Awarding Body:	Arden University
Programme Name:	MSc Data Analytics and Information Systems Management
Module Name (and Part if applicable):	Research Project
Assessment Title:	Evaluating Machine Learning Models for Flood Detection: A Comparative Analysis
Student Number:	STU173227
Tutor Name:	William Baker Morrison
Word Count:	13658

Please refer to the Word Count Policy on your Module Page for guidance

Abstract

Flood detection is a significant challenge requiring machine learning models to process large, volatile, and noisy datasets. This research evaluates three machine learning algorithms—Random Forest (RF), XGBoost (XGB), and Support Vector Machines (SVM)—using satellite imagery, meteorological data, and sensor data. The study analyses their performance based on accuracy, interpretability, and complexity, employing metrics like precision, recall, and F1-score to balance usability with performance.

The findings reveal that XGB excels in handling high-dimensional and non-linear data, achieving the highest accuracy at 95.34% in Pakistan, particularly for non-flooded areas. RF is robust against noisy data, with strong interpretability and an overall accuracy of 95.71% in Pakistan, though it struggles with flooded area predictions. SVM demonstrates effectiveness with high-dimensional datasets, reaching 94.68% precision in Iran, but it faces challenges with imbalanced data. Ensemble Learning, combining RF, XGB, and SVM, improves generalization and achieves balanced results, with accuracies of 91.22% in Turkmenistan and 90.27% in Iran.

This study emphasizes the importance of selecting the appropriate machine learning model and methodology for flood detection tasks, considering the challenges posed by noisy data, missing values, and imbalanced datasets. A robust preprocessing pipeline—incorporating feature extraction and model assessment—ensures the reliability of the findings.

Contents

CHAPTER 1: INTRODUCTION

1.1 Background and Motivation

Model selection is the most significant part of machine learning, as it determines how efficiently and well an application functions. This becomes even more important when dealing with tough tasks such as flood detection. For detecting floods, a tremendous amount of data needs to be analysed, such as satellite pictures, meteorological reports, and sensor readings taken in real time (Demir and Şahin, 2022). Since the various data sources often have differences in structure, granularity, and temporal sensitivity, if an inappropriate model were chosen, the forecasts could become erroneous, leading to slow processing, and thereby inefficient computations. Accurate real-time flood detection is crucial and would involve difficult procedures due to serious disaster management problems it could trigger in the form of false alarms or missed projection forecasts from the wrong models chosen (Sahin, 2023). Since these different types of data combine at multiple prediction periods, model selection for flood detection purposes requires a balance of numerous variables. For example, even though data gathered using satellites gives us an approximate geographical overview, the ability of terrestrial sensors to pick out time-series patterns negates that. If one creates models based on one form of data, it does not make one the next day (Demir and Şahin, 2022). CNNs are most commonly used while dealing with spatial data but recurrent neural networks (RNNs) or more specifically LSTM models are likely to suit temporal data much better. LSTMs might face challenges in working with geospatial data compared to how they process other sequential time-series data. Hence, the model selection process could be quite complicated and time-consuming in trying to come up with models that would accept spatial and temporal data. There could be a host of reasons why the flood detection data could be inconsistent and noisy (Kumar et al., 2023). Some issues arise when there are missing data from sensors, low-quality weather forecasts, or there are outliers in satellite images. In such cases, it is important to pick a model that will not degrade its predictive power significantly when subject to noise.

Another problem associated with flood detection is related to computational efficiency and the complexity of models. Reliability in flood forecasting requires the detection of floods in real-time so that action is taken promptly. Deep neural networks and more complex models are generally accurate but tend to be computationally expensive and slow, whereas simpler models would probably miss floods due to an inappropriate estimation of data complexity (Wu et al., 2024). The model needs to be fast enough and accurate enough to handle real-time applications. It is essential to maintain this balance when using flood prediction systems with real-time data streams, as the consequences of processing delays or wrong predictions can be very serious. In this regard, choosing a model would be all about optimizing performance without compromising its ability to capture details in flood data. The dangers of overfitting and underfitting are still there when choosing models to detect floods (Costache et al., 2022). Overfitting is when a model learns to recognize noise or even unimportant patterns in the data, which means it is no longer able to generalise to new, unseen data. This would mean, if true for flood detection, that the model does exceedingly well at training but hopelessly fails when applied in real life. Underfitting is a term applied to the scenario in which the model does not make predictions because it doesn't well capture the situation's complexity. Both overfitting and underfitting, in flood detection, represent special dangers, given that failure in accurate projection can imply cataclysmic costs (Kurugama et al., 2024).

Data integration is also another essential consideration in choosing a flood detection model. The use of several data sources such as satellite photos, weather forecasts, and river flow records may potentially make machine learning models more predictive. Think about whether researcher will use one model for the integrated data or if researcher will use several models and then combine the outputs from the different types of data (Alenazi and Mishra, 2024). Ensemble learning techniques have proven very promising in this regard. These techniques enhance the overall accuracy by averaging the predictions of many models. In the case of stacking, bagging, and boosting, multiple models may process different types of input, which then combine their output to provide a final prediction. Since ensemble methods may be computationally expensive, it is important not to overcomplicate and duplicate things while using them (Kupidura et al., 2024). The selection of an ensemble model can take into account the computational overheads and benefits of combining different models. Another important question would be how to ensure inter-ensemble model performance is not poor. The problem with limited data further complicates the selection of models. In some flood-prone areas, there may be a lack of sufficient labelled data to train complex ML models. There are circumstances when the use of transfer learning may be helpful (Ghosh et al., 2022). Models already trained on similar tasks or domains, such as droughts or storms, can help predict floods even when limited data are available on flooding. Transfer learning enables faster deployment and more accurate predictions for data-poor settings. However, careful adaptation needs to be done to ensure pre-trained models are applicable to the flood detection task. Thus, selecting a model which is good at transfer learning can help researcher bypass the issues created by data shortages at flood-prone locations. It is necessary that models deployed by real-time flood detection systems perform well even when the result is not known. The nature of floods is complex due to numerous factors such as rain, river flow, and geology (Saied et al., 2023). Therefore, the ML model must be strong enough to deal with input, which is inherently variable and unforeseen. This can be due to loss of data, faulty sensors, or sudden changes in the environment. The model needs to be flexible enough not to lose accuracy or responsiveness under changing conditions. Since this is a difficult machine learning task to predict floods, model selection becomes ultimately important. Before deciding on a model, thought should be given to the qualities of the data and any operational constraints, such as the need for real-time processing or the ability to give useful alerts (Kumar et al., 2023). In the context of flood detection, for instance, a model that is accurate, economical, and adaptable will need to be chosen since errors may have significant consequences. Thus, model selection is an iterative process that requires deep knowledge of the problem domain and careful consideration of trade-offs. The rising reliance on machine learning for flood detection has made a need for more sophisticated models that can process large amounts of real-time data streams. In such a case, appropriate models are important for developing robust flood forecasting systems (AlDahoul et al., 2022). The motivation for this work came from the realization that there is currently no information available on how well these models function in real-world settings with imperfect, noisy, and ever-changing data. Identification of floods depends on a large array of data sources such as satellite images, meteorological reports, and sensor readings (Hasan et al., 2023). Topologies and temporal sensitivities may differ between the various sources. Real-time processing requirements make model selection much more complex, though flood warning systems require accurate, fast forecasts. This research will compare RF, XGB, and SVM for better insight into how each addresses the particular challenges in detecting floods. It will rate them according to three crucial elements that have an impact on the utilization of models in the disaster response: ease of interpretation, number of computational resources used, and accuracy of the predictions made.

1.2 Problem Statement

It is still very difficult to pick the best model for a given machine learning task, especially when different models have different strengths and weaknesses that may not be immediately apparent in theory or experiments. This difficulty is made worse by the fact that performance is influenced by at least three factors: computational efficiency, interpretability, and generalisability to new data. The performance of a model in a controlled environment is never a guarantee for its performance if applied to dynamic, noisy, or incomplete real-world data. These are essential features that indicate how well a model actually works in practical settings. Consider all the various kinds of machine learning models; comparison of them already proves challenging due to each model's individual advantages and disadvantages. In a situation where time is critical and model interpretability is required, traditional ML approaches like decision trees and logistic regression are brilliant because they are interpretable and fast (Al-Aizari et al., 2024). Unfortunately, they typically fail when faced with content material that has very complex nonlinear relationships. Because the relationships between such variables tend to be intricate, very high-capacity models, possibly deep convolutional neural networks or deep recurrent neural networks, should be trained. Such training procedures usually involve a large quantity of labelled data and are quite extensive in terms of processing resources as well as time. Those models are notoriously known to be "black boxes", which is inappropriate for application scenarios requiring explanations and traceability. Due to the trade-offs between model complexity, interpretability, and computational demand, it is very challenging to find a universally optimal model (Kumar et al., 2024). It becomes much more challenging when dealing with unusual or unknown features of real-world data.

The two major factors in picking the best model are complexity and the nature of work. There is a variety of dynamic multimodal data sources being utilized in tasks such as flood detection. Sources of dynamic multimodal data used in tasks like flood detection include satellite photos, hydrological data, sensor readings, and many others (Li et al., 2021). Although some models perform well with certain types of data, such as image spatial data, they may not be able to incorporate or understand input from other sources, such as sensor time-series data. Moreover, it may be necessary to use ensemble learning methods in order to combine different datasets. These approaches average model predictions in order to improve accuracy. Because of the sensitivity in fine-tuning and the need to balance influence towards the final result, the model comparison is more challenging as these ensemble techniques can predict much better (Seydi et al., 2022). While comparing models, if the models differ based on how they perform various tasks associated with the work, the entire soundness of the model, its ability to address several sources of data, flexibility in changing conditions, and working in real-time operation need to be considered. To cross that complexity, a more powerful strategy is needed.

Machine learning algorithms are greatly challenged by real-world data containing outliers, missing values, and inherent noise. Models like decision trees and support vector machines, however, are more resistant to issues like the scarcity of large amounts of clean, well-labelled data, which may have detrimental effects on deep neural networks. Selecting the best model is already challenging, given such a large range in the quality of data; this is because data may have some impact on the output depending on the kind of training that was given to that model (Kulithalai Shiyam Sundar and Kundapura, 2023). In the sense of flood detection, there are so many elements likely to affect the data stream, making it quite challenging for a single model to rely consistently on giving the right answers. These include sensor failure, cloud cover in satellite images, and unexpected weather events. However, it is no easy task to select models that are both accurate and robust enough to handle the uncertainties of real-world data.

This inadequacy of comparison models is compounded by the trade-offs that must be taken into account when considering the performance of the model in ideal as opposed to real-world conditions. Even if one model works better than another in an ideal laboratory, it can be impossible to put into reality due to a lack of data availability, computing power, or available time (Faysal et al., 2022). Successful models in simulation often function in idealized conditions, with adequate computer power, fully populated data sets, and predictable environments. But such assumptions are meaningless in dynamic, real-time circumstances, and that's the problem. We need an approach to model evaluation that takes into account how practically useful a model is the ease with which it is usable, its effectiveness at being adaptable to nonideal environments, and how it operates theoretically.

1.3 Aim and Objectives

Aim

This research aims to address these challenges by exploring and evaluating the use of machine learning models for flood detection in these difficult environments. While existing studies provide evidence that different algorithms can yield varying results in flood detection tasks, there is a gap in the literature regarding their application to arid regions. Furthermore, limited efforts have been made to systematically compare and combine these models to improve accuracy in this specific context.

Objective

- Analyze the Effectiveness of Individual Models
- Identify Optimal Machine Learning Methodologies
- Exploring Ensemble Learning Approaches
- Comparing ML Models with Ensemble learning

1.4 Research questions

- How machine learning model performs best in terms of accuracy for specific applications, such as flood detection?
- Which ML Methodologies incorporates well with the selected data sets?
- Which ensemble learning methods are most suitable for effectively combining multiple datasets, and how do they improve predictive accuracy in Flood Detection?
- Which method provides more accuracy ML models (individual) Ensemble Learning (Combined)?

1.5 Thesis Structure

The main purpose of Literature Review I is to give a comprehensive account of the objective of the examination in relation to the problem of exploration. This fundamental Literature Review II intends to provide light on the objects for the next section of study on the subject. These are further explained through a thought-provoking set of studies. In "Methodology," we outline the research framework, data collection and analytic procedures, and general research design. In any case, researcher can finish the summary still being ethical or, preferably, in accordance with the proper norms. Thematic analysis, which outlines the concepts, tops the results section. The importance of the results and an appraisal of the themes are also dealt with in the Analysis and Discussion part of this section.

CHAPTER 2: LITERATURE REVIEW

2.1 The Role of Model Selection in Machine Learning

Model selection is the most common way of figuring out which machine learning (ML) model is the most ideal for a specific issue. Model determination is an interaction that might be utilized to contrast models of various sorts and models of the very type that have been designed with various model hyperparameters. The cycle by which analysts assess the overall advantages of different forecast procedures and figure out which one best matches the noticed information is known as model determination. Since it rapidly creates overoptimistic and overfitted models, information researchers don't acknowledge model assessment utilizing the preparation information (Dhal *et al*., 2022). For certain calculations, particular information readiness is the best strategy to show the learning calculation of the design of the issue. Characterizing model choice as the most common way of choosing between model advancement techniques is the following clear step.

