ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24JUN1770

Enhancing Flood Management Through Machine Learning: A Comprehensive Analysis of the CatBoost Application

Ogundolie O. I. [1,2]; Olabiyisi S. O. [1]; Ganiyu R. A [3], Jeremiah Y. S[1]; Ogundolie F. A.[4]
[1]Department of Computer Science, Ladoke Akintola University of Technology, Ogbomoso, Nigeria
[2]African Regional Center for Space Science and Technology Education in English (ARCSSTE-E), Ile-Ife, Nigeria
[3]Department of Computer Engineering, Ladoke Akintola University of Technology, Ogbomoso, Nigeria
[4]Department of Biotechnology, Faculty of Computing and Applied Sciences, Baze University,Abuja, Nigeria

Abstract:- River flooding is a major natural disaster that has caused enormous damage to our environment, infrastructure and human life. River flooding has led to flooding in river basins which has disrupted human activities and fatalities. This study is a review of river basin flooding, the impact of machine learning techniques in flood prediction in river basins, flood management in the past and the impact of machine learning in flood management. This review further examined how the Categorical boosting algorithm (CatBoost) which is a machine learning technique, could improve flood prediction in river basins and its applications in flood management. Several case studies of how CatBoost models have been used to predict flooding and enhance early warning systems were also reviewed in this study. CatBoost has been recognized to be excellent in working on categorical variables making it efficient in handling datasets with complex relationships. This makes it applicable for flood prediction in river basins considering the factors involved in flooding. CatBoost's effectiveness in flood forecasting and flood susceptibility modelling was demonstrated in some case studies. CatBoost has the potential to change flood management, minimize the disastrous impacts of floods, and enhance sustainable development, regardless of its limits. The review highlights the importance of machine learning to improve flood protection and the need for concerted efforts to get beyond implementation obstacles and take full advantage of CatBoost's flood management capabilities.

Keywords:- *Flooding, CatBoost, Flood Management, Machine Learning.*

I. INTRODUCTION

As significant losses in terms of property, infrastructure, and human life occur all over the world, flooding has been recognized as one of the most damaging natural catastrophes (Canadell *et al*., 2023). Climate change is causing extreme weather events to occur more frequently and with greater intensity, which highlights the urgent need for effective flood management strategies (Canadell *et al*, 2023). River basin flooding is a complex and multifaceted phenomenon that results from a number of factors, including intense precipitation, melting of snow, changing land usage, and variations in the climate (Dierauer *et al*., 2021; Rajkhowa and Sarma, 2021). Flooding can have disastrous effects, including community uprooting, infrastructure destruction, human casualties, and environmental deterioration (Petrucci, 2022)). Effective flood management requires proactive measures to anticipate and mitigate flood risks, as well as robust decision-making processes to respond to flood events in real-time (Molinari *et al.*, 2020). The integration of machine learning algorithms has demonstrated the potential to improve flood management techniques recently (Mosavi *et al.*, 2018).

In recent years, river basin flood control strategies have benefited from the application of machine learning (ML) (Nguyen *et al.*, 2024). The utilization of machine learning algorithms in flood prediction offers several advantages, including the ability to assess complex spatiotemporal data, detect nonlinear relationships, and adapt to changing environmental conditions (Mosavi et al., 2018). By utilizing the power of data analytics, machine learning algorithms can analyze enormous volumes of hydrological data, identify patterns and trends, and generate insights to support informed decision-making (Mosavi *et al.*, 2018). Supervised learning methods such as support vector machines (SVM), random forests, and gradient boosting machines (GBM) have been extensively employed for flood prediction tasks (Mosavi et al., 2018; Tehrany et al., 2015). These algorithms are able to generate predictive models that forecast future flood events based on historical data on rainfall, river flow, soil moisture, and other hydrological variables (Mosavi et al., 2018; Tehrany et al., 2015). This study explores the application of the robust machine learning algorithm Categorical Boosting (CatBoost) in flood management and how it might significantly change current practices.

This in-depth investigation examines the application of CatBoost in river basin management, taking special emphasis on critical domains such as early warning systems, flood prediction, and decision support. This review thoroughly examines the corpus of existing literature and case studies in an effort to provide insights into the potential benefits, challenges, and future directions of using machine learning for flood management in river basins.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24JUN1770

II. RIVER FLOODING

A river is a flowing watercourse that runs into another river, the ocean, or a lake. Usually, fresh water runs through it. The formation of ecosystems, human civilizations, and landscapes depends on rivers. These are vital water sources for drinking, transportation, agriculture, and industry. There are numerous sources of rivers, such as lakes, springs, and even glaciers. They follow the least-obstructed path, sculpting the landscape over time. Rivers dynamically alter their course as they move downstream, taking up water from tributaries, runoff, and rainfall. Many different species find homes in rivers, which support a diverse array of aquatic and terrestrial life. They also provide routes for the migration and dispersal of plants and wildlife. River basins provide several benefits such as water supply.

The land area that a river and its tributaries drain is referred to as a river basin, sometimes called a watershed or catchment region. It includes every surface water and groundwater movement that eventually feeds into a single river or network of rivers, moving from high points like mountains to low points like valleys or coastlines. River basins are naturally occurring hydrological units that are linked networks in which gravity causes water to flow downward. They are essential to the water cycle because they control the availability and distribution of freshwater resources. The size of river basins varies; they can be small, local drainage zones or vast, transboundary regions that cut across several nations. There are some types of river basin such as endorheic, exorheic, ephemeral and perennial River Basins as shown in Figure 1.

