Summer 5-17-2021

Bibliometric Survey on Flood Prediction using Machine Learning

seema patil Prof.
Symbiosis International University, seemap@sitpune.edu.in

Daksh Khurana Mr.
Symbiosis International University, daksh.khurana.btech2018@sitpune.edu.in

Kartik Rao Mr
Symbiosis International University, kartik.rao.btech2018@sitpune.edu.in

Priyanshu Meena Mr
Symbiosis International University, priyanshu.meena.btech2018@sitpune.edu.in

Shivendra Singh Mr
Symbiosis International University, shivendra.singh.btech2018@sitpune.edu.in

Bibliometric Survey on Flood Prediction using Machine Learning

Seema Patil[1], Daksh Khurana[2], Kartik Rao[3], Priyanshu Meena[4], Shivendra Singh[5]

1 Professor, Symbiosis Institute of Technology (SIT) affiliated to Symbiosis International (Deemed University), Pune, India.

Email: seemap@sitpune.edu.in

2 Student, Symbiosis Institute of Technology (SIT) affiliated to Symbiosis International (Deemed University), Pune, India.

Email: daksh.khurana.btech2018@sitpune.edu.in

3 Student, Symbiosis Institute of Technology (SIT) affiliated to Symbiosis International (Deemed University), Pune, India.

Email: kartik.rao.btech2018@sitpune.edu.in

4 Student, Symbiosis Institute of Technology (SIT) affiliated to Symbiosis International (Deemed University), Pune, India.

Email: priyanshu.meena.btech2018@sitpune.edu.in

5 Student, Symbiosis Institute of Technology (SIT) affiliated to Symbiosis International (Deemed University), Pune, India.

Email: shivendra.singh.btech2018@sitpune.edu.in

Abstract

Floods are one of the most devastating natural hazards, and modelling them is extremely difficult. Flood prediction model advancement study led to factors such as loss of human and animal life, property damage, and risk mitigation. The focus of this bibliometric survey is to recognise the few studies which have upheld on the factors affecting the floods. The analysis is done based on 254 documents such as articles, conference papers, article reviews and some reviews and notes. India contributes to the maximum number of documents followed by China and the United States of America. This bibliometric survey is conducted using Scopus. The survey includes analysis by the type of country or territory the documents are written in, authors contributing to the area of research, and statistical analysis based on citations, subject areas, and source types. This bibliometric survey revealed that the maximum numbers of publications on flood prediction using machine learning are from articles, etc, affiliated with Duy Tan University in the year 2020. Most of the research (i.e., 22.6%) is carried out by the Environmental Sciences department. Citation's graph shows that highest numbers of citations are in the year 2020.

Keywords: Floods, Machine Learning, Bibliometric survey

1. Introduction

Among the natural disasters, floods are the most dangerous and destructive natural disasters causing massive damage to environment and are difficult to model. The most destructive flood ever took place in the history of humanity was in the year 1931 known as China floods or the Yangtze–Huai River floods. The flood inundated approximately 180,000 square kilometres which is equivalent to the size of England and half of Scotland resulting in an estimation of 3.7 to 4 million deaths. Estimated death tolls vary widely.

The research on the factors affecting the floods can be contributed to loss of human and animal life, property damage, and risk reduction. Therefore, governments are under pressure to design a model and a platform to predict floods and prevent the country or town from this disaster. The prediction of floods is highly complex due to the dynamic nature of climate condition. The focus of this bibliometric survey is to recognise the studies which upheld that the effect of climate factors is playing a crucial role for the floods to take place. The below table depicts some of the top incidents and the deaths caused by the floods in various countries with respect to the year in which the incident took place.

Death Toll	Event	Location	Year
500,000–4,000,000	1931 China Floods	China	1931
900,000–2,000,000	1887 Yellow River flood	China	1887
500,000–800,000	1938 Yellow River flood	China	1939
229,000	Typhoon Nina	China	1975
145,000	1935 Yangtze flood	China	1935
100,000+	St. Felix's Flood, storm surge	Netherlands	1530
up to 100,000	The flood of 1099	Netherlands & England	1099
5,748	2013 North India floods	India	2013
3,076	2004 Eastern India, Bangladesh monsoon rain	India & Bangladesh	2004
1,000-8,000	2016 Indian floods by monsoon rain	India	2016

Table 1: The Deadliest floods

Source: https://en.wikipedia.org

As disasters lead to great loses, Internet based technology could help in prediction of floods. As internet is being used by majority of 2016 person in the world, most of the information and happenings in the world can be accessed by them anywhere they want to. Similarly, the forecast or an alert can be communicated to them on their devices. These include sensors like the Global Positioning System (GPS), Global Navigation Satellite System (GNSS), and ultrasonic sensors, all of which can be easily installed on mobile phones and provide extremely precise and reliable location information. All critical information can be quickly conveyed to them, allowing them to prepare ahead of time.

The Internet of Things, or IoT, is now widely used in a variety of fields of science, including flood prediction. Wireless sensor networks are one of the innovations (WSN). The physical conditions of a given area are monitored and recorded by many of these instruments, which are made up of spatially distributed sensors. Floods can be predicted using a combination of IoT and neural networks. This is a vital technology in many developing countries to reduce the after-effects of flooding, such as significant loss of life and property. Large losses can be avoided with the use of flood prediction models and systems. During risky weather conditions around the world, these systems provide early warning and accurate data collection.

2. Precursory Data Collection

The Internet holds a vast database of published papers which are useful for this bibliometric survey. The most widely accessed platforms from which these publications can be viewed are Google Scholar, Scopus,ResearchGate and ScienceDirect. While some of these publications are Open Access but most of them require registration to these websites or subscription to their plans. Our institute's library allowed us access to Scopus and ScienceDirect.

Scopus host a huge repository of scientific journals research papers and articles from publishers around the world. It has been proven as a trusted source of accurate information internationally.

3. Primary Keywords

The following table represents the keywords used to perform the search of required publications and data. The keywords are split into two parts that is master key word and primary keyword which were formulated together in a query to give us the much-needed results for this bibliometricanalysis.

Master Keyword	"Floods"
Primary Keywords Using (OR)	"Prediction", "Machine Learning"

Table 2: Planned Search Tactics

Thus, the query used to search the documents in the Scopus are: "Floods" OR "Prediction" OR "Machine Learning".

3.1.Key Findings

We have used Scopus database for this bibliometric paper. The search conducted using the search keyword tactic resulted in 254 papers and these papers are in many different languages. It includes consideration of both published and unpublished publications.