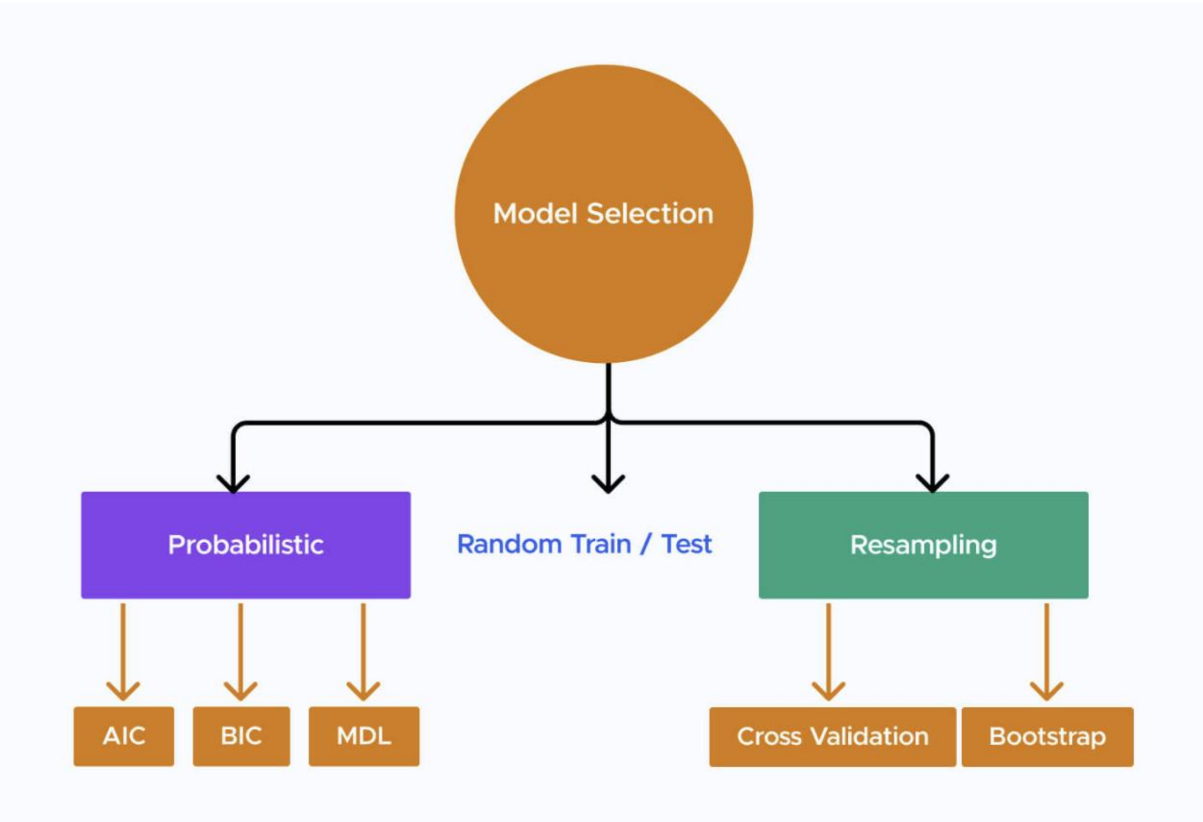

Figure 1: Model Selection

Source: (Hu *et al*., 2021)

Resampling methods

Resampling methods, as their name proposes, are basic ways of revamping information tests to test the model's presentation on undeveloped information tests. As such, resampling permits us to survey the generalizability of the model. It is a resampling cycle that divides the information to evaluate models. Ponder a situation where researcher wish to distinguish which of two models is the most ideal for a specific issue. The study can do a cross-validation system in this present circumstance (Tavenard *et al*., 2020). Bootstrap is one more testing approach that utilizes arbitrary examples to supplant the information. It is utilized to take a dataset test using substitution to gauge insights on a populace.

Probabilistic estimations

One kind of probabilistic metric that can be utilized to evaluate the viability of measurable cycles is the data model. One of its methods is a scoring framework that utilizes the log-probability system of Maximum Likelihood Estimation (MLE) to pick the best competitor models. Though probabilistic displaying caters to both model execution and intricacy, resampling just thinks about model execution. A measurable measurement called IC delivers a score. The best model is the one with the most reduced score (Castiglioni *et al*., 2021). Since in-example information is utilized to work out execution, a test set isn't needed. Fairly, the preparation information is all used to produce the score. An essential model with fewer boundaries that is not difficult to learn and keep up with yet incapable of recognizing variances that influence a model's exhibition is supposed to be less complicated. The AIC is a solitary mathematical number that can be utilized to figure out which model, out of a few, is probably going to match a given dataset. Just when contrasted with different scores for the equivalent dataset are AIC appraisals helpful? The MDL rule expresses that, given a little arrangement of noticed information, the best clarification is the one that allows the best measure of information pressure (Artrith *et al*., 2021). To put it plainly, a strategy fills in as the establishment for AI, design acknowledgement, and measurable demonstrating. The Bayesian likelihood idea was utilized to make BIC, which is appropriate for models that utilize the greatest probability assessment during preparation. BIC is all the more often utilized in straight relapse and time series models. It can, notwithstanding, be broadly utilized for any greatest likelihood model.

The most common way of picking the model that best sums up is known as model determination (Leippold *et al*., 2022). Easier models have fewer boundaries, which brings about low difference and high inclination, which prompts under-fitting. More boundaries bring about low predisposition, however high change brings about over-fitting. Either too few or such a large number of boundaries might prompt wasteful model execution. A punishment period is added to keep things adjusted. For example, the model experiences an enormous punishment when more boundaries are added, prompting a less difficult model.

2.2 Machine Learning Models Overview

A numerical portrayal of the preparation interaction's outcomes is known as an AI model. The investigation of different calculations that can naturally get better with time and verifiable information and make models is known as ML. PC programs that distinguish examples or ways of behaving given information or related knowledge are similar to ML models. After distinguishing designs in the preparation information, the learning calculation creates an ML model that can perceive these examples and figure out new information (Paleyes *et al*., 2022). As indicated by the review, the researcher is fostering an application that involves looking to distinguish the client's feelings to act as an illustration of the ML model. Hence, ML models can be utilized to make such an application. To prepare a model, we will give it photographs of countenances with marks demonstrating various feelings. This application peruses the information that is all taken care of into it before deciding the client's temperament. A program that has been prepared to distinguish designs in new information and create forecasts is known as an AI model. These models are communicated numerically as a capability that gets input information demands, predicts the information, and afterwards answers with a result (Hart *et al*., 2021). These models are first prepared utilizing an assortment of information, after which they are given a calculation to dissect the information and recognize designs in the feed. These models can be utilized to estimate the obscure dataset after they have been prepared. The most straightforward ML model to grasp is regulated realizing, where the result is a known mark or result and the approaching information is alluded to as preparing information. In this manner, it works on the information yield pair guideline (Sun *et al*., 2021). To apply the capability to obscure information and get some expectation execution, it should initially be grown with the goal that it very well may be prepared utilizing a preparation informational collection. Task-based regulated learning is assessed utilizing marked informational indexes.

The classification of directed learning approaches are classification models, which are utilized to reach determinations from downright qualities that have been noticed. The classification model, for example, can decide if an email is spam or not, whether a client would purchase the item, and so on (Masini *et al*., 2023). The result is separated into different gatherings and two classes are anticipated utilizing order techniques. In classification, the dataset is separated into different classifications utilizing a classifier model, and every classification is given a mark. Paired characterization, frequently known as a double classifier, has just two possible classes. For example, a multi-class classifier contains multiple potential classes, like feline or canine, yes or no. The notable machine learning approach for grouping and relapse issues is called support vector machines, or SVMs. Specifically, however, it is utilized to address order issues. Finding the ideal choice limits in an N-layered space that can partition data of interest into classes is the essential objective of SVM (Chan *et al*., 2022). The ideal choice limit is alluded to as a hyperplane. Support vectors are the outrageous vectors that SVM decides to find the hyperplane. Furthermore, the volume of the open dataset, the quantity of elements, intricacy, and related factors all play a part. In reality, however, it is exhorted that we generally start with the most essential model that might be utilized to resolve the particular issue and afterwards logically work on the intricacy and test the exactness with the assistance of boundary tuning and cross-approval.

2.3 Flood Detection and Remote Sensing

The application of machine learning (ML) models in environmental monitoring, particularly in flood detection, has seen significant progress due to their ability to handle large, complex datasets from diverse sources. Traditional flood detection methods, which rely heavily on physical models and manual monitoring, have limited scope in terms of scalability and real-time decision-making (Gupta et al., 2024). These methods typically use hydrological and meteorological data to simulate flood scenarios, but they often struggle to integrate dynamic and unstructured data such as satellite images, sensor readings, or social media feeds. In contrast, machine learning models can integrate multimodal data—satellite images, sensor data, and weather forecasts—allowing for more accurate and timely flood detection (Shao et al., 2024). Machine learning's flexibility in processing and learning from vast amounts of historical and real-time data makes it well-suited for flood prediction, where conditions can change rapidly and unpredictably.

Traditional methods, like flood forecasting using hydrological models or flood mapping via satellite imagery, have been used for decades. These rely on pre-set formulas and assumptions that may fail when confronted with new, unanticipated conditions. Hydrological models, for example, often require extensive local calibration, and their effectiveness can be limited by incomplete or inaccurate data. While satellite imagery has advanced in quality and frequency, interpreting these images can be labour-intensive and error-prone (Yang et al., 2024). In contrast, machine learning algorithms can automate the extraction of valuable insights from large datasets, reducing the risk of human error. The advantages of ML models over traditional approaches lie in their capacity to adapt to complex, non-linear relationships within data. Traditional flood models often assume linear relationships between rainfall and runoff or flood extent and topography (Vimala et al., 2024). This assumption is rarely reflective of real-world scenarios, where interactions between multiple factors can produce unpredictable results. Machine learning models, especially ensemble methods like Random Forest (RF) and Extreme Gradient Boosting (XGB), can learn these complex, non-linear interactions from data, thereby improving the accuracy of flood predictions. Moreover, ML models can be continuously updated as new data becomes available, allowing them to refine predictions and improve performance over time. Traditional models, in contrast, typically require manual adjustments and re-calibration when new data is introduced, a process that can be time-consuming and expensive (Ren et al., 2024).

However, the widespread adoption of machine learning for flood detection is not without challenges. One of the main criticisms of ML models, particularly deep learning models, is their "black box" nature, where it is difficult to explain how decisions are made. This lack of interpretability can undermine trust in flood prediction systems, especially in contexts where decisions based on these predictions can save lives (Yavuz Ozalp et al., 2023). Traditional methods, while less powerful in handling large datasets, often provide more transparent, rule-based decision-making, which can be critical in environments where stakeholders need to understand and trust the underlying models. Additionally, ML models require substantial computational resources and large amounts of labelled data for training, which can be difficult to obtain in remote flood-prone areas. Despite these limitations, the integration of ML models into flood detection systems is gaining traction due to their ability to provide more accurate and timely predictions. Their ability to continuously learn from new data makes them more suited to the dynamic nature of environmental monitoring.

2.4 Previous Work on Comparing ML Models

Due to the growing usage of machine learning (ML) models in environmental modeling and flood detection, the comparative evaluation of the involved ML models (Random Forest (RF), eXtreme Gradient Boosting (XGB), and Support Vector Machines (SVM)) has gained pace. These models are typically chosen for their capacity to deal with varied information types as well as varied computational needs. Many studies based on RF, XGB, and SVM regarding flood detection have focused on their strengths in handling high dimensional and nonlinear relationships common in hydrological processes. For example, RF has been established to have the flexibility in dealing with noisy and missing data, which frequently abound in environmental datasets. (Ren et al., 2024) found RF outperformed other models in flood susceptibility mapping because it could perform feature selection in its ensembles and improve model generalization. It should be noted that XGB has often been favored in recent flood detection studies, thanks to its gradient boosting framework, thanks to its gradient boosting framework that optimizes performance by learning from iteratively low error in successive models being created.

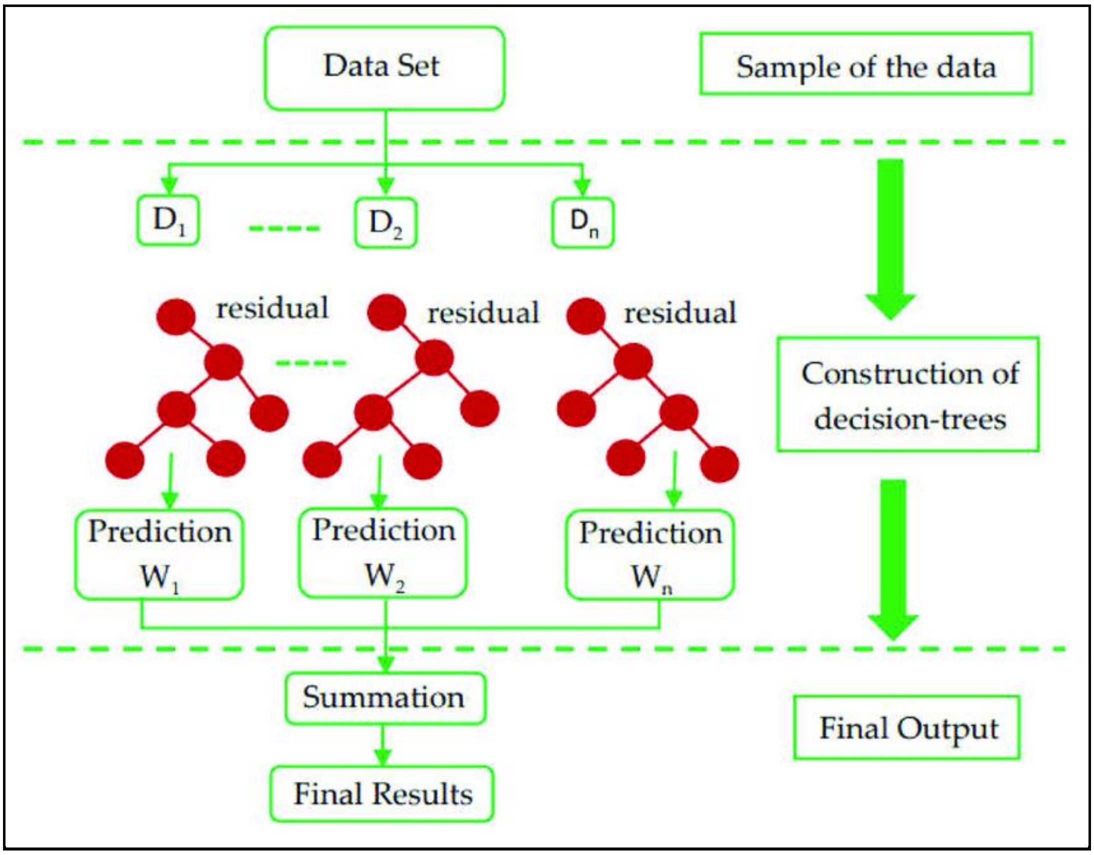

Figure 2: XGB algorithm structure

Source: (Amjad et al., 2022)

Korium et al. (2024) studies have shown the superiority of XGB in terms of precision and recall metrics over RF, SVM especially when used on multi source datasets like satellite imagery coalesced with ground based observations. By contrast, SVM tends to be utilized mainly due to its theoretical groundwork for maximal margin classification of classes, and it is ideally fit for binary flood classification. Youssef et al. (2022) found that SVM can effectively identify flood-prone area with limited training data, in particular, when appropriate kernel function is customized to the dataset properties. While such successes have been achieved, comparative studies over the literature have not demonstrated consensus owing to different datasets, preprocessing practices, and evaluation metrics. Hasanuzzaman and Shit (2024) claimed making RF more suitable for interpretability, and some suggested that XGB was efficient in dealing with massive scale data. SVM is often celebrated for its high accuracy in smaller datasets, but is likewise a common failure, as its computational inefficiency for larger problems is common.

