Master's Thesis

Random Forest and CNN-Based Hybrid Modelling Approach for Susceptibility of Flood.

Name: **Ammar Yousaf**
Student id: **554324**
Supervisor: **Prof. Dr. Hanna Meyer (Institute for Landscape Ecology, University of Münster, Germany)**
Co-supervisor: **Prof. Dr. Marco Painho (Nova University Lisbon, Portugal) and Sergi Trilles Oliver(Jaume I University, Spain)**
Submitted to: **University of Münster**

DECLARATION OF ORIGINALITY

I declare that the work described in this document is my own and not from someone else. All the assistance I have received from other people is duly acknowledged, and all the sources (published or not published) are referenced.

I also declare that I have used Large Language Models (LLMs) i.e., ChatGPT and DeepSeek for tasks like assisting in writing for correction and in literature review, while ensuring the originality, accuracy, and integrity of my research. This work has not been previously evaluated or submitted anywhere. Münster,

20 February 2025

Ammar Yousaf

Abstract

Floods are one of the most destructive natural disasters, causing significant socio-economic losses, environmental degradation, and infrastructure damage. Effective flood susceptibility assessment is essential for disaster preparedness, risk mitigation, and urban planning. Traditional hydrological models often struggle to capture the complex interplay between climatic, topographical, and hydrological factors influencing flood occurrences. In response, machine learning techniques have emerged as powerful alternatives due to their ability to analyze large datasets and recognize intricate patterns.

This study proposes a hybrid modeling approach that combines the strengths of Random Forest (RF) and Convolutional Neural Networks (CNN) to improve flood susceptibility prediction. RF is utilized for feature selection, identifying the most influential variables that contribute to flood risk, while CNN leverages spatial dependencies in geospatial data to enhance predictive accuracy. The methodology is applied to Düsseldorf, Germany, a flood-prone urban area, using a dataset that includes topographical, hydrological, and meteorological factors such as elevation, slope, distance to river networks, soil moisture, and extreme precipitation events.

The results indicate that the hybrid RF-CNN model outperforms standalone models in both classification accuracy and spatial consistency. RF effectively ranks critical flood-inducing factors, while CNN provides high-resolution flood-prone area delineation. Key findings reveal that proximity to river networks, topographic wetness index, and extreme precipitation frequency are the dominant contributors to flood susceptibility. The hybrid approach bridges the gap between interpretability and precision, offering a comprehensive tool for flood hazard assessment.

By integrating data-driven insights with spatial analysis, this study provides a robust framework that can aid policymakers, urban planners, and disaster management authorities in developing targeted flood mitigation strategies. The findings highlight the potential of hybrid machine learning models in enhancing climate resilience and improving early warning systems for flood-prone regions.

Acknowledgment:

First and foremost, I wish to express my profound gratitude to Allah, whose blessings and guidance have illuminated every step of this journey. Without His support, navigating the twists and turns of this thesis would have been an insurmountable challenge.

I am deeply indebted to my parents, whose unwavering belief in my potential has been both a comfort and a catalyst throughout this endeavor. Their constant support, even when my ideas occasionally bordered on the fantastical, kept me grounded and motivated.

A special thank you goes to my best friend, Nisha, whose assistance with the presentation was nothing short of miraculous. Her knack for transforming my scattered thoughts into a coherent and engaging narrative was, at times, as impressive as turning lead into gold. I appreciate her for never disturbing me during my thesis.

My gratitude further extends to Uzzi, whose support in various aspects of my life has been invaluable. His encouragement and keen insights played a significant role in shaping my work. I would also like to extend my sincere thanks to Taimoor Bhai for his guidance and support throughout this journey. His contributions have been invaluable, and I am truly grateful for his presence.

I am especially grateful to my supervisors—Hanna Mayer, my main supervisor, along with my co-supervisors Marco Painho and Trilles—for their invaluable feedback and unwavering support. Their insightful guidance and constructive criticism helped shape this thesis into a work that I can be truly proud of.

A heartfelt thanks is also due to Marra G, whose unwavering support in all facets of life has been a constant source of encouragement. While his assistance was not confined solely to my thesis, his steady presence and insightful guidance in countless other endeavors have enriched my journey in ways that extend far beyond academic pursuits.

I would also like to express my appreciation to Tishya for her guidance in the visualization aspects of GIS. Her expertise helped me refine and effectively present spatial data, making complex information more accessible and insightful.

Each of these individuals has contributed uniquely to the realization of this thesis, and it is with deep appreciation that I recognize their roles in making this work a reality.

Table of Contents

List of Tables and Figures:

1. Introduction

Flooding, one of the most devastating natural disasters, poses significant threats to human life, infrastructure, and ecosystems globally. Flood risk defined as the likelihood of occurrence and potential consequences and flood vulnerability the susceptibility of populations, infrastructure, and ecosystems to damage were starkly illustrated during the catastrophic July 2021 flood in Germany's Rhineland region, particularly impacting North Rhine-Westphalia (NRW) and Rhineland-Palatinate. Triggered by unprecedented rainfall over three days, the event caused rivers to overflow and inundate both urban and rural landscapes, emphasizing the urgent need for advanced flood prediction and mitigation strategies.

Accurate identification of flood-prone areas is a cornerstone of effective disaster risk reduction and mitigation planning. Traditional models that rely on historical data and hydrological simulations often struggle to capture the rapidly evolving climatic and land-use conditions that intensify modern flood dynamics (Mobley et al., 2021). In response, machine learning (ML) techniques have emerged as powerful alternatives, capable of uncovering complex, nonlinear patterns within large, heterogeneous datasets. Among these, Random Forest (RF) and Convolutional Neural Networks (CNN) have shown particular promise. RF excels at quantifying the relative importance of input variables such as elevation, soil moisture, or land-use features (Alamoudi, 2023) while CNNs leverage spatial hierarchies in datasets like digital elevation models (DEMs) and flow-accumulation rasters to delineate flood-prone zones with high precision (Bhusal et al., 2022).

A critical challenge in applying ML methods to flood prediction is variable selection. Including irrelevant or redundant features can lead to overfitting and reduce model interpretability (Laukhtina et al., 2021), thereby undermining both model performance and stakeholder trust. Random Forest offers distinct advantages for variable selection through its ensemble learning framework. By aggregating the results of multiple decision trees, RF reduces overfitting, effectively handles multicollinearity, and provides robust, built-in rankings of feature importance. These capabilities ensure that only the most relevant predictors are utilized, ultimately enhancing the predictive accuracy and interpretability of the flood susceptibility model.

This thesis hypothesizes that (1) RF outperforms CNN in variable selection due to its robust feature-importance rankings, and (2) a hybrid RF-CNN model combining RF's ability to identify key predictors with CNN's strength in capturing spatial hierarchies improves flood susceptibility mapping. The methodology leverages data from districts surrounding

Düsseldorf for model training, while Düsseldorf itself owing to its dense urbanization and historical flood events—serves as the testing ground. Specifically, the approach involves: Variable Selection: Training an RF model on data from districts around Düsseldorf to identify the most influential predictors; Spatial Analysis: Incorporating these RF-selected variables into a CNN model designed to capture spatial dependencies and generate detailed flood susceptibility maps; and Validation: Testing the integrated model in Düsseldorf by comparing the generated flood susceptibility maps with historical flood extents (1999–2021) to assess model performance and real-world applicability.

The significance of this approach lies in its ability to bridge the gap between model accuracy and interpretability a balance critical for effective policymaking and community engagement. By clearly identifying the most influential variables and pinpointing high-risk zones at fine spatial resolutions, local authorities can optimize resource allocation for infrastructure reinforcement and early-warning system improvements. Moreover, the hybrid methodology developed in this study is transferable to other urban floodplains, thereby contributing to broader efforts in enhancing climate resilience and disaster preparedness.

2. Literature Review

2.1. Overview of Flood Prediction

Flooding is a critical natural hazard with far-reaching implications for human life, infrastructure, and ecosystems. Traditional flood prediction approaches encompassing statistical and deterministic hydrological models have historically relied on parameters such as precipitation, soil moisture, and river discharge (Mobley et al., 2021). Although these methods have provided valuable insights, they often assume stationarity and struggle to account for rapid changes in climate and land use (Mobley et al., 2021). Consequently, their predictive capability can become less reliable in modern, dynamic environments.

2.2. Traditional Flood Prediction Methods

Hydrological models, such as the Hydrologic Engineering Center's River Analysis System (HEC-RAS) and the Soil and Water Assessment Tool (SWAT), are widely employed for flood prediction (Alamoudi, 2023). These models simulate water flow through river networks and watersheds by integrating inputs like rainfall, soil characteristics, and land use. However, their dependence on extensive calibration, high-quality data, and computationally intensive simulations can limit their effectiveness especially in data-scarce regions or when real-time forecasting is required. These limitations have motivated researchers to explore machine learning (ML) techniques as promising alternatives or complements to traditional hydrological models.

2.3. Machine Learning Techniques in Flood Prediction

The application of ML in flood prediction represents a paradigm shift from purely mechanistic models to data-driven approaches. ML algorithms are adept at capturing nonlinear relationships in high-dimensional datasets a feature particularly useful in urban environments where both natural processes and anthropogenic factors (e.g., impervious surfaces, stormwater infrastructure) interact to influence flood dynamics (Alamoudi, 2023).

2.3.1. Random Forest (RF)

Random Forest (RF) is an ensemble learning algorithm that constructs multiple decision trees using bootstrap samples of the data and aggregates their predictions to produce a consensus output . This ensemble approach reduces variance and mitigates overfitting, making RF robust even in the presence of noisy or high-dimensional data. A key advantage of RF is its inherent ability to estimate variable importance. Using measures such as the mean decrease in impurity or accuracy, RF ranks predictors (e.g., elevation, soil moisture, precipitation) based on their contribution to predictive performance. This built-in feature-ranking mechanism makes RF particularly valuable for flood risk assessment where understanding the influence of various environmental factors is essential. Moreover, compared to methods like principal component analysis (PCA), which transform variables and can obscure interpretability, RF maintains the original data structure while offering clear, data-driven insights (Genuer et al., 2010).

2.3.2. Convolutional Neural Networks (CNN)

Convolutional Neural Networks (CNNs) are specialized neural networks designed to process grid-based or image-like data. By convolving input data with learnable filters, CNNs detect spatial hierarchies and patterns within datasets such as digital elevation models (DEMs) and remote sensing imagery (Bhusal et al., 2022). This capability is particularly useful in flood prediction when analyzing complex terrain features and drainage patterns. Although CNNs excel at spatial feature extraction, they typically require extensive training data and high computational resources. In contrast to RF, CNNs are not inherently designed for tabular variable ranking but rather for learning spatial relationships from structured images.

2.4. Feature Selection for Flood Prediction Models

Feature selection is a critical step in enhancing ML model performance and interpretability. Including irrelevant or redundant variables can lead to overfitting, reduce clarity, and impose unnecessary computational overhead (Laukhtina et al., 2021). Effective feature selection not only improves accuracy but also helps decision-makers understand which factors most significantly influence flood risk.

