Hybrid Panel Data Forecasting for Coastal Flood Hazard Assessment: A Spatial Data Mining Approach Using Exponential Smoothing and Neural Networks

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

Keywords: flood hazard, panel data, forecast, coastal, exponential smoothing, neural network

Received: Januar 30, 2025

Due to climate change–induced sea level rise, coastal flood risk is increasing significantly, creating an urgent need for improved flood risk assessment and mitigation strategies. Effective flood analysis requires both time series forecasting and hazard risk evaluation, including the determination of appropriate weights for flood-related parameters such as elevation, runoff, distance to the shoreline, and sea level. Traditional forecasting methods like exponential smoothing are limited in capturing relationships between variables, while neural networks can model non-linear interactions but are less commonly applied to long-term forecasting. To address these limitations, this study proposes a hybrid method that integrates Exponential Smoothing (ES) and Neural Networks (NN) for panel data analysis, where ES identifies trends and seasonal patterns and its output is used as additional input for NN. The NN is also employed to objectively determine parameter weights, reducing the subjectivity of conventional AHP-based approaches. The method integrates geographic and climatic variables, including wind speed, temperature, sea surface pressure, and rainfall, and is applied to coastal areas in Semarang, Demak, and Jepara, Indonesia. Results show that the hybrid model outperforms standard ES and NN methods, achieving flood forecasting errors (MAPE) between 3.5% and 8.3% and parameter weighting accuracy of 88–94%, contributing to a more holistic and reliable flood risk analysis.

Povzetek: Raziskava predlaga hibridni pristop, ki združuje eksponentno glajenje in nevronske mreže za natančnejše napovedovanje obalnih poplav ter objektivno določanje uteži poplavnih dejavnikov, s čimer izboljša oceno poplavnega tveganja na podlagi geografskih in podnebnih podatkov.

1 Introduction

Rob floods are one type of flood that increasingly threaten coastal areas worldwide, occurring as a phenomenon when high tide sea water overflows onto land. Global warming causes rising sea levels, increase the risk of coastal flooding by up to five times this century worldwide, putting more than 70 million people in danger in increasingly expanding floodplain areas, with projections of loss of land and critical infrastructure due to permanent inundation [1]. The fact that many major cities and communities are situated along the shore, making more people and infrastructure susceptible to coastal flooding, makes this situation much worse. By 2050, hundreds of densely populated coastal cities are expected to face increased flood risk, with that risk doubling by 2100 [1]. This increase could lead to a setback in human development globally, especially in coastal areas that are centers of important social and economic activity. In addition to being caused by human activities such as urbanization and land-use change, climate change factors also exacerbate these conditions through rising sea levels and increasing frequency of floods. Therefore, a deep understanding of flood risk and developing effective mitigation strategies are critical to protecting coastal communities from such negative impacts.

Hazard risk analysis and time-series forecasting are important approaches in flood mitigation [2], [3], [4]. This involves assessing the flood potential in an area, with: evaluation of historical flood data [5], [6], [7], [8], weight determination [2], [9], [10], and future climate projections to predict where and when floods might occur [11], [12]. Predictive models in this field are increasingly being enhanced by the application of Spatial Data Mining (SDM), which enables the analysis of large datasets to identify trends and improve forecast accuracy [13], [14]. SDM plays a crucial role in enhancing predictive models for flood forecasting by enabling the analysis of large datasets to identify trends and improve forecast accuracy. It integrates various data types, including historical flood data and geographical,

climatic, and land use variations, essential for practical flood risk analysis. The application of SDM allows for identifying non-linear correlations between variables that traditional forecasting methods, such as univariate approaches, often overlook. By combining SDM with techniques like Artificial Neural Networks (ANN) and Exponential Smoothing (ES), predictive models can achieve significantly improved accuracy, as evidenced by a hybrid model that achieves an error smaller than single methods [1], [2], [6]. This methodological advancement is significant in regions like Indonesia, where geographical and climatic variances can significantly affect flood dynamics.

Long-term forecasting techniques used in predicting floods in time series generally employ a univariate approach [15], with time series data, such as the exponential smoothing method, which has the advantage of simplicity in the mathematical formulas used. However, exponential smoothing has the disadvantage of being limited in depicting relationships with other influencing variables [16], [17]. The solution to the problem of the relationship between variables causing floods can use non-linear approach neural networks, with cross-sectional data, which has the advantage of making fairly accurate estimates in the short term but does not have the advantage of long-term forecasting like the univariate approach [18]. Both types of data, time-series and cross-section data, when combined in the form of panel data, can be used for time-series forecasting analysis that links the relationships between variables in making forecasts, by combining both approaches (time-series forecast and non-linear estimation) with hybrid method (Triple Exponential Smoothing and Neural Network). The hybrid method approach is expected to offer solutions that not only improve the accuracy of predictions but also adaptive flexibility to environmental dynamics, making them a relevant and strategic topic in today's era of climate change [13], [19], [20], [21].

Previous studies with various methods (ARIMA, SARIMA, LSTM, etc.) have highlighted the effectiveness of different forecasting models in coastal flood forecasting but have not consistently reported MAPE values as part of their evaluation metrics, focusing only on RMSE and MAE [22], [23], [24]. Although Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) are valuable tools for assessing model performance in a particular context or study, they have limitations in terms of cross-study comparisons due to their subjectivity related to the characteristics of the data [25], [26]. In contrast, MAPE provides a more standardized approach that facilitates comparisons across multiple scenarios [27], [28], so the use of MAPE in coastal flood forecasting studies is a novelty.

To provide accurate visualization in the form of a flood hazard map, it is critical to determine the proper weight of flood risk parameters so that each pixel on the map can be given a representative risk score. This weight reflects the degree of influence of each variable, such as elevation, runoff, distance to water sources, and sea level, on the potential for flooding in the location. Weighting is essential because of geographic variation. In flood hazard risk analysis, it is necessary to determine the weight values for various parameters that influence flood risk. Geographic, climatic, and land use variations in different regions of Indonesia cause the flood hazard weights to differ in each area. The Analytic Hierarchy Process (AHP) method, which is often used to determine parameter weights, is quite effective in empirical analysis [29], [30]. However, this method needs to be improved due to subjectivity and the potential bias of experts [31]. AHP relies heavily on subjective judgments for assigning weights to different criteria, which can introduce bias and inconsistency [32], [33]. Additionally, AHP may struggle with handling complex, nonlinear relationships among flood risk factors and can be limited in capturing the dynamic interactions between variables over time and space.

In this study, we propose an innovative solution using Historical Flood Data-Based Neural Networks to determine parameter weights more objectively and accurately. This approach is expected to reduce bias and enhance the reliability of flood risk analysis through data-based optimization [2], [33]. Neural network-based approach addresses these limitations by learning directly from data, which allows it to model complex, nonlinear patterns in flood risk factors without requiring subjective weight assignments. Neural networks can integrate multiple influential variables and their interactions more effectively, improving prediction accuracy and robustness. Moreover, when combined with GIS and remote sensing data, neural networks provide a more data-driven, adaptive, and precise flood susceptibility mapping compared to the more rigid and expert-opinion-dependent AHP.

The objective of this study is to evaluate the effectiveness of combining Neural Networks and Exponential Smoothing in improving long-term and short-term flood forecasting accuracy in coastal areas. We hypothesize that combining Neural Networks and Exponential Smoothing improves long-term and short-term flood forecasting accuracy compared to single-model approaches. The study focuses on Semarang, Demak, and Jepara due to their significant exposure to flooding events, as evidenced by historical data indicating a high frequency of flood occurrences in these areas. Additionally, these locations are characterized by unique environmental conditions that make them particularly susceptible to the impacts of climate change, thus providing a relevant context for this research.

2 Data

The study was carried out in coastal areas representing tropical climates in Indonesia, specifically in Semarang City, Demak, and Jepara, covering an area of approximately 2,291.456 km² with a total population of 4,819,953 people, where Semarang City has 373.78 km² and 1,694,743 people, Demak has 897.43 km² and 1,252,970 people, and Jepara has 1,020.246 km² and 1,872,240 people based on the latest measurements [34]. The three areas were chosen because they are prone to

hydrometeorological disasters on the north coast of Java, especially coastal floods, caused by a combination of high rainfall, extreme sea tides, and land subsidence due to both human activities and natural processes, and they also have a long history of flood events [35], [36], [37].

For the analysis of coastal flood hazards, several variables are used: sea level data, elevation, runoff (from land cover), and distance to the coastline. Elevation, runoff, and distance are variables that tend to remain constant (not experiencing significant changes), whereas sea level data is a variable that varies over time.

In contrast to other variables, sea level is changing continuously over time as a water surface along the coast. Tidal gauge records and sea tide models are used to recreate normal high tide conditions as well as infrequent but significant extreme events [38]. The study makes use of historical sea level data for the years 1995–2024 from the Indonesian Geospatial Reference System (SRGI) [45]. Many time-based weather variables, such as U10 & V10 (10m Wind Speed), D2M (2m Dew Point Temperature), T2M (2m Temperature), MSL (Mean Sea Level Pressure), SP (Surface Pressure), TP (Total Precipitation), and SWH (Significant Wave Height), are used to forecast sea level with the assumption that all of these variables have an impact on sea level [47], [48], [49]. Wind speed has an impact on water flow and can lead to storm surges, which raises the possibility of coastal flooding [39], [40]. Rainfall patterns and cloud formation are influenced by atmospheric humidity and condensation, which are influenced by the dew point and air temperature [40]. In addition, surface pressure and average sea surface pressure have a direct impact on ocean and atmospheric dynamics that affect water level variations [41], [42]. Both the intensity of tropical storms and the thermal expansion of seawater are influenced by sea surface temperature [42]. Meanwhile, total precipitation increases the amount of water on land [42], which can make runoff worse. Information about wave patterns that can either increase or lessen the impact of extreme tides in coastal areas can be found in the direction of the waves [42]. By combining these variables, the prediction model is able to account for the intricacy of the atmosphere-ocean interaction, producing estimates of flood risk and sea level rise that are more accurate. Hourly weather data recorded as time series with a resolution of roughly 27.83 meters is the source of ERA5 hourly data on single levels (Copernicus), from which weather data is derived [46].

Elevation is fundamental in flood threat analysis because it determines which areas are physically vulnerable to inundation. For example, low-lying areas are more susceptible to flooding due to their proximity to sea level, directly affecting the level and depth of potential flood inundation [38]. Elevation data is obtained from DEMNAS Indonesia (National Digital Elevation Model), provided by the Geospatial Information Agency (BIG).

Runoff values, derived from land cover data conversion, represent how much rainfall becomes surface runoff rather than infiltrating into the ground [38]. Therefore, areas with impermeable surfaces or specific land uses generate higher runoff, increasing the flood threat. Runoff data is derived from land cover: Copernicus Global Land Cover Layers: CGLS-LC100 Collection 3 [43], which is then converted based on the Rational Method Runoff Coefficient [44].

Distance to the coastline is essential because closeness increases exposure: locations closer to the coast are at greater risk from storm surges and sea level rise events. A spatial relationship between the location of each pixel and its distance from the coastline is concluded when creating flood threat maps [38]. The distance to the coastline is essential because proximity increases exposure: locations closer to the coast are at greater risk from storm surges and sea level rise events, concluding that there is a spatial relationship between the location of each pixel and its distance from the coastline when creating flood threat maps. Distance to the coastline is calculated by measuring the distance of each pixel from Shapefile to the nearest coastline in each area.

3 Method

3.1 State of the art and related works

Forecasting, especially related to flooding, is a research field that continues to be discussed, with various approaches being used. One of the main focuses is on coastal locations vulnerable to rising sea levels and land subsidence, which significantly increases the risk of flooding [52] and ultimately negatively impacts the local economy [53]. Some topics related to Coastal Flood Forecasting (Fig. 1) are flood risk and adaptation response modeling, hydrodynamic modeling, and flood risk assessments.

Coastal flood risk modeling and adaptation response assesses future flood risk and formulates effective adaptation strategies. By analyzing climate change scenarios and their consequences on flood risk, this method integrates physical and social models to understand potential impacts comprehensively [54].