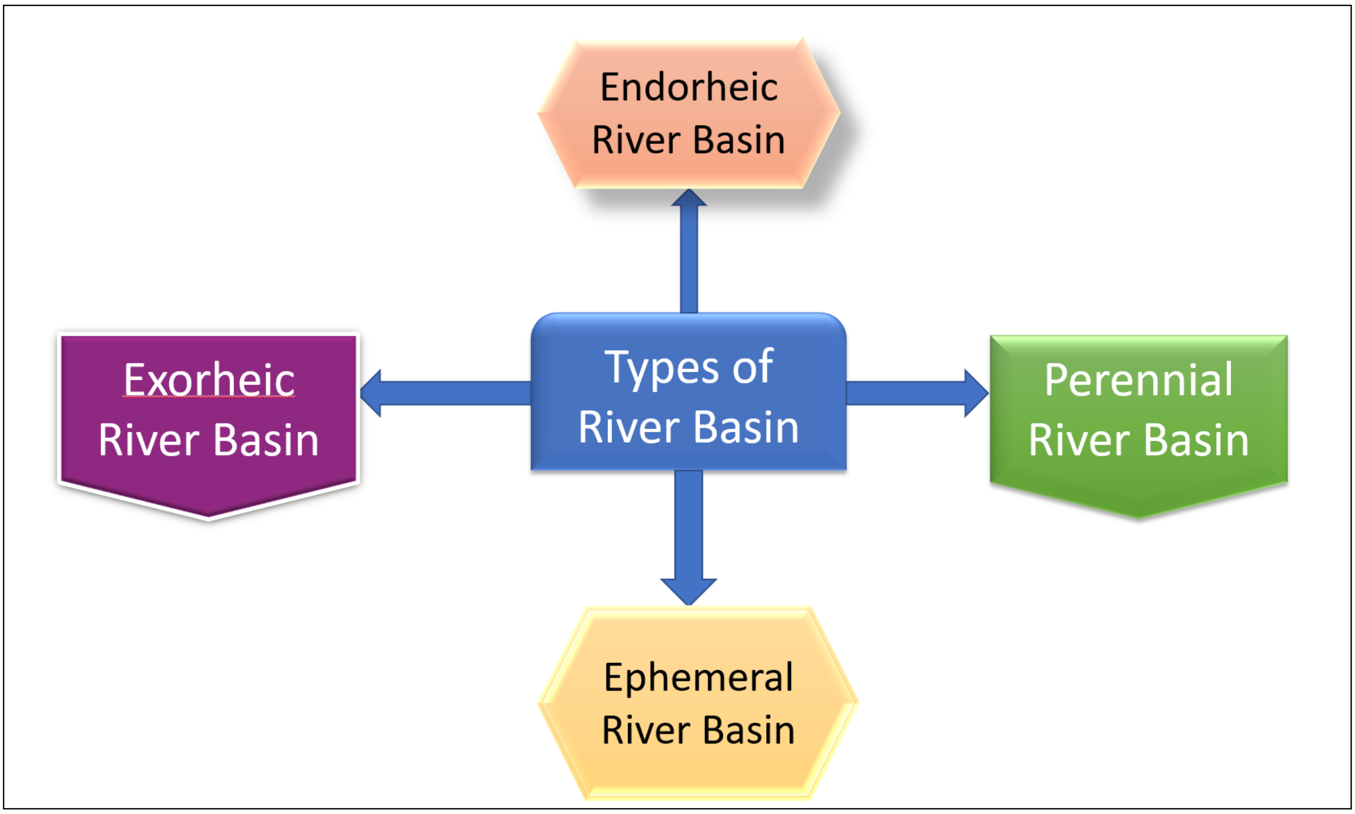

Fig 1: Types of River Basin.

Endorheic basins are closed drainage systems in which water evaporates, and accumulates in internal lakes, or marshes rather than flowing to the sea or ocean. These basins, which might contain salt flats or playas, are frequently found in arid or semi-arid areas. Open drainage systems called exorheic basins are where water eventually empties into the sea or ocean. These river networks eventually merge and discharge into bigger bodies of water, making them the most prevalent sorts of river basins. Ephemeral basins, which are frequently found in arid areas with irregular rainfall, are distinguished by transient water flow. Periods of high runoff and flooding may occur in these basins, followed by protracted dry spells. Regular rainfall, groundwater supplies, or glacial melt provide perennial basins with year-round water flow. These basins are frequently connected to bigger rivers and tributaries and maintain more stable ecosystems.

Fig 2: Benefits of River Basins

Figure 2 shows some of the benefits of river basins such as enabling economic activities, and maintaining diverse ecosystems river basins act as ecological reservoirs, collecting and holding water necessary for agriculture, ecosystem support, and human drinking and sanitation needs. Numerous plant and animal species are supported by the different ecosystems found in river basins, which include riparian zones, wetlands, and aquatic habitats. For wildlife, these areas offer refuge, food sources, and nesting sites. Recreational opportunities in river basins include boating, fishing, hiking, and wildlife observation. They draw visitors to beautiful scenery, national parks, and locations related to river valleys and waterfalls that are part of cultural heritage. River basins are crucial for managing floodwaters because they absorb surplus rainfall, reduce runoff, and lessen the chance of flooding downstream. Wetlands and floodplains act as natural barriers within river basins, reducing the impact of flooding. River basins facilitate various economic activities such as agriculture, fishing, transportation, and hydropower generation. They provide fertile soils for farming, navigable waterways for trade, and renewable energy resources through hydroelectric dams. Flooding in river basins is a frequent natural event that has important socioeconomic and environmental ramifications. It is essential to comprehend the causes and effects of flooding in river basins in order to implement effective flood management and catastrophe risk reduction strategies (Saber *et al.*, 2023; Jia *et al.*, 2022). One of the primary causes of flooding in river basins is heavy rainfall, which increases river discharge and water levels (Merz et al., 2021). High precipitation events in river basins have the potential to quickly submerge low-lying areas due to runoff (Wu et al., 2023). Furthermore, during the warmer months in some parts of the globe, river flow increases due to melting from highland areas, raising the possibility of floods (Zeleňáková et al., 2015). Usually, flooding surpasses the ability of man-made or natural drainage systems to hold and redirect water (Glago et al., 2021). Flash floods and quick runoff can be caused by prolonged or severe rainfall, particularly in places with impermeable surfaces like metropolitan areas (Zhao et al., 2020; Prokešová et al., 2022). A region's vulnerability to floods is further impacted by variables like terrain, soil composition, changes to land use, and the unpredictable nature of the climate (Roy *et al*., 2020). Increased floods in river basins are also largely caused by human activities such as urbanization, deforestation, and changes in land use (Handayani et al., 2020; Chakraborty and Chakraborty, 2021). Deforestation increases sediment loads in rivers and reduces water infiltration because it increases soil erosion and reduces forests' capacity to absorb rainfall (Nasirzadehdizaji and Akyüz 2022).

