University of Ibadan
Journal of Science and Logics in ICT Research

Performance Analysis of Fuzzified Machine Learning Algorithm for Flood Risk Assessment

[1]✉Oladapo, K. A., [2]Akinsola, J. E.T. and [3]Akinbiola, S. A.

[1]Department of Computer Science, McPherson University, Ogun State, Nigeria.
[2]Department of Computer Science, First Technical University, Oyo State, Nigeria.
[3]Institute of Sustainable Development, First Technical University, Oyo State, Nigeria
[1]oladapoka@mcu.edu.ng, [2]akinsolajet@gmail.com, [1]sehindeakinbiola@gmail.com

Abstract

Pluvial flooding is a type of flood that occurs when high-force precipitation surpasses the limit of drainage framework which has become a threat to human life and the global economy, thus this study proposes a fuzzified Machine Learning (ML) applications that can be used to reduce this risk. However, less attention has been paid to the use of a fuzzy rule-based classification to appraise the performance of ML applications, based on pluvial flood Conditioning Variables (CVs) for training a classifier. This research proposes a fuzzified classifier models and a performance analysis of the five ML algorithms namely K-Nearest Neighbours (KNN), Random Forest (RF), Classification and Regression Trees (CART), Naïve Bayes (NB) and Artificial Neural Network (ANN) algorithms to detect and predict pluvial flood risk. The performance analysis was evaluated using the 10-fold cross-validation and hold-out techniques, based on accuracy, sensitivity, specificity, precision and Area Under Receiver Operating Characteristics (AUROC) metrics. The performance evaluation results for each algorithm, using hold-out techniques in respect of accuracy, sensitivity, specificity, precision, and AUROC for KNN were 95.3%, 95.3%, 92.7%, 93.8% and 92.2% respectively; for RF, 72.8%, 73.0%, 73.2%, 73.0% and 83.6% respectively; for NB, 71.0%, 77.0%, 73.7%, 84.7% and 72.7% respectively; for CART, 98.4%, 98.4%, 98.3%, 98.4% and 98.6% respectively; and for ANN, 83.6%, 84.0%, 96.9%, 74.0% and 87.9% respectively. In addition, results obtained for using 10-fold cross-validation method for KNN were 96.4%, 96.4%, 94.1%, 96.6% and 93.7% respectively; for RF, 95.2%, 95.2%, 93.7%, 94.3% and 94.6% respectively; for NB, 77.3%, 77.3%, 74.7%, 84.3% and 89.5% respectively; for CART, 95.5%, 99.5%, 99.4%, 99.5% and 97.6% respectively; and for ANN, 89.5%, 89.5%, 89.7%, 89.1% and 89.9% respectively. Thus, this study shows that the fuzzified ML application can be used in detecting and predicting pluvial floods. Consequently, CART which had the best results, when compared to the rest of the classifier models, is recommended for use by experts.

Keywords: *Risk assessment, Pluvial flood, Fuzzy logic, Machine learning algorithms, Performance analysis*

1. INTRODUCTION

The idea of digitizing everything is now a reality, where artificial intelligence, internet of things, machine learning and other advanced technologies can capture and analyse vast amount of data. This has positively affected various industries and increasingly transformed how business is done globally. Machine learning, which is one of the fundamental constituents of artificial intelligence, portrays the capacity of computers to basically instruct themselves by predicting and deciding on their own based on the data they have at any given time [1].

An illustration of a routine task in machine learning is the accurate prediction of events. Customarily, it takes human insights to perform this basic task but machine learning has the capability to imitate the same logical construction. In effect, this has reduced the threat to human life and the future of human society which has been confronted by the increasing sum of occupants, unmanageable suburbanization and spatial progress [2], infrastructural aging [3] and in addition the change in climate which has given rise to

Oladapo K. A., Akinsola J. E.T. and Akinbiola S. A. (2024). Performance Analysis of Fuzzified Machine Learning Algorithm for Flood Risk Assessment,*University of Ibadan Journal of Science and Logics in ICT Research (UIJSLICTR),* Vol. 12 No. 1, pp. 89 – 97

different levels of risk from water-related natural disaster as a result of the above mentioned challenges.

Handling flood risk with the intention of safety and comfort of the citizens as well as saving their environments is one of the major responsibilities of each country's leadership especially in flood prone areas [4]. At large, flood management has advanced from flood control tactics to flood risk administration. Governments, are therefore under obligation to advance consistent and precise maps of flood susceptible area, advance strategy for maintainable flood risk management with focus on preparation, prevention and protection. Although this is not an effort to eradicate flood risk, its goal is to alleviate it. Flood is foreseen to transpire more sternly in urban areas across the globe [5] [6].