Hydrodynamic modeling examines and predicts water flow behavior and its interactions with the environment [55]. Using principles of physics, mathematics, and engineering, these models accurately simulate water movement under various conditions, including floods, waves, currents, and other hydrologic phenomena.

Flood risk assessment systematically identifies, analyzes, and evaluates the risks associated with flooding [56], [57], [58]. This approach seeks to understand the potential hazards of flooding, evaluate their likely impacts, and assess the vulnerability of infrastructure and populations to flood events.

These approaches can stand alone or be enhanced through integration with statistical or data mining techniques [56], strengthening the ability to predict and respond to coastal flood hazards. This study seeks to explore flood risk assessment through an innovative approach that leverages data mining for accurate estimates.

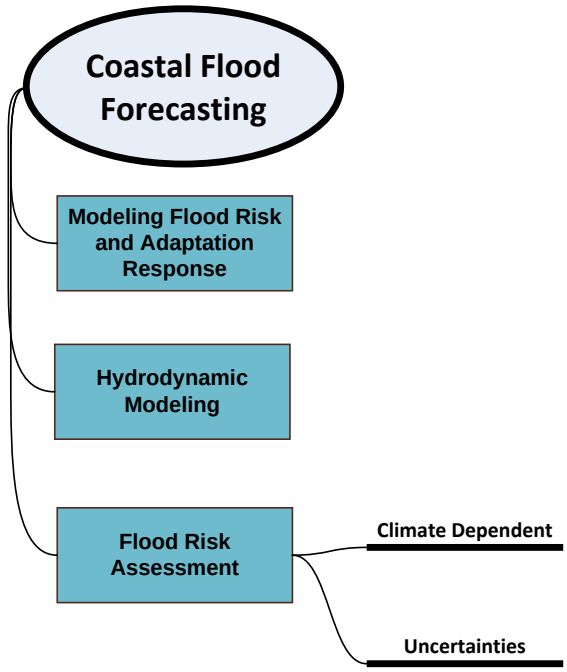

Figure 1: Coastal flood forecasting mind map

Several methods have been used in sea level forecasting associated with risk assessment (Table 1), with various advantages and disadvantages. Using a single method in prediction has the advantages of simplicity and clarity. However, this method has several limitations: limitations in capturing data complexity, risk of overfitting or underfitting, and difficulty handling data variability. In terms of data complexity, forecasting methods are generally univariate (e.g., ARIMA, ES, etc.) which do not consider the influence between variables as in regression approaches (linear regression, neural networks, etc.), so a hybrid approach is proposed by combining forecasting and the influence between variables.

Table 1: Sea level forecasting related works

Data	Techniques	Accuracy Metrics	Limitations
Tide Gauge Data	ARIMA [59], [60], [61]	RMSE	Limited to linear relationships, struggles with non-stationary data
Satellite Altimetry	Neural Network Auto-Regressive (NNAR) [62], [63]	MSE, MAE, RMSE	Requires large datasets, computationally intensive
Tide Gauge + Satellite Data	Extreme Gradient Boosting (XGBoost)	MSE, MAE, RMSE	Sensitive to outliers, overfitting risk
Tide Gauge Data	Facebook Prophet (FProphet) [64]	MSE, MAE, RMSE	Assumes additive seasonality, less effective for multiplicative trends
Tide Gauge Data	Exponential Smoothing [65]	RMSE	linear trends, sensitive to sudden changes
Tide Gauge + Satellite Data	Long Short-Term Memory (LSTM)	MSE, MAE, RMSE	Requires large datasets, computationally intensive
Tide Gauge Data	Moving Average	RMSE	Limited to short-term forecasting, less effective for capturing seasonality
Tide Gauge Data	ARMA	RMSE	Limited to linear relationships, struggles with non-stationary data
Tide Gauge + Satellite Data	Recurrent Neural Network (RNN)	MSE, MAE, RMSE	Requires large datasets, computationally intensive

In several recent studies, hybrid approaches can improve prediction accuracy by reducing bias and variance, thus handling non-linear relationships and complex interactions among environmental variables [66], as well as capturing temporal patterns and long-term dependencies in data [67], which are very important for hydrological predictions. This approach adds an essential dimension to flood vulnerability analysis by emphasizing the use of predictive machine learning to assess relationships among variables [68], making it possible to predict spatial vulnerability levels in areas.

Exponential Smoothing effectively captures trends and seasonality in time series data, providing stable and interpretable predictions. However, this method struggles with non-linear patterns and complex interactions between variables. On the other hand, neural networks excel at modeling non-linear relationships. They can learn complex patterns from large datasets, but they often operate as a black box, making interpretation difficult. Integration of Exponential Smoothing and Neural Networks by combining the strengths of both methodologies is expected to improve the accuracy of sea level forecasting for coastal flood hazard prediction [69].

Integrating these two methods is expected to leverage the stability and interpretability of Exponential Smoothing while leveraging the predictive power of Neural Networks, ultimately resulting in more robust and reliable flood hazard forecasts that can adapt to changing environmental conditions.

3.2 Flood hazard in coastal area

In several disaster risk studies, especially the flood hazard assessment conducted in this study, hazards are measured using a multi-criteria approach, where several variables that make up hazards are given scores and weights according to the level of impact, such as in Indonesia, which uses the Indonesia Disaster Risk Index [70], [71]. The formula used is the classic weighted sum, the sum of the multiplication results between the weight and the value/score of each variable. It is commonly used in risk analysis, multi-criteria decision-making, and hazard evaluation (Eq.1). The Variable Score (S_i) is the value of the intensity of the hazard based on various

factors. Weights (w_i) can be determined in various ways: expert judgment, AHP, or data-driven (used in this study) to assess the contribution of each factor to the overall hazard.

$$\text{Hazard Index} = \sum_i^n w_i S_i \quad (1)$$

$$\text{H} = w_1 S_1 + w_2 S_2 + w_3 S_3 + w_4 S_4 \quad (2)$$

The Hazard Index (H) assesses an area's danger level based on several factors (Eq.2). The first factor, elevation (S_1), is the height of a location from sea level that affects the risk of flooding [71]. The second factor, runoff (S_2), is the water flow above the surface that exceeds the soil's capacity to absorb water [72]. The third factor, distance to the coastline (S_3), is the distance of a location from the coastline, where locations closer to the coast may be more vulnerable to hazards such as tsunamis and hurricanes [73]. The last factor, sea level (S_4), measures the influence of tides that can affect flood risk in coastal areas [74]. Each of these factors is weighted (w_1, w_2, w_3, w_4) according to the degree of its influence on the overall Hazard Index (H).

All variables are stored in raster format (GeoTIFF), and calculations are performed on each pixel with a resolution of about 8 meters (adjusting to the initial Raster: DEM). Each raster has been cut and re-sampled to ensure consistency and accuracy of the data.

3.3 Neural network data driven weighting

The use of feature importance in determining flood hazard weights provides a more objective and data-driven approach compared to traditional methods such as the Analytic Hierarchy Process (AHP), which heavily relies on the subjective judgments of experts. With feature importance, the relative contribution of various parameters is assessed. Differences between regions significantly affect the results of flood hazard assessments because each area has unique geographic characteristics [75], [76], climate, land use, and socio-economic conditions, all of which influence how the flood risk parameters play a role.

Weighting using neural networks can be done by calculating feature importance and measuring the variance of the weights connected to each input feature during the training phase [77], [78]. This process involves assessing how much each feature affects the output by observing the changes in related weights. The more significant the weight change, the more important the corresponding feature influences the output [77]. This method is beneficial for understanding which input variables impact the model's predictions most, thereby allowing for better interpretation and optimization of model performance. The final feature importance scores are calculated based on the variance of the weights combined with the final weights for each input feature; this helps identify the most critical factors affecting the model's output, such as flood hazard assessment.

Several feature variables are used to calculate feature importance in the flood depth prediction model: elevation, runoff, distance to the coastline, and sea level [79], [80]. The flood depth class variable plays a role as the target to be predicted. The value of the flood depth class is obtained from the flood depth classification process, which is measured on a meter scale and then categorized into certain classes to facilitate analysis and modeling.

Each of these feature variables has an important role in flood modeling. Elevation determines the height of the land relative to sea level, which directly affects the risk and level of flooding [81]. Runoff describes the flow of surface water from rainfall and runoff, contributing to the water volume that can cause flooding [82]. Distance to the coastline affects the potential impact of coastal flooding because the closer a location is to the coast [82], the more likely it is to be flooded due to rising sea levels. Sea level is in the form of sea level fluctuations that can worsen flood conditions, especially in coastal areas [82].

Calculating feature importance in neural network [77] starts with data preparation, building a model, training, evaluation, and the last is feature importance analysis (Fig. 2). Building model involves: determining the number of hidden layers and neurons, using ReLU (Eq.3) and softmax (Eq.4) activation functions, and finally initializing weights and biases.

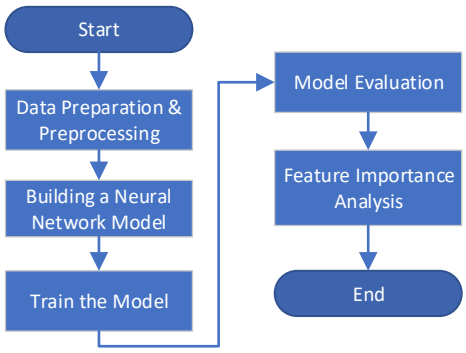

Figure 2: Weighting with NN

$$f(x) = \max(0, x) \quad (3)$$

$$P_i = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}} \quad (4)$$

The ReLU (Rectified Linear Unit) activation function is a straightforward function in neural networks, which converts all negative input values to zero and maintains positive values as they are (Eq.3), with a graph in the form of a straight line on the positive side and a flat line on zero for the negative side, so that non-linear relationships are captured yet efficiently while keeping computation light [83]. The softmax activation function (Eq.4) is used on the output layer of the multiclass classification model to convert the raw output value (Z_i) into an interpretable probability (P_i) by converting the input vector ($z = (z_1, z_2, \ldots, z_K)$ into a probability distribution ($p\ [p_1, p_2, \ldots\ldots, p_K]$), each output element is in the range of 0 to 1 and the sum of all components is 1, so that softmax provides a prediction in the form of the odds

of each class (the class with the highest probability is selected as the prediction result) [83].

During the training phase (Fig.2), backpropagation is performed to calculate the output, calculate MSE & MSE, and update the weights based on the loss gradient with the Adam optimizer (Adaptive Moment Estimation) to adaptively adjust the learning level for each parameter to accelerating convergence and improving model performance [84]. Adam is an optimization algorithm that combines the concepts of momentum and adaptive learning rate concepts by calculating the moving average of the first gradient (momentum) and the second quadratic gradient to adjust parameter updates adaptively. At iteration t, Adam calculates the biased first moment estimate m_t (Eq. 5), which is a moving average of current gradients gt weighted by decay rate β_1; similarly, it computes the biased second raw moment estimate v_t (Eq. 6), a moving average of squared gradients weighted by decay rate β_2. These estimates are then bias-corrected to obtain unbiased moments (Eq. 7 and Eq. 8). Finally, parameters are updated (Eq. 9) where $\alpha > 0$ is the learning rate controlling step size, and small constant ($\epsilon > 0$) ensures numerical stability to avoid division by zero. This mechanism allows Adam to efficiently adapt to complex loss landscapes, accelerating convergence during training.

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1)\mathrm{g_t} \quad (5)$$

$$v_t = \beta_2 v_{t-1} + (1 - \beta_2)\mathrm{g}_t^2 \quad (6)$$

$$\widehat{m_t} = \frac{m_t}{1 - \beta_1^t} \quad (7)$$

$$\widehat{v_t} = \frac{v_t}{1 - \beta_2^t} \quad (8)$$

$$\theta_{t+1} = \theta_t - \alpha \frac{\widehat{m_t}}{\sqrt{\widehat{v_t}} + \epsilon} \quad (9)$$

The Magnitude of Weights (Absolute Weights) on the Input Layer is used to calculate feature importance, which is often used to measure feature importance in neural networks. It is done by referring to the absolute value of the weights (w_{ij}) that connect the input neuron to the neuron in the next layer (Eq. 10), where the i index indicates the i input feature and the j index indicates the neuron in the hidden layer. By calculating the sum or average of the absolute value of the weight, the contribution value (influence) of the feature on the output of the model is obtained. The greater the magnitude of a feature's weight, the more important the role it plays in the model's prediction process because small changes to that feature will significantly impact subsequent neuron activation.

$$\mathrm{FI_i} = \sum_{\mathrm{j=1}}^{\mathrm{n}} |W_{ij}| \quad (10)$$

3.4 Time series, cross-section, and panel data

Time series data is a sequence of data points collected or recorded at consecutive time points (e.g., Table 2), with the characteristics of a date (period) and one actual variable [85] , which are analyzed using a univariate time series forecasting approach such as the Triple Exponential Smoothing (TES) method [17].

