Detection of Rainfall-Runoff Using Support Vector Machine (SVM) for Sustainable Water Resource Management

Ramli I.[1,5,6*], Izzaty A.[3], Rusdiana S.[4] and Ferijal T.[2]

[1]Professor, Agricultural Engineering Department, Universitas Syiah Kuala, Indonesia

[2]Head Lecturer, Agricultural Engineering Department, Universitas Syiah Kuala, Indonesia

[3]Master Student, International Master Program on Natural Hazard Mitigation and Management, Department of Hydraulic and Ocean Engineering, National Cheng Kung University, Taiwan

[4]Head Lecturer, Departement of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Syiah Kuala, Indonesia

[5]Department of Environmental Master's Program, Universitas Syiah Kuala, Indonesia

[6]Research Center for Environmental and Natural Resources, Universitas Syiah Kuala, Indonesia

*ichwana.ramli@usk.ac.id

Abstract *One of the natural disasters frequently experienced is flooding, which is a consequence of climate change phenomena resulting from unstable rainfall patterns. Flood risk can be assessed based on rainfall, land use conditions, and river change density, which are crucial factors for flood-prone areas. These locations are often near coastlines, especially in urban areas situated in coastal zones. In the Krueng Pase Watershed, it has been found that land use near urban areas, such as wetlands, has increased annually. Land use modeling with an accuracy test of 83% has shown a continuous decrease in forest areas. This correlates with the population growth in city centers, leading to increased construction activities. Using a support vector machine based on machine learning, flood risk mapping was conducted around the Krueng Pase Watershed to monitor satellite images from Sentinel-2 through the Google Earth Engine (GEE) cloud platform. The results related to flood risk areas in the Krueng Pase Watershed show that in 2016, the area at risk was 527.4 hectares. By 2020, the flood risk area had sharply increased to 2,037.8 hectares, with many areas mapped as flood-affected. This increase was due to extensive construction in the city center and the reduction of forest function areas based on the 2020 land use change results. Furthermore, the Support Vector Machine (SVM) model observed that in 2023, the flood risk area in this study location decreased significantly to 919.5 hectares, showing a substantial reduction from the previous year. This indicates unpredictable climate change affecting rainfall patterns. Therefore, as seen in 2020, increased rainfall can significantly elevate the flood risk area, which requires careful monitoring and mapping for the Krueng Pase Watershed.*

Keywords: Flood Risk, Climate Change, Rainfall, Sentinel-2, Support Vector Machine (SVM).

Introduction

The study of water on Earth, including its presence, flow, and distribution, as well as its physical and chemical properties and interactions with the environment, particularly its relationships with living things, is the main emphasis of hydrology (Beven 2012). It also looks at the interactions that water has with the environment at each stage of the hydrologic cycle, including the connections between different elements like streamflow and precipitation (Yang, Li et al. 2015). For instance, modeling runoff is crucial for gaining a comprehensive understanding of how various changes affect hydrological phenomena (Xu 2002). Runoff models illustrate the effects of modifications in impervious surfaces, vegetation, and weather conditions on water systems (Devia, Ganasri and Dwarakish 2015). Surface runoff happens when rainwater cannot penetrate the soil due to its saturation, causing the water to flow across the land surface into surface water bodies like rivers, streams, reservoirs, and lakes. The advancement of information for detecting or predicting runoff is related to the intensity of precipitation occurring in a given area. Runoff modeling is employed to gain a deeper understanding of watershed productivity and responses, as well as to forecast water availability, monitor changes over time, and predict extreme events such as floods and droughts. The variation in rainfall intensity is another concurrent aspect pertaining to runoff and its evolution. A wide range of natural events may result from this. Certain locations may flood during extended rainy seasons due to climate change's unexpected effects on rainfall patterns. The correlation between rainfall and the likelihood of flood discharge is significant from both practical and theoretical standpoints. The design storm method, commonly used in engineering, involves estimating a hydrograph with a specific peak discharge probability based on a synthetic rainstorm with the same probability, utilizing a rainfall-runoff model (Breinl, Lun et al. 2021); (Packman and Kidd 1980). Rainfall-runoff relationships are intricate hydrological phenomena affected by the temporal and spatial variability of watershed characteristics, uncertainties in rainfall patterns, and alterations in soil cover and morphological parameters.