In recent years, methods of mitigating and preventing flood disasters have moved from defending approach to management approach. This is based on the comprehensive risk assessment findings and cost with benefit analysis. Machine learning can improve the risk management [7]. Number of machine learning classification algorithm has been purpose to classify, detect and predict pluvial flood risk, but the comparative study of the performance of a fuzzified algorithms has not been studied effectively. It has not been studied which among the available fuzzified classifier model can provide the best prediction for pluvial flood risks.

Therefore, this study proposes a classifier model based on fuzzy rule-based classification and implementing these models for predicting pluvial flood risk on the available conditioning variables and then makes comparative analysis of the performances of the classifier model in predicting with better accuracy. The ML algorithms used in this study for designing classifier models are K-Nearest Neighbours (KNN), Random Forest (RF), Classification and Regression Trees (CART), Naïve Bayes (NB) and Artificial Neural Network (ANN).

The rest of the paper is organised as follows: The related work section studies previous literatures for existing methods proposed for flood risk. Then, the methodology section describes the data preparation, performance metrics and technique applied on the conditioning variables as well as presents the performance analysis of algorithms. The Section 5 provides the discussion on results of the study and finally Section 6 concludes the study.

2. Related Works

Chang *et. al* [8] proposed region flood susceptibility model using fuzzy logic, multi-criteria positioning and weight linear combination approach to produce vulnerability model. The use of advanced optimization methods has improved the susceptibility map and also, improved the interdependencies of flood generation variables. The assumption of the natural features being constant may affect the accuracy of the model

Nasiri *et. al.* [9] developed a real time multistep-ahead forecast model. The model adopts one static–back propagation neural network and two dynamic – Elman Neural Network with Non-linear autoregressive network (three artificial neural networks) with statistical techniques (correlation analysis and Gamma test) to make water level prediction for urban pluvial flood control. It has the ability to resolve the issue of long-term dependencies in a time series. It effectively discovers the long-term dependencies through the recursive results and alleviates the variability issue in the results. The duration of rainfall affects the accuracy and reliability of the forecast model.

Rashidi *et. al.* [10] proposed a multi-criteria decision-making method (MCDM) for flood susceptibility mapping. The results show that 75% of the study area are highly susceptible to flooding and that SVM model perform best and the integration of ML and MCDM results show that 6% of the study area is at high flood risk, it also reveals that population density and area density influence the vulnerability of flood.

Noymanee *et al.* [11] developed a water level machine learning based model (Bayesian linear model) for open date in predicting flood peak in urban regions. This model could be used for short term warning system and have the ability to handle complex task. It is limited by time and resources and alteration from data set can create a radical variation for the model.

Lee *et. al.* [12] used frequency ratio and logistic ratio models as data mining techniques with geographic information system (GIS) tools in generating susceptibility map to correlate between flood data and related factors. This model was useful in clarifying the mechanism amid flood occasions and related variables. The model validation was affected due to difficulty of acquiring data.

Seyoum *et. al.* [13] proposed an interactive and cooperative framework (data driven - a multilayer perception Artificial Neural Networks and Random Forest) in refining the monitoring and managing pluvial flood. It shows promise in the absence of hydraulic model. To improve the prediction of pluvial flows, test of different data transformation technique will be needed.

3. Methodology

The methodology implemented for the study is summarized in Figure 1 as a proposed workflow of this study.

3.1 Datasets And Attributes

The pluvial flood dataset generation was based on the identified conditioning variables related to pluvial flood, with the interpretation of Shuttle Radar Topography Mission (SRTM DEM) Digital Elevation Model land imagery. The data were interpreted to form a basis for geo-spatial database of the dataset using the python module of the ARCGIS software. To elucidate the advantages of the developed model, the dataset was generated from one of the South West States of Nigeria. This is Oyo State which is one of the leading urban areas in Nigeria.

The study was finally conducted in Ibadan metropolis, which comprises of 11 local government areas (LG) at the outskirts and 5 local government areas at the urban areas. The latter are: Ibadan North LG, Ibadan North-East LG, Ibadan North-West LG, Ibadan South-West LG and Ibadan South-East LG Areas. However, the justification for selecting the five local government areas was based on the fact that it is fast growing in terms of level of physical development and characterised with various commercial activities and due to large dataset generated from the whole state. The geo-spatial database consists of eight conditioning variables namely slope, aspect, curvature, flow accumulation, rainfall, topographic wet index, drainage network and drainage density (Table 1). After which the filter feature selection method was adopted to ascertain the features which contribute most to the expected outcome. The description of data is presented in Table 1.