Flooding in river basins affects human populations and ecosystems in a major way (Merz et al., 2021). Floods have the ability to destroy infrastructure, including homes, roads, and farms, as well as result in fatalities and population displacement (Chandrathilake, 2022; Qian and Eslamian 2022). A variety of socioeconomic effects are associated with flooding, including impaired transportation networks, employment losses, and increased risks of waterborne infections (Jonkman, 2005). Flooding can lead to habitat changes, biodiversity loss, and ecological destruction (Sedighkia et al., 2023). Furthermore, flooding can have severe and long-lasting financial repercussions, including neglected productivity, cleanup and recovery expenses, and property damage (Tanoue *et al*., 2020)

Furthermore, according to Hooper and Lloyd (2011), flooding in river basins can negatively impact aquatic habitats, biodiversity, and water quality. Floodwaters can also carry pollutants like pesticides, fertilizers, and silt, endangering aquatic life and contaminating water supplies (Mushtaq *et al.,* 2020).

III. FLOOD MANAGEMENT

The process of mitigating flood disasters by regulating floods' natural state through artificial techniques is known as flood control. Flood control was implemented by humanity after they realized that floods were inevitable but controllable. However, because of the increased likelihood of floods brought on by climate change, flood management measures must also change. However, flood tragedies persisted even after a number of flood control measures were put into place, and people started to recognize the limitations of these initiatives. Furthermore, it is challenging to increase the standard for flood control projects without taking cost-effectiveness into account (Abdi-Dehkordi *et al.*, 2021). The phrase "flood management" originated as a result of these realizations and refers to the ability to live with flooding, reduce its damages, and sometimes even benefit from it (Wang *et al.,* 2022).

In order to minimize the consequences of floods in river basins, integrated flood control strategies are essential (Xia and Chen, 2021). Flood management strategies usually aim to prevent, lessen, or eliminate effects and activities before flood occurrence. In order to mitigate the detrimental impacts of a flood event, flood control techniques include both structural and non-structural solutions. Implementing both structural and non-structural interventions—like community-based adaptation plans, early warning systems, ecosystem restoration, and floodplain zoning—is essential (Shrivastava *et al.* 2020). Structural measures like floodwalls, reservoirs, and levees can help manage river flow and lessen the risk of flooding in densely populated regions (Hooper and Lloyd, 2011). Non-structural solutions focus on land-use planning, watershed management, and public awareness campaigns to increase community resilience and reduce vulnerability to flood hazards (Ansari *et al.,* 2022)

Structural interventions are more expensive to implement than non-structural ones. Long-term structural measure maintenance can be highly costly and lead to large losses if done improperly or insufficiently (Wang *et al.*, 2022). In addition, there may be more ecological effects. Non-structural measures are more extensive and have fewer adverse impacts than structural ones, but they are also less costly and more sustainable. Society has countless years of expertise in water management in an attempt to minimize the impact of flood disasters on human life and property. Through the use of both non-engineering and engineering solutions to keep people and floodwaters separate, societies have steadily raised the standards for flood management.

However, despite all efforts, the economic losses brought on by flood disasters have not decreased; as a result, one of the major subjects in decreasing the damage caused by flood disasters is figuring out the best mix of engineering and non-engineering approaches. However, from the perspective of disasters reduction, the probability of natural disasters occurring is hardly impacted by human intervention. However, by lowering the vulnerability of disaster victims, reducing the amount of property exposed in flooded areas, and enhancing disaster prevention and mitigation capacities, humans may minimize the losses caused by natural disasters. As a result, risk-based flood control strategies have replaced flood management systems focused on both structural and non-structural measures.

IV. MACHINE LEARNING IN FLOOD MANAGEMENT

Artificial intelligence (AI) has a branch called machine learning (ML) that can automatically and intuitively identify patterns in a dataset without the need for explicit programming. Less computing is required, training, validation, testing, and assessment processes are completed more quickly, the model performs better than physical models, and there is a noticeable reduction in complexity when applying complex real-world scenarios (Mosavi *et al.*, 2018; Wagenaar *et al.,* 2020).

By analyzing hydrological data and identifying patterns and trends, machine learning techniques such as support vector machines (SVM), random forests, and gradient boosting machines (GBM) can forecast floods. On the basis of historical rainfall, river flow, and soil moisture data, models for future flood events are developed using supervised learning techniques. Additionally, machine learning algorithms enable risk assessment, early warning systems, and decision support systems by integrating several data sources and streamlining the decision-making process.

