Classification of Flood Disasters Severity Levels by Employing Machine Learning Techniques

S Rajeshbabu, G Sakthivel, A Arulkumar and K Kannan

January 12, 2023

CLASSIFICATION OF FLOOD DISASTERS SEVERITY LEVELS BY EMPLOYING MACHINE LEARNING TECHNIQUES

S.Rajeshbabu[1*], G.Sakthivel[2*] A.Arulkumar[3*], K.Kannan[4*]

1* Assistant Professor, Mechatronics Department, Kamaraj College of Engineering and Technology, K.Vellakulam, Madurai District-625701, Tamilnadu, India,
2* Assistant Professor, Mechatronics Department, Kamaraj College of Engineering and Technology, , K.Vellakulam,Madurai District-625701, Tamilnadu, India,
3* Assistant Professor, Mechatronics Department, Kamaraj College of Engineering and Technology, , K.Vellakulam,Madurai District-625701, Tamilnadu, India,
4* Professor, Mechatronics Department, Kamaraj College of Engineering and Technology, K.Vellakulam, Madurai District-625701, Tamilnadu, India,
rajeshbabumtr@kamarajengg.edu.in , sakthivelmtr@kamarajengg.edu.in , arulkumarmtr@kamarajengg.edu.in , hodmtr@kamarajengg.edu.in

Abstract

Natural disasters can happen at any time and pose a risk to individuals. Numerous disasters have the potential to affect the economy of the country as well as the people. The most frequent disaster that happens worldwide and severely affects people is flooding. The article provides machine learning classification techniques, like decision trees , Naive Bayes, support vector machines (SVM), and K-Nearest Neighbour (KNN) with severity index,. The severity index was established using economic loss as a starting point. To define the damage, however, a variety of factors were taken into consideration, such as the total number of fatalities, injuries, and flood victims, as well as the costs associated with reconstruction and the overall economic losses. The study reveals that, the Ensemble and decision tree classifiers provide better ideal categorization in terms of accuracy and error.

Keywords: Natural disasters, Decision trees, Support Vector Machines, K-Nearest Neighbour Naïve Bayes and Ensemble classifiers

1. Introduction

Natural disasters have occurred mostly in recent years, usually causes significant loss to both individuals and entire nation. The big disaster that may occur mostly in India includes floods, tsunamis, earthquakes, cyclones, and famine. One of the most frequent disasters among such occurrences is flooding. The states that are most affected by flooding include Kerala, Kashmir, and Uttarakhand. Based on a study, the event has results in more than 10,000 fatalities and more than $10 million in economic losses. As a result of such a natural disaster, there is infrastructure damage, food and water shortages, public health difficulties, environmental impact, and economic impact [1]. Since natural disasters were unpredictable that may affect any type of concern a fuzzy keywords-driven Natural Disasters Warning is introduced for the enterprises [2]. Some real time disaster events and their analysis were discussed [3-6]. Though we have some analysis forecasting and monitoring is also essential for such disaster events there are several methods adopted [7-9].The above mentioned article initiated to implement machine learning based classification method for the flood disaster event. However some classification method applications like fuzzy logic, support vector machines and neural networks led the further development in the work [10-13].Apart from the other types of classification techniques decision tree classifier and KNN Classifier have given promising classification levels [14&15].

1.1 FLOOD DISASTER DATA EVENT CLASSIFICATION

The flood disaster data events were sampled and processed in three processing units. Initially the disaster events were sampled and processed using principal component analysis and classified using different types of classifiers. The entire progression of classification is revealed in the Figure.1.1.