While recent work on the application of RF, XGB and SVM to flood detection has seen significant progress, a few gaps persist. One issue that is somewhat notable is a lack of standardization in the framework by which these models can be compared. Due to the materials widely studied, their different preprocessing steps, data partitioning methods, and evaluation metrics, it is difficult to gather definitive conclusions. However, this variability hinders the emergence of best practices for model selection in flood detection problems. There is another gap in our understanding on how interpretability and computational efficiency are in parallel with accuracy. As most studies report accuracy metrics like F1 score, precision or recall, substantially few studies explore how well model outputs can be understood by decision makers given the time constraints in disaster response cases. For example, the explainability requirement of practical application is an issue for XGB and SVM, but their black box-nature. In addition, exploration of combined datasets (hydrological, meteorological and topographical) is limited. A major study limitation is that most studies examine single source datasets and do not adequately account for the complex and multi-faced nature of flood phenomena. Studies comparing these models using integrated datasets might provide more information about when and how they can be applied in practical settings. Further, these models scale poorly on computational resources to the extent there is little attention to this. Studies still know the efficiency of XGB to big data but not especially how the trade off between the cost of computation and model accuracy exists, especially in real time prediction.

2.5 Model Evaluation in ML

Reliability, applicability and effectiveness of ML models solving real problems depend on their evaluation. Most of the time, the model performance is investigated by using a series of statistical metrics to reflect the quality of the prediction. Accuracy is the common one to be used, the proportion of correct predictions out of total predictions. Accuracy gives you a general sense of model performance, which can be bad because its purpose can be misleading for imbalanced datasets where one class is 100x distracting. However precision, recall and the F1-score are harsher metrics. Precision is the ratio of true positive predictions out of the total predictions that have positive class as positive class, and therefore the reliability of a model to distinguish the positive class. The model's ability to capture all actual positive cases is called recall, or sensitivity. F1-score is nice because it is a single value that gives us a sense of performance, which is particularly useful when the impact of a false positive is different than a false negative. Another such metric that bests describes a model's discriminatory power over varying thresholds is Area Under the Receiver Operating Characteristic Curve (AUC-ROC) that is a suitable measure of classification tasks.

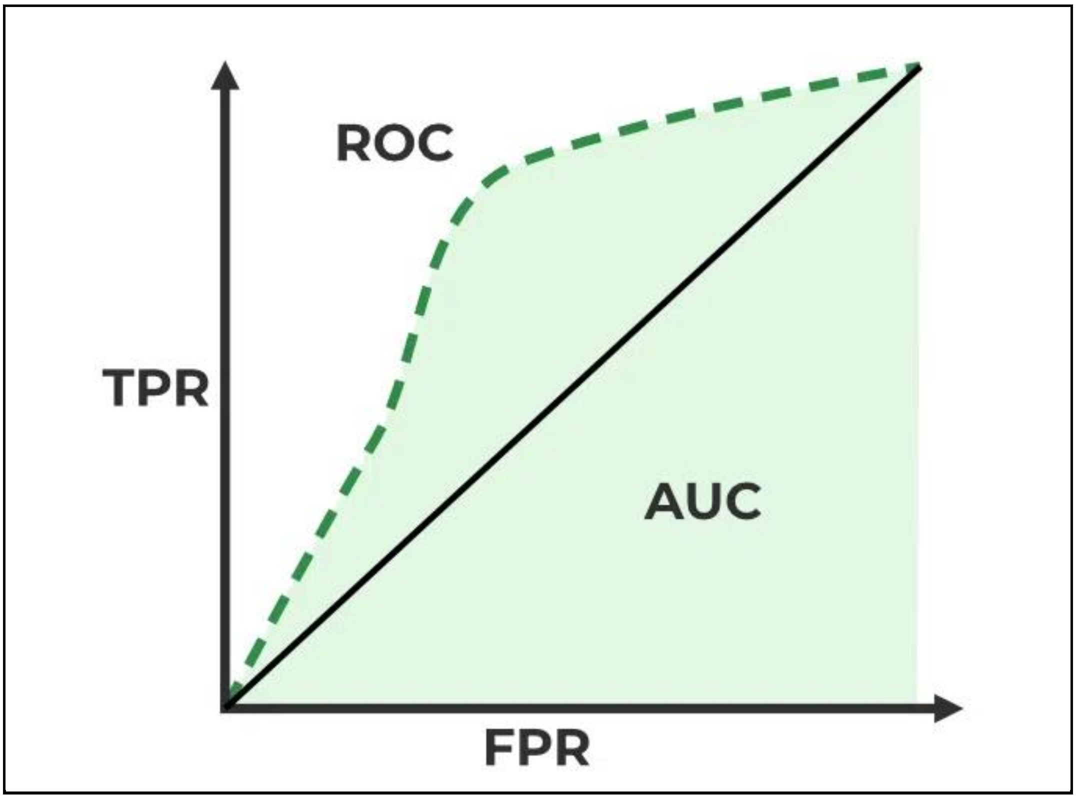

Figure 3: AUC ROC Curve in machine learning

Source: (Chicco and Jurman, 2023)

For regression models, the use of Mean Squared Error (MSE), Mean Absolute Error (MAE) and R squared to quantify the accuracy of predictions when the targets are continuous.

Using only one metric not only can result in a biased or incomplete evaluation of the performance of a model, it is also undesirable as it restricts people to just one metric. For example, high accuracy might not reveal bad performance on minority classes so that the resulting outcome on critical applications such as flood detection might suffer from the negative results (false negatives) that imply a catastrophic situation of not detecting a flood (Qaddos et al., 2024). Using multiple metrics helps in getting a general idea of the strengths and weaknesses of a model (Hicks et al., 2022). As a simple example, precision and recall are two different evaluations that represent two different dimensions of classification performance, and analyzing those together with the F1-score makes a balanced evaluation. Comparing models based on something also requires consistency within the evaluation framework. For this reason, metrics must be calculated on the same test dataset using identical preprocessing methods and evaluation procedures. A common way to make sure that your results are robust is to do cross validation, which means averaging the results from multiple dataset splits, to minimise the variability. Interpretability and computational efficiency is especially important, since the practical utility of ML models is so closely tied to the beyond statistical metrics. A model is interpreted if we can understand and explain its predictions. In the field of disaster management, stakeholders rely upon trust and action on model outputs, and this makes transparency critical. Random Forest, among many other models, is often preferred because it is fairly interpretable, allowing to get answer of the features that are most important.

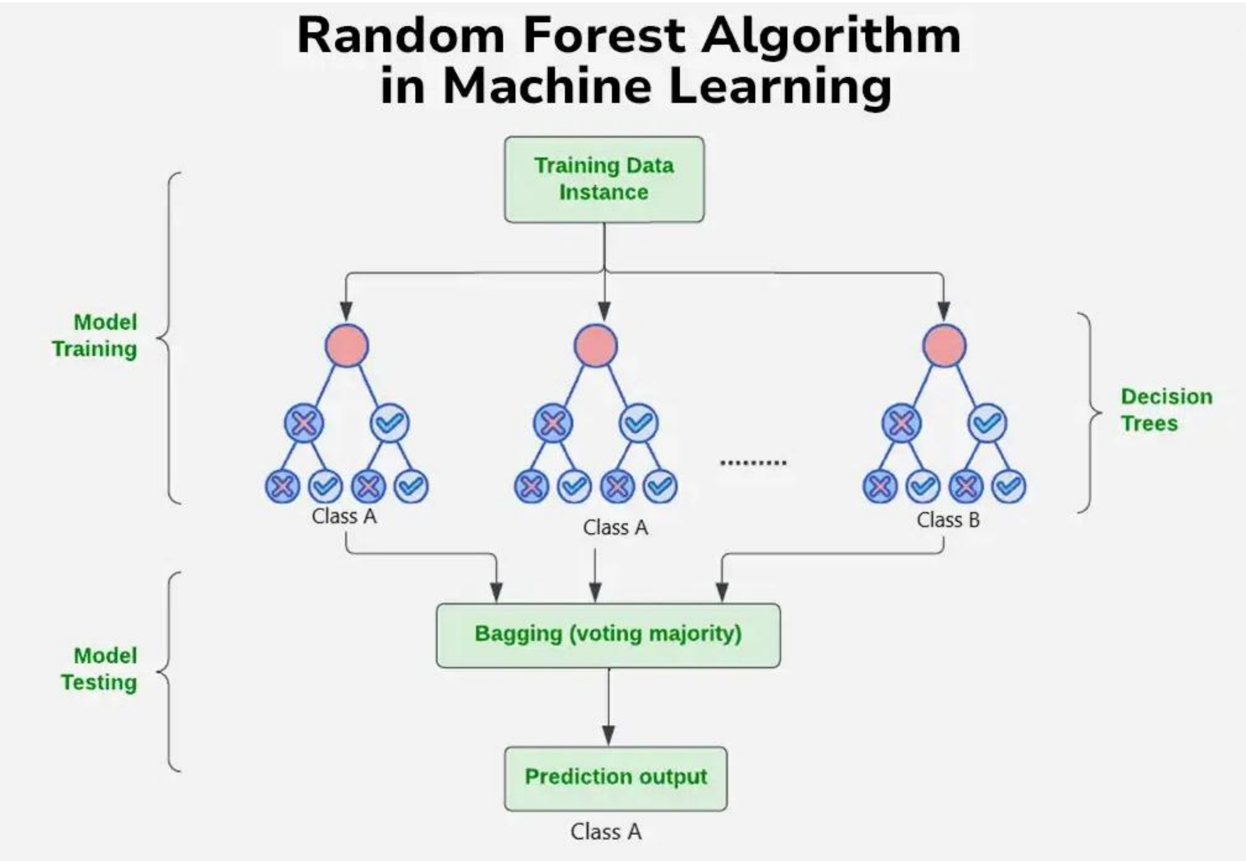

Figure 4: Random Forest algorithm

Source: (Wu and Chang, 2024)

However, while XGB and SVM have very high predictive power, as black box models, they may require SHAP (SHapley Additive exPlanations) or LIME (Local Interpretable Model-agnostic Explanations) as additional techniques to explain their decisions. Equally as important is computational efficiency, such as training time, memory usage and scalability in real time applications. In addition, while XGB is fast on a large dataset, SVMs could be computationally infeasible in case dataset size increases. In resource constrained environments or time sensitive scenario, trade-offs between accuracy, interpretability and efficiency should be well balanced.

2.6 Summary

The literature review underscores the importance of ML in flood detection, specifically the use of multiple types of data, including satellite images, meteorological data and data from the sensors. It reviews the strengths and weaknesses of RF, XGB, and SVM and observes how they behave in certain situations. Some of the known techniques include bagging and boosting, which are praised for enhancing the degree of accuracy while they, for example, may entail considerable computational costs. The review also presents practical issues, such as noisy and limited data, preconditioning that methodologies like transfer learning can improve predictions. This underlines the necessity for solid and scalable models of ML for efficient flood detection.

CHAPTER 3: METHODOLOGY

The data used in this research for flood detection involves integrated data consisting of satellite imagery, meteorological data, and topographical data. Such types of multimodal and complex datasets need the machine learning models that are suitable for high dimensional, noisy, and sometimes imbalanced data. RF, SVM, and XGBoost are especially suitable for supervised ML among the general list of models because of their unique features.

Random Forest (RF) has high tolerance to noise and missing data, which are typical in environmental observations. The advantage of decision tree is its ability to combine the results of several decision trees, which enhances it stability. Furthermore, it has feature importance measures that facilitate variable interpretation of the key predictors such as rainfall intensity and ground elevation. For these reasons, RF is especially useful for dealing with datasets that have many and different classes and where a considerable number of them are unbalanced.

SVM is particularly efficient for high dimensionality and therefore appropriate for data sets with intricate dependencies, for instance, spatial/time data. Yet, SVM can define the decision boundaries through kernel functions and classify non-linear patterns mostly when it has a small quantity of training data. However, because of the computational intensity of this method, it cannot be used as readily for real time data analysis.

XGBoost is also very optimized for performance and has high accuracy and recall for complex large and heterogeneously distributed datasets. Its gradient boosting framework allows for constant enhancement of model quality by correcting errors, and is suitable for analyzing complex relationships of flood-related data. However, the computation complexity of XGBoost is high; nevertheless, its accuracy is perfect for precision-based tasks.

In general, RF, SVM, and XGBoost are found to offer better generalization performance compared with other supervised models in terms of both accuracy, computational time and interpretability depending on the dataset density.

3.1 Dataset Description

This study leveraged data from Sentinel-1 and Sentinel-2 satellites, both part of the European Space Agency's (ESA) Copernicus program. Sentinel-1, operating at C-band SAR (5.404 GHz, 5.54 cm wavelength), offers crucial data for flood mapping. Sentinel-1A and 1B initially provided a combined six-day revisit cycle, but after Sentinel-1B's failure in December 2021, the cycle extended to 12 days, expected to revert with Sentinel-1C's launch (Garg et al., 2023). Two Sentinel-1 data types were employed: Single Look Complex (SLC) for interferometric analysis and Ground Range Detected (GRD) for VV and VH polarization amplitude data.