Conventional flood mapping techniques use optical and radar satellite sensors and detect floods using band thresholding and normalized differencing algorithms. MODIS provides daily global water detection using SWIR or NIR spectra. These methods are also used by medium-resolution sensors like Sentinel-2 and Landsat, although misclassifications result from their low near- and mid-infrared reflectance values. SAR sensors, like Sentinel-1, can detect floods through clouds by identifying water with lower backscatter values. However, these methods often rely on user-defined thresholds, which may lead to an overestimation or underestimation of flooded areas. Physically-based models can forecast short-term floods, but only to the extent that substantial hydrological measurements and computer power are available. Their dependence on hydrological knowledge and susceptibility to systematic errors further undermines their integrity. Thus, while physically based models can predict different types of flooding, their limitations emphasize the need for further advancements in flood detection methods.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24JUN1770

Machine learning (ML) offers several benefits in flood management, notable among them the capacity to analyze massive volumes of data and identify nonlinear relationships. More accurate forecasts and proactive mitigation techniques are made possible by this capacity. Additionally, by enabling stakeholders to understand the factors driving flood intensity and vulnerability, the interpretability offered by ML algorithms aids in their ability to make better decisions. Furthermore, machine learning-based techniques can improve over time by adjusting to changing environmental variables through iterative learning processes.

V. INTRODUCTION TO CATEGORICAL BOOSTING ALGORITHM (CATBOOST)

Gradient Boosting Decision Trees (GBDT) is a tool used in the Machine Learning (ML) algorithm known as Categorical Boosting (CatBoost). It can handle categorical features efficiently and benefits from handling them during training instead of preprocessing time (Dorogush et al. 2018). CatBoost is a popular algorithm for predicting, recommendation-making, and ranking tasks (Peretz, 2018). It is broad and applicable to many different contexts and issues. Compared to other GBDT methods, the CatBoost algorithm works better with the default parameters; but, when certain crucial parameters are adjusted, the algorithm performs much better (Peretz, 2018). The training and optimization times of the CatBoost algorithm are among its drawbacks. Instead of dealing with categorical characteristics during preprocessing time, the CatBoost method handles them during training. The dataset can be randomly permuted, allowing the entire dataset to be used for training. This is achieved by calculating the average label value for the example with the same category value placed before the supplied one in the permutation (Xu *et al*., 2023). Each of the category features can be joined to create a new one. The CatBoost algorithm uses a greedy approach to evaluate the combinations while creating a new split for a tree. For the second and following splits in the tree, it will combine all combinations present with all categorical characteristics in the dataset rather than combining for the first split (Zhong *et al.,* 2023). Every split that is listed in the tree is regarded as a category with two values that are combined (Huang *et al.,* 2019). The CatBoost algorithm demonstrates unbiased boosting with categorical features. it has two modes for choosing the tree structure, Ordered and Plain. Plain mode corresponds to a combination of the standard GBDT algorithm with an ordered Target Statistic. Prokhorenkova et al. (2018) used theoretical analysis to create an ordered boosting method that addresses gradient bias. The training data often undergo random variations as a result of the CatBoost algorithm. Consequently, by picking a random permutation and determining gradients based on it, several permutations can be employed to increase the algorithm's robustness.

VI. APPLICATION OF CATBOOST IN FLOOD MANAGEMENT

CatBoost can be used to manage floods in a number of important areas, such as data collection and preprocessing, real-time flood monitoring, predictive modeling for flood forecasting, and decision support systems (Seydi et al., 2022b). Using historical data on weather patterns, water levels, and geographic factors, CatBoost can be utilized to produce accurate flood event forecasts (Seydi et al., 2022a). This facilitates the early implementation of mitigation measures and evacuation of communities that are vulnerable to flooding by the authorities. Because of its ability to handle categorical variables, it is also well-suited for integrating different data sources and optimizing decision-making processes (Kulkarni 2022).

According to Hammami et al. (2019), CatBoost models have the ability to precisely and accurately geographically reference point-based data with high spatial and temporal precision, which will open up a lot of applications. Geographic coordinates are incorporated into the input qualities by the models, which makes this possible. High-quality benchmark flood simulation and forecast data is readily available to stay up with the latest developments in nowcasting systems. It can be used in a calibration and verification process by comparison with real-time data (Tounsi, 2023). Flood hazard maps, which display the areas subject to various levels of flood danger, can be made using point-based data (Ajibade *et al.*, 2021). Finding and gathering the pertinent data will be the initial step in this procedure, and it usually doesn't cost much to do. New flood hazard maps that are created using this added value data can be used in future flood risk management plans and more informed decision-making processes (Maskrey *et al*., 2022).

A wide range of immediate information regarding flood events, including water level information, is available attributable to the real-time data that is currently available from various sensors, electronic devices, web-based systems, and social media (Van Ackere *et al.*, 2019). The utilization of this data stream, which is increasing in volume and visibility, could yield a variety of novel forecasting and nowcasting applications for customers (Boone *et al*., 2019). Nevertheless, little research has been done so far to integrate this heterogeneous data into an organized, real-time data processing process (Yao *et al*., 2023).

Decision support systems (DSS) can be used to predict floods by leveraging CatBoost's ability to take a range of data types and capture complex correlations between variables (Xiang 2022). CatBoost can provide accurate predictive models that assist proactive flood mitigation strategies and early warning systems by analyzing meteorological, geographical, and hydrological factors in combination with historical flood data (Al-Kindi and Alabri, 2024). Furthermore, the integration of CatBoost into the present flood control systems facilitates enhanced risk assessment and real-time decision-making (Saber et al., 2023). CatBoost models can be integrated into decision support platforms to provide stakeholders with quick and

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24JUN1770

accurate assessments of the risk of flooding. This makes the execution of mitigation measures and resource allocation more effective.

Moreover, the interpretability features of CatBoost might facilitate the dissemination of information and stakeholder participation in flood management protocols. With CatBoost models, decision-makers can better understand the factors influencing flood vulnerability and severity, enabling them to develop targeted interventions and flexible policies. As a result, the general resistance to flood disasters will rise. The use of CatBoost in flood management has the potential to significantly enhance decision support systems and integrate with existing frameworks, leading to improved flood response, prediction, and mitigation in years to come.

VII. CASE STUDIES

The effectiveness of CatBoost in flood prevention has been demonstrated by numerous case studies. In areas where the majority of the land is vulnerable to flooding, researchers have developed CatBoost models to assess river water levels and forecast potential flood threats with extreme accuracy. In a similar vein, CatBoost has been applied to assess complex hydrological data and improve early warning systems, reducing the vulnerability of communities to catastrophic flood events. For instance, with an AUC of 79%, Catboost was used to map the flood vulnerability in Kerala, India's Idukki area (Saravanan et al., 2023). Van Phong et al. (2023) used CatBoost to estimate and map the flood vulnerability of the Que Son district in Quang Nam province, Vietnam. The geospatial database was created using 96 flood and non-flood locations as well as a set of 10 conditioning factors. According to Van Phong et al. (2023), CatBoost performed admirably in this study's flood susceptibility modeling, with an AUC of 0.94 for testing and 0.96 for training datasets. The generated flood susceptibility map, where the majority of historical flood pixels were situated in high and very high susceptibility classes, demonstrated the model's ability to predict flood susceptibility with accuracy. CatBoost effectively identified the study area's most flood-prone places by taking into account a variety of parameters, including terrain, precipitation patterns, and land cover. This demonstrated the utility of CatBoost as a tool for assessing flood susceptibility and developing mitigation plans. (Van Phong *et al.*, 2023)

In addition, Seydi et al. (2022) created a model and evaluated its overall accuracy (OA) in the Gorganrud basin in Iran against other boosting algorithms such as XGBoost, CatBoost, and LightGBM. The model's OA of 92.40% was more than the other models. When compared to LightGBM, CatBoost performed slightly more effectively at identifying flooded areas but was not as effective at detecting non-flooded areas. The developed model and CatBoost models were highlighted as the most efficient with AUC values of 0.954 and 0.959, respectively, in the Gorganrud basin, surpassing other models (Seydi *et al*., 2022).

Another study applied a light gradient boosting machine (LightGBM) and categorical boosting (CatBoost), to predict flash flood susceptibility (FFS) in the Wadi System Hurghada, Egypt (Saber *et al.,* 2022). Fourteen flood-controlling factors were selected and evaluated for their relative importance in flood occurrence prediction. The performance of the two models was assessed using various indexes in comparison to the common random forest (RF) method. The results show areas under the receiver operating characteristic curves (AUROC) of above 97% for all models and that LightGBM outperforms other models in terms of classification metrics and processing time (Saber *et al*., 2022). The developed FFS maps demonstrate that highly populated areas are the most susceptible to flash floods. This particular study proved that the employed algorithms (LightGBM and CatBoost) can be efficiently used for FFS mapping (Saber *et al*., 2022).

Another research tested CatBoost, LightGBM and XGBoost for daily streamflow forecasting in the mountainous Skawa River catchment, Poland. CatBoost provided the best results among the three models (Szczepanek, 2022). The XGBoost did not turn out to be the best model for the daily flow forecast, although it is the most used model. Assuming the use of models with their default parameters, the best results were obtained with CatBoost (Szczepanek, 2022). By optimizing the hyperparameters, the best forecast results were obtained by LightGBM. The gradient boosting algorithms provide a good streamflow prediction in mountainous rivers. All tested models achieved Nash-Sutcliffe model efficiency (NSE) in the range of 0.85–0.89 and RMSE in the range of 6.8–7.8 m. To obtain an NSE above 0.8, the recommended period of training data should be not less than 12 years. The differences in model results were smaller than the differences within the models themselves when suboptimal hyperparameters were used, emphasizing the importance of proper tuning for model performance (Szczepanek, 2022).

Predicting rainfall can apply to predicting flood events, Kumar *et al.*, 2023 carried out a performance evaluation on some machine learning models such as CatBoost, XGBoost, Lasso, Ridge, Linear Regression, and LGBM for predicting rainfall in urban metropolitan Cities. CatBoost was identified in the research as an effective model for predicting rainfall in urban metropolitan areas. It demonstrated the highest accuracy with the fewest errors during the training, validation, and testing phases, outperforming the other models. Daily rainfall data has distinct temporal patterns that CatBoost effectively captures, exhibiting good accuracy with low MAE, RMSE, and high R2 scores. CatBoost and XGBoost outperformed traditional linear regression-based techniques, continuously sustaining low prediction errors (Kumar *et al*., 2023).

VIII. EARLY WARNING SYSTEMS

CatBoost is an effective tool for analyzing a wide range of data sources in early warning systems, including historical river levels, meteorological data, soil moisture levels, land use patterns, and geographic features. These large datasets enable CatBoost to identify complex relationships and patterns that result in river floods, improving the accuracy and timeliness of predictions (Kumar et al., 2023a). CatBoost's ability to handle outliers and missing data, which is a major constraint in flood prediction, is one of its key advantages (Kumar et al., 2023b). Its gradient-boosting structure allows it to learn from errors and progressively improve prediction accuracy iteratively. Moreover, CatBoost's interpretability features facilitate stakeholders' understanding of the factors influencing flood risk, hence fostering informed decision-making and adaptable response strategies. By providing insights into the relative relevance of different variables, CatBoost facilitates the effective allocation of resources and the prioritization of mitigation actions.

IX. CHALLENGES AND LIMITATIONS

Notwithstanding CatBoost's obvious benefits, there are certain challenges associated with its use in flood control. Problems with data availability and quality, interpretability of models, processing resources, and ethical considerations severely limit the widespread usage of models. Furthermore, the implicit nature of machine learning algorithms like CatBoost raises concerns about decision-making transparency and accountability, necessitating a thorough assessment and validation of results. Using machine learning algorithms for flood management raises additional ethical questions of algorithmic bias, data privacy, and responsibility. Stakeholders need to address these ethical issues in order to ensure that machine learning-based solutions benefit all communities and minimize any drawbacks.

X. FUTURE CONSIDERATIONS

The effectiveness of CatBoost and other machine-learning algorithms in flood management will have multiple opportunities to be enhanced in the near future. The precision and quality of input data could be improved by new advancements in data collection technologies, such as Internet of Things devices and satellites for remote sensing. likewise, interdisciplinary partnerships comprising data scientists, hydrologists, and policymakers are required to create complete solutions that address the effects of floods on society and the environment.

XI. CONCLUSION

In conclusion, machine learning approaches offer valuable insights into early warning systems, flood prediction, and decision support, and they also provide efficient strategies for managing flooding in river basins. One significant advancement in the application of machine learning to reduce the risk of disaster is the use of CatBoost in flood management. With the use of predictive modeling and real-time monitoring, CatBoost is a helpful tool for building community resilience against the growing risk of floods. However, problems with data quality, interpretability, computational resources, and ethical considerations need to be tackled in order to fully reap the benefits of machine learning in flood prevention. Through persistent research, innovation, and collaboration, machine learning approaches can play a key role in the creation of more resilient and adaptable flood management systems to address the challenges posed by urbanization and climate change. The complex phenomena of flooding in river basins is caused by a multitude of factors, including precipitation patterns, changes in land use, and human activity. Integrated flood management solutions that consider both structural and non-structural measures are necessary due to the significant impact that flooding has on ecosystems and populations. By researching the origins and consequences of flooding in river basins, stakeholders can develop effective strategies to lower risks and strengthen resistance to subsequent flood events. However, in order to overcome the challenges and limitations associated with the application of CatBoost for flood prediction, stakeholders from a range of industries must collaborate. With further research and development, CatBoost has the potential to totally change flood management practices and reduce the disastrous consequences of floods globally. Lastly, by utilizing machine learning algorithms, decision-makers can improve the accuracy, efficiency, and sustainability of flood management systems, reducing the impact of floods on the most vulnerable and promoting sustainable development.

REFERENCES

[1]. Abdi-Dehkordi, M., Bozorg-Haddad, O., Salavitabar, A., Mohammad-Azari, S., and Goharian, E. (2021). Development of flood mitigation strategies toward sustainable development. Natural hazards, 108(3), 2543-2567.
[2]. Abedi, R., Costache, R., Shafizadeh-Moghadam, H., and Pham, Q. B. (2022). Flash-flood susceptibility mapping based on XGBoost, random forest and boosted regression trees. Geocarto International, 37(19), 5479-5496.
[3]. Ajibade, F. O., Ajibade, T. F., Idowu, T. E., Nwogwu, N. A., Adelodun, B., Lasisi, K. H., and Adewumi, J. R. (2021). Flood-prone area mapping using GIS-based analytical hierarchy frameworks for Ibadan city, Nigeria. Journal of Multi-Criteria Decision Analysis, 28(5-6), 283-295.
[4]. Al-Kindi, K. M., and Alabri, Z. (2024). Investigating the Role of the Key Conditioning Factors in Flood Susceptibility Mapping Through Machine Learning Approaches. Earth Systems and Environment, 1-19.
[5]. Ansari, M. S., Warner, J., Sukhwani, V., and Shaw, R. (2022). Implications of flood risk reduction interventions on community resilience: An assessment of community perception in Bangladesh. Climate, 10(2), 20.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24JUN1770