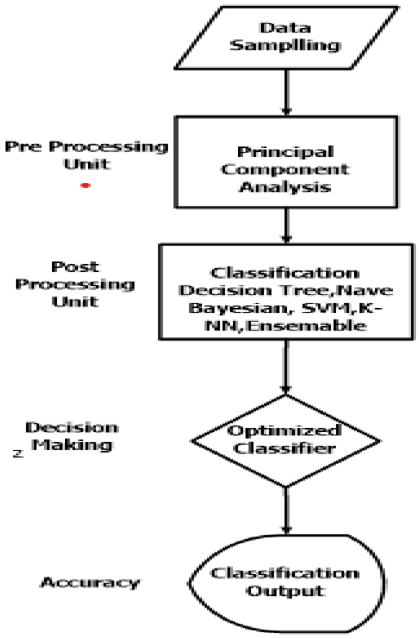


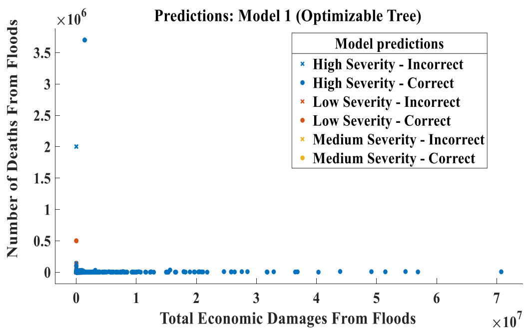
Figure.1.1.Process Flow Chart of Flood Disaster Event Classification

2. Methods of Machine learning based Classifications

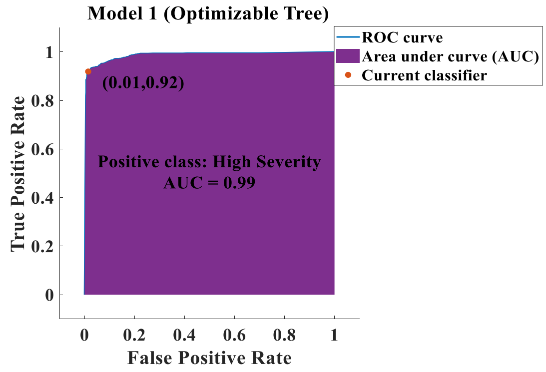
A classification model seeks to infer some significance from the values that were observed. It will attempt to forecast the value of one or more outputs given one or more inputs. Various categorization models exist. Like Logistic regression, decision trees, random forests, support vector machines, multilayer perceptron's, and Naive Bayes are some of classification models.

2.1. Decision trees

The decision tree is a technique used to create regression models in the form of a tree structure. Decision trees are frequently used in machine learning and data mining applications to divide data into small groups. The scatter plot of the Flood data, Confusion Matrix and SOC were depicted in the fig.2.1 (a-c).



(a)



(b)

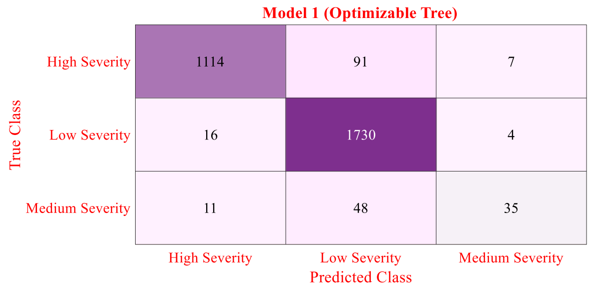


Figure 2.1. (a) Decision Tree Representation of Flood Data, (b) ROC (c) Confusion Matrix

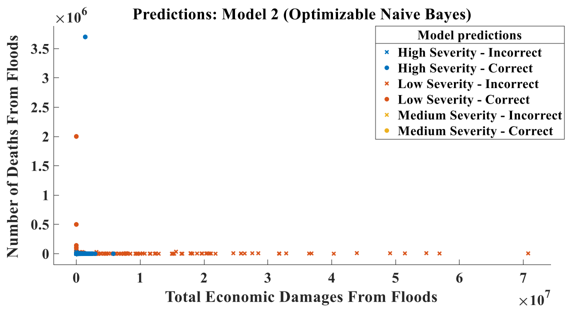
2.2. Naive Bayes Classifiers

A group of classification algorithms built on the Bayes' Theorem are known as naive Bayes classifiers. It is a family of algorithms rather than a single algorithm, and they all adhere to the same basic idea. The mathematical formulation of Bayes' theorem is given by the equation.1.