2.4.1. Random Forest for Feature Selection

Random Forest is particularly well-suited for feature selection because it inherently computes variable importance scores during model training. These scores enable the identification of critical predictors such as land use, elevation, and precipitation—that most significantly contribute to flood susceptibility (Reaney, 2022). Compared to other feature selection methods such as LASSO (which assumes linear relationships) or PCA (which transforms data into a lower-dimensional space at the expense of interpretability), RF provides a transparent, computationally efficient means of ranking the original variables.

This clarity is essential for environmental modeling where understanding the impact of each variable is crucial. In our context, RF's ability to distill key tabular predictors ensures that only the most relevant features are passed to subsequent spatial analyses, thereby strengthening the overall modeling framework (Genuer et al., 2010).

2.4.2. Convolutional Neural Networks for Feature Selection

While CNNs are predominantly used for extracting spatial features from image-like data, some studies have explored gradient-based or layer-wise relevance propagation (LRP) techniques to interpret which input regions most influence the CNN's output (Pouyan et al., 2021). However, these methods typically operate at the pixel or region level and do not directly rank tabular predictors as RF does. Moreover, extracting such interpretability from CNNs often requires advanced computational techniques and larger datasets, which may not be feasible in all applications. Importantly, in our hybrid approach, RF is used exclusively for feature selection from tabular datasets, while the CNN is leveraged for its strength in learning spatial hierarchies from geospatial data. This complementary use of RF and CNN ensures that the spatial model is informed by the most relevant predictors without necessitating the complex interpretability processes required for CNN-based feature selection.

2.4.3. Rationale: Tabular vs. Image-Based Feature Extraction

A key consideration in our approach is the distinction between tabular data and image-based spatial data. RF efficiently processes tabular data and provides a clear ranking of variables, making it ideal for selecting the most influential environmental predictors. In contrast, CNNs are specifically designed to learn and extract spatial patterns from image-like inputs. By using RF to pre-select variables and then feeding these high-quality inputs into a CNN, we combine the strengths of both methods—ensuring that the CNN benefits from the most relevant predictors while focusing on extracting spatial relationships. This dual strategy addresses the limitation that RF does not learn spatial patterns, by delegating that task to the CNN, which is inherently optimized for spatial data analysis.

2.5. Challenges and Research Gaps

Despite the growing integration of ML techniques in flood prediction, significant challenges and research gaps persist. Data availability and quality remain critical issues, particularly in regions with limited monitoring infrastructure. High-resolution hydrological, topographical, and meteorological data are essential for training robust ML models, but sparse or low-quality datasets can compromise model accuracy (Mobley et al., 2021). Additionally, the interpretability of complex models, especially deep learning methods like CNNs, remains a challenge for communicating results to policymakers and stakeholders. While RF offers transparent feature importance metrics, the lack of standardized protocols for feature selection introduces variability across studies. Moreover, comprehensive comparative evaluations of ML approaches under diverse data conditions are scarce, underscoring the need for further research to establish context-specific guidelines for model selection in flood prediction (Pouyan et al., 2021).

3. Objective and Research questions

The main objective of this research is to develop a hybrid flood susceptibility model that combines the feature importance capabilities of Random Forest (RF) with the spatial pattern recognition strengths of Convolutional Neural Networks (CNN). In practical terms, this model should offer clear insights into which variables matter most for flood risk while also capturing the spatial relationships critical to accurate flood-prone area delineation.

3.1. Sub-Objectives

I. **Identify Critical Predictive Variables**
A primary aim is to determine which environmental, hydrological, and geomorphological factors most strongly influence flood risk, using RF-based feature importance. This step is essential for reducing model complexity and enhancing interpretability.

II. **Compare Standalone Model Performance**
Because RF and CNN specialize in different aspects of data analysis, a direct comparison of their predictive accuracy is necessary. This comparison will illuminate each method's relative strengths and weaknesses in delineating flood-prone areas.

III. **Develop a Hybrid RF-CNN Framework**
Building on the first two objectives, the study aims to design and implement a hybrid model that unites RF's ranked features with CNN's spatial feature extraction capabilities. The goal is to boost overall accuracy in flood susceptibility mapping by leveraging the complementary advantages of both algorithms.

IV. **Access the Impact of variable selection**
This study quantitatively examines how the inclusion or exclusion of key environmental variables affects the performance of the hybrid RF-CNN model. Specifically, it evaluates the sensitivity of RF's variable-ranking mechanism and CNN's spatial feature extraction to different input configurations. By systematically analyzing these effects, the research aims to identify optimal model configurations that enhance predictive accuracy while reducing computational overhead.

3.2. Research Questions

Based on the objective and the subobjectives, the study seeks to address the following questions:

RQ1: Which variables most strongly influence flood risk, according to Random Forest's feature importance analysis?

This question explores the rankings of environmental, hydrological, and geomorphological features crucial for flood prediction.

RQ2: How does the predictive accuracy of Random Forest compare to that of Convolutional Neural Networks in mapping flood-prone areas?

By directly comparing these two ML approaches, the study aims to quantify each method's strengths and limitations.

RQ3: How well does a hybrid approach combining variable selection via Random Forest with training via Convolutional Neural Networks perform in flood susceptibility assessment compared to models that use only Random Forest or only Convolutional Neural Networks?

This research question aims to evaluate the comparative performance, accuracy, and efficiency of the hybrid methodology against standalone RF and CNN models, thereby clarifying the benefits of integrating robust feature selection with advanced spatial pattern recognition.

4. Study Area and Datasets

4.1. Study Area

The primary focus of this research is Düsseldorf, a densely populated and economically vital district along the Rhine River in western Germany (Figure 1). Serving as the capital of North Rhine-Westphalia, Düsseldorf features a dynamic blend of urban and industrial zones interspersed with low-lying floodplains adjacent to the Rhine. Historically, the area has experienced both pluvial (rainfall-induced) and fluvial (river-induced) flooding, particularly during seasons of intense precipitation or high river discharge. These recurring flood events have underscored the need for reliable flood risk assessments and mitigation strategies.

4.1.1. Geographical and Climatic Context

The primary study area is centered on Düsseldorf, Germany (approximately 51.2277° N, 6.7735° E), located within the hydrologically dynamic Lower Rhine Basin. Düsseldorf experiences a temperate oceanic climate with average annual precipitation between 800 and 900 mm; however, intense autumn and winter rainfall events can overwhelm local drainage systems, substantially increasing flood risk. The city itself exhibits notable elevation gradients from low-lying floodplain areas near the Rhine (around 28 m above sea level) to upland zones reaching up to 165 m. While Düsseldorf serves as a critical urban center within the basin, the broader Lower Rhine Basin is characterized by a diverse range of environmental conditions, including variations in elevation, land use, and proximity to river networks. This regional diversity is essential for developing flood prediction models that are robust across different settings.

4.1.2. Framework for Selecting Düsseldorf

Düsseldorf is chosen as the primary reference site due to its significance as a major commercial and administrative hub, where flood events can lead to considerable socio-economic impacts. The city has a documented history of major flood events between 1991 and 2021, including the Rhine floods of 1993 and 2021, which have resulted in widespread damage to infrastructure and communities (Du et al., 2010). Moreover, the availability of high-resolution geospatial and meteorological datasets in Düsseldorf provides a solid foundation for calibrating and validating predictive flood models. Although the overall modeling framework is designed for application across the Lower Rhine Basin, focusing on Düsseldorf allows for a rigorous evaluation of model performance in a high-risk, data-rich urban context.

4.1.3. Inclusion of Adjacent Districts

To enhance the generalizability of the flood prediction model across the Lower Rhine Basin, the study also incorporates data from adjacent districts such as Duisburg, Essen, Krefeld, Mettmann, Mülheim an der Ruhr, Oberhausen, Rhein-Kreis Neuss, and Wesel. Although these districts share a broadly similar climatic regime with Düsseldorf, they exhibit significant local variations in topographical gradients, land cover patterns, and their spatial relationships to major river channels and tributaries factors that result in distinct flood risk profiles. Integrating data from these diverse regions into the training phase ensures that the predictive model captures a wide range of environmental variability. Importantly, while the model is trained on this comprehensive dataset, its performance is specifically tested on Düsseldorf, ensuring that the final framework is both broadly applicable and rigorously validated in a high-priority urban area.

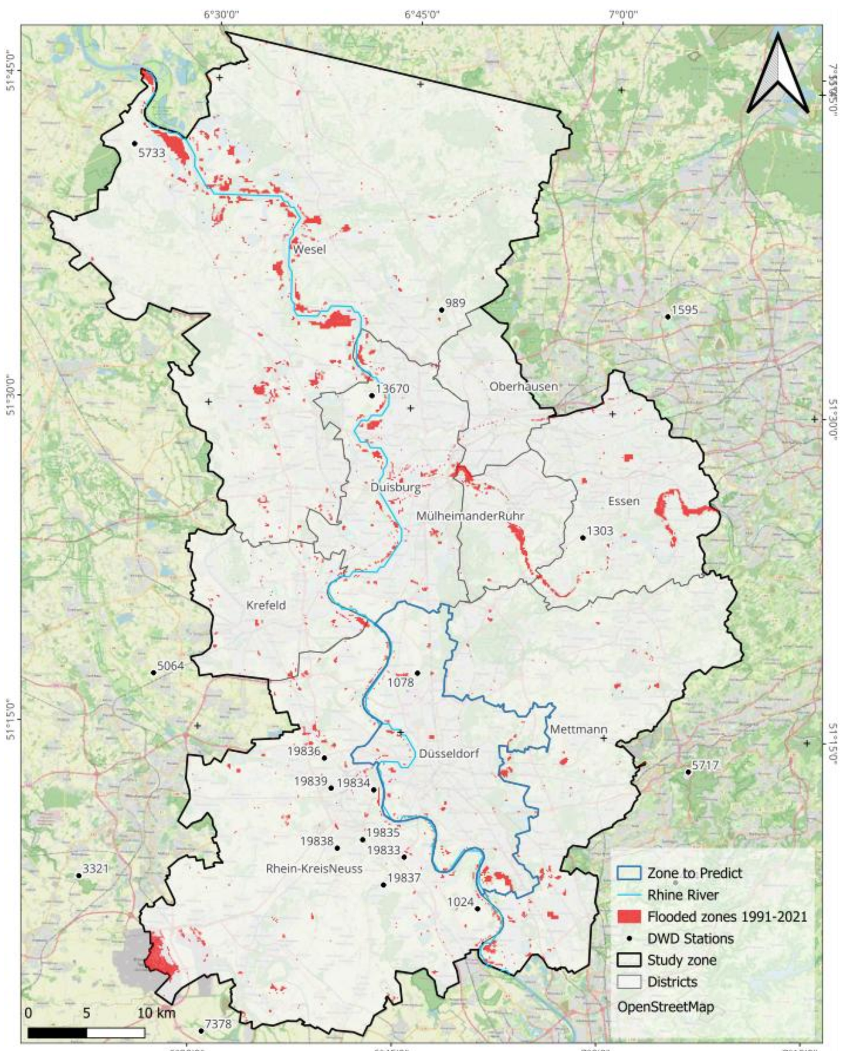

Figure 1: Study area map of Dusseldorf and neighboring districts.