Rainfall-runoff relationships are intricate hydrological phenomena affected by the temporal and spatial variability of watershed characteristics, uncertainties in rainfall patterns, and alterations in soil cover and morphological parameters (Adamowski 2013);

(Poff, Tokar and Johnson 1996). Since insufficient rainfall can result in droughts and excessive rainfall can create floods, rainfall is a crucial factor in agricultural planning and management worldwide. Disaster management, environmental sustainability, and the planning of water resources all depend on an understanding of rainfall patterns and behavior (Riza and Nuryadi 2023); (Manzanas, Amekudzi et al. 2014). Rainfall is a crucial component in water resource management. Stakeholders can plan for efficient water use, such as agricultural irrigation, urban water delivery, and hydropower generation, by understanding rainfall patterns. It's critical to predict high rainfall events that could cause flooding in order to reduce the risk of flooding. More access to measured rainfall data from satellite sensors and weather stations has been made possible by recent technological breakthroughs (Macdonald, Redfern et al. 2022). However, it might be difficult and time-consuming to manually analyze and categorize rainfall data into different groups. Appropriate classification of rainfall facilitates the creation of efficient early warning systems that aid in mitigating losses caused by flooding. Three main types of runoff models (RR) exist: physically-based, conceptual, and data-driven models (Devia, Ganasri and Dwarakish 2015). Data-driven models, often called empirical or black-box models, use machine learning techniques, statistical tools, and vast datasets of historical records to build RR correlations. In essence, they enable runoff prediction based on a collection of variables across time by statistically capturing the RR dynamics that dictate how a catchment reacts to rainfall (Kwon, Kwon and Han 2020); (Young, Liu and Wu 2017). Conceptual models, as opposed to data-driven models, depict hydrological processes through a network of interrelated systems at various temporal and spatial scales (Paik, Kim et al. 2005). On the other hand, mathematical equations that explain the interconnections of the hydrological processes involved in transforming rainfall into runoff are the source of physically based models.

Hydrological modeling has increasingly incorporated machine learning techniques over the past few decades, including fuzzy logic models, Artificial Neural Networks (ANN), Support Vector Machines (SVM), and other genetic algorithms, with numerous studies demonstrating the successful application of SVM in runoff prediction and hydrologic classification (Kwon, Kwon and Han 2020); (Granata, Gargano and De Marinis 2016). Data-driven modeling, which integrates well-established conceptual and physically-based models with machine learning, provides a robust framework for runoff modeling. Simultaneously, the use of satellite imagery to analyze geographical patterns, trends, and environmental impacts has gained significant attention, as high-resolution sensors enable more precise detection of environmental changes. The selection of satellite imagery depends on the specific needs of the research, though accuracy can vary due to factors such as data quality and sampling uncertainties (Alquraish and Khadr 2021); (Nijssen and Lettenmaier 2004). In regions with high rainfall and humid climates, tools that predict rainfall-runoff are crucial. Given the unpredictability of climate change, detecting flood events in river basin areas through SVM and Sentinel-2 data from Google Earth Engine (GEE) is an effective approach, helping to identify flood-affected areas. Previous parameters are then used to assess the extent of land impacted by flooding, providing valuable insights into areas affected by the rainfall-runoff phenomenon.

Literature Review

The biophysical characteristics of a river basin (DAS) are factors that influence the output of the hydrological cycle process within a river basin, meaning that river flow is heavily affected by the biophysical conditions of the basin. The analysis of the biophysical conditions of a river basin includes the analysis of soil type, land use, topographical conditions such as slope gradient and contour, and river water quality (Harisagustinawati et al., 2020). Within a river basin, the upstream and downstream regions are biophysically connected through the hydrological cycle in relation to the hydrological system. A river basin has specific characteristics closely related to key elements such as soil, land use, topography, slope gradient, and slope length. These factors influence the degree of evapotranspiration, infiltration, percolation, runoff, surface flow, groundwater content, and river flow in response to rainfall. The most easily obtainable and relatively unchanging parameters of a river basin are its geographic features and morphometric characteristics. The biophysical conditions of a river basin comprise a set of geospheric phenomena, consisting of interconnected physical and biological elements that functionally interact with one another. Slope length affects the flow of water on the land surface, resulting in more water flowing at higher speeds at the base of the slope compared to the upper part of the slope (Fadhil & Wilis, 2019). Elevation also plays a role in flood occurrence; the lower the area, the higher the flood potential, and vice versa. The higher the elevation, the safer it is from flood disasters (Darmawan et al., 2017). The greater the soil's water absorption or infiltration capacity, the lower the flood risk, and conversely, the lower the infiltration capacity, the higher the flood risk (Matondang, 2013). Rainfall and temperature conditions in a region influence the hydrological state, ultimately affecting flood discharge formation (Knighton et al., 2017). Fundamentally, floods are caused by abnormally high rainfall, overwhelming the drainage system, including rivers, streams, and artificial channels, leading to water overflow. Increased flood discharge is not only due to rainfall but also changes in land use. The higher the drainage density (Dd), the more efficient the drainage system, indicating greater surface runoff and less stored groundwater (Matondang, 2013). River density refers to the ratio of the total river length (in kilometers) to the area of the river basin (Harisagustinawati et al., 2020). In determining flood-prone areas based on physical characteristics, several key parameters are required, which will be compared using the scoring method.

