RÉPUBLIQUE ALGÉRIENNE DÉMOCRATIQUE ET POPULAIRE
MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR
ET DE LA RECHERCHE SCIENTIFIQUE

المدرسة الوطنية المتعددة التقنيات
Ecole Nationale Polytechnique

LRS EAU
ENP - ALGERIE

ÉCOLE NATIONALE POLYTECHNIQUE
DÉPARTEMENT HYDRAULIQUE
LABORATOIRE DE RECHERCHE DES SCIENCES DE L'EAU

FINAL PROJECT THESIS
PRESENTED WITH A VIEW TO OBTAIN THE DIPLOMA OF
STATE ENGINEER IN HYDRAULICS

Mapping Flood Susceptibility Areas And Assessing Influential Factors -Case Of The Chellif Basin-

Prensted by:
Mr. **CHAGROUNE** Abdessamed & Mr. **HALFAOUI** Mustapha
Presented and publicly defended on 06/07/2023

JURY COMPOSITION :

President	Mrs.TCHEKIKEN Chahinez	MCB	ENP
promoter	Mr.TACHI Salah Eddine	MCA	ENP
Co-promoter	Mr.HASNAOUI Yacine	PHD Student	ENP
Examiner	Mr.BENZIADA Salim	MAA	ENP

ENP - 2023

RÉPUBLIQUE ALGÉRIENNE DÉMOCRATIQUE ET POPULAIRE
MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR
ET DE LA RECHERCHE SCIENTIFIQUE

المدرسة الوطنية المتعددة التقنيات
Ecole Nationale Polytechnique

LRS EAU
ENP - ALGERIE

ÉCOLE NATIONALE POLYTECHNIQUE
DÉPARTEMENT HYDRAULIQUE
LABORATOIRE DE RECHERCHE DES SCIENCES DE L'EAU

FINAL PROJECT THESIS
PRESENTED WITH A VIEW TO OBTAIN THE DIPLOMA OF
STATE ENGINEER IN HYDRAULICS

Mapping Flood Susceptibility Areas And Assessing Influential Factors -Case Of The Chellif Basin-

Prensted by:
Mr. **CHAGROUNE** Abdessamed & Mr. **HALFAOUI** Mustapha
Presented and publicly defended on 06/07/2023

JURY COMPOSITION :

President	Mrs.TCHEKIKEN Chahinez	MCB	ENP
promoter	Mr.TACHI Salah Eddine	MCA	ENP
Co-promoter	Mr.HASNAOUI Yacine	PHD Student	ENP
Examiner	Mr.BENZIADA Salim	MAA	ENP

ENP - 2023

RÉPUBLIQUE ALGÉRIENNE DÉMOCRATIQUE ET POPULAIRE
MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR
ET DE LA RECHERCHE SCIENTIFIQUE

المدرسة الوطنية المتعددة التقنيات
Ecole Nationale Polytechnique

LRS EAU
ENP - ALGERIE

ÉCOLE NATIONALE POLYTECHNIQUE
DÉPARTEMENT HYDRAULIQUE
LABORATOIRE DE RECHERCHE DES SCIENCES DE L'EAU

MÉMOIRE DE PROJET DE FIN D'ÉTUDES
POUR L'OBTENTION DU DIPLÔME D'INGÉNIEUR D'ÉTAT EN HYDRAULIQUE

La Cartographie Des Zones De Susceptibilité Aux Inondations Et Évaluation Des Facteurs Influents -Cas Du Bassin Du Chellif-

Présenté par :
Mr. **CHAGROUNE** Abdessamed & Mr. **HALFAOUI** Mustapha
Présenté et soutenu publiquement le 06/07/2023

COMPOSITION DU JURY:

Présidente	Mrs.TCHEKIKEN Chahinez	MCB	ENP
promoteur	Mr.TACHI Salah Eddine	MCA	ENP
Co-promoteur	Mr.HASNAOUI Yacine	PHD Student	ENP
Examinateur	Mr.BENZIADA Salim	MAA	ENP

ENP - 2023

Acknowledgement

First and foremost, I would like to express my gratitude to God Almighty, who has given us the strength and patience to complete this humble work.

Our heartfelt thanks go to; our beloved parents for their support and patience.

Our thesis supervisors, Mr. TACHI Salah Eddine and Mr. HASNAOUI Yacine, for their invaluable support throughout our project.

We also thank the president and members of the jury who will have the honor of evaluating our work.

Our thanks also extend to all our teachers throughout the years of study. We appreciate every teacher who has aided us with their knowledge from the early stages of school until this moment.

We would like to extend our sincerest gratitude to all those who have contributed, directly or indirectly, to the development of this thesis. To all our close friends and the entire hydraulic class of 2023, we express our gratitude.

ملخص

تُعَتَبَرُ الفيضانات واحدة من أكثر الظواهر الكارثية تدميرًا. يُعَرَف عرض الفيضانات على أنه ميل لتعرُضْ الضرر الناجم عن هذه الظاهرة. ومع ذلك، فإن التنبؤ الدقيق بالفيضانات الفجائية لا يزال أمرًا صعبًا بسبب تعقيد الظاهرة. في هذه الدراسة، اعتمدنا نهجًا يستند إلى أنظمة المعلومات الجغرافية(SIG) وتقنيات الاستشعار عن بعد (RS) ونماذج التصنيف في التعلم الآلي مثل $LGBM$ و $AdaBoost$ وتقنية التعلم الآلي الجديدة المسماة بـ $Stacking$ ، لإنشاء خريطة عرض الفيضانات في حوض شليف. تم استخدام خمسة عشر عاملًا هيدرولوجيًا وتوبوغرافيًا كمدخلات لنماذج عرض الفيضانات. أظهرت النتائج أن نموذج الـ $Stacking$ كان الأكثر تحسينًا، بقيمة AUC تبلغ $0,99$ ، تليه $LGBM$ بقيمة 0.98 و $AdaBoost$ بقيمة 0.96 . تُستخدم نتائج هذه الدراسة للتخطيط وتنفيذ استراتيجيات التخفيف من الفيضانات في المنطقة.

الكلمات الرئيسية: الفيضانات، أنظمة المعلومات الجغرافية ،الاستشعار عن بعد، التعلم اآللي، $LGBM$ ، $AdaBoost$ ، $Stacking$ ، AUC

Résume

Les inondations sont considérées comme l'un des phénomènes catastrophiques les plus destructeurs. La susceptibilité aux inondations est définie comme la propension à subir des dommages causés par ce phénomène. Cependant, la prédiction précise des crues éclair reste difficile en raison de la complexité du phénomène. Dans cette étude, nous avons adopté une approche basée sur les systèmes d'information géographique (SIG), les techniques de télédétection (RS) et les modèles de classification de l'apprentissage automatique, tels que LGBM, AdaBoost et la nouvelle technique de l'apprentissage automatique appelée Stacking, afin de créer une carte de susceptibilité aux inondations dans le bassin versant de Chellif. Quinze facteurs hydrologiques et topographiques ont été utilisés comme entrées pour les modèles de susceptibilité aux inondations. Les résultats ont montré que Stacking était le modèle le plus optimal, avec une valeur AUC de 0,99, suivi de LGBM avec 0,98 et AdaBoost avec 0,96. Les résultats de cette étude sont utilisés pour la planification et la mise en œuvre de stratégies d'atténuation des inondations dans la région.

Mots clés : La susceptibilité aux inondations , crues , SIG , RS , apprentissage automatique ,LGBM, AdaBoost , Stacking , AUC

Abstract

Floods are considered one of the most destructive catastrophic phenomena.Flood susceptibility is defined as the tendency to suffer damage caused by this phenomenon. However, accurately predicting flash floods remains challenging due to the complexity of the phenomenon. In this study, we adopted an approach based on geographic information systems (GIS), remote sensing techniques (RS), and machine learning classification models such as LGBM, AdaBoost, and the new machine learning technique called Stacking, to create a flood susceptibility map in the Chellif watershed. Fifteen hydrological and topographic factors were used as inputs for the flood susceptibility models. The results showed that Stacking was the most optimal model, with an AUC value of 0.99, followed by LGBM with 0.98 and AdaBoost with 0.96. The findings of this study are used for planning and implementing flood mitigation strategies in the region.

Key words: Flood susceptibility, flash floods, GIS, RS, machine learning, LGBM, AdaBoost, Stacking, AUC

Contents

List of Figures

List of Tables

List of Abreviation

AdaBoost : Adaptive Boosting

ANRH : Agence Nationale des Ressources Hydraulique(french)

AGIRE : Agence Nationale de Gestion Integree des Ressources en Eau (french)

AUC : Area Under the Curve

DEM : digital elevation model

FN: False negative

FP :]False positive

GIS: :Geographic Information System

GPS: Global Positioning System

LiDar : Light Detection and Ranging

LightGBM : Light Gradient Boosting Machine

ML: Machine learning

NDVI : Normalized Difference Vegetation

NDWI : Normalized Difference Water Index

ROC : Receiver Operating Characteristic

SPI : Stream Power Index

TPR : True positive rate

TNR : True negative rate

TWI : Topographic wetness index

TP : True positive

TN : True negative

GENERAL INTRODUCTION

Floods are the result of a combination of various environmental, meteorological, hydrological, and geomorphological factors, as well as human intervention. The current global climate change is the primary cause behind the increasing frequency and magnitude of flood risks. Furthermore, large-scale human interventions in the environment, such as deforestation, unregulated construction, dam construction, and urbanization near rivers, are additional factors that can contribute to flooding, alongside climate change, as influential causes.

Studying floods on a large scale is challenging due to the influence of numerous local factors, including rainfall, land slope, river systems, and land cover. Assessing flood risks typically involves hydrological and hydraulic modeling, such as HEC-RAS modeling, to estimate flood intensity and extent for different return periods. However, these modeling techniques rely on observational data that may not always be available.

In current research, there is a focus on developing models that leverage geographic information systems (GIS),machine learning (ML) methods and remote sensing (RS) techniques, as well as high-performance artificial intelligence methods, to understand the factors influencing floods and their interrelationships. These advancements are crucial for the preparation of susceptibility maps, which serve as a fundamental step in flood risk management.By integrating data from various sources and employing advanced modeling techniques, researchers aim to enhance our understanding of flood phenomena and improve flood risk assessment and management strategies.