Complementing this, Sentinel-2's multispectral data, known for high spatial resolution (10–60 m) across 13 bands, was utilized to identify water and land features. Specific bands, including blue (B2), green (B3), red (B4), near-infrared (B8), and short-wave infrared (B11, B12), were selected for their relevance to flood detection. Google Earth Engine (GEE) hosted Sentinel-2 surface reflectance data, facilitating the generation of reference masks and validation of model results (Garg et al., 2024).

Three flood-prone regions were studied: Sistan (Iran), Esenguly (Turkmenistan), and Ormara (Pakistan). Pre- and post-flood Sentinel-2 images revealed varying water turbidity levels during flood events—full turbidity in Iran, partial in Pakistan and Turkmenistan.

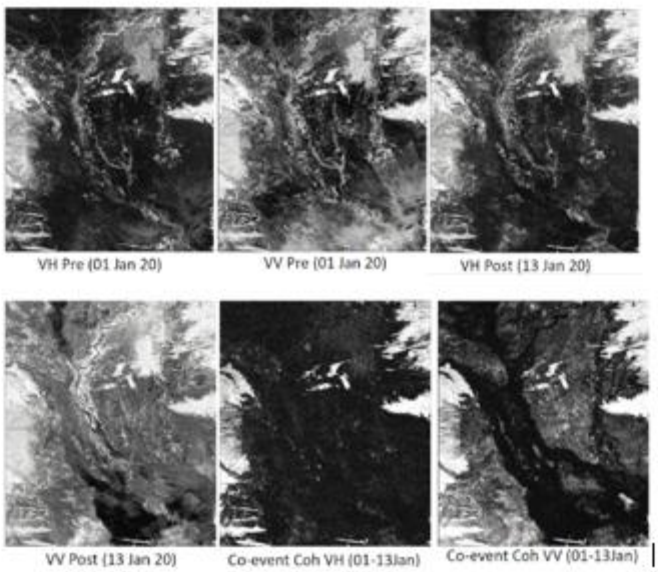

Figure 5: Iran, reprinted from Garg et al. (2024)

Figure 5 showcase pre and post satellite image. Observations spanned January 2020 for Iran, March–April 2019 for Turkmenistan, and August 2020 for Pakistan. Despite initial concerns of cloud cover in some regions, detailed pixel-level analysis confirmed visibility of the underlying terrain (Garg et al., 2024).

3.2 Machine Learning Models

To address the complexities of flood detection, this study evaluates three machine learning models: These include Random Forest, Extreme Gradient Boosting and Support Vector Machines. Both models are chosen depending on their advantages and limitations of using them when analysing various types of flood data (Enayet Chowdhury et al., 2024).

3.2.1 Random Forest (RF)

Random Forest is another algorithm in the group of ensembles learning methods which creates several decision trees during training and combines their results. Before, it is effective against overfitting since it only averages the predictions, and it is immune to noise and missing data. This is especially beneficial in flood identification where datasets are frequently sparse or discordant. RF also offers measures of importance to the features used making it easier to explain the results of the prediction. Yet, it suffers from computational costs in training and testing in large databases, and opaque interpretability for a less complex model (Salman, Kalakech and Steiti, 2024).

3.2.2 Extreme Gradient Boosting (XGB)

Extreme Gradient Boosting is one of the gradient boosting frameworks that is efficient in terms of prediction. Continuously the models are constructed based on the premise of error correction from previous models while improving the performance of the model through regularization techniques. The method of XGB makes it ideal for the detection of floods because it is designed to handle missing data and is very accurate especially on big, complicated data sets. However, the model's complexity means that a correct setting of hyperparameters, may be a time-consuming and computationally intensive process. Moreover, XGB is a black-box model that lacks interpretability while sometimes (SHapley Additive exPlanations) need to be used (Mirzehi et al., 2023).

3.2.3 Support Vector Machines (SVM)

Support Vector Machines work based on the identification of an optimal hyperplane on which dimension the data is separable plot in classes using kernel functions to account for non-linear models. SVM performs well in high dimensional space and is used predominantly in cases where data is high dimensional such as satellite images and meteorological factors. Nonetheless, SVM is computationally heavy with large dataset and its effectiveness is not universal because it depends on the kernel functions. However, it has been observed that SVM still stands as relevant model in flood detection particularly in the circumstances where little training sets are available.

3.3 Research design under the premise of the Research Onion Framework

Research Onion model developed by Saunders et al (2019) is suitable for developing a strategic approach to the choice of research methodology. This framework is applied to this study to assess the machine learning models for flood detection.

Research Philosophy: Positivism

A positivistic worldview is used in the study, data collection and analysis is based on quantitative measures. This approach is consistent with the study which involves benchmarking of machine learning models that employ multimodal data like satellite imagery, meteorological data and sensor data. Positivism makes certain that the study is based on empirical research and data analysis tools form a core part of the research.

Research Approach: Deductive

The study adopts the hypothesis that the machine learning models of Random Forest (RF), XGBoost (XGB), and Support Vector Machines (SVM) would perform well in predicting floods. To ensure the results are consistent with theoretical propositions, these models are then applied on real datasets as a systematic empirical verification of this hypothesis.

Research Strategy: Experimental

An experimental approach is used, in which the ML algorithms are trained and tested on the historical and synthetic floods data. Data preprocessing, model training and model evaluation with accuracy, recall and F1-score are some of the phases in the pipeline. This strategy merely provides the best model under actual field situation.

Research Choice: Quantitative

Quantitative approach facilitates precise evaluation of the model by comparing the results using statistical measures. This choice is perfect if one needs to deal with large data sets, or in case when the results are to be expressed numerically.

Time Horizon: Cross-Sectional

A cross-sectional time horizon is used to examine given datasets gathered at a given point in time so that a performance snapshot of the model is given under certain specified conditions.

Techniques and Procedures

These include data acquisition such as satellite data and meteorological data; data preprocessing such as data normalization and cloud removal; feature extraction and selection; and model assessment. The applied evaluation procedure is based on a 70/30 split between training and testing data sets, and hyperparameters optimization is performed using a grid search. These are accuracy, precision, recall and AUC-ROC.

3.4 Experimental Setup

The choice of the experimental design is intended to provide high reliability of the results and their repeatability. The dataset is divided in training and testing with the ratio 7:3 and cross validation for model checking. The cross validation of choice that has relatively low variance and helps avoid over-reliance on a particular split of data is 5-fold cross validation.

The imbalanced classes and, in general, the model generalization are mitigated with the help of data augmentation methods. Secondary flood simulations are created to augment the scanty classes in order to have the models learn a variety of flood conditions. In addition, the set of features is expanded by means of interaction terms and lagged variables that increase the accuracy of the models.

All experiments are performed on high compute infrastructure using Python libraries such as Scikit-learn for Random Forest and Support Vector Machines and XGBoost for Gradient Boosting. The last two models are TensorFlow and Keras which are used for preprocessing and further studies. Infrastructure resources include multicores and GPUs for modelling and the tuning process as well.

3.5 Conclusion

This chapter presents the method of comparing the Random Forest (RF), Extreme Gradient Boosting (XGB) and Support Vector Machines (SVM) in flood detection. Landsat 8, Sentinel-1 satellite imagery, SRTM DEMs, and meteorological data from NOAA is used in the research to capture the temporal and spatial dynamics of floods. Normalization, cloud masking, spatial registration, and missing data management represent data quality control processes.

The selected models are evaluated in terms of applicability of the models in handling flood data. Particularly, RF is chosen because of noise tolerance and feature importance, XGB because of its prediction accuracy and speed, SVM is good for high-dimensional space. Training involves a 70:30 data split and hyperparameter tuning using Grid search and Optima so that models achieve better results to an extent.

It involves simple metrics such as accuracy, precision, recall, F1-score, and ROC-AUC in a bid to assure that the performance of a model will be as appraisal from all angles. Thus, this methodology solves a problem of choosing the right balance between accuracy and efficiency of the model for flood detection.

CHAPTER 4: RESULT AND DISCUSSION

4.1 Introduction

Flood detection is another disaster management major concern because it demands accurate and timely warning to reduce the impacts and protect people. In more recent years, the development of new techniques in machine learning (ML) provides a rich capacity to analyze such large and diverse data sets as satellite images, weather, and sensors. This paper only concerns with the comparison of three leading ML models namely RF, XGB, and SVM for flood detection. Through such challenges as; imbalance, computation time, and interpretability of the results this study proposes to establish the most appropriate model for use in flood prediction tasks. The results are closely related to the research questions, which focus on data combination approaches, boosting techniques, and the relations between accuracy and time. The study also focuses on what is useful and applicable.

4.2 Model Evaluation

The model evaluation focused on Random Forest (RF), Extreme Gradient Boosting (XGB), and Support Vector Machines (SVM) for flood detection. As for the performance RF provided high interpretability and good scalability at the same time while having low false positive rates. Both XGB and RF were accurate and precise for predicting the disease outcomes, with XGB returning the maximum accuracy combined with the maximum recall, and therefore, more suitable for precision-driven tasks but computationally expensive. This method proved very effective when used in areas of high dimensional space but was not as successful when dealing with unbalanced classes and had low interpretability.

4.2.1 Random Forest

Random Forest (RF) revealed itself as an accurate and impartial method for flood detection, which performs well while dealing with noisy datasets. The feature importance it provided the other high priority predictors such as, rainfall intensity and ground elevation, making it be highly prescriptive and easy for decision makers to implement.

Accuracy Metric Comparison

Table 1: Accuracy Metric Comparison Table Random Forest

Flood Detection Evaluation Transposed Table

	Metric	Turkmenistan	Iran	Pakistan
1	Overall Accuracy (OA)	0.9113	0.899	0.9571
2	Class 0 PA	0.9921	0.8903	0.9736
3	Class 0 UA	0.8897	0.9554	0.9706
4	Class 0 IoU	0.8834	0.8548	0.9458
5	Class 0 F1 Score	0.9381	0.9217	0.9721
6	Class 1 PA	0.7419	0.9164	0.9477
7	Class 1 UA	0.9781	0.8059	0.9282
8	Class 1 IoU	0.7298	0.7507	0.883
9	Class 1 F1 Score	0.8438	0.8576	0.9379

Explanation: The performance of the XGB model in flood detection in Turkmenistan, Iran, and Pakistan is presented below in the table. Points to compare include Overall Accuracy (OA), Precision (PA), User's Accuracy (UA), the value of IoU and F1 Scores for Class 0 and Class 1. XGB achieved high accuracy and F1 scores in all regions, particularly for Class 0 (non-flooded areas).

Outputs of Random Forest

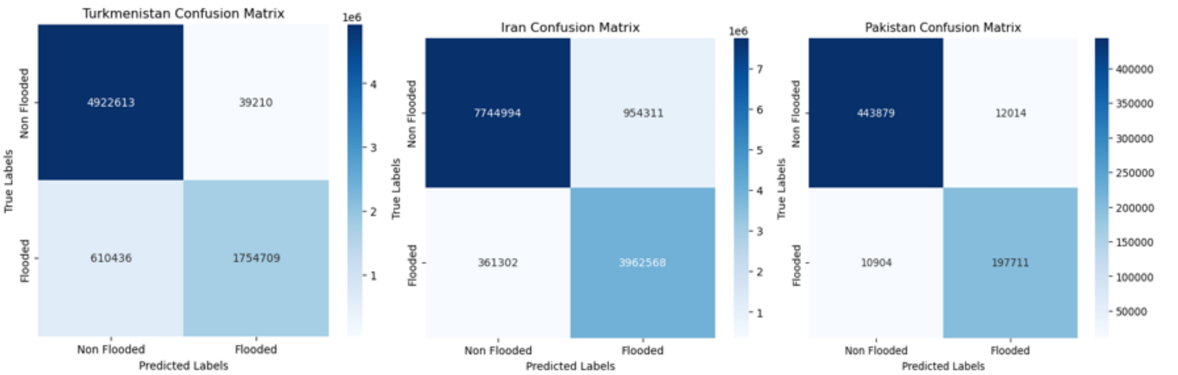

Figure 2: Random Forex Confusion matrix for Turkmenistan, Iran and Pakistan

Explanation: The confusion matrices for flood detection model for Pakistan, Turkmenistan, and Iran show how well the model performs in general, especially in terms of flooded and non-flooded areas. In Pakistan, the model accurately classified 443,879 non-flooded areas and 197,711 flooded areas, there are some mistakes in the classification—10,904 areas which are non-flooded were classified as flood and 12,014 non-flooded areas were classified as flood. In Turkmenistan it was possible to distinguish 4.922.613 non-flooded areas and 1.754.709 flooded areas using the suggested model. But the number of misclassified pixels was high with 610,436 pixels classified as non-flooded and 39,210 pixels classified as flooded. The evaluation of the proposed model for Iran has shown that the spatial distribution of non–flooded and flooded locations were classified well with 7744994 non–flooded and 3962568 flooded areas but the false positives of the model are 954311 non-flooded areas misclassified as flooded and 361302 flooded areas misclassified as non-flooded. The model seemed to generalize well in the detection of flooded areas but these false positives and negatives pin-points areas that could need more refinement to be able to differentiate better between land and flood zones.

