MACHINE LEARNING OPTIMIZATION IN THE FLOOD PRONE AREA MAPPING SYSTEM

Zulham Sitorus[1]*, Eko Hariyanto[2], Fahmi Kurniawan[3]
[1,2,3] Computer System, Universitas Pembangunan Panca Budi, Indonesia

*Corresponding author: zulhamsitorus@dosen.pancabudi.ac.id

ARTICLE INFO

Date received : 28 Oct 2022
Revision date : 21 Nov 2022
Date received : 27 Nov 2022

Keywords:
Machine Learning, Optimization, Flood

ABSTRACT

One of the areas in Batu Bara Regency, Pahlawan Village, Tanjung Tiram District, has an area of 173.79 km² and is located in a lowland area with an altitude of 0.-4.5m which is directly adjacent to the Malacca Strait to the east. Where almost half of the area is affected by sea tides, Hero Village has a tropical climate with two seasons namely the rainy season and the dry season. The people who live in Pahlawan Village, Tanjung Tiram District. There are so many obstacles faced by the people of Pahlawan Village, including the problem of flooding which has an impact on the health and the economy of the community. Lack of counseling and knowledge, as well as public awareness of the occurrence of flooding during high tides, and when the rainy season will increase the water discharge at sea level will rise so that it can cause flooding. Due to the occurrence of floods, and the impact of losses that affect material and non-material, it is very important to map flood-prone areas for regional development planning. Identification of potential floods involves machine learning using the Random Forest method, taking into account the factors that trigger floods. The Random Forest method also provides sensitivity parameters using the Receiver Operating Characteristic (ROC) curve which shows flood-prone areas such as Pahlawan Village, Tanjung Tiram District.

INTRODUCTION

Floods and inundations are an annual problem and have a major impact on the condition of society both socially, economically, and environmentally. Floods are not a personal problem that is researched based on scientific disciplines, but floods are caused by damaged environmental systems and disrupted environmental physical chains so to overcome the problem of flooding need to be studied in an integrated manner. Flooding in Pahlawan Village, Tanjung Tiram District, Batu Bara Regency or other areas was caused by siltation of the water channel and close to the coastline, as well as accumulation of garbage in the waterways in Pahlawan Village. The increased flow rate in seawater and falling rainfall as a result of not/less absorption into the ground, so that rainfall becomes surface runoff. Surface runoff that moves on the ground surface erodes the soil and carries it into the river body, therefore the flow of the river not only increases the discharge but also adds material resulting from erosion. The big focus of this research is machine learning which can apply how to periodically detect complex patterns and make intelligent decisions based on data available in previous years. Machine learning can learn patterns from historical data contained in previous years' databases for mapping floods due to rainfall and tides from the sea, this can be done to find out how big the function of the flood mapping results is in the Pahlawan Village area. Irregular floods occur due to tsunamis, tidal waves, river overflows, or rob, (Niode et al, 2016).

Machine learning has the meaning of a method used to create programs that can learn from available data from several sources that can be trusted to use that data. In contrast to ordinary computer programs that are carried out statically because there are already provisions for what must be done first, however, it is different from machine learning programs, which are programs designed to be able to learn on their own with data that is already available.

Flooding is one of the stories that is often heard either in rumors or online news that currently exists, where it occurs during the rainy season with frequent, heavy, and relatively long-term water discharge intensities, areas that are customers Floods during the rainy season are around river currents that actively flow from upstream to downstream or dams whose depth has decreased sufficiently as a result of human negligence to obey in disposing of garbage, (Amril Mutoi Siregar, 2018). The existence of disaster risk mapping is very important in structuring disaster management that is directed, appropriate, and (Nugraha, 2013).

Mapping flood-prone areas in Pahlawan Village is one way to control floods sustainably. Analysis of flood-prone areas in this study used the help of a Geographic Information System (GIS) with four parameters, namely (1) Rainfall Level, (2) Land Plain, (3) Garbage Population (4) Village Distance to the Sea. Previously several studies raised the theme of flood vulnerability, one of which was by Ligal Sebastian. There were two categories of causes of flooding, namely natural causes and human activities. With the occurrence of floods, there are flood control methods, namely structural and non-structural methods. There are two types of structural methods, namely repair and regulation of river systems which include river network systems, river normalization, embankment protection, flood embankments, short cuts and floodways; and Construction of flood control which includes dams, retention ponds, construction of check dams (sediment catchers), river slope reduction buildings, ground sills, retarding basins and polder construction, in flood prevention.

LITERATURE REVIEW

Machine Learning

In data science, to be able to find the pattern behind a dataset so that it can be even more useful, a machine learning method is needed. Machine learning itself discusses how machines can learn on their own so that the machine can perform certain tasks without being programmed explicitly.

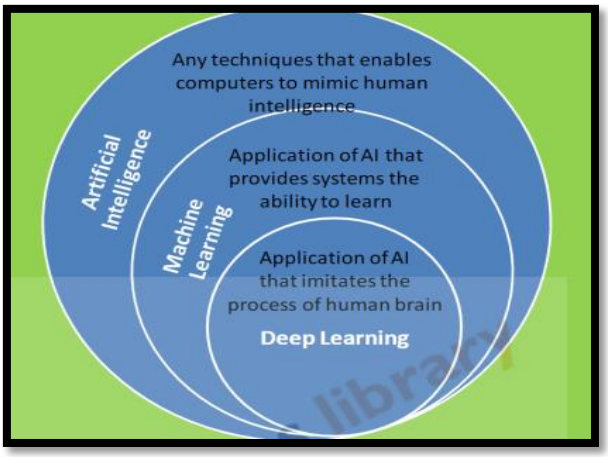

Figure 1. Hubungan AI, Machine Learning, Dan Deep Learning

According to Neden Siti Fatonah (2021), this study succeeded in implementing a classification method using the Naive Bayes algorithm for implementing flood disaster detection that occurred in the previous several years to be able to mitigate flood disasters in subsequent years using Rapidminner tools, and obtained a level of compatibility/accuracy of 76.73%. Meanwhile, according to Arya Febriansya et al (2020), - Natural disasters are one of the most frequent disaster events in Indonesia lately, especially flood disasters. Floods are one of the natural disasters that come and cannot be prevented. minus the impact of losses that result and result in a large impact on the human population. To minimize risk or loss to humans, knowledge, understanding, skills preparedness are needed to prevent, detect and anticipate earlier about various kinds of disasters or better known as disaster mitigation. The application for the emergency response stage for floods is still not much done and still needs to be developed. Machine learning is a technological innovation that has been widely applied in various fields of life and can also be used to improve flood disaster mitigation. This is very much needed for the effective and efficient management of a disaster so that risk reduction of the occurrence of a disaster can be achieved. According to Ahmed M. Al-Areeq (2022) Floods, one of the most common natural hazards globally are challenging to predict and estimate accurately. This study aims to demonstrate the predictive ability of four ensemble algorithms for assessing flood risk. Bagging

ensemble (BE), logistic model tree (LT), kernel support vector machine (k-SVM), and k-nearest neighbor (KNN) are the four algorithms used in this study for flood zoning in Jeddah City, Saudi Arabia. The 141 flood locations have been identified in the research area based on the interpretation of aerial photos, historical data, Google Earth, and field surveys. For this purpose, 14 continuous and different categorical factors are identified to examine their effect on flooding in the study area. The dependency analysis (DA) was used to analyze the strength of the predictors. The study comprises two different input variable combinations (C1 and C2) based on the features of sensitivity selection. The under-the-receiver operating characteristic curve (AUC) and root mean square error (RMSE) were utilized to determine the accuracy of a good forecast. The validation findings showed that BE-C1 performed best in terms of precision, accuracy, AUC, and specificity, as well as the lowest error (RMSE). The performance skills of the overall models proved reliable with a range of AUC (89–97%).

Convolutional Neural Network

The Convolutional Nueral Network (CNN) is a development of the MLP (Multilayer Perceptron) method which was originally named NeoCognitron by a researcher named Kunihiko Fukushima from NHK Broadcasting Science Research Laboratories, Tokyo, Japan. Then YanLecun and friends matured MLP into a LeNet architecture with their research on handwriting recognition (LeCun, Bottou, Bengio, & Haffner, 1998). The application of CNN continues to grow and Alex Krizhevsky successfully implemented CNN by winning the ImageNet Large Scale Visual Recognition Challenge competition in 2012 with a model called AlexNet. In its development, there are many CNN architectures that are often used in research such as AlexNet, VGGNet, GoogLeNet, ResNet etc.

How to calculate the convolution process is illustrated in the following figure where there is a multiplication of the image pixels with the filter to get the output feature maps. In the convolution operation, it consists of padding, stride, and filter operations as parameters. To calculate the size of feature maps or feature maps, calculations are carried out using Equation (2.1).

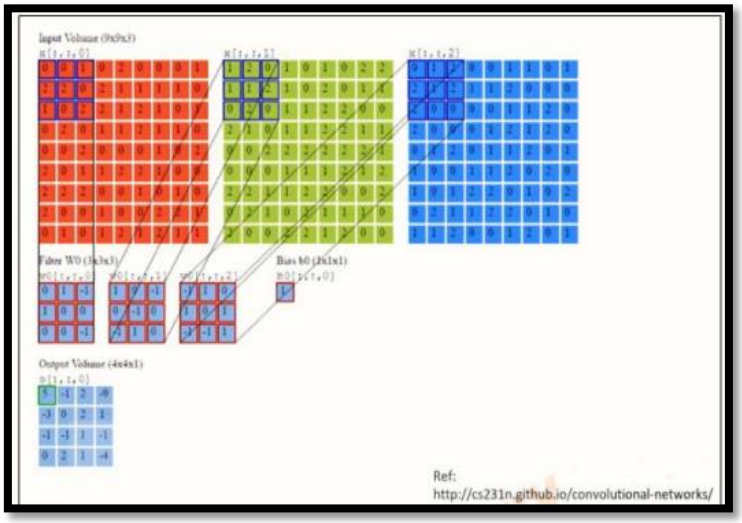

Figure 2. Ilustration of Operation Convolution

$$\text{Output} = \frac{n+2p-f}{S} + 1 \quad \text{.....(2.1)}$$

Information:
n = Input Length / Heigth
F = Karnel Filter Length / Heigth
P = Padding
S = stride

METHOD

This research is a descriptive survey research to find out which areas are classified as flood-prone zones in Pahlawan Village, Batu Bara Regency, North Sumatra. In order for the research to run in a structured manner and be completed on time, a research phase was created. The research steps can be seen in the following figure.

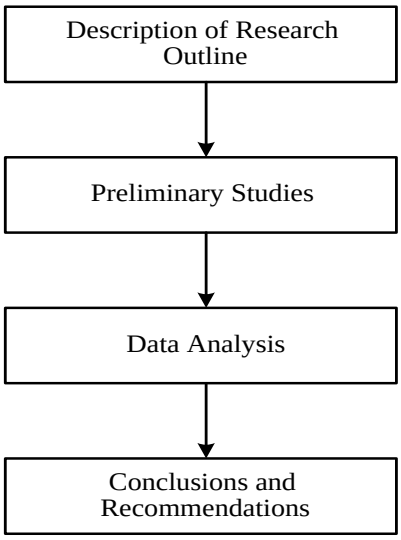

Figure 3. Research Stages

1. Description of Research Outline
   Based on the research flowchart, we can divide this research study into the following stages/steps.
2. Preliminary Studies
   The preliminary study carried out in this research was in the form of collecting literature on flood-prone areas in Pahlawan Village, as well as other sources related to research.
3. Data Collection
   The data collection carried out in this study was directly taken or collected from the field, namely in the form of data from surveys and field observations. Data collection was carried out by direct observation of the location of the floods in several places to find out the location of the coordinates of the flood-affected areas so that they are positioned in the mapping process.
4. Data Analysis
   After conducting a survey in the field, the existing data is collected and processed and then analyzed to obtain conclusions that are in accordance with the actual conditions that exist in the survey location. The data analysis stage was carried out by managing data from the results of site reviews and collecting data related to flood problems and then managing it into the Microsoft Excel program, to find out the percentage level of flood-affected areas. The analytical method used is Descriptive Analysis. Descriptive research is research that is intended to collect information about the status of an existing symptom, namely the condition of the symptoms according to what they were at the time the research was conducted.

RESULTS AND DISCUSSION

Results

This research was conducted in many flood-prone areas in Pahlawan Village, Batu Bara Regency, North Sumatra in 2022 at a time of high rain intensity which could cause flooding which included Hamlet I, Hamlet II, Hamlet III, Hamlet IV, Hamlet V, and Hamlet VI.

Based on the results of data analysis and mapping of flood-prone areas in Pahlawan Village and reviews in several locations, four hamlets are flood-prone zones. Pahlawan Village in Batu Bara Regency, which is located at 2º03'00”- 3º26'00” North Latitude and 99º01'-100º00' East Longitude, is recorded as having an area of 922.20 square km with altitudes varying between 0 – 30 meters above sea level. The sea is one of the causes of flooding in this village.

Figure 4. Pahlawan Village

The map above shows that the western part to the north is relatively low, close to the coast, which is in Hamlet I to Hamlet VI, while in other parts it tends to be plains away from the coast. As a result, flooding is more dominant in the six hamlets because of overflowing tides from the sea and also the occurrence of dead tides which causes rainwater not to be collected into the sea. For more details, the following is the percentage of flood-prone areas in Pahlawan Village.

Discussion

From the results of the analysis, several areas that are classified as flood-prone zones in Pahlawan Village include Hamlet I, Hamlet II, Hamlet III, Hamlet IV, Hamlet V, and Hamlet VI with the area affected as divided by hamlet as follows.

Table 1. Flood Prone Presentation Vilage Pahlawan

No.	Discraption	Hamlet Name	An Area (km^2)	Persentage
1.	Affected	Hamlet I, II, III, IV, V and VI	599.43	65%
2.	The Unaffected	Hamlet VII, VIII, IX, X and XI	322.77	35%

The table above, shows the area of flood-prone areas in six hamlets with a total area of 599.43km2 or around 65 percent of the area of Pahlawan Village, and 6 hamlets are categorized as flood-prone zones.

CONCLUSION

After obtaining the results from data processing and data analysis, the researcher can draw conclusions that are answers to scientific questions that exist in the research objectives. After that, the researcher was able to contribute in the form of suggestions to readers regarding the obstacles and solutions related to the problems in this study. Frequent floods in Pahlawan Village pose a serious threat to people's lives and infrastructure. Designing the scope of a flood map is an important first step to minimizing damage from a flood disaster. Flood continuity maps help highlight areas at risk of flooding, categorize possible flooding based on interactions between several conditioning factors, ranking between low to high-risk areas. In this study, we use machine learning (CNN Artificial Neural Networks). The flood-prone zones in Pahlawan Village include Hamlet I, Hamlet II, Hamlet III, Hamlet IV, Hamlet V and Hamlet VI with an area affected by floods with an area of 599.43 km each, a percentage value of 65%, an area of 322.77 km, the percentage value is 35%.

From this research, it can be understood that although the CNN model can simulate flood-prone areas, because of this, the design of this flood-prone map can help flood risk management and land use planning in Pahlawan Village, important in reducing flood disasters. Furthermore, from these results, we can quickly identify areas with high flood hazards and encourage the development of policies and infrastructure that can reduce the potential impact of flooding.

REFERENCES

Ameliola, S, Nugraha. 2013. Perkembangan Media Informasi dan Teknologi TerhadapAnak dalam Era Globalisasi. Malang : Universitas Brawijaya. http://icssis.files.wordpress.com/2013/09/2013-0229.pdf. Diakses 12 januari 2017.

Amril Mutoi Siregar, 2018 “Penerapan Algoritma K-Means Untuk Pengelompokan Daerah Rawan Bencana di Indonesia” (Information System Journal), 2018.

Arya Febriansyah1 , Alhuda Ramadhan2 , Mokhamad Gustiawan3 , Muhammad Revin R4 , Rahmaniza Maulana5 , Ristu Juli Y6 , Rollando G.E 7 , Ricky Firmansyah, (2020) Penerapan Machine Learning Dalam Mitigasi Banjir Menggunakan Data Mining, Jurnal Nasional Komputasi dan Teknologi Informasi Vol. 3 No. 3, Desember 2020 P-ISSN 2620-8342 E-ISSN 2621-3052.

Ahmed M. Al-Areeq, S. I. Abba, Mohamed A. Yassin, Mohammed Benaafi, Mustafa Ghaleb and Isam H. Aljundi 2022, Computational Machine Learning Approach for Flood Susceptibility Assessment Integrated with Remote Sensing and GIS Techniques from Jeddah, Saudi Arabia, Remote Sens. 2022, 14, 5515. https://doi.org/10.3390/rs14215515.

Hahn, P. (2019). Artificial intelligence and machine learning. Handchirurgie Mikrochirurgie Plastische Chirurgie, 51(1), 62–67. https://doi.org/10.1055/a-0826-4789.

Ligal Sebastian, 2008, Pendekatan Pencegahan Dan Penanggulangan Banjir Flood Prevention and Control Approach, Dinamika TEKNIK SIPIL, Volume 8, Nomor 2, Juli 2008 : 162 – 169.

LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. Proceedings of the IEEE. 86(11):2278-232.

Niode, F Dennis, Rinde Yaulie, dan Karouw Stanley. 2016. “Geographical Information System (GIS) untuk Mitigasi Bencana Alam Banjir di Kota Manado”. Manado: Universitas Sam Ratuangi. Jurnal Teknik Elektro dan Komputer. Vol.5 No.2 Januari-Maret, ISSN:2301-8402.

Nenden Siti Fatonah, 2021 “Penerapan Deteksi Bencana Banjir Menggunakan Metode Machine Learning”, Jurnal Format Volume 10 Nomor 2 Tahun 2021 :: ISSN : 2089 – 5615 :: E-ISSN : 2722 – 7162.

http://cs231n.github.io/convolutional-networks