The objective of this study is to produce a flood susceptibility map for the Chellif basin using machine learning models, specifically LGBM; AdaBoost and the new model in machine learning Stacking. The methodology followed in this study is as follows:
Preparation and extraction of influential factor maps,in this step maps related to factors influencing floods are prepared and extracted. These factors could include rainfall patterns, topography, land use, and other relevant variables. Next,analysis and data processing of factor data,the collected data on these influential factors are analyzed and processed to ensure their quality and suitability for further analysis. After that the application of machine learning to assess the impact of each factor on floods,machine learning techniques such as LGBM,AdaBoost and Stacking are applied to evaluate the impact of each factor on flood occurrence and severity. These models are trained using the prepared factor maps and relevant flood data. Also,Validation of machine learning model results,the results obtained from the machine learning models are validated to assess their accuracy and reliability. This validation process involves comparing the model predictions with observed flood events or other reference data.

Finally,Combination and mapping of factor results,the results obtained from analyzing each factor are combined and mapped to create the final flood susceptibility map. This map represents the spatial distribution of flood susceptibility levels in the study area, integrating the influence of multiple factors.

By following this methodology, the study aims to provide a comprehensive flood susceptibility assessment using machine learning models.The resulting susceptibility map can be a valuable tool for understanding and managing flood risks.

This study is divided into four chapters:

1.The first chapter includes a literature review on floods.

2.The second chapter presents a global vision of the study area.

3.The third chapter is devoted the methodology to be followed.

4.The fourth chapter concerns the results of the study and their interpretations.

And finally a general conclusion which summarizes the results obtained in the chapters previous.

Chapter I

Literature Review

Floods are a widespread and devastating natural disaster that can have adverse impacts on human health, natural and artificial environments . They pose a major risk to human life(loss of life, injury),assets (agriculture area, yield production, homes, and buildings), communication systems (urban infrastructure,bridges, roads, and railway lines), culture heritage, and ecosystems [28, 30].Flood-related incidents cause more than 2000 deaths every year, and more than 75 million people are adversely affected across the globe [44].Floods are often triggered by many factors, including both natural and anthropogenic factors. Heavy rainfall or snow melt that overflows to adjacent areas or floodplains and temporarily inundates surrounding areas is a common cause of floods [4,5].Recent studies have indicated that climate change is a fundamental factor that induces floods in various parts of the world [39]

Charlton et al.[17] indicate that flood disasters in a region can be considerably influenced by changes in land use patterns forming an impermeable surface, which may increase flow velocity. Other factors that contribute to flood occurrence include slope, elevation, land use, curvature, Normalized Difference Vegetation Index (NDVI), proximity to rivers, among others.

Due to the complex nature of floods and their frequent occurrence and extensive destruction across the globe, many scientists have devoted significant effort to investigate and understand flood hazards for better mitigation and management [20, 22]

Algeria is prone to floods, and in the last 20 years, it has experienced severe flooding that resulted in the loss of hundreds of lives, thousands of injuries, and property damage worth over a billion dollars.[10] The majority of these flood disasters occurred in the arid and semi-arid regions, where there are intense storm events that trigger flash floods, especially during the autumn season. Flash floods are characterized by a sudden rise and fall of water with little to no warning, usually caused by heavy rainfall in a relatively small area,as defined[12]

The impact of floods in Bangladesh is widespread and multifaceted. Firstly, floods in Bangladesh result in the loss of lives and injuries, with vulnerable communities bearing the brunt of these disasters. Due to its unique hydro-geographic location and low-lying floodplains, Bangladesh experiences annual floods from June to September during the rainy season. Flash floods, particularly in the northern region of Bangladesh and downstream areas like the lower Teesta River basin, cause significant economic loss, fatalities, damage to infrastructure, housing, and bridges, often exceeding US$20 million, as well as affecting a significant number of people, according to Fao and Unicef (2017)[33]. In August 2017, a flash flood occurred in the lower Teesta River basin, affecting Kurigram, Lalmanirhat, and Nilphamari districts, resulting in a vast landmass being submerged and five individuals swept away. According to NIRAPAD, this flood had adverse impacts on about 6.8 million people and over 560,000 hectares of cropland, resulting in damages of up to US$10 million (FAO and Unicef, 2017).

Iran is particularly prone to flash floods. In 2019 alone, flash floods killed 78 people, injured 1076 others, and displaced about 300,000 people; in total, some 10 million people were affected (UN Office for the Coordination of Humanitarian Affairs, 2019). One of the most susceptible areas to flash flooding is the Haraz watershed in Mazandaran Province[34]. Twenty-eight villages in this area weredestroyed during flash floods[16].

Mapping flash floods involves the use of remote sensing technologies and geographic information systems (GIS) to detect changes in water levels and identify flooded areas. Remote sensing technologies, such as satellite imagery and airborne sensors, are employed to detect alterations in water levels and assess the extent of flooding. The data collected is then processed using GIS software, which incorporates information on topography, land use, and other factors influencing water flow, to generate maps.[9]

This section presents a comprehensive description of the GIS environment and geospatial data preprocessing methods employed in the research. The application of GIS facilitated the efficient analysis and processing of diverse data, encompassing climatological, morphological, hydrological, and geological factors. By integrating these factors within the GIS framework, thematic maps were created to visually represent their spatial distribution across the study area.To ensure the reliability and accuracy of the analysis, a comprehensive approach to geospatial data preprocessing was adopted. The initial step involved the extraction of relevant factors using an inventory map that classified sites as either experiencing land degradation or non-degradation. This inventory map served as the fundamental basis for collecting essential information for subsequent analysis.[46]

Following the extraction of factors, a pre-treatment analysis of the statistical data was conducted. This involved applying appropriate statistical techniques to evaluate and preprocess the extracted values, ensuring data quality, consistency, and suitability for further modeling. Classification modeling techniques were then employed to develop models capable of predicting flood occurrence based on the identified factors. Machine learning algorithms were integrated into the GIS environment to facilitate this modeling process, enabling the capture of complex relationships and patterns between the factors and flood susceptibility.[56]

The performance of the developed models was assessed using performance criteria, which evaluated the accuracy and reliability of the models in predicting flood susceptibility within the study area. This performance evaluation provided valuable insights into the strengths, limitations, and practical applicability of the models. Furthermore, feature importance analysis was conducted to determine the relative significance of each factor in contributing to the flood susceptibility models. This analysis prioritized and assigned weights to different factors in the final susceptibility maps generated by the models, thereby enhancing the accuracy and effectiveness of their predictions.

The integration of GIS and machine learning techniques in this research harnessed spatial data to develop highly accurate flood susceptibility models. This integration facilitated a comprehensive analysis of the study area, yielding valuable insights into the factors influencing flood occurrence. The combination of GIS and machine learning techniques showcases their potential synergy in addressing complex environmental issues and supporting informed decision-making processes.[49].

While much research has been conducted on the factors that contribute to flash flood initiation, there are still areas that require more attention. These factors include the impact of microtopography and soil moisture on flash flood initiation. The role of microtopography is complex due to the terrain's interaction with precipitation, making it difficult to fully understand its impact[13]. Soil moisture plays a critical role in the initiation of flash floods, but it is often oversimplified or ignored in hydrological models. More research is needed to better understand these factors and their impact on flash floods, especially in regions where flash floods are common but data is limited[38].

There are also several other factors related to flash floods that require further investigation. These include the effects of climate change, land use and land cover changes, and human interventions on flash floods. Identifying triggers and thresholds for flash floods is also an area that needs more attention. The role of groundwater in flash flood initiation and propagation, as well as the effects of sediment transport and vegetation cover, are additional factors that require further study. Understanding the interactions between flash floods and other natural disasters such as landslides and debris flows is also important[31].

Despite the existence of flash flood models, much about the phenomenon remains poorly understood. While rainfall intensity, catchment topography, and soil properties have been identified as key factors affecting flash flood generation, other factors such as antecedent soil moisture, land use, and river morphology require further investigation. The transferability of flash flood models to other regions has not been widely evaluated, and model parameterization remains a significant challenge. Therefore, more research is needed to better understand the factors contributing to flash floods and to improve flash flood prediction, forecasting, and management.

Deep learning techniques have gained significant traction in flood management as they offer potential solutions to the limitations of traditional numerical models, providing accurate results at a faster pace. These models have demonstrated improved accuracy compared to conventional approaches and faster computation times compared to numerical methods. While various applications have emerged in flood susceptibility, inundation, and hazard mapping, further research is required to explore how deep learning can support real-time flood warning systems during emergencies and enhance flood risk estimation[9].

One major challenge is the development of deep learning models that can generalize effectively to new case studies. To address this, recent advancements in deep learning can be leveraged, drawing inspiration from developments in other domains. For instance, models based on graph neural networks and neural operators can handle data with arbitrary structures, allowing for generalization across diverse case studies and considering complex interactions with the natural and built environment. Moreover, to account for uncertainties and provide probabilistic predictions, it is essential to move beyond deterministic models. Deep Gaussian Processes or Bayesian neural networks can be employed to construct probabilistic models, enabling the incorporation of uncertainties in outcomes[29].

Additionally, integrating physics-based principles into deep learning approaches can help preserve underlying physical equations, resulting in more reliable alternatives to accelerate numerical models. This approach, known as physics-based deep learning, ensures that the models align with known physical laws and principles. By addressing these identified gaps and leveraging recent advancements, deep learning techniques can play a crucial role in enhancing flood management, real-time flood warning systems, and accurate flood risk estimation.

Despite the numerous models developed for predicting flash floods, hydrodynamic modeling still faces various challenges in flash flood mapping. Limited data availability poses one of the difficulties, which is vital for hydrodynamic model calibration and validation[37]. Moreover, the flash flood phenomenon's complexity involves many variables that interact with each other in intricate ways, and most existing models assume a uniform rainfall distribution over the catchment, which may not always be the case. Additionally, deterministic approaches are widely used in most models, and they do not consider the uncertainties related to the input data and model parameters. These shortcomings underscore the need for further research to improve hydrodynamic modeling in flash floods[5].

Another significant challenge is the lack of high-resolution topographic data, which is crucial for precise modeling of flow dynamics. There is also a need for more comprehensive models that account for the interactions between various hydrological processes like surface runoff, infiltration, and subsurface flow. In addition, the current models make assumptions about the soil's hydraulic properties and flow's boundary conditions, leading to considerable errors in model predictions. Further research is necessary to understand the uncertainties related to input data and model parameters and develop robust approaches for model calibration and validation. These research gaps emphasize the importance of further research to enhance hydrodynamic modeling for flash flood mapping and understand the flash flood phenomenon's complex nature[57].

Chapter II

Presentation of the Study Area

II.1 Introduction

A thorough understanding of the physical environment of the studied area is essential for comprehending the phenomenon under investigation and interpreting the results. Since each site has its own specificities, studying the site is necessary to understand the physical characteristics of the location. This chapter aims to present, in its first part, the study area and its characteristics, and in its second part, the issue of floods in this area.