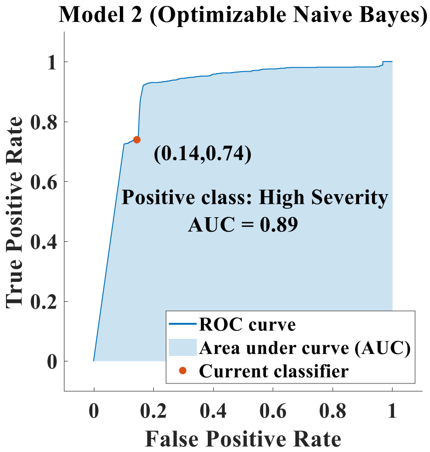
$$F(X,Y) = \frac{F(Y,X)F(X)}{F(Y)} \qquad (2.1)$$

Where X and Y are events and F(Y) ≠ 0.

The scatter plot of the Flood data, Confusion Matrix and ROC were depicted in the fig.2.2 (a-c).



(a)



(b)

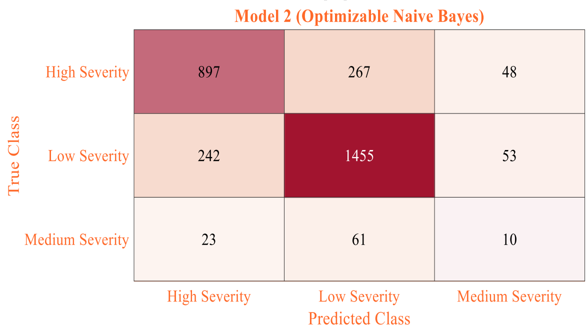


Figure 2.2. (a) Naive Bayes Representation Of Flood Data, (b) ROC and (c) Confusion Matrix

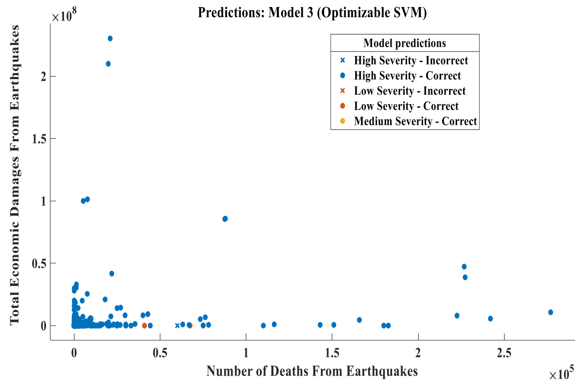
2.3. *Support Vector Machines (SVM)*

A SVM sorts data into classes by locating the optimal hyper plane that divides all of the data points in one class from those in the other. The hyper plane with the biggest margin between the two classes is the optimum hyper plane for a SVM. Margin is the maximum thickness of the slab that is perpendicular to the hyper plane however has no interior data points.

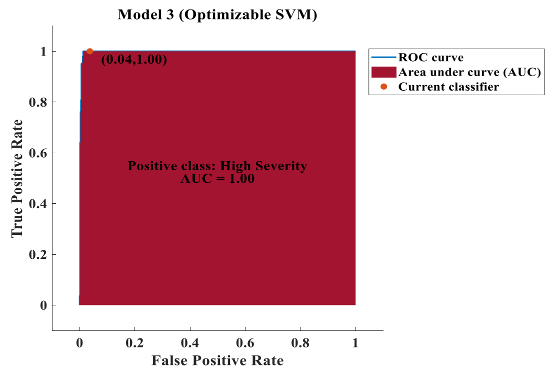
The equation of a hyper plane is given by the equation.2.2

$$F(A) = A'\gamma + b = 0 \qquad (2.2)$$

where $\gamma \in R^d$ and b is a real number. The scatter plot of the Flood data, Confusion Matrix and ROC were depicted in the fig.2.3 (a-c).