The parameters used in this study for weighting include slope gradient, land elevation, soil type, rainfall, land use, and river density. Inappropriate land-use changes also contribute to making land surfaces impermeable, preventing rainwater from infiltrating the soil. Land-use changes that reduce soil porosity, such as from vegetative (agricultural) to non-vegetative (non-agricultural) use, or from open land to paved surfaces, lead to reduced water infiltration into the soil. According to a case study by Dammalage & Jayasinghe (2019), land-use changes in river basin areas are predicted to have a significant impact on flood inundation. Significant land-use changes over the years, along with a sharp increase in built-up areas and a decrease in agricultural and green areas, have been the main contributors to the rising flood inundation incidents. According to research by Pattison & Lane (2012), to understand how local land management effects on runoff are distributed through drainage networks to downstream settlements, the spatial distribution of land management changes is crucial, as land use affects the volume and timing of runoff. Thus, the relative timing of each subcatchment's contribution to the main channel, which influences water volume at specific tributary peaks in relation to the main channel, plays a key role in how localized runoff changes are amplified at the catchment outlet.

Support Vector Regression (SVR) is an extension of SVM for regression cases. The goal of SVR is to find a function as a hyperplane in $f(x)$ the form of a regression function that fits all input data with a certain error (ε) while making the function as flat as possible. (Scholkopf & Smola, 2002). Abe (2005) The purpose of SVR is to map the input vectors into a higher-dimensional space. For instance, the following function represents the regression line as the optimal hyperplane:

$$f(x) = \mathbf{w}^T \varphi(x) + b \tag{1}$$

The SVM modeling in this study uses the regression line equation as the optimal *hyperplane*:

$$f(x) = \mathbf{w}^T \varphi(x) + b \tag{1.2}$$

The function of the kernel:

1. Kernel Linier

$$\varphi(x) = K(x, x') = x^T x \tag{2}$$

2. Kernel Polynomial

$$\varphi(x) = K(x, x') = (\gamma(x^T x) + 1)^d \tag{3}$$

3. Radial Basis Function (RBF)

$$\varphi(x) = K(x, x') = \exp\left(-\gamma \left\| x - x_i \right\|^2\right) \tag{4}$$

Based on the data forecasting error criterion, the optimal model can be chosen. A better degree of precision is indicated by a smaller or minimal error value. The model that performs best on the testing data or the out-of-sample data is selected as the best model. The error rate between the outcomes of two model experiments is determined using the Root Mean Square Error (RMSE). As the aim of forecasting is to produce predictions with the least amount of error, the best model is determined by comparing its RMSE value to that of the other models. The following formula can be used to determine the RMSE value (Wei, 2006).

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{M} \sum_{l=1}^{M} \left(Z_{n+l} - \hat{Z}_n(l) \right)^2} \tag{4.41}$$

With M are the possibilities, Z_{n+l} is the actual data dan $\hat{Z}_n(l)$ adalah prediction data.

Methodology

The process of predicting floods with Support Vector Machine (SVM) entails a classification method in which the algorithm is trained to distinguish between flood and non-flood situations using input features. SVM is frequently used in hydrological modeling in conjunction with historical flood records, environmental factors, and satellite data (like Sentinel-2). In order to divide data points into places that are likely to experience floods and those that are not, the SVM algorithm finds the best hyperplane. In the training phase, the algorithm learns patterns and correlations among the input variables by utilizing labeled datasets as flood and non-flood examples. After training, the SVM model uses the features it has discovered to forecast the occurrence of floods in fresh data. This method works especially well for managing flood risks and can be included into real-time monitoring and spatial analytic systems for river basins, such as in Google Earth Engine (GEE). The SVM model is a powerful tool that can be utilized with Sentinel-2 data in Google Earth Engine (GEE) for predicting and detecting floods. Sentinel-2 offers high-resolution optical images, making it ideal for tracking changes in surface water, vegetation, and land cover all of which are essential for identifying flood events. Within GEE, Sentinel-2 satellite data, combined with additional information such as elevation, rainfall, and land use, can be leveraged to build a dataset for training the SVM classifier. The SVM model in GEE helps distinguish between flooded and non-flooded areas by evaluating spectral indices like the Normalized Difference Water Index (NDWI) or the Modified Normalized Difference Water Index (MNDWI).

a. Data Acquisition

The first step involves acquiring relevant Sentinel-2 imagery. Sentinel-2 offers high-resolution optical data, which is crucial for monitoring changes in surface water bodies. To begin, appropriate time frames and geographic locations, particularly flood-prone areas such as river basins, are selected. Preprocessing includes filtering out cloud cover using available cloud-masking techniques in GEE, as cloud interference can distort the analysis. This preprocessing ensures that only high-quality, usable images are retained. The focus is then placed on key regions of interest, such as low-lying areas or those adjacent to rivers, where flooding is likely to occur.