II.2 Geographical location and delimitation

The Cheliff basin covers an area of 43,794 km² and is divided into two distinct regions:

1. The upstream part of the Cheliff, with an area of 20,500 km², is bordered to the south by the Djebel Amour mountains and to the north by the Ouarsenis mountains.

2. The downstream part of the Cheliff, spanning an area of 23,250 km², is bordered to the south by the Tiaret, Saida, and Ouarsenis massifs, and to the north by the Dahra and Beni Menacer.[36]

The Cheliff basin extends between longitudes 0° 7' and 3° 31' East and latitudes 33° 53' and 36° 26' North. It is adjacent to:

- The Algiers and Oran coastal areas to the north.
- The Issue, Hodna, Zahrez, and Constantine high plateau basins to the east.
- The Macta and Oran high plateau basins to the west.
- The Saharan Atlas to the south.

The location of the study basin is shown in Figure (II.1)

The Cheliff basin itself consists of three major hydrological units or main sub-basins:

1.Upper Cheliff River (upstream of Boughezoul) covering an area of 19,710 km².

2.Middle and Upper Cheliff sub-basin (Middle and Upper Cheliff) covering 13,870 km².

3.Lower Cheliff and Oued Mina sub-basin (Lower Cheliff and Mina River) covering 10,170 km².[36]

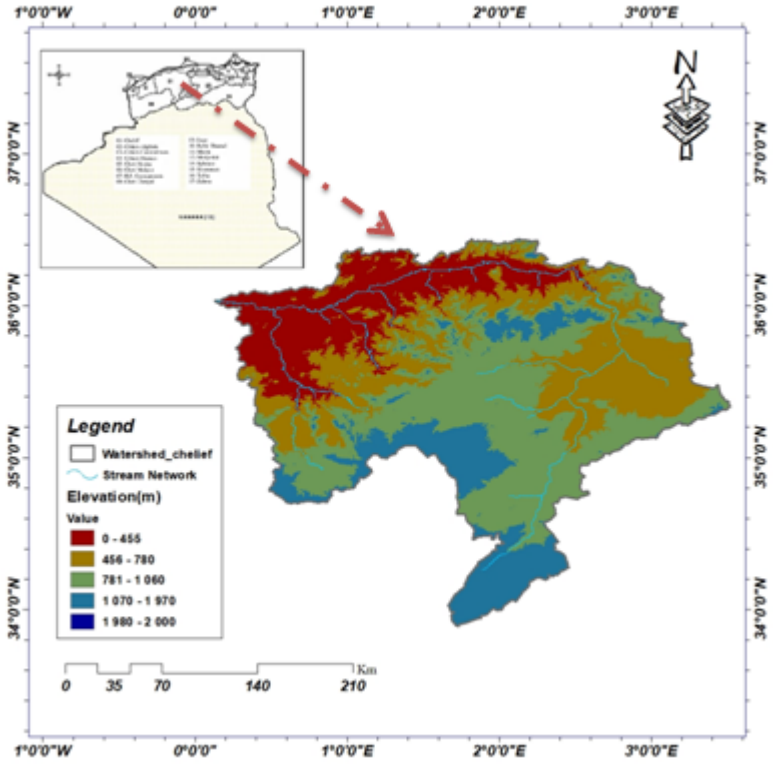


Figure II.1: Geographical Location of the Cheliff Watershed

II.3 Topography

The topography of the Chellif River watershed is varied, with the highest elevations located in the Tell Atlas Mountains in the south and the lowest elevations located near the coast in the north. The watershed is characterized by several mountain ranges, including the Ouarsenis, the Dahra, and the Djurdjura. The highest point in the watershed is Djebel Toubkal, which rises to an elevation of 4,167 meters in the Tell Atlas Mountains. The Cheliff wadi, located in the northern part of the watershed, flows westward through the province of Chlef before eventually reaching the Mediterranean Sea near Mostaganem..[36]

II.4 General Characteristics of the Cheliff Watershed

II.4.1 Geology

The Cheliff basin belongs to the elongated East-West sub littoral sedimentary basins, which were formed after the last phase of Alpine tangential tectonics To the north, this depression is separated from the sea by the northern Tell, represented by a series of parallel reliefs mainly composed of Jurassic-Cretaceous formations, which are also found in the plain (Dahra and epimetamorphic massifs with schistosity of Doui, Rouina, and Témoulga). To the south, the Cheliff basin is limited by the southern Tell, characterized by a set of mountainous massifs where the substrate consists mainly of marl-limestone and corresponds to the allochthonous Tellian zone with its various nappe structures.[36]

The Geological map ofthe Cheliff watershed is shown in Figure (II.2):

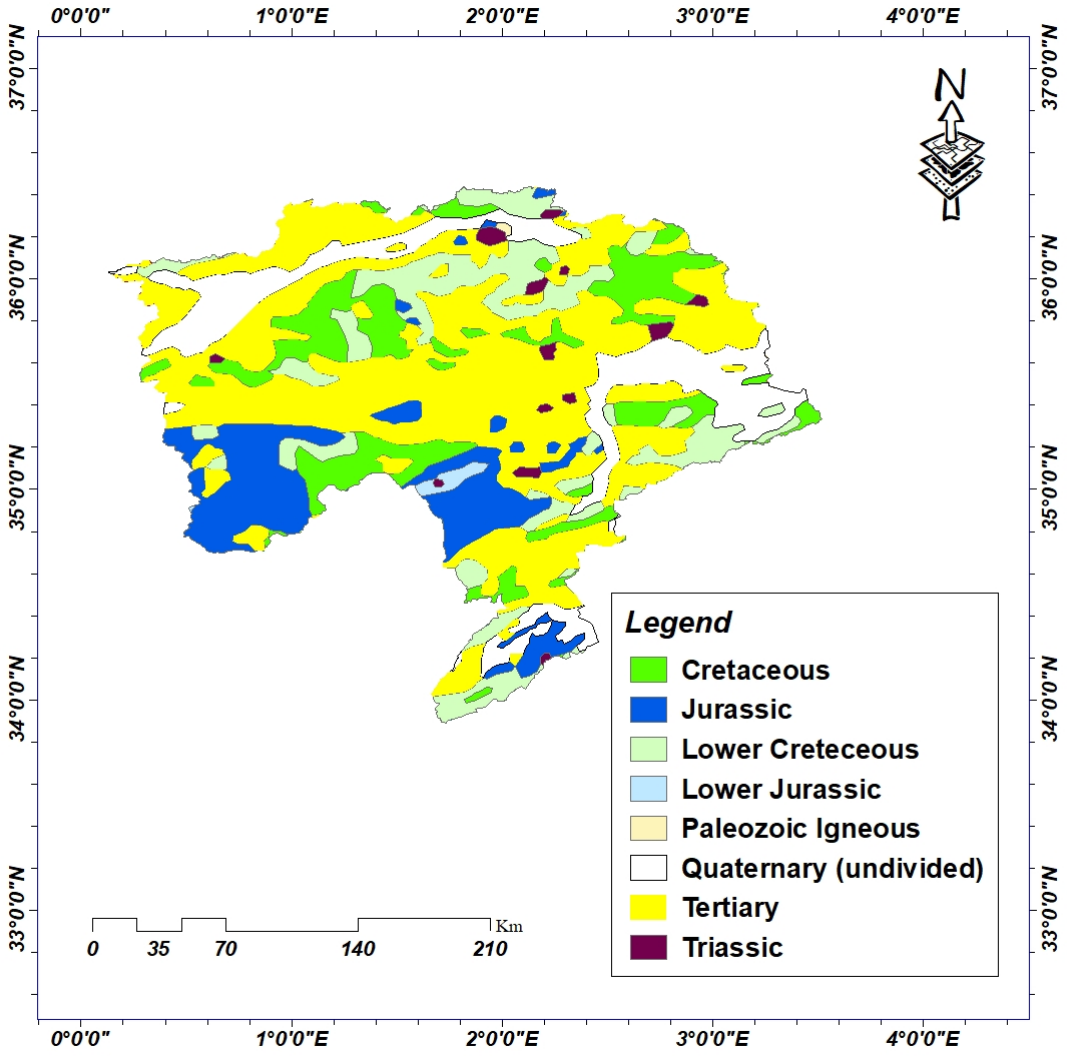


Figure II.2: Geological situation of the Cheliff watershed

II.4.2 Vegetation and Land Use

The spatial distribution of land use in the Cheliff basin exhibits several categories of land occupation. Rangelands dominate the basin, covering 28550,57 km², which represents 65.19% of the total surface area. Crops also occupy a significant portion, 7124.80 km², accounting for 16.27% of the basin.Bare ground covering 5669.83 km²(12.94%). Forest formations cover approximately 4% of the total area and are primarily found in regions with moderately rugged terrain. The remaining portion of the basin comprises unproductive lands such as Built Area, water bodies (Table 1).

Table II.1: Distribution of Land Use / Land Cover area in the Cheliff basin.

LULC Class	Area(km²)	Area(%)
water	62.43	0.14
Trees	1354.28	3.09
Crops	7124.81	16.27
Built Area	1032.07	2.36
Bare ground	5669.84	12.95
Rangeland	28550.58	65.19
Total	43494	100

The LandUse/LandCover Map is shown in Figure (II.3):

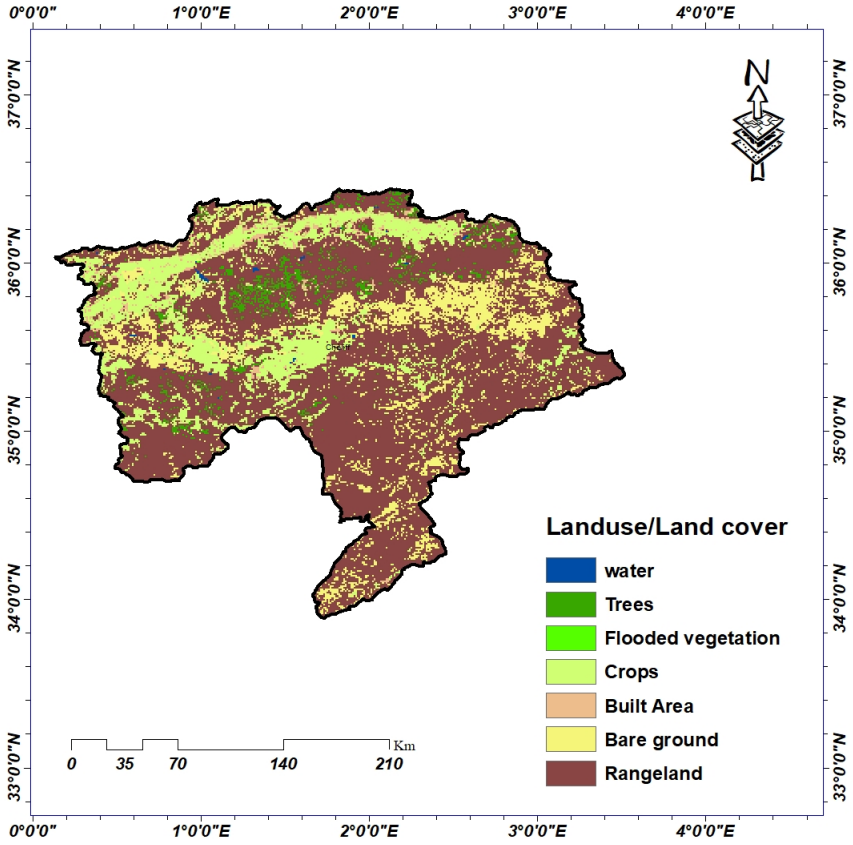


Figure II.3: Land use/Land cover map

II.4.3 Agricultural Activity

The Cheliff basin is primarily characterized by agricultural activities. Upstream of the Boughazoul dam, olive cultivation is predominant. In the upper and middle regions of the Cheliff, the main agricultural practices include cereal crops, forage production, horticulture, industrial crops, fruit tree cultivation, and legumes.