(a)



(b)

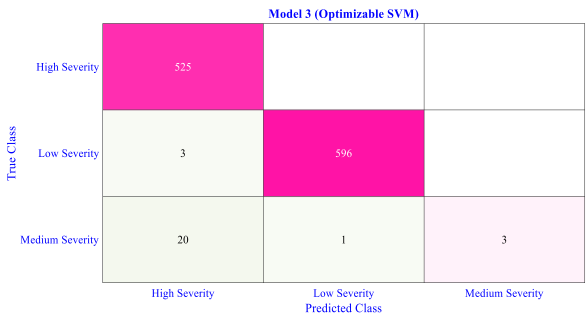


Figure 2.3. (a) SVM Representation of Flood data, (b) ROC and (c) Confusion Matrix

2.4. *K- Nearest-Neighbor (K-NN)*

The number of nearest neighbours and the distance metric may both be changed in the nearest-neighbor classification model known as K-NN.A Classification KNN classifier may be used to make reconstitution predictions since it saves training data. Alternately, employ the predict technique to categorise fresh observations using the model. The scatter plot of the

Flood data, Confusion Matrix and SOC were depicted in the fig.2.4 (a-c).

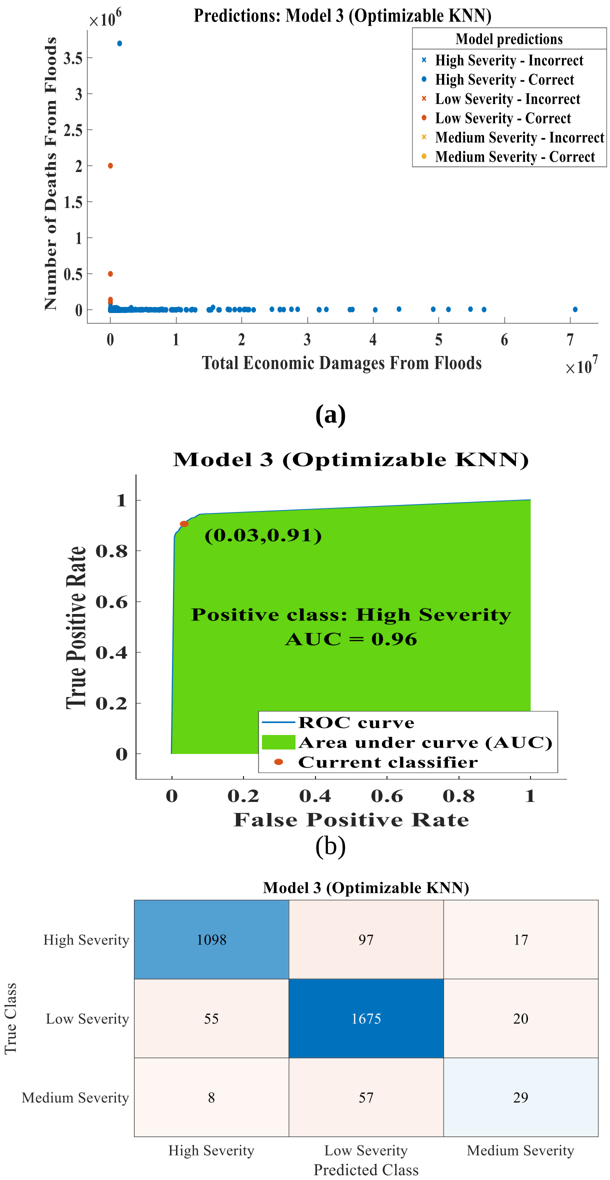


Figure 2.4. (a) K-NN Representation of Flood data, (b) SOC, (c) Confusion Matrix

2.5. Ensemble learning

A prediction model made up of a weighted mixture of many classification models is called a classification ensemble. In general, improving forecast accuracy by integrating several categorization models. The scatter plot of the Flood data, Confusion Matrix and SOC were depicted in the fig.2.5 (a-c).