Table 2: Sample of time series data

Date	Sea level (cm)
2022-11	54.3
2022-12	55.1
2023-01	49.8
2023-02	51.3

Cross-sectional data is data collected by capturing snapshots of various variables without considering changes over time. Cross-sectional data (example: Table 3) are analyzed using a non-linear estimation approach [86], [87] like Neural Network (NN), which is effective in capturing complex and non-linear relationships in the data. NN utilizes learning algorithms to model intricate patterns and interactions among variables, making it suitable for handling cross-sectional data.

Table 3: Sample of Cross-sectional data

Location	Temperature	Wind Speed	Sea Level
Demak	30.82	1.1302	54.3
Demak	30.47	1.4842	55.1
Jepara	29.87	0.8074	49.8
Jepara	30.48	1.2403	51.3
Semarang	29.54	1.1908	44.7
Semarang	30.45	1.6641	55.6

Panel Data, which is a combination of time series and cross-sectional data, for example, sea level data collected from several locations at different points in time (example: Table 4).

Table 4: Sample of panel data

Date	Location	Temperature	Wind Speed	Tide
2022-12	Demak	30.82	1.1302	54.3
2023-01	Demak	30.47	1.4842	55.1
2022-12	Jepara	29.87	0.8074	49.8
2023-01	Jepara	30.48	1.2403	51.3
2022-12	Semarang	29.54	1.1908	44.7
2023-01	Semarang	30.45	1.6641	55.6

The data panel has a key advantage over purely time-series or cross-sectional data by integrating diverse data sources from time and space in flood risk assessments, emphasizing the need for dynamic models that can incorporate changing variables such as land use, climate, etc [88]. The evolution of flood modeling techniques increasingly prioritizes integrating spatial and temporal

data, supporting models that can dynamically respond to changes in environmental and anthropogenic factors [2]. Various case studies and methodologies also highlight the importance of applying advanced data techniques, including spatial-temporal modeling, to reduce uncertainty in flood risk assessments more accurately and effectively [89]. Panel data enables evaluation how flood risk mitigation measures perform over time and across regions.

In this study, UTC was used as a global time standard, which provides a consistent reference, which is officially abbreviated from Coordinated Universal Time as a neutral compromise between the English terms (Coordinated Universal Time) and the French term (Temps Universel Coordonné)—used [90]. Panel data is stored in Excel format and then imported into Pandas DataFrame using Python to facilitate data analysis and processing.

3.5 TES & NN combination to forecast panel data

Temporal dependencies in the panel data are included in two main steps [91] that carried out on the panel data; the first step focuses on the classic statistical method of Triple Exponential Smoothing (TES) to capture trend and seasonal patterns in the data, while the second step combines the approach with a NN model that utilizes feature selection [92] in the form of the selection of relevant weather features as inputs to capture complex non-linear relationships. The combination of the results of the two and aims to produce predictions that utilize each method's strengths and optimize input features [93], which models non-linear patterns and complex interactions between variables. Thus, this integration allows the model to effectively accommodate temporal dependencies while improving the accuracy of predictions on multivariate panel data.

The use of the TES for time series data is characterized by calculations: Level, Trend and Seasonal [94], which is a moving forecasting method that gives weight in stages to the latest data, so that if the data is new, the forecasting value will be updated.

The data is divided into a period (annual / biennial / five years / etc.), the initial level value is obtained from the average actual value in the first period (Fig.2), the initial trend value is obtained from the average value of the difference in the actual value of each month in the same 2 periods and divided by the number of months, and the initial seasonal value for each month is the actual value divided by the Initial Level value. Then the Level (L_t), Trend (T_t), Seasonal (S_t) and Forecast (F_{t+m}) are calculated. Smoothing parameters used: α as level parameters, β as trend parameters, and γ as seasonal parameters. Y_t is the value of observation at time (t). S_{t-m} is a seasonal factor in the previous period, L_t is the smoothed level, T_t is a smoothed trend, and (m) is the length of the season.

$$L_t = \alpha\left(\frac{Y_t}{S_{t-m}}\right) + (1-\alpha)(L_{t-1} + T_{t-1}) \quad (6)$$

$$T_t = \beta(L_t - L_{t-1}) + (1-\beta)T_{t-1} \quad (7)$$

$$S_t = \gamma\left(\frac{Y_t}{L_t}\right) + (1-\gamma)S_{t-m} \quad (8)$$

$$F_{t+m} = (L_t + mT_t)S_{t-m+k} \quad (9)$$

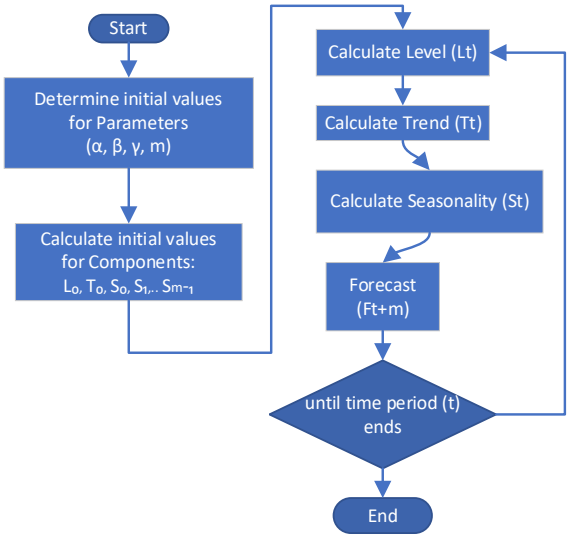

Figure 3: TES

The calculation of some of these parameters (Fig.3) is repeated until a minimum error is obtained through the optimization process up to the specified period (t). Each iteration is performed by updating the model parameters using the L-BFGS-B algorithm that efficiently searches for the optimal solution and ends when the convergence criteria are met (a slight change in the value of the error function or reaching a predetermined maximum number of iterations). The resulting model is expected to have optimal performance in predicting data in a certain period. Various optimization methods have been tried to perform smoothing parameter optimization, including popular algorithms such as Adam, Stochastic Gradient Descent (SGD), etc. However, the results of the experiments showed that L-BFGS-B provided the highest accuracy compared to these alternatives. L-BFGS-B consistently produces more optimal objective function values and predictions with the highest precision [95], so it is considered superior in smoothing parameter optimization. L-BFGS-B uses Hessian matrix approximation efficiently, allowing faster and more stable convergence than gradient-based methods such as SGD or Adam [84].

To make predictions from the forecast results (cross-sectional data), NN is used on weather data to predict sea level (Fig.4). The data is divided into training and testing data, each covering a specific year period. The features used to train the model are several columns that have been normalized. Then, several hidden layers, neurons, and the ReLU activation function (Eq.10) are used. The prediction results are evaluated using the MAPE (Eq.11) and MSE (Mean Squared Error) (Eq.12) metrics [94].

$$\mathrm{ReLU}(x) = \max(0, x) \quad (10)$$

$$\text{MAPE} = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{A_t - F_t}{A_t}\right| \times 100 \tag{11}$$

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2 \tag{12}$$

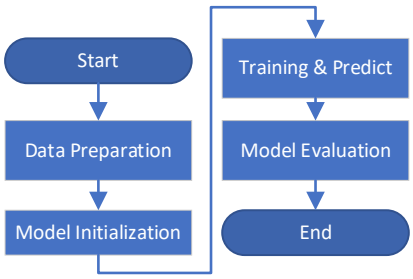

Figure 4: NN for prediction

There is no set rule for choosing the number of hidden layers or neurons. This decision is highly dependent on the data's nature and the modeling's purpose, so it is often trial-and-error or validation-based tuning [96], [97]. The NN architecture design chosen in this research, which is six hidden layers with 50 neurons each, is based on empirical experiments and the complexity of the modeled data. The architecture was chosen to balance the capacity of the model to study complex patterns while avoiding over-installing. Using multiple hidden layers allows the model to capture hierarchical features in the data, which is crucial in flood prediction by considering the relationships between variables.

The activation functions, ReLU, were selected based on performance in capturing non-linear relationships [98]. ReLU was chosen for its ability to reduce the problem of disappearing gradients [99], allowing for faster convergence during training [100], [101].

The hybrid forecasting method that combines TES and NN aims to solve panel data problems, improving prediction accuracy by leveraging the strengths of each technique, resulting in forecasting data that considers the influence between variables.

The first step is to forecast each variable using TES, effectively capturing seasonal patterns and trends in the data. The forecast results from TES are combined into a single dataset used as input for the NN model. NN is trained using this dataset to capture non-linear patterns and complex interactions between variables that TES may not capture. After the NN model is trained, predictions are made on the testing data, and the results are evaluated using metrics such as MAPE and MSE to measure the model's accuracy. By combining these two methods, the forecasting results are expected to be better than the univariate approach, as TES handles seasonal patterns and trends while Neural Networks capture complex prediction patterns.

3.6 TES & NN combination to forecast panel data with predicted bias

Bias refers to systematic errors that occur during training, leading to an inaccurate representation of the studied population [102], [103]. To improve prediction accuracy, the value of bias is optimized during the training process. This optimized bias can then be used in ensemble methods, combining predictions from multiple models to enhance overall accuracy [104]. Specifically, the output from neural networks, including the optimized bias, can be combined with the output from other models to generate final predictions. This approach aims to correct bias in time series data and improve the accuracy of forecasting models.

Analyzing the panel data with Average Bias involves two key stages: TES is used to identify trends and seasonal patterns within the dataset; subsequently, this method is integrated with the NN model that employs feature selection by choosing pertinent weather variables as inputs to model complex non-linear relationships. By combining the outputs of both processes and applying bias correction, the approach aims to generate predictions that leverage the strengths of each technique while optimizing the selected input features.

$$Bias = \frac{1}{N}\sum_{i=1}^{N}(\mathrm{y_i} - \hat{y}_\mathrm{i}) \tag{13}$$

Bias in data mining refers to systematic errors that occur during training, leading to an inaccurate representation of the population being studied. Bias can arise from the data used or from the algorithm itself that shows a certain tendency. Bias assessment can be done data-driven using an estimation approach, such as the NN method. In a NN, bias is a parameter that is added to the input of each neuron before an activation function is applied. The Bias prediction carried out in this study is post-processing by adjusting the model output to be fairer or correcting the prediction based on specific metrics [105]. In detail, it is calculated by predicting the bias value for each sample in the test data using a trained model and then taking the average of all these predictions ((Eq.13)) to obtain one representative value [106]. This average value is then normalized by dividing it by the number of features (variables) used in the model. The result is a scalar value representing the overall bias estimate after normalization. The use of a single bias value from a NN (or similar method) can be used in ensemble methods, which combine predictions from multiple models to improve accuracy. In this case, the output from the NN (including the bias that has been optimized during training) can be combined with the output from other models to make the final prediction.

3.7 Forecasting model accuracy

To evaluate how well a model can predict the true values, including in the context of flood forecasting (sea level rise), several metrics are often used: MSE, RMSE and MAPE.

MSE (Mean Squared Error) measures the average of the squared errors between the predicted and actual values. While RMSE (Root Mean Square Error) measures the average of the squared errors between the predicted and actual values, then takes the square root. MSE and RMSE give a larger penalty for larger errors and are the basis for calculating RMSE [107]. MSE is the basis for calculating RMSE which is more sensitive to outliers [13], [108] so it is important in flood forecasting because flood events often involve extreme values (severe flood events).