These publications have different categories including article, review, conference papers, conference reviews, and note. All the data related to the publications are given below:

Sr. No	Publication Type	Number of Publications
1	Article	182
2	Conference Paper	61

3	Conference Review	7
4	Review	3
5	Note	1

Table 3: Publication Types

Source: https://www.scopus.com

These are the publications of different types of keywords like machine learning, floods, prediction, forecasting, etc. All the data related to the keywords is shown in the table below:

Keywords	Number of Publication
Machine Learning	173
Floods	134
Forecasting	97
Flood Control	66
Learning Systems	56
Prediction	56
Decision Trees	45
Learning Algorithms	42
Rain	38
Artificial Neural Network	36

Table4: Keywords

Source: https://www.scopus.com

3.2.Exploratory Data Features

The publications considered for this bibliometric review have a count of 254. Out of which 78 have been published in 2019, 119 in 2020 and 57 in 2021. Scopus Database was used for the data.

Year	Publications Count

2021	57
2020	119
2019	78

Table 5: Publications per Year

Source: https://www.scopus.com

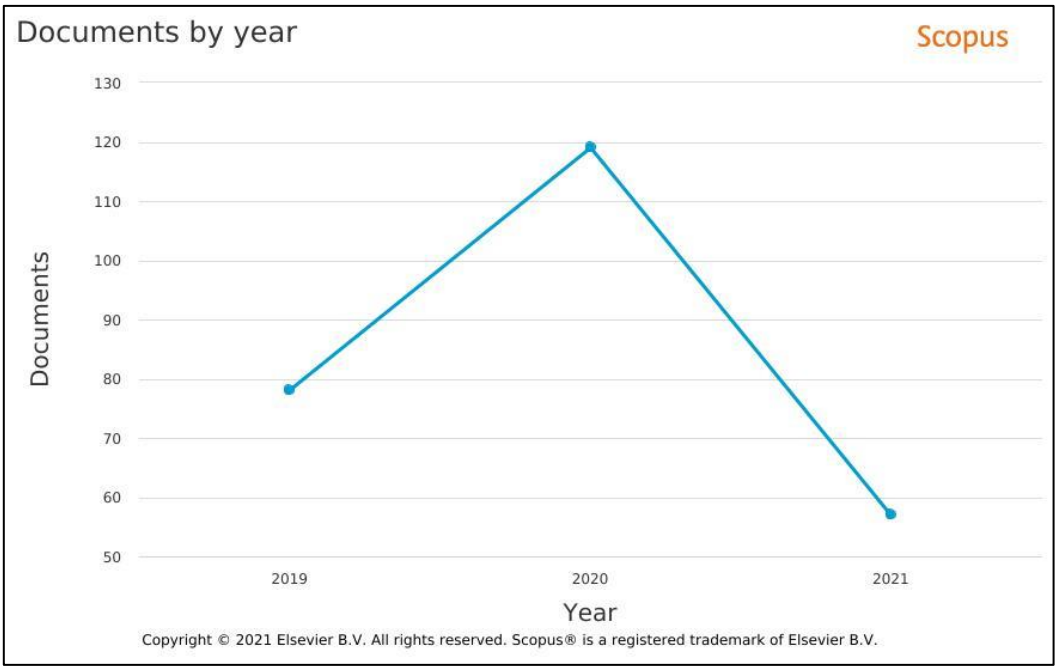

Figure 1: Documents by Year

Source: https://www.scopus.com(accessed on 2nd May 2021)

4. Bibliometric Study

To perform the bibliometric analysis for finding out factors affecting floods and its prediction, we applied the following two ways-:

- Geographic Region, Network and Citation Analysis
- Demography about the keyword, affiliation, author, and subject areas

4.1.Country or Territory based Statistical analysis

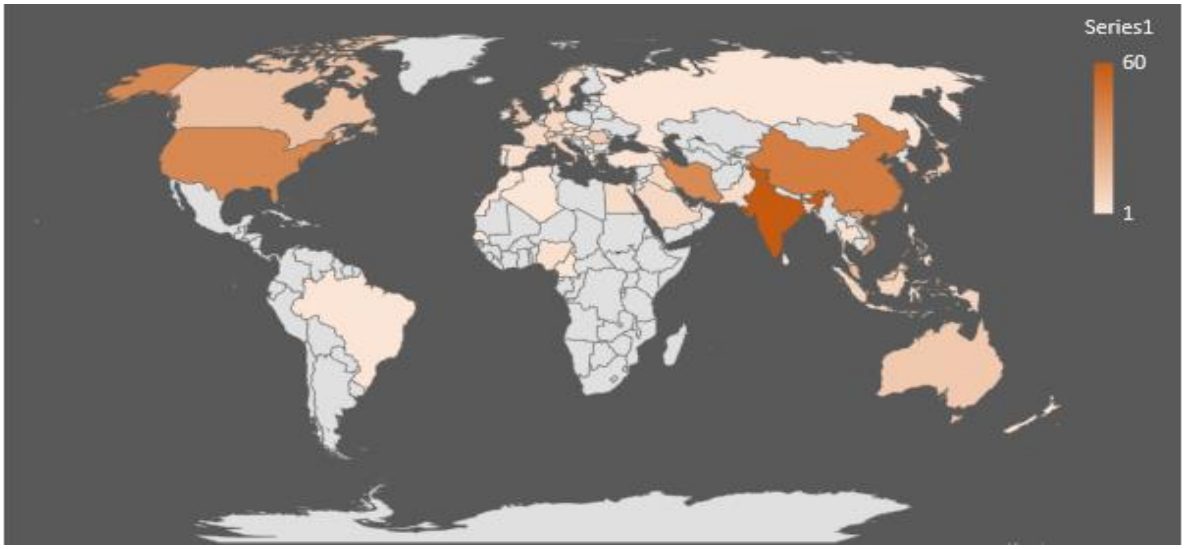

Figure 2: Topographic Provincial Analysis

Source: https://www.scopus.com (accessed on 2nd May 2021)

The above map chart shows the distribution analysis of the research carried out by countries all over the world. The higher the concentration of the colour means the larger contributions over the past couple of years related to the topic of flood prediction using machine learning.

The bar chart given below shows the top ten nations that have contributed to the research. The length of the bar from left to right indicates the total number of publications from a specific country which is labelled beside the graph. The longer the bar means the higher number of published documents from that country. In our research we found out that the greatest numbers of related publications are from India, then followed by China and the United States. Australia and the United Kingdom has the least number of contributions to this field of research.

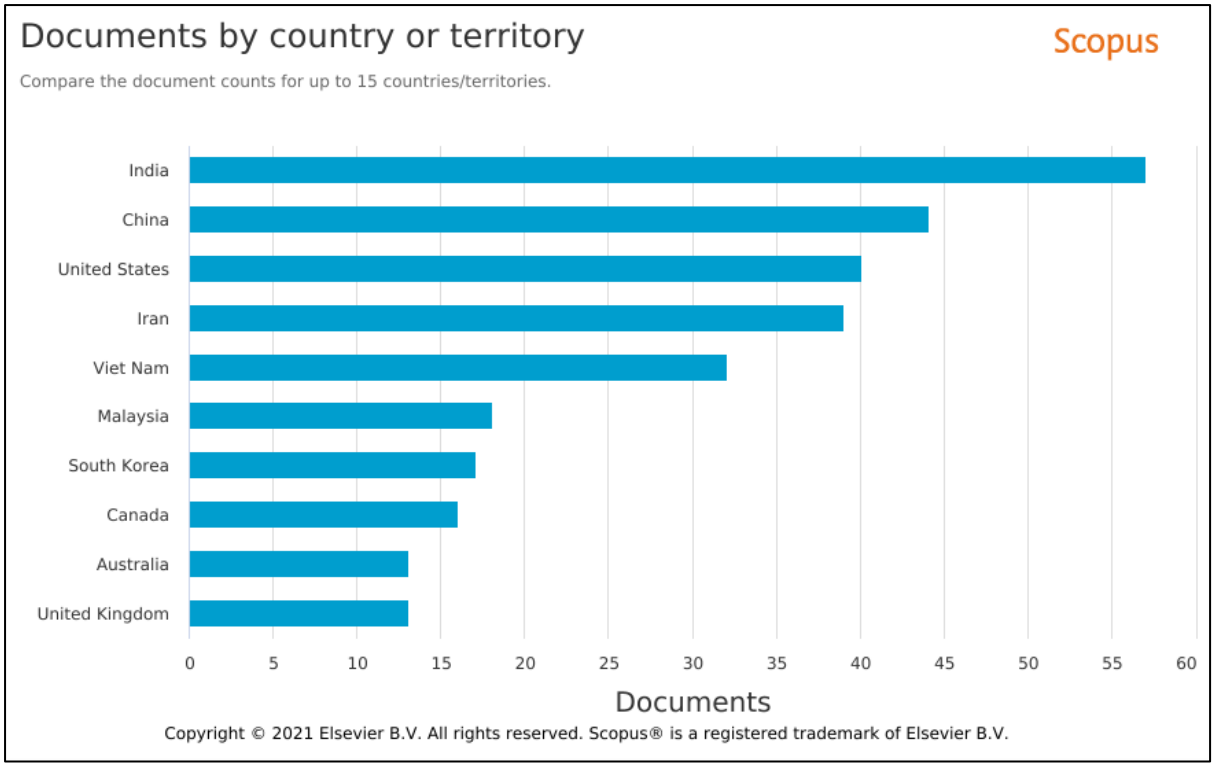

Figure 3: Documents by Country

Source: https://www.scopus.com(accessed on 2nd May 2021)

4.2.Network Analysis

Network analysis is used for graphical representation of the relationship between various computable attributes. Various tools available in market for the same purpose both paid and free. In this bibliometric paper we have used VOS viewer tool to for making the network analysis graphs.VOS viewer is a free app that can be downloaded. The tool's main aim is to evaluate the bibliometric network based on the parameters. VOS viewer is given the.csv extension file from Scopus as input. Based on network, overlay, and density, three forms of visualisation analysis are performed.

Figure 4 shows a network visualisation chart based on a combination of Scopus keywords and source names. The keywords used in the source titles of extracted documents are represented by circles on the diagram. The larger the circle, the more often the keyword is encountered. The distance between two keywords is represented by the links between the circles.The greater the association between the keywords, the smaller the relation size. Keywords with the same colour describe groups of keywords that are closely related. The diagram is divided into eight clusters, each of which is represented by a different colour.