b. Feature Extraction

The crucial step that converts unprocessed satellite data into useful variables for the SVM model is feature extraction. Spectral indices that react to moisture content and the presence of water are essential for flood detection. The Normalized Difference Water Index (NDWI) and Modified NDWI (MNDWI) are two often used indices that compare reflectance in particular spectral bands to highlight water bodies. To improve the model's performance, other features like height, slope, and rainfall may also be included. Since water tends to accumulate in lower altitudes, slope and elevation, for example, can be used to better understand the role of terrain in flood risk. By offering thorough details on landscape elements, the combination of these features improves the SVM model's accuracy and resilience. Model Training: Use GEE's SVM classifier to train the model based on the labeled dataset.

c. Model Training and Detection

The SVM classifier in GEE is then trained using the labeled dataset. Since SVM is a supervised learning method, it needs labeled data in order to identify the boundaries of areas that are flooded and those that are not. In order to divide data points into two groups inside the feature space, the SVM algorithm looks for the best hyperplane. In order to increase the model's capacity to generalize to previously untested data, the objective is to maximize the margin between the classes. When it comes to classification tasks, where the model parameters like the feature set, labels, and kernel type (linear, radial, or polynomial) are supplied, the built-in functions of GEE make it simple to use SVM. During the training phase, the model is regularly given labeled samples to help it learn how to distinguish between safe and flood-prone places. This procedure could take some time, depending on the quantity of the dataset, but it is essential to guarantee accurate flood prediction.

d. Prediction: The SVM classifier is used to make predictions using fresh, unseen Sentinel-2 imagery when the model training is complete. Using the patterns it learnt during the training phase, the model uses the new data to classify each pixel as either flooded or non-flooded. The ability to provide these forecasts almost instantly makes it possible to detect floods in a timely manner. Using mapping technologies, which highlight flooded areas for simple identification, it is feasible to see the expected outcomes. To evaluate prediction accuracy, the model's outputs can also be verified using publicly available datasets or actual flood incidents. Retraining with fresh data or adding new features allows for ongoing model improvement, ensuring increased performance and flexibility over time.

Results and Discussion

Areas with high rainfall will have a greater influence on the occurrence of flooding, where the higher the rainfall in an area, the higher the potential for flooding. The rainfall data for the Krueng Pase watershed (DAS) in Figure 1, rainfall ranges from 1000-2000 mm/year.

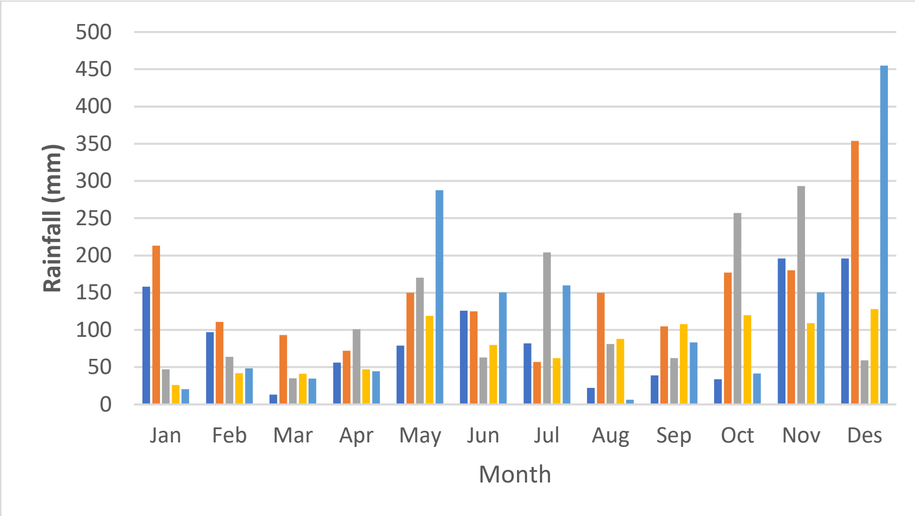

Figure 1. Rainfall in the Krueng Pase watershed

Rainfall greatly influences river water levels (Figure 1) apart from being influenced by the type of soil to absorb water and the slope of the land and land use. Rainfall and runoff data (in this study, water level is observed) is very important for sustainable management of clean water resources.