MAPE (Mean Absolute Percentage Error) measures the relative error between predicted and actual values in percentage form. Unlike other metrics, MAPE is objective to the data and easy to interpret, as it is expressed as a percentage. It can be classified into the following categories [107]: MAPE < 10% is considered very good, 10% ≤ MAPE < 20% is good, 20% ≤ MAPE < 50% is sufficient, and MAPE ≥ 50% is deemed poor. All metrics are used, as an approach to evaluate how well the model can predict flood peaks or flood occurrence times.

A hybrid method between NN and TES is used, aiming to combine the advantages of both techniques to improve forecasting accuracy. Then, the gradient descent method is used as an optimization to obtain optimal weight values for NN and alpha (level), beta (trend), and gamma (seasonal) values for TES. The result of the analysis of the gathered data will be a geographic information system that shows maps of Flood Hazards (H).

3.8 Raster visualization for flood hazard analysis

Raster visualization for flood hazard analysis plays a crucial role in disaster risk management by providing a detailed, spatially explicit representation of flood hazard levels across a region [109]. Flooding poses significant threats to lives, infrastructure, and economies, especially in coastal and low-lying areas [110]. Accurate flood hazard maps enable policymakers, urban planners, and emergency responders to identify vulnerable zones, prioritize mitigation efforts, and design effective evacuation plans, ultimately reducing the impact of flood events on communities.This process begins with creating a raster for elevation by combining and cropping several GeoTIFF files from the Digital Elevation Model (DEMNAS) according to the region's boundaries [111], [112]. Using raster-based methods for final visualization allows for a fine-grained depiction of flood hazard levels, where each cell is assigned a score based on predicted or observed values, effectively highlighting areas with varying flood hazard [110], [113], [114].

The coordinate reference system (CRS) is standardized to EPSG:4326, and the spatial resolution is set to 8.35 meters, matching the original resolution of the DEMNAS data [115]. Maintaining this high resolution preserves critical topographic details for accurately analyzing elevation variations—a key determinant of flood risk. Furthermore, consistent resolution facilitates the integration of diverse datasets, enhancing the reliability and informativeness of the resulting flood hazard map.

Elevation data undergoes min-max normalization, where higher elevations correspond to lower flood risk scores, reflecting the inverse relationship between elevation and flood susceptibility. Land cover data is converted into runoff values using the Rational Method Runoff Coefficient, contributing to the overall flood hazard score by accounting for surface permeability and water retention capacity. Distance to the coastline is calculated using a shapefile from the Indonesian Landmark, with an Euclidean distance transform applied to measure proximity to the sea. This distance is normalized and inverted, as areas closer to the coastline generally face higher flood risks due to storm surges and sea-level rise.

A sea level raster is created by combining sea level scores with distance scores linked to estimated sea level values. The method produces a comprehensive flood hazard map by integrating all these factors—elevation, land cover runoff, coastline proximity, and sea level—into a weighted composite raster. This nuanced visualization supports targeted flood risk assessment and informs sustainable coastal management strategies, helping to safeguard vulnerable populations and infrastructure from future flood events.

4 Result

The analysis begins by determining the weight of each coastal flood hazard factor, with each factor assigned a weight based on its level of influence on the coastal flood risk. After the weights of the hazard factors are determined, the next step is to perform univariate forecasting for each variable using the TES. The results of this univariate forecasting are then used as input for the next prediction step using NN. This prediction aims to determine the flood depth class while considering the complex interactions between variables.

4.1 Weight determination with NN

Determining flood hazard weights using neural networks involves several steps, from data collection and processing to model training and evaluation. The first step is to determine the coastal flood database, which contains a history of flood and non-flood events. This data includes several parameters that occur during floods (average height, average runoff, average closest distance to shoreline, and flood depth). Once the data was collected, average values for elevation and runoff parameters were calculated, which were then saved to the flood database. In addition, distance to coastline measurements were taken and saved in raster form for further analysis.

Next, the features used for training and testing the NNmodel were determined. These features include all the parameters that have been collected (elevation, surface flow, distance to coastline, and tide). The target is then the flood depth class. The data is then divided into training and testing sets to train the model and evaluate its performance.

The first layer has 64 neurons, and the second layer also has 32 neurons, both using the sigmoid activation function, which helps capture non-linear patterns in the data. The last layer contains one neuron with a linear activation function, which is used to produce continuous output. Backpropagation was performed using the ADAM optimizer the MSE loss function, and the model was trained for 50 epochs.

This step presents a flood hazard analysis for three regions: Jepara, Semarang, and Demak. The analysis consists of two primary components: an evaluation of the predictive model's performance, assessed using key classification metrics (Accuracy, Precision, Recall, and F1-Score) to ensure its reliability; and the calculation of Relative Feature Importance Weights, which ultimately serve as the final flood hazard weights for each location.

Table 5: Flood hazard weight

		Jepara	Semarang	Demak
Evaluation	Accuracy	0.9474	0.9333	0.8788
	Precision	0.9649	0.8800	0.8970
	Recall	0.9474	0.9333	0.8788
	F1-Score	0.9502	0.9037	0.8697
Relative Feature Importance Weight	Elevation	0.2278	0.2321	0.1679
	Runoff	0.3616	0.1998	0.3235
	Distance	0.2264	0.2775	0.3220
	Sea Level	0.1842	0.2906	0.1866

Based on the evaluation metrics (Table 5), the predictive models demonstrate results ranging from good to excellent across all three regions. The model for Jepara delivers the strongest outcome, achieving an exceptional F1-Score of 95.02% and an accuracy of 94.74%. This indicates an outstanding balance between precision and recall. Similarly, Semarang's model is also highly effective, yielding a strong F1-score of 90.37% and an accuracy of 93.33%. Meanwhile, the Demak model operates at a competent level, although slightly lower than the others, registering an F1-score of 86.97% and an accuracy of 87.88%. The high accuracy scores signify that the models are reliable in distinguishing between flood-prone and non-flood-prone areas. For instance, the 94.74% accuracy in Jepara means the model correctly classified approximately 95 out of every 100 conditions. It is important to note that while a predictive model was developed in this study, its primary output is the calculation of these feature weights, which will be utilized for hazard visualization and mapping.

Regarding the Determining Factors of Flood Hazard (Relative Feature Importance Weight), the analysis reveals that the primary drivers of flood risk vary by location, highlighting the unique characteristics of each region. In Jepara, flood hazard is primarily driven by Runoff (surface runoff), which holds a weight of 0.3616. For the Semarang region, the most significant variable is Sea Level (weight: 0.2906), closely followed by Distance from a river/sea at 0.2775. In Demak, the risk is most influenced by two nearly equally powerful factors: Runoff (0.3235) and Distance (0.3220). Overall, these results indicate that although all features (Elevation, Runoff, Distance, and Sea Level) are relevant, the mitigation priorities should differ for each regency. Jepara needs to focus on surface runoff management, while Semarang must prioritize mitigation related to sea-level rise, and Demak should address runoff and distance from water sources in a balanced manner.

The differences in weights reflect each location's unique environmental conditions and flood dynamics, which necessitate tailored approaches to flood risk analysis and management.

4.2 TES & NN combination to forecast panel data

To create the first model: forecasting with TES, starting with the processing of sea level and weather data. This data processing includes: converting dates from UTC to UTC+7, pre-processing each variable (converting sea level values from meters to centimeters, converting temperature from Kelvin to Celsius, etc.), grouping data by month for maximum values (sea level, V10, U10), average values (D2M, T2M), and minimum values (MSL, SP), then normalizing the data using the decimal scaling method. The forecasting results in three regions (Semarang, Demak, and Jepara) show different error rates for each variable. Table 6 presents the performance evaluation of the initial forecasting model, which was created as the first model in the study. The model's performance is measured using error metrics: MSE, RMSE, & MAPE, with lower error values indicating higher model accuracy. The results show the model's high accuracy (MAPE in 1.6%-2.2%) in predicting the weather variables: U10, V10, D2M, T2M, MSL, and SP. Conversely, the model faces significant challenges in predicting TP (MAPE≈70%), which has the highest error values, so this variable is excluded in the next process, also with SWH which actually has a fairly good error (MAPE≈18%) but not as high as other variables. For the Z variable (sea level) at three locations (Jepara, Semarang, and Demak), the prediction errors range from 22% to 25%, indicating moderate accuracy. Overall, this table indicates that the initial model is already very reliable for some meteorological variables but requires substantial improvements for other variables, especially rainfall and wave height. Accurate initial predictions for these variables are the reason they are included in the next stage. Z (sea level) has a moderate error rate, indicating that this model has fairly good accuracy in predicting sea level (Table 6, Figure 5), but it needs to consider the influence of weather variables to improve accuracy results.

Table 6: Initial Forecast, to create first model

Variables	MSE	RMSE	MAPE (%)
U10	0.0011	0.0327	1.67
V10	0.0008	0.0282	2.01
D2M	0.3895	0.6241	1.94
T2M	0.4683	0.6843	1.71
MSL	610.8699	24.7158	2.20
SP	626.0488	25.0210	2.23
TP	0.0062	0.0790	72.36
SWH	0.0253	0.1590	18.67
Z Jepara	174.7511	13.2193	22.87
Z Semarang	177.2125	13.3121	23.71
Z Demak	196.6765	14.0241	24.75

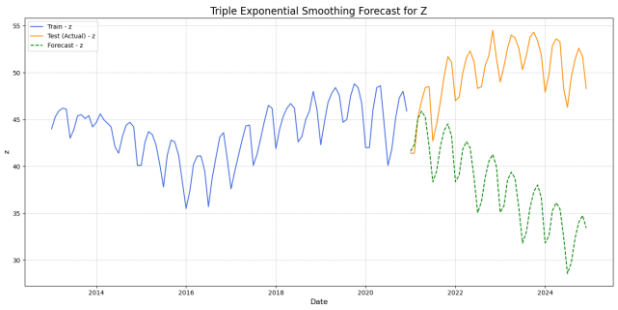

Figure 5: Initial forecast using TES

Training the NN model (for the second model, hybrid), by combining two methods: Triple Exponential Smoothing (TES) and Neural Network (NN), using sea level data influenced by several weather variables, starting with reading the data. Then, the data is divided into training data (1995-2020) and testing data (2021-2024). Next, two models are trained separately using the training data: (1) The TES model is trained on the time series data z to capture trend and seasonality patterns, (2) The NN model is trained to learn the relationship between selected feature variables (U10, V10, D2M, T2M, MSL, and SP) and the target variable z. These feature data are first scaled using MinMaxScaler. The NN model was created using MLPRegressor with six hidden layers containing 50 neurons.

The model performance evaluation (Table 7 & Figure 6) clearly shows that the Hybrid Method consistently outperforms the single method in all three regions. This superiority is evidenced by a drastic reduction in error values, with the MAPE error rate successfully reduced to less than half in all locations (e.g., from 20% to 4%). The primary reason behind this improvement in accuracy is that the hybrid method has higher accountability by incorporating the influence of weather variables into the modeling, a factor not considered by single methods. As a result, the hybrid approach produces forecasts that are not only more statistically accurate but also more reliable and comprehensive.

Table 7: Comparison between single method (TES) and hybrid method (TES & NN)

		Jepara	Semarang	Demak
Single Method	MSE	174.7511	177.2125	196.6765
	RMSE	13.2193	13.3121	14.0241
	MAPE	22.8745 %	23.7123 %	24.7514 %
Hybrid	MSE	25.1946	22.3257	22.6336
	RMSE	5.0194	4.7250	4.7575
	MAPE	8.6873%	8.3628%	8.3404%
Hybrid with Predicted Bias	MSE	4.8567	5.3047	5.0739
	RMSE	2.2038	2.3032	2.2525
	MAPE	3.5509%	3.7958%	3.6839%

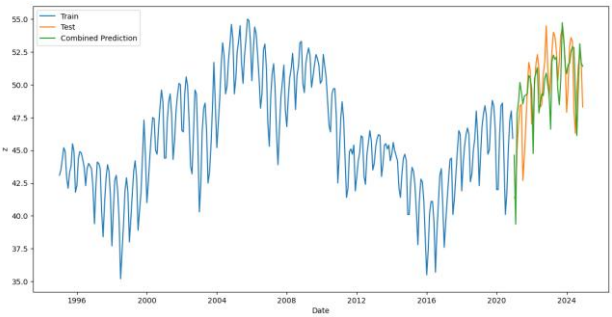

Figure 6: Forecast training & testing using hybrid method

4.3 Forecast with TES, NN & predicted bias

NN training was carried out to correct bias in time series data. The data was divided into training sets and test sets, with the features (X) and target (Y) used. The feature is then normalized using the MinMaxScaler into a range of 0 to 1. The NN model is made with three layers: an input layer with 64 neurons, a hidden layer with 32 neurons, and an output layer with 1 neuron. The model was compiled with the ADAM optimizer and MSE loss function, and then trained using training data that had been scaled over 50 epochs with a batch size of 32 and validation of 20% of the training data. The trained model is used to predict bias on the test set. The predicted average bias is calculated and adjusted by dividing by the number of features and factors.