3.1.1 Pluvial Flood Data Classification

The concept of fuzzy logic is a computerised thinking technique, which can imitate compound human ideas. The strength lies in the addition of logics (Boolean) to a fuzzy set of partial truths, whose outputs are continually explained within 1 and 0. It comprises three main operations as shown in Figure 2. Firstly, is the fuzzification which draws an input example to a membership importance using the membership function and was implemented using the triangular type. This was followed by inference, in this section, the fuzzified data were deduced and analysed considering some set of fuzzy rules. Lastly, defuzzification was used to assign the analysed output variables with the precise decision.

3.1.2 Pluvial Flood Conditioning Variable

The three-categorization constructed on the tool of pluvial flooding was carefully chosen as reported in [4] and considered useful for this study with eight sub variables including slope, curvature, aspect, topographic wet index, flow accumulation, drainage density, drainage network and rainfall as shown in Figure 3 – 10.

3.2 Evaluation Parameters

In this study, we designed the several classifier models using different machine learning classification algorithm namely, K-Nearest Neighbours (KNN), Random Forest (RF), Classification and Regression Trees (CART), Naïve Bayes (NB) and Artificial Neural Network (ANN) algorithms to classify the fuzzified conditioning variables and to predict pluvial flood risk. . In evaluating the model performance, on the obtained set of conditioning variables, the study employed five measures of performance which comprised the sensitivity, specificity, accuracy percentages, precision and the area under receiver operating characteristics for choosing the best fit classifier. Kappa statistics was used to provide valuable information on the reliability of the performance metrics. Sensitivity expresses the correctly categorized positive instances, where specificity expresses the correctly categorized negative instances and accuracy is the proportion of suitably categorized instances. Precision is the amount of categorized defective instances, which are effectively defective instances. Kappa Statistics measures the comparison or relationship (similarity) of ensemble in multi-classifier systems. Receiver Operating Characteristic demonstrates the compromise between the true positive (TP) and the false positive (FP) rates, the accuracy of a classifier is represented by the area under curve, the bigger the region covered by the area under curve, the efficient the classifier. The value of the sensitivity, specificity, accuracy percentages, precision and kappa statistics are as shown in equation 1 to 5.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

$$\text{Sensitivity} = \frac{TP}{TP+FN} \quad (2)$$

$$\text{Specificity} = \frac{TN}{FP+TN} \quad (3)$$

$$\text{Precision} = \frac{TP}{TP+FP} \quad (4)$$

Where TN and TP stand for true negative and positive whereas FN and FP denote false negative and positive.

$$\text{Kappa Statistics} = \frac{P(A)-P(E)}{1-P(E)} \quad (5)$$

Table 1: Data Type and Source

S/N	Type	Scale Resolution	Period	Source
1	Slope	30 x 30	2019	SRTM DEM (USGS)
2	Aspect	30 x 30	2019	SRTM DEM (USGS)
3	Curvature	30 x 30	2019	SRTM DEM (USGS)
4	Flow Accumulation	30 x 30	2019	SRTM DEM (USGS)
5	Rainfall	30 x 30	2019	Copernicus Climate Data Store
6	Topographic Wet Index	30 x 30	2019	SRTM DEM (USGS)
7	Drainage Density/Network	30 x 30	2019	SRTM DEM (USGS)

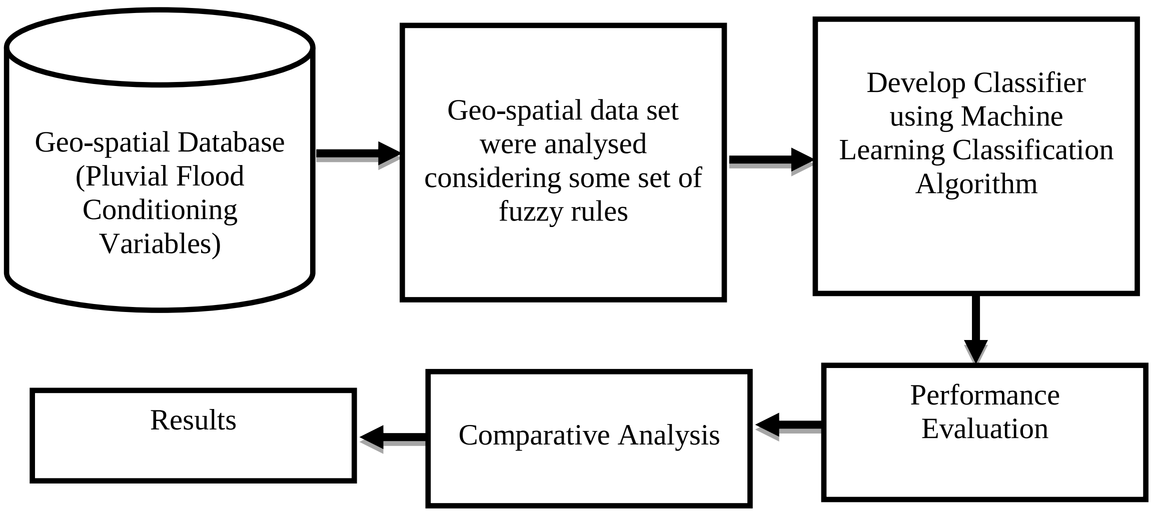

Figure 1: Methodology Workflow

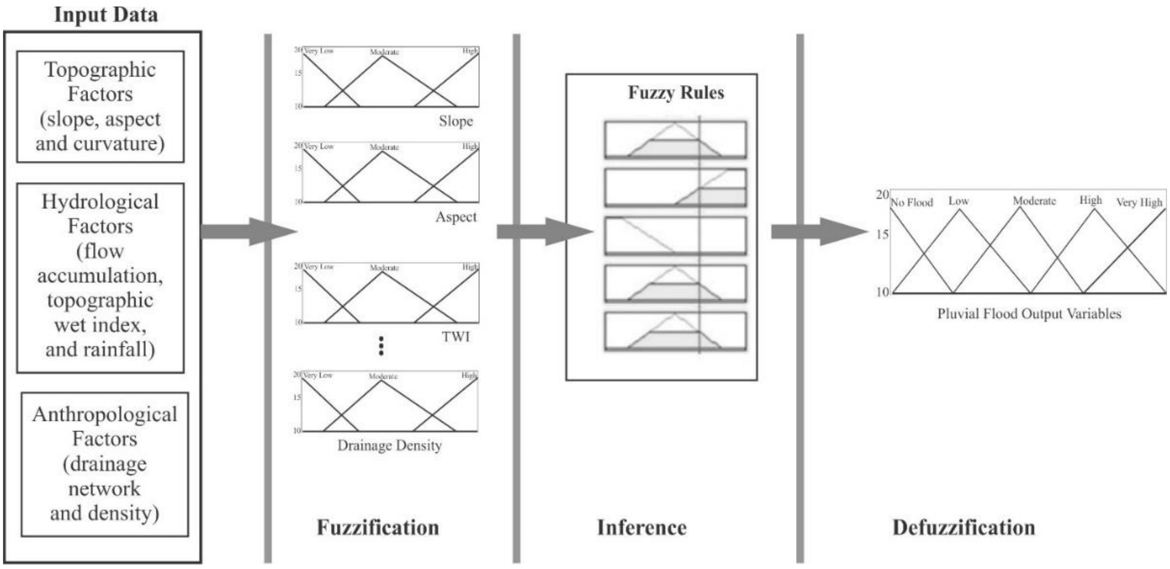

Figure 2: Fuzzy-Rule Based Approach

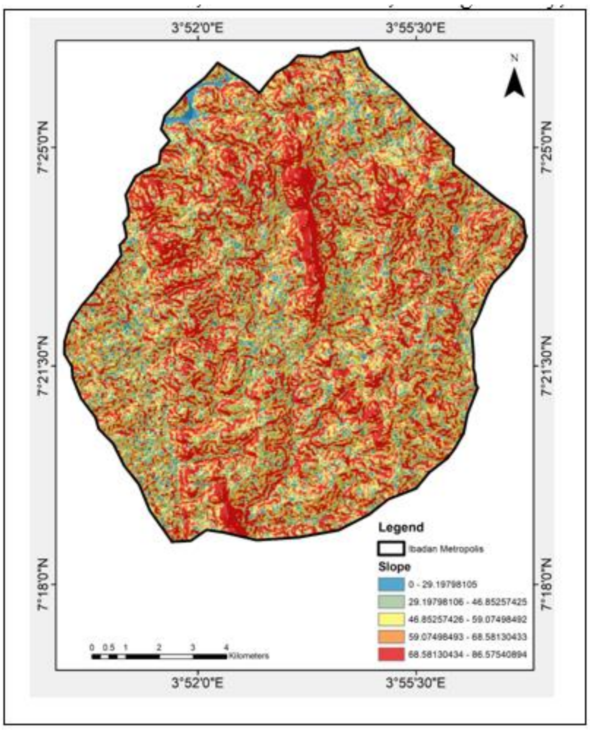

Figure 3: Slope

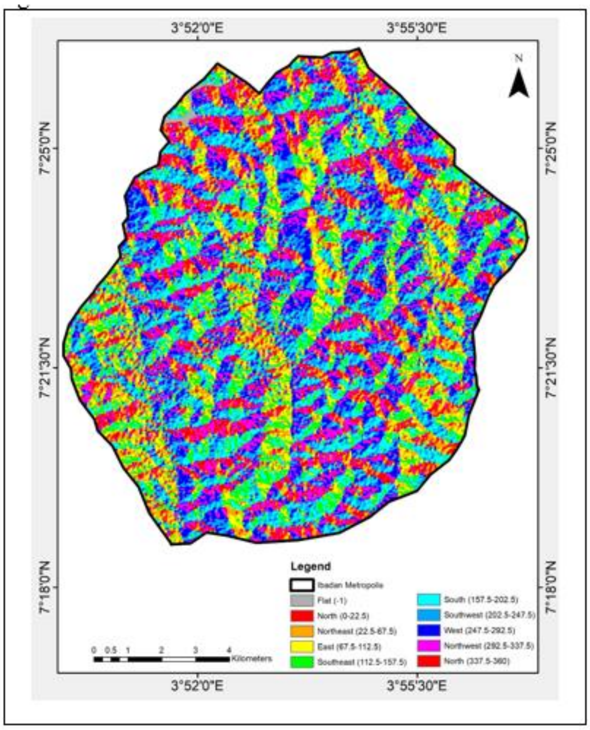

Figure 4: Aspect

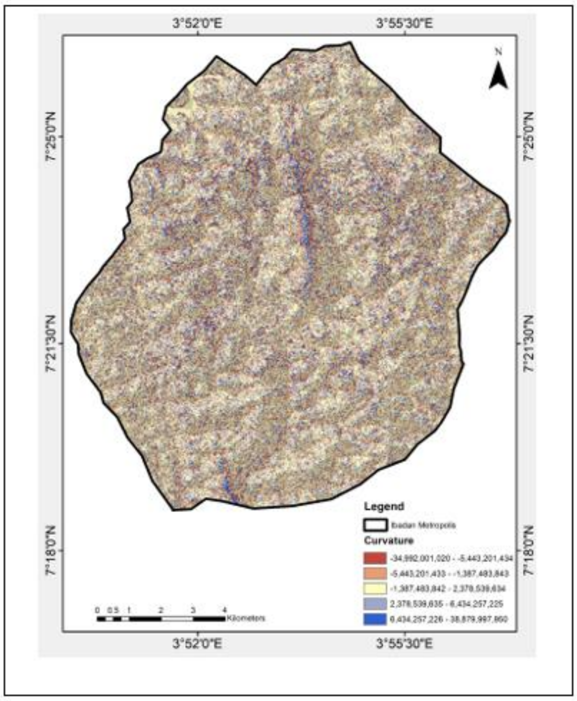

Figure 5: Curvature

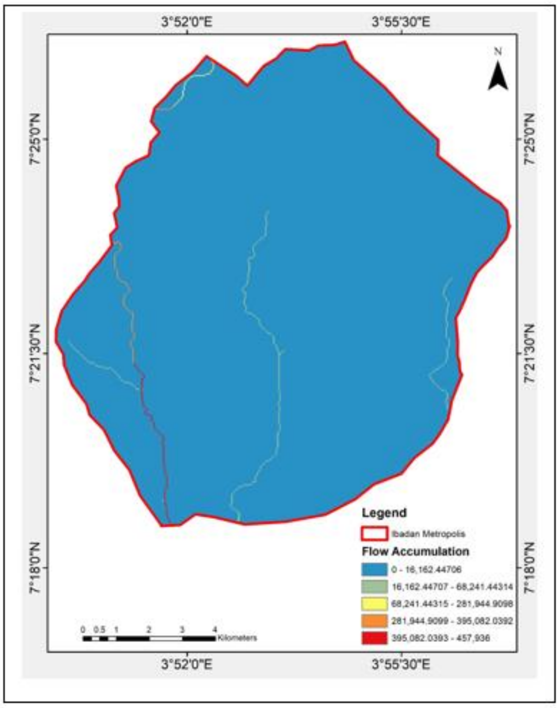

Figure 6: Flow Accumulation

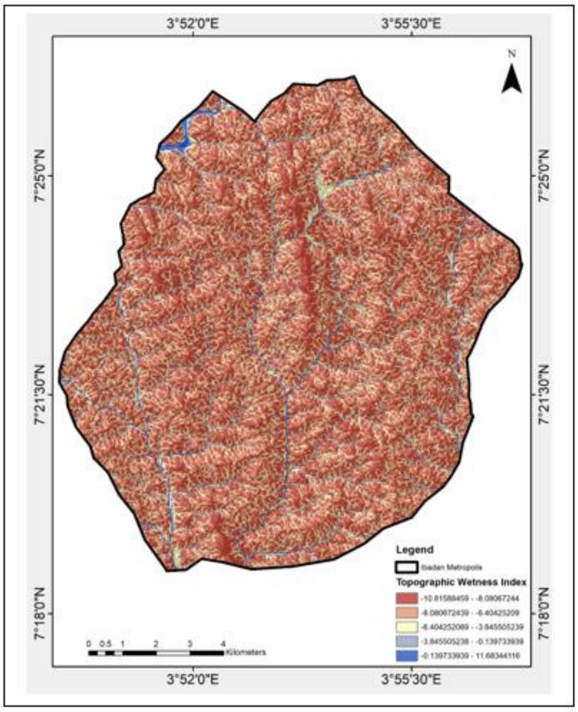

Figure 7: Topographic Wet Index

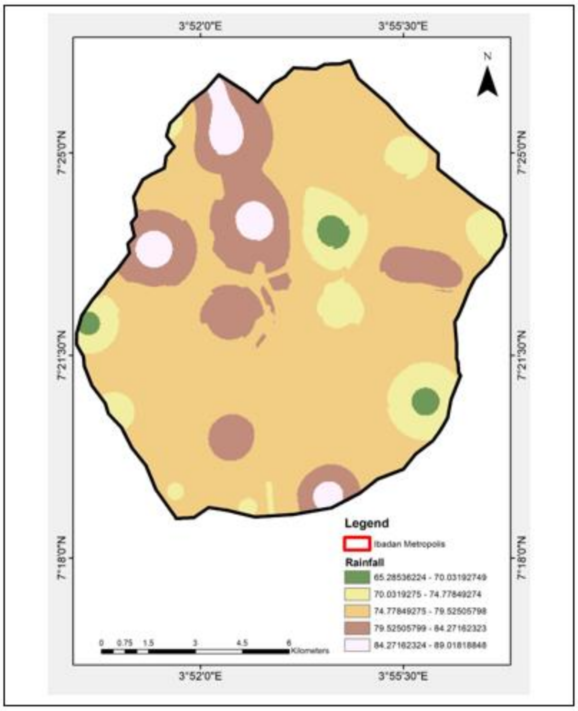

Figure 8: Rainfall Map

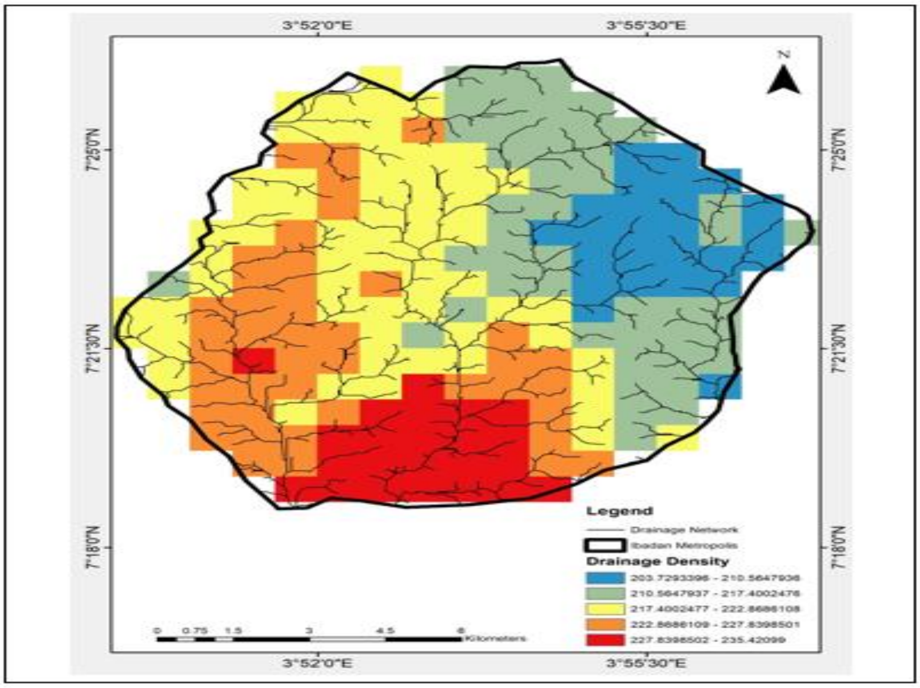

Figure 9: Drainage Density and Network

To analyse the performance of the classifier models in detecting and predicting pluvial flood risk, first evaluating criteria has been defined. The performance evaluation of the five machine learning algorithms was done using the 10-Fold Cross-Validation (10-F C-V) as well as Holdout owing to the various location points of each conditioning