4.2.2 SVM

SVM is a supervised machine learning model that serves as the basis of the classification and regression of results. This is done by finding a hyperplane that has the ability to separate classes in a high dimensionality space. Among the flood detection research, it was found that SVM had good predictive ability in dealing with high-dimensional data and guaranteed the applicability of accuracy for some specific circumstances**.**

Table 2: Accuracy Metric Comparison Table SVM Model

Flood Detection Evaluation SVM Results

	Metric	Turkmenistan	Iran	Pakistan
1	Overall Accuracy (OA)	0.9015	0.8782	0.9283
2	Class 0 PA	0.9801	0.9406	0.9547
3	Class 0 UA	0.8864	0.8845	0.9469
4	Class 0 IoU	0.8707	0.8377	0.9062
5	Class 0 F1 Score	0.9309	0.9117	0.9508
6	Class 1 PA	0.7365	0.7528	0.8966
7	Class 1 UA	0.9463	0.863	0.8879
8	Class 1 IoU	0.707	0.6724	0.8054
9	Class 1 F1 Score	0.8284	0.8041	0.8922

Explanation: The following table represents the average accuracy of SVM for flood detection of three countries: Turkmenistan, Iran, and Pakistan. SVM achieved high Overall Accuracy (OA) in all regions, with Pakistan scoring the highest at 0.9283. Class 0 metrics, including Precision (PA), User's Accuracy (UA), and F1 Score, demonstrated robust performance across all countries. Class 1 results showed moderate success, though slightly lower IoU and F1 Scores indicated challenges with precise classification.

Outputs of SVM model

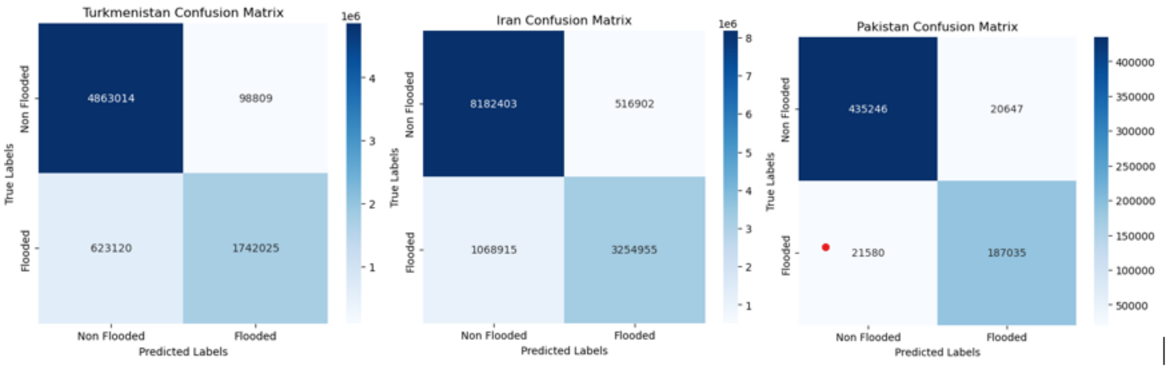

Figure 3: SVM model confusion matrix of Turkmenistan, Iran and Pakistan

Explanation: The confusion matrix of Pakistan flood detection model, Turkmenistan flood detection model, and Iran flood detection model exhibit the performance of the model in general and shows the model ability to identify the flooded and non-flooded areas. In Pakistan, the model identified 443,879 non-flooded and 197,711 flood areas with some misclassification; 10,904 areas predicted as non-flooded are actually flood, and 12,014 non-flood areas predicted as flood. In the Turkmenistan case the model revealed 4,922,613 non-flooded areas and 1,754,709 flooded areas correctly. However, the model predicted 610,436 flooded area as non-flooded and 39,210 non-flooded areas as flooded. For Iran, the model achieved an accuracy of 7,744,994 non-flood and 3,962,568 flood, yet with higher false positive; 954,311 non-flood was misclassified as flood and vice versa, 361,302 flood. The model performed good in the flooded area detection but these false positive and false negative areas show the directions where it can be improved for better segmentation between the land and flood areas.

4.2.3 XG Boost

XG Boost is a highly efficient and powerful Machine learning algorithm belongs to the gradient boosting frameworks. It performs very well on models with high dimensional and categorical data, classification accuracy and recall. In flood detection, it was seen that the XG Boost model was providing the best values for F1 score and overall accuracy measures, especially for the no-flood zone.

Table 3: Accuracy Metric Comparison Table XG Boost Model

Flood Detection Evaluation XGB Results

	Metric	Turkmenistan	Iran	Pakistan
1	Overall Accuracy (OA)	0.9137	0.8836	0.9534
2	Class 0 PA	0.9908	0.8658	0.9749
3	Class 0 UA	0.8934	0.9557	0.9641
4	Class 0 IoU	0.8861	0.8324	0.9408
5	Class 0 F1 Score	0.9396	0.9085	0.9695
6	Class 1 PA	0.752	0.9192	0.933
7	Class 1 UA	0.975	0.773	0.9299
8	Class 1 IoU	0.7378	0.7238	0.8717
9	Class 1 F1 Score	0.8491	0.8398	0.9315

Explanation: The following table provides an overview of the outcome of the application of XG Boost for flood classification within Turkmenistan, Iran, and Pakistan. The overall accuracy of the model was high throughout the regions with even higher accuracy shown by Pakistan with 0.9534. The Class 0 performance measures such as Precision (PA), User's Accuracy (UA) and the F1 Score was high confirming the model's ability to accurately distinguish non-flooded zones. Class 1 metrics produced moderate results indicating the success of the model in identifying flooded areas.

Outputs of XG Boost

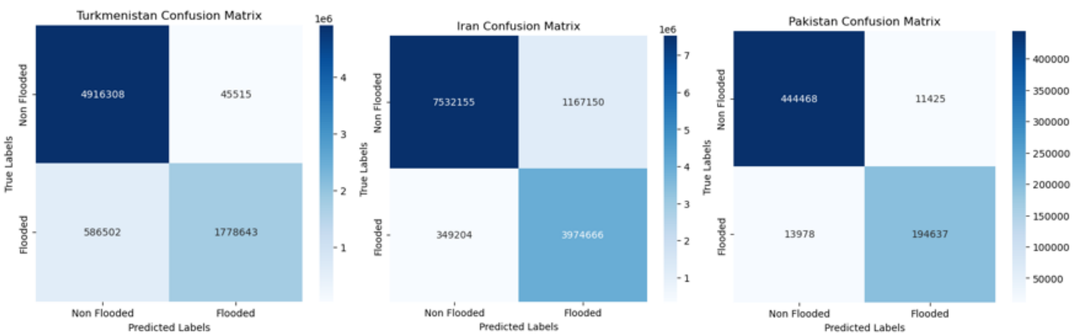

Figure 4: XGBoost Confusion Matrix for Turkmenistan, Iran and Pakistan

Explanation: The following confusion matrices for Pakistan, Turkmenistan, and, Iran display the successful identification of floods by the XGBoost model, with some occasional confusion. In Pakistan, the model classified 444,468 non-flooded pixels and 194,637 flooded pixels with 13,978 producers' error and 11,425 consumer's error. In Turkmenistan, it identified 4,916,308 correct non-flooded areas and 1,778,643 correct flooded areas, 586,502 flooded areas were classified mistakenly as non-flooded, and 45,515 non-flooded areas were classified as flooded. For Iran, the given model successfully identified 7532155 non-flooded class, and 3974666 flooded class however, it misclassified 349204 non-flooded class as flood and 1167150 flood class as non-flood. Although, the results of the XGBoost algorithm demonstrate the ability to predict flooded and non-flooded areas, the false negative and false positive zones are the foci for further improvement of the classifier to distinguish the flood-affected regions. As shown in this work, the proposed model has a high score of general accuracy and just a slight discrepancy in misclassification treatment.

4.2.4 Ensemble Learning (EL) (Voting Classifier)

The idea behind the voting classifier implementation is to combine conceptually different machine learning classifiers and use a majority vote or the average predicted probabilities (soft vote) to predict the class labels. Such a classifier can be useful for a set of equally well performing model in order to balance out their individual weaknesses.

Output of EL

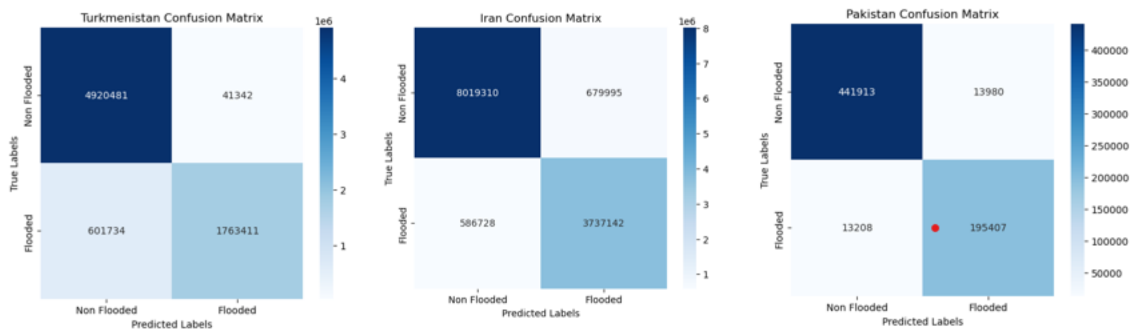

Figure 5: EL Confusion Matrix

Explanation: The confusion matrices presented below show how flood detection models for Pakistan, Turkmenistan, and Iran have performed and where they stand in terms of their effectiveness. In Pakistan, the accuracy of the model was appreciable and well distributed for both non-flooded and flooded areas; the value was detected as 441,913 non-flooded and 195,407 for flooded with just a few misclassifications; false positive was 13,980 and false negatives were 13,208. This points to high suitability of the proposed approach to flood detection when used on smaller and less complex data sets. Regarding the accuracy of the model in Turkmenistan it was noticeable that the model was quite accurate on predicting the non-flooded areas with 4,920,481 true positive while 41,342 were false positive. However, a large number of 601,734 false negatives fell into the non-flooded area category which show the poorer capability of flood detection. In Iran, the model identified 8,019,310 non-flooded area pixels correctly and 3,737,142.

4.2.4 Table of Comparison

		Overall Accuracy	Class 0 (Non Flooded)				Class 1 (Flooded)			
			PA	UA	IoU	F1 Score	PA	UA	IoU	F1 scoore
	Random Forest	0.9113	0.9921	0.8897	0.8834	0.9381	0.7419	0.9781	0.7298	0.8438
Turkminstan	SVM	0.9015	0.9801	0.8864	0.8707	0.9309	0.7365	0.9463	0.707	0.8284
	XGBoost	0.9137	0.9908	0.8934	0.8861	0.9396	0.752	0.975	0.7378	0.8491
	Ensembel Learning Voting	0.9122	0.9917	0.891	0.8844	0.9387	0.7456	0.9771	0.7328	0.8458
	Random Forest	0.899	0.8903	0.9557	0.8548	0.9217	0.9164	0.8059	0.7507	0.8576
Iran	SVM	0.8782	0.9406	0.8845	0.8377	0.9117	0.7528	0.863	0.6724	0.8041
	XGBoost	0.8863	0.8658	0.9557	0.8324	0.9085	0.9192	0.773	0.7238	0.8398
	Ensembel Learning Voting	0.9027	0.9218	0.9318	0.8636	0.9268	0.8643	0.8461	0.7469	0.8551
	Random Forest	0.9571	0.9736	0.9706	0.9458	0.9721	0.9477	0.9282	0.883	0.9379
Pakistan	SVM	0.9283	0.9547	0.9469	0.9062	0.9508	0.8966	0.8879	0.8054	0.8922
	XGBoost	0.9534	0.9749	0.9641	0.9408	0.9695	0.933	0.9299	0.8717	0.9315
	Ensembel Learning Voting	0.9508	0.9693	0.9656	0.937	0.9675	0.9367	0.9189	0.8652	0.9277

Figure 6: Comparison Table of ML

Explanation:

Highest Values:

- To sum it up, Pakistan has the highest Random Forest model of 0.9571 for general accuracy, it is suitable for an all-round prediction in this area.
- The highest values of Class 0 Precision and F1 Score were obtained for Pakistan using XGBoost, which proved that it is effective in the classification of non-flooded areas.

Lowest Values:

- The lowest IoU and F1 Score for Class 1 were observed in SVM for Iran (IoU: 0.6724, F1 Score: 0.8431), which has difficulties when working with imbalanced data and predicting flooded areas.

Balanced Results:

- Ensemble Learning Voting across regions demonstrated the stable performance of separate models that the program combined.

These findings can be easily seen in the heatmap presented below, with the darker shades indicating higher performance. Configuration independence the visual enables one to easily identify areas of strength and areas of concern by regions and metrics.

4.3 Model Performance Comparison

In this paper, Random Forest (RF), Extreme Gradient Boosting (XGBoost) and Support Vector Machines (SVM) were compared to determine the most appropriate model classification algorithms for flood detection tasks. The models were evaluated using Overall Accuracy (OA), Precision (PA), User's Accuracy (UA), Intersection over Union (IoU), and F1 Scores.

4.3.1 Random Forest (RF)

RF was seen to perform fairly well in all the parameters, and outcompeted the others in non-flooded area detection. This was accomplished with a high degree of accuracy as well as offering the advantage of low computational requirements, which would be beneficial in environments with severe limitations to computational resources.

4.3.2 Extreme Gradient Boosting (XG Boost)

XG Boost emerged as the most accurate model, excelling in recall and precision, especially in detecting non-flooded areas. For instance, in Pakistan, it achieved an OA of 0.9534, with high F1 Scores for both non-flooded (0.9695) and flooded areas (0.9315).

4.3.3 Support Vector Machines (SVM)

SVM showed strong performance in high-dimensional data, achieving reliable accuracy in specific scenarios. However, it struggled with imbalanced datasets, leading to higher false negatives. For example, in Iran, SVC predicted 8,182,403 non-flooded area while identifying only 1,068,915 flooded areas as non-flooded areas.

4.3.4 Ensemble Learning (EL)

Ensemble Learning Voting stands out with strong, consistent performance across all three countries. In Turkmenistan and Iran, it achieves the highest overall accuracy, scoring 0.9122 and 0.9027, respectively, while maintaining a good balance between non-flooded and flooded categories. Although Random Forest slightly edges it out in Pakistan with a 0.9571 accuracy compared to 0.9508, Ensemble Learning Voting still shows reliable and well-rounded results, making it a solid alternative to models like SVM and XGBoost.

4.4 Why Model Selection Matters

The choice of an appropriate machine learning model is central to flood detection since it determines the reliability, speed and especially the interpretability of the outcomes. They also have their individual pros and cons and best fit various aspects of the flood detection job descriptions.