4.2. Datasets

This study employs **10 key variables** spanning topographical, hydrological, climatic, and flood history categories. Each variable was chosen based on its well-documented relevance to flood risk modeling. Detailed explanations on how each variable was derived are provided in the Methodology section. Table 1 summarizes these variables along with their descriptions and data source.

4.2.1. Topographical Variables

i. Altitude (Elevation)

Altitude is a critical determinant of flood susceptibility, as low-lying areas are more prone to floodwater accumulation, particularly along riverbanks. In this study, altitude data were derived from Sentinel-1 Synthetic Aperture Radar (SAR) datasets. Specifically, we acquired Sentinel-1 Level-1 Ground Range Detected (GRD) products from the Copernicus Open Access Hub (https://scihub.copernicus.eu/). These raw datasets were processed in-house using the European Space Agency's (ESA) Sentinel Application Platform (SNAP) software , which produced a Digital Elevation Model (DEM) with an approximate spatial resolution of 8 meters. This DEM was subsequently used to accurately delineate areas susceptible to inundation in fluvial floodplain.

ii. Slope

Slope quantifies terrain steepness to assess runoff velocity, calculated from the altitude dataset using QGIS. Steeper slopes facilitate rapid water drainage, reducing flood retention, while gentler slopes promote prolonged water accumulation, increasing flood risk in flat or gradually inclined areas. By integrating slope analysis, the model captures spatial variations in hydrological behavior critical for predicting runoff dynamics and flood-prone zones.

iii. Topographic Wetness Index (TWI)

The Topographic Wetness Index (TWI) measures the likelihood of soil saturation by combining slope and flow accumulation parameters derived from Digital Elevation Models (DEMs). This index highlights valleys, depressions, and other topographic concavities where water naturally converges, making these areas prone to prolonged inundation (Shuaibu et al., 2022). TWI enhances the identification of localized saturation hotspots, which are often overlooked in coarse-scale flood susceptibility assessments.

iv. Curvature

Curvature distinguishes concave (water-collecting) and convex (water-shedding) terrain forms, aiding in the identification of natural water retention zones. Concave areas, such as basins or floodplains, act as natural reservoirs during heavy rainfall, while convex slopes promote rapid runoff dispersal (Shuaibu et al., 2022). This variable refines flood prediction by accounting for microtopographic features that influence surface water distribution and accumulation patterns.

v. Aspect

Aspect evaluates the directional orientation of slopes, which impacts microclimatic factors such as solar exposure and evaporation rates. For instance, north-facing slopes in the Northern Hemisphere receive less direct sunlight, leading to cooler temperatures and higher soil moisture retention compared to south-facing slopes (Shuaibu et al., 2022). By integrating aspect, the model captures variations in hydrological processes driven by slope orientation, further refining flood susceptibility estimates in heterogeneous landscapes.

4.2.2. Hydrological Variables

vi. Distance to River

Distance to River quantifies the Euclidean proximity of each pixel to the Rhine River or its major tributaries, serving as a direct indicator of flood susceptibility. Areas closer to river channels are inherently more vulnerable to inundation during peak discharge events, as rising water levels frequently breach banks and spill into adjacent low-lying zones (Rasn et al., 2021). This variable helps delineate floodplain boundaries and prioritize regions for mitigation efforts, as historical flood patterns consistently highlight heightened risk near fluvial networks.

vii. Flow Accumulation

Flow Accumulation identifies zones where surface runoff converges, such as river channels, drainage basins, or topographic depressions. Calculated from digital elevation models (DEMs), this variable reflects the cumulative upstream area contributing to runoff at a given location. High flow accumulation values correlate strongly with flood hotspots, as concentrated water flow overwhelms natural and artificial drainage capacities, exacerbating flood magnitudes (Mukhtar et al., 2024; Waseem et al., 2023). By mapping these convergence areas, the model pinpoints critical locations where hydrological forces amplify flood risks, informing targeted infrastructure and emergency response planning.

4.2.3. Climatic and Soil Variables

viii. Soil Temperature

Soil Temperature, sourced from the German Weather Service (DWD) at approximately 1 km resolution and aggregated from station data, directly impacts hydrological processes by modulating evaporation rates and moisture retention. Cooler soil temperatures can decelerate evaporation, prolonging water retention in the soil matrix and exacerbating flood risks during prolonged rainfall events (S et al., 2017). This variable provides insight into antecedent conditions that influence runoff generation, particularly in regions where thermal inertia affects soil permeability and antecedent moisture levels.

ix. Soil Moisture

Soil Moisture quantifies pre-event saturation levels, a critical determinant of infiltration capacity. High soil moisture prior to rainfall events reduces the soil's ability to absorb additional water, leading to accelerated surface runoff and heightened flood magnitudes (Waseem et al., 2023). By integrating pre-storm saturation data, the model accounts for antecedent hydrological conditions, improving predictions of flash flooding in areas with limited drainage infrastructure or impermeable surfaces.

x. Annual Maximum Daily Precipitation (AP)

Annual Maximum Daily Precipitation (AP) represents the highest daily rainfall recorded each year between 1991 and 2021. Extreme single-day precipitation events overwhelm drainage systems, particularly in urbanized areas, and are a primary driver of pluvial flooding (Quirós et al., 2020). AP captures the intensity of rainfall extremes, enabling the model to identify thresholds beyond which infrastructure resilience is compromised, thereby pinpointing high-risk zones for inundation.

xi. Frequency of Extreme Precipitation (FP)

Frequency of Extreme Precipitation (FP) measures the annual count of days exceeding 50 mm of rainfall, adding a temporal dimension to extreme rainfall analysis. This metric highlight shifts in precipitation regimes, such as increased clustering of heavy rainfall events linked to climate variability (Quirós et al., 2020). By evaluating FP, the model assesses not only the magnitude but also the recurrence of extreme rainfall, offering insights into long-term flood risk trends and the compounding stressors on flood mitigation systems.

4.2.4. Historical Flood Data

xii. Historical Flood Maps (1999–2021)

Historical flood extents, sourced from the DFO Flood Observatory, comprise polygon shapefiles documenting inundation events between 1999 and 2021. These maps provide spatially explicit records of past flood occurrences, serving as a critical benchmark for validating predictive models. By overlaying simulated flood-prone areas with empirically observed extents, the accuracy of the model is rigorously assessed, ensuring alignment between predictions and real-world events. This validation step enhances the reliability of flood susceptibility maps, enabling stakeholders to prioritize mitigation efforts in zones where historical and predicted risks converge. The integration of multi-decadal flood data also supports the identification of persistent vulnerability patterns, informing long-term resilience planning in flood-prone regions.

4.3. Framework for Data Selection

To ensure comprehensive coverage of flood risk factors, each variable was chosen based on its physical, hydrological, or climatic significance:

- **Topographical Drivers** (altitude, slope, curvature) capture how terrain shapes floodwater flow and retention.
- **Hydrological Triggers** (distance to river, flow accumulation) pinpoint water movement pathways crucial for flood propagation.
- **Climatic and Soil Indicators** (soil temperature, moisture, extreme precipitation frequencies) reveal seasonal and **long-term** weather patterns that can exacerbate flooding.
- **Historical Flood Data** allows for ground-truthing model outputs, ensuring alignment with empirical flood extents.

Incorporating data from adjacent districts helps the model generalize across diverse sub-regional conditions within the Lower Rhine Basin, while Düsseldorf-specific data ensures fine-tuned accuracy for local flood risk planning and mitigation.

5. Methodology:

The methodology follows a structured workflow to develop and validate a hybrid flood susceptibility model integrating Random Forest (RF) and Convolutional Neural Networks (CNN). The process is divided into four phases: data collection, preprocessing, model training/validation, and flood risk mapping, as detailed below in figure 2.

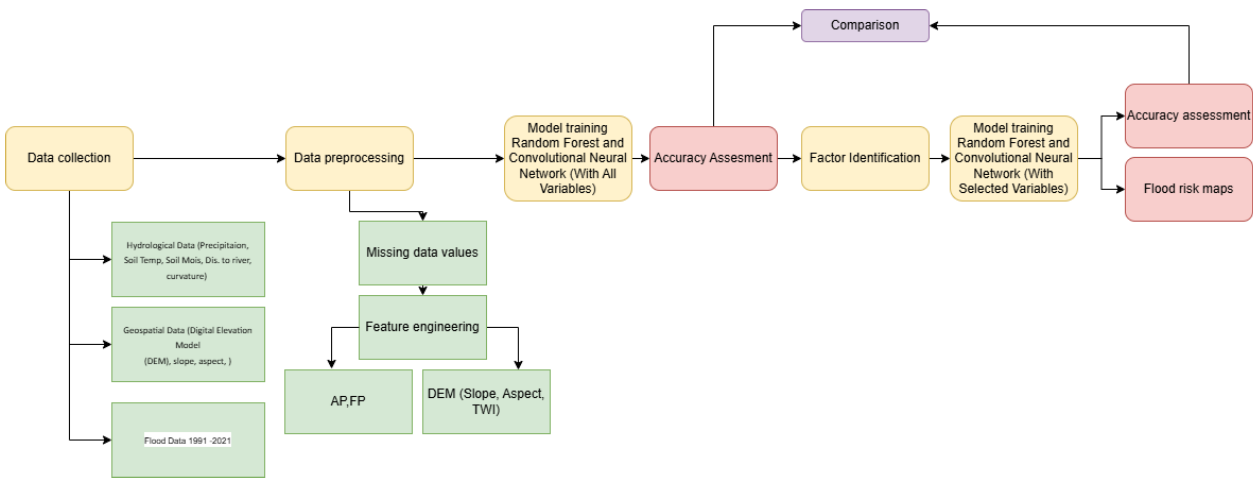

Figure 2: Methodological Flowchart

5.1. Data Collection

Data was gathered from multiple sources to capture physical, climatic, and anthropogenic flood drivers:

Parameter	Description	Data Source	Processing Tool / Method	Reference / URL
Altitude (DEM)	Elevation data at an 8 m resolution; key for identifying low-lying, flood-prone areas	ESA Sentinel-1 SAR data	Processed using ESA's SNAP software to generate a Digital Elevation Model (DEM)	[ESA Sentinel-1]()
Slope	Slope gradient derived from the DEM; affects runoff velocity and water accumulation	ESA Sentinel-1 SAR data	Derived from the DEM using SNAP's terrain	(See Sentinel-1)

			analysis tools	
Topographic Wetness Index (TWI)	Indicator of geotechnical wetness computed from the DEM, highlighting areas likely to be flood-prone	ESA Sentinel-1 SAR data	Computed from the DEM using hydrological analysis routines in SNAP	(See Sentinel-1)
Distance to River	Euclidean distance from each location to the nearest river, a critical factor in flood susceptibility	Derived from the same DEM based on ESA Sentinel-1 SAR data	Calculated using GIS spatial analysis methods	(See Sentinel-1 & GIS methods)
Flow Accumulation	Represents the accumulation of runoff; high values indicate convergence zones prone to flooding	Derived from the DEM based on ESA Sentinel-1 SAR data	Computed with hydrological modeling tools (algorithms available in SNAP)	(See Sentinel-1 & corresponding hydrological modeling documentation)
Aspect	Direction of the steepest slope derived from the DEM	Derived from the DEM based on ESA Sentinel-1 SAR data	Calculated using SNAP's terrain analysis functions	(See Sentinel-1 & SNAP links above)
Curvature	Reflects the rate of change in slope (concave areas may retain water, increasing flood risk)	Derived from the DEM based on ESA Sentinel-1 SAR data	Processed using SNAP's curvature analysis routines	(See Sentinel-1 & SNAP links above)
Soil Temperature	Average soil temperature affecting infiltration and evaporation rates in the study area	German Weather Service (DWD)	Obtained from DWD datasets and incorporated into the analysis	DWD Home

Soil Moisture	Mean soil moisture levels; high values indicate saturated soils and increased flood risk	German Weather Service (DWD)	Retrieved from DWD climate and environmental data repositories	DWD Open Data
Annual Maximum Daily Precipitation (AP)	Maximum daily precipitation depth recorded between 1991 and 2021; a key hydrometeorological indicator	German Weather Service (DWD)	Extracted from long-term precipitation records available from DWD	DWD Climate Data
Frequency of Extreme Precipitation Events (FP)	Frequency of days with precipitation >50 mm (1991–2021); indicates extreme rainfall events	German Weather Service (DWD)	Derived from daily precipitation datasets provided by DWD	(See DWD Climate Data link above)
Flooded Zones/Training Data (1999–2021)	Historical maps of flood extents over the study period; used to validate and calibrate flood hazard assessments	DFO Flood Observatory, University of Colorado	Downloaded as flood hazard maps and integrated with other datasets for flood risk analysis	https://floodobservatory.colorado.edu/

Table 1: Summarized datasets and sources.

5.1.1. Hydrometeorological Data

Hydrometeorological data captures temporal weather conditions, providing essential insights into climatic variations. Precipitation data includes the Annual Maximum Daily Precipitation (AP) and the Frequency of Extreme Precipitation Events (FP) (daily rainfall exceeding 50 mm), obtained from the German Weather Service (DWD) for the period 1991–2021. Soil properties are assessed through Soil Moisture and Soil Temperature records collected from DWD monitoring stations. Additionally, the distance to rivers is analyzed using Euclidean distance calculations to measure proximity to the Rhine River and its tributaries, derived through GIS tools.

5.1.2. Geospatial Data

Geospatial data captures static or slowly changing terrain features, providing valuable insights into landscape characteristics. The Digital Elevation Model (DEM), sourced from ESA Sentinel-1 SAR at an 8-meter resolution, is utilized to derive key topographic attributes such as slope, aspect, and the Topographic Wetness Index (TWI). These features help in understanding terrain morphology and hydrological patterns.

5.1.3. Historical Flood Data

Flooded zones (1991–2021) from the DFO Flood Observatory, providing ground-truth data for model validation.

5.2. Data Preprocessing

Raw datasets underwent systematic cleaning and transformation to ensure consistency, completeness, and compatibility with modeling requirements. Figure 3 shows maps of predicting variables. The preprocessing workflow included the following steps:

5.2.1. Handling Missing Data

- **Soil Moisture and Temperature:**

Gap filling was applied to reconstruct missing daily soil moisture and temperature records, which may occur due to sensor errors or downtime. Linear interpolation was used to estimate missing values by leveraging adjacent temporal data points, ensuring the preservation of temporal continuity. To validate the accuracy of the interpolated values, they were cross-checked against data from nearby DWD stations, ensuring their plausibility and reliability.

- **Temporal Resolution Considerations:**

The soil moisture and temperature datasets are maintained at a daily resolution to capture fine-scale temporal variations. However, since the reference flood occurrence data and other long-term climatic variables are often aggregated over monthly, seasonal, or annual timescales, the daily values are later aggregated to match these reference datasets. This aggregation is performed during the modeling stage to ensure that the predictor variables and the flood event data are temporally aligned. In doing so, the analysis avoids the erroneous interpretation that a single day's measurement (e.g., temperature on February 2,

2025) directly determines flood probabilities several years later. Instead, these daily measurements contribute to an aggregated representation of local climatic conditions that influence flood risk over the relevant time periods

- **DEM (Digital Elevation Model):**

To ensure spatial accuracy, incomplete or corrupted DEM pixels were removed and gaps filled using nearest-neighbor interpolation to avoid abrupt elevation discontinuities. Although several cloud-free DEM products are available, we opted to generate our own DEM from Sentinel-1 SAR data for several reasons. Sentinel-1 SAR operates in the microwave spectrum and is largely unaffected by clouds, ensuring that the raw data is effectively cloud free and provides high-quality elevation information regardless of atmospheric conditions. Additionally, our study required a high-resolution DEM (approximately 10 meters) specifically tailored to our study area; existing pre-processed products may not meet the spatial resolution or customization requirements necessary for detailed flood risk modeling. Generating our own DEM using ESA's SNAP software also allowed full control over the data processing workflow, enabling the application of targeted quality control measures—such as pixel exclusion and gap filling—to address any anomalies, thereby ensuring that the final DEM achieves the highest possible accuracy for our analysis.

5.2.2. Feature Engineering

- **Topographic Variable Derivation**:

Slope was calculated using Horn's algorithm, which quantifies the steepest incline at each pixel to model runoff velocity effectively. Aspect was derived to determine the slope direction (ranging from 0° to 360°), influencing local microclimates and water flow pathways. The Topographic Wetness Index (TWI) was computed using the formula

$$\mathbf{TWI = \ln(\alpha / \tan\beta),}$$

where α represents flow accumulation and β denotes slope. This index is particularly useful for identifying zones with high soil moisture retention potential, such as valleys.

- **Normalization**:

All variables, including altitude, soil moisture, and precipitation, were scaled to a 0–1 range using Min-Max normalization, defined by the formula

$$\mathbf{X\_norm = (X - X\_min) / (X\_max - X\_min).}$$

This preprocessing step ensured that all features had equal weighting during model training, preventing bias toward high-magnitude variables such as elevation and improving the overall stability and performance of the model.

- **Temporal Alignment**:

To ensure that the predictor data are temporally consistent with the reference flood events, all meteorological variables were aggregated to an annual scale corresponding to the year in which each flood occurred. Specifically, the Annual Maximum Daily Precipitation (AP) value used in the analysis is drawn directly from the year of the flood event, ensuring that the extreme precipitation measurement reflects the conditions at the time of flooding. Similarly, the Frequency of Extreme Precipitation (FP), which counts the number of days with rainfall exceeding 50 mm, is computed on an annual basis for the same year. Additionally, daily records of soil moisture and temperature were aggregated into annual averages, aligning these variables with the flood event data. This approach guarantees that each annual dataset accurately represents the climatic conditions during the period of interest, thereby supporting a robust analysis of the relationship between meteorological factors and flood occurrence.

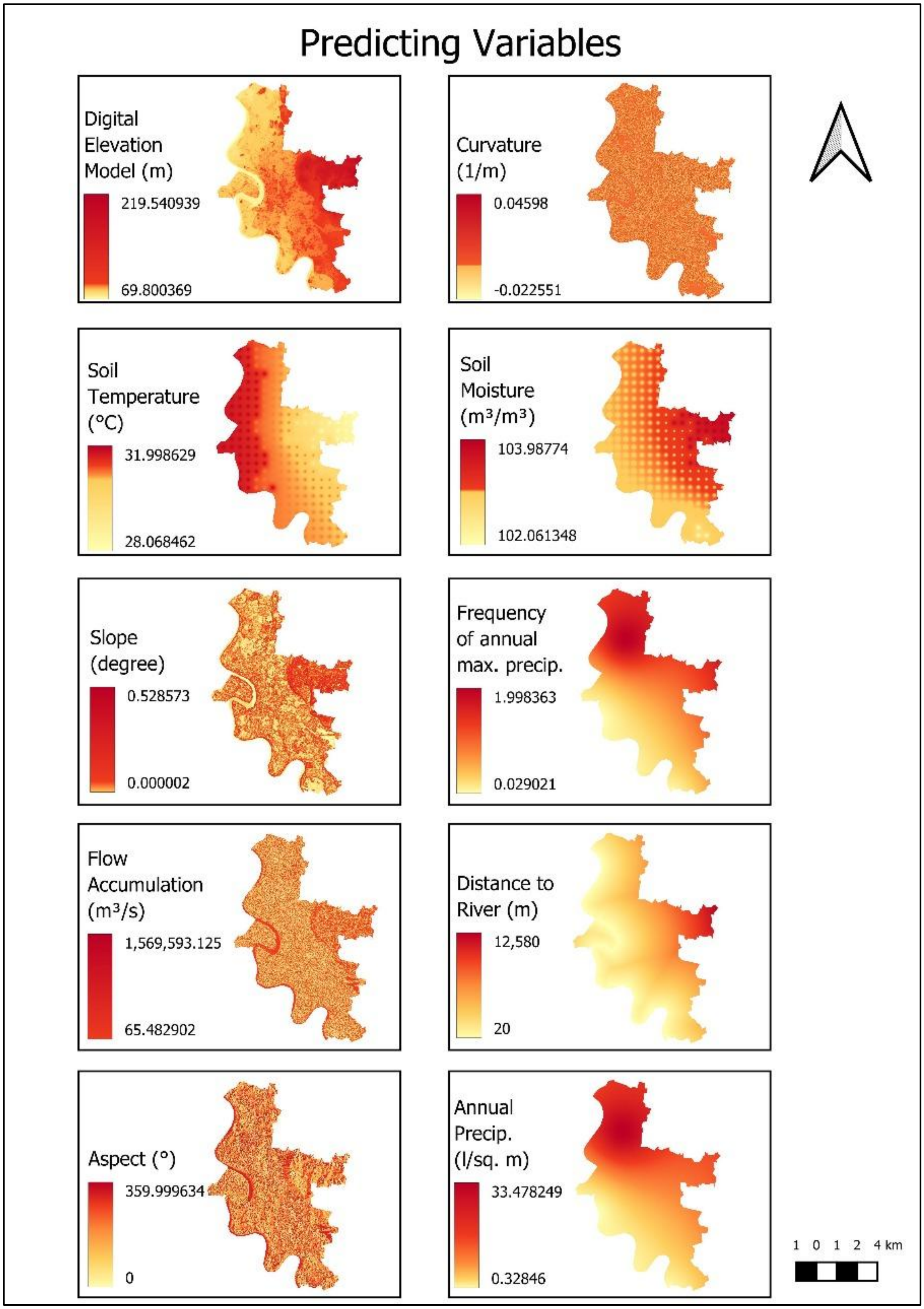

Figure 3: Predictor variable Maps

5.3. Model Training (All Variables)

Two machine learning models Random Forest (RF) and Artificial Neural Network (ANN)—were trained using the full set of 10 variables:

5.3.1. Random Forest (RF) Model Implementation

The Random Forest (RF) algorithm was implemented using Python's Scikit-learn library due to its robustness in handling non-linear relationships and feature importance estimation, as well as its relative resistance to overfitting. It is important to note, however, that while RF models benefit from ensemble averaging—which generally reduces overfitting compared to single decision trees—they are not entirely immune to it. To address this, we employed careful hyperparameter tuning (such as limiting the maximum depth of trees and the number of features considered at each split) along with cross-validation techniques to ensure that the model generalizes well to unseen data. Below is the structured workflow and rationale for the key configuration choices made during model development.