[6]. Boone, T., Ganeshan, R., Jain, A., and Sanders, N. R. (2019). Forecasting sales in the supply chain: Consumer analytics in the big data era. International journal of forecasting, 35(1), 170-180.
[7]. Canadell, J. G., Monteiro, P. M., Costa, M. H., Cotrim da Cunha, L., Cox, P. M., Eliseev, A. V., and Zickfeld, K. (2023). Intergovernmental Panel on Climate Change (IPCC). Global carbon and other biogeochemical cycles and feedbacks. In Climate change 2021: The physical science basis. Contribution of working group I to the Sixth Assessment Report of the intergovernmental panel on climate change (pp. 673-816). Cambridge University Press.
[8]. Chakraborty, S. K., and Chakraborty, S. K. (2021). Land-use changes: floodplains, dams, and reservoirs–integrated river basin management. Riverine Ecology Volume 2: Biodiversity Conservation, Conflicts and Resolution, 531-607.
[9]. Chandrathilake, T. (2022). The need for ecohydrological research in tropical forests for healthy watersheds. Journal of Tropical Forestry and Environment, 12(02).
[10]. Dierauer, J. R., Allen, D. M., and Whitfield, P. H. (2021). Climate change impacts on snow and streamflow drought regimes in four ecoregions of British Columbia. Canadian Water Resources Journal/Revue canadienne des ressources hydriques, 46(4), 168-193.
[11]. Dorogush, A. V., Ershov, V., and Gulin, A. (2018). CatBoost: gradient boosting with categorical features support. arXiv preprint arXiv:1810.11363.
[12]. Glago, F. J. (2021). Flood disaster hazards; causes, impacts and management: a state-of-the-art review. Natural hazards-impacts, adjustments and resilience, 29-37.
[13]. Hammami, S., Zouhri, L., Souissi, D., Souei, A., Zghibi, A., Marzougui, A., and Dlala, M. (2019). Application of the GIS based multi-criteria decision analysis and analytical hierarchy process (AHP) in the flood susceptibility mapping (Tunisia). Arabian Journal of Geosciences, 12, 1-16.
[14]. Handayani, W., Chigbu, U. E., Rudiarto, I., and Putri, I. H. S. (2020). Urbanization and Increasing flood risk in the Northern Coast of Central Java—Indonesia: An assessment towards better land use policy and flood management. Land, 9(10), 343.
[15]. Hooper, B. P., and Lloyd, G. J. (2011). Report on IWRM in transboundary basins. UNEP-DHI Cent. Water Environ.
[16]. Huang, G., Wu, L., Ma, X., Zhang, W., Fan, J., Yu, X., Zeng, W. and Zhou, H., (2019). Evaluation of CatBoost method for prediction of reference evapotranspiration in humid regions. *Journal of Hydrology*, **574**, 1029-1041.
[17]. Jia, H., Chen, F., Pan, D., Du, E., Wang, L., Wang, N.,andYang, A. (2022). Flood risk management in the Yangtze River basin—Comparison of 1998 and 2020 events. International Journal of Disaster Risk Reduction, 68, 102724.
[18]. Jonkman, S. N. (2005). Global perspectives on loss of human life caused by floods. Natural hazards, 34(2), 151-175.
[19]. Kulkarni, C. S. (2022) Advancing Gradient Boosting: A Comprehensive Evaluation of the CatBoost Algorithm for Predictive Modeling. J Artif Intell Mach Learn & Data Sci 2022, 1(1), 54-57.
[20]. Kumar, V., Kedam, N., Sharma, K. V., Khedher, K. M., and Alluqmani, A. E. (2023). A comparison of machine learning models for predicting rainfall in urban metropolitan Cities. Sustainability, 15(18), 13724.
[21]. Kumar, V., Kedam, N., Sharma, K. V., Mehta, D. J., and Caloiero, T. (2023b). Advanced machine learning techniques to improve hydrological prediction: A comparative analysis of streamflow prediction models. Water, 15(14), 2572.
[22]. Maskrey, S. A., Mount, N. J., & Thorne, C. R. (2022). Doing flood risk modelling differently: Evaluating the potential for participatory techniques to broaden flood risk management decision-making. Journal of Flood Risk Management, 15(1), e12757.
[23]. Merz, B., Blöschl, G., Vorogushyn, S., Dottori, F., Aerts, J. C., Bates, P., and Macdonald, E. (2021). Causes, impacts and patterns of disastrous river floods. Nature Reviews EarthandEnvironment, 2(9), 592-609.
[24]. Molinari, D., Scorzini, A. R., Arrighi, C., Carisi, F., Castelli, F., Domeneghetti, A., ... and Ballio, F. (2020). Are flood damage models converging to “reality”? Lessons learnt from a blind test. Natural Hazards and Earth System Sciences, 20(11), 2997-3017.
[25]. Mosavi, A., Ozturk, P.,and Chau, K. W. (2018). Flood prediction using machine learning models: Literature review. Water, 10(11), 1536.
[26]. Mushtaq, B., Bandh, S. A., Shafi, S., Mushtaq, B., Bandh, S. A., and Shafi, S. (2020). Management of water resources. Environmental Management: Environmental Issues, Awareness and Abatement, 1-46.
[27]. Nasirzadehdizaji, R., and Akyüz, D. E. (2022). Application of swat hydrological model to assess the impacts of land use change on sediment loads. International Journal of Agriculture Environment and Food Sciences, 6(1), 108-120.
[28]. Nguyen, H. D., Dang, D. K., Nguyen, N. Y., Pham Van, C., Van Nguyen, T. T., Nguyen, Q. H., and Bui, Q. T. (2024). Integration of machine learning and hydrodynamic modeling to solve the extrapolation problem in flood depth estimation. Journal of Water and Climate Change, 15(1), 284-304.
[29]. Peretz T. (2018). Mastering The New Generation of Gradient Boosting. https://towardsdatascience.com/https-medium-com-talperetz24-mastering-the-new-generation-of-gradient-boosting-db04062a7ea2 .Accessed on 29th December 2021