Next, a combination of the TES and NN methods to make a combined prediction. The data was divided into training sets and test sets. The TES model is trained using training data and is used to make predictions on the test set. The NN model is trained with training data and is also used to make predictions on the test set. The predictions from both models were combined by taking

the mean from the predictions of the TES and the NN and then subtracting the bias. Actual data is combined with forecast data. MSE and MAPE are calculated to evaluate the performance of the model on the test set. In addition, a forecast for the 2024-2044 period was also made (Figure 7). In the context of this long-term flood hazard forecast, model deviations are of significant concern. To take this into account, this hybrid forecasting method allows for the integration of time series data and complex interactions between various influencing factors. This approach helps in adapting the model to better reflect changing conditions over time, thereby reducing the risk of model deviation.

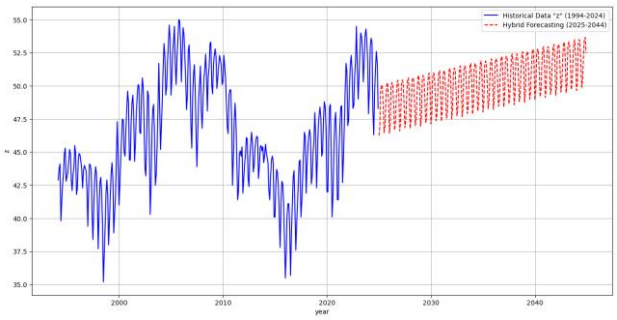

Figure 7: Sea Level Forecast 2025-2044

4.4 Visualization (Flood Hazard)

The visualization is provided in the form of a raster map created by calculating the flood hazard levels. Starting with creating a raster for elevation, several GeoTIFF files from DEMNAS, which form each region (Jepara/Semarang/Demak), are merged and then clipped according to the regional boundaries (shapefile of Indonesia's topographical boundary). The coordinate reference system (CRS) was changed to EPSG:4326, and the spatial resolution was converted to 8.35 meters. Then, for visualization purposes on the flood hazard map in the form of scores, the elevation was transformed using min-max normalization, and an inversion was performed (the higher the elevation, the lower the score).

Determination of CRS, clipping, and resampling to achieve the exact resolution is also performed on the landcover raster. The landcover values are converted to runoff based on the Rational Method Runoff Coefficient. This runoff value is considered as the score.

To create a raster distance to the coastline, the shapefile of the Indonesian Landform Boundary was used, and then the Euclidean distance transformation was calculated to determine the distance from each point of each pixel in the raster to the nearest coastline (which has been defined previously). This distance is then converted to meters. For score determination, min-max normalization is performed, and since this distance has a negative correlation with flood risk, inversion is carried out.

To create the sea level raster, each pixel is calculated: Sea level = Sea level Score × Distance Score, which means using the distance raster and linking it with the sea level value from the forecast (or actual value) at a specific time (month and year) determined as a parameter.

Finally, the flood hazard map is created by calculating the pixel values on all rasters associated with the predetermined weight values (different in each region).

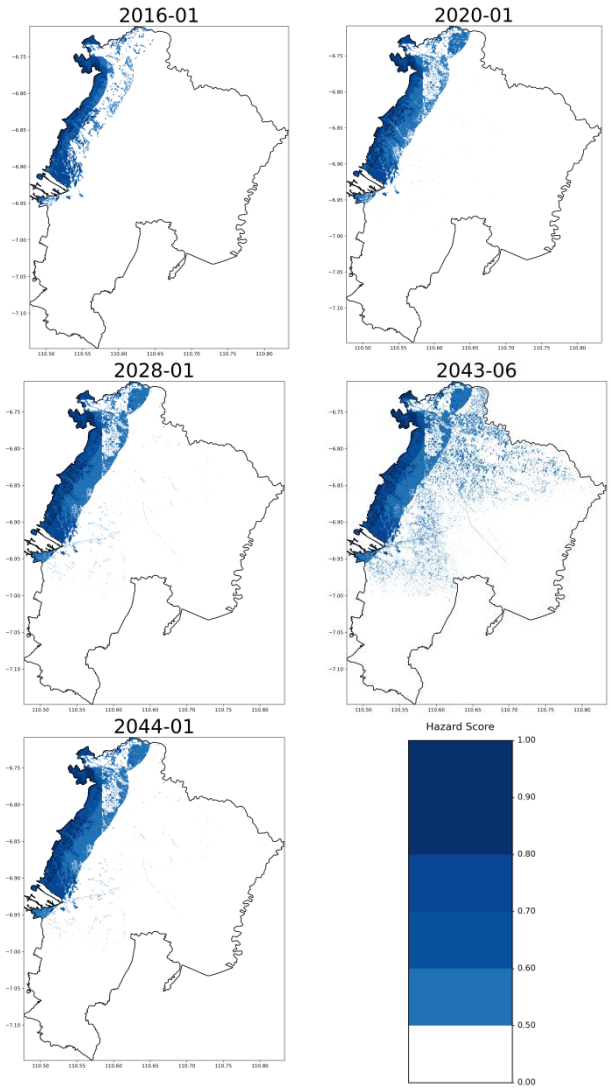

Figure 8: Flood Hazard: Demak

The projected future growth in flood-affected areas and increased hazard in the Demak area are depicted in Figure 8. Greater danger is indicated by darker blue hues, and the extent of the at-risk areas is depicted by larger blue regions. The northwestern coastal region was the primary location for flood hazards in January 2016 and January 2020. From 2016 to 2020, the number of flood-prone areas increased slightly, with areas at risk of flooding (light blue) seemingly moving a little farther inland. This means that the risk has already shown an upward trend during those four years. Predictions for January 2028 indicate a marked expansion of the at-hazard area in comparison to 2020. The light-blue patches extend farther into the interior, suggesting that areas that were previously low-hazard or not at risk are now starting to face the threat of flooding. In coastal

areas (dark blue), the level of hazard intensity is still high and has somewhat increased. The most dire scenario is projected for June 2043 (Long-Term Prediction), when areas at risk of flooding will have greatly expanded. Flood risk is expected to be low to high in nearly half of the northern portion of the region. According to this, the impact of climate change or other factors (like land subsidence and changes in land use) is expected to significantly worsen the flood situation within a 20-year period starting in 2028.

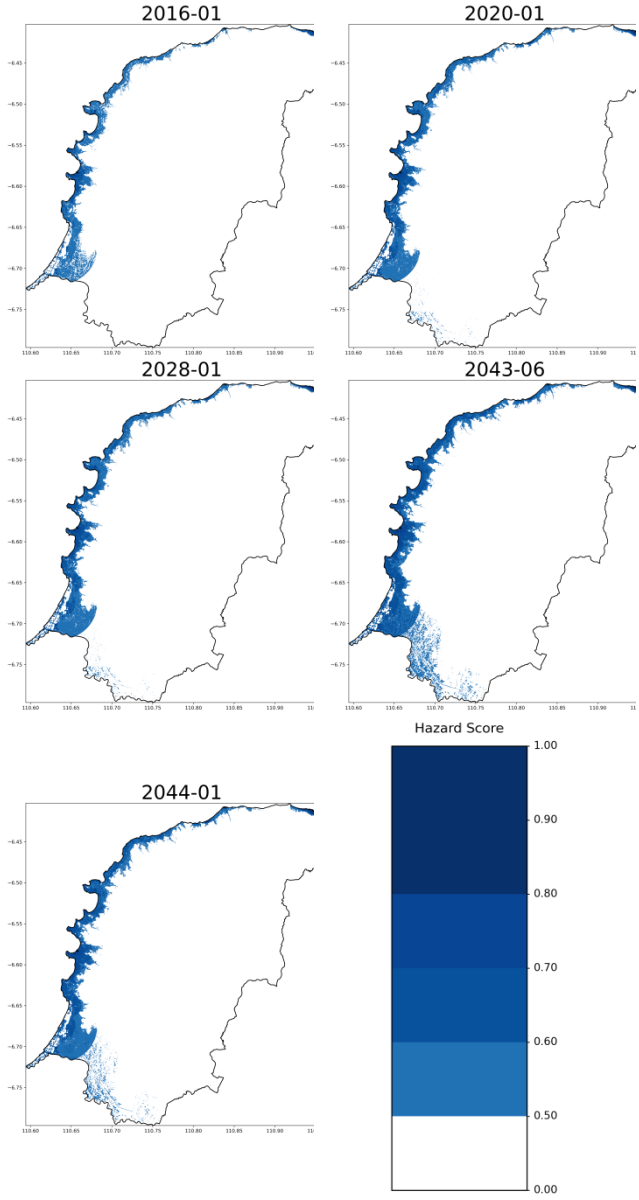

Figure 9: Flood Hazard: Jepara

The Jepara area, which has a long coastline, is predicted to have tidal flood or coastal inundation hazards, as shown in Figure 9. This set particularly emphasizes the dangers along the coast, in contrast to the Demak area, which depicts the spread of floods inland. Flood hazards were concentrated in the bay areas, river estuaries, and lowlands along the coastline during both of the 2016–2020 timeframes. The nearly identical hazard pattern during this time suggests that the risk of coastal inundation is comparatively constant. The most dangerous areas (dark blue) are found in the coastal basins that are most at risk. Hazardous area extent is predicted to slightly increase in 2028. The inundation areas (light blue) have slightly expanded at a number of locations along the coast, though not significantly, suggesting that the sea level may be rising or that high tides may be occurring more frequently. This long-term forecast shows a more pronounced increase in danger than in prior years. Particularly in the southern portion and in a number of coastal areas in the north, the inundation areas (light blue) seem to be more widespread. This suggests that low-lying coastal areas will experience more widespread and long-lasting flooding by 2043 as a result of the anticipated rise in sea levels. Wider low-risk zones now encircle the highest-risk areas (dark blue), which stay in the same place.

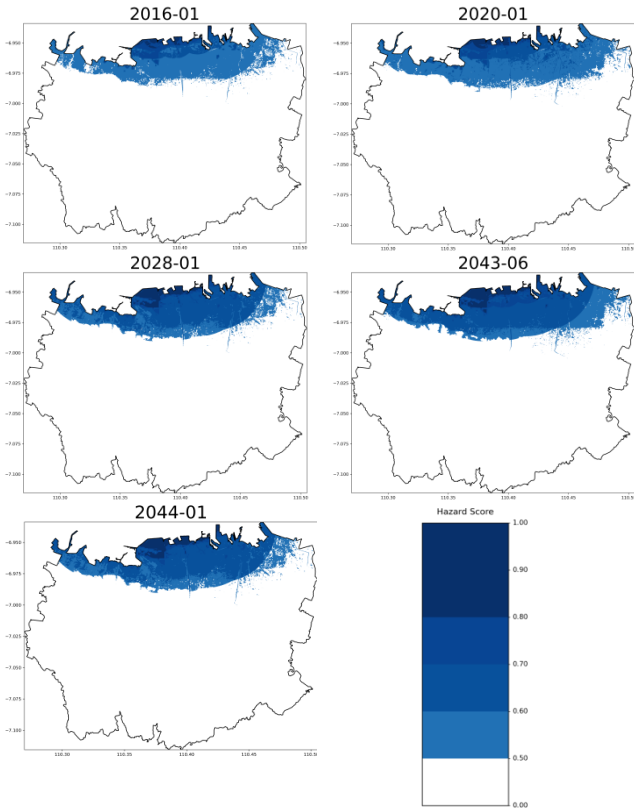

Figure 10: Flood Hazard: Semarang

The city of Semarang is seriously threatened by coastal flooding (rob), which is expected to get worse and spread more widely in the future, according to Figure 10. The majority of Semarang's northern coastal regions already had a high risk of flooding in 2016. While the light blue (lower risk) areas have already spread quite far southward (inland), the dark blue (highest risk) areas are concentrated close to the coast. Comparing 2020 to 2016, there was a minor decline. Flooded areas appear slightly larger and have spread farther inland, particularly those at lower risk (light blue). The impacted area is expected to significantly expand by 2028. A larger area to the

south is expected to be inundated by flooding than in 2020. Deeper or more frequent flooding is indicated by the dark blue areas' apparent slight increase in intensity. The Long-Term Prediction for 2043 is a worrying situation. The map displays a significant and striking increase in tidal flooding. It is anticipated that the inundation area will "consume" a significant portion of North Semarang and extend deep into the city center. Nearly the whole coastal region has turned into a high-risk zone (dark blue), meaning that flooding is likely to occur there either permanently or very frequently.