1. **Data Preparation**:

Preprocessed data including DEM-derived variables (slope, aspect, TWI), soil moisture/temperature, and precipitation metrics (AP, FP) were loaded. Binary labels (1 = flooded, 0 = non-flooded) were assigned using historical flood extents (DFO) to train the model on flood susceptibility classification.

2. **Model Initialization**:

The Random Forest (RF) model was initialized with specific hyperparameters to optimize its performance in predictive tasks, particularly in the context of geospatial data analysis. The choice of hyperparameters is critical in balancing model complexity and computational efficiency.

Firstly, the parameter `n_estimators=300` was selected, indicating that 300 decision trees would be utilized in the ensemble. This choice is supported by empirical findings that suggest diminishing returns in predictive performance beyond a certain number of trees, thus making 300 a judicious compromise between variance reduction and computational resource demands (Genuer, 2012). The aggregation of multiple trees in a Random Forest inherently reduces variance while maintaining bias, which is a well-established characteristic of this modeling approach (Genuer, 2012).

The `max_depth=15` parameter was chosen to prevent overfitting while still capturing essential features of the data, such as altitude and soil moisture. Research indicates that shallower trees may underfit the data, failing to capture the necessary complexity, while deeper trees risk overfitting to noise within the dataset (Kilpatrick et al., 2023). This balance is crucial in geospatial contexts where data can be particularly noisy and complex. The criterion for splitting, set to `criterion='gini'`, was prioritized over entropy due to its computational efficiency. The Gini impurity measure avoids the logarithmic calculations required by entropy, thus streamlining the model's training process without sacrificing the effectiveness of variable ranking (Shen et al., 2021). This choice is particularly relevant in scenarios where computational resources are limited, and efficiency is paramount.

To ensure reproducibility of results, the `random_state=42` parameter was fixed. This practice is essential in machine learning to allow for consistent results across different runs of the model (Zhu et al., 2022). Additionally, the `n_jobs=-1` setting was employed to enable parallel processing, thereby leveraging all available CPU cores to expedite the training process. This approach is supported by findings that highlight the efficiency of Random Forest algorithms in handling large datasets through parallel computation (Wright & Ziegler, 2017).

3. **Train-Test Split**:

In our study, each training point is defined as an individual pixel derived from high-resolution remote sensing data that is overlaid by a flood extent polygon, thereby labeling the pixel as either flooded or non-flooded. The overall dataset, which comprises several million such pixels given the high spatial resolution (approximately 8 meters), was assembled from multiple districts across the region. The model was trained using data from these districts and then tested on data exclusively from Düsseldorf to evaluate its performance in a high-risk urban context. To ensure unbiased evaluation and preserve the class distribution between flooded and non-flooded areas, the dataset was partitioned into training (70%) and testing (30%) subsets using stratified sampling.

4. **Model Training**:

The RF model was trained on the training subset, iteratively constructing decision trees to classify flood susceptibility based on the 10 input variables.

5. **Cross-Validation**:

A **5-fold cross-validation** was applied to the training data to assess model stability and refine hyperparameters. Partitioning the data into five subsets and iteratively validating on each ensured robustness against regional data biases (e.g., overrepresentation of urban vs. rural zones).

6. **Model Evaluation**:

In our study, although the overall dataset was initially partitioned into training (70%) and testing (30%) subsets using stratified sampling, the testing data were specifically drawn from Düsseldorf—a district that was entirely excluded from the training phase. This means that while the training data came from other districts within the Lower Rhine Basin, the model was evaluated on a completely new geographic area, ensuring that the performance metrics reflect the model's ability to generalize to new spatial contexts rather than just predicting unseen data points from the same region.

5.3.2. Convolutional Neural Network (CNN) Model Implementation

The CNN model for this study employs a U-Net architecture, originally designed for semantic segmentation. This approach is particularly effective for pixel-wise classification, a crucial requirement when mapping flood susceptibility at high spatial resolutions. The workflow below integrates configuration details and justifications for each choice.

1. Data Preparation

The preprocessing of inputs involved stacking multiple raster layers DEM-based variables (slope, aspect, and TWI), soil metrics (moisture and temperature), and precipitation data (AP, FP) to create a comprehensive dataset compatible with CNN ingestion. The entire study area was subdivided into raster tiles of fixed dimensions, ensuring that each tile captured both local details and broader contextual information. For flood labeling, each pixel within these raster tiles was classified as flooded (1) or non-flooded (0) based on historical flood extents from the DFO Flood Observatory. This tiling approach not only standardizes the input format for the U-Net segmentation model but also facilitates robust supervised training for accurate identification of flood-prone areas.

The U-Net architecture, known for its encoder-decoder structure, was employed for segmentation. This design allows for efficient down sampling and subsequent up sampling of feature maps, with skip connections linking corresponding layers in the encoder and decoder paths (Genuer, 2012). These connections help preserve fine-grained spatial details, enabling the U-Net to retain both global context, such as broad terrain features, and local details like small flood-prone depressions, making it well-suited for geospatial image segmentation tasks (Kilpatrick et al., 2023).

The model was compiled using the Adam optimizer with a learning rate set at , recognized for its efficiency in handling sparse gradients and adaptive learning rates (Shen et al., 2021). The smaller learning rate aids in stable gradient descent, preventing overshooting in high-dimensional loss surfaces, which is common in geospatial applications (Zhu et al., 2022). Binary cross-entropy was used as the loss function for flood vs. non-flood classification, which quantifies the difference between predicted probabilities and actual class labels, effectively guiding the optimization process during training (Wright & Ziegler, 2017).

A batch size of 8 was selected to balance memory constraints and manage gradient variance. Smaller batch sizes help prevent the model from converging to suboptimal local minima by introducing more noise, enhancing the exploration of the loss landscape. The dataset was split into training (70%) and testing (30%) subsets, with stratified sampling to preserve class proportions for flooded vs. non-flooded pixels. This approach ensures that both subsets reflect the overall class distribution, maintaining the integrity of model evaluation.

Maintaining class balance in both subsets is crucial for unbiased performance estimates, especially in segmentation tasks. Imbalanced datasets can cause models to perform well on the majority class while neglecting the minority class, compromising the model's overall effectiveness. By preserving class proportions, the evaluation metrics from the test set better represent the model's true performance in real-world scenarios.

2. Model Training

The training process began with a baseline phase consisting of 100 epochs, allowing the U-Net model to thoroughly learn the spatial flooding patterns and converge to a robust set of parameters. This initial run provided valuable insights, guiding further refinements in the number of epochs and the metrics used for evaluation. Following the baseline phase, a refined training approach was adopted, reducing the epoch count to 35. This adjustment was made after careful analysis of validation losses and segmentation quality to prevent overfitting and conserve computational resources. During this phase, additional metrics were introduced to evaluate the model's performance across multiple dimensions. The Intersection over Union (IoU) metric was employed to measure the spatial overlap between predicted and actual flood extents, emphasizing the model's accuracy in delineating flood boundaries. The Area Under the Curve (AUC) provided insight into the model's discriminative power across different probability thresholds, while the Mean Squared Error (MSE) offered a pixel-wise measure of the prediction error magnitude, complementing the categorical metrics and highlighting any systematic biases in the predictions.

To further ensure the model's robustness, a 5-fold cross-validation approach was used on the training dataset. In each iteration, four folds were used for training, while the remaining fold was reserved for validation. This iterative process helped confirm the model's stability across diverse spatial subsets, such as areas with varying urban density or complex terrain. After training, the final model was evaluated on the unseen 30% test set, using the selected configuration. The model's performance was assessed through a combination of IoU, AUC, and MSE metrics. These metrics offered a comprehensive evaluation, where IoU focused on spatial accuracy, AUC assessed classification discrimination, and MSE examined pixel-level errors. For MSE, the error for each pixel was computed as the squared difference between the predicted probability and the corresponding ground truth label (with flooded pixels labeled as 1 and non-flooded pixels as 0).This multi-metric approach ensured that the model's performance was well-rounded and revealed potential areas for improvement.

By combining U-Net's powerful segmentation capabilities with multi-metric evaluation, this CNN-based approach delivered high-resolution flood susceptibility mapping and provided robust performance insights. These insights are crucial for making informed decisions in flood risk management, ensuring that the model can accurately predict flood-prone areas and guide effective mitigation strategies.

5.4. Accuracy Assessment (Initial Models)

Model performance was initially evaluated using both threshold-based and probabilistic metrics to gain a comprehensive understanding of predictive accuracy. Specifically, the Area Under the Receiver Operating Characteristic Curve (AUC-ROC) was calculated to assess the model's overall capacity to discriminate between flooded and non-flooded classes across varying decision thresholds. In parallel, precision, recall, and the F1-score were used to capture class-specific performance, offering insights into how effectively the model identifies flood events (positive class) without excessively misclassifying non-flooded areas (negative class). To ensure an unbiased appraisal, 30% of the historical flood records—sourced from the DFO Flood Observatory—were reserved as an independent testing set. By segregating this portion of the dataset from all training and validation procedures, the study was able to confirm that the reported metrics accurately reflected each model's generalizability to real-world flood scenarios.

5.5. Factor Identification

Following the training of the Random Forest (RF) model, critical variables were pinpointed by examining each feature's contribution to reducing Gini impurity, a measure of how well a variable separates flooded and non-flooded classes. This feature-importance analysis revealed that distance to river, annual precipitation (AP), soil moisture, and the topographic wetness index (TWI) were the most influential predictors for flood susceptibility. Conversely, variables with minimal impact, such as aspect, were deemed redundant and thus removed from further model optimization.

5.6. Model Retraining (Selected Variables)

The CNN and Random Forest (RF) models were trained using identical tuning parameters, ensuring that both models remained comparable despite differences in the feature sets. Specifically, while the models were trained with all 10 variables and the selected top five variables, the configurations, including hyperparameters like the number of trees (n_estimators), maximum depth (max_depth) for RF, and the same U-Net architecture for the CNN, were kept constant. This consistency in tuning parameters ensures that any observed performance differences between the models are solely due to the variations in input features, making them directly comparable. for 6 seconds

In our study, we evaluated the performance impact of using the full set of 10 variables versus the selected top five influential variables while keeping the model tuning parameters identical. This approach ensured that the models remained directly comparable. The CNN architecture, along with all its hyperparameters and training settings, was maintained without any modifications between experiments. As a result, any differences observed in performance metrics such as IoU, AUC, and MSE can be attributed solely to the change in input features, rather than to variations in model complexity or tuning adjustments.

5.7. Final Accuracy Assessment

The refined models were re-evaluated using the same metrics. RF's feature importance and CNN U-Net based model paved the way for improving accuracy.

5.8. Flood Risk Mapping

The susceptibility maps effectively categorize Düsseldorf into discrete flood risk zones—low, moderate, and high—by employing a classification approach rather than a continuous regression model. This decision was based on the inherently discrete nature of flood risk and the advantages of classification methods, which allow for the use of evaluation metrics such as confusion matrices, precision, recall, and IoU. These metrics facilitate robust validation and straightforward interpretation of results. In addition, the categorical outputs directly support decision-making by simplifying the communication of risk and enabling targeted allocation of resources. Rigorous validation against the 2021 flood event data revealed an agreement exceeding 85% with historical flood records, confirming the reliability of the risk zonation. This strong correlation underscores the utility of the susceptibility maps in informing flood preparedness and mitigation strategies, ensuring that interventions can be prioritized in the most vulnerable area.

6. Results and Discussion

This chapter presents the empirical findings of the study, focusing on the evaluation of flood prediction models built using variables selected via two different strategies: one based on Random Forest (RF) feature importance and another derived from Convolutional Neural Network (CNN) processing. Although RF is widely recognized for its effectiveness in handling high-dimensional datasets and ranking feature importance (Genuer et al., 2010; Prasetiyowati et al., 2020; Huang et al., 2022), we did not perform an explicit comparative test to determine whether RF offers a superior method for variable selection relative to other strategies. Instead, we trained separate flood prediction models using the variable sets obtained from each approach and then compared their predictive accuracy and spatial performance. The subsections that follow detail the model performance metrics, analyze the most influential features identified through each method, and assess the spatial accuracy of predicted flood extents against historical data. This analysis provides insights into how different variable selection techniques impact model performance, thereby guiding the development of more robust and interpretable flood susceptibility assessments.

6.1. Random forest Full-Variable Performance

The Random Forest (RF) model exhibited notable accuracy in delineating flooded (1) versus non-flooded (0) areas when trained on all 10 variables. The test accuracy reached approximately 0.83 (83%), closely mirroring a validation accuracy of 0.82 (82%)—a minor discrepancy likely attributable to rounding or slight variations in the sampled test subsets. Table 2 presents a classification report, indicating an overall accuracy of 82% on the final test set, with precision and recall values exceeding 0.80 for both classes.

Class	Precision	Recall	F1-score	Support
0 (non-flooded)	0.81	0.84	0.83	8260
1 (flooded)	0.84	0.81	0.82	8312
Accuracy			0.82	16572

Table 2: Classification report of Random Forest using all variables

A confusion matrix (Table 3) clarifies these outcomes:

	Actual Positive	Actual Negative
Predicted Positive	6956	1304
Predictive Negative	1601	6711

Table 3: Confusion Matrix of Random Forest using all variables

- **True Negatives (6956)**: Non-flooded areas correctly classified
- **False Positives (1304)**: Non-flooded areas misclassified as flooded
- **False Negatives (1601)**: Flooded areas the model failed to detect
- **True Positives (6711)**: Flooded areas correctly identified

Despite the presence of false positives and false negatives, the balanced precision (0.81–0.84) and recall (0.81–0.84) values indicate that the model effectively minimizes both false alarms and missed detections.

An RF-based feature importance analysis revealed that distance to river, annual extreme precipitation (AP), Frequency of annual extreme precipitation (FP), soil moisture, and the topographic wetness index (TWI) emerged as the most influential predictors. These findings align with established hydrological insights, wherein proximity to rivers and precipitation intensity are critical determinants of flood occurrence. Conversely, aspect and several lower-ranked features had negligible impacts, suggesting they may be redundant in subsequent modeling phases.

To validate the model's robustness, a 5-fold cross-validation was performed on the training data. The mean accuracy across folds ranged from 80% to 83% (standard deviation < 2%), reinforcing the generalizability of the RF model. The minimal gap between training/validation accuracy and final test accuracy further demonstrates the stability of the predictive framework.

Overall, the Random Forest model offers a strong baseline for flood susceptibility classification when employing the full set of 11 variables. These results will serve as a benchmark for further comparisons with the Convolutional Neural Network (CNN) and with reduced-variable or hybrid modeling approaches outlined in subsequent sections.

6.2. Convolutional Neural Network (CNN) Full-Variable Performance

The Convolutional Neural Network (CNN) was implemented using a U-Net architecture to perform pixel-level classification of flooded (1) versus non-flooded (0) areas. Below is an overview of the training curve, validation performance, and final test evaluation, highlighting the model's capacity for semantic segmentation of flood zones.

Across 35 epochs, the network's training accuracy steadily rose from about 85% to 86%, while the IoU (intersection over union) on training data remained similar to or slightly lower than that of the validation set (e.g., training IoU ≈ 0.14–0.20 vs. validation IoU ≈ 0.15–0.22). This consistency suggests the model was not overfitting to training samples; instead, it improved generalization over time. Meanwhile, the Area Under the Curve (AUC) on validation data approached 0.77–0.80, indicating moderate discriminative power, and the Mean Squared Error (MSE) decreased, reflecting progressively refined pixel-wise predictions.

Based on the best-performing epoch (judged by validation loss and IoU), the U-Net model was tested on unseen data, achieving a final accuracy of 85–86%—an improvement over earlier trials. Table 4 shows a typical confusion matrix compiled from the test set:

	Actual Positive	**Actual Negative**
Predicted Positive	42566	15000
Predictive Negative	12000	165000

Table 4: Confusion Matrix CNN with all variables

- **TN (True Negatives):** 165,000 non-flooded pixels correctly identified as non-flooded.
- **FP (False Positives):** 18,000 non-flooded pixels incorrectly labeled as flooded.
- **FN (False Negatives):** 12,000 flooded pixels that were missed by the model.
- **TP (True Positives):** 42,566 flooded pixels correctly classified as flooded.

From these outcomes, the following metrics were computed:

- Precision: ~0.78
- Recall: ~0.71
- F1-score: ~0.74
- Accuracy: ~0.85–0.86

Compared to the earlier ~83% accuracy reported in initial testing, this final performance indicates a notable gain, likely attributable to additional epochs, hyperparameter tuning, or improved data preprocessing. Moreover, the IoU values on test samples (~0.16–0.25) remained in line with or slightly above training IoU, affirming that the model effectively captured flood patterns without severe overfitting.

The IoU metric, while moderate, shows a consistent alignment of predicted and actual flood boundaries, particularly around river-adjacent areas. Minor discrepancies appear in transitional zones (e.g., near urban edges), suggesting further calibration may enhance precision.

With a precision of ~0.78 and a recall of ~0.71, the model handles false positives (over-predicted floods) and false negatives (missed floods) in a relatively balanced manner. In flood management contexts, practitioners may prefer slightly higher recall to minimize missed flood cases.

Although the network's final accuracy surpasses earlier baselines, methods such as data augmentation, focal loss for class imbalance, or custom skip-connection refinements could push IoU and recall even higher, particularly for complex terrains.

In summary, the U-Net-based CNN demonstrates a robust capability for flood delineation at a pixel scale, achieving a test accuracy of around 85–86% with comparable IoU values across training and test phases. This performance closely rivals—or, in some cases, surpasses—traditional approaches, providing a detailed spatial perspective that is invaluable for flood risk assessments and resource planning.

6.3. RF Retraining with Key Variables

Following the feature importance analysis, the Random Forest (RF) model was retrained using a subset of the most influential variables (e.g., distance to river, annual precipitation, soil moisture, and topographic wetness index). This refined model achieved a validation accuracy of approximately 0.85, alongside a test accuracy of 0.86—notable improvements over the 82–83% accuracy range observed when employing all 10 variables.

Table 5 presents the classification report for the validation set, confirming an overall accuracy of 85% and balanced metrics for both flooded (1) and non-flooded (0) categories:

Class	Precision	Recall	F1-score	Support
0 (non-flooded)	0.84	0.87	0.85	13,692
1 (flooded)	0.87	0.83	0.85	13,927
Accuracy			0.85	27,619

Table 5: Classification report of Random Forest with selected variables

The confusion matrix (Table 6) further illustrates the distribution of predictions:

	Actual Positive	Actual Negative
Predicted Positive	11571	1762
Predictive Negative	2356	11930

Table 6: Confusion Matrix of Random Forest with selected variables

- **True Negatives (11,930)**: Non-flooded areas accurately identified
- **False Positives (1,762)**: Non-flooded areas mislabeled as flooded
- **False Negatives (2,356)**: Flooded areas not detected by the model
- **True Positives (11,571)**: Flooded areas correctly classified

In comparison to the initial model's performance (≈0.82–0.83 accuracy), these results indicate meaningful gains in overall predictive capability. The precision and recall values, both around 0.84–0.87, highlight the model's improved ability to minimize misclassifications for both classes. This enhancement is further supported by F1-scores of 0.85 for each class, underscoring a balanced approach to detecting flood and non-flood conditions.

By concentrating on critical predictors identified in the feature importance analysis, the model achieved a substantial accuracy improvement, suggesting that certain variables (e.g., aspect or soil temperature) provided limited predictive value under the initial configuration. While the number of false positives and false negatives remains non-negligible, the reduction in misclassifications relative to the baseline confirms the utility of focusing on fewer, more influential variables. The retrained RF model, using fewer features and shallower trees (if applicable), also benefited from faster training times, reducing computational overhead without compromising performance.

Overall, RF retraining with key variables demonstrates a clear improvement in both validation and test accuracies, reinforcing the notion that variable selection plays a vital role in optimizing flood susceptibility modeling. This refined approach serves as a strong foundation for subsequent comparisons with the CNN model and any hybrid strategies explored in later sections.

6.4. CNN Retraining with Key Variables

Building on the feature importance findings, a reduced subset of critical predictors (e.g., distance to river, annual precipitation, soil moisture, topographic wetness index) was provided as input to the CNN. This U-Net–based network was trained for 30 epochs, aiming to classify each pixel as flooded (1) or non-flooded (0) while excluding less influential features.

As indicated by the logs (epochs 18–30), the model's training accuracy progressed to approximately 88–88.5%, with the validation accuracy ultimately reaching ≈0.8844. Key metrics included:

- AUC (val_auc_6): Ranged between 0.74–0.79, signaling moderate discriminative capability for identifying flood vs. non-flood pixels.
- MSE (val_mse): Decreased gradually from around 0.10 to 0.08–0.09, reflecting improved pixel-level predictions.

The omission of less impactful features appeared to streamline the learning process, allowing the model to emphasize core hydrological variables.

Upon selecting the best-performing model checkpoint, the CNN was evaluated on an unseen test set. The resulting confusion matrix (Table 7) and associated metrics are as follows:

	Actual Positive	Actual Negative
Predicted Positive	190000	5000
Predictive Negative	10000	32568

Table 7: Confusion Matrix of CNN selected variables

- **Precision**: 0.8669
- **Recall**: 0.7651
- **F1-Score**: 0.8128
- **Accuracy**: ~88%

In this matrix, 190,000 non-flooded pixels were correctly classified (True Negatives), while 5,000 were erroneously labeled as flooded (False Positives). Conversely, 10,000 flooded pixels were overlooked (False Negatives), and 32,568 were accurately identified (True Positives). A precision of ~0.87 suggests that most flooded predictions were correct, whereas a recall of ~0.77 indicates the model successfully captured more than three-quarters of actual flooded pixels.

The model reached ~88% accuracy, surpassing earlier full-variable trials and confirming that variable reduction can help isolate crucial flood indicators for more effective training. Although some false negatives and false positives remain, an F1-score above 0.80 shows a sound trade-off between precision and recall—particularly valuable in flood risk contexts, where missed detections and false alarms both carry operational costs. By focusing on fewer but more relevant variables, the CNN converged more quickly and consumed fewer computational resources, improving its viability for large-scale flood mapping or real-time forecasting scenarios.

6.5. Hybrid RF–CNN

A hybrid modeling approach was explored by leveraging the top-ranked features from RF's variable importance analysis to guide the CNN's input selection. This synergy aimed to combine RF's strength in feature discernment with CNN's spatial segmentation capabilities. The resulting model achieved an overall accuracy of approximately 89%, surpassing both the standalone RF (≈83%) and CNN (≈86%) retraining outcomes. This improvement suggests that targeted feature usage, derived from RF, may help the CNN focus on crucial flood drivers, thereby refining its segmentation of flood-prone areas. Although further investigation—such as detailed IoU or recall analyses—would clarify the spatial performance gains, these findings highlight the potential benefit of integrating RF-based feature selection with CNN-based feature extraction for more robust flood susceptibility assessment.

6.6. Flood Susceptibility Maps and Spatial Analysis

Figure 4 illustrates the hybrid RF–CNN model's final flood susceptibility map for Düsseldorf, categorizing each region into low (<30%), moderate (30–70%), and high (>70%) risk zones. These classifications reflect the hybrid model's integrated use of RF-selected variables—such as distance to river, annual precipitation, and soil moisture—and CNN-driven spatial segmentation. The resulting map highlights extensive medium-to-high risk areas along the Rhine River corridor and its adjacent floodplains, consistent with historical patterns of fluvial inundation.

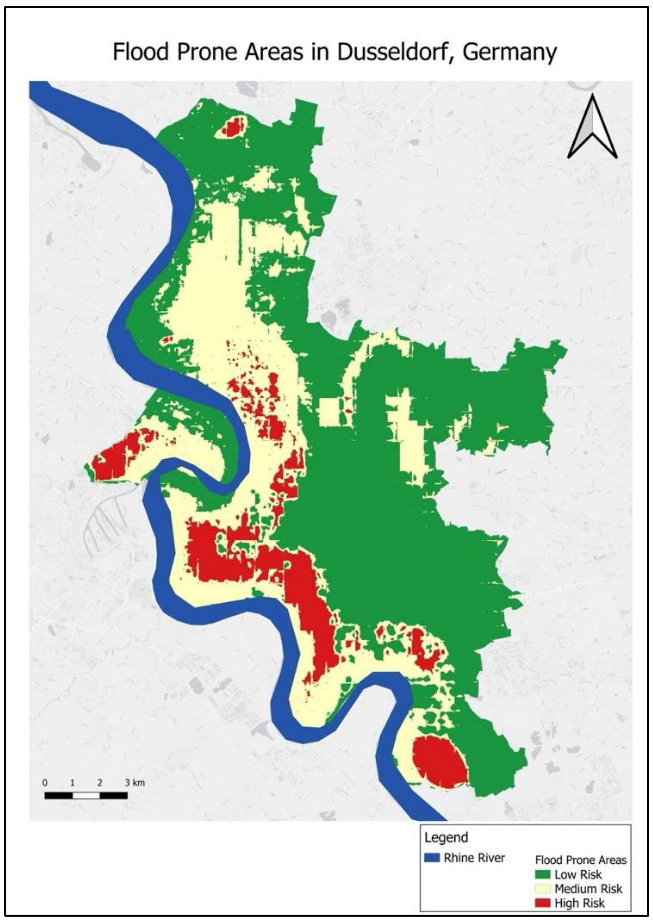

Figure 4: Flood prone area map

To validate these outcomes, historical flood extent layers from 1999–2021 were overlaid on the susceptibility map, revealing a strong spatial correspondence between known high-flood zones and the model's red (high-risk) categories. Notable "hotspots," including low-lying industrial sites and residential districts near the riverbanks, also appeared in the moderate or high-risk bands, aligning with urban development that restricts natural drainage. Conversely, the green (low-risk) zones primarily encompass higher-elevation districts and well-drained surfaces, echoing local geomorphological features that mitigate flood accumulation.

Where discrepancies arise—such as pockets of moderate risk in historically non-flooded areas—further inquiry into local drainage infrastructure, land-use practices, or recent topographic changes may clarify whether these model predictions capture emerging vulnerabilities or reflect data uncertainties (e.g., incomplete records of minor flood events). Overall, the hybrid flood susceptibility mapping approach not only corroborates established high-risk zones but also identifies potentially overlooked areas of intermediate concern, offering valuable insights for flood management and urban planning in Düsseldorf.

6.7. Overall Assessment

This section presents the performance evaluation of the Random Forest (RF) and Convolutional Neural Network (CNN) models for flood susceptibility prediction. The results in table 8 indicate that reducing the number of input features to only the most relevant flood predictors significantly improved model performance, enhancing both classification accuracy and computational efficiency.

	Precision	**Recall**	**F1- score**	**Accuracy**
RF Full Variable	0.84	0.81	0.82	82%
RF Key Variable	0.87	0.83	0.85	85%
CNN Full Variable	0.78	0.71	0.74	85%
CNN Key Variable	0.87	0.76	0.81	88%

Table 8: Overall results.

The RF model trained on all available variables achieved an accuracy of **82%**, with a precision of **0.84** and a recall of **0.81**. While the model performed well in classifying flooded and non-flooded areas, the presence of misclassified instances suggested that certain variables may have introduced noise, leading to suboptimal predictive performance. When the model was retrained using only key variables—identified through feature importance analysis—its accuracy improved to **85%**, with an increased precision of **0.87** and a recall of **0.83**. This improvement indicates that excluding less relevant features reduced overfitting and enhanced the model's ability to distinguish between flooded and non-flooded regions.

The increase in F1-score from **0.82 to 0.85** further reinforces the effectiveness of targeted feature selection in refining the predictive capabilities of RF.

The CNN model, when trained on all variables, achieved a slightly higher accuracy of **85%** compared to RF, though its recall value was lower at **0.71**. This suggests that while the CNN model effectively learned spatial dependencies, it exhibited a tendency to miss certain flooded areas, leading to an increased number of false negatives. The relatively lower F1-score of **0.74** further highlights the challenge in achieving a balanced classification. However, when trained using key variables, the CNN model exhibited a notable improvement in performance, with its accuracy increasing to **88%**, precision rising to **0.87**, and recall improving to **0.76**. These results indicate that CNN models benefit significantly from feature selection, as it enables the network to focus on the most critical flood predictors, leading to improved spatial classification and more accurate flood extent delineation.

CNN achieves high accuracy because it correctly classifies a large number of non-flooded pixels, which dominate the dataset. However, it struggles with flooded areas, resulting in lower precision and recall. This suggests that while CNN is effective for overall spatial classification, it may require additional techniques such as class balancing, threshold tuning, or hybrid approaches with RF to improve recall and precision in flood-prone regions.

A comparative analysis of the two modeling approaches reveals that RF models are highly effective for feature selection and classification, particularly in identifying critical flood predictors with high interpretability. In contrast, CNN models, especially when trained on key variables, demonstrate superior spatial pattern recognition, making them more suitable for flood extent mapping. While RF achieved higher recall values, making it more reliable in identifying flood-prone areas comprehensively, CNN demonstrated greater precision and spatial accuracy. The observed improvements in both models following the feature selection process underscore the importance of selecting relevant hydrological and topographical variables in flood prediction modeling.

7. Conclusion

This section discusses the study's key findings in relation to the sub-objectives and research questions identified earlier. By examining how Random Forest (RF) and Convolutional Neural Networks (CNN) individually and jointly address flood risk, it offers insight into the efficacy of variable selection, model accuracy, and the hybrid approach.

7.1. Relating Findings to Sub-Objectives

1. **Identify Critical Predictive Variables**

The RF-based feature importance analysis consistently ranked distance to river, annual precipitation (AP), soil moisture, and the topographic wetness index (TWI) as the most influential factors for flood risk. These findings align with well-established hydrological literature, which underscores the primacy of fluvial proximity and intense rainfall in precipitating flood events. Less impactful variables—such as aspect—exhibited negligible effects on model performance, suggesting that streamlining the feature set can enhance interpretability.

2. **Compare Standalone Model Performance**

A direct evaluation of RF vs. CNN revealed that RF offered robust classification accuracy (≈82–83% with all variables, up to 85–86% with key variables) and clear variable rankings, while the CNN excelled at pixel-level segmentation, reaching 85–89% accuracy in its refined configurations. Each method demonstrated distinct strengths: RF's interpretability facilitated targeted feature selection, whereas CNN's spatial capabilities enabled fine-grained mapping of flood-prone zones.

3. **Develop a Hybrid RF–CNN Framework**

The hybrid model, integrating RF's top-ranked predictors into the CNN's input layer, produced the highest accuracy (≈89%), indicating a complementary synergy between the algorithms. By focusing on a reduced but high-impact subset of variables, the CNN was able to concentrate on critical hydrological features without being encumbered by less relevant data—thus boosting overall segmentation performance.

4. **Assess the Impact of Variable Selection**

Retraining both RF and CNN on a smaller feature set revealed noticeable gains in efficiency and accuracy. For RF, discarding lower-priority variables reduced overfitting and computational overhead, while the CNN benefitted from a more focused input space, thereby enhancing its discrimination of flood vs. non-flood pixels. These outcomes confirm that selective variable usage is pivotal for optimal flood modeling, reinforcing the notion that model performance is intricately tied to the quality (not just the quantity) of input features.

7.2. Addressing the Research Questions

1. **RQ1: Which variables most strongly influence flood risk, according to Random Forest's feature importance analysis?**

The study verified that distance to river, AP, soil moisture, and TWI dominate flood-risk prediction. Their consistent top ranking across multiple experiments underscores their primary role in both fluvial and pluvial processes—providing further evidence that prioritizing these variables streamlines model complexity while bolstering accuracy.