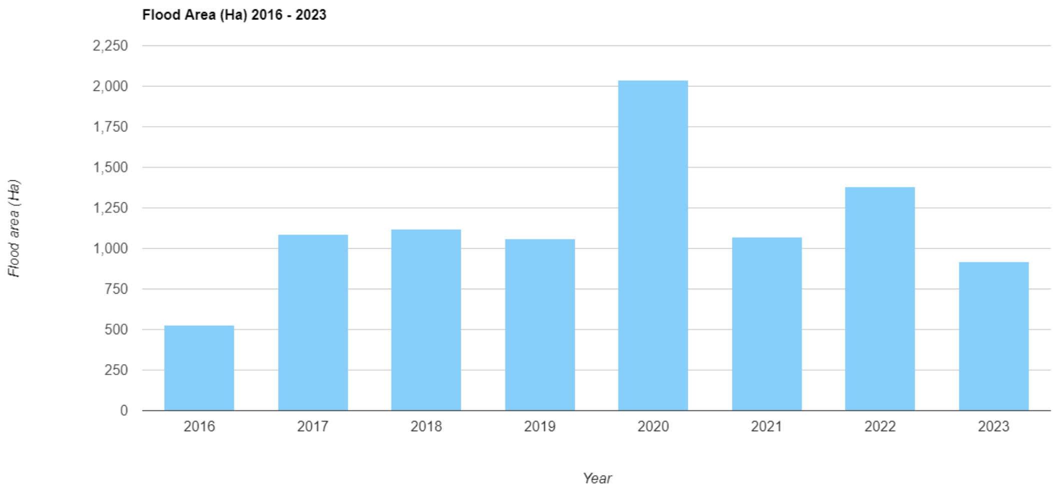

Figure 2. Flood Area Chart

Above is the distribution of flood-affected areas, showing the extent of the area in hectares each year. The analysis of flood risk areas in the Krueng Pase Watershed reveals that in 2016, the area at risk was 527.4 hectares. By 2020, this risk area had surged to 2,037.8 hectares, with numerous areas identified as flood-affected. This increase was attributed to extensive urban development in the city center and the loss of forested areas, as indicated by the 2020 land use changes. However, the Support Vector Machine (SVM) model indicated a significant reduction in flood risk areas to 919.5 hectares in 2023, reflecting a notable decrease from the previous year. This suggests that unpredictable climate changes are impacting rainfall patterns. Consequently, as observed in 2020, heightened rainfall can considerably increase the flood risk area, necessitating ongoing monitoring and mapping for the Krueng Pase Watershed.

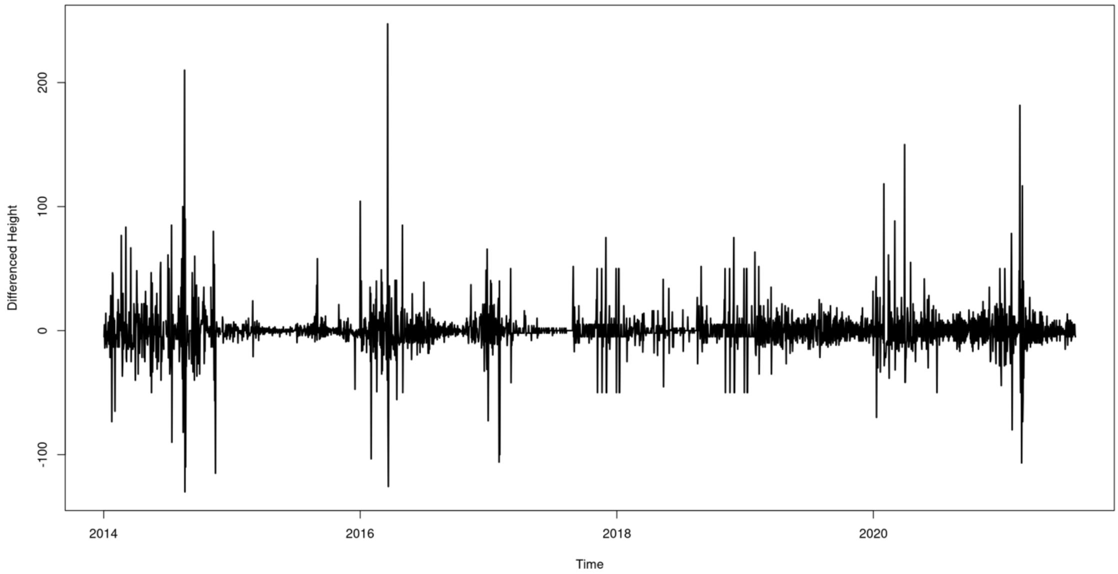

Figure 2 River water level in the downstream area of the Krueng Pase watershed

Based on map analysis and parameter scoring, an overlay of the physical vulnerability map is carried out. The parameters that are overlaid are rainfall, land use, soil type, slope slope, land height (elevation) and river density. The highest weight value is given to land use parameters because whether or not an area is prone to flooding is largely determined by the land use at that location, the more open the land is, the higher the potential for flooding. Followed by the parameters of elevation and river density which have the same weight because both have the same influence, where the lower the height of an area and the closer the distance of the area to the river, the greater the potential for flooding. The weight of rainfall is not too high but is quite influential in the problem of flooding, where the higher the rainfall the higher the probability of flooding, however the influence of the amount of rainfall on flooding will not apply to highland areas because there is little chance of causing flooding so the weight is not too high. Soil type and slope slope do not have a greater influence on flooding compared to other parameters, so the weight given is the lowest compared to other weights. Based on the map analysis and parameter scoring, the first overlay of the flood vulnerability map was conducted. The overlaid parameters include rainfall, land use, soil type, slope gradient, elevation, and river density. Each parameter is assigned a weight, rainfall (15%), slope (10%), soil type (10%), land use (25%), elevation

(20%) and river density 20%). After performing the overlay, the flood vulnerability map, as seen in Figure 1, was generated. This map, based on the overlay of the six parameters, classifies the area into five levels of vulnerability.