5 Discussion section

This study examines how rising sea levels made worse by climate change are increasing the risk of coastal floods in Indonesia's coastal regions, especially in Semarang, Demak, and Jepara, using spatial data mining forecasting techniques. Compared to conventional approaches, the hybrid model combining TES and NN significantly improves flood prediction accuracy. The hybrid model achieved a MAPE value of 8.3% to 3.5%, indicating that it is better at predicting floods more accurately compared to single methods [12]. Incorporating sea level data, elevation, runoff, distance to the coastline, and other parameters into the model allows for a comprehensive analysis of flood hazards. By using historical flood data to measure the weight of these parameters through NN, the assessment becomes more objective and subjectivity in the Analytic Hierarchy Process (AHP) is reduced [8]. For areas like Indonesia, where geographical and climatic variances can greatly affect flood dynamics, this methodological development is essential. This study also highlights how crucial it is to analyze very large datasets using spatial data mining to spot trends and raise the precision of predictive models. In addition to identifying non-linear correlations, the hybrid approach is expected to overcome the weaknesses of traditional forecasting techniques, which often ignore the influences between variables [6]. The results indicate that the risk of flooding will increase in hundreds of densely populated cities by 2050, potentially doubling by 2100 [1].

The trade-off between NN, exponential smoothing, and hybrid approaches (Table 10) involves many considerations. NN provide scalability and the ability to model complex relationships at the expense of computational efficiency and interpretability. Exponential Smoothing offers a more straightforward and interpretable approach that is computationally efficient but may struggle with scalability and complexity. Hybrid Approaches aim to combine the strengths of both methods, providing a balanced solution that improves scalability, maintains a degree of interpretability, and improves accuracy.

Table 10: Trade-offs between all the methods used

Aspect	NN	ES	Hybrid
Computational Efficiency	High computational demand; longer training times	Low computational demand; quick execution	High computational demand; resource-intensive
Scalability	Highly scalable; adapts to large datasets	Limited scalability; struggles with complexity	Highly scalable; flexible adaptation
Interpretability	Low interpretability (black box nature)	High interpretability; straightforward mechanics	Moderate interpretability; combines strengths
Model Accuracy	High accuracy potential; risk of overfitting	Good accuracy for simple patterns; limited in complexity	Enhanced accuracy; robust to variability

Based on the weight analysis derived from Feature importance (Table 5), it can be concluded that the main triggers of coastal flooding vary significantly in each location, indicating unique local vulnerabilities [116]. High water runoff from the land poses the biggest risk in Jepara. Semarang, on the other hand, is more susceptible to direct sea influences, especially those related to sea level and land subsidence (elevation). In the meantime, water runoff and Demak's closeness to the coast pose a nearly equal threat. This key distinction highlights the fact that mitigation strategies must be tailored to each region and cannot be applied universally.

6 Conclusion

The hybrid method applied has successfully addressed the panel data problem with relatively better MAPE values, ranging from 8.3% to 3.5%. Additionally, the flood hazard weights have also been well identified, with accuracy values varying between 88% and 94%, regional differences can affect results due to the different geographical and climatic characteristics in each region, which can affect flood dynamics.

The findings of this study contribute to the increasing knowledge of flood risk assessment and management, providing a solid framework for future studies and practical applications in flood-prone areas.

The success of the hybrid model in predicting coastal floods can serve as a valuable tool for policymakers and urban planners in developing effective flood management strategies. It is hoped that decision-makers can benefit from appropriate flood mitigation techniques for coastal areas, as infrastructure and lives are seriously threatened by this condition.

Some points that can be concluded:

1. Integrating a hybrid approach with climate model projections is recommended in climate change uncertainty with significant variations due to differences in physics assumptions, parameterization, and emission scenarios by capturing a more comprehensive range of possible future conditions.
2. Predictions become more resilient to biases or model-specific errors by using multiple projections at once, resulting in more reliable estimates of flood risk or related variables under climate change scenarios.
3. Probabilistic information from hybrids supporting risk-based decision-making can be used to design flexible adaptation strategies for future possibilities.

This research opens up essential insights and broader implications for coastal policy and risk management:

1. Scalability of the System
   Hybrid models can be widely applied in various coastal areas with different characteristics as long as adequate data is available.
2. Real-World Application
   The model must be followed by real implementation, including integration with early warning systems and regional spatial planning.
3. Implementation Challenges
   Field implementation may face data limitations, human resources, and technological infrastructure. Therefore, adequate policy support and investment must be needed to overcome these barriers.

7 Acknowledgements

This study received support from Doctoral Program of Information System, School of Postgraduate Studies, Universitas Diponegoro.

References

[1] UNDP, "Climate change's impact on coastal flooding to increase 5-times over this century, putting over 70 million people in the path of expanding floodplains, according to new UNDP and CIL data." Accessed: Nov. 25, 2024. [Online]. Available: https://www.undp.org/press-releases/climate-changes-impact-coastal-flooding-increase-5-times-over-century-putting-over-70-million-people-path-expanding-floodplains

[2] V. Kumar, K. Sharma, T. Caloiero, D. Mehta, and K. Singh, "Comprehensive Overview of Flood Modeling Approaches: A Review of Recent Advances," *Hydrology*, vol. 10, no. 7, p. 141, Jun. 2023, doi: 10.3390/hydrology10070141.

[3] F. Aureli, A. Maranzoni, and G. Petaccia, "Advances in Dam-Break Modeling for Flood Hazard Mitigation: Theory, Numerical Models, and Applications in Hydraulic Engineering," *Water*, vol. 16, no. 8, p. 1093, Apr. 2024, doi: 10.3390/w16081093.

[4] P. Sibandze, A. M. Kalumba, A. H. Aljaddani, L. Zhou, and G. A. Afuye, "Geospatial Mapping and Meteorological Flood Risk Assessment: A Global Research Trend Analysis," *Environ. Manage.*, Oct. 2024, doi: 10.1007/s00267-024-02059-0.

[5] FEMA, "Guidance for Flood Risk Analysis and Mapping." [Online]. Available: https://www.fema.gov/sites/default/files/documents/fema_flood-risk-assessment-guidance.pdf

[6] 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.

[7] M. Lang *et al.*, "Use of Systematic, Palaeoflood and Historical Data for the Improvement of Flood Risk Estimation. Review of Scientific Methods," *Nat. Hazards*, vol. 31, no. 3, pp. 623–643, Mar. 2004, doi: 10.1023/B:NHAZ.0000024895.48463.eb.

[8] K. Yao *et al.*, "A Novel Flood Risk Analysis Framework Based on Earth Observation Data to Retrieve Historical Inundations and Future Scenarios," *Remote Sens.*, vol. 16, no. 8, p. 1413, Apr. 2024, doi: 10.3390/rs16081413.

[9] Y. Chen, D. Wang, L. Zhang, H. Guo, J. Ma, and W. Gao, "Flood risk assessment of Wuhan, China, using a multi-criteria analysis model with the improved AHP-Entropy method," *Environ. Sci. Pollut. Res.*, vol. 30, no. 42, pp. 96001–96018, Aug. 2023, doi: 10.1007/s11356-023-29066-8.

[10] S. R. Singh, E. Harirchian, C. E. F. Monjardin, and T. Lahmer, "GIS-Based Risk Assessment of Building Vulnerability in Flood Zones of Naic, Cavite, Philippines Using AHP and TOPSIS," *GeoHazards*, vol. 5, no. 4, pp. 1040–1073, Oct. 2024, doi: 10.3390/geohazards5040050.

[11] R. Oyelakin, W. Yang, and P. Krebs, "Analysing Urban Flooding Risk with CMIP5 and CMIP6 Climate Projections," *Water*, vol. 16, no. 3, p. 474, Jan. 2024, doi: 10.3390/w16030474.

[12] Z. Yin *et al.*, "Assessing the economic impacts of future fluvial flooding in six countries under climate change and socio-economic development," *Clim. Change*, vol. 166, no. 3–4, 2021, doi: 10.1007/s10584-021-03059-3.

[13] 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.

[14] R. Lan, "Spatial Data Mining and Knowledge Discovery," in *Advances in Cartography and Geographic Information Engineering*, Singapore: Springer Singapore, 2021, pp. 503–530. doi: 10.1007/978-981-16-0614-4_14.

[15] J. Henson, "Univariate vs Multivariate Time Series Forecasting." Accessed: Oct. 18, 2023. [Online]. Available: https://medium.com/@jesse.henson/univariate-vs-multivariate-time-series-forecasting-cfcc4150e20a

[16] J. Brownlee, "Taxonomy of Time Series Forecasting Problems," Machine Learning Mastery. Accessed: Oct. 18, 2023. [Online]. Available: https://machinelearningmastery.com/taxonomy-of-time-series-forecasting-problems/

[17] F. Petropoulos and E. Spiliotis, "The Wisdom of the Data: Getting the Most Out of Univariate Time Series Forecasting," *Forecasting*, vol. 3, no. 3, pp. 478–497, Jun. 2021, doi: 10.3390/forecast3030029.

[18] N. Rankovic, D. Rankovic, M. Ivanovic, and L. Lazic, "Improved effort and cost estimation model using artificial neural networks and taguchi method with different activation functions," *Entropy*, vol. 23, no. 7, 2021, doi: 10.3390/e23070854.

[19] A. Sahoo, S. Samantaray, and D. K. K. Ghose, "Prediction of Flood in Barak River using Hybrid Machine Learning Approaches: A Case Study," *J. Geol. Soc. India*, vol. 97, no. 2, pp. 186–198, 2021, doi: 10.1007/s12594-021-1650-1.

[20] A. H. Zuhairi, F. Yakub, S. A. Zaki, and M. S. M. Ali, "Review of flood prediction hybrid machine learning models using datasets," 2022. doi: 10.1088/1755-1315/1091/1/012040.

[21] 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.

[22] T. Canchala, W. Alfonso-Morales, Y. Carvajal-Escobar, W. L. Cerón, and E. Caicedo-Bravo, "Monthly Rainfall Anomalies Forecasting for Southwestern Colombia Using Artificial Neural Networks Approaches," *Water*, vol. 12, no. 9, p. 2628, Sep. 2020, doi: 10.3390/w12092628.

[23] K. Xu, Z. Han, L. Bin, R. Shen, and Y. Long, "Rapid forecasting of compound flooding for a coastal area based on data-driven approach," *Nat. Hazards*, vol. 121, no. 2, pp. 1399–1421, Jan. 2025, doi: 10.1007/s11069-024-06846-0.

[24] B. Victor *et al.*, "Off to new Shores: A Dataset & Benchmark for (near-)coastal Flood Inundation Forecasting," 2024. [Online]. Available: https://arxiv.org/abs/2409.18591

[25] J. Chou, Y. Cao, and P. Minford, "Evaluation and indirect inference estimation of inattentive features in a New Keynesian framework," *J. Forecast.*, vol. 42, no. 3, pp. 530–542, Apr. 2023, doi: 10.1002/for.2950.

[26] X.-P. Li *et al.*, "The Study of Biological Glue Droplet Impact Behavior of Bioceramic Powders Applied in 3D Printing of Bone Scaffolds," *Appl. Sci.*, vol. 12, no. 4, p. 1898, Feb. 2022, doi: 10.3390/app12041898.