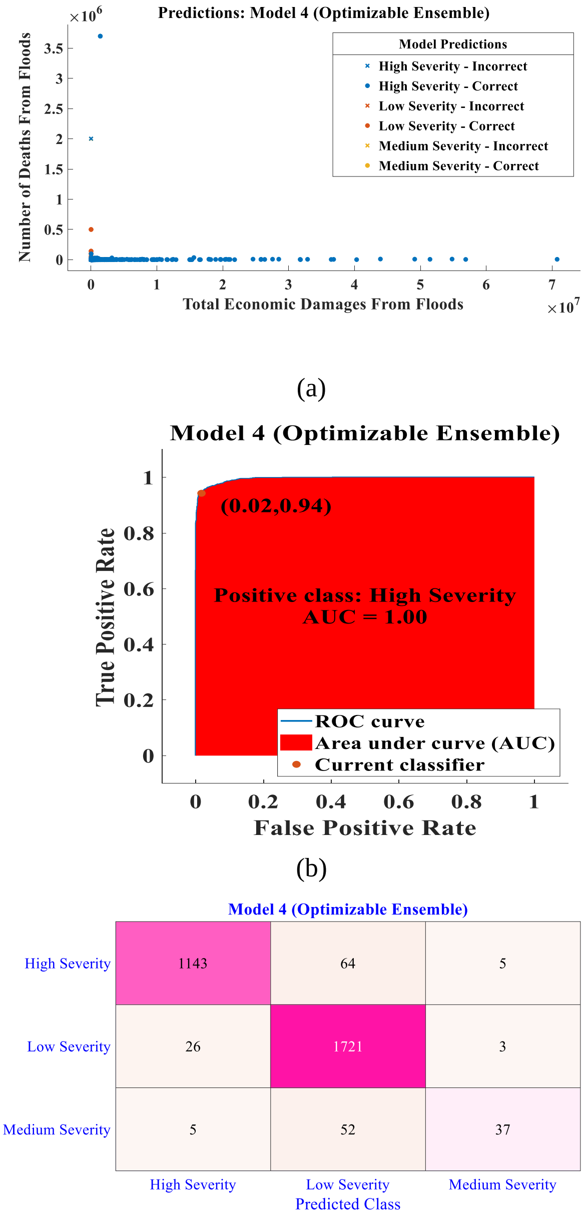


Figure 2.5. (a) Ensemble Learning Representation Of Flood Data, (b) ROC And (c) Confusion Matrix

3. SIMULATION RESULTS OF OPTIMIZED CLASSIFIERS

A simulation was carried out for 50 numbers of iterations. It is identified that the optimized decision tree and ensemble classifiers outperformed and SVM classifiers provided a low accuracy.

3.1 OPTIMIZED DECISION TREE CLASSIFIERS

The optimized tree classifiers provide the good accuracy level. The best point hyper parameters are achieved at the iteration range of 15. The maximum classification error plot is as shown in figure 3.1

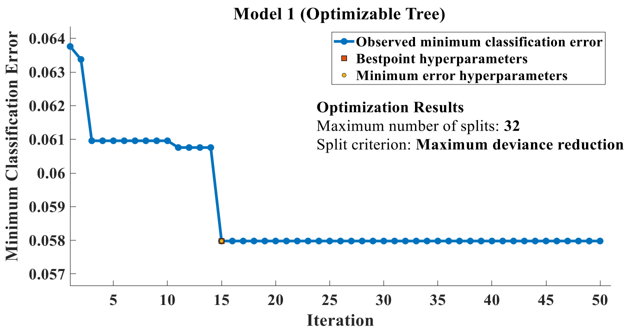
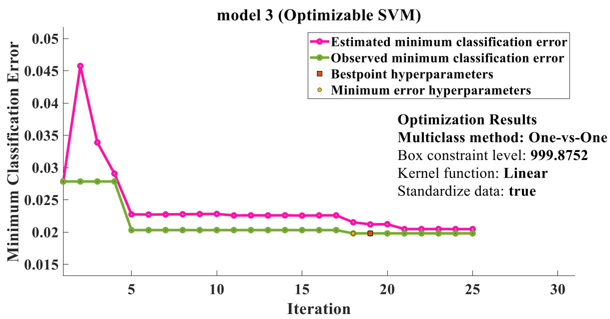


Figure 3.1.Maximum Classification Error Plot At Different Iterations

3.2 OPTIMIZED SVM CLASSIFIERS

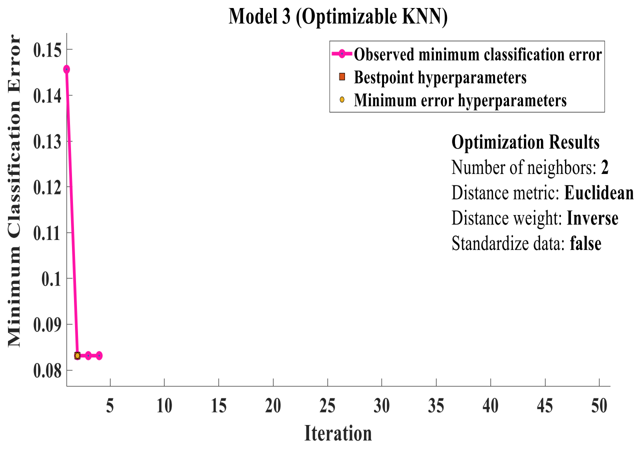
SVM classifiers offer a poor level of accuracy in the form of classification and offer a minimum level of error classification with a greater training time. The minimum classification error obtained by the SVM classifier is as shown in figure.3.3



3.2. Maximum Classification Error Plot at Different Iterations

3.3 OPTIMIZED ENSEMBLE CLASSIFIERS

The optimized ensemble classifiers provide a good accuracy level with minimum number of iterations. The minimum classification error of the optimized ensemble classifier is as shown in figure3.3.



3.3. Maximum Classification Error Plot At Different Iterations

Table.1. Performance Metrics of Different Types of Classifiers

Parameters	Classifiers				
	Optimized Decision tree	Optimized Naive Bayesian	Optimized SVM	Optimized K-NN	Optimized Ensemble
Accuracy (%)	94.2	77.3	42.1	91.7	94.9
Misclassification Cost	177	694	1768	254	155
Prediction Speed (obs/Sec)	8900	1600	1800	4100	2100
Training Time	176.6	401.15	315.17	354.32	462.45

The Misclassification cost is low in optimized ensemble classifier while compared to all the other types of classifiers. The performance of the different types of classifiers is as shown in the table.1.

3.4 TECHNICAL FINDINGS

1. The accuracy differs in all the classifiers however decision tree and ensemble classifiers play a lead role.
2. The misclassification cost varies with respect to the accuracy level.
3. The prediction speed is independent with accuracy and misclassification costs.
4. Based on the training time the accuracy level got altered.

4. CONCLUSION

The article proposes different types of classifier performance metrics for a flood disaster with three severity levels. The accuracy of the various classifier shows that an optimized decision tree and ensemble classifiers provides a good level of classification in terms of accuracy, training time and prediction speed. However the optimized SVM Classifier provides has a low level of accuracy. It is predicted that the decision tree and ensemble classifiers may provide the same form of classification for other types of disasters like earthquake, Tsunami, volcano cyclones.

5. ACKNOWLEDGEMENTS

The authors alone would be responsible for said statements made in the article.

6. REFERENCES

[1] List of Major Natural Disasters That Occurred in India till Date report on Maps of India, published By Rumani Saikia Phukan, May 11, 2022.

[2] Zewei Sun;Hanwen Liu;Chao Yan;Ran,'Natural disasters warning for enterprises through fuzzy keywords search',Tsinghua Science and Technology, pp.558 – 564, Volume: 26, Issue: 4, Year: 2021.

[3] Lulwa Ahmed;Kashif Ahmad;Naina Said;Basheer Qolomany;Junaid Qadir;Ala Al-Fuqaha,'Active Learning Based Federated Learning for Waste and Natural Disaster Image Classification', IEEE Access, pp. 208518 – 208531, Volume: 8, Year: 2020.

[4] Y. Roh, G. Heo and S. E. Whang, "A survey on data collection for machine learning: A big data—AI integration perspective", IEEE Transactions on . Knowl. Data Eng., Oct. 2019.

[5] K. Ahmad, K. Pogorelov, M. Riegler, O. Ostroukhova, P. Halvorsen, N. Conci, et al., "Automatic detection of passable roads after floods in remote sensed and social media data", Signal Process. Image Commun., vol. 74, pp. 110-118, May 2019.

[6] N. Said, K. Ahmad, M. Riegler, K. Pogorelov, L. Hassan, N. Ahmad, et al., "Natural disasters detection in social media and satellite imagery: A survey", Multimedia Tools Appl., vol. 78, no. 22, pp. 31267-31302, Nov. 2019.