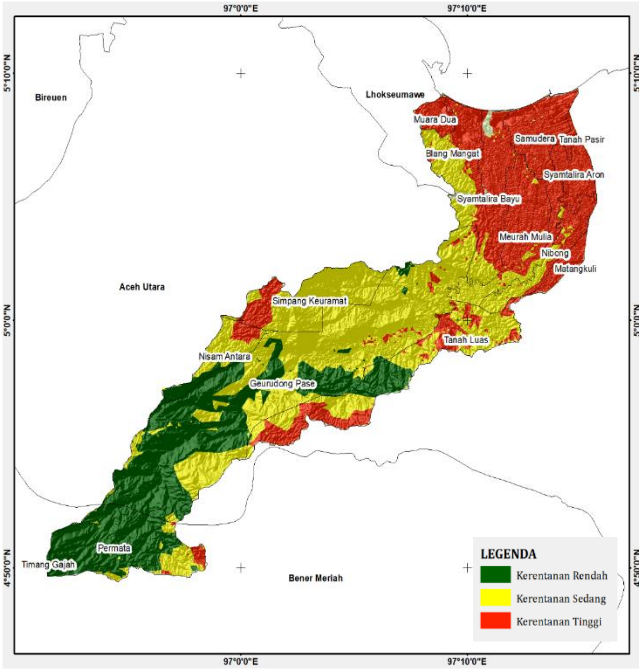

Figure 3. Physical Vulnerability Map of the Krueng Pase River Basin

The Pase River basin has blue-colored patches that are indicative of water, especially in the coastal parts, according to the NDWI results below. This results was separated between the NDWI with the land and vegetation part. According to this, regions shown in blue are thought to represent water bodies with values lower than 0. For the bodies of water, the maximum value is 0.7. The area that is tinted blue is a low-lying section of the coastal zone that is primarily made up of ponds and provides wet regions for the Pase River basin.

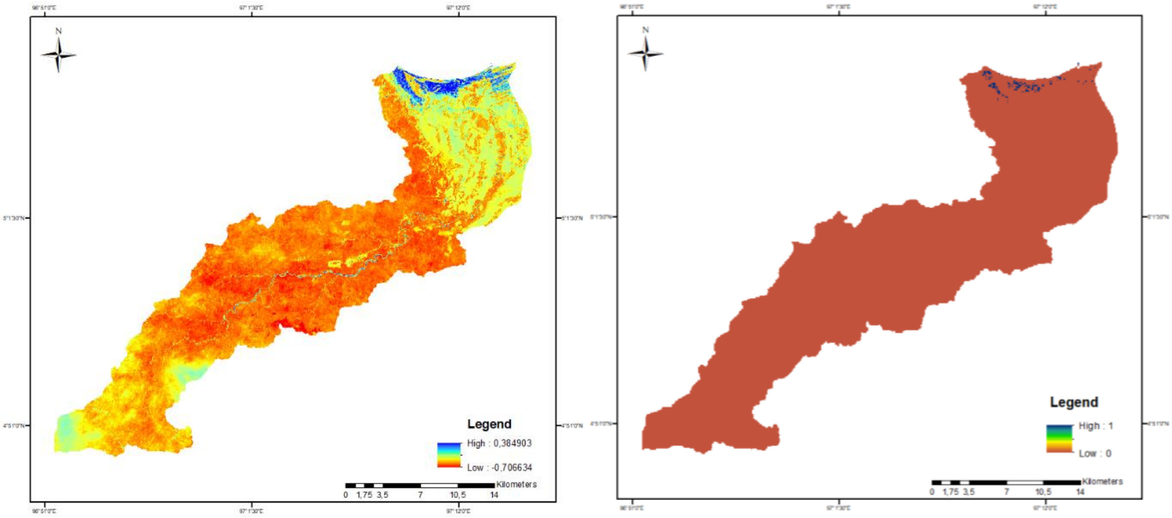

Figure 4. (a) NDWI Detection. (b) Permanent Water in Pase Basin

The NDWI analysis's findings show that river basins and permanent water bodies have quite different methods for detecting water. The NDWI readings in river basins varied greatly with the seasons and with variations in water levels brought on by precipitation, melting, or human activity. These variations demonstrate how dynamic river basins are due to the great variability of their water extent. The identification of flood-prone locations within the basin was made possible by the effectiveness of the NDWI in detecting transient water bodies, such as those created during flooding episodes. But places with a lot of riparian vegetation showed contradictory signals, therefore more precise thresholds were needed to distinguish between water and non-water features. Permanent water bodies showed long-term stability by maintaining high and consistent NDWI values. These water bodies appear to be less susceptible to short-term hydrological fluctuations or seasonal variations based on the stability of the NDWI readings. Furthermore, for permanent water bodies, the NDWI research can offer insights into possible alterations in water quality. Such as in Fig 2 (b), it shown that the high value is 1 which show the permanent water with the blue color.