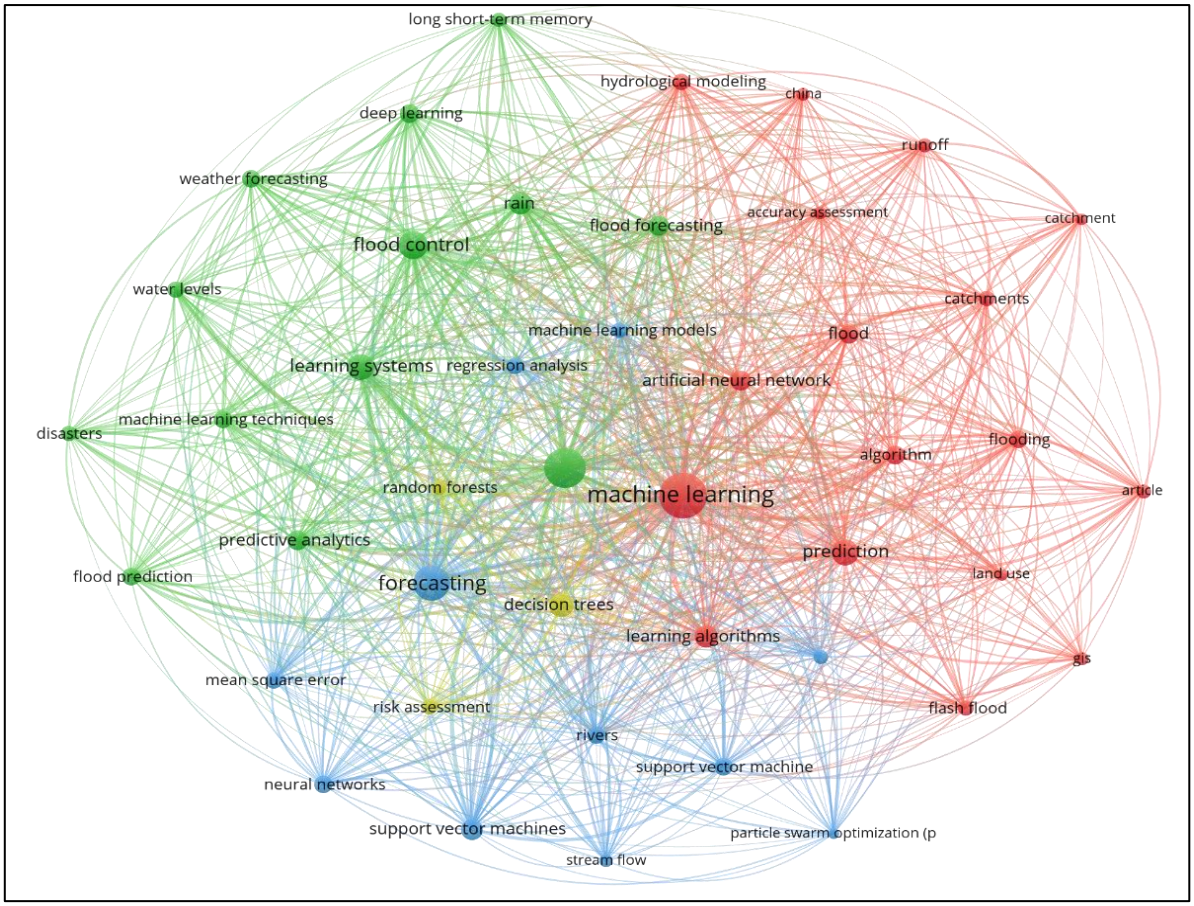

Figure 4: Network Visualization based on Keywords and Source Title

Data Source: https://www.scopus.com (accessed on 2nd May 2021)

Tool Used: https://www.vosviewer.com/

Figure 5 shows a network visualization chart as a combination of author's names and their coappearance as authors on different research that are published.Each circle in the figure represents the name of an author. Bigger circles like Costache R. means they have higher number of contributions.The aim of this visualization map is to show collaboration between the authors. Here, the link between the different author circles show collaboration between the respective authors.

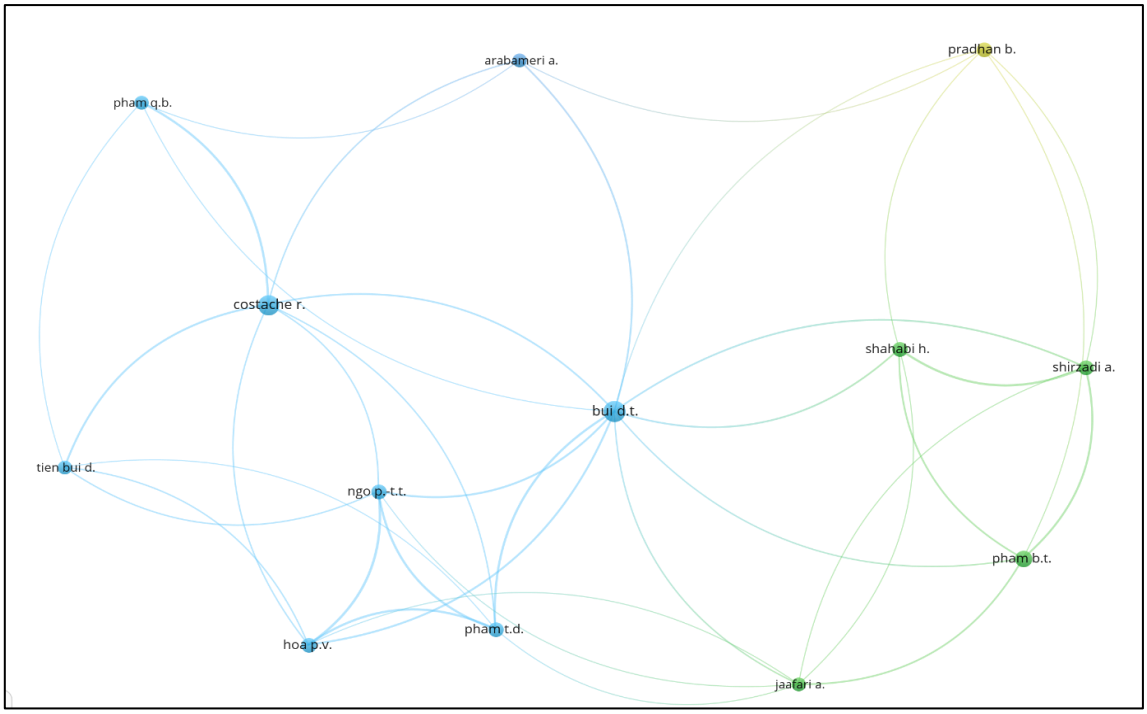

Figure 5: Network Visualization based on Coappearance as authors

Data Source: https://www.scopus.com (accessed on 2nd May 2021)

Tool Used: https://www.vosviewer.com/

Figure 6, below consists of a network visualization chartof the country collaboration. The software used for this diagram is Gephi which is a free software. The Fruchterman-Reingold layout is used to plot this diagram. The intensity of a concerted effort between two nations is measured by the breadth of their relation.The quality of the circle around the nation determines how commited the nation is which means the bigger the circlemore is the commitment of the nation towards the research.