[30]. Petrucci, O. (2022). Factors leading to the occurrence of flood fatalities: a systematic review of research papers published between 2010 and 2020. Natural hazards and earth system sciences, 22(1), 71-83.
[31]. Prokešová, R., Horáčková, Š., and Snopková, Z. (2022). Surface runoff response to long-term land use changes: Spatial rearrangement of runoff-generating areas reveals a shift in flash flood drivers. Science of the Total Environment, 815, 151591.
[32]. Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., and Gulin, A. (2018). CatBoost: unbiased boosting with categorical features. *In Advances in neural information processing systems* 6638-6648.
[33]. Qian, Q. and Eslamian, S. (2022). Impact of urbanization on flooding. In Flood Handbook (pp. 97-112). CRC Press.
[34]. Rajkhowa, S., and Sarma, J. (2021). Climate change and flood risk, global climate change. In Global climate change (pp. 321-339). Elsevier.
[35]. Roy, P., Pal, S. C., Chakrabortty, R., Chowdhuri, I., Malik, S., and Das, B. (2020). Threats of climate and land use change on future flood susceptibility. Journal of Cleaner Production, 272, 122757.
[36]. Saber, M., Boulmaiz, T., Guermoui, M., Abdrabo, K. I., Kantoush, S. A., Sumi, T., ... and Mabrouk, E. (2022). Examining LightGBM and CatBoost models for Wadi flash flood susceptibility prediction. Geocarto International, 37(25), 7462-7487.
[37]. Saber, M., Boulmaiz, T., Guermoui, M., Abdrabo, K. I., Kantoush, S. A., Sumi, T., ... and Mabrouk, E. (2023). Enhancing flood risk assessment through integration of ensemble learning approaches and physical-based hydrological modeling. Geomatics, Natural Hazards and Risk, 14(1), 2203798.
[38]. Saravanan, S., Abijith, D., Reddy, N. M., Parthasarathy, K. S. S., Janardhanam, N., Sathiyamurthi, S., and Sivakumar, V. (2023). Flood susceptibility mapping using machine learning boosting algorithms techniques in Idukki district of Kerala India. Urban Climate, 49, 101503.
[39]. Sedighkia, M., Jahanshahloo, M., and Datta, B. (2023). Evaluating minimum environmental flow requirements in rivers: a combined decision-tree approach integrating hydrological, physical habitat, and water quality indexes. Journal of Sustainable Water in the Built Environment, 9(4), 04023006.
[40]. Seydi, S. T., Kanani-Sadat, Y., Hasanlou, M., Sahraei, R., Chanussot, J., and Amani, M. (2022a). Comparison of machine learning algorithms for flood susceptibility mapping. Remote Sensing, 15(1), 192.
[41]. Shrivastava, P., Smith, M. S., O'Brien, K., and Zsolnai, L. (2020). Transforming sustainability science to generate positive social and environmental change globally. One Earth, 2(4), 329-340.
[42]. Szczepanek, R. (2022). Daily streamflow forecasting in the mountainous catchment using XGBoost, LightGBM and CatBoost. Hydrology, 9(12), 226.
[43]. Tanoue, M., Taguchi, R., Nakata, S., Watanabe, S., Fujimori, S., and Hirabayashi, Y. (2020). Estimation of direct and indirect economic losses caused by a flood with long-lasting inundation: application to the 2011 Thailand flood. Water Resources Research, 56(5), e2019WR026092.
[44]. Tehrany, M. S., Pradhan, B., and Jebur, M. N. (2015). Flood susceptibility analysis and its verification using a novel ensemble support vector machine and frequency ratio method. Stochastic environmental research and risk assessment, 29(4), 1149-1165.
[45]. Tounsi, A. (2023). Enhancing Flood Forecasting Systems Through the Integration of Machine Learning and Advanced Sensing and Modeling Techniques (Doctoral dissertation, Stevens Institute of Technology).
[46]. Van Ackere, S., Verbeurgt, J., De Sloover, L., Gautama, S., De Wulf, A., & De Maeyer, P. (2019). A review of the internet of floods: Near real-time detection of a flood event and its impact. Water, 11(11), 2275.
[47]. Van Phong, T., Nguyen, D. D., and Pham, B. T. (2023). Modeling and Mapping of Flood Susceptibility at Que Son District, Quang Nam Province, Vietnam using CatBoost. In IOP Conference Series: Materials Science and Engineering (Vol. 1289, No. 1, p. 012019). IOP Publishing.
[48]. Wagenaar, D., Curran, A., Balbi, M., Bhardwaj, A., Soden, R., Hartato, E., ... and Lallemant, D. (2020). Invited perspectives: How machine learning will change flood risk and impact assessment. Natural hazards and earth system sciences, 20(4), 1149-1161.
[49]. Wang, L., Cui, S., Li, Y., Huang, H., Manandhar, B., Nitivattananon, V., Fang, X., and Huang, W. (2022). A review of the flood management: From flood control to flood resilience. Heliyon, 8(11). https://doi.org/10.1016/j.heliyon.2022.e11763
[50]. Wu, C., Ju, Y., Yang, S., Zhang, Z., and Chen, Y. (2023). Reconstructing annual XCO2 at a 1 km× 1 km spatial resolution across China from 2012 to 2019 based on a spatial CatBoost method. Environmental Research. [HTML]
[51]. Wu, Y., Ju, H., Qi, P., Li, Z., Zhang, G., and Sun, Y. (2023). Increasing flood risk under climate change and social development in the Second Songhua River basin in Northeast China. Journal of Hydrology: Regional Studies, 48, 101459.
[52]. Xia, J., and Chen, J. (2021). A new era of flood control strategies from the perspective of managing the 2020 Yangtze River flood. Science China Earth Sciences, 64(1), 1-9.
[53]. Xiang, W., Xu, P., Fang, J., Zhao, Q., Gu, Z., and Zhang, Q. (2022). Multi-dimensional data-based medium-and long-term power-load forecasting using double-layer CatBoost. Energy Reports, 8, 8511-8522.
[54]. Xu, Y., Rao, C., Xiao, X., and Hu, F. (2023). Novel Early-Warning Model for Customer Churn of Credit Card Based on GSAIBAS-CatBoost. CMES-Computer Modeling in Engineering and Sciences, 137(3).

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24JUN1770

[55]. Yao, S., Du, W., Chen, N., Wang, C., & Chen, Z. (2023). An integrated process-based framework for flood phase segmentation and assessment. International Journal of Geographical Information Science, 37(6), 1315-1337.
[56]. Zeleňáková, M., Gaňová, L., Purcz, P., and Satrapa, L. (2015). Methodology of flood risk assessment from flash floods based on hazard and vulnerability of the river basin. Natural Hazards, 79, 2055-2071.
[57]. Zhao, D. H., Shen, H. W., Tabios III, G. Q., Lai, J. S., and Tan, W. Y. (1994). Finite-volume two-dimensional unsteady-flow model for river basins. Journal of Hydraulic Engineering, 120(7), 863-883.
[58]. Zhao, Z., Wang, H., Zhang, Y., Deng, C., Xie, Q., and Wang, C. (2020). Problems and countermeasures of river management in the process of rapid urbanization in China. Water, 12(8), 2260.
[59]. Zhong, W., Zhang, D., Sun, Y., and Wang, Q. (2023). A CatBoost-Based Model for the Intensity Detection of Tropical Cyclones over the Western North Pacific Based on Satellite Cloud Images. Remote Sensing, 15(14), 3510.