[27] M. Habulan, B. Đurin, A. P. Siročić, and N. Sakač, "Analysis of the Particulate Matter Pollution in the Urban Areas of Croatia, EU," in *The 3rd International Electronic Conference on Atmospheric Sciences*, Basel Switzerland: MDPI, Nov. 2020, p. 9. doi: 10.3390/ecas2020-08145.

[28] M.-J. Park and H.-S. Yang, "Comparative Study of Time Series Analysis Algorithms Suitable for Short-Term Forecasting in Implementing Demand Response Based on AMI," *Sensors*, vol. 24, no. 22, p. 7205, Nov. 2024, doi: 10.3390/s24227205.

[29] R. V. Vargas and P. M. P. IPMA-B, "Using the analytic hierarchy process (AHP) to select and prioritize projects in a portfolio," in *PMI global congress*, PA: Project Management Institute Washington, DC, 2010, pp. 1–22.

[30] M. Tavana, M. Soltanifar, and F. J. Santos-Arteaga, "Analytical hierarchy process: revolution and evolution," *Ann. Oper. Res.*, vol. 326, no. 2, pp. 879–907, Jul. 2023, doi: 10.1007/s10479-021-04432-2.

[31] P. Madzík and L. Falát, "State-of-the-art on analytic hierarchy process in the last 40 years: Literature review based on Latent Dirichlet Allocation topic modelling," *PLoS One*, vol. 17, no. 5, p. e0268777, May 2022, doi: 10.1371/journal.pone.0268777.

[32] S. Mourato, P. Fernandez, L. G. Pereira, and M. Moreira, "Assessing Vulnerability in Flood Prone Areas Using Analytic Hierarchy Process—Group Decision Making and Geographic Information System: A Case Study in Portugal," *Appl. Sci.*, vol. 13, no. 8, p. 4915, Apr. 2023, doi: 10.3390/app13084915.

[33] K. C. Swain, C. Singha, and L. Nayak, "Flood Susceptibility Mapping through the GIS-AHP Technique Using the Cloud," *ISPRS Int. J. Geo-Information*, vol. 9, no. 12, p. 720, Dec. 2020, doi: 10.3390/ijgi9120720.

[34] BPS, "Provider of Quality Statistical Data For Advanced Indonesia." Accessed: Dec. 05, 2024. [Online]. Available: https://www.bps.go.id/id

[35] E. Hermawan *et al.*, "Large-Scale Meteorological Drivers of the Extreme Precipitation Event and Devastating Floods of Early-February 2021 in Semarang, Central Java, Indonesia," *Atmosphere (Basel).*, vol. 13, no. 7, 2022, doi: 10.3390/atmos13071092.

[36] A. Chrysanti, A. Adhani, I. N. Azkiarizqi, M. B. Adityawan, M. S. B. Kusuma, and M. Cahyono, "Assessing Compound Coastal–Fluvial Flood

Impacts and Resilience Under Extreme Scenarios in Demak, Indonesia," *Sustainability*, vol. 16, no. 23, p. 10315, Nov. 2024, doi: 10.3390/su162310315.

[37] N. A. Akhirianto, K. S. Wardani, A. Ma'rufatin, A. Nurwijayanti, F. M. G. Putra, and F. Chusnayah, "Sustainable regional development planning based on coastal disaster risk assessment (Case study: North coast of Central Java)," *IOP Conf. Ser. Earth Environ. Sci.*, vol. 1479, no. 1, p. 012040, Apr. 2025, doi: 10.1088/1755-1315/1479/1/012040.

[38] M. Imani *et al.*, "Risk Assessment of Coastal Flooding under Different Inundation Situations in Southwest of Taiwan (Tainan City)," *Water*, vol. 13, no. 6, p. 880, Mar. 2021, doi: 10.3390/w13060880.

[39] M. Tu'uholoaki, A. Espejo, M. Wandres, A. Singh, H. Damlamian, and Z. Begg, "Quantifying Mechanisms Responsible for Extreme Coastal Water Levels and Flooding during Severe Tropical Cyclone Harold in Tonga, Southwest Pacific," *J. Mar. Sci. Eng.*, vol. 11, no. 6, p. 1217, Jun. 2023, doi: 10.3390/jmse11061217.

[40] J.-H. Yuk, J.-S. Kang, and H. Myung, "Applicability Study of a Global Numerical Weather Prediction Model MPAS to Storm Surges and Waves in the South Coast of Korea," *Atmosphere (Basel).*, vol. 13, no. 4, 2022, doi: 10.3390/atmos13040591.

[41] W. Costa, D. Idier, J. Rohmer, M. Menendez, and P. Camus, "Statistical Prediction of Extreme Storm Surges Based on a Fully Supervised Weather-Type Downscaling Model," *J. Mar. Sci. Eng.*, vol. 8, no. 12, p. 1028, Dec. 2020, doi: 10.3390/jmse8121028.

[42] L. Fan and L. Du, "Combined effects of climatic factors on extreme sea level changes in the Northwest Pacific Ocean," *Ocean Dyn.*, vol. 73, no. 3–4, pp. 181–199, 2023, doi: 10.1007/s10236-023-01543-1.

[43] 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.

[44] 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/

[45] SRGI, "Tide Prediction: Indonesian Geospatial Reference System." Accessed: Nov. 01, 2024. [Online]. Available: https://srgi.big.go.id/map/pasut-prediction-active

[46] J.-N. Hersbach, H., Bell, B., Berrisford, P., Biavati, G., Horányi, A., Muñoz Sabater, J., Nicolas, J., Peubey, C., Radu, R., Rozum, I., Schepers, D., Simmons, A., Soci, C., Dee, D., Thépaut, "ERA5 hourly data on single levels from 1940 to present," 2023. doi: 10.24381/cds.adbb2d47.

[47] D. F. Hill, "Spatial and Temporal Variability in Tidal Range: Evidence, Causes, and Effects," *Curr. Clim. Chang. Reports*, vol. 2, no. 4, pp. 232–241, Dec. 2016, doi: 10.1007/s40641-016-0044-8.

[48] A. T. Devlin, D. A. Jay, E. D. Zaron, S. A. Talke, J. Pan, and H. Lin, "Tidal Variability Related to Sea Level Variability in the Pacific Ocean," *J. Geophys. Res. Ocean.*, vol. 122, no. 11, pp. 8445–8463, Nov. 2017, doi: 10.1002/2017JC013165.

[49] J. Challis, D. Idier, G. Wöppelmann, and G. André, "Atmospheric Wind and Pressure-Driven Changes in Tidal Characteristics over the Northwestern European Shelf," *J. Mar. Sci. Eng.*, vol. 11, no. 9, p. 1701, Aug. 2023, doi: 10.3390/jmse11091701.

[50] Copernicus Climate Change Service, "ERA5: data documentation," 2025. [Online]. Available: https://confluence.ecmwf.int/display/CKB/ERA5%3A+data+documentation

[51] Copernicus Climate Change Service, "ERA5: What is the spatial reference," 2025. [Online]. Available: https://confluence.ecmwf.int/display/CKB/ERA5%3A+What+is+the+spatial+reference

[52] K. Park and E. H. Lee, "Urban flood vulnerability analysis and prediction based on the land use using Deep Neural Network," *Int. J. Disaster Risk Reduct.*, vol. 101, p. 104231, Feb. 2024, doi: 10.1016/j.ijdrr.2023.104231.

[53] A. Nirwansyah, *Geodata dan Pemanfaatannya dalam Isu Kebencanaan*. 2024.

[54] M. Lorie *et al.*, "Modeling coastal flood risk and adaptation response under future climate conditions," *Clim. Risk Manag.*, vol. 29, p. 100233, 2020, doi: https://doi.org/10.1016/j.crm.2020.100233.

[55] H. Moftakhari, J. E. Schubert, A. AghaKouchak, R. A. Matthew, and B. F. Sanders, "Linking statistical and hydrodynamic modeling for compound flood hazard assessment in tidal channels and estuaries," *Adv. Water Resour.*, vol. 128, pp. 28–38, Jun. 2019, doi: 10.1016/j.advwatres.2019.04.009.

[56] J. A. Pollard, T. Spencer, and S. Jude, "Big Data Approaches for coastal flood risk assessment and emergency response," *WIREs Clim. Chang.*, vol. 9, no. 5, pp. 1–14, Sep. 2018, doi: 10.1002/wcc.543.

[57] M. U. Parodi, A. Giardino, A. van Dongeren, S. G. Pearson, J. D. Bricker, and A. J. H. M. Reniers, "Uncertainties in coastal flood risk assessments in small island developing states," *Nat. Hazards Earth Syst. Sci.*, vol. 20, no. 9, pp. 2397–2414, Sep. 2020, doi: 10.5194/nhess-20-2397-2020.

[58] D. L. Anderson *et al.*, "Projecting Climate Dependent Coastal Flood Risk With a Hybrid Statistical Dynamical Model," *Earth's Futur.*, vol. 9, no. 12, 2021, doi: 10.1029/2021EF002285.

[59] A.-L. Balogun and N. Adebisi, "Sea level prediction using ARIMA, SVR and LSTM neural network: assessing the impact of ensemble Ocean-Atmospheric processes on models' accuracy," *Geomatics, Nat. Hazards Risk*, vol. 12, no. 1, pp. 653–674, Jan. 2021, doi: 10.1080/19475705.2021.1887372.

[60] Y. N. Hilal, G. D. A. Nainggolan, S. H. Syahputri, and F. Kartiasih, "Comparison of ARIM and LSTM Methods in Predicting Jakarta Sea Level," *J. Ilmu dan Teknol. Kelaut. Trop.*, vol. 16, no. 2, pp. 163–178, Oct. 2024, doi: 10.29244/jitkt.v16i2.52818.

[61] M. F. Marsani, V. Someetheram, M. S. Mohd Kasihmuddin, S. Z. Mohd Jamaludin, M. A. Mansor, and B. Badyalina, "The application of seasonal autoregressive integrated moving average (SARIMA) model in forecasting Malaysia mean sea level," 2024, p. 070005. doi: 10.1063/5.0223836.

[62] E. Ayitey, F. Ayiah-Mensah, S. Nunoo, and J. A. Addor, "Applying machine learning techniques for sea level rise forecasting in Axim: tackling missing data and outliers," *Model. Earth Syst. Environ.*, vol. 11, no. 3, p. 164, Jun. 2025, doi: 10.1007/s40808-025-02314-1.

[63] X. Long *et al.*, "Evaluating Current Statistical and Dynamical Forecasting Techniques for Seasonal Coastal Sea Level Prediction," *J. Clim.*, vol. 38, no. 6, pp. 1477–1503, Mar. 2025, doi: 10.1175/JCLI-D-24-0214.1.

[64] L. Elneel, M. S. Zitouni, H. Mukhtar, and H. Al-Ahmad, "Examining sea levels forecasting using autoregressive and prophet models," *Sci. Rep.*, vol. 14, no. 1, p. 14337, Jun. 2024, doi: 10.1038/s41598-024-65184-0.

[65] P. K. Srivastava, T. Islam, S. K. Singh, G. P. Petropoulos, M. Gupta, and Q. Dai, "Forecasting Arabian Sea level rise using exponential smoothing state space models and ARIMA from TOPEX and Jason satellite radar altimeter data," *Meteorol. Appl.*, vol. 23, no. 4, pp. 633–639, Oct. 2016, doi: 10.1002/met.1585.

[66] M. Jahanbani, M. H. Vahidnia, H. Aghamohammadi, and Z. Azizi, "Flood susceptibility mapping through geoinformatics and ensemble learning methods, with an emphasis on the AdaBoost-Decision Tree algorithm, in Mazandaran, Iran," *Earth Sci. Informatics*, vol. 17, no. 2, pp. 1433–1457, Apr. 2024, doi: 10.1007/s12145-023-01213-2.

[67] G. D. Kumar, S. Tyagi, and K. C. Pradhan, "Predictive ML analysis: Rainfall & flood vulnerability in Bihar, India," in *Artificial Intelligence and Information Technologies*, London: CRC Press, 2024, pp. 447–453. doi: 10.1201/9781032700502-71.

[68] G. D. Kumar, K. C. Pradhan, and S. Tyagi, "Deep Learning Forecasting: An LSTM Neural Architecture based Approach to Rainfall and Flood Impact Predictions in Bihar," *Procedia Comput. Sci.*, vol. 235, pp. 1455–1466, 2024, doi: 10.1016/j.procs.2024.04.137.