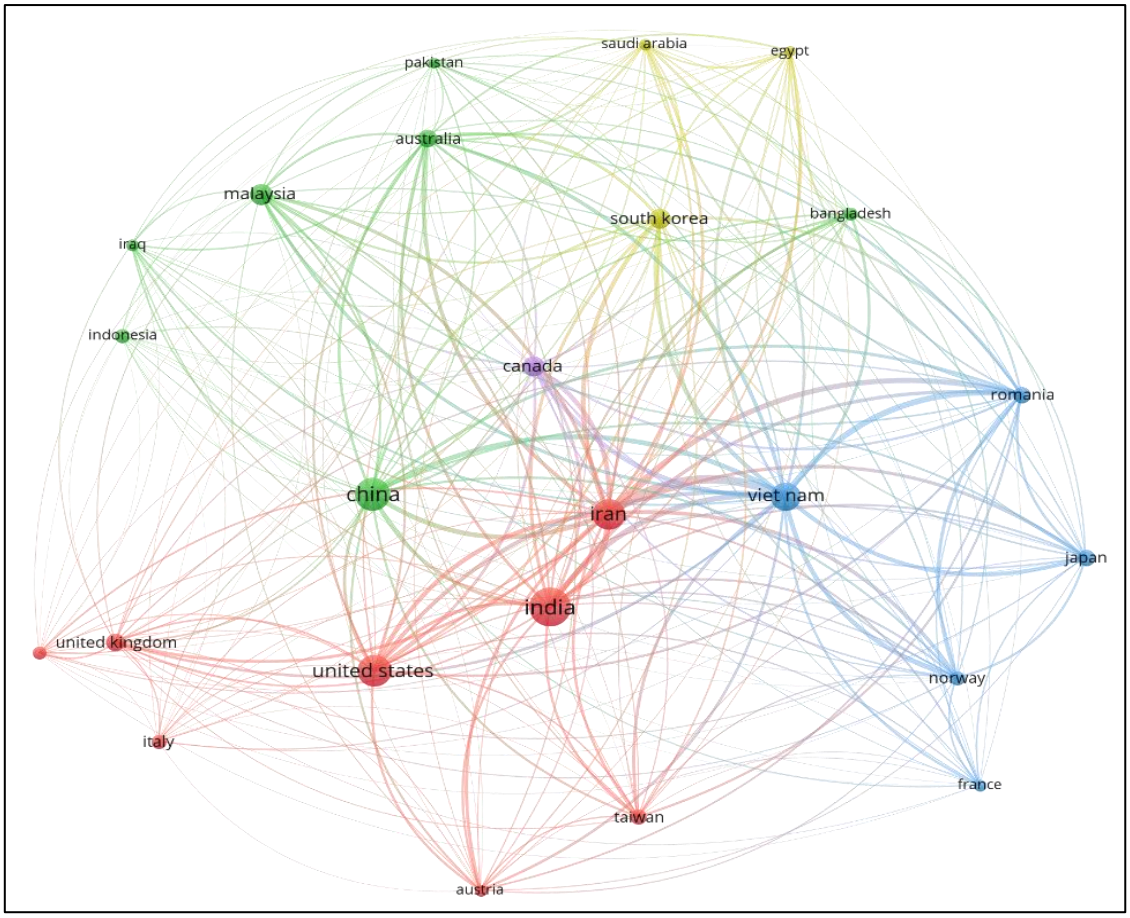

Figure 6: Network Visualization based on Country

Data Source: https://www.scopus.com (accessed on 2nd May 2021)

Tool Used: https://www.vosviewer.com/

4.3. Affiliation measurements-based Statistical Analysis

Graph below stipulates the top ten bestowed organizational affiliations. Maximum number of affiliations is from Duy Tan University, Vietnam with 28 affiliated documents and Ton-Duc-Thang University, Vietnam with 17 affiliated documents followed by Universitatea din Bucur Esti, Romania with 12 affiliations and National Institute of Hydrology and Tarbiat Modares University have 10 affiliated documents each.

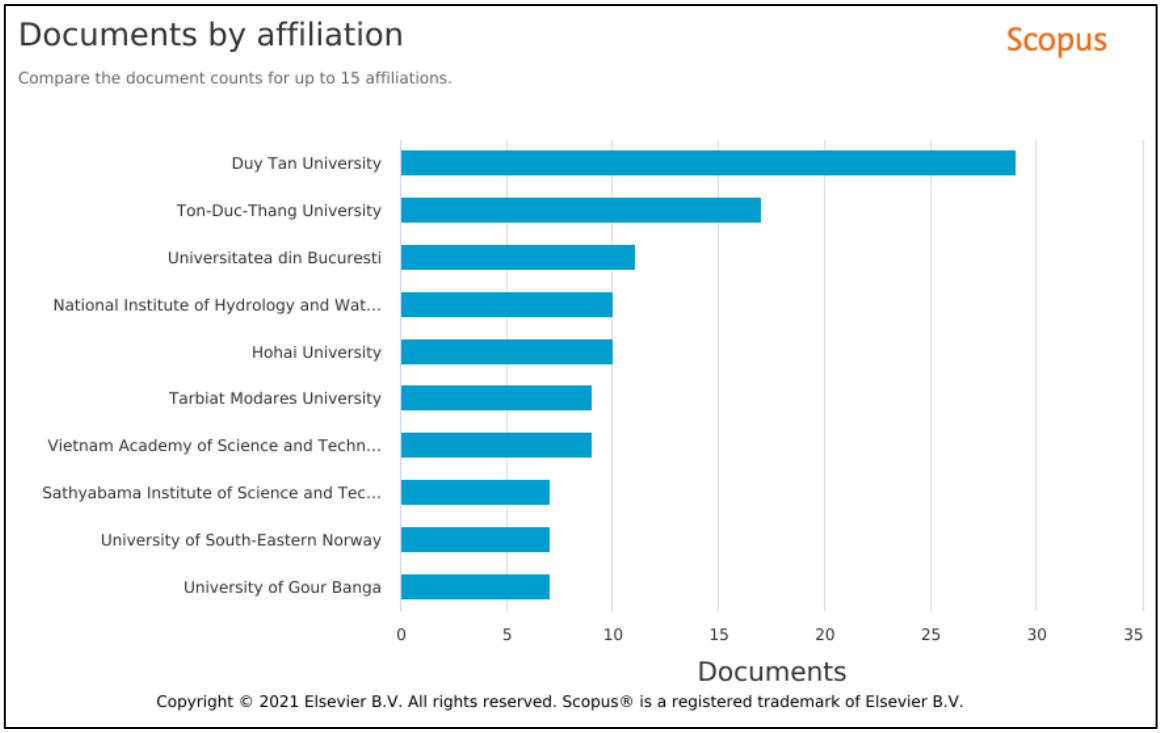

Figure 7: Documents byAffiliation

Source: https://www.scopus.com(accessed on 2nd May 2021)

4.4.Author Contributions based Statistical Analysis

Figure below depicts the top 10 authors who have given their contribution in flood predictions using machine learning. It is clearly visible that Costache, R., Bui, D.T., and Ngo, P.T.T have given the maximum number of contributions towards the field of flood prediction.