In the lower Cheliff area, the primary agricultural practices revolve around cereal crops, forage production, horticulture, fruit tree cultivation, and legumes.[51]

II.4.4 Hydrographic Network

All the water from the main tributaries is collected towards the center of the Cheliff plain, flowing into the main Oued (river). This river drains all the water from the Cheliff basin towards the outlet located near Mostaganem (Figure II.4). The Cheliff originates from the Djebel Amour, located at the edge of the Sahara. Initially, there are several tributaries, including Oued Berkana, Sabgague, Mekta, and Oued Namous.

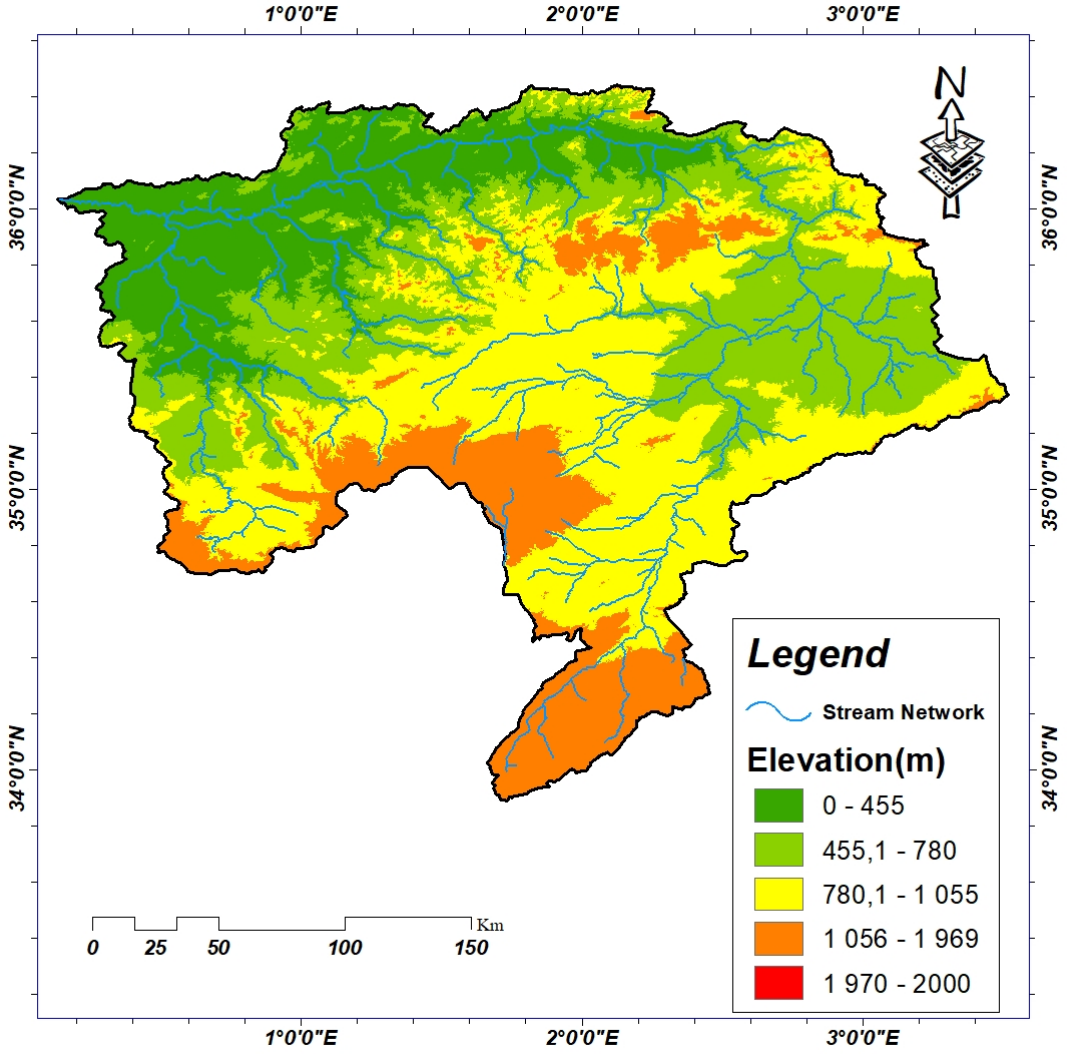


Figure II.4: Hydrographic network of chellif basin.

The Cheliff River, resulting from the confluence of these two watercourses, crosses a gap between Ksar El Boughari and Boughar, leaving the high plateaus and entering the Tell region. From Boughari to Amoura (formerly known as Dollfuss city).

it flows through deep gorges, and starting from the Djebel Amoura, it follows an east-west-oriented plain. Finally, after separating the Mostaganem plateau from the Abdel Malek Ramdan plateau (formerly known as Willis), the Cheliff discharges its waters into the sea after a journey of over 720 km.

II.5 Morphological characteristics of the watershed

The characteristics of a watershed can have significant effects, particularly on the variation of flow during flood periods. Both external factors such as precipitation and climatic conditions, as well as internal factors related to the morphological characteristics of the watershed, notably the size (surface area), shape, altitude, slope, and orientation, have an impact on the time of concentration. The time of concentration partly describes the speed and intensity of the watershed's response to precipitation. The Chellif watershed encompasses an area of 43,794 km^2 and has a perimeter of approximately 1438 km.

II.5.1 compactness index

The compactness index is a shape parameter used to characterize the shape of a watershed. It is calculated as follows:

$$C = 0.28\frac{P}{\sqrt{S}} \tag{II.1}$$

P: the perimeter of the watershed (Km).

S: the surface of the watershed (Km^2).

When the compactness index is greater than 1, it indicates that the watershed's perimeter is longer than that of a circle with the same area. In this case, a value of 1.92 suggests that the watershed's shape is roughly 92% longer or more intricate than a perfect circle with the same area.The compactness coefficient of the Chellif watershed is 1.92 indicating its elongation and the development of linear erosion.

II.5.2 Equivalent rectangle

An equivalent rectangle is a rectangle that has the same area (or sometimes the same perimeter) as the original shape but does not necessarily have the same proportions or orientation. It is a way of reducing a complex shape to a simpler geometric form while preserving a specific characteristic, such as area or perimeter. To determine the equivalent rectangle, the length and width of the rectangle are adjusted to match the area of the original shape. The resulting rectangle may have different dimensions and aspect ratios depending on the shape being approximated

a. Length of equivalent rectangle:
The length of the equivalent rectangle is given by the following expression:

$$L = \frac{C\sqrt{S}}{1.12}\left(1 + \sqrt{1 - \left(\frac{1.12}{C}\right)^2}\right) \tag{II.2}$$

C: compactness index.
S: surface of the watershed in km².
L: length of the equivalent rectangle in km.

b. Width of equivalent rectangle:
The width of the equivalent rectangle is given by the following expression:

$$l = \frac{C\sqrt{S}}{1.12}\left(1 - \sqrt{1 - \left(\frac{1.12}{C}\right)^2}\right) \tag{II.3}$$

C: compactness index.
S: surface of the watershed in km².
l: width of the equivalent rectangle in km.

II.5.3 Average slope index:

The average slope index is a measure used to quantify the steepness or gradient of a particular terrain or surface. It provides information about the rate of change in elevation over a given distance. It plays a vital role in hydrological modeling and watershed management. It affects surface runoff, water infiltration, and the formation of drainage networks. Steeper slopes generally result in faster runoff and can contribute to increased erosion and flood risks

$$I = \frac{H_{\max} - H_{\min}}{L} \tag{II.4}$$

These parameters are summarized in the following table:

Table II.2: Summary of the morphological characteristics of the Chellif watershed

Parameter	**Symbol**	**Value**	**Unit**
Surface	S	43794	Km²
Perimeter	P	1438	km
compactness index	C	1,92	-
Length of equivalent rectangle	L	650,13	Km
Width of equivalent rectangle	l	67,36	Km
Minimum altitude	Hmin	0	m
Maximum altitude	Hmax	1800	m
Average altitude	Hmoy	900	m
Average slope index	I	2,76	%

II.6 The nature of watershed surfaces

The nature of watershed surfaces plays a crucial role in their hydrological behavior. Parameters such as slope, lithology, soil characteristics, and vegetation cover are influential factors. These parameters greatly affect surface permeability and roughness, which in turn

determine the speed of runoff and the ratio of runoff to infiltration.

A significant portion of the valley's area is occupied by rainfed crops, with fruit trees covering a very small area. Vegetation cover influences evapotranspiration, delays runoff, and helps in conserving the soil surface against erosion.

The topography is highly varied, characterized by pronounced relief and steep slopes that promote runoff velocity over infiltration.

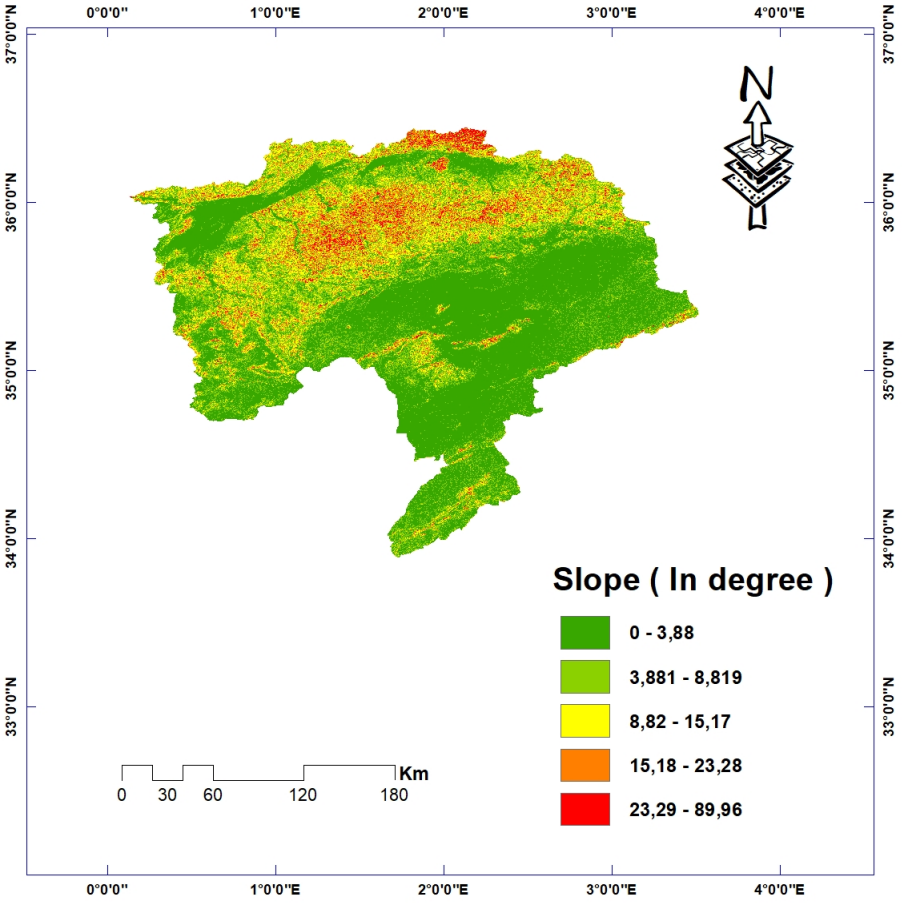


Figure II.5: Slope variation Map

The classification of soil and vegetation cover is depicted on maps that are subject to possible transformations or changes resulting from alterations in forested areas or forest fires. It should be noted that this region has a Mediterranean climate, and vegetation is often adapted to withstand fires or possesses good regenerative capacity. The regeneration of vegetation cover is particularly rapid in areas with high rainfall and much slower in areas with low rainfall.

Areas that may undergo changes due to forest fires or logging are eventually recolonized, within a period of 7 or 8 years, by shrubs as a substitute for the original forest cover. However, complete regeneration may require 30 to 40 years in humid Mediterranean zones. When considering calculations involving areas with vegetation cover or erosion risks,

significant changes should be avoided, as their accidental disappearance or alteration should not imply modifications in the future purpose and occupation of these areas.

II.7 climatology

II.7.1 Climat:

The climate of the Chellif plain is typically Mediterranean, characterized by hot and dry summers and cool and rainy winters. Despite this, the rainfall is relatively low, with average annual precipitation of 473mm for El Khemis, 400mm for Chellif, and 325mm for Ghilizane. The annual potential evapotranspiration (calculated using the Turc formula) is 1145mm for High Chelif, 1276mm for the central area, and 1300 for the low

II.7.2 Precipitaion:

Precipitation is one of the most important elements that define the climate of a given location. Since the study region is characterized by a semi-arid climate, it has experienced prolonged periods of drought in recent years. The average annual precipitation varies between 400 and 600mm. The first rains usually occur between September and November, while the last ones fall from April to May. Although rainfall events are infrequent and short-lived, they are intense, which means that floods can be dangerous and cause significant damage to both human lives and property.(site:power.larc.nasa.gov)

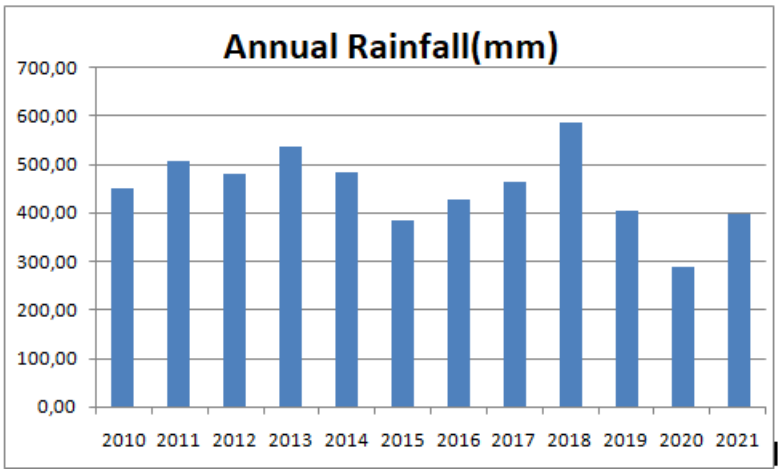


Figure II.6: Annual rainfall variation in chellif watershed (2010-2021)

II.7.3 Temperature:

The upper and middle Cheliff basin is characterized by an average annual temperature ranging from 13 to 20°C. The monthly maximum temperature exceeds 30.6°C in July at the Ain Defla station, while the monthly minimum temperature is 4.6°C in January at the Medea station (ANRH, 2014). The observation period for temperature, similar to rainfall, needs to be specified. In the lower Cheliff plain and Oued Mina, the average annual temperature varies from 14 to 29°C. The monthly maximum temperature reaches 29.5°C in August at the Relizane station, and the monthly minimum temperature is 6.1°C

in January at the Tiaret station (ANRH, 2014). In the Cheliff basin near Boughazoul, the average annual temperature ranges from 13 to 16°C. The monthly maximum temperature is 27°C in August at the Ksar Chellala station in the Tiaret province. All these temperature variations are illustrated in Table (II.3), which indicates the average annual temperature in four stations distributed throughout the Cheliff region.(ANRH, 2014)
These parameters are summarized in the following table:

Table II.3: Mean annual temperature (C°) for the Cheliff regions, during the 2004/2014 period. (ANRH, 2014)

Années	**Tiaret**	**Chlef**	**Miliana**	**k-chellala**
2004	14,3	20,9	16,01	12,99
2005	15,7	19,6	16,79	13,85
2006	15,7	20	15,98	13,85
2007	15,9	20	18,34	13,85
2008	15,4	20,3	16,38	13,85
2009	15,6	19,7	17,19	15,03
2010	15	20,2	16,09	19,29
2011	15	19,8	18,34	19,29
2012	15	19,8	16,09	13,75
2013	15	19,8	16,09	14,02
2014	15,8	19,8	16.02	16,60
Max	15,9	20,9	18.34	19,29
Min	14,3	19 ,6	15,98	12,99
Ecart	1,6	1,9	3,05	6, 3
Moy	15,3	19,9	16 ,4	15,5

II.7.4 Relative humidity (Hr%)

Refers to the amount of water vapor present in a certain volume of air and is considered one of the essential elements of the hydrological cycle. The relative humidity data for the Tiaret station, located in the center of the Cheliff basin, are provided in Table (II.4)(Source ANRH 2008).

Table II.4: Relative humidity recorded at the Tiaret station (1980-2008)

M	Sep	Oct	Nov	Dec	Jan	Fev	Mar	Avr	Mai	Juin	Juil	Août	**Annuel**
Hr %	55	64	73	78	76	69	66	56	54	48	38	45	60,17

II.8 Mobilization of surface water resources

The Cheliff watershed is equipped with 16 operational dams, which are referenced as follows: Merdja Sidi Abed, Gargar, Sidi M'Hamed Ben Aouda , Bakhada, Dahmon, Colonel Bougara, Sidi Yacoub, Oued elFouda, Derdour, Harezza, Ghrib, Sidi M'Hamed Ben Taiba, Ouled Mellouk, Koudiat Rosfa

These dams serve various purposes within the watershed, such as water storage, flood control, irrigation, and hydroelectric power generation. the characteristics of these dams are listed in the table

Figure II.7: Dams in the Chellif watershed (Source: AGIRE)

Table II.5: Dams in the Chellif watershed

dam	Oued	Type Work	Y_c	$I_c(HM^3)$	$L_{sc}(HM^3)$	L (HM^3)	coordinates	
							Latitude	Longitude
Merdja S.Abed	Chlef	Terre	1984	54,9	44,86	10,04	36° 0' 38" north	0° 59' 26" east
Gargar	Rhiou	Terre	1988	450	283,49	166,51	35° 25' 40"north	1° 33' 47" east
Sidi M.B Aouda	Mina	Terre	1978	235	125,32	109,68	35° 36' 12" north	0° 35' 25" east
Bakhadda	Mina	En roche	1959	56	37,26	18,74	35° 20' 0" north	1° 3' 37" est
Dahmouni	Nahr- Ouassel	Terre	1987	41	35,51	5,49	35° 25' 40" north	1° 33' 47" east
C.Bougara	Nahr- Ouassel	Terre	1989	69,50	60,65	8,85	35° 33' 41" north	1° 54' 3" east
Sidi Yacoub	Ardjen	Terre	1985	280	224,064	55	35° 57' 21" north	1° 19' 31" east
Oued el Fodda	Fodda	Poids	1932	228	96,79	131,21	36° 2' 44" north	1° 36' 41" east
Derdeur	Ain Defla	Terre	1984	115	107,54	7,46	36° 19' 46" north	2° 1' 43" est
Harreza	Hareza	Terre	1984	76	74,61	1,19	33° 27' 38" north	1° 11' 0" east
Ghrib	Chlef	En roche	1939	350	169,35	180,65	36° 9' 46" north	2° 33' 38" east
Sidi M.B Taiba	BDA	En roche	2005	75	70,22	4,78	36° 19' 46" north	2° 1' 43" east
Ouled Mellouk	Rouina	Terre	2003	127	114,10	12,9	36° 11' 18" north	1° 50' 22" east
Koudiat Rosfa	Fodda	Terre	2004	75	66,04	8,96	/	/

L_{on}: Longitude

L_{at}:Latitude

L:Loss (HM^3)

L_{sc}: Last survey capacity (HM^3)

Ai:Ability initial

Y_{ss} : Year setting service

II.9 Groundwater resources:

The Cheliff watershed contains several hydrogeological units that contain significant aquifers, suitable for both drinking water supply (AEP) and irrigation purposes. In the northern part of the Cheliff basin, located between the two Tellian mountain ranges of Mont du Dahra and the Ouarsenis massif, numerous geological formations hold underground water resources. The oldest groundwater is attributed to the Jurassic period, while the most recent corresponds to Quaternary alluvium.

The Cheliff basin is divided into three sub-basins: upper, middle, and lower Cheliff, separated by two thresholds known as the Aïn Defla threshold and the Oum Drou threshold. The Cheliff region is home to 42 aquifers, with a potential of 330 million cubic meters per year (ANRH, 2015). These aquifers represent valuable water resources for various uses in the region, including drinking water supply and agricultural irrigation.

II.10 History of flash floods in the region

Flash floods have had a significant impact on the Cheliff watershed throughout its history. The region's geographical characteristics, coupled with climatic factors, contribute to the occurrence of devastating flash floods. The Cheliff watershed, located in a mountainous region with steep slopes and narrow valleys, is prone to intense rainfall events, particularly during the winter season. These events, combined with poor soil infiltration capacity due to the presence of impermeable rock layers, amplify the risk of flash floods.

Over the years, the Cheliff watershed has experienced several major flash floods that have resulted in catastrophic damage. One notable example is the flash flood of 1987,This event is one of the most devastating flash floods in the history of the Cheliff watershed. It occurred after heavy rainfall, primarily in the mountainous areas surrounding the watershed.

The sudden surge of water caused flash flooding in the valleys and downstream areas. The floodwaters inundated villages, towns, and agricultural lands along the Cheliff River, resulting in significant loss of life and widespread destruction. Infrastructure such as roads, bridges, and buildings were severely damaged or completely swept away. The economic impact was substantial, affecting agricultural activities and disrupting the livelihoods of the local population, another significant event occurred in 2001.
This flash flood event occurred following intense and prolonged rainfall in the Cheliff watershed. The heavy downpours saturated the soil, leading to rapid runoff and a surge of water in the river system. The flash flood affected various regions within the watershed, including cities and towns such as Ain Defla and Chlef. The floodwaters submerged urban areas, causing extensive damage to houses, public infrastructure, and businesses. Agricultural lands were also severely impacted, leading to significant losses in crop production. The flash flood of 2001 resulted in the displacement of people, loss of lives, and a long-term recovery and reconstruction process for the affected communities.

Efforts to mitigate the risk of flash floods in the Cheliff watershed have been ongoing. Construction of reservoirs and dams, implementation of early warning systems, and improved land management practices are among the measures undertaken to minimize the impact of future flash flood events. However, the history of flash floods in the Cheliff watershed serves as a reminder of the need for continued vigilance, preparedness, and adaptation to ensure the safety and well-being of the communities living in this vulnerable region.

Table II.6: History of flooding in the cheliff watershed

Locale	Data	Cause
Boumedfaa	01/01/1969	/
Chiffa	01/03/1973	/
Sidi Aissa	27/10/1985	/
Oued Fodda	06/10/1986	Flash flood wadi
sidi bouzid	05/03/1988	/
Ksar El Boukhari	05/10/1991	/
tiaret	19/10/1991	/
Berrouaghia	05/10/1991	Flash flood wadi
Taougrit	10/11/2001	/
Hoceinia	03/05/2006	/
Ain Tork	03/05/2006	Flash flood wadi
El Hamdania	09/04/2007	/
Ain Boucif	30/09/2009	Flash flood wadi
Zeboudja	01/01/2009	Urban runoff
Zeboudja	01/10/2009	Urban runoff
Ouled Fares	01/10/2009	Urban runoff
Soumaa	24/09/2009	Urban runoff
Blida	29/09/2009	Urban runoff
Blida	24/09/2009	Urban runoff
Blida	28/09/2009	Flash flood wadi
Damous	07/11/2010	Flash flood wadi
Taougrit	10/11/2010	/
El Abadia	02/02/2011	/
djelfa	01/10/2011	/
Labiod Medjadja	01/04/2012	Flash flood wadi
Taougrit	01/10/2012	/
Ouled Yaich	11/03/2012	/
Blida	11/03/2012	/
Bouaarfa	21/05/2013	/
El-Affroun	21/05/2013	Flash flood wadi
Ain Romana	21/05/2013	Flash flood wadi
Taougrit	01/04/2013	/
msila	25/04/2013	/
Taougrit	01/01/2014	/

II.11 Flood Inventory Map

The flood inventory map serves as a fundamental tool for assessing flood susceptibility and predicting future flood occurrences in a region. An accurate analysis of flood susceptibility requires a precise flood inventory map that identifies the locations where flooding has occurred.

A flood inventory map records the locations of flood events and provides detailed information on the characteristics of historical floods [18]. The locations of flood events have been obtained through historical records and extensive field surveys conducted by the Ministry of Water Resources and Hydraulic Security[4].
In this study, a total of 256 points have been identified as flood-prone areas, while 256 no-flood-prone points have been randomly selected from no-flooded areas.

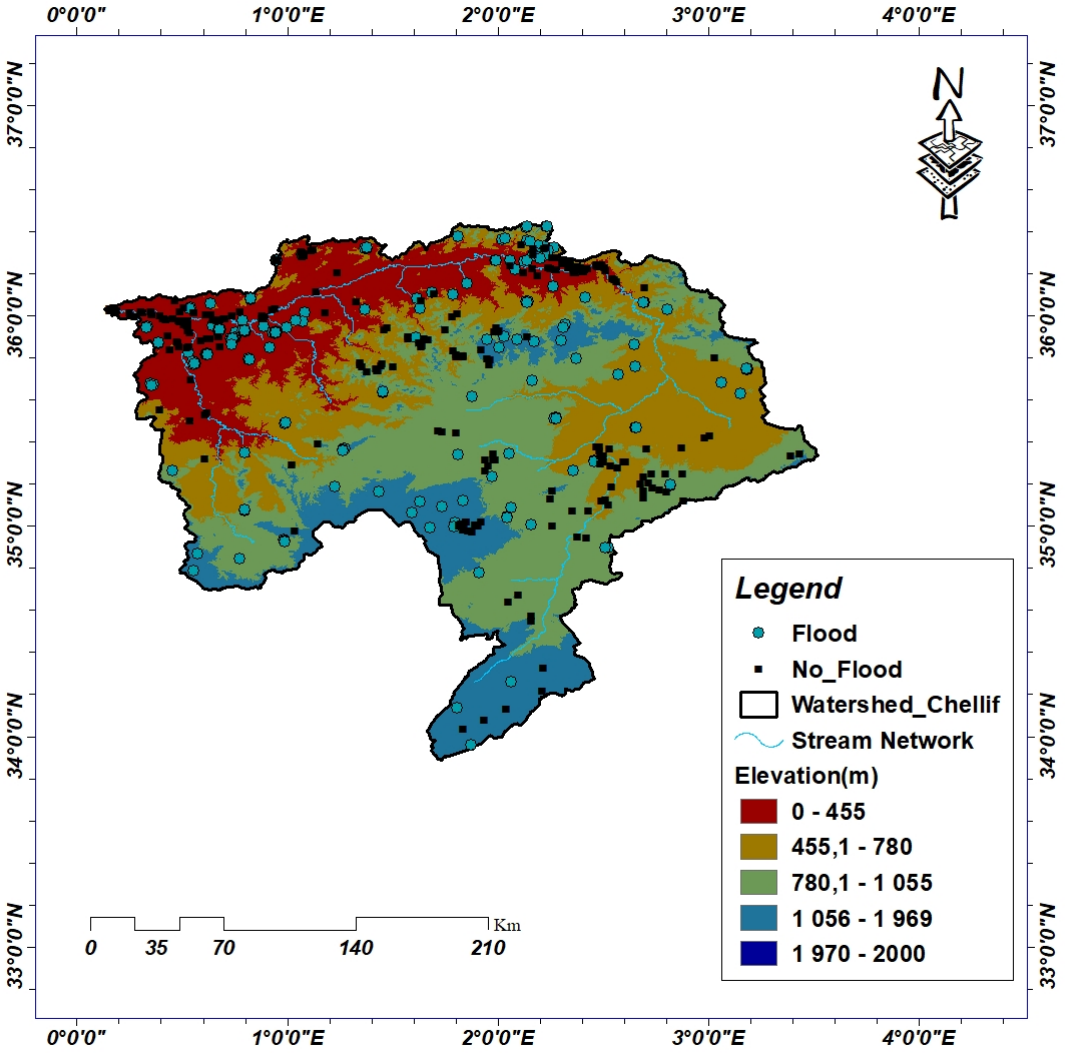


Figure II.8: Inventory map of flood points and non-flood points

II.12 Conclusion

The main purpose of this chapter is to present data that give an overview of the geographical, topographical, climatological, geological, and other characteristics of our study area such as land use, water resources, hydrographic network ..etc

This chapter also allowed us to develop knowledge on the problem of floods within the Chellif Watershed, which will help us to develop a study against flooding in times of flood.

Chapter III

Methodology for Flood Risk Mapping

III.1 Introduction

The present study aims to define the mapping of flood susceptibility. Specifically, it involves producing a flood susceptibility map using a modeling approach with the use of Geographic Information Systems (GIS) and machine learning tools .
The flood mapping methodology consists of four main steps :

1. Construction of a geospatial database for influencing factors and historical flood events.

2. Development and application of machine learning models.

3. Validation of results and evaluation of model performance.

4. Generation of the flood susceptibility map by overlaying the previously determined conditioning factor maps.

To begin, a geospatial database is created, which includes data on factors that influence flood susceptibility, such as topography, land cover, and hydrological characteristics.

Historical flood events are also incorporated into the database.
Next, machine learning models are developed and applied to analyze the relationship between the influencing factors and flood occurrences. These models are trained using the available data and then used to predict flood susceptibility across the study area.

The results of the machine learning models are then validated to assess their accuracy and reliability. Performance evaluation metrics are used to determine the effectiveness of the models in capturing flood susceptibility patterns.

Finally, the flood susceptibility map is generated by overlaying the maps of the previously determined conditioning factors. This map provides a visual representation of areas at higher risk of flooding based on the identified factors.

By following this methodology, the study aims to provide a comprehensive understanding of flood susceptibility and produce a reliable flood susceptibility map that can support decision-making processes related to flood risk management and mitigation.

III.2 Application of GIS in the Chellif Watershed

In the present work, our aim is to create a flood risk mapping in the Chelief Watershed using GIS, remote sensing and AI techniques. The purpose of this chapter is to determine the factors influencing flood risks in the Chellif Watershed using remote sensing techniques in GIS environment.
The methodology followed consists of the following four steps:

1. Firstly, integrating relevant geographical information obtained from satellite images into a GIS.

2. Utilizing the functionalities of ArcGIS on the data integrated into the GIS to calculate the factors influencing floods in the Chellif Watershed. The results will be presented in the form of detailed maps.

The calculation of factors is performed using different methods, which are as follows:

- The hydrographic network was automatically generated from the Digital Elevation Model (DEM) using GIS extension.

- Vegetation cover (NDVI),Normalized Difference Water Index(NDWI) and precipitation data are derived from satellite image processing using ArcGIS.

- Other factors such as slope, SPI, TWI, etc., are calculated based on the DEM.

3. We standardized the resulting factors using ArcGIS. Standardization is a strategy to convert different inputs of a decision problem to a common scale for comparison [8]. In our case study, we standardized the factors so that their values range between 0 and 1.

4. Finally, from the resulting maps, we extracted characteristic data of flooded and non-flooded points in the form of a numerical table. These data will be used in the next step of our work.

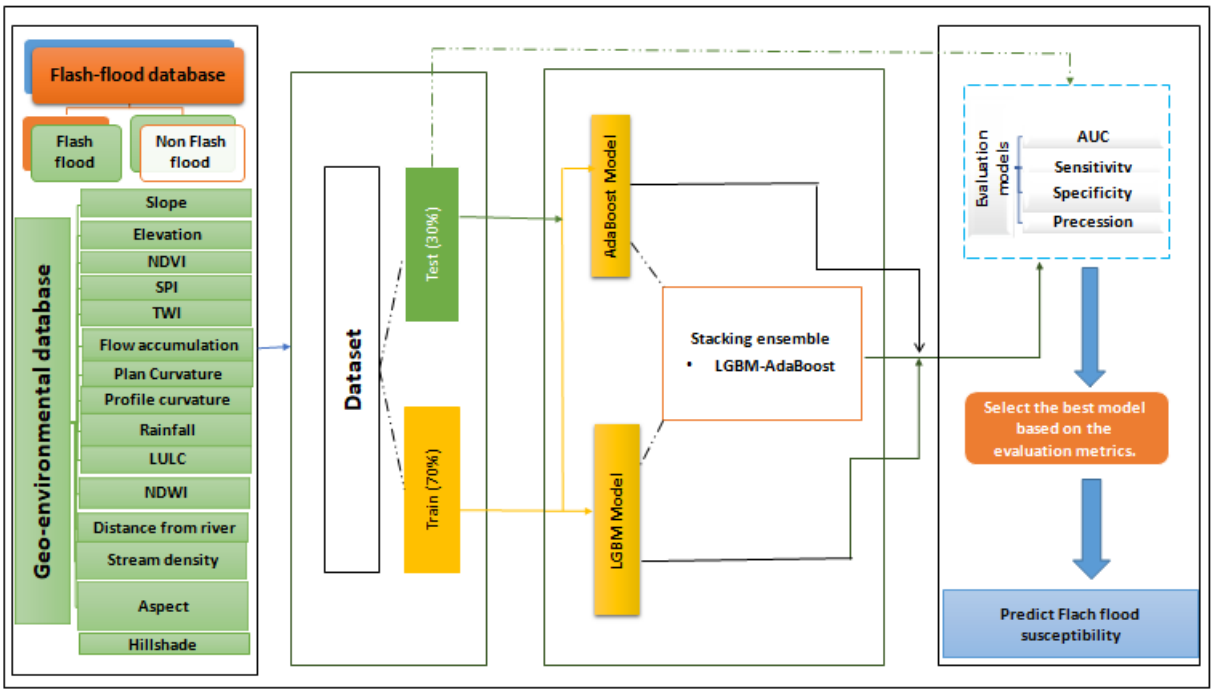


Figure III.1: Flowchart methodology for Flashflood susceptibility mapping.

III.3 Factors influencing flood

The modeling of floods requires the characterization of geospatial information on key hydrological properties of the study area. Therefore, a series of conditioning factors and their relationship with floods need to be studied to create flood susceptibility maps[14].

There are no universal guidelines for selecting flood conditioning factors. The selection of causal factors generally varies from one location to another, depending on the study area and data availability.

In this study, 15 factors have been chosen for flood mapping: elevation,slope, aspect of elevation, Rainfall,Hillshade ,Landuse/Landcover(LULC),Profile curvature,Plan curvature,Stream density,Distance from river,Flow accumulation,Stream power index (SPI),Topographic wetness index (TWI),vegetation cover (NDVI),Water Index(NDWI).

These factors are determined using GIS tools and remote sensing, and they are defined as follows:

Elevation:

The elevation factor assumes paramount importance in the modeling of flood phenomena [42]. In particular, a reciprocal relationship exists between flooding occurrences and elevation levels. Greater elevation diminishes the likelihood of flooding incidents, while lower elevation heightens the risk of such events transpiring . As elevation decreases, the terrain typically assumes a flatter profile, resulting in augmented water flow within streams and rivers [4].

Elevation data acquisition techniques, such as GPS, LiDAR, and satellite measurements, are commonly employed to gather information pertaining to the vertical positioning of the Earth's surface. This collected data serves as the basis for constructing digital elevation models (DEMs) or elevation maps, which effectively depict the diverse topographical elevations found across the terrain. By virtue of these models, terrain characteristics can be visualized and analyzed.

Slope:

Slope is a fundamental topographical attribute that reflects the degree of spatial irregularity across a surface. It plays a pivotal role in governing both the velocity of surface runoff and the process of percolation, thereby exerting influence over the flow dynamics of water, particularly during flood events . A steeper slope angle corresponds to a decreased likelihood of water accumulation, reduced infiltration rates, and elevated flow velocities. Consequently, areas characterized by lower and flatter terrains are more prone to experiencing flooding incidents[52].

Aspect:

The selection of this index stems from its association with the convergence of water flow and the determination of flow directions .Aspect plays a significant role in influencing the paths of flooded water flow, while also influencing the distribution of soil moisture.

Consequently, aspect indirectly impacts the occurrence and extent of flooding events.[50]

Rainfall:

The occurrence of flash flooding is closely associated with rainfall patterns. Both short-duration, heavy rainfall events and long-duration, lower intensity rainfall can contribute to flooding incidents [49]. In the case of the Chellif watershed, a rainfall map was constructed utilizing a dataset spanning 10 years (2010-2021) derived from 13 rain gauges. Various interpolation methods were employed, including simple kriging, ordinary kriging, inverse distance weighting (IDW) with powers ranging from 1 to 5, a radial basis function (RBF) employing a completely regularized spline, and a spline with a tension kernel function. Among these methods, the IDW approach with a power value of "1" was selected due to its lowest root mean square error (RMSE) value. The resulting rainfall map was classified into 5 distinct classes that represent it in the next chapter.

Hillshade:

Hillshade is an important hydrological parameter that captures the shading and shadowing effects on a terrain model. It provides valuable information about slope and aspect, influencing water flow and accumulation patterns. Incorporating hillshade into analyses enhances our understanding of landscape dynamics and aids in various water management applications.

LULC:

Land use and land cover (LULC) exert a significant influence on surface runoff and sediment transportation, thereby directly impacting the frequency of flood occurrences (Benito et al., 2010). This influence stems from the fact that LULC patterns govern the generation of surface runoff and the extent of infiltration. Built-up areas, characterized by limited opportunities for water infiltration and high surface water generation, are more prone to frequent flooding incidents.Conversely,forested areas facilitate water infiltration, resulting in reduced flooding events [53].The relationship between flood events and vegetation density exhibits an inverse correlation when considering hydrological responses across different temporal scales[23].

Profile curvature:

The velocity of runoff can be influenced by the curvature of a profile aligned parallel to the direction of the steepest slope. The role of profile curvature as a significant factor in flood dynamics has been recognized by several researchers.Convex slopes tend to expedite overland flow and can impact processes such as infiltration, soil saturation, and surface runoff[1].

Plan curvature:

According to certain researchers, plan curvature, akin to profile curvature, is regarded as a significant factor in influencing floods . A convex slope contributes to the acceleration of overland flow and can have implications for processes such as infiltration, soil saturation, and surface runoff[2].

Stream density:

River density refers to the measurement of total stream length (m) within a specific area divided by the corresponding watershed area (km2). When comparing all other factors, areas with higher stream densities are generally more prone to experiencing flooding [55].Fraser and Schumer (2012) have observed a correlation between larger flood peaks and volumes with higher stream densities in perennial watersheds, whereas ephemeral watersheds tend to exhibit lower flood peaks. In order to visualize river density, a river density map was generated utilizing the "Line density" tool in ArcGIS 10.4, with the natural breaks classification method applied to select five distinct classes.

Distance from river:

The proximity to the river plays a crucial role in determining the areas most susceptible to flood inundation, with the most heavily affected regions typically located adjacent to the river . Distance from the river serves as a vital conditioning factor for assessing flood vulnerability within a basin, as it influences both the occurrence of flooding incidents and the interaction between river flow and other factors.
Greater distances from the river correspond to a reduced likelihood of flood events.At the basin scale, regional flooding can be attributed to the storage of terrestrial water[24].

Flow accumulation:

Flow accumulation is recognized as a significant conditioning factor for flood susceptibility mapping, according to certain researchers.This variable is derived from the combination of flow direction data and the calculation of the number of upstream pixels that contribute flow to each pixel. Specifically, each pixel is assigned a value equivalent to the number of pixels draining through it[35]. The calculation of flow accumulation was performed using the eight flow direction algorithm, and the "Arc Hydro Tools" in ArcGIS 10.4 were utilized for this purpose .

Stream power index (SPI):

The Stream Power Index (SPI) plays a crucial role in shaping fluvial systems (Knighton, 1999). It is calculated using Equation(III.1) :

$$\mathrm{SPI} = \mathrm{A} \cdot \tan(\beta) \tag{III.1}$$

A: flow accumulation

β(in radians) : the slope gradient

The SPI is indicative of both the sediment transport capacity and the erodibility of the stream bed, collectively referred to as the total SPI

Topographic wetness index (TWI):

The Topographic Wetness Index (TWI) is a significant factor in the occurrence of floods,as it spatially represents the variation in basin wetness[49].This index quantifies the water content within each pixel of the region and is computed using Equation (III.2):

$$\text{TWI} = \ln\left(\frac{\text{A}}{\tan(\beta)}\right) \tag{III.2}$$

A:flow accumulation(m^2/m)

β:denotes the slope gradient (in degree).

Higher TWI values are generally associated with increased likelihood of flooding [49].

NDVI:

The Normalized Difference Vegetation Index (NDVI) is a widely used metric for assessing the greenness of the land surface and detecting the presence of water bodies.Changes in NDVI serve as an indicator of variations in vegetation and surface water coverage over time[3], enabling the examination of the relationship between flooding and vegetation within a watershed.Higher vegetation densities are generally associated with lower probabilities of flooding in the study area . The NDVI metric ranges from +1, indicating the highest vegetation density, to -1, representing the lowest vegetation density[54].This index is produced using the combination of image bands 4 and 5 from the Landsat 8 OLI satellite It is calculated using Equation(III.3) :

$$\text{NDVI} = \frac{\text{Bande 5} - \text{Bande 4}}{\text{Bande 5} + \text{Bande 4}} \tag{III.3}$$

NDWI:

NDWI stands for Normalized Difference Water Index. It is an index used to assess the presence and extent of water bodies or the moisture content of vegetation.

Since water absorbs light in the visible to infrared electromagnetic range, the index wears out the green bands (Band 3) and near infrared (band 5) images from (Landsat 8/OLI). Bodies of water generally have a higher reflectance on the blue spectrum (0.4 - 0.5 µm) than on the green (0.5 -0.6 µm) and red (0.6 - 0.7 µm) spectrum) (Claudia, et al., 2019). The NDWI index is calculated using this equation:

$$\text{NDWI} = \frac{\text{Bande 3} - \text{Bande 5}}{\text{Bande 3} + \text{Bande 5}} \tag{III.4}$$

The index values range from -1 to +1, where higher values indicate a higher presence of water or moisture content. NDWI is commonly employed in various applications, including water resource management, drought monitoring, and flood mapping[41].

III.4 Machine Learning Models:

Machine learning is the process of solving practical problems by collecting a set of data and then constructing a statistical model algorithmically based on that data. The main objective of this chapter is to first introduce the tools and strategies of machine learning. Subsequently, the application of machine learning programs and algorithms for preparing a susceptibility map is used as an initial step in flood management in the Chellif Watershed.

Finally, the obtained results from the machine learning models are presented through the final susceptibility map. The results are interpreted, and the reliability of the employed models is discussed at the next chapter.

III.4.1 Definition

Machine Learning (ML) is a branch of modern computer science that involves analyzing databases or interacting with the environment using algorithms to identify patterns and make predictions based on past statistics. It is based on data exploration and pattern recognition to derive predictive analyses.

ML demonstrates its potential in cases where trends need to be identified within vast and diverse datasets, often referred to as big data. It is more effective in terms of accuracy and speed compared to traditional methods of analyzing massive datasets because ML algorithms can learn and adapt their results based on new acquired data without requiring reprogramming. While traditional analysis methods often struggle with immense volumes of data, ML, on the other hand, thrives when databases are growing, as it can learn and detect trends with significantly improved accuracy [6].

III.4.2 Establishing a Machine Learning Model

The four key processes involved in developing a machine learning model are as follows:

1. First, it is necessary to select and organize a dataset. These data will be fed into the machine learning model, which will then learn how to solve the problem it was designed for.

2. The next step involves choosing an algorithm to apply to the extracted training data. The choice of algorithm depends on two factors: the type and volume of the training data and the problem to be solved.

3. The algorithm then needs to be trained in an iterative process. Variables are passed to the algorithm, and the results are compared to what the algorithm should have produced. To increase the accuracy of the result, the variables can be modified before the algorithm is retrained until the desired outcome is achieved. Once trained, the algorithm takes the form of a machine learning model.

4. Finally, the model needs to be put to work and improved upon. The model is applied to new data, the source of which is determined by the current situation. The accuracy of the model may also vary over time

By following these processes, a machine learning model can be developed, trained, and applied to real-world data to solve specific problems. The continuous improvement and adaptation of the model contribute to its effectiveness and performance in various domains.

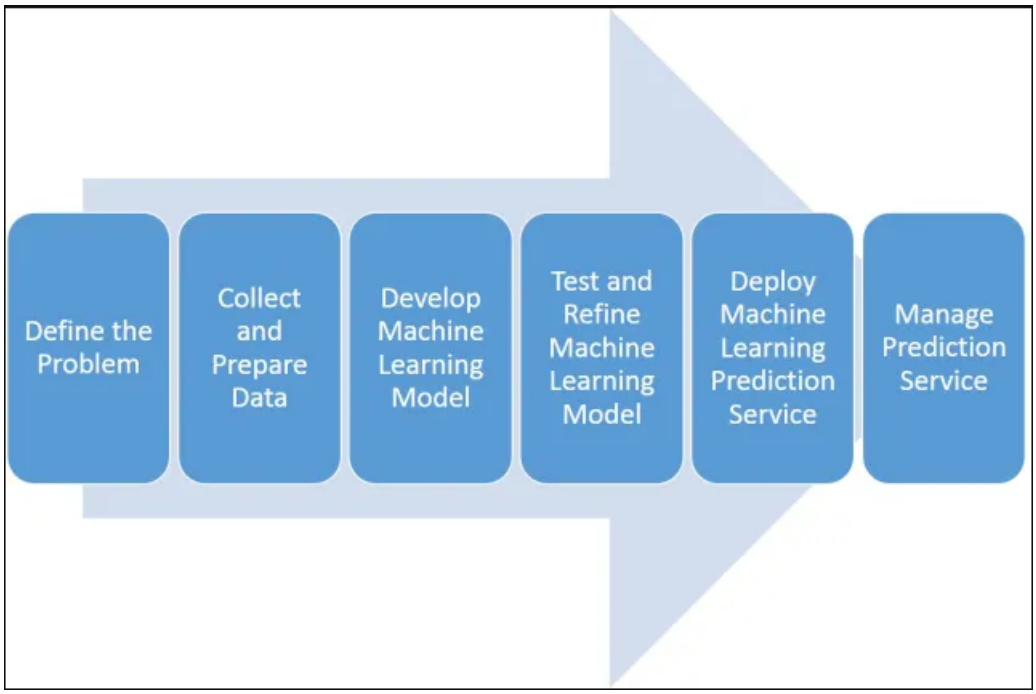


Figure III.2: Step of building a machine learning model

III.5 Machine Learning-Based Modeling in the Chellif Watershed

III.5.1 The models used

Among the most frequently used machine learning models that have shown their performance in flood mapping, we will choose three models following:

Light Gradient Boosting model (LGBM) :

LGBM is a boosting ensemble model that transforms weak learners into a powerful model. It was made open-source by Microsoft in 2017. LGBM improves the performance of Gradient Boosted Decision Trees (GBDT) models by accelerating runtime and reducing memory consumption while maintaining high accuracy[7] .Traditional GBDT-based models tend to experience decreased accuracy and slower forecasting speed when dealing with large volumes of data[15].
LGBM addresses these issues by employing a histogram-based algorithm that handles high-dimensional data effectively, speeds up computation, and prevents overfitting. This boosting technique involves transforming continuous floating-point values into integers and constructing a histogram shape with depth and width restrictions. Unlike XGB, LGBM utilizes a pre-sorted based Decision Trees (DT) technique and incorporates parallel learning through a parallel voting DT during the training process. This enables parallel learning for the model, where initial samples are distributed to multiple trees to select the top-k samples using Local Voting Decision (LVD). The global voting decision then collects the top-k LVD attributes to compute the top-2k attributes for k iterations. In the optimization process, LGBM employs the Leaf-wise method to find suitable leaves[40].

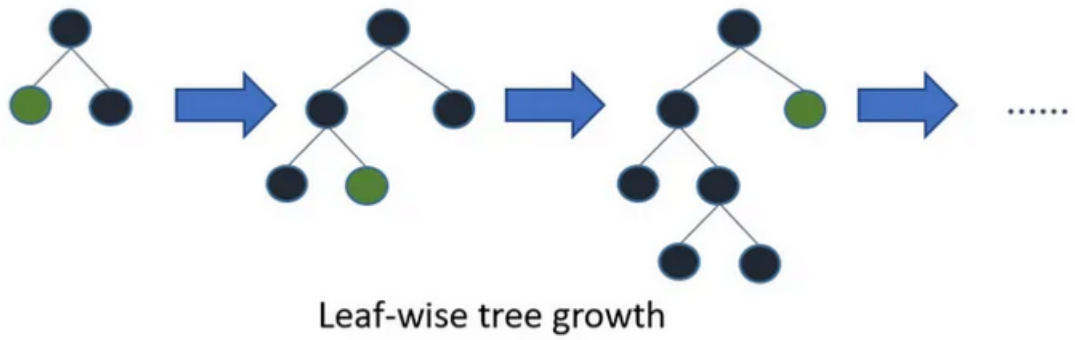


Figure III.3: Leaf wise growth

Adaptive Boosting (AdaBoost):

AdaBoost is machine learning technique initiated by Freund and Schapire[26], many algorithms are derived from AdaBoost either for classification or applied to regression .The AdaBoost algorithm is an iterative approach that seeks to construct a robust classifier

through the combination of weak learners generated in prior iterations. The algorithm modifies the learning pattern in accordance with the error returned by the weak learners, with the ultimate goal of achieving a final hypothesis that exhibits low error relative to a given distribution[27].

Stacking Model(AdaBoost and LGBM):

Stacking is one of the popular ensemble modeling techniques in machine learning. Various weak learners are ensemble in a parallel manner in such a way that by combining them with Meta learners, we can predict better predictions for the future.

This ensemble technique works by applying input of combined multiple weak learners' predictions and Meta learners so that a better output prediction model can be achieved.In stacking, an algorithm takes the outputs of sub-models as input and attempts to learn how to best combine the input predictions to make a better output prediction[43].

Stacking is also known as a stacked generalization and is an extended form of the Model Averaging Ensemble technique in which all sub-models equally participate as per their performance weights and build a new model with better predictions. This new model is stacked up on top of the others; this is the reason why it is named stacking.In the context of stacking AdaBoost model and LGBM model,it involves training both types of models separately and then combining their predictions to make a final prediction[45].
By stacking AdaBoost model and LGBM model, you can leverage the strengths of both types of models and potentially improve predictive performance. This approach allows for capturing complex temporal dependencies with AdaBoost and utilizing the boosting capabilities of LGBM models.

III.5.2 Measurement of performance and validation

The ROC curve :

The Receiver Operating Characteristic (ROC) curve, also known as the performance characteristic curve, is a graph that represents the performance of a classification model. This model aims to classify elements into two distinct categories based on one or more properties of these elements. In our study, we represent the positive pixels (flooded) and negative pixels (No-flooded), which are precisely classified into true positives, true negatