Data Mining Approach for River Flood Hazard Time-Series: Using a Combination of Triple Exponential Smoothing and Neural Networks, in Demak

Dimara Kusuma Hakim[1*], Rahmat Gernowo[2], Anang Widhi Nirwansyah [3], Totok Haryanto[4]
[1]Department of Informatics, Universitas Muhammadiyah Purwokerto, 53182, Indonesia
[2]Department of Physic Faculty of Mathematics and Natural Sciences, Diponegoro University, Indonesia
[3]Department of Geography Education, Universitas Muhammadiyah Purwokerto, 53182, Indonesia
[4]Bachelor of Management Program, Faculty of Economics and Business, Universitas Muhammadiyah Purwokerto, 53182, Indonesia
E-mail: dimarakusumahakim@gmail.com
[*]Corresponding author

Keywords: triple exponential smoothing, neural network, time-series, flood forecast, flood hazard

Received: October 9, 2024

Conventional flood hazard maps depict a static perspective on flood risk and are helpful in flood forecasting. Then, time-series flood forecasting research was developed and widely conducted. Still, it has limitations because the univariate method used does not consider the influence of other variables, unlike regression. Also, the determination of flood hazard weights, which is usually conducted through empirical studies using AHP, has a good level of accuracy but has the drawback of expert subjectivity.
A new approach is proposed for flood hazard forecasting, combining the Triple ES method with NN and determining weights in a data-driven manner using NN. A combination of time series and regression methods (significantly non-linear), namely Triple ES with NN, results in good accuracy with an MSE value of 3.03, MAE 1.20, RMSE 1.74, R^2 0.54, and a MAPE of 32.93%.
The evaluation results for flood hazard weight determination with an MSE of 0.0111 and a MAPE of 7.81% show promising results, and the weights can be used in Hazard Flood GIS. Visualization in the form of a GIS Hazard Map can be done after all related raster data have been combined. The computational outcomes, particularly the MSE and MAPE values, demonstrate the effectiveness of the proposed approach, providing a clear understanding of the model's performance.

Povzetek: Z razvojem modela Triple ES in nevronskih mrež učinkovito napoveduje poplavno nevarnost z vizualizacijo v obliki GIS.

1 Introduction

There is no doubt that river flooding is one of the most destructive natural disasters [1], [2], [3], as it encompasses impacts that include social, economic, and environmental dimensions. The development of flood control efforts and disaster reduction has successfully reduced some casualties due to flooding; however, this remains a global concern [4]. The flood in Demak that occurred some time ago has had a very significant impact on the local community. The flood in Demak that occurred some time ago has significantly impacted the local community. The flood caused by the breach of six levees on the Wulan River and heavy rainfall has resulted in 89 villages in 11 sub-districts submerged with water levels ranging from 30 to 80 cm, affecting approximately 93,149 [5]. This incident is not new for Demak, as it has frequently happened in previous years [6], [7], [8], [9], [10], [11].
The National Disaster Management Agency (BNPB) and the Meteorology, Climatology, and Geophysics Agency (BMKG) have implemented mitigation strategies to reduce the impact of flood disasters [12], [13], [14], and they also conduct flood risk projections based on weather data analysis and local geographical conditions. However, the predictions have a time limit for various reasons (weather and climate variability, environmental changes, and climate change). The BNPB and BMKG do not provide long-term flood predictions that extend beyond the current year [15], [16], [17], which presents research opportunities outside of these government agencies.
Conventional flood hazard maps present a static view of flood risks. Still, they are less effective in capturing the dynamic characteristics of the variables that influence these risks over time, which are essential [3], [18], [19], especially in terms of long-term forecasting. Previous research has been conducted to perform long-term forecasting using ARMA and ARIMA methods [20], [21], and their weaknesses due to not considering seasonality components [22], [23], [24] have been addressed with SARIMA and Triple Exponential Smoothing [25]. However, the univariate nature of these methods still does not resolve the issue of the influence of other variables, so a combination of techniques is needed to simultaneously address several issues: trends, seasonality, and the impact between variables.
In flood hazard mapping, an index must be multiplied by the variables influencing flooding. The determination of the flood hazard index, which has been carried out by BNPB in the form of the Indonesian Disaster Risk Index

(IRBI) using the Analytic Hierarchy Process (AHP) method, provides a guideline for the index used by decision-makers. However, it has several limitations [26], [27]: the necessity of having experts, difficulties in using comparative scales, potential low consistency, sensitivity to data changes, limitations in handling changes in priorities, and a limited number of variables. Therefore, research to create a separate index driven by data is an interesting endeavor to pursue [28].

This research aims to develop a flood hazard information system using a data mining approach with time series forecasting through a combination of the Triple Exponential Smoothing (Triple ES) method and Artificial Neural Networks (ANN). It is expected to address issues such as trends, seasonality, and the influence between variables simultaneously and determine the optimal weight of flood hazard variables using a data-driven approach with ANN. Optimization with gradient-based algorithms is used in every method to achieve the lowest possible error value.

2 Data

The data is in tabular, raster, and vector (shapefile) form. Tabular data contain relevant attributes or variables, such as rainfall, temperature, humidity, etc. Raster data (Figure 1) consists of a grid of pixels representing spatial information: land cover (COPERNICUS 100m Proba-V-C3 Global) & elevation (DEMNAS). The shapefiles used represent regional boundaries and river lines.

This study uses three data types for deeper analysis and understanding of the phenomenon studied: flood database (Table 1), meteorological data (Table 2), and raster data. Flood event information is collected, and we create a Flood database that contains information related to flood events and variables at the location (depth, runoff, river distance, land cover, wind speed, rainfall, etc.). This database looks at flood risk, finds places likely to flood, and plans how to protect those places. Including temperature, precipitation, humidity, surface pressure, and wind speed, meteorological data helps us understand how the atmosphere and weather trends affect flooding and rain. A thorough understanding of how meteorological conditions interact with flood episodes and how they affect ecosystems and populations is obtained by combining these two forms of data.

Elevation in raster form (satellite imagery) and river area (in shapefile/SHP format) of Demak Regency was obtained from the Geospatial Information Agency of Indonesia. River distance is obtained by performing Euclidean calculations for each pixel in the river area data and finding the nearest distance. The raster land cover data was obtained by accessing the API (Application Programming Interface) to the Copernicus Global Land Cover Layers (a product of Sentinel-2), which has a relatively high resolution (100 meters).

Runoff is obtained by converting land cover according to the Runoff Coefficient Table [29]. All this data is stored and converted into raster format and resampled to approximately 30 meters.

Then, the following data was obtained from ERA5-Land Hourly over 14 years (2011-2024) from the upstream area of the research location: precipitation, temperature, dew point, humidity, surface pressure, and wind speed. This data is stored as a database.

Additionally, data on flood events sourced from mass media and the National Disaster Management Agency is collected as material for determining the flood hazard weight. All data is processed to assess flood hazard weight and forecast time series.

Table 1: Flood event

id	date	location	coord	depth
1	2023-01-01	Betokan	-6.87, 110.63	55
2	2023-01-01	Bintoro	-6.89, 110.63	83
3	2023-01-01	Mangun jiwan	-6.88, 110.59	48
…	…	…	…	…

Table 2: Meteorological data

YEAR	MO	DY	T2M	T2MDEW	QV2M	PS	WS10M	PRECTOTCORR
2011	1	1	25.6	23.81	18.75	99.03	4.88	63.42
2011	1	2	25.2	23.09	17.92	99.14	4.4	37.1
2011	1	3	25.4	23.42	18.29	99.13	3.85	26.04
2011	1	4	25.3	23.4	18.27	99.08	4.73	11.12
2011	1	5	25.8	23.94	18.88	99.04	4.3	10.82
2011	1	6	25.9	23.82	18.73	99.13	3.42	3.62
2011	1	7	25.6	23.6	18.49	99.11	3.59	1.39
…	.	.	…	…	…	…	…	…

3 Method

3.1 Related works

Based on various previous studies on the topic of river floods (Fig 1), two things are often discussed: risk maps [30], [31] and prediction models [32], [33], [34]. This study attempts to combine the two domains with a data mining approach.

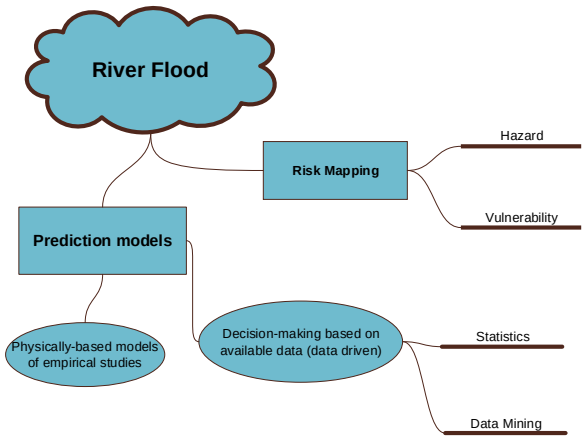


Figure 1: State of the art mind map

This summary table (Table 3) organize the findings from various previous studies and existing limitations.

Table 3: Related works

Method	Key findings	Limitations
Projected flood risk based on analysis of weather data and local geographical conditions [7], [8], [9]	Implementation of mitigation strategies to reduce the impact of flood disasters	Does not provide long-term flood predictions beyond the current year
ARMA & ARIMA [15], [16],	Long-term forecasting using ARMA and ARIMA methods	Doesn't consider seasonal components
SARIMA & Triple ES [17], [18], [19]	Overcoming the disadvantages of ARMA and ARIMA by considering seasonal components	Univariate properties do not resolve the influence of other variables
Analytic Hierarchy Process (AHP) [20], [21]	Determination of Indonesia's disaster risk index (IRBI)	Requires expertise, difficulty in using comparative scales, low potential for consistency, sensitivity to data changes, limitations in handling priority changes, and limited number of variables

This study develops a flood hazard information system using a data mining approach with time series forecasting through Triple Exponential Smoothing (Triple ES) and Artificial Neural Networks (ANN) methods. This approach is expected to overcome the problems of trends, seasonality, and influence between variables simultaneously and determine the optimal weight of flood hazard variables using a data-based approach with ANN. Each method is optimized with a gradient-based algorithm to achieve the lowest possible error values.

3.2 Determination of river flood hazard weight with ANN

Determining the weight of flood hazards, which is generally carried out using an empirical approach by using the Analytic Hierarchy Process method (AHP), is a very relevant research area in dealing with disaster threats by examining the relationship between hydrological variables in the past (such as rainfall, river flow, and topography), has provided insight into flood patterns [35], [36]. However, this has limitations because it heavily relies on the subjective assessments of experts; there is a significant possibility of personal bias that can influence the final results [37], [38]. Another data-driven approach is considered to be able to overcome this problem [39], [40]. Data-driven approaches, especially data mining with various methods, can utilize large amounts of data and identify complex patterns to determine the weight of flood hazards [41], [42]. In this task, the Artificial Neural Network (ANN) method is used, a computational model inspired by how the brain works [43]. Technically, it utilizes several processing units called neurons, processed through several hidden layers, requires an activation function to operate, and employs a specific optimization technique to achieve optimal predictions. Several steps were taken in this research to perform weight calculations using ANN.

3.2.1 Create a flood and precipitation database, download raster data

Historical data on flood events in the Demak region over the last 12 years has been collected, including the location and timing of the flood occurrences. Data on precipitation history was created for the upstream area (Grobogan Rain Station) over 14 years to complement the flood database and to understand the rainfall patterns that influence flood occurrences in the downstream area (Demak).

Raster data covering river elevation and area were downloaded from the Indonesian Geospatial Information Agency and land cover data from Sentinel (using the Copernicus API). The Surface Runoff band is created based on the land cover by converting the Runoff Coefficient Table [23]. In contrast, the river distance band is calculated using Euclidean distance (Eq 1) for each pixel relative to the nearest river area [44].

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \quad (1)$$

Flood data is updated based on raster: elevation, river distance, runoff, and land cover at the time and location of the flood event. The average value is calculated by reading the raster (GeoTIFF), then setting a radius of 500 meters from the flood location point, creating 50 random points, taking each value from those random points, and calculating the average value. The depth of flooding is one of the leading indicators in flood risk assessment.

3.2.2 Conducting modeling with ANN for calculating weights for each variable

Conducting modeling using Artificial Neural Network (ANN) to predict flood risk based on input variables (Fig 2) involves training the model, validation, and optimization to achieve accurate results. Calculating the weights for each variable involves the following steps: calculating the average weight, normalizing the weights, and determining the percentage weights. These weights indicate the extent of each variable's influence on flood risk and are used in creating the flood hazard index.

To perform the calculations, the flood event database was updated with average elevation, average river distance, average surface runoff, and average rainfall.

Then, outlier values in the data were removed using Z-score, and the data was divided into training and testing sets. Data is standardized using StandardScaler to ensure all features are at the same scale. The artificial neural network model was built using Sequential, with multiple layers of Dense and Dropout to prevent overfitting. The model is compiled using an ADAM optimizer with an adjusted learning rate and an MSE loss function. To improve the model's performance, EarlyStopping stops training early if there is no increase in value loss after a few epochs. The model is trained with training data and validated with a subset of training data. After training, the model is evaluated with test data to obtain optimal MSE values. The predictions are made on a test set, the prediction results are classified into the nearest bins, and finally, the MAPE is calculated to measure the accuracy of the predictions. The prediction results and actual values are used to generate a comparison graph.

Several layers are used to determine flood weights using ANN. First layer: 64 neurons with ReLU activation function. After this stage, 20% of the neurons will be dropped using the Dropout layer to prevent overfitting and improve the model's generalization. The second layer contains 32 neurons with a ReLU activation function. After this stage, 20% of the neurons will be dropped using the Dropout layer to prevent overfitting and improve the model's generalization—the third layer is one neuron with a linear activation function. The Adam optimizer (Adaptive Moment Estimation) is used with a learning rate of 0.001 because Adam combines the advantages of two optimizers, AdaGrad and RMSProp, to adjust the learning rate during training, so it requires less memory and is efficient. The MSE loss function was chosen because it is suitable for regression problems. The EarlyStopping callback automatically stops the training if the validation loss does not improve after five consecutive epochs and the best weight achieved during the training is selected. Model training using training data is carried out for a maximum of 100 epochs, with a batch size of 10. 20% of the training data will be used as validation data.

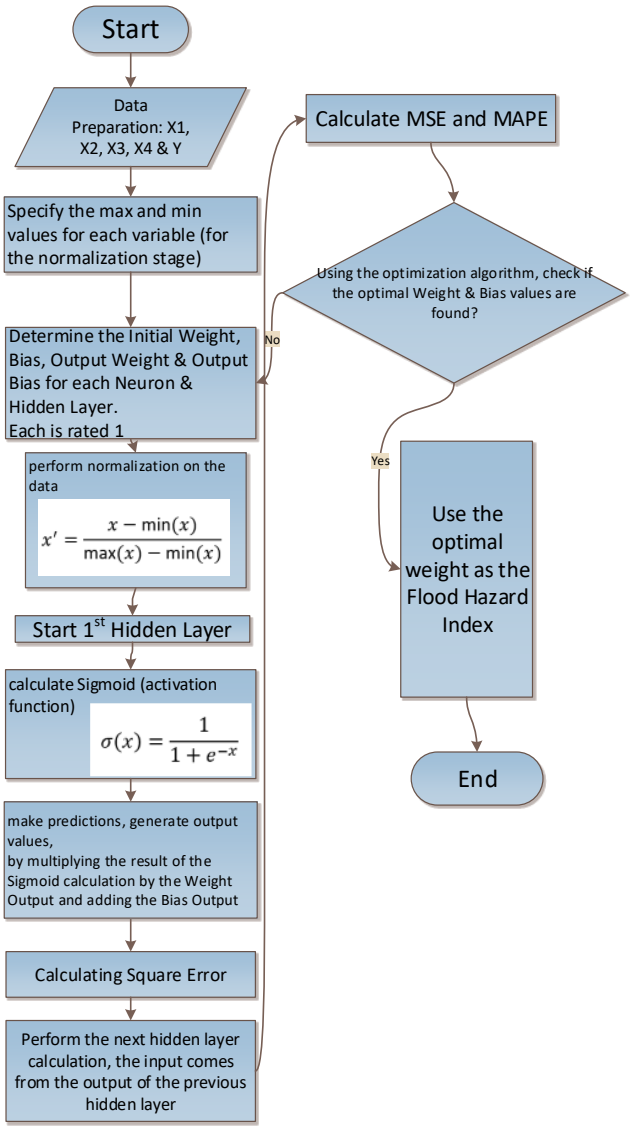


Figure 2: Determination of river flood hazard weight with ANN

3.3 Build a flood forecasting model with a Combination of Triple ES, ANN, and gradient-based optimization

The Triple ES method (Fig 3.a) is a technique for forecasting that considers level, trend, and season. This method is considered a univariate time series approach because it only predicts one variable. Meanwhile, the Neural Network method (Fig 3.b) predicts one dependent variable influenced by various independent variables but has no relationship with time.

The univariate time series approach, which has limitations in considering the influence between variables (because the forecast is only for one variable), has been improved with the multivariate approach (such as Vector Autoregression), which involves the influence between time series variables [45]. However, it still has limitations in handling situations where there are independent variables (X1, X2) that affect the dependent variable (Y) while also involving time [45].

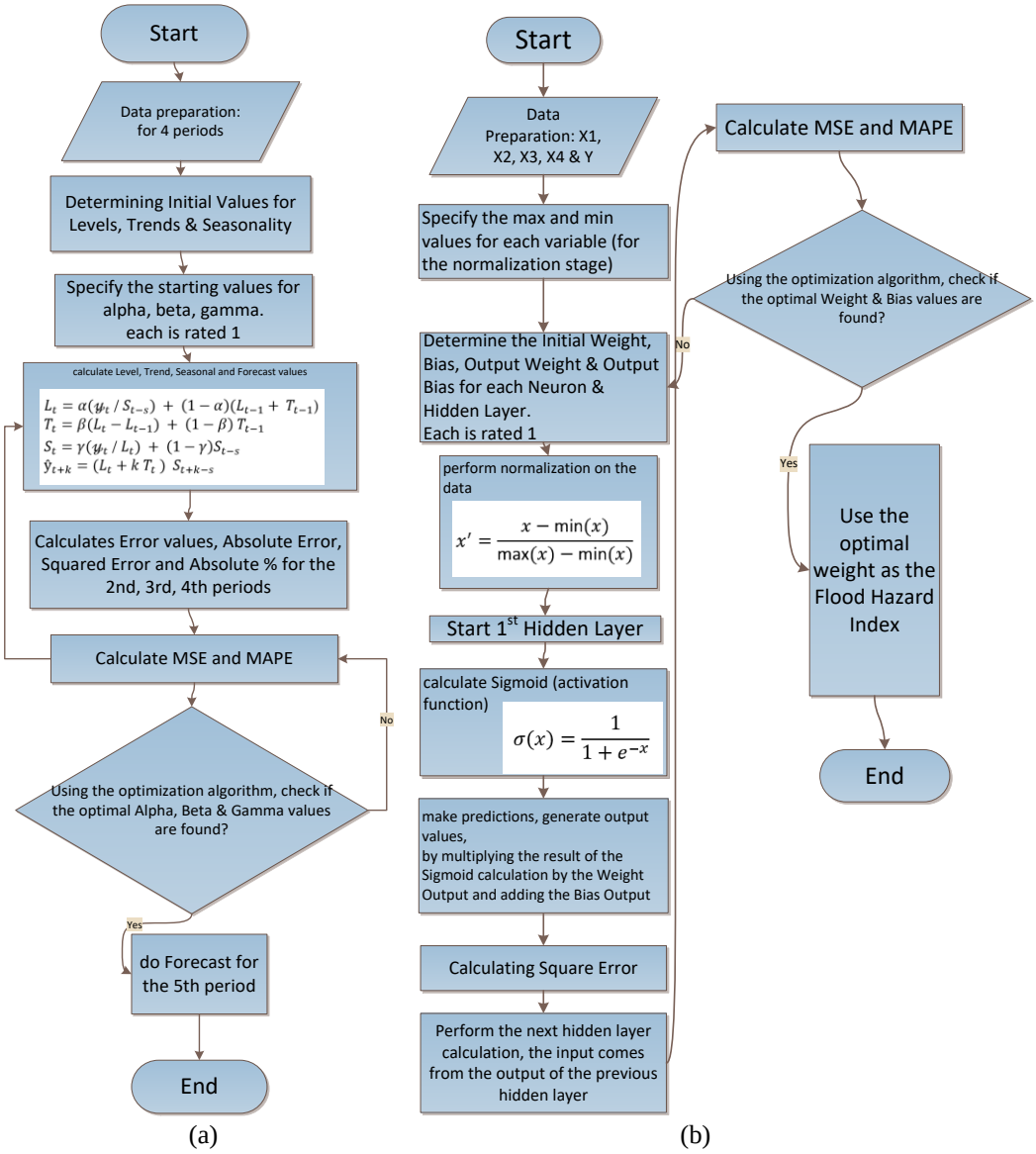


Figure 3: Basic methods: triple ES (a) and neural network (b)

The combination of the Triple ES method and ANN (Fig. 4) is carried out in this research to simultaneously address several issues: level, trend, seasonality, and the influence between variables. Forecasting each variable uses Triple ES first, and estimation is done using Neural Network.

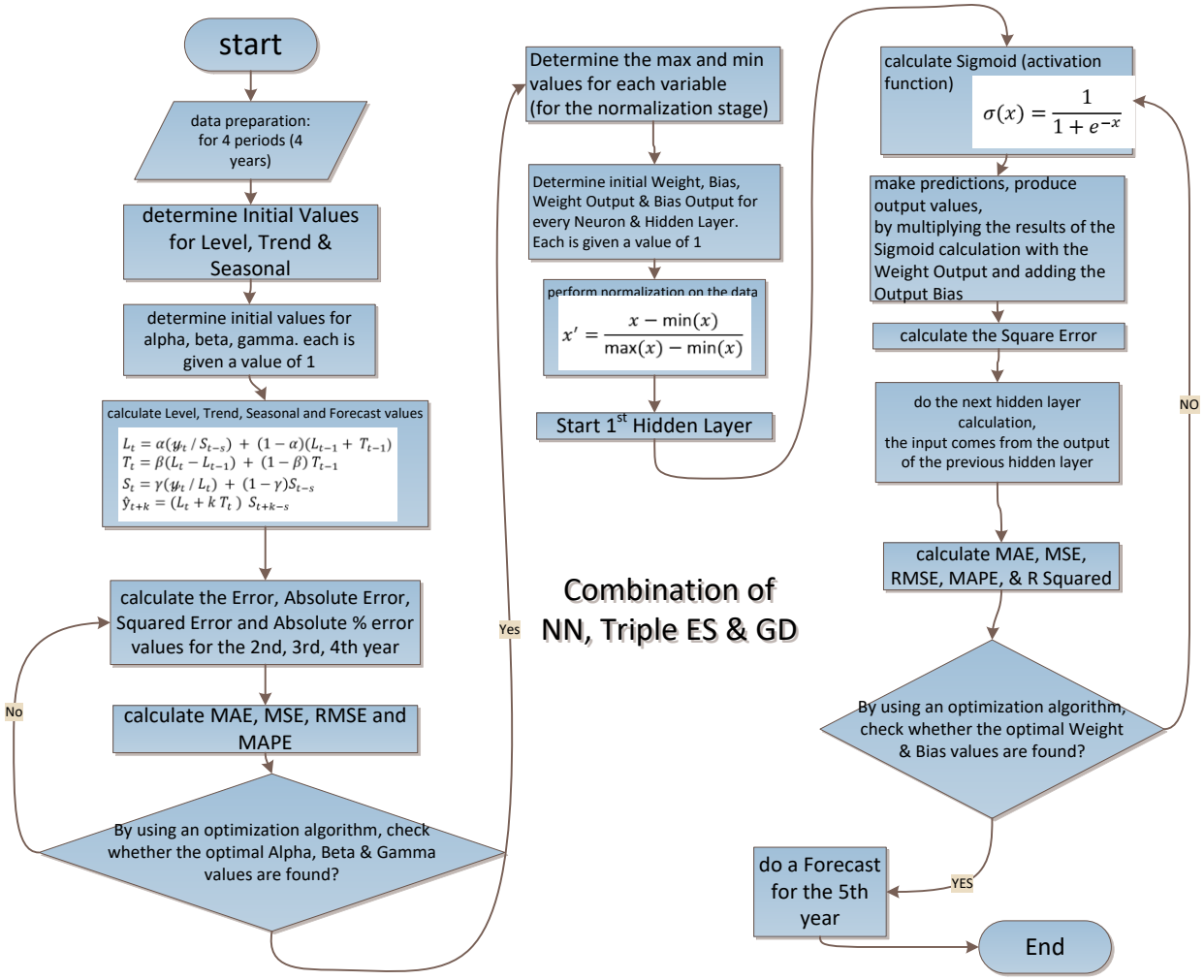


Figure 4: Forecast with hybrid approach

3.3.1 Preparation

Some data that affect rainfall are grouped by year and month: Temperature (C), Dewpoint (C), Humidity (g/kg), Surface Pressure (kPa), Wind Speed (m/s), and Total Precipitation (mm/day). The aggregation functions used in this data processing are max and sum. The sum function is used for total precipitation to obtain the total rainfall over a specific period, which is essential for predicting rainfall accumulation. Temperature, dew point, Humidity, surface pressure, and wind speed each use the max function, identifying extreme conditions that may affect rainfall. This aggregation function is beneficial in achieving accuracy in forecasting and prediction.

3.3.2 Prediction of each influential variable

Triple ES performs initial forecasting for each variable using gradient-based optimization: L-BFGS-B with loss: MSE. Then, evaluation is conducted by calculating MSE and MAPE for each variable and comparing the forecast results with the test data. At the end of this activity, model 1 for forecasting is saved.