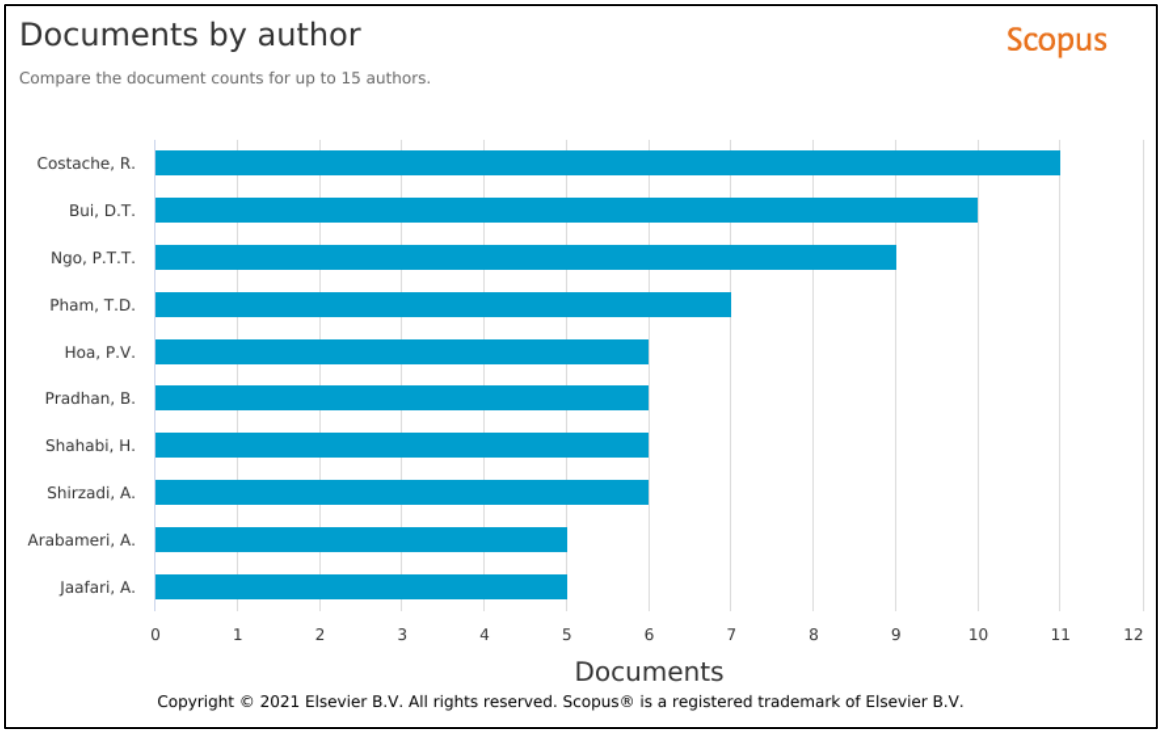

Figure 8: Documents byAuthor

Source: https://www.scopus.com(accessed on 2nd May 2021)

4.5.Subject Areas based Statistical Analysis

The following pie chart shows the subject area wise segregation of the work done adjacent with our search of flood prediction over the span of last two years. One can easily interpret from the graph that most of the research (i.e., 22.6%) is carried out by the Environmental Sciences department which is then followed by the Computer Science Department (16.5%). A significant amount of research is also carried in areas of Engineering and Earth and Planetary Science. The Physics and Astrophysics department is the one which has least explored the field of flood prediction.

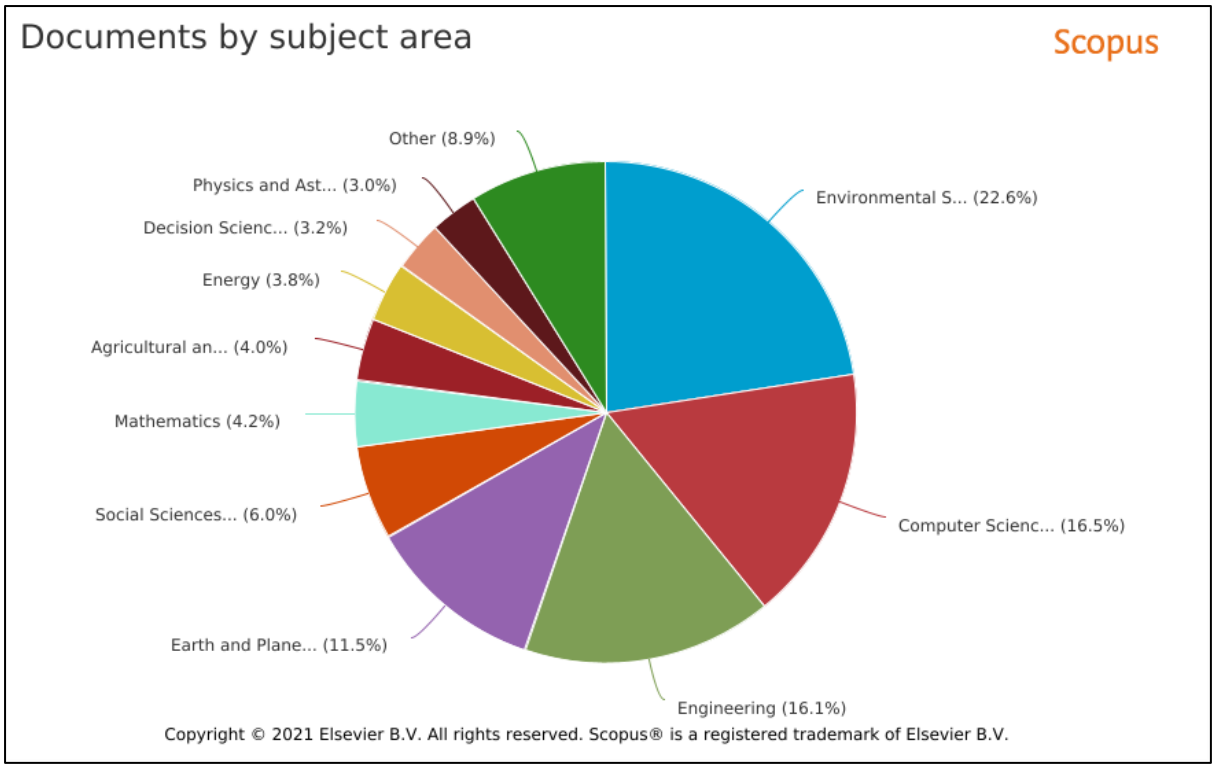

Figure 9: Documents bySubject Area

Source: https://www.scopus.com(accessed on 2ndMay 2021)

4.6.Source Types based Statistical Analysis

The following chart shows the distribution of research related to the keyword's “flood”, “prediction” and “machine learning” by the type of research documents published in the last couple of years. It is intelligible that many of these documents are research articles followed by conference papers and notes are the least explored among all the research options.

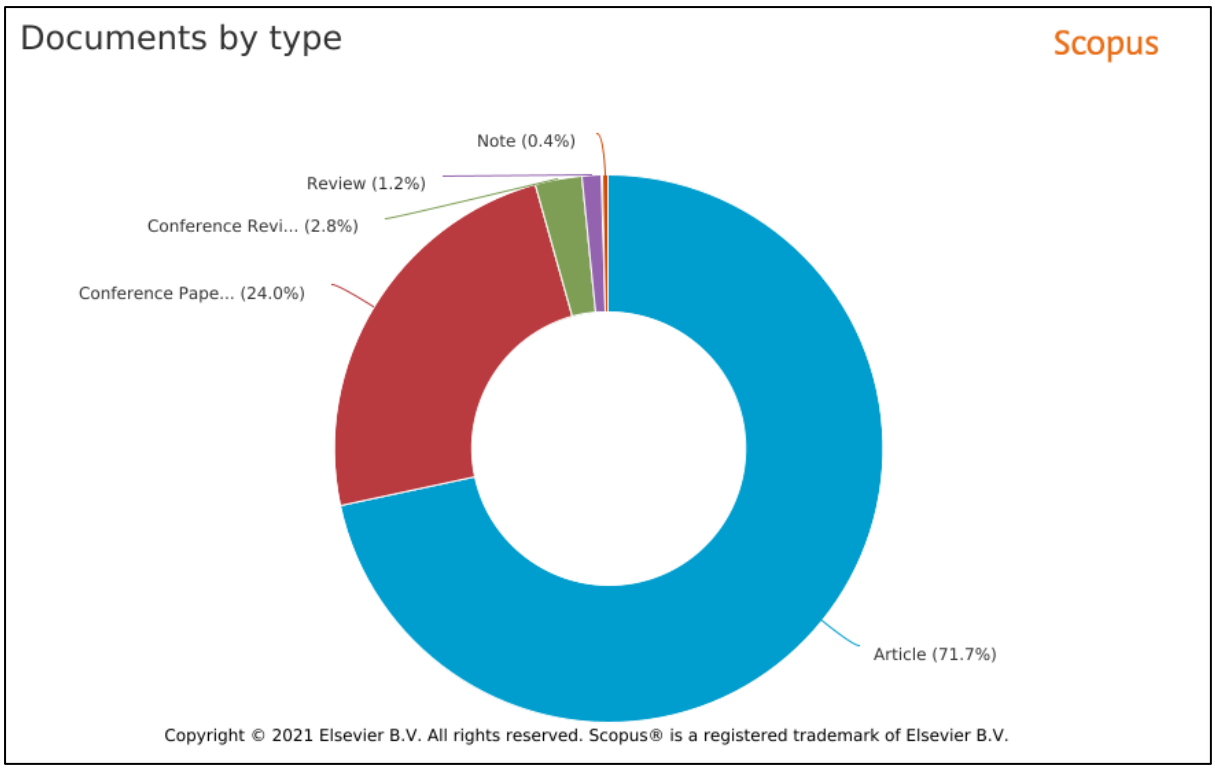

Figure 10: Documents by Type

Source: https://www.scopus.com(accessed on 2nd May 2021)

4.7.Citations based Statistical Analysis

From the graph below we can see the citations received through these documents that are drawn in study of flood predictions using machine learning. As it is legible by the graph the year 2020 showed the highest number of citations in the past two years while the number of citations were the lowest in the year 2018.

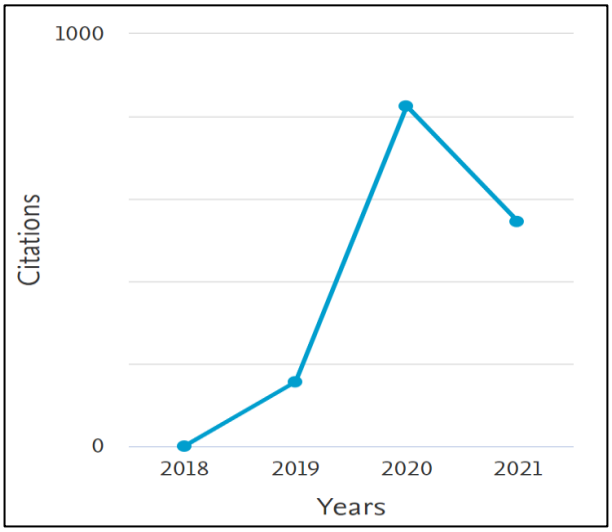

Figure 11: Citations

Source: https://www.scopus.com(accessed on 2nd May 2021)

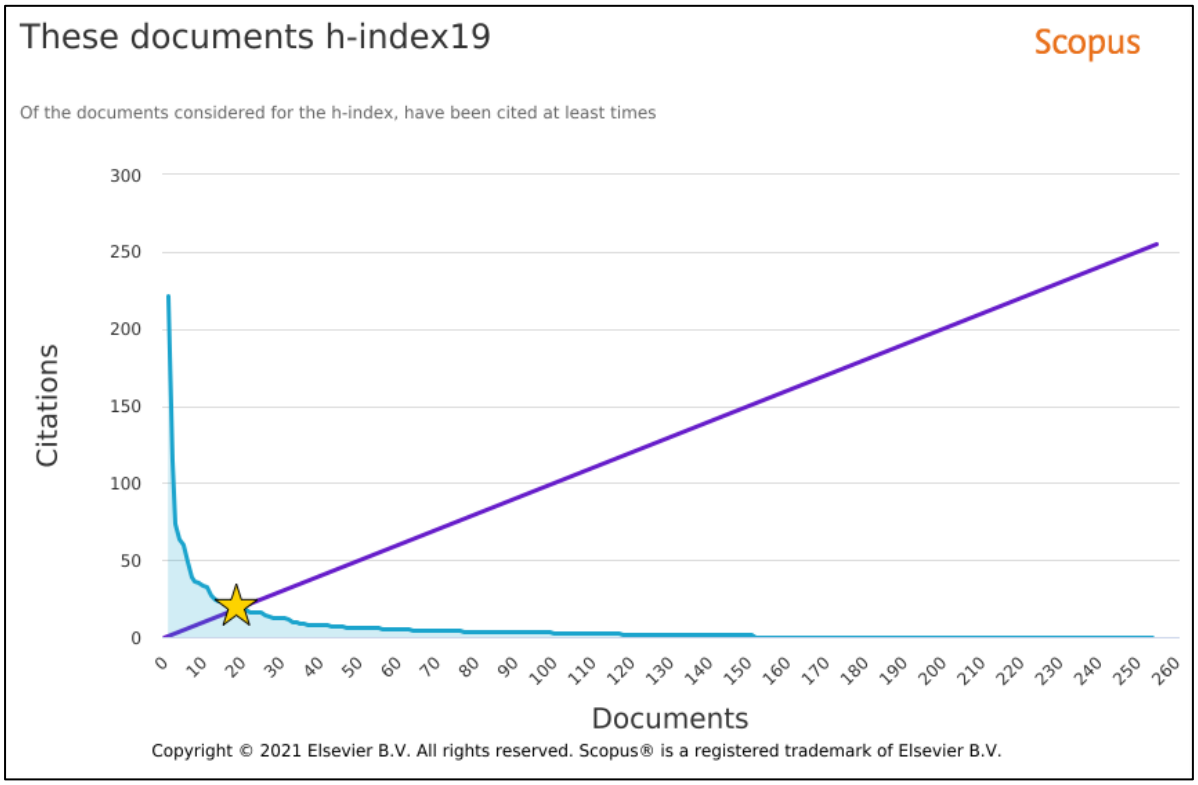

Figure 12: H-Index 19

Source: https://www.scopus.com(accessed on 2nd May 2021)

5. Research Ramifications of Study

Flooding poses enormous challenges to all rain surplus nations across the globe causing huge loss to human life and property. Different organisations in the world are constantly working on efficient flood prediction techniques using evolving technologies. By analysing the various meteorological parameters flood prediction could be made more accurate hence averting major natural disaster and reducing its impact to a great extent. The main keywords used in bibliometric analysis are of flood prediction, model used in prediction, analysis of various parameters. This shows that there is a lot of work on making this model more accurate and precise. This would help in better prediction of floods.

The main keywords used in the bibliometric analysis of flood prediction using machine learning are machine learning, floods, forecasting and flood control. A lot of research and analysis have been done by researchers to improve and avoid such severe event.China is one of the rising economies and is also the one which got affected by the floods much more.

6. Confining the Recent Study

The current investigation has some limits. The current research carried out on the Scopus database using certain keywords for doing an informative analysis. Due to only utilizing Scopus Directory some important publications such as Web of Science and Google Scholar, which contain scientific documents were not included for the current research study. The

search term "Floods", “Prediction” and “Machine Learning” was used to conduct the survey due to which some other important publications and articles may be missed and weren’t incorporated. The literature studies were also confined to English language. The articles and documents from the publication year 2019 to 2021 were considered for analysing the results.

7. Conclusion

The purpose of the bibliometric research is to bring out useful and necessary information about a particular topic in a research field. We studied literature on impact and prediction of floods. This will help for successful research for finding effective algorithms and parameters required for prediction with greater accuracy. The greatest numbers of documents were published by the country India followed by China and United States, respectively. Most contributing subject area was Environmental Science followed by Computer Science then Engineering. Recent drastic changes in climate have increased the number of floods by a lot than in the last 10-15 years. Therefore, there is lot of work being done in this area hence we felt there was a need for a bibliometric study about this topic.

8. References

1. Anaraki, M.V., Farzin, S., Mousavi, SF. et al. Uncertainty Analysis of Climate Change Impacts on Flood Frequency by Using Hybrid Machine Learning Methods. Water Resource Manage 35, 199–223 (2021). https://doi.org/10.1007/s11269-020-02719-w
2. Suresh Sankaranarayanan, Malavika Prabhakar, Sreesta Satish, Prerna Jain, Anjali Ramprasad, Aiswarya Krishnan; Flood prediction based on weather parameters using deep learning. *Journal of Water and Climate Change* 1 December 2020; 11 (4): 1766–1783. https://doi.org/10.2166/wcc.2019.321.
3. Pankaj Singha, Priyanka Das, Swapan Talukdar, Swades Pal, “Modeling livelihood vulnerability in erosion and flooding induced river island in Ganges riparian corridor, India” in Ecological Indicators, Volume 119, 2020,https://doi.org/10.1016/j.ecolind.2020.106825
4. Mohammadtaghi Avand, Hamidreza Moradi, Mehdi Ramazanzadeh lasboyee, using machine learning models, remote sensing, and GIS to investigate the effects of changing climates and land uses on flood probability, Journal of Hydrology, 2020, 125663, ISSN 0022-1694, https://doi.org/10.1016/j.jhydrol.2020.125663.
5. Saiesh Naik, Srushti Ashok Patil, Anang Verma, Anil Hingmire, 2021, Flood Prediction using Logistic Regression for Kerala State, INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT) NTASU – 2020 (Volume 09 – Issue 03),
6. Nadia Zehra, Department of Computer Sciences, Allama Iqbal Open University, Islamabad, Pakistan Prediction Analysis of Floods Using Machine Learning Algorithms (NARX & SVM) 2020-01-30. Page: -24-34
7. Mahyat Shafapour Tehrany, Simon Jones, Farzin Shabani, Identifying the essential flood conditioning factors for flood prone area mapping using machine learning techniques, CATENA, Volume 175, 2019, Pages 174-192, ISSN 0341-8162, https://doi.org/10.1016/j.catena.2018.12.011
8. Michael Bruen, Jianqing Yang, Functional networks in real-time flood forecasting—a novel application, Advances in Water Resources, Volume 28, Issue 9, 2005, Pages 899-909, ISSN 0309-1708, https://doi.org/10.1016/j.advwatres.2005.03.001

9. Khosravi, K., Shahabi, H., Pham, B. T., Adamowski, J., Shirzadi, A., Pradhan, B., . . . Prakash, I. (2019). A comparative assessment of flood susceptibility modeling using multi-criteria decision-making analysis and machine learning methods. Journal of Hydrology, 573, 311-323. doi:10.1016/j.jhydrol.2019.03.073
10. Janizadeh, S., Avand, M., Jaafari, A., Van Phong, T., Bayat, M., Ahmadisharaf, E., . . . Lee, S. (2019). Prediction success of machine learning methods for flash flood susceptibility mapping in the tafresh watershed, iran. Sustainability (Switzerland), 11(19) doi:10.3390/su1119542
11. Dodangeh, E., Choubin, B., Eigdir, A. N., Nabipour, N., Panahi, M., Shamshirband, S., & Mosavi, A. (2020). Integrated machine learning methods with resampling algorithms for flood susceptibility prediction. Science of the Total Environment, 705 doi:10.1016/j.scitotenv.2019.135983
12. Rahman, M., Ningsheng, C., Islam, M. M., Dewan, A., Iqbal, J., Washakh, R. M. A., & Shufeng, T. (2019). Flood susceptibility assessment in bangladesh using machine learning and multi-criteria decision analysis. Earth Systems and Environment, 3(3), 585-601. doi:10.1007/s41748-019-00123-y
13. Costache, R., & Tien Bui, D. (2020). Identification of areas prone to flash-flood phenomena using multiple-criteria decision-making, bivariate statistics, machine learning and their ensembles. Science of the Total Environment, 712 doi:10.1016/j.scitotenv.2019.136492
14. Senent-Aparicio, J., Jimeno-Sáez, P., Bueno-Crespo, A., Pérez-Sánchez, J., & Pulido-Velázquez, D. (2019). Coupling machine-learning techniques with SWAT model for instantaneous peak flow prediction. *Biosystems Engineering, 177*, 67-77. doi:10.1016/j.biosystemseng.2018.04.022
15. Costache, R., Hong, H., & Pham, Q. B. (2020). Comparative assessment of the flash-flood potential within small mountain catchments using bivariate statistics and their novel hybrid integration with machine learning models. Science of the Total Environment, 711 doi:10.1016/j.scitotenv.2019.134514