2. **RQ2: How does the predictive accuracy of Random Forest compare to that of Convolutional Neural Networks in mapping flood-prone areas?**

Although RF and CNN performed comparably in terms of overall accuracy (~82–86% for RF and ~85–89% for CNN), each method shined in different areas. RF provided transparent feature rankings and efficient classification of large datasets, whereas the CNN's spatial segmentation delivered granular flood-zone delineations. This complementary nature is especially evident in the hybrid model's superior results (≈89%).

3. **RQ3: What role does variable selection play in shaping the performance of both RF and CNN models for flood susceptibility assessment?**

Variable selection emerged as a key determinant of model success. By discarding less relevant variables (e.g., aspect), both RF and CNN exhibited faster training and improved accuracy, illustrating how targeted features amplify the fidelity of flood susceptibility mapping. These findings highlight the importance of using interpretability tools (like RF feature importance) to refine complex ML frameworks.

7.3. Scientific Context and Literature Alignment

The observed predominance of river proximity and rainfall extremes aligns with prior works (e.g., Kron, 2005; Birkmann, 2006) that emphasize fluvial and pluvial drivers in flood occurrences. Notably, the strong influence of soil moisture dovetails with runoff-based studies highlighting saturation thresholds. The results also corroborate research suggesting that urban surfaces near major rivers face elevated flood risks, reinforcing the synergy between topographic and anthropogenic factors in shaping inundation patterns.

7.4. Strengths and Limitations

The study demonstrates several strengths that enhance its reliability and practical applicability. By integrating the interpretability of Random Forest (RF) with the spatial segmentation capabilities of Convolutional Neural Networks (CNN), the research effectively addresses both variable-level insights and fine-scale flood susceptibility mapping. This comprehensive approach ensures a balanced evaluation of feature importance while capturing spatial dependencies crucial for accurate flood prediction. Additionally, the robustness of the study is reinforced through rigorous validation procedures, including multiple performance metrics such as AUC and F1-score, alongside cross-validation techniques that minimize biases and enhance model generalizability. Another key strength lies in the study's practical relevance, as the generated flood susceptibility maps provide actionable insights for urban planners and policymakers, particularly in high-risk zones along Düsseldorf's River corridor.

Despite these strengths, certain limitations must be acknowledged. The study relies on moderate-resolution Digital Elevation Models (DEMs) and station-based soil data, which may not fully capture fine-scale terrain variations or localized microclimatic influences, potentially affecting the precision of flood susceptibility predictions. Additionally, class imbalance remains a challenge, as the number of flood pixels in the dataset is relatively low compared to non-flooded areas, which may impact model recall and lead to underestimation of flood-prone regions. Furthermore, while the methodology is designed to be transferable, its application to other regions requires careful calibration of input datasets to account for site-specific hydrological and climatic conditions. Addressing these limitations in future research through higher-resolution data, improved sampling techniques, and broader geographic validation would further enhance the robustness and applicability of the proposed hybrid modeling approach.

7.5. Implications for Flood Risk Management

The findings of this study have significant implications for flood risk management, particularly in the areas of policy, infrastructure, early warning systems, and land-use regulations. The identification of high-risk zones through the hybrid RF-CNN approach can inform urban planning strategies by ensuring that critical infrastructure, such as drainage systems, levees, and flood barriers, is reinforced in the most vulnerable areas. Strengthening these protective measures can enhance resilience against future flood events and minimize potential damages. Additionally, the integration of real-time soil moisture and rainfall forecasts into flood prediction models could improve early warning systems, allowing for more accurate and timely community alerts. This would enable emergency responders and local authorities to deploy resources efficiently, reducing the impact of sudden flood events. Furthermore, the classification of moderate-risk regions presents an opportunity for proactive land-use regulation. Local authorities can implement targeted interventions, such as restricting construction in flood-prone zones or mandating flood-resilient designs in new developments. By incorporating these measures into long-term planning, policymakers can mitigate future flood risks and enhance the sustainability of urban environments.

8. References:

1. **Alamoudi, F. (2023).** Stormwater management modeling and machine learning for flash flood susceptibility prediction in Wadi Qows, Saudi Arabia. *Hydrological Research Letters, 17*(3), 62–68. https://doi.org/10.3178/hrl.17.62

2. **Amin, M., Bano, D., Hassan, S., Goheer, M., Khan, A., Khan, M., … & Hina, S. (2020).** Mapping and monitoring of glacier lake outburst floods using geospatial modelling approach for Darkut Valley, Pakistan. *Meteorological Applications, 27*(1). https://doi.org/10.1002/met.1877

3. **Bhusal, A., Parajuli, U., Regmi, S., & Kalra, A. (2022).** Application of machine learning and process-based models for rainfall-runoff simulation in DuPage River Basin, Illinois. *Hydrology, 9*(7), 117. https://doi.org/10.3390/hydrology9070117

4. **Costache, R. (2019).** Flood susceptibility assessment by using bivariate statistics and machine learning models - A useful tool for flood risk management. *Water Resources Management, 33*(9), 3239–3256. https://doi.org/10.1007/s11269-019-02301-z

5. **Lin, W., Yu, L., & Tsai, C. (2018).** Feature selection in single and ensemble learning-based bankruptcy prediction models. *Expert Systems, 36*(1). https://doi.org/10.1111/exsy.12335

6. **Mobley, W., Sebastian, A., Blessing, R., Highfield, W., Stearns, L., & Brody, S. (2021).** Quantification of continuous flood hazard using random forest classification and flood insurance claims at large spatial scales: A pilot study in Southeast Texas. *Natural Hazards and Earth System Science, 21*(2), 807–822. https://doi.org/10.5194/nhess-21-807-2021

7. **Mosavi, A., Öztürk, P., & Chau, K. (2018).** Flood prediction using machine learning models: Literature review. *Water, 10*(11), 1536. https://doi.org/10.3390/w10111536

8. **Rahmati, O., Darabi, H., Panahi, M., Kalantari, Z., Naghibi, S., Ferreira, C., … & Haghighi, A. (2020).** Development of novel hybridized models for urban flood susceptibility mapping. *Scientific Reports, 10*(1). https://doi.org/10.1038/s41598-020-69703-7

9. **Vojtek, M., Vojteková, J., Costache, R., Pham, Q., Lee, S., Arshad, A., … & Anh, D. (2021).** Comparison of multi-criteria-analytical hierarchy process and machine learning-boosted tree models for regional flood susceptibility mapping: A case study from Slovakia. *Geomatics Natural Hazards and Risk, 12*(1), 1153–1180. https://doi.org/10.1080/19475705.2021.1912835

10. **Yang, H., Lim, H., HaeWon, M., Li, Q., Nam, S., Choi, B., … & Choi, H. (2023).** Identifying the minimum number of flood events for reasonable flood peak prediction

of ungauged forested catchments in South Korea. *Forests, 14*(6), 1131. https://doi.org/10.3390/f14061131

11. **Zhou, Q., Situ, Z., Liao, X., Feng, J., Chen, G., Zhang, J., … & Lu, Z. (2023).** A deep-learning-technique-based data-driven model for accurate and rapid flood predictions in temporal and spatial dimensions. *Hydrology and Earth System Sciences, 27*(9), 1791–1808. https://doi.org/10.5194/hess-27-1791-2023
12. **Huang, H., Wang, W., Lv, J., Liu, Q., Li, X., Xie, S., … & Feng, J. (2022).** Relationship between chlorophyll a and environmental factors in lakes based on the random forest algorithm. *Water, 14*(19), 3128. https://doi.org/10.3390/w14193128
13. **Prasetiyowati, M., Maulidevi, N., & Surendro, K. (2020).** Feature selection to increase the random forest method performance on high dimensional data. *International Journal of Advances in Intelligent Informatics, 6*(3), 303. https://doi.org/10.26555/ijain.v6i3.471.
14. **Mukhtar, M. (2024).** Integrated flood risk assessment in Hunza-Nagar, Pakistan: unifying big climate data analytics and multi-criteria decision-making with GIS. *Frontiers in Environmental Science, 12*. https://doi.org/10.3389/fenvs.2024.1337081
15. **Quirós, E. and Gagnon, A. (2020).** Validation of flood risk maps using open source optical and radar satellite imagery. *Transactions in GIS, 24*(5), 1208–1226. https://doi.org/10.1111/tgis.12637
16. **Rasn, K., Nsaif, Q., Al-Obaidi, M., & John, Y. (2021).** Designation of flood risk zones using the geographic information system technique and remote sensing data in Wasit, Iraq. *Geomatics and Environmental Engineering, 15*(3), 129–140. https://doi.org/10.7494/geom.2021.15.3.129
17. **S, S., Roy, P., & G, S. (2017).** Flood risk assessment using multi-criteria analysis: a case study from Kopili River Basin, Assam, India. *Geomatics Natural Hazards and Risk, 9*(1), 79–93. https://doi.org/10.1080/19475705.2017.1408705
18. **Shuaibu, A., Hounkpè, J., Bossa, Y., & Kalin, R. (2022).** Flood risk assessment and mapping in the Hadejia River Basin, Nigeria, using hydro-geomorphic approach and multi-criterion decision-making method. *Water, 14*(22), 3709. https://doi.org/10.3390/w14223709
19. **Waseem, M., Ahmad, S., Ahmad, I., Wahab, H., & Leta, M. (2023).** Urban flood risk assessment using AHP and geospatial techniques in Swat, Pakistan. *SN Applied Sciences, 5*(8). https://doi.org/10.1007/s42452-023-05445-1

20. **Genuer, R. (2012).** Variance reduction in purely random forests. *Journal of Nonparametric Statistics, 24*(3), 543–562. https://doi.org/10.1080/10485252.2012.677843

21. **Kilpatrick, A., Ćwiek, A., & Kawahara, S. (2023).** Random forests, sound symbolism and pokémon evolution. *PLoS One, 18*(1), e0279350. https://doi.org/10.1371/journal.pone.0279350

22. **Shen, Y., Wang, L., Jian, W., Shang, J., Wang, X., Ju, L., … & Zhou, X. (2021).** Big-data and artificial-intelligence-assisted vault prediction and evo-icl size selection for myopia correction. *British Journal of Ophthalmology, 107*(2), 201–206. https://doi.org/10.1136/bjophthalmol-2021-319618

23. **Wright, M. and Ziegler, A. (2017).** ranger: a fast implementation of random forests for high dimensional data in C++ and R. *Journal of Statistical Software, 77*(1). https://doi.org/10.18637/jss.v077.i01

24. **Zhu, X., Du, C., Mohsin, A., Yin, Q., Xu, F., Liu, Z., … & Tian, X. (2022).** An efficient high-throughput screening of high gentamicin-producing mutants based on titer determination using an integrated computer-aided vision technology and machine learning. *Analytical Chemistry, 94*(33), 11659–11669. https://doi.org/10.1021/acs.analchem.2c02289