[7] Mohammed Moishin;Ravinesh C. Deo;Ramendra Prasad;Nawin Raj;Shahab Abdulla ,'Designing Deep-Based Learning Flood Forecast Model With ConvLSTM Hybrid Algorithm',IEEE Access, Volume: 9 Year: 2021.

[8] Scott Miau,Wei-Hsi Hung,'River Flooding Forecasting and Anomaly Detection Based on Deep Learning', IEEE Access, Volume: 8, Year: 2020.

[9] SupattraPuttinaovarat,Paramate Horkaew,'Flood Forecasting System Based on Integrated Big and Crowdsource Data by Using Machine Learning Techniques,IEEE Access, Volume: 8, Year: 2020.

[10] P.-S. Yu, T.-C. Yang, S.-Y. Chen, C.-M. Kuo, and H.-W. Tseng, "Com-parison of random forests and support vector machine for real-time radar-derived rainfall forecasting," J. Hydrol., vol. 552, pp. 92–104, Sep. 2017.

[12] M. Dehghani, B. Saghafian, F. Nasiri Saleh, A. Farokhnia, and R. Noori, "Uncertainty analysis of streamflow drought forecast using artificial neural networks and Monte–Carlo simulation," Int. J. Climatol., vol. 34, no. 4, pp. 1169–1180, Mar. 2014.

[13] A. K. Lohani, N. Goel, and K. Bhatia, "Improving real time flood fore-casting using fuzzy inference system," J. Hydrol., vol. 509, pp. 25–41, Feb. 2014.

[14] Leonardo Cañete-Sifuentes,Raúl Monroy,Miguel Angel Medina-Pérez,'A Review and Experimental Comparison of Multivariate Decision Trees', IEEE Access, pp. 110451 – 110479,Volume: 9, Year: 2021.

[15] Omar Kherif, Youcef Benmahamed, Madjid Teguar, Ahmed Boubakeur, Sherif S. M. Ghoneim,' Accuracy Improvement of Power Transformer Faults Diagnostic Using KNN Classifier With Decision Tree Principle, IEEE Access,pp. 81693 – 81701, Volume: 9, June 2021.

[16] www.ourworldindata.org

Biographies

Rajeshbabu.S was born in India in 1985. He received his B.E. (EEE), M.E. (Power System Engineering), and Ph.D. (Electrical Engineering) degrees from Anna University in 2006, 2008, and 2019. He is an assistant professor in the department of mechatronics engineering at Kamaraj College of Engineering and Technology, Madurai, India. His areas of interest are renewable energy systems, power quality. He is a chartered engineer as well as a member of the Institute of Engineers (India) and the Fluid Power Society of India (FPSI).

Sakthivel. G. received his B.E. degree in Electrical and Electronics Engineering from Anna University, Chennai, India, in 2009 and his M.E. degree in Power Electronics and Drives from Anna University, in 2012. He completed his Ph.D. degree from Anna University. Presently, he is working as an Assistant Professor in the Department of Mechatronics Engineering at Kamaraj College of Engineering and Technology, Madurai, India. His current research interests includes Solar PV,Energy conservation in electric motor drives, Power converters, Power quality, and Electric vehicles.

Arulkumar.A. was born in India in 1987. He received his B.E. (EEE), M.E. (Power Electronics and Drives), from Anna University in 2010, 2014 and currently pursuing Ph.D. (Electrical Engineering) under part time category in Anna university. He is an assistant professor in the department of mechatronics engineering at Kamaraj College of Engineering and Technology, Madurai, India. His areas of interest are renewable energy systems, power quality. He is a member of Indian Society for Technical Education.(ISTE).

Kannan. K. was born in India in1971. He received his B.E. (ECE) from Madurai Kamaraj University in the year 1992,M.E. (Industrial Engineering) from Anna Univerisity in 2004 and Ph.D. (Industrial Engineering) from Manonmanium Sundaranar University in 2014. He is now working as a professor in the department of mechatronics engineering at Kamaraj College of Engineering and Technology, Virudhunagar, India. His area of interest includes Robotics, Mechatronics, Image Processing .