[69] D. K. Hakim, R. Gernowo, A. W. Nirwansyah, and T. Haryanto, "Data Mining Approach for River Flood Hazard Time-Series: Using a Combination of Triple Exponential Smoothing and Neural Networks, in Demak," *Informatica*, vol. 49, no. 19, Apr. 2025, doi: 10.31449/inf.v49i19.7288.

[70] BNPB, *IRBI 2023*. 2023. [Online]. Available: https://inarisk.bnpb.go.id/IRBI-2023/mobile/index.html

[71] C. Buta, G. Mihai, and M. Stănescu, "Flood Risk Assessment Based on Flood Hazard and Vulnerability Indexes," *Ovidius Univ. Ann. Constanta - Ser. Civ. Eng.*, vol. 22, no. 1, pp. 127–137, Dec. 2020, doi: 10.2478/ouacsce-2020-0014.

[72] FEMA, "Guidance for Flood Risk Analysis and Mapping. Hydrology: Rainfall-Runoff Analysis," 2019. [Online]. Available: https://www.fema.gov/sites/default/files/2020-02/Hydrologic_Rainfall_Runoff_Analysis_Feb_2019.pdf

[73] T. Tingsanchali and T. Promping, "Comprehensive Assessment of Flood Hazard, Vulnerability, and Flood Risk at the Household Level in a Municipality Area: A Case Study of Nan Province, Thailand," *Water*, vol. 14, no. 2, p. 161, Jan. 2022, doi: 10.3390/w14020161.

[74] UNISDR, "Flood Hazard and Risk Assessment," 2017. [Online]. Available: https://www.unisdr.org/files/52828_04floodhazardandriskassessment.pdf

[75] Q. Wang, H. Gu, X. Zang, M. Zuo, and H. Li, "Flood resilience in cities and urban agglomerations: a systematic review of hazard causes, assessment frameworks, and recovery strategies based on LLM tools," *Nat. Hazards*, May 2025, doi: 10.1007/s11069-025-07285-1.

[76] S. Zhang, Z. Yang, and C. Wang, "Environmental risk assessment of earthen heritage sites in arid regions: a case study of the Great Wall in Xinjiang," *npj Herit. Sci.*, vol. 13, no. 1, p. 200, May 2025, doi: 10.1038/s40494-025-01736-5.

[77] C. R. de Sá, "Variance-Based Feature Importance in Neural Networks," 2019, pp. 306–315. doi: 10.1007/978-3-030-33778-0_24.

[78] M. Saarela and S. Jauhiainen, "Comparison of feature importance measures as explanations for classification models," *SN Appl. Sci.*, vol. 3, no. 2, p. 272, Feb. 2021, doi: 10.1007/s42452-021-04148-9.

[79] K. Xu, J. Fang, Y. Fang, Q. Sun, C. Wu, and M. Liu, "The Importance of Digital Elevation Model

Selection in Flood Simulation and a Proposed Method to Reduce DEM Errors: A Case Study in Shanghai," *Int. J. Disaster Risk Sci.*, vol. 12, no. 6, pp. 890–902, 2021, doi: 10.1007/s13753-021-00377-z.

[80] G. Schumann, L. Giustarini, A. Tarpanelli, B. Jarihani, and S. Martinis, "Flood Modeling and Prediction Using Earth Observation Data," *Surv. Geophys.*, vol. 44, no. 5, pp. 1553–1578, Oct. 2023, doi: 10.1007/s10712-022-09751-y.

[81] D. B. Gesch, "Best Practices for Elevation-Based Assessments of Sea-Level Rise and Coastal Flooding Exposure," *Front. Earth Sci.*, vol. 6, Dec. 2018, doi: 10.3389/feart.2018.00230.

[82] H. Liang and X. Zhou, "Impact of Tides and Surges on Fluvial Floods in Coastal Regions," *Remote Sens.*, vol. 14, no. 22, 2022, doi: 10.3390/rs14225779.

[83] H. Zhou, *Learn Data Mining Through Excel: A Step-by-Step Approach for Understanding Machine Learning Methods*. Berkeley, CA: Apress, 2020. doi: 10.1007/978-1-4842-5982-5.

[84] R. Abdulkadirov, P. Lyakhov, and N. Nagornov, "Survey of Optimization Algorithms in Modern Neural Networks," *Mathematics*, vol. 11, no. 11, p. 2466, May 2023, doi: 10.3390/math11112466.

[85] S. Prasad, "Time Series," in *Advanced Statistical Methods*, Singapore: Springer Nature Singapore, 2024, pp. 109–162. doi: 10.1007/978-981-99-7257-9_3.

[86] C. Fieberg, D. Metko, T. Poddig, and T. Loy, "Machine learning techniques for cross-sectional equity returns' prediction," *OR Spectr.*, vol. 45, no. 1, pp. 289–323, Mar. 2023, doi: 10.1007/s00291-022-00693-w.

[87] F. Duval, J.-P. Boucher, and M. Pigeon, "Telematics combined actuarial neural networks for cross-sectional and longitudinal claim count data," *ASTIN Bull.*, vol. 54, no. 2, pp. 239–262, May 2024, doi: 10.1017/asb.2024.4.

[88] N. S. Grigg, "Comprehensive Flood Risk Assessment: State of the Practice," *Hydrology*, vol. 10, no. 2, p. 46, Feb. 2023, doi: 10.3390/hydrology10020046.

[89] A. Díez-Herrero and J. Garrote, "Flood Risk Assessments: Applications and Uncertainties," *Water*, vol. 12, no. 8, p. 2096, Jul. 2020, doi: 10.3390/w12082096.

[90] A. Buckle, "Why is it Called UTC – not CUT," timeanddate.com. Accessed: Jan. 02, 2025. [Online]. Available: https://www.timeanddate.com/time/utc-abbreviation.html

[91] A. Babii, R. T. Ball, E. Ghysels, and J. Striaukas, "Panel Data Nowcasting: The Case of Price-Earnings Ratios," 2023. [Online]. Available: https://arxiv.org/abs/2307.02673

[92] S. F. Crone and N. Kourentzes, "Feature selection for time series prediction – A combined filter and wrapper approach for neural networks," *Neurocomputing*, vol. 73, no. 10–12, pp. 1923–1936, Jun. 2010, doi: 10.1016/j.neucom.2010.01.017.

[93] N. J. Clark and K. Wells, "Dynamic generalised additive models (DGAMs) for forecasting discrete ecological time series," *Methods Ecol. Evol.*, vol. 14, no. 3, pp. 771–784, Mar. 2023, doi: 10.1111/2041-210X.13974.

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

[95] L. Bottou, F. E. Curtis, and J. Nocedal, "Optimization Methods for Large-Scale Machine Learning," *SIAM Rev.*, vol. 60, no. 2, pp. 223–311, Jan. 2018, doi: 10.1137/16M1080173.

[96] M. T. Hagan, H. B. Demuth, and M. Beale, *Neural Network Design*, 2nd ed. Martin Hagan, 2014.

[97] O. A. Montesinos López, A. Montesinos López, and J. Crossa, "Fundamentals of Artificial Neural Networks and Deep Learning," in *Multivariate Statistical Machine Learning Methods for Genomic Prediction*, Cham: Springer International Publishing, 2022, pp. 379–425. doi: 10.1007/978-3-030-89010-0_10.

[98] A. A. Waoo and B. K. Soni, "Performance Analysis of Sigmoid and Relu Activation Functions in Deep Neural Network," 2021, pp. 39–52. doi: 10.1007/978-981-16-2248-9_5.

[99] J. Urenda, O. Kosheleva, and V. Kreinovich, "Fuzzy Techniques Explain the Effectiveness of ReLU Activation Function in Deep Learning," 2024, pp. 425–430. doi: 10.1007/978-3-031-55684-5_28.

[100] Y. Kim, I. Ohn, and D. Kim, "Fast convergence rates of deep neural networks for classification," 2019. [Online]. Available: https://arxiv.org/abs/1812.03599

[101] V. M. Vargas, D. Guijo-Rubio, P. A. Gutiérrez, and C. Hervás-Martínez, "ReLU-Based Activations: Analysis and Experimental Study for Deep Learning," 2021, pp. 33–43. doi: 10.1007/978-3-030-85713-4_4.

[102] B. N. Oreshkin, D. Carpov, N. Chapados, and Y. Bengio, "N-BEATS: Neural basis expansion analysis for interpretable time series forecasting," 2020. [Online]. Available: https://arxiv.org/abs/1905.10437

[103] M. Shah and N. Sureja, "A Comprehensive Review of Bias in Deep Learning Models: Methods, Impacts, and Future Directions," *Arch. Comput. Methods Eng.*, vol. 32, no. 1, pp. 255–267, Jan. 2025, doi: 10.1007/s11831-024-10134-2.

[104] H. Abbasimehr and R. Paki, "Improving time series forecasting using LSTM and attention models," *J. Ambient Intell. Humaniz. Comput.*, vol. 13, no. 1, pp. 673–691, Jan. 2022, doi: 10.1007/s12652-020-02761-x.

[105] S. Siddique, M. A. Haque, R. George, K. D. Gupta, D. Gupta, and M. J. H. Faruk, "Survey on Machine Learning Biases and Mitigation

Techniques," *Digital*, vol. 4, no. 1, pp. 1–68, Dec. 2023, doi: 10.3390/digital4010001.

[106] R. Wang, P. Chaudhari, and C. Davatzikos, "Bias in machine learning models can be significantly mitigated by careful training: Evidence from neuroimaging studies," *Proc. Natl. Acad. Sci.*, vol. 120, no. 6, Feb. 2023, doi: 10.1073/pnas.2211613120.

[107] R. J. Hyndman and G. Athanasopoulos, *Forecasting: principles and practice, 3rd edition*. Melbourne, Australia.: OTexts, 2021. [Online]. Available: https://otexts.com/fpp3/

[108] M. M. Msabi and M. Makonyo, "Flood susceptibility mapping using GIS and multi-criteria decision analysis: A case of Dodoma region, central Tanzania," *Remote Sens. Appl. Soc. Environ.*, vol. 21, p. 100445, Jan. 2021, doi: 10.1016/j.rsase.2020.100445.

[109] D. Kvočka, R. A. Falconer, and M. Bray, "Flood hazard assessment for extreme flood events," *Nat. Hazards*, vol. 84, no. 3, pp. 1569–1599, Dec. 2016, doi: 10.1007/s11069-016-2501-z.

[110] N. Tabasi, M. Fereshtehpour, and B. Roghani, "A review of flood risk assessment frameworks and the development of hierarchical structures for risk components," *Discov. Water*, vol. 5, no. 1, p. 10, Feb. 2025, doi: 10.1007/s43832-025-00193-2.

[111] M. Pronk *et al.*, "DeltaDTM: A global coastal digital terrain model," *Sci. Data*, vol. 11, no. 1, p. 273, Mar. 2024, doi: 10.1038/s41597-024-03091-9.

[112] M. Yalcin, "A GIS-Based Multi-Criteria Decision Analysis Model for Determining Glacier Vulnerability," *ISPRS Int. J. Geo-Information*, vol. 9, no. 3, p. 180, Mar. 2020, doi: 10.3390/ijgi9030180.

[113] UN-SPIDER, "Flood Mapping and Damage Assessment Using Sentinel-1 SAR Data in Google Earth Engine: Step-by-Step," 2025. [Online]. Available: https://www.un-spider.org/advisory-support/recommended-practices/recommended-practice-google-earth-engine-flood-mapping/step-by-step

[114] UN-SPIDER, "Flood Hazard Assessment: Step-by-Step," 2025. [Online]. Available: https://www.un-spider.org/advisory-support/recommended-practices/recommended-practice-flood-hazard-assessment/step-by-step

[115] BIG, "Demnas." [Online]. Available: https://www.big.go.id/content/produk/demnas

[116] M. Karpowicz and J. Ejsmont-Karabin, "Diversity and Structure of Pelagic Zooplankton (Crustacea, Rotifera) in NE Poland," *Water*, vol. 13, no. 4, p. 456, Feb. 2021, doi: 10.3390/w13040456.