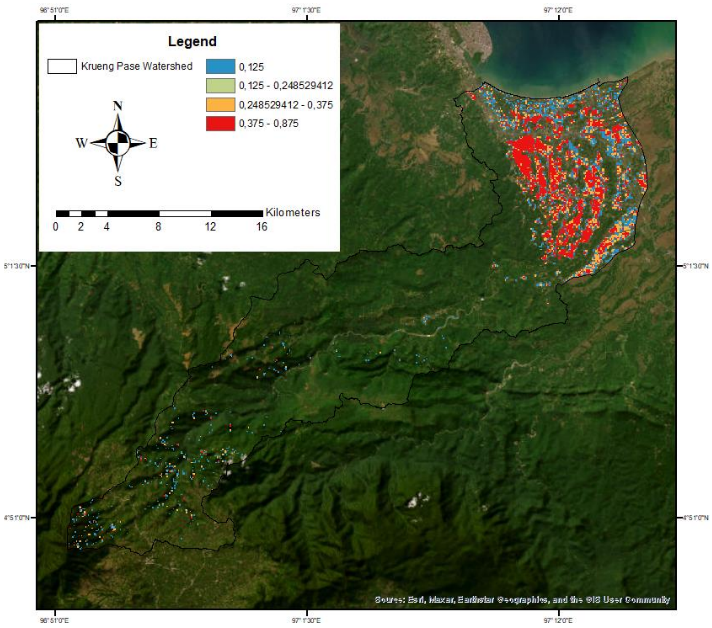

Figure 5. The Flood Maps Detection

The areas affected by floods are primarily found in low-elevation regions and in close proximity to coastal zones, as indicated by the flood maps created using the SVM model. The digital elevation models (DEM) employed in the investigation indicated that low-lying areas are more vulnerable to floods because of their propensity to accumulate water during periods of high rainfall or storm surges. Furthermore, being close to the coast increases the danger of flooding greatly because these places are more likely to experience floods because of their susceptibility to storm surges and sea level rise. These results highlight how crucial it is to take coastal proximity and elevation into account when assessing flood risk and developing management plans in order to properly address vulnerabilities. According to the results of the flood maps, four classes of flood-affected areas are indicated. The first classification represents areas with the lowest flood risk, with values ranging from 0 to 0.125. This is followed by areas colored light green with values from 0.125 to 0.240. Next, the orange-colored areas indicate regions with significant flood exposure, with values from 0.248 to 0.375. The highest flood risk is represented by red-colored areas, which are subject to overflow and runoff leading to flooding, with values from 0.375 to 0.875.

Finding and mapping flood-prone areas has been made easier with the use of the Support Vector Machine (SVM) model for flood detection in river basins. This method has shown to be reliable and efficient. The SVM model did a good job of differentiating between the water and non-water classes inside the river basin. It is renowned for its capacity to solve challenging classification tasks with high accuracy. The model's strength is its ability to work with big datasets, like Sentinel-2 satellite images, and its flexibility when it comes to different input parameters, like spectral indices like the Normalized Difference Water Index (NDWI) and other pertinent flood-related characteristics. By using this method, the SVM provided a dependable flood monitoring tool by precisely identifying flooding episodes and defining the regions most impacted by water inundation.

The SVM model is useful for regional and global flood risk assessments because of its efficacy in generalizing across various flood scenarios. The model exhibits great performance and robustness in complex contexts, as evidenced by the accuracy with which it detects flooded areas even in the presence of a variety of land cover types, including plant, urban infrastructure, and bare soil. Furthermore, the SVM model was able to accurately describe the dynamics of floods throughout time by capturing both transient and permanent water bodies. Integrating physical factors, such as digital elevation models (DEM) and land cover data, that are associated with flood susceptibility was a crucial component of this research. By taking topographical and environmental elements that affect flood risk into consideration, these parameters significantly improved the accuracy of the model. The SVM model was able to more accurately determine risky locations and measure flood susceptibility by integrating these physical characteristics. Nonetheless, significant restrictions were noted, especially when there was just a partial cloud cover and satellite imagery interference decreased the accuracy of flood detection. The quality and resolution of the input data were also critical to the SVM model's performance, highlighting the necessity for ongoing advancements in satellite sensor capabilities and preprocessing methods. Notwithstanding these difficulties, the SVM model is still a useful tool for flood detection since it has a number of advantages over more conventional approaches, including speed, scalability, and accuracy.

The RMSE value of 0.414136 shows that the average error in prediction of rainfall and runoff for sustainable clean water resource management from the SVM model to the actual value is very good. This shows that if the RMSE value approaches zero, the geometric correction will be better. The Support Vector Machine (SVM) method shows that the prediction model has a high level of accuracy and very low error. With small MSE and RMSE values, the SVM model is able to predict rainfall and water runoff with good precision, so that the difference between predicted results and actual data is very minimal. This accuracy is critical in sustainable water resource management, as precise predictions of rainfall and runoff enable more efficient water management. For example, with accurate predictions, authorities can better regulate water distribution for irrigation, manage reservoirs to avoid excess or shortage of water, and reduce the risk of flooding. SVM as a prediction method has been proven to be able to handle data complexity well, especially in dealing with non-linear patterns that are often encountered in hydrological predictions, thus providing a strong basis for long-term water resources planning and environmental sustainability.

Conclusion and Recommendation

a. Conclusion
    1. The SVM model did a good job of differentiating between the water and non-water classes inside the river basin. It is renowned for its capacity to solve challenging classification tasks with high accuracy. The SVM model was able to more accurately determine risky locations and measure flood susceptibility by integrating these physical characteristics. Nonetheless, significant restrictions were noted, especially when there was just a partial cloud cover and satellite imagery interference decreased the accuracy of flood detection.
    2. The quality and resolution of the input data were also critical to the SVM model's performance, highlighting the necessity for ongoing advancements in satellite sensor capabilities and preprocessing methods.

b. Recommendation

1. Adding More Hydrological and Physical Parameters although spectral indices like NDWI have shown to be useful, the predictive power of the model can be further increased by including additional hydrological parameters like soil moisture and precipitation data, as well as physical parameters like DEM. With a thorough grasp of flood dynamics and vulnerability provided by these inputs, the SVM is better equipped to distinguish between flooded and non-flooded locations and determine susceptibility based on topographical and environmental characteristics.
2. Model Validation and Calibration need to preserve accuracy over time, the SVM model must be calibrated on a regular basis using updated ground truth data. To verify the model's results and guarantee accurate flood detection, this entails incorporating in-situ flood observations, river gauge measurements, and other real-time hydrological data.

References

Adamowski, J. (2013). "Using support vector regression to predict direct runoff, base flow and total flow in a mountainous watershed with limited data in Uttaranchal, India." Annals of Warsaw University of Life Sciences-SGGW. Land Reclamation **45**(1).

Alquraish, M. M. and M. Khadr (2021). "Remote-Sensing-Based Streamflow Forecasting Using Artificial Neural Network and Support Vector Machine Models." Remote Sensing **13**(20): 4147.

Beven, K. J. (2012). Rainfall-runoff modelling: the primer, John Wiley & Sons.

Breinl, K., et al. (2021). "Understanding the relationship between rainfall and flood probabilities through combined intensity-duration-frequency analysis." Journal of Hydrology **602**: 126759.

Devia, G. K., et al. (2015). "A review on hydrological models." Aquatic procedia **4**: 1001-1007.

Granata, F., et al. (2016). "Support vector regression for rainfall-runoff modeling in urban drainage: A comparison with the EPA's storm water management model." Water **8**(3): 69.

Kwon, M., et al. (2020). "A hybrid approach combining conceptual hydrological models, support vector machines and remote sensing data for rainfall-runoff modeling." Remote Sensing **12**(11): 1801.

Macdonald, N., et al. (2022). "Understanding the impact of the built environment mosaic on rainfall-runoff behaviour." Journal of Hydrology **604**: 127147.

Manzanas, R., et al. (2014). "Precipitation variability and trends in Ghana: An intercomparison of observational and reanalysis products." Climatic change **124**: 805-819.

Nijssen, B. and D. P. Lettenmaier (2004). "Effect of precipitation sampling error on simulated hydrological fluxes and states: Anticipating the Global Precipitation Measurement satellites." Journal of geophysical research: Atmospheres **109**(D2).

Packman, J. and C. Kidd (1980). "A logical approach to the design storm concept." Water Resources Research **16**(6): 994-1000.

Paik, K., et al. (2005). "A conceptual rainfall-runoff model considering seasonal variation." Hydrological Processes: An International Journal **19**(19): 3837-3850.

Poff, N. L., et al. (1996). "Stream hydrological and ecological responses to climate change assessed with an artificial neural network." Limnology and Oceanography **41**(5): 857-863.

Riza, O. S. and A. Nuryadi (2023). "Bibliometric Study: Rainfall Classification-Prediction using Machine Learning Methods." Digital Zone: Jurnal Teknologi Informasi dan Komunikasi **14**(2): 206-218.

Xu, C.-y. (2002). "Hydrologic models." Textbooks of Uppsala University. Department of Earth Sciences Hydrology.

Yang, W.-Y., et al. (2015). "Saturation-excess and infiltration-excess runoff on green roofs." Ecological Engineering **74**: 327-336.

Young, C.-C., et al. (2017). "A physically based and machine learning hybrid approach for accurate rainfall-runoff modeling during extreme typhoon events." Applied Soft Computing **53**: 205-216.