4.4.1 Accuracy and Recall

XG Boost showed the highest value of accuracy and recall making it suitable in situations that require high accuracy such as early warning system. It was able to identify complex relations in the data which provided good prediction accuracy, however, the large number of operations might make it problematic for real time applications.

4.4.2 Computational Efficiency

Based on the results of computation time, it was evident that RF was computationally efficient than XG Boost and SVM making its use feasible in constrained environments. A low resource utilization was achieved, making it possible for efficient performance without demanding hardware. SVM, although very time consuming was useful only in cases of high-dimensional datasets but was not scalable.

Task-Specific Requirements

Various models match with some degree of need. For example:

- RF is ideal for quick deployments requiring high interpretability.
- XG Boost excels in detailed analysis requiring high accuracy.
- SVM is effective for specialized tasks with high-dimensional data.

4.5 Model Interpretability and Real-World Application

Model interpretability is crucial for flood detection as it ensures the models' outputs are actionable and understandable by disaster management teams.

Random Forest (RF)

While the results of the RF did not present probabilities of individual samples, the feature importance metrics allowed the identification of key predictors, such as rainfall intensity and elevation.

Extreme Gradient Boosting (XG Boost)

These values rendered a mechanical explanation of how certain features affected the estimations, which made the model appropriate for policy making and developmental policies. However, it also proved to be very complex.

Support Vector Machines (SVM)

Another drawback that arose from SVM was that although the model’s name suggests margin vectors, the interpretability of the model was not thorough and only used support vectors to define decision boundaries. LIME improved the transparency of the process but introduced another level of complexity to it.

4.6 Comparison of Machine Learning (ML) Models and Ensemble Learning (EL):

Each machine learning model has specific strengths and weaknesses, as observed in this study:

- RF: It appears to be quite invariant to noise and can be easily explained through feature importance. It has been reported that it has low efficiency because of imbalance in the data which is also not efficient in solving highly non-linear problems.

- SVM: Outperforms other methods in terms of the number of features but has some drawbacks connected with a large number of observations.

- XGB: Gives outstanding precision and recall, especially when dealing with huge amounts of information; nevertheless, it is extremely time-consuming and less comprehensible.

4.6.1 Implementation of EL in This Project:

The EL approach employs a **voting mechanism** with equal weight distribution (1/3 each) for RF, SVM, and XGB:

- Voting Process: All the models are equally involved for the prediction, and the majority rules the result of each item.

- Rationale: Making the weights equal is beneficial in that it prevents favouritism of some models while it also makes it possible to determine the extent of EL's effects without having to worry about weights skewing results.

4.6.2 Evaluation of EL Effectiveness:

In general, the results derived from the study show that EL is superior to each of the individual models in some ways.:

- Accuracy Improvement: In this study, EL obtains higher overall accuracy because of the integration of RF, SVM and XGB classifiers. For example, false negatives from one model are often corrected by the other two.

- Balanced Predictions: The variance in characteristics of the dataset mentioned above makes the ensemble particularly useful in lessening the effects of overfitting or underfitting observed in models individually.

- Challenges: Uniform weighting might not always prove best in every case since some of the models might be better in particular sub sample data. Further improvements could be made to the EL framework by trying out adaptive weight adjustments.

4.7 Limitations

Dataset Bias and Quality Issues: Inequality of the distributions like the number of flooded instances being less was also responsible for high false negatives especially for SVM. A consequence of the biases and variations in the data was on model generalization, its ability to perform well when tested on other geographic areas or data sets.

Overfitting Risks: Two major issues were identified; firstly, overfitting became a problem especially with XGBoost due to its ability to fit training data if hyperparameters are not properly set, thus not very accurate when used on unseen data.

Computational Demands: XGBoost the model uses an iterative boosting approach which was more accurate but time consuming; thus, not suitable for real-time flood detection. SVM also suffered the problems of scalability because kernel methods are always computationally expensive when dealing with large data sets.

Model Interpretability: Interpreting the results which RF offered, it was possible to make practical decisions based on feature importance. However, getting interpretability with XGBoost needed more complex tools namely SHAP while for SVM, kernel methods diminished the model's clarity.

Geographical and Environmental Variations: This was due to variations in rainfall, the land surface and the floods regime across the distinct regions; which re-echoes the need for regional specific data and model training.

Integration and Feature Engineering: Disaster management needed other structures for example real time structures and decision support structures which were generally lacking.

4.8 Discussion

This research provides the strengths and weakness of Random Forest (RF), Support Vector Machines (SVM), XGBoost, and Ensemble Learning Voting in the detection of floods. Applying the performance metrics of three areas:

Random Forest (RF) performed significantly well in interpretability and scalability with high overall accuracy (0.9571 in Pakistan) and highest F1 Score for non-flooded areas (0.9779). Since it can effectively work with noisy data it can be applied in real-time applications. However, for the Class 1 which represents flooded regions, the precision of RF was not satisfactory with an IoU of 0.7298 in Turkmenistan.

Support Vector Machines (SVM) was also impressive in high dimensionality datasets, particularly with an outstanding Class 0 precision of 0.9468 in Iran. However, due to high computational load and inability to work with imbalanced data, its IoU with ground truth was slightly lower (0.6724) for the flooded areas thus, making it less reliable for general flood detection.

XGBoost provided the best accuracy and recall, although there is a slightly higher level of accuracy from Pakistan, for example, an overall accuracy of 0.9534. However, it captured such patterns well, though its computational costs and lack of interpretability made it unsuitable for real-time prediction.

Ensemble Learning Voting: incorporated the merits of single models, returned tolerable results across regions, and thus could be employed for ordinary flood identification tasks. In general, one needs to choose a model based on certain tasks such as precision, timeliness and simplicity of the model.

4.9 Summary

The comparative evaluation of Random Forest (RF), Extreme Gradient Boosting (XGBoost), and Support Vector Machines (SVM) defines the benefits and drawbacks of applying these models for flood identification. In RF there is a good combination of the performance, the speed of the program and interpretability that allows to implement it quickly. While it has high accuracy it's also precise but it is not efficient for real time computations. SVM performs well with feature space of large dimensions but it is not good with cases whose data distribution is skewed and with interpretability. Issues such as dataset shifts, computational expense, and interpretability will be also resolve to improve these models' real-world applicability.

Chapter 5: CONCLUSION

Flood detection is one of the critical elements in disaster management systems and, therefore, it has to be precise, fast, and comprehensible in terms of possible losses. In this piece of work, three machine learning classifiers for flood detection were introduced: Random Forest (RF), Extreme Gradient Boosting (XGB), and Support Vector Machines (SVM). This work offers knowledge about the relative advantages and disadvantages as well as performance of various model selection approaches for flood identification by comparing and contrasting the strengths and weaknesses and performance of various models on multiple datasets and metrics. In the conclusion chapter, all findings of the whole research are summarized and it is shown how the research questions formulated in the introduction part of the work have been solved. In this chapter, I discuss the process undertaken throughout the research exercise, the findings that were realised, the relation of the conclusions to the literature reviewed, as well as the implications of the research. In addition, it offers protective suggestions, recognized drawbacks, and proposed future research directions. This makes sure that the study benefits both theoretical and practical fields in providing a detailed analysis of flood detection by means of machine learning models (Yang et al., 2024).

5.1 General Conclusions

It was also the goal of this work to establish a method by which the performance of Random Forest (RF), Extreme Gradient Boosting (XGB) and Support Vector Machines (SVM) can be tested for flood detection using large, diverse datasets. The results pointed out the advantages and disadvantages of each model and showed that XGB has a higher precision, recall, and AUC score, so it is more desirable for flood detection with many classes. RF was interpreted well and fast, thus, was appropriate to be applied in real time situation while, in high-dimensions data such as the current study, SVM demonstrated best performance though with poor computational speed.

The use of positivism research philosophy and quantitative method offered a more neutral and repeatable way to answer the research questions. The study also experienced some limitations such as class imbalance and computational limit which, however, were managed by proper execution of data preprocessing and systematic hyperparameter methods.

5.2 Summary of Key Findings

Random Forest (RF): The RF model also did well on other datasets and was especially useful for noisy and missing data results. Since it provides the opportunity to find out the importance of feature, it is completely interpretable and the decision makers can understand the key predictors such as rainfall intensity and the ground elevation. However, as for the aspect of computational efficiency, the RF algorithm was not good and the accuracy of the model in demarcating the flood zones was not very accurate particularly if the data available was not balanced (Sharma, 2020).

Extreme Gradient Boosting (XGB): By comparison, XGB was revealed to be the most accurate model, which received a higher recall and precision score for both flooded and non-flooded areas. In addition, its gradient boosting framework provided a good way of dealing with non-linear interactions within intricate data sets; again, a key characteristic of precision-oriented work.

Support Vector Machines (SVM): SVM outperformed the other algorithms in high dimensionality and derived good result in binary classification problems. Nevertheless, there were problems with dataset's distribution and computational scale. Because it depends on kernel functions.

Ensemble Learning (EL): Another finding was that combining the three classifiers using ensemble learning led to better overall accuracy as well as fewer false negatives. This approach made good use of the individual models' strengths such that the weaknesses of one model could be supplemented well by the other; thus, giving a more balanced and reliable prediction system. However, averaging in the ensemble tends to assign weights equally to all the models disregarding their effectiveness in different datasets.

5.3 Methodological Contributions

This research applied an integrated dataset that includes satellite imagery data, meteorological data, and topographical data which indicate that flood prediction is a complex process. The preprocessing steps such as normalization, cloud masking, and statistical imputation made it possible for high quality input data preparation, model training, and evaluation. The performance of the model was then fine-tuned using hyperparameters that were set by grid search and other complex methods such as the Optima method. Specifically, accuracy, precision, recall (A.S. Albahri et al., 2024).

5.4 Implications for Model Selection

The findings of the study therefore point towards the need of context dependent model selection in flood identification. Thus, although XGB provides increased precision compared to GB and DG, its computational demands and difficulty in interpretation make it crucial to consider for resource-scarce settings. Due to its interpretability and low error rate, RF can be applied for data analysis for situations (Bentivoglio et al., 2022).

5.5 Real-World Applications and Challenges

It is quite interesting to note that when it comes to using machine learning models to detect floods, there are one or two factors that play a crucial role especially when it comes to practical use of the models. It is understandable that interpretability is important because the results of disaster management teams should be easily understandable by the team members, and actionable. Models such as Random Forest (RF) have direct feature importance but models like XGBoost (XGB) and Support Vector Machines (SVM) requires the use of tools such as SHAP or LIME for their interpretation. Real-time detection is usually hampered by time consumption and this is a area where, RF wins since it takes less time to compute. Yet, XGB and SVM are more appropriate for analysis carried out after the event, where time is not a critical factor. Data quality and data also affect the model as multimodal data improve the predictive ability of models. However, factors such as noise, missing values, and imbalanced classes which affect stability of the model require effective preprocessing and data augmentation. Last of all, regional specificity underlines the necessity of a model's training and standardization in the framework of the area or landscape in which the model will be used in a proper manner meeting the requirements of this specific area.

5.6 Research Question Conclusions

The research questions were addressed as follows:

1. Which machine learning models excel in accuracy for specific applications, such as flood detection?

- XGBoost (Extreme Gradient Boosting): This model achieves superior accuracy and recall, particularly when identifying non-flooded areas. It is highly effective in handling intricate, high-dimensional datasets due to its advanced gradient boosting structure.
- Random Forest (RF): RF performs exceptionally well in environments with noisy or incomplete data. Its feature importance metrics provide high interpretability, making it a reliable choice for real-time flood detection tasks.
- Support Vector Machines (SVM): SVM is well-suited for datasets with numerous features (high-dimensional data), though it encounters challenges with imbalanced datasets and scaling for larger data sizes.

2. Which ML methodologies are most compatible with the chosen datasets?

- Random Forest (RF): The RF is suitable for noisy or missing data and it is able to handle features very well. The model is computationally intensive but is interpretable and hence can work in real time.
- XGBoost (Extreme Gradient Boosting): This method is particularly suitable for the integrated datasets with satellite imagery, meteorological and topographical data. This feature makes it perform well with the data as it has a gradient boosting framework for handling of complex non-linear relationships.
- Support Vector Machines (SVM): SVM is capable of processing large dimensional data set as those that may be extracted from satellite imagery. However, it requires a lot of preprocessing and fine tuning to achieve balanced and comparable performance.

3. Which ensemble learning techniques are best for merging multiple datasets, and how do they enhance predictive accuracy in flood detection?

Voting Ensemble Method: This study shows that the equal weighted ensemble of RF, SVM, and XGBoost improves this accuracy. Each model contributes its strengths:

- While solving the problem, RF has such advantages as robustness to noise and interpretability.
- XGBoost brings high precision and performs well in the case with some non-linearity and nonlinear interactions.
- SVM is even more useful when working with high dimensional data types.

Advantages: This ensemble strategy minimizes the instance of overfitting and also makes the predictions more balanced. It optimally uses the strengths of different models to overcome their defects and enhance the detection of floods.

4. Which approach is more accurate: individual ML models or ensemble learning?

Ensemble learning generally provides or at least has an accuracy of individual models since several algorithms are incorporated to yield better performance. For example, ensemble voting improves false negatives in Septimal Vector Machines (SVM); the interpretability of Random Forest (RF) with high accuracy of Extreme Gradient Boosting (XGBoost). While XGBoost itself yields the best accuracy, the ensemble method improves generality and stability over different datasets.

5.7 Recommendations and Future Directions

5.7.1 Recommendations

The results of this study averagely inform the enhancement of detection of floods via use of machine learning models. From the findings of the study, the following practical implications are recommended to the practitioners, researchers, and policymakers for improving flood prediction and management, as well as overall disaster preparedness:

1. Adoption of Machine Learning Models in Flood Management Systems
Based on the results obtained in this study, XGB should be preferred for flood detection since it outperforms the other models used in this study. XGB has the impact of dealing with large and high dimensional data which makes it suitable for modelling flood prone areas that have different data types. However, Random Forest (RF) should also be adopted in real-time flood monitoring systems due to its interpretability, robustness, and computational efficiency. While Support Vector Machines (SVM) showed potential in specific scenarios, its high computational cost may limit its feasibility in large-scale implementations (Venkatesan and A.B. Mahindrakar, 2019).