variables. The 10-F C-V was used one time for each fold of each cross-validation and then for a final time for the complete geo-spatial dataset for each of the five machine learning algorithms in total of eleven times for each algorithm. The hold method was used as test data in the percentage-split ratio 80:20.

4. Results and Discussion

4.1 Results

The performance analysis of the considered classification algorithms over the sensitivity, specificity, accuracy percentages, precision, kappa statistics and the area under receiver operating characteristics are shown in the Table 2 and 3.

Table 2: Performance Analysis of classifier model in Cross Validation Method

MLAs	Accuracy (%)	TP_Rate (Sensitivity) (%)	TN_Rate (Specificity) (%)	Precision (%)	Kappa_Statistics (%)	AUROC (%)
KNN	96.4	96.4	94.1	96.6	95.4	93.7
RF	95.2	95.2	93.7	94.3	86.3	94.6
NB	77.3	77.3	74.7	84.3	71.1	89.5
CART	99.5	99.5	99.4	99.5	99.3	97.6
ANN	89.5	89.5	89.7	89.1	89.3	89.9

Table 3: Performance Analysis of classifier model in Hold-Out Method

MLAs	Accuracy (%)	TP_Rate (Sensitivity) (%)	TN_Rate (Specificity) (%)	Precision (%)	Kappa_Statistics (%)	AUROC (%)
KNN	95.3	95.3	92.7	93.8	94.6	92.2
RF	72.8	73.0	73.2	72.0	77.2	83.6
NB	77.0	77.0	73.7	84.7	70.6	72.7
CART	98.4	98.4	98.3	98.4	99.2	98.6
ANN	83.6	84.0	96.9	74.0	84.7	87.9

4.2 Discussion

The performance analysis based on cross validation method, conducted on the algorithms as shown in Table 3. It is obvious that CART was positioned at top level in respect of classification accuracy percentage, sensitivity, specificity, precision, kappa statistics and AUROC with 99.5%, 99.5%, 99.4%, 99.5%, 99.3% and 97.6% respectively. KNN, RF, ANN and NB fell in the latter category respectively. The performance analysis based on hold-out method, steered on the algorithms as shown in Table 4. It is obvious that CART was positioned at top level in respect of classification accuracy percentage, sensitivity, specificity, precision, kappa statistics and AUROC with 98.4%, 98.4%, 98.3%, 98.4%, 99.2% and 98.6% respectively. KNN, ANN, NB and RF fell in the latter category respectively. Thus, we can say that the classifier model designed using CART classification algorithm can provide efficient prediction of pluvial flood risk among others.

The five machine learning algorithms performance were verified in relation to the obtainable standard which includes accuracy percentages, sensitivity, specificity, precision, kappa statistics and AUROC and comparison were made on their performances. The results of the evaluation showed that CART outperformed other algorithms in all metrics for both hold-out and 10F cross validation methods followed by KNN which has a close range in all metrics for both hold-out and 10F cross validation methods. This indicated that CART was the best classifier for detecting and predicting pluvial flood when dealing with multi-class classification for predictive analytics.

