 records. The dataset is split into 70:30 ratio for training and testing. Holdout methods were used dividing the dataset to training (70%) and testing (30%). The dataset is tested in batches and remaining random data have been used for testing. The top performing model in terms of sensitivity and accuracy is Multilayer Perceptron model which produced an accuracy of 79.119% shown in figure 2. The worst performing model is AdaNavie which produced an accuracy of 56.719%. In the proposed experiment the models AdaNaive, J48, Adasvm, IMLP produced ACC values of 56.25, 54.1667,52.0833, 79.1667. However, the ensemble models produced improved results. The improved results are obtained by the combination of MLP and K-Means clustering. The performance graph of the above-mentioned models is shown in Figure 3. ANN and RF ended up in accuracy of 0.956. Along with this, LSTM and RF which are the ensemble classifiers produce 0.714 of the average sensitivity.

When SVM is not combined with any other model the performance was not so good during training and testing stages, but when combined with LSTM it produced a better outcome.

The correctness can be calculated by the formula given below
\[
\text{Accuracy} = \frac{TP}{(TP+TN)} \times 100\%
\]

While comparing with individual ML techniques ensemble models produced greater performance and accuracy and also worked faster than some entity samples. As compared to individual models the ensemble classifier performed better and the main reason is the variable distribution of data. Some strong non-linearity is there in the flood sensor data in the variable so in order to build an accurate and performable model some parameters have to be altered. In order to handle large instances of data neural network and random forest are combined together. Finally, the ensemble models provided an optimal and robust solution when compared with linear models. Table 2 displays the performance evaluation table for the proposed method. In order to show that the algorithm used in proposed method has greater accuracy a comparison graph have been displayed in Figure 4. The ROC graph for the testing data is shown in Figure 5 which proves that the proposed model is more accurate and optimized.

Table 2. Performance Evaluation Table

| Model       | Class (C1, C2, C3) | Sensitivity | Accuracy |
|-------------|--------------------|-------------|----------|
| ANN         | C1                 | 0.535       | 0.558    |
|             | C2                 | 0.529       | 0.533    |
|             | C3                 | 0.691       | 0.568    |
| RF          | C1                 | 0.795       | 0.693    |
|             | C2                 | 0.682       | 0.671    |
|             | C3                 | 0.564       | 0.51     |
| ANN + KMC(MLP) | C1        | 0.705       | 0.919    |
|             | C2                 | 0.706       | 0.814    |
|             | C3                 | 0.582       | 0.889    |
Figure 3. Execution time comparison

Figure 4. Accuracy value comparison

Figure 5. ROC for testing data
5. Conclusion and future work

The data is collected with the help of sensors mounted on the rivers is used as input for machine learning algorithms. The proposed method uses ensemble machine learning models which produces promising results for the flood severity prediction and also has better accuracy. By this proposed method early warning regarding flood severity can be obtained thus human lives and properties can be prevented. As a future work genetic algorithm can be used for optimization and much deep learning models can also be used for better performance and accuracy.

References

[1] Al-Fuqaha, M. Guizani, M. Mohammadi, M. Aledhari and M. Ayyash, Internet of Things: A survey on enabling technologies protocols and applications, IEEE Commun. Surveys Tuts., vol. 17, pp. 2347-2376, 4th Quart. 2015.
[2] M. Mohammadi, A. Al-Fuqaha, S. Sorour and M. Guizani, Deep learning for IoT big data and streaming analytics: A survey, IEEE Commun. Surveys Tuts., vol. 20, no. 4, pp. 2923-2960, 4th Quart. 2018.
[3] Moraru, M. Pesko, M. Porcius, C. Fortuna and D. Mladenic, Using machine learning on sensor data, J. Comput. Inf. Technol., vol. 18, no. 4, pp. 341-347, 2010.
[4] E. Borgia, The Internet of Things vision: Key features applications and open issues, computer. Common., vol. 54, pp. 1-31. Dec. 2014.
[5] Y. Sun and B. R. Scanlon, 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, pp. 28, 2019.
[6] G. Furquim, G. Filho, R. Jalali, G. Pessin, R. Pazzi and J. Ueyama, How to improve fault tolerance in disaster predictions. A case study about flash floods using IoT ML and real data, Sensors, vol. 18, no. 3, pp. 907, 2018.
[7] K. Tuberson and K. Tremaine, How the Internet of Things can prepare cities for natural disasters, 2017.
[8] S. Zafar, S. Jangsher, O. Bouachir, M. Aloqaily and J. B. Othman, QoS enhancement with deep learning-based-interference prediction in mobile IoT, Comput. Commun., vol. 148, pp. 86-97, Dec. 2019.
[9] T. Khan, S. Ghosh, M. Iqbal, G. Ubakanma and T. Dagiukslas, RESCUE: A resilient cloud based IoT system for emergency and disaster recovery, Proc. IEEE 20th Int. Conf. High Perform. Computer. Commun. IEEE 16th Int. Conf. Smart City IEEE 4th Int. Conf. Data Sci. Syst. (HPCC/Smart City/DSS), pp. 1043-1047, Jun. 2018.
[10] Sinha, P. Kumar, N. P. Rana, R. Islam and Y. K. Dwivedi, Impact of Internet of Things (IoT) in disaster management: A task-technology fit perspective, Ann. Oper. Res., vol. 283, no. 1, pp. 759-794, Dec. 2019.
[11] M. Suganya and H. Anandakumar, Handover based spectrum allocation in cognitive radio networks, 2013 International Conference on Green Computing, Communication and Conservation of Energy (ICGCC), Dec. 2013. doi:10.1109/icgcce.2013.6823431. doi:10.4018/978-1-5225-5246-8.ch012
[12] Haldorai and A. Ramu, An Intelligent-Based Wavelet Classifier for Accurate Prediction of Breast Cancer, Intelligent Multidimensional Data and Image Processing, pp. 306-319.
[13] K. R. Kavitha, A. V. Ram, S. Anandu, S. Karthik, S. Kailas and N. M. Arjun, PCA-based gene selection for cancer classification, Proc. IEEE Int. Conf. computer. Intel. computer. Res. (ICCIC), pp. 1-4, Dec. 2018.
[14] V. Lempitsky, M. Verhoek, J. A. Noble and A. Blake, Random forest classification for automatic delineation of myocardium in real-time 3D echocardiography, Proc. Int. Conf. Funct. Imag. Modeling Heart, pp. 447-456, 2009.
[15] R. A. Sheikh and A. Bhardwaj, A review on integrating intrusion detection model using classifiers and chi-square feature selection, Int. J. Advance Res. Ideas Innov. Technol., vol. 5, no. 5, pp. 230-233, 2019.

[16] M. Bilal, A. Gani, M. Marjani and N. Malik, A study on detection and monitoring of water quality and flow, Proc. 12th Int. Conf. Math. Actuarial Sci. Comput. Sci. Statist. (MACS), pp. 1-6, Nov. 2018.