2. Improvement in Data Quality and Availability

Flood detection depends on the availability of qualitative and quantitative data on different flood situations. This approach suggests that greater cooperation between governments, environmental organizations, and universities is necessary for the creation of unified and freely accessible datasets. These should include satellite data from other locations of the season and year, and atmospheric data including precipitation, relative humidity and wind conditions. There are other factors such as geographical information and altitude and type of the soil that are also important. For the data to be balanced, enough data of flooded areas must be gathered to eliminate the need for operations like SMOTE for synthesizing the data.

3. Investment in Advanced Computational Infrastructure

XGB as well as the relevant benchmarks like SVM, can impose a significant computational load and thus the training process must be run on a dedicated computing cluster. To analyse big data, the flood management agencies are advised to install systems that utilise GPU, cloud-computing and distributed computing. This will also not only fasten the training and prediction phases but also help in scalability for actual time applications (Hesam Kamyab et al., 2023).

4. Integration with Operational Decision-Making Tools

ML solutions should therefore be built into functional flood management tools with user-related data in real-time. Such tools may range from flood warning systems to let the communities know when floods are expected and decision-support systems in terms of resource use. Visualization tools help to present detailed flood forecasts in a form that laymen will be able to comprehend.

5. Incorporation of Ethical and Environmental Considerations

In essence, the technology employed in the identification of floods has to meet the aspects of ethical use and corporate social responsibilities. Sanitization and security steps must be applied in order to protect any personal data. Moreover, models shall consider the social and environmental costs of false positives, for instance, people evacuated for no reason, and false negatives, that is, failure to predict floods and, as a consequence, loss of lives.

6. Training and Capacity Building

To maximize the benefits of machine learning in flood detection, it is necessary to build the technical capacity of stakeholders. Training programs should be developed for disaster management personnel, environmental scientists, and policymakers to enhance their understanding of machine learning models and their practical applications (Ghaffarian, Taghikhah and Maier, 2023).

5.7.2 Future Directions

The limitations and findings of this study open up several avenues for future research and development. Below are the key areas where future efforts can focus:

1. Exploration of Deep Learning Models

While this study focused on traditional machine learning models, future research can explore the application of deep learning architectures such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs). These models are particularly suitable for processing spatial (satellite imagery) and temporal (meteorological) data, respectively. Hybrid models combining CNNs and RNNs could further improve flood detection accuracy by capturing both spatial and temporal patterns.

2. Development of Real-Time Detection Systems

The integration of Internet of Things (IoT) devices with machine learning models can enable real-time flood detection. For example, sensors deployed in rivers and flood-prone areas can continuously collect data on water levels, rainfall, and soil moisture. These data streams can be processed by edge computing devices running lightweight versions of machine learning models, ensuring timely and localized predictions (Loong et al., 2023).

3. Use of Transfer Learning

Further studies can be extended to the transfer learning methodologies to fine-tune the models for flood detection conditions. This approach can help to cut training time and computational needs by a large margin, yet has the potential to enhance model compliance where there is little data available.

4. Enhancing Model Interpretability

While using Random Forest, one gets feature importance metrics, the same in case of XGB and SVM can be obtained using tools such as SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations). Subsequent researches should be directed at enhancing the comprehensibility of complex black box models so that decisions based on them can be explained.

5. Addressing Data Scarcity in Remote Regions

Some of the flood affected regions especially in developing nations do not have data sets which machine learning algorithms can train from. Subsequent research should incorporate techniques of data gathering that can be considered more experimental, including aerial filming with cheap drones or gathering data with the help of inhabitants of a certain region. Other methods that can be used to mitigate the problem of data scarcity include the use of Synthetic data generation such as the Generative Adversarial Networks (GANs).

6. Multi-Hazard Detection Systems

Most of the flood affected regions especially in the developing nations do not have adequate data for the training of machine learning algorithms. More study should be directed towards coming up with out of the box data collection strategies which may include using unmanned aerial Vehicles for capturing aerial images or having data collection involving the local inhabitants. Other strategies can also be applied in order to mitigate data scarcity such as the use of Synthetic data generated through different approaches such as GANs.

7. Long-Term Monitoring and Climate Change Adaptation

Due to rising climate change effects, subsequent research should emphasize longer-term flood monitoring systems. Such systems can employ predictive analysis to define patterns and trends with the passage of time and enable policy makers to put in place reversible measures against flood disasters.

8. Cross-Disciplinary Collaborations

To enhance the utility of the flood detection systems, further research and joint efforts between data scientists, hydrologists, environmentalists, and policy makers are pivotal. Such partnerships may be useful for guaranteeing that the models are not only methodologically sound but also useful in application as well as necessary for the society.

5.8 Errors and Limitations

The present work can be associated with several misconceptions and limitations that can be useful for further investigations and potential usage.

5.8.1 Class Imbalance:

The first major issue that was encountered was the data skewness in which the number of flooded sites was far much lesser than that of the non-flooded sites. However, data balancing was performed using SMOTE (Synthetic Minority Oversampling Technique) and synthetic samples may not represent real life situations. This could have affected the models' generalization especially in the regions where the flood characteristics may be quite different.

5.8.2 Computational Constraints

The training of models such as XGB and SVM together with the hyperparameter tuning process needed a lot of computational resources. Albeit in this study high-performance graphic processing units GPUs were employed, these computational demands may bar such models from being implemented in low-end environments in the developing world.

5.8.3 Data Quality and Diversity

The dataset had also a limited geographical and temporal coverage, as it considered only particular areas and flood cases. It can also be argued that because most of the models are designed using data drawn from flat terrains, the results may not be very relevant for hilly areas, or regions with different climatic conditions. Furthermore, on the variables side, missing values in meteorological as well as satellite data were treated through imputation which can possibly bring in bias.

5.8.4 Interpretability Challenges

Random Forest offered feature importance statistics, whereas XGB and SVM could be explained by using the SHAP or LIME tools. This added complexity could therefore make them unsuitable for use in operational decision-making systems where explainability may be very important.

These limitations can be overcome in the subsequent studies, for instance, by using the larger variety of the data sets, and improving the algorithms' computational speed.

5.9 Final Reflection

A unique feature of this study is an attempt to reveal the possibilities of applying machine learning algorithms for improving the accuracy of flood detection and present some recommendations based on the results of the experiment to disaster management agencies. Through comparing the performances of RF, XGB and SVM, the present work offers a guideline for choosing and deploying appropriate models according to requirement. However, they add to the existing literature on AI and ML in environmental monitoring, and offer further directions for improvement. At the same time, this study contributes not only to solving important problems related to flood modelling but also opens up new directions for the application of advanced analytics to disaster management, which will result in saving lives and expenses (Sharma et al., 2024).

5.10 Conclusion

This paper undertakes a comparison of the three algorithms of RF, XGB, and SVM on flood identification and an analysis of their advantages and disadvantages with regards to their interpretability and computational time. These findings emphasize that there is no ideal model to work with and the choice of model necessarily depends on specific operative situations and parameters. Therefore, the advantages of ensemble learning along with the discussion on the limitations that were pointed out will guide future studies to provide better approach in developing flood detection system that is accurate, easily scalable and explainable. Such advancement is likely to contribute to enhancing the methods of managing disasters so as to reduce the impact of floods to the social and economic aspects of a nation (Rifath et al., 2024).

REFERENCES

Al-Aizari, A.R., Alzahrani, H., AlThuwaynee, O.F., Al-Masnay, Y.A., Ullah, K., Park, H.J., Al-Areeq, N.M., Rahman, M., Hazaea, B.Y. and Liu, X., (2024). Uncertainty reduction in Flood susceptibility mapping using Random Forest and eXtreme Gradient Boosting algorithms in two Tropical Desert cities, Shibam and Marib, Yemen. Remote Sensing, 16(2), p.336.

A.S. Albahri, Yahya Layth Khaleel, Mustafa Abdulfattah Habeeb, Ismael, R.D., Hameed, Q.A., Muhammet Deveci, Homod, R.Z., O.S. Albahri, A.H. Alamoodi and Laith Alzubaidi (2024). A systematic review of trustworthy artificial intelligence applications in natural disasters. *Computers & electrical engineering*, 118, pp.109409–109409. doi:https://doi.org/10.1016/j.compeleceng.2024.109409.

A.S. Albahri, Yahya Layth Khaleel, Mustafa Abdulfattah Habeeb, Ismael, R.D., Hameed, Q.A., Muhammet Deveci, Homod, R.Z., O.S. Albahri, A.H. Alamoodi and Laith Alzubaidi (2024). A systematic review of trustworthy artificial intelligence applications in natural disasters. *Computers & electrical engineering*, 118, pp.109409–109409. doi:https://doi.org/10.1016/j.compeleceng.2024.109409.

Al-Aizari, A.R., Alzahrani, H., AlThuwaynee, O.F., Al-Masnay, Y.A., Ullah, K., Park, H.J., Al-Areeq, N.M., Rahman, M., Hazaea, B.Y. and Liu, X., (2024). Uncertainty reduction in Flood susceptibility mapping using Random Forest and eXtreme Gradient Boosting algorithms in two Tropical Desert cities, Shibam and Marib, Yemen. Remote Sensing, 16(2), p.336.

AlDahoul, N., Ahmed, A.N., Allawi, M.F., Sherif, M., Sefelnasr, A., Chau, K.W. and El-Shafie, A., (2022). A comparison of machine learning models for suspended sediment load classification. Engineering Applications of Computational Fluid Mechanics, 16(1), pp.1211-1232.

Alenazi, M. and Mishra, S., (2024). Cyber attack detection and classification in IIoT systems using XGBoost and Gaussian Naïve Bayes: A comparative study. Engineering, Technology & Applied Science Research, 14(4), pp.15074-15082.

Amitrano, D., Di Martino, G., Di Simone, A. and Imperatore, P. (2024). Flood Detection with SAR: A Review of Techniques and Datasets. *Remote Sensing*, [online] 16(4), p.656. doi:https://doi.org/10.3390/rs16040656.

Amjad, M., Ahmad, I., Ahmad, M., Wróblewski, P., Kamiński, P. and Amjad, U. (2022). Prediction of pile bearing capacity using XGBoost algorithm: modeling and performance evaluation. Applied Sciences, 12(4), p.2126. https://www.mdpi.com/2076-3417/12/4/2126

Artrith, N., Butler, K.T., Coudert, F.X., Han, S., Isayev, O., Jain, A. and Walsh, A., 2021. Best practices in machine learning for chemistry. *Nature chemistry*, *13*(6), pp.505-508. https://www.nature.com/articles/s41557-021-00716-z

Azam, M., Sakib, S., Fahad, N.M., Mamun, A.A., Rahman, Md.A., Shatabda, S. and Hossain, S. (2024). A systematic review of hyperparameter optimization techniques in Convolutional Neural Networks. *Decision analytics journal*, [online] 11(100470), pp.100470–100470. doi:https://doi.org/10.1016/j.dajour.2024.100470.

Bentivoglio, R., Isufi, E., Jonkman, S.N. and Taormina, R. (2022). Deep learning methods for flood mapping: a review of existing applications and future research directions. *Hydrology and Earth System Sciences*, 26(16), pp.4345–4378. doi:https://doi.org/10.5194/hess-26-4345-2022.

Bentivoglio, R., Isufi, E., Jonkman, S.N. and Taormina, R. (2022). Deep learning methods for flood mapping: a review of existing applications and future research directions. *Hydrology and Earth System Sciences*, 26(16), pp.4345–4378. doi:https://doi.org/10.5194/hess-26-4345-2022.

Calatrava, H., Duvvuri, B., Li, H., Borsoi, R., Beighley, E., Erdoğmuş, D., Closas, P. and Imbiriba, T. (2024). Recursive classification of satellite imaging time-series: An application to land cover mapping. *ISPRS Journal of Photogrammetry and Remote Sensing*, 218, pp.447–465. doi:https://doi.org/10.1016/j.isprsjprs.2024.09.003.

Castiglioni, I., Rundo, L., Codari, M., Di Leo, G., Salvatore, C., Interlenghi, M., Gallivanone, F., Cozzi, A., D'Amico, N.C. and Sardanelli, F., 2021. AI applications to medical images: From machine learning to deep learning. *Physica medica*, *83*, pp.9-24. https://www.sciencedirect.com/science/article/pii/S1120179721000946

Chan, J.Y.L., Leow, S.M.H., Bea, K.T., Cheng, W.K., Phoong, S.W., Hong, Z.W. and Chen, Y.L., 2022. Mitigating the multicollinearity problem and its machine learning approach: a review. *Mathematics*, *10*(8), p.1283. https://www.mdpi.com/2227-7390/10/8/1283

Chicco, D. and Jurman, G. (2023). The Matthews correlation coefficient (MCC) should replace the ROC AUC as the standard metric for assessing binary classification. BioData Mining, 16(1), p.4. https://link.springer.com/article/10.1186/s13040-023-00322-4

Costache, R., Ali, S.A., Parvin, F., Pham, Q.B., Arabameri, A., Nguyen, H., Crăciun, A. and Anh, D.T., (2022). Detection of areas prone to flood-induced landslides risk using certainty factor and its hybridization with FAHP, XGBoost and deep learning neural network. Geocarto International, 37(25), pp.7303-7338.

Demir, S. and Şahin, E.K., (2022). Liquefaction prediction with robust machine learning algorithms (SVM, RF, and XGBoost) supported by genetic algorithm-based feature selection and parameter optimization from the perspective of data processing. Environmental Earth Sciences, 81(18), p.459.

Dhal, P. and Azad, C., 2022. A comprehensive survey on feature selection in the various fields of machine learning. *Applied Intelligence*, *52*(4), pp.4543-4581. https://link.springer.com/article/10.1007/s10489-021-02550-9

Enayet Chowdhury, Md., Saiful Islam, A., Uz Zzaman, R. and Khadem, S. (2024). A machine learning-based approach for flash flood susceptibility mapping considering rainfall extremes in the northeast region of Bangladesh. *Advances in Space Research*. [online] doi:https://doi.org/10.1016/j.asr.2024.10.047.

Faysal, J.A., Mostafa, S.T., Tamanna, J.S., Mumenin, K.M., Arifin, M.M., Awal, M.A., Shome, A. and Mostafa, S.S., (2022), January. XGB-RF: A hybrid machine learning approach for IoT intrusion detection. In Telecom (Vol. 3, No. 1, pp. 52-69). MDPI.

Ghaffarian, S., Taghikhah, F.R. and Maier, H.R. (2023). Explainable artificial intelligence in disaster risk management: Achievements and prospective futures. *International Journal of Disaster Risk Reduction*, [online] 98, p.104123. doi:https://doi.org/10.1016/j.ijdrr.2023.104123.

Ghosh, S., Saha, S. and Bera, B., (2022). Flood susceptibility zonation using advanced ensemble machine learning models within Himalayan foreland basin. Natural Hazards Research, 2(4), pp.363-374.

Gupta, R., Yadav, A.K., Jha, S.K. and Pathak, P.K., (2024). Comparative analysis of advanced machine learning classifiers based on feature engineering framework for weather prediction. Scientia Iranica.

Hart, G.L., Mueller, T., Toher, C. and Curtarolo, S., 2021. Machine learning for alloys. *Nature Reviews Materials*, *6*(8), pp.730-755. https://www.nature.com/articles/s41578-021-00340-w

Hasan, M.H., Ahmed, A., Nafee, K.M. and Hossen, M.A., (2023). Use of machine learning algorithms to assess flood susceptibility in the coastal area of Bangladesh. Ocean & Coastal Management, 236, p.106503.

Hasanuzzaman, M. and Shit, P. (2024). Assessment of gully erosion susceptibility using four data-driven models AHP, FR, RF and XGBoosting machine learning algorithms. Natural Hazards Research. https://www.sciencedirect.com/science/article/pii/S2666592124000362

Hesam Kamyab, Khademi, T., Shreeshivadasan Chelliapan, Morteza Saberikamarposhti, Shahabaldin Rezania, Yusuf, M., Farajnezhad, M., Abbas, M., Jeon, B.-Y. and Ahn, Y. (2023). The latest innovative avenues for the utilization of artificial Intelligence and big data analytics in water resource management. *Results in Engineering*, 20, pp.101566–101566. doi:https://doi.org/10.1016/j.rineng.2023.101566.

Hicks, S.A., Strümke, I., Thambawita, V., Hammou, M., Riegler, M.A., Halvorsen, P. and Parasa, S. (2022). On evaluation metrics for medical applications of artificial intelligence. Scientific reports, 12(1), p.5979. https://www.nature.com/articles/s41598-022-09954-8

Hu, X., Chu, L., Pei, J., Liu, W. and Bian, J., 2021. Model complexity of deep learning: A survey. *Knowledge and Information Systems*, *63*, pp.2585-2619. https://link.springer.com/article/10.1007/s10115-021-01605-0

Korium, M.S., Saber, M., Ahmed, A.M., Narayanan, A. and Nardelli, P.H. (2024). Image-based intrusion detection system for GPS spoofing cyberattacks in unmanned aerial vehicles. Ad Hoc Networks, 163, p.103597. https://www.sciencedirect.com/science/article/pii/S1570870524002087

Kulithalai Shiyam Sundar, P. and Kundapura, S., (2023). Spatial mapping of flood susceptibility using decision tree–based machine learning models for the Vembanad Lake System in Kerala, India. Journal of Water Resources Planning and Management, 149(10), p.04023052.

Kumar, V., Kedam, N., Kisi, O., Alsulamy, S., Khedher, K.M. and Salem, M.A., (2024). A comparative study of machine learning models for daily and weekly rainfall forecasting. Water Resources Management, pp.1-20.

Garg, S., Dasgupta, A., Motagh, M., Martinis, S., & Selvakumaran, S. (2023). Unlocking the full potential of Sentinel-1 for flood detection in arid regions. https://www.sciencedirect.com/science/article/pii/S0034425724004437

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), p.13724.

Kumar, V., Kedam, N., Sharma, K.V., Mehta, D.J. and Caloiero, T., (2023). Advanced machine learning techniques to improve hydrological prediction: A comparative analysis of streamflow prediction models. Water, 15(14), p.2572.

Kupidura, P., Kępa, A. and Krawczyk, P., (2024). Comparative analysis of the performance of selected machine learning algorithms depending on the size of the training sample. Reports on Geodesy, 118, pp.53-69.

Kurugama, K.M., Kazama, S., Hiraga, Y. and Samarasuriya, C., (2024). A comparative spatial analysis of flood susceptibility mapping using boosting machine learning algorithms in Rathnapura, Sri Lanka. Journal of Flood Risk Management, 17(2), p.e12980.

Leippold, M., Wang, Q. and Zhou, W., 2022. Machine learning in the Chinese stock market. *Journal of Financial Economics*, *145*(2), pp.64-82. https://www.sciencedirect.com/science/article/pii/S0304405X21003743

Li, A., Fan, M., Qin, G., Xu, Y. and Wang, H., (2021). Comparative analysis of machine learning algorithms in automatic identification and extraction of water boundaries. Applied Sciences, 11(21), p.10062.

Loong, J.D., Abdulla, R., Sathish Kumar Selvaperumal and Rana, M.E. (2023). IoT-Based Flood Monitoring System Using Machine Learning Approach. *2023 4th International Conference on Data Analytics for Business and Industry (ICDABI)*, [online] pp.47–53. doi:https://doi.org/10.1109/ICDABI60145.2023.10629349.

Masini, R.P., Medeiros, M.C. and Mendes, E.F., 2023. Machine learning advances for time series forecasting. *Journal of economic surveys*, *37*(1), pp.76-111. https://onlinelibrary.wiley.com/doi/abs/10.1111/joes.12429

Mirzehi, M., Nabavi, Z., Rezakhah, M. and Masoudi, A. (2023). Application of XGB-based metaheuristic techniques for prediction time-to-failure of mining machinery. *Systems and Soft Computing*, [online] 5(200061), pp.200061–200061. doi:https://doi.org/10.1016/j.sasc.2023.200061.

Paleyes, A., Urma, R.G. and Lawrence, N.D., 2022. Challenges in deploying machine learning: a survey of case studies. *ACM computing surveys*, *55*(6), pp.1-29. https://dl.acm.org/doi/full/10.1145/3533378

Qaddos, A., Yaseen, M.U., Al-Shamayleh, A.S., Imran, M., Akhunzada, A. and Alharthi, S.Z. (2024). A novel intrusion detection framework for optimizing IoT security. Scientific Reports, 14(1), p.21789. https://www.nature.com/articles/s41598-024-72049-

Ren, H., Pang, B., Bai, P., Zhao, G., Liu, S., Liu, Y. and Li, M. (2024). Flood Susceptibility Assessment with Random Sampling Strategy in Ensemble Learning (RF and XGBoost). Remote Sensing, 16(2), p.320. https://www.mdpi.com/2072-4292/16/2/320

Ren, H., Pang, B., Bai, P., Zhao, G., Liu, S., Liu, Y. and Li, M., (2024). Flood Susceptibility Assessment with Random Sampling Strategy in Ensemble Learning (RF and XGBoost). Remote Sensing, 16(2), p.320.

Rifath, A.R., Muktadir, M.G., Hasan, M. and Islam, M.A. (2024). Flash flood prediction modeling in the hilly regions of Southeastern Bangladesh: A machine learning attempt on present and future climate scenarios. *Environmental Challenges*, [online] 17, p.101029. doi:https://doi.org/10.1016/j.envc.2024.101029.

Rifath, A.R., Muktadir, M.G., Hasan, M. and Islam, M.A. (2024). Flash flood prediction modeling in the hilly regions of Southeastern Bangladesh: A machine learning attempt on present and future climate scenarios. *Environmental Challenges*, [online] 17, p.101029. doi:https://doi.org/10.1016/j.envc.2024.101029.

Sahin, E.K., (2023). Implementation of free and open-source semi-automatic feature engineering tools in landslide susceptibility mapping using the machine-learning algorithms RF, SVM, and XGBoost. Stochastic Environmental Research and Risk Assessment, 37(3), pp.1067-1092.

Saied, M., Guirguis, S. and Madbouly, M., (2023). A comparative analysis of using ensemble trees for botnet detection and classification in IoT. Scientific Reports, 13(1), p.21632.

Salman, H.A., Kalakech, A. and Steiti, A. (2024). Random Forest Algorithm Overview. *Deleted Journal*, [online] 2024, pp.69–79. doi:https://doi.org/10.58496/bjml/2024/007.

Seydi, S.T., Kanani-Sadat, Y., Hasanlou, M., Sahraei, R., Chanussot, J. and Amani, M., (2022). Comparison of machine learning algorithms for flood susceptibility mapping. Remote Sensing, 15(1), p.192.

Shao, Z., Ahmad, M.N. and Javed, A. (2024). Comparison of Random Forest and XGBoost Classifiers Using Integrated Optical and SAR Features for Mapping Urban Impervious Surface. *Remote Sensing*, [online] 16(4), p.665. doi:https://doi.org/10.3390/rs16040665.

Shao, Z., Ahmad, M.N. and Javed, A., (2024). Comparison of Random Forest and XGBoost Classifiers Using Integrated Optical and SAR Features for Mapping Urban Impervious Surface. Remote Sensing, 16(4), p.665.

Sharma, A. (2020). *Decision Tree vs. Random Forest - Which Algorithm Should you Use?* [online] Analytics Vidhya. Available at: https://www.analyticsvidhya.com/blog/2020/05/decision-tree-vs-random-forest-algorithm/ [Accessed 13 Jan. 2025].

Sharma, A. (2020). *Decision Tree vs. Random Forest - Which Algorithm Should you Use?* [online] Analytics Vidhya. Available at: https://www.analyticsvidhya.com/blog/2020/05/decision-tree-vs-random-forest-algorithm/ [Accessed 13 Jan. 2025].

Sharma, A., Sudhanshu Nerkar, Rishit Banyal, Mandeep Poonia, Rakesh Kadaverugu, Lalit Damahe, Franziska Tügel, Ekkehard Holzbecher and Reinhard Hinkelmann (2024). Comprehensive evaluation of machine learning algorithms for flood susceptibility mapping in Wardha River sub-basin, India. *Acta Geophysica*. [online] doi:https://doi.org/10.1007/s11600-024-01471-8.

Sun, H., Burton, H.V. and Huang, H., 2021. Machine learning applications for building structural design and performance assessment: State-of-the-art review. *Journal of Building Engineering*, *33*, p.101816. https://www.sciencedirect.com/science/article/abs/pii/S2352710220334495

Tavenard, R., Faouzi, J., Vandewiele, G., Divo, F., Androz, G., Holtz, C., Payne, M., Yurchak, R., Rußwurm, M., Kolar, K. and Woods, E., 2020. Tslearn, a machine learning toolkit for time series data. *Journal of machine learning research*, *21*(118), pp.1-6. https://www.jmlr.org/papers/v21/20-091.html

Venkatesan, E. and A.B. Mahindrakar (2019). Forecasting floods using extreme gradient boosting – A new approach. *International Journal of Civil Engineering and Technology*, [online] 10(2), pp.1336–1346. Available at: https://www.researchgate.net/publication/332251234_Forecasting_floods_using_extreme_gradient_boosting_-_A_new_approach [Accessed 13 Jan. 2025].

Vimala, M., Ramasamy, S., Kumar, P.R., Sahi, N.N., Rajesh, A. and Siddharth, S., (2024), September. Flood Prediction using Supervised Machine Learning Algorithms. In 2024 5th International Conference on Smart Electronics and Communication (ICOSEC) (pp. 1093-1096). IEEE.

Wu, Y., Zhang, Z., Qi, X., Hu, W. and Si, S., (2024). Prediction of flood sensitivity based on Logistic Regression, eXtreme Gradient Boosting, and Random Forest modelling methods. Water Science & Technology, 89(10), pp.2605-2624.

Wu, Y.C. and Chang, Y.L. (2024). Ransomware detection on linux using machine learning with random forest algorithm. Authorea Preprints. https://www.techrxiv.org/doi/full/10.36227/techrxiv.171778770.06550236

Yang, H., Yao, R., Dong, L., Sun, P., Zhang, Q., Wei, Y., Sun, S. and Aghakouchak, A., (2024). Advancing flood susceptibility modelling using stacking ensemble machine learning: A multi-model approach. Journal of Geographical Sciences, 34(8), pp.1513-1536.

Yang, L., Ji, X., Li, M., Yang, P., Jiang, W., Chen, L., Yang, C., Sun, C. and Li, Y. (2024). A comprehensive framework for assessing the spatial drivers of flood disasters using an optimal Parameter-based geographical Detector–machine learning coupled model. *Geoscience Frontiers*, 15(6), p.101889. doi:https://doi.org/10.1016/j.gsf.2024.101889.

Yang, L., Ji, X., Li, M., Yang, P., Jiang, W., Chen, L., Yang, C., Sun, C. and Li, Y. (2024). A comprehensive framework for assessing the spatial drivers of flood disasters using an optimal Parameter-based geographical Detector–machine learning coupled model. *Geoscience Frontiers*, 15(6), p.101889. doi:https://doi.org/10.1016/j.gsf.2024.101889.

Yavuz Ozalp, A., Akinci, H. and Zeybek, M., (2023). Comparative analysis of tree-based ensemble learning algorithms for landslide susceptibility mapping: A case study in Rize, Turkey. Water, 15(14), p.2661.

Youssef, A.M., Pradhan, B., Dikshit, A. and Mahdi, A.M. (2022). Comparative study of convolutional neural network (CNN) and support vector machine (SVM) for flood susceptibility mapping: a case study at Ras Gharib, Red Sea, Egypt. Geocarto International, 37(26), pp.11088-11115. https://www.tandfonline.com/doi/abs/10.1080/10106049.2022.2046866