5. CONCLUSION

In conclusion, this study involves comparing the performances of five fuzzified prediction models for detecting and predicting of pluvial flood using a spatial database with 144, 401 location points and 8 conditioning variables. Feature transformation, feature selection/classification were carried out on the generated dataset which was pre-processed with fuzzy logic. The fuzzified prediction models were developed using five different types of machine learning classification algorithms namely: Naïve Bayes, Random Forest, Classification and Regression Tree, K-Nearest Neighbour and Artificial Neural Network. After the performance analysis, Classification and Regression Tree (CART) was established to be the best classifier out of the five and was used in building the model for pluvial flood detection and prediction. Fuzzy logic can emulate complex human thoughts and ease the decision-making process particularly when dealing with multi-class classification issues. This study helps in selecting best classifier for detecting and predicting pluvial floods.

References

[1] Pham, B. T., Luu, C., Phong, T. V., Nguyen, H. D., Le, H. V., Tran, T. Q., Ta, H. T., & Prakash, I. (2020). Flood risk assessment using hybrid artificial intelligence models integrated with multi-criteria decision analysis in Quang Nam Province, Vietnam. *Journal of Hydrology*, *592*, 125815. https://doi.org/10.1016/j.jhydrol.2020.125815

[2] Szewrański, S., Chruścinśki, J., Kazak, J., Swíader, M., Tokarczyk-Dorociak, K., & Zmuda, R. (2018). Pluvial Flood Risk Assessment Tool (PFRA) for rainwater management and adaptation to climate change in newly urbanised areas. Water (Switzerland), 10(4), 1–20. https://doi.org/10.3390/w10040386

[3] Noymanee, J., Nikitin, N. O., & Kalyuzhnaya, A. V. (2017). Urban Pluvial Flood Forecasting using Open Data with Machine Learning Techniques in Pattani Basin. Procedia Computer Science, 119(2017), 288–297. https://doi.org/10.1016/j.procs.2017.11.187

[4] Oladapo K.A., Idowu S.A., Adekunle, Y.A., and Ayankoya, F.A (2021). Categorization of Conditioning Variables for Pluvial Flood Risk Assessment. International Journal of Scientific and Engineering Research 11(8):355-367, August 2020.

[5] Nasiri, H., Mohd-Yusof, M. J., & Mohammad-Ali, T. A. (2016). An overview to flood vulnerability assessment methods. Sustainable Water Resources Management, 2(3), 331–336. https://doi.org/10.1007/s40899-016-0051-x

[6] Parizi, E., Bagheri-Gavkosh, M., Hosseini, S. M., & Geravand, F. (2021). Linkage of geographically weighted regression with spatial cluster analyses for regionalization of flood peak discharges drivers: Case studies across Iran. *Journal of Cleaner Production*, *310*, 127526. https://doi.org/10.1016/j.jclepro.2021.127526

[7] Yeganeh, N., & Sabri, S. (2014). Flood vulnerability assessment in Iskandar Malaysia using multi-criteria evaluation and fuzzy logic. Research Journal of Applied Sciences,

Engineering and Technology, 8(16), 1794–1806. https://doi.org/10.19026/rjaset.8.1167

[8] Chang, F. J., Chen, P. A., Lu, Y. R., Huang, E., & Chang, K. Y. (2014). Real-time multi-step-ahead water level forecasting by recurrent neural networks for urban flood control. Journal of Hydrology, 517, 836–846. https://doi.org/10.1016/j.jhydrol.2014.06.013

[9] Nasiri, H., Mohd-Yusof, M. J., & Mohammad-Ali, T. A. (2016). An overview to flood vulnerability assessment methods. Sustainable Water Resources Management, 2(3), 331–336. https://doi.org/10.1007/s40899-016-0051-x

[10] Rashidi Shikhteymour, S., Borji, M., Bagheri-Gavkosh, M., Azimi, E., & Collins, T. W. (2023). A novel approach for assessing flood risk with machine learning and multi-criteria decision-making methods. *Applied Geography*, *158*, 103035. https://doi.org/10.1016/j.apgeog.2023.103035

[11] Noymanee, J., Nikitin, N. O., & Kalyuzhnaya, A. V. (2017). Urban Pluvial Flood Forecasting using Open Data with Machine Learning Techniques in Pattani Basin. Procedia Computer Science, 119(2017), 288–297. https://doi.org/10.1016/j.procs.2017.11.187

[12] Lee, S., Lee, S., Lee, M. J., & Jung, H. S. (2018). Spatial assessment of urban flood susceptibility using data mining and geographic information system (GIS) tools. Sustainability (Switzerland), 10(3). https://doi.org/10.3390/su10030648

[13] Seyoum, S., Willems, P., & Verbeiren, B. (2019). Data-Driven Models for Pluvial Flood Forecasting in Brussels. NOVATECH, 1–4.