Triple ES is a time series forecasting method that is an advanced form of exponential smoothing because it involves three parameters: α (alpha) as the level, β (beta) as the trend, and γ (gamma) as the seasonal component [46]. The steps taken in Triple ES are to calculate Level, Trend, and Seasonality and make forecasts (Eq 2-5). Where: L_t is the level at time (t), T_t is the trend at time (t), S_t is the seasonal component at time (t), Y_t is the actual value at time (t), (m) is the seasonal period, (k) is the forecast horizon.

$$L_t = \alpha \frac{Y_t}{S_{t-m}} + (1-\alpha)(L_{t-1} + T_{t-1}) \quad (2)$$

$$T_t = \beta(L_t - L_{t-1}) + (1-\beta)T_{t-1} \quad (3)$$

$$S_t = \gamma \frac{Y_t}{L_t} + (1-\gamma)S_{t-m} \quad (4)$$

$$F_{t+k} = (L_t + kT_t)S_{t-m+k} \quad (5)$$

3.3.3 Creating and training a neural network model

$$z = \frac{x - \mu}{\sigma} \quad (6)$$

The standard scaler shifts the feature distribution so that it has a mean of zero and a standard deviation of one (Eq 6), ensuring that all features are on the same scale. Where: x

is the original value of the feature, μ is the mean of the feature, σ is the standard deviation of the feature, and z is the normalized value of the feature.

3.3.4 Conducting testing for combination method

It begins with forecasting the variables using Model 1 (forecasting) that has been created before, forecasting several months in the testing data as variable X. Then, it continues with the use of Model 2 (prediction) that has been developed to predict variable Y. The prediction results are then processed by rounding to the nearest class to round the predictions to the values closest to the score class. The prediction results are saved for evaluation. A join of actual data and predicted data was performed for evaluation, and then MSE, RMSE, R Squared and MAPE were calculated.Then, the optimal model produced is used to make forecasts for a period of 20 years (2025-2045).

3.4 Visualization and calculation of river flood hazards based on raster

Visualizing river flood hazards using rasters requires integrating geospatial data and raster processing (especially starting with DEM). Raster processing allows for data analysis in a grid format, which is very useful for illustrating variations in topography and land use that can influence water flow.

Several steps are taken to produce a Time Series Hazard Map for a particular month and year. This process begins with collecting relevant geospatial data, such as elevation maps, rainfall data, and land use information (to determine runoff values). Once the data is collected, the next step is preprocessing, which includes cleaning and normalizing the data to ensure consistency and accuracy. Next, a hydrological analysis is carried out to determine areas at high risk of flooding by considering various variables.

First, the prepared raster data will be analyzed using a hydrological model to simulate water flow and potential inundation. This model will produce a flood risk map showing the most vulnerable areas. Then, the analysis results will be visualized as a thematic map that is easy to understand.

3.4.1 DEM (digital elevation model)

To be able to retrieve DEM Raster data, DEMNAS Data Download is carried out from the Geospatial Information Agency (BIG) via tanahair.indonesia.go.id. Then, Raster Data Processing (Merging & Clipping) is carried out. Merging is carried out because the raster covering the Demak area consists of several GeoTIFF files, which are combined into one raster file. The Clipping process is carried out according to the boundaries of the study area (Shapefile, which is also obtained from BIG). Coordinate Transformation (resolution 30 m for each pixel) and Coordinate System equalization (EPSG: 4326) are needed because data integration is carried out, namely the combination of data from various sources; in this study, the integration of elevation raster with raster: rainfall, runoff, and river distance will be carried out. Coordinate Transformation and Coordinate System equalization are carried out by changing all rasters into the same coordinate system so that a more comprehensive analysis can be carried out. Normalization and inverting are carried out at the analysis stage. Because the higher the location, the lower the risk of flooding, so when using the Min-Max Normalization method as a score, the invert is carried out (Fig 5.a).

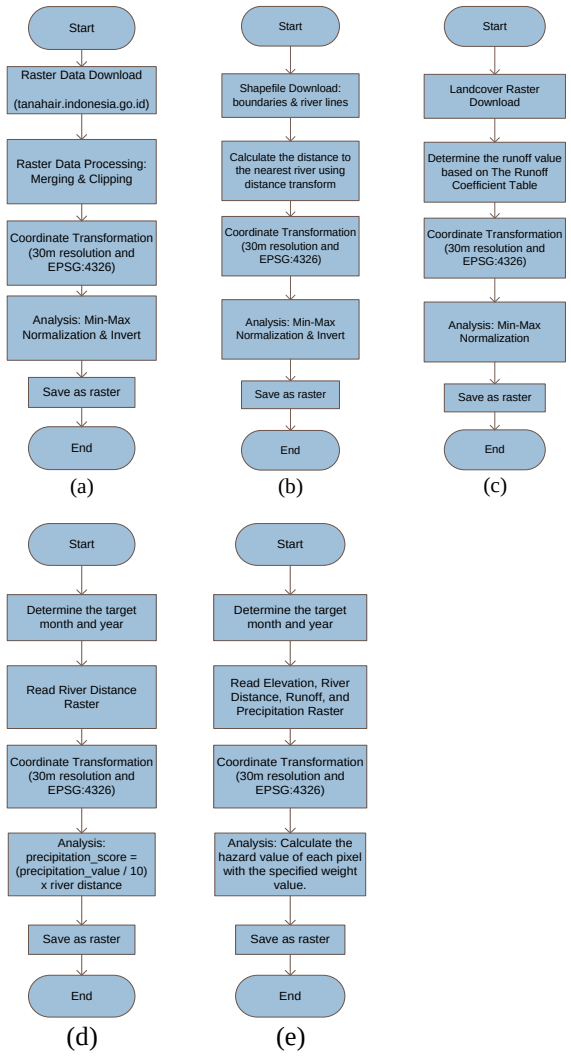


Figure 5: Raster Processing: a) DEM, b) River Distance, c) Runoff, d) Precipitation, e) Hazard

3.4.2 River distance raster creation

Two shapefiles are first downloaded from BIG (boundary and river line) to perform river distance raster processing (Fig 5.b), which is needed for the analysis. After that, the distance to the nearest river is calculated using the distance transformation method (euclidean distance) so that each pixel will contain the distance value to the nearest river line. Next, a coordinate transformation is performed to ensure that the raster has a resolution of 30 meters and uses the EPSG:4326 coordinate reference system, which is essential for the suitability of geospatial data. After the distance raster is generated, perform a Min-Max normalization analysis to transform the distance values into a more standardized range, followed by value inversion so that areas closer to the river have higher values. Finally, the results of this processing are saved as a raster for further analysis.

3.4.3 Runoff raster creation

Landcover Raster (Copernicus Global Land Cover Layers) taken within a 40 km radius of the research location is used in the Runoff Raster Creation (Fig 5.c), which aims to produce a map of potential water runoff. This raster has a resolution of 100 m [47], which is good enough to be used in flood analysis. It describes the land cover that affects the value of water runoff: forest, agricultural land, urban, etc. The runoff value is determined for each pixel based on the Runoff Coefficient Table [29], which provides different coefficients for different land cover types, reflecting how much rainwater will flow as runoff compared to what the soil absorbs. Furthermore, a coordinate transformation is carried out to ensure that the runoff raster has a resolution of 30 meters and uses the EPSG:4326 coordinate reference system, which is essential to maintain the consistency of geospatial data and facilitate integration with other data. After obtaining the runoff value, a Min-Max normalization analysis is carried out to transform the runoff value into a more standardized range, making it easier to interpret the data. The last step is to save the results of this processing as a raster.

3.4.4 Precipitation raster

The relationship between rainfall in upstream areas and the distance to the river of a location provides a score for the influence of rainfall at a location. Based on actual and forecast data on rainfall, the target month and year for analysis are first determined because rainfall can vary significantly over time. After that, the previously created river distance raster (the distance between the nearest rivers from each location) is read. To maintain the consistency of geospatial data, coordinate transformation is carried out to create a rainfall raster with a resolution of 30 meters using the EPSG:4326 coordinate reference system. The rainfall value is divided by 10 (to adjust the scale) and then multiplied by the distance to the river. The last step is to save the processing results in a raster (Fig 5.d).

3.4.5 Hazard raster

In Hazard Raster processing (Fig 5.e), the first step is determining the target month and year for the analysis because risk conditions can vary depending on a specific time. After that, read the previously generated elevation, river distance, runoff, and rainfall rasters, which provide essential information about the factors that affect the risk. The analysis is carried out by calculating the hazard value of each pixel by multiplying it by the weight value determined for each factor, which can include elevation, distance to river, runoff, and rainfall. The last step is to save the results of this processing as a raster.

4 Result and discussion

This section explains all the results that have been obtained, in the form of numbers, tables, and map visualizations.

4.1 Raster data processing

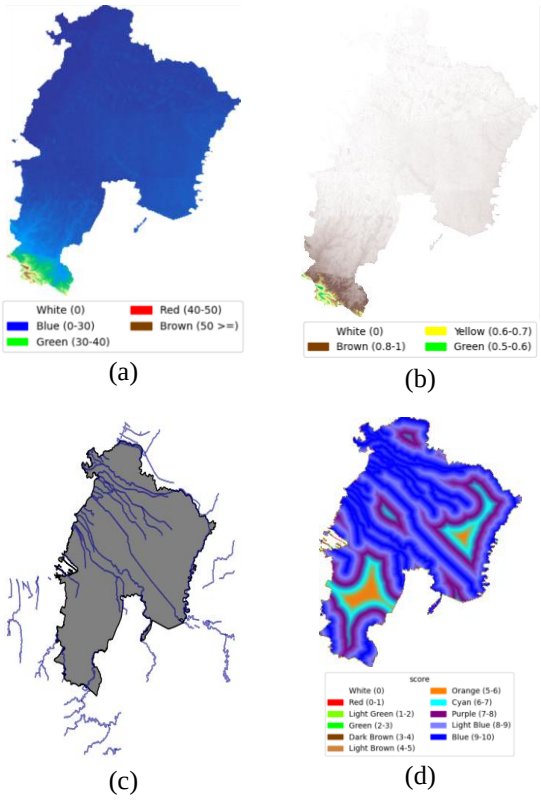
The raster calculation process is designed to produce several key parameters: elevation, distance to river, and runoff analysis. This process is continued by creating a raster to assess the impact of upstream rainfall on downstream areas based on total monthly rainfall and river distance.

The elevation raster describes the height of the land surface relative to sea level, starting with the raster merging process and continuing through resampling (Fig 6.a), truncation, and finally, normalization, which includes an inversion step (Fig 6.b). This is important because higher locations tend to have less impact on flooding.

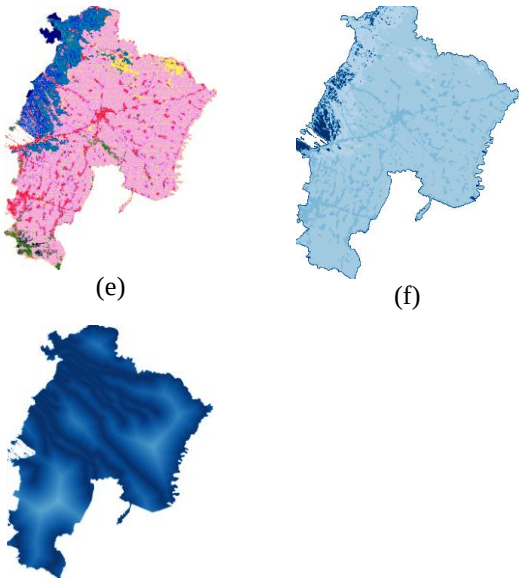
The distance to the river raster reflects the shortest distance from each pixel to the river. It starts with a base map depicting the river flow (Fig 6.c) and then calculates the shortest distance for each pixel using the Euclidean distance method (Fig 6.d). Pixels closer to the river are shown in bluer colors.

Precipitation, the impact of upstream rainfall on downstream areas, is calculated by involving the distance of the nearest river to each pixel (Fig 6.g).

The land cover raster is the raw data that provides information about the land cover at each pixel (Fig 6.e). The color code for the legend is according to the Discrete Classification in the Copernicus Global Land Service [48]. Then, the land cover value is converted into runoff (Fig 6.f) based on the runoff coefficient table [29]. In the raster visualization, a bluer color indicates more water, which is considered runoff.



(a) (b) (c) (d)



(e) (f) (g)

Figure 6: Raster Processing: a) Raw Elevation, b) Inverted Elevation, c) River Flow, d) River Distance, e) Land Cover, f) Runoff, g) Precipitation

4.2 River flood hazard weight

This study aims to assess river flood hazard weights using Artificial Neural Networks (ANN), offering advancements over traditional methods like the Analytic Hierarchy Process (AHP). By creating a flood and rainfall database and utilizing relevant raster data, ANN modeling calculates the weight of variables influencing flood hazards. This data-driven approach is expected to yield more accurate results and enhance the effectiveness of flood hazard information systems.

4.2.1 Flood database and raster data

A comprehensive database was constructed to develop a data-driven flood hazard weighting system. The first data includes flood events with the following structure: id, flood, region, year, date, information, and coordinates (Table 1). The second data comprises 11 years (2013-2023) of precipitation from the upstream region of Grobogan. Subsequently, raster data was collected. The obtained Raster Elevation data was integrated with river distance, land cover, and runoff information. The flood database was updated using raster data for the column's: runoff, river distance, precipitation, and elevation. Precipitation data was populated by querying Grobogan's upstream precipitation database.

Table 4: Flood database

id	f	date	loc	coord	precp	depth	runoff	distc	elv
1	1	2023-01-01	Betokan	-6.87, 110.63	149.54	55.68	0.38	0.56	4.48
2	1	2023-01-01	Bintoro	-6.89, 110.63	149.54	83.66	0.47	0.50	1.68
3	1	2023-01-01	Mangunjiwan …	-6.88, 110.59 …	149.54	48.40	0.33	0.50	5.21
15911	0	2013-04-10	Ngelo Kulon	-6.79, 110.68	21.76	0	0.6	0.46	3.60
15912	0	2013-04-10	Pecuk	-6.79, 110.71	21.76	0	0.51	0.54	4.84
15913	0	2013-04-10	Pasir	-6.79, 110.67	21.76	0	0.56	0.44	4.70

Table 4: Flood database (continued)

Temp	Dewp	Humd	Pressure	Wind Speed	Precipitation	datetime
24.71	23.28	18.13	99.16	2	0.32	2013110
24.7	23.18	18.01	99.12	2.14	0.28	2013111
24.62	23.09	17.94	99.11	2.25	0.23	2013112
…	…	…	…	…	…	…
27.01	25.33	20.45	99.49	1.48	0.59	2023123
26.78	25.26	20.39	99.46	1.61	0.64	2023122
26.55	25.16	20.26	99.4	1.71	0.7	2023123

4.2.2 Flood hazard model with ANN

The data is prepared before applying ANN. When a negative weight is found at the end of the process, the data preparation process is corrected by inverting the variables with negative weights. A negative weight indicates an inverse relationship between the input and target variables. Higher elevations and greater distances from the river are related to a reduced risk of flooding. This makes sense, as areas that are elevated or farther from the river are less prone to being impacted by floodwaters. The variables were defined as X1: elevation, X2: river distance, X3: runoff, X4: precipitation, and Y: flood depth (target). Outliers were removed using the Z-score method to ensure data quality. The dataset was split into training and testing sets with an 80:20 ratio. Next, the data was standardized to improve the model's performance.

The neural network model was constructed using a sequential architecture. The model included an input layer with 64 neurons and a ReLU activation function, followed by a dropout layer with a 20% dropout rate. This was followed by a hidden layer with 32 neurons and another dropout layer with a 20% dropout rate. The output layer consisted of a single neuron with a linear activation function. The model was compiled using the Adam optimizer with a learning rate 0.001 and MSE loss function. Early stopping was incorporated to prevent overfitting, monitor the validation loss with five epochs, and restore the best weights.

The model was then trained on the prepared dataset, and its performance was evaluated using MSE and MAPE.

The model achieved an optimal MSE of 0.0111, indicating a high level of accuracy in its predictions. Additionally, MAPE was 7.81%, reflecting the model's precision in forecasting relative to the actual values.

4.2.3 Weight

The weight calculation results show elevation: -0.0435, river distance: -0.0492, runoff: 0.0411, and precipitation: 0.0623. At this stage, negative weight results for elevation and river distance indicate that both variables negatively influence flood predictions. In the context of flood hazard weighting, this is considered invalid, so the data preparation is repeated. The data for both variables is inverted.

After inverting both variables and repeating the process, the output shows that elevation has the highest average weight of 0.2096, followed by river distance at 0.0899, precipitation at 0.0729, and runoff at 0.0124. After normalization, the elevation weight becomes 0.5499, indicating that elevation is the most significant factor in the flood hazard model. The distance from the river and rainfall also have a considerable influence, with weights of 0.2360 and 0.1913, respectively, while surface runoff has the least impact, with a weight of 0.0325. Thus, elevation is the most influential variable in this model, followed by the distance from the river and rainfall, while surface runoff has the most minor contribution.

4.3 Combination of triple ES and ANN

To forecast the precipitation, the first step is to group historical statistics on rainfall by year and month and then find the highest value for each variable. Then, we used Triple ES to make predictions about each variable. The data was split into two groups: training data (from 2011 to 2019) and testing data (2020–2024). Using L-BFGS-B optimization, weather factors like temperature, humidity, pressure, and rainfall are used for forecasting. Once the model is made, each variable's MSE and MAPE values are found by comparing the predicted results with the actual data. The model is then saved as Model 1.

Based on the test results, the optimal seasonal period is 48 months (4 years). Based on the test results, the optimal seasonal period is 48 months (4 years); the test results for all periods are reported in Table 2. Periods from 12 to 36 show increasing error values, but it turns out that the 48-season period shows the lowest error (MAPE: 32.93%). The 60 period failed because the training data was insufficient for training for two full seasons (120).

The initial forecasting results (Triple ES) show that temperature (2.15%), dewpoint (2.89%), humidity (4.0%), and surface pressure (0.12%) have very low error values (MAPE) and are suitable for use in the following prediction stage. Wind speed has an error that is relatively low (12.8%) and acceptable in time-series analysis. Precipitation has a very high error (66.51%), which is addressed in the following prediction stage.

To help improve prediction accuracy, separate variables were created: normalized and binned, with Min-Max normalization used for all variables except for precipitation, which used equal-width binning. The variables Temperature, Dewpoint, Humidity, Surface Pressure, and Wind Speed receive scores from normalization within the range of 0-1. The total precipitation variable also gets a score of 0-1 but as a result of binning into four categories: light rain (score: 0.25), moderate rain (0.5), heavy rain (0.75), and hefty rain (1). The data is divided into two parts: training data and testing data.

The next step with ANN is to achieve optimal results; several treatments are carried out to train the model. It starts with the use of a standard scaler for data normalization, the use of 2 hidden layers (the first hidden layer using 100 neurons, and the second hidden layer using 50 neurons), a maximum of 500 iterations, activation function: ReLU, and optimizer: ADAM. At the end of this activity, model 2 for prediction is saved. The model generated with ANN shows a MAPE evaluation result of 0.3585, which means an average difference of 32.93% between the model's predictions and the actual data. The mean squared error between the expected and actual values is represented by an MSE of 0.11, indicating a small prediction error, meaning the model's performance is quite good.

The final step is a combination of methods (Fig 7), using models 1 and 2 sequentially, to forecast variable X for several years using model 1 for testing purposes. Then, the prediction of the Y variable is based on the forecast results of the X variable. Final evaluation results of the

combination of methods in MSE and MAPE are obtained; MSE: 3.03 shows that the average square error is relatively low, and MAPE 0.3293 means that the forecast misses by 32.93% of the actual value). In the context of time-series forecasting, it isn't easy to obtain high accuracy compared to prediction techniques, so this value indicates a reasonable level of accuracy.

Table 5 presents a comparison between the Triple Exponential Smoothing (ES) method, which uses a seasonal period of 48 and achieves a Mean Absolute Percentage Error (MAPE) of 66.51%, and the Ensemble Method, which was tested with seasonal periods of 12, 24, 36, and 48. The optimal performance was observed with a seasonal period 47, resulting in a MAPE of 32.93%. It is important to note that a seasonal period 60 could not be applied as it failed to meet the requirement of at least two complete cycles.

Table 5: Evaluation

Metric	Triple ES (48)	Ensemble (12)	Ensemble (24)	Ensemble (36)	Ensemble (48)
MAPE	66.51%	35.85%	35.10%	36.73%	32.93%
MAE	79.44	1.28	1.23	1.30	1.20
MSE	12472.07	3.28	3.33	3.10	3.03
RMSE	111.68	1.81	1.83	1.76	1.74
R^2	0.29	0.51	0.50	0.53	0.54

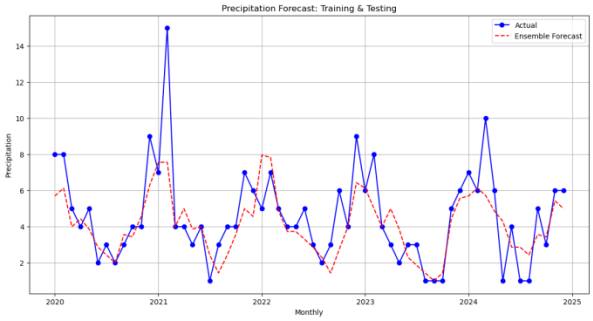


Figure 7: Precipitation forecast, training & testing

Once the model's accuracy is deemed sufficient, a forecast is made using this model for the next 20 years (2025-2045) in Fig 8. The results of this forecast are used for visualization purposes.

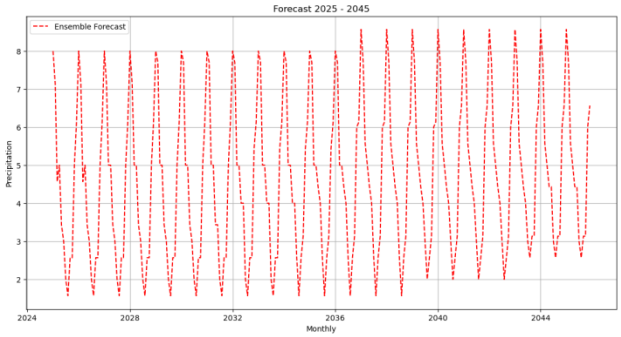


Fig 8: Precipitation forecast

4.4 Visualization in the form of a flood hazard map

Visualization in the form of a hazard map can be done using year and month parameters. Fig 2 shows the flood hazard calculation results with the variables elevation, river distance, runoff, and precipitation in a time-series manner. Because elevation, river distance, and runoff are relatively constant, rainfall provides the main difference in this time-series-based flood hazard research. When preparing all raster data, a transformation is performed to set the coordinate reference system (CRS) to EPSG:4326 with the desired resolution.

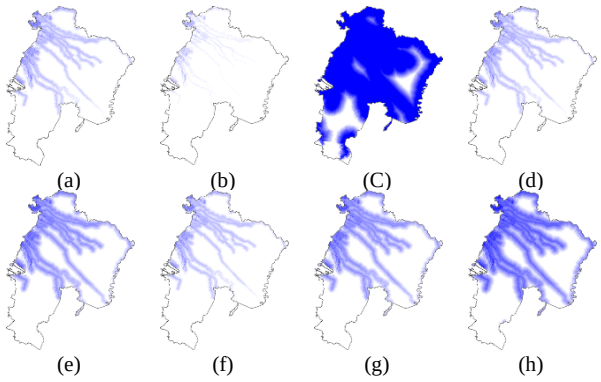
Elevation data was created by combining several rasters that form the Demak Regency and then clipped according to the regional boundaries. Because the elevation of the sea area is ignored, the empty values are changed to NaN, and the values are also set to 0 if they are negative (beneath the surface of the sea). The flood hazard weight calculation results indicate that the elevation variable's weight is negative. A negative weight for the elevation variable means that an increase in elevation tends to reduce the value calculated in the model. In the context of flood danger, the higher an area is, the less likely it is to be affected by floods. The maximum elevation value of 254.16 meters indicates that Demak has relatively low topography, making it more at risk of flooding, mainly since Demak is located in a coastal area with many rivers. The low elevation increases the likelihood of flooding during heavy rainfall or rising sea levels. Because the location of Demak Regency is generally close to the sea, scoring does not use height categories as usual for visualization purposes. Instead, it uses min-max normalization, which means the score ranges from 0 to 1 relative to the area. In performing this normalization, because negative elevation values (below sea level) are replaced with zero, the zero value is considered the minimum elevation value. The maximum elevation value is calculated; if NaN values are in the pixels, they will be ignored. Normalization is done by dividing each value by the maximum value that has been calculated. The normalized values are then inverted so that the highest value becomes the lowest and vice versa.

A distance transformation creates a raster for river distance, and the results are converted from degrees to meters. Like elevation, min-max normalization is also used to score for river distance. After normalization, the values are reversed so that the most significant value becomes the smallest and vice versa. Using inverse values in the score of the nearest distance to the river is related to the method of flood risk assessment. The closer an area is to the river, the higher the flood risk. By reversing the distance value, areas closer to the river will receive a higher score, indicating a more significant flood risk. Conversely, areas further from the river will receive a lower score, indicating a more negligible risk. The maximum distance to the river of 12,807.14 meters means that the farthest point in the analyzed area is approximately 12.8 km from the river.

Raster Land Cover was obtained from Copernicus Global Land Cover Layers: CGLS-LC100 Collection 3 [48], by

setting a coordinate point in Demak Regency and then defining a polygon with a radius of 50 km from the center point. Then, the Land Cover value is converted into Runoff according to the runoff coefficient table [23]. The runoff value is already in score form, so normalization, binning, or scoring is not needed again. A positive weight on this variable indicates that a higher runoff value means a higher risk of flooding. A higher runoff value means more surface water flows rather than being absorbed into the ground. This increases the potential flood risk because the ground surface cannot effectively absorb rainwater, causing water to flow and accumulate in certain areas. A high runoff value can be caused by various factors, such as excessive rainfall, impermeable soil, or land use changes like roads and buildings that increase the area of impermeable surfaces. Runoff forecasting uses double exponential smoothing to obtain the trend component and forecast results. However, the percentage change from the forecast results shows minimal annual variation, especially in the early years, with a slight upward trend in the following years; this may be due to the limited data. This increase in value is not significant enough to be considered a major factor in flood hazard assessment. Given the slight fluctuations and relatively stable trends, this runoff value is regarded as constant in the context of flood hazards, which is why this study only uses data from 2019, the most recent data available.

The use of Triple ES and NN causes the total rainfall forecast to produce a repeating pattern (Fig 9), the trend component remains but does not dominate, so there is no forecast of annual rainfall increase. The use of forecast data for creating flood hazard maps is done by providing the year and month as parameters, then the rainfall values are taken from the forecast data to form rainfall raster data. In determining the rainfall value for each pixel in Demak Regency as the downstream area, a multiplication is performed between the rainfall value of the upstream area (Grobogan) and the river distance that has been normalized and inverted. The closer to the river, the more the rainfall influence value is considered to be the same as the upstream area.



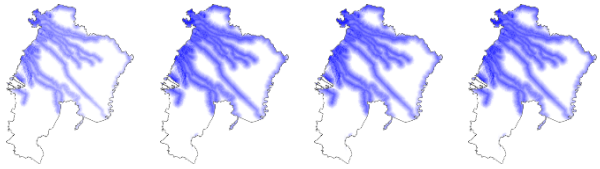


Figure 9: Flood hazard map, based on historical data and forecast results:
(a) 2012-02, (b) 2012-12, (c) 2021-02, (d) 2021-11, (e) 2026-01, (f) 2026-02, (g) 2030-01, (h) 2039-01, (i) 2042-02, (j) 2043-01, (k) 2044-01, (l) 2045-01

Flood hazards were historically high in 2021, particularly in February, when alarming risk figures were recorded, indicating the potential for significant environmental and community impacts. However, compared to the previous year, the flood probability has decreased significantly in 2022 and beyond, raising hopes that mitigation efforts are starting to bear fruit.

This study's projection analysis reveals a worrying upward trend in flood risk despite a decline in 2022. Flood risks are projected to increase from 2025 to 2025, with January and February of 2026 being the high-risk months. There is a possibility that the hazard will continue to grow as this trend is not limited to 2026 but will continue beyond (the projections of this study are limited to 2045). This suggests the need to consider future flood risks further and implement mitigation strategies to save communities and the environment from adverse impacts.

5 Conclusion

The conclusion drawn from this research is that the combination method of Triple ES and ANN is capable of predicting rainfall while considering influencing factors, with a MAPE of 32.93%.

Based on a 20-year forecast, flood risk can be estimated on a seasonal basis. Because of the use of seasonal components, the trend component is used only to support the seasonal component and does not represent the overall trend prediction (up and down).

For future research, it is recommended to use methods that only consider the trend component, such as Double Exponential Smoothing.

The flood danger in Demak Regency is not only from river factors but also from tidal flooding, so it is recommended to consider coastal flooding in future research.

Acknowledgements

This research is funded by the Ministry of Education, Culture, Research, and Technology, Directorate General of Higher Education, Research, and Technology of Indonesia. In accordance with the decision letter from the Directorate of Research, Technology, and Community Service Number 0459/E5/PG.02.00/2024.

References

[1] D. K. Hakim, R. Gernowo, and A. W. Nirwansyah, "Flood prediction with time series data mining: Systematic review," *Nat. Hazards Res.*, vol. 4, no. 2, pp. 194–220, Oct. 2023, doi: 10.1016/j.nhres.2023.10.001.

[2] I. Ardiansah *et al.*, "Integrated Streamflow Forecasting System: A Step Towards Smart Flood Management," *Inform.*, vol. 47, no. 9, pp. 109–121, 2023, doi: 10.31449/inf.v47i9.4890.

[3] S. Parvaze, J. N. Khan, R. Kumar, and S. P. Allaie, "Temporal flood forecasting for trans-boundary Jhelum River of Greater Himalayas," *Theor. Appl. Climatol.*, vol. 144, no. 1–2, pp. 493–506, Apr. 2021, doi: 10.1007/s00704-021-03562-8.

[4] Q. Yu, Y. Wang, and N. Li, "Extreme Flood Disasters: Comprehensive Impact and Assessment," *Water*, vol. 14, no. 8, p. 1211, Apr. 2022, doi: 10.3390/w14081211.

[5] B. Jo, "Peta Banjir Demak 2024, Jumlah Korban, & Update Kondisi Terkini," Tirto. Accessed: Mar. 30, 2024. [Online]. Available: https://tirto.id/peta-banjir-demak-dan-update-kondisi-terkini-gW6F

[6] DetikJateng, "Banjir Demak 2023 Rendam Seratusan Desa: Data Lokasi dan Penyebab," Detik. Accessed: Mar. 30, 2024. [Online]. Available: https://news.detik.com/berita/d-6497484/banjir-demak-2023-rendam-seratusan-desa-data-lokasi-dan-penyebab

[7] S. Afifah and D. R. Hizbaron, "Vulnerability assessment of residential buildings to tidal flood hazards in Sriwulan Village, Sayung District, Demak Regency," *E3S Web Conf.*, vol. 200, p. 01008, Oct. 2020, doi: 10.1051/e3sconf/202020001008.

[8] S. Dewi Novita, A. Alif Nor, T. Taryono, M. Muchamad Farid, and K. Dinda Nur Fadila, "Detection Of Flood Hazard Potential Zones By Using Analytical Hierarchy Process In Tuntang Watershed Area, Indonesia," *Geogr. Tech.*, vol. 19, no. 1/2024, pp. 61–78, Dec. 2023, doi: 10.21163/GT_2024.191.05.

[9] A. A. Atmaja, M. Asa, K. A. Alam, D. D. Putra, C. N. Yuanita, and C. Patricia, "Strengthening Resilience in the Face of Extremes: Lessons Learned from the 2024 Demak Flood," RDI Gobal.

[10] Mirza Hanif Al Falah, Tri Retnaningsih Soeprobowati, H. Hadiyanto, A. Rahim, Barra Madini Noor, and N. Permatasari, "Diatom Stratigraphy as a Flood Record in the Lower Tuntang River, Demak, Central Java," *Evergreen*, vol. 10, no. 1, pp. 272–282, Mar. 2023, doi: 10.5109/6781082.

[11] A. Purwanto, P. Paiman, and A. Sudiro, "The Use of Sentinel-2A Images to Estimate Potential Flood Risk With A Multi-Index Approach in The Mempawah Watershed," *Geosfera Indones.*, vol. 8, no. 1, p. 83, Apr. 2023, doi: 10.19184/geosi.v8i1.37156.

[12] K. Prasetyaningtyas, "Prakiraan Daerah Potensi Banjir Bulan Juli-September 2024." Accessed: Jun. 10, 2024. [Online]. Available: https://www.bmkg.go.id/iklim/potensi-banjir.bmkg?p=prakiraan-daerah-potensi-banjir-bulan-juli-september-2024&tag=&lang=ID

[13] Kompas, "BMKG Prediksi Banjir Bandang di Sumbar sampai 22 Mei, Imbau Warga Hindari Lereng Bukit." Accessed: May 14, 2024. [Online]. Available: https://nasional.kompas.com/read/2024/05/14/11021731/bmkg-prediksi-banjir-bandang-di-sumbar-sampai-22-mei-imbau-warga-hindari

[14] AntaraNews, "BMKG prediksi 12 daerah berpotensi alami banjir bandang kategori siaga." Accessed: May 15, 2024. [Online]. Available: https://www.antaranews.com/berita/2504189/bmkg-prediksi-12-daerah-berpotensi-alami-banjir-bandang-kategori-siaga#:~:text=Miming menyebutkan untuk periode tiga hari ke depan%2C,Barat%2C Kalimantan Selatan%2C Kalimantan Timur%2C dan Kalimantan Tengah.

[15] Tempo, "Aceh Flooding Subsides, BNPB Warns of Subsequent Downpour." Accessed: May 15, 2024. [Online]. Available: https://en.tempo.co/read/1867865/aceh-flooding-subsides-bnpb-warns-of-subsequent-downpour

[16] H. Cloke, "Why flooding is still so difficult to predict and prepare for," theconversation.com. Accessed: Jun. 15, 2024. [Online]. Available: https://theconversation.com/why-flooding-is-still-so-difficult-to-predict-and-prepare-for-126866

[17] Tempo, "Great Losses over 4 Years of Hydrometeorological Disasters: BNPB," Tempo. Accessed: Jun. 16, 2024. [Online]. Available: https://en.tempo.co/read/1689461/great-losses-over-4-years-of-hydrometeorological-disasters-bnpb

[18] A. Al-Fugara, A. N. Mabdeh, S. Alayyash, and A. Khasawneh, "Hydrological and Hydrodynamic Modeling for Flash Flood and Embankment Dam Break Scenario: Hazard Mapping of Extreme Storm Events," *Sustainability*, vol. 15, no. 3, p. 1758, Jan. 2023, doi: 10.3390/su15031758.

[19] M. Shariati, M. Kazemi, R. Naderi Samani, A. Kaviani Rad, M. Kazemi Garajeh, and N. Kariminejad, "An integrated geospatial and statistical approach for flood hazard assessment," *Environ. Earth Sci.*, vol. 82, no. 16, p. 384, Aug. 2023, doi: 10.1007/s12665-023-11077-w.

[20] N. Umar and A. Gray, "A water level prediction using ARMA and ARIMA models: A case study of the river Niger," in *GRASPA 2021: The International Environmetric Society*, 2021. [Online]. Available: https://pureportal.strath.ac.uk/en/publications/a-water-level-prediction-using-arma-and-arima-models-a-case-study

[21] W. M. Wong, M. Y. Lee, A. S. Azman, L. A. F. Rose, and F. P. T. dan Teknousahawan, "Development of Short-term Flood Forecast Using ARIMA," *Int. J. Math. Model. Methods Appl. Sci.*, vol. 15, pp. 68–75, 2021.

[22] E.-I. Koutsovili, O. Tzoraki, N. Theodossiou, and G. E. Tsekouras, "Early Flood Monitoring and Forecasting System Using a Hybrid Machine Learning-Based Approach," *ISPRS Int. J. Geo-Information*, vol. 12, no. 11, p. 464, Nov. 2023, doi: 10.3390/ijgi12110464.

[23] M.-J. Chang, I.-H. Huang, C.-T. Hsu, S.-J. Wu, J.-S. Lai, and G.-F. Lin, "Long-Term Flooding Maps Forecasting System Using Series Machine Learning and Numerical Weather Prediction System," *Water*, vol. 14, no. 20, p. 3346, Oct. 2022, doi: 10.3390/w14203346.

[24] A. Borucka, "Seasonal Methods of Demand Forecasting in the Supply Chain as Support for the Company's Sustainable Growth," *Sustainability*, vol. 15, no. 9, p. 7399, Apr. 2023, doi: 10.3390/su15097399.

[25] D. K. Hakim, R. Gernowo, and A. W. Nirwansyah, "Rainfall, Wind Speed, and Temperature Forecast

Using Triple Exponential Smoothing and Gradient Descent," *KnE Soc. Sci.*, Mar. 2024, doi: 10.18502/kss.v9i6.15311.

[26] N. Munier and E. Hontoria, "Shortcomings of the AHP Method," 2021, pp. 41–90. doi: 10.1007/978-3-030-60392-2_5.

[27] P. Li, "Design of AHP-Delphi Emergency Capability Evaluation Index System Model in Management Stage," 2022, pp. 1208–1214. doi: 10.1007/978-981-16-8052-6_170.

[28] J. Lin, C. Sreng, E. Oare, and F. A. Batarseh, "NeuralFlood: an AI-driven flood susceptibility index," *Front. Water*, vol. 5, Oct. 2023, doi: 10.3389/frwa.2023.1291305.

[29] CivilWeb, "Rational Method Runoff Coefficient." Accessed: Aug. 09, 2024. [Online]. Available: https://civilweb-spreadsheets.com/drainage-design-spreadsheets/runoff-and-rainfall-intensity-calculator-spreadsheet/rational-method-runoff-coefficient/

[30] R. B. Mudashiru, N. Sabtu, I. Abustan, and W. Balogun, "Flood hazard mapping methods: A review," *J. Hydrol.*, vol. 603, p. 126846, Dec. 2021, doi: 10.1016/j.jhydrol.2021.126846.

[31] N. I. Saikh and P. Mondal, "GIS-based machine learning algorithm for flood susceptibility analysis in the Pagla river basin, Eastern India," *Nat. Hazards Res.*, vol. 3, no. 3, pp. 420–436, Sep. 2023, doi: 10.1016/j.nhres.2023.05.004.

[32] A. B. Adeyemi and A. A. Komolafe, "Flood Hazard Zones Prediction Using Machine-Learning-Based Geospatial Approach in Lower Niger River Basin, Nigeria," *Nat. Hazards Res.*, Jan. 2025, doi: 10.1016/j.nhres.2025.01.002.

[33] M. Nhangumbe, A. Nascetti, S. Georganos, and Y. Ban, "Supervised and unsupervised machine learning approaches using Sentinel data for flood mapping and damage assessment in Mozambique," *Remote Sens. Appl. Soc. Environ.*, vol. 32, 2023, doi: 10.1016/j.rsase.2023.101015.

[34] Z. Demissie, P. Rimal, W. M. Seyoum, A. Dutta, and G. Rimmington, "Flood susceptibility mapping: Integrating machine learning and GIS for enhanced risk assessment," *Appl. Comput. Geosci.*, vol. 23, p. 100183, Sep. 2024, doi: 10.1016/j.acags.2024.100183.

[35] H. El-Bagoury and A. Gad, "Integrated Hydrological Modeling for Watershed Analysis, Flood Prediction, and Mitigation Using Meteorological and Morphometric Data, SCS-CN, HEC-HMS/RAS, and QGIS," *Water*, vol. 16, no. 2, p. 356, Jan. 2024, doi: 10.3390/w16020356.

[36] H.-J. Wang, R. Merz, S. Yang, and S. Basso, "Inferring heavy tails of flood distributions through hydrograph recession analysis," *Hydrol. Earth Syst. Sci.*, vol. 27, no. 24, pp. 4369–4384, Dec. 2023, doi: 10.5194/hess-27-4369-2023.

[37] E. Lechowska, "Approaches in research on flood risk perception and their importance in flood risk management: a review," *Nat. Hazards*, vol. 111, no. 3, pp. 2343–2378, Apr. 2022, doi: 10.1007/s11069-021-05140-7.

[38] A. Singh, D. Dawson, M. Trigg, and N. Wright, "A review of modelling methodologies for flood source area (FSA) identification," *Nat. Hazards*, vol. 107, no. 2, pp. 1047–1068, Jun. 2021, doi: 10.1007/s11069-021-04672-2.

[39] S. Sandiwarno, "Penerapan Machine Learning Untuk Prediksi Bencana Banjir," *J. Sist. Inf. Bisnis*, vol. 14, no. 1, pp. 62–76, Jan. 2024, doi: 10.21456/vol14iss1pp62-76.

[40] F. Karim, M. A. Armin, D. Ahmedt-Aristizabal, L. Tychsen-Smith, and L. Petersson, "A Review of Hydrodynamic and Machine Learning Approaches for Flood Inundation Modeling," *Water*, vol. 15, no. 3, p. 566, Feb. 2023, doi: 10.3390/w15030566.

[41] B. K. Panigrahi, S. Das, T. K. Nath, and M. R. Senapati, "An Application of Data Mining Techniques for Flood Forecasting: Application in Rivers Daya and Bhargavi, India," *J. Inst. Eng. Ser. B*, vol. 99, no. 4, pp. 331–342, Aug. 2018, doi: 10.1007/s40031-018-0333-9.

[42] M. Nayak, S. Das, and M. R. M. R. R. Senapati, "Improving Flood Prediction with Deep Learning Methods," *J. Inst. Eng. Ser. B*, vol. 103, no. 4, pp. 1189–1205, Aug. 2022, doi: 10.1007/s40031-022-00720-y.

[43] I. N. da Silva, D. Hernane Spatti, R. Andrade Flauzino, L. H. B. Liboni, and S. F. dos Reis Alves, *Artificial Neural Networks*. Cham: Springer International Publishing, 2017. doi: 10.1007/978-3-319-43162-8.

[44] J. Thioulouse, S. Dray, A.-B. Dufour, A. Siberchicot, T. Jombart, and S. Pavoine, *Multivariate Analysis of Ecological Data with ade4*. New York, NY: Springer New York, 2018. doi: 10.1007/978-1-4939-8850-1.

[45] M. Akkaya, "Vector Autoregressive Model and Analysis," in *Handbook of Research on Emerging Theories, Models, and Applications of Financial Econometrics*, Cham: Springer International Publishing, 2021, pp. 197–214. doi: 10.1007/978-3-030-54108-8_8.

[46] F. Kronthaler, *Statistics Applied With Excel*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2023. doi: 10.1007/978-3-662-64319-8.

[47] M. Buchhorn, M. Lesiv, N.-E. Tsendbazar, M. Herold, L. Bertels, and B. Smets, "Copernicus Global Land Cover Layers—Collection 2," *Remote Sens.*, vol. 12, no. 6, p. 1044, Mar. 2020, doi: 10.3390/rs12061044.

[48] M. Buchhorn *et al.*, "Copernicus Global Land Service: Land Cover 100m: collection 3: epoch 2015: Globe," Sep. 2020, *Zenodo*. doi: 10.5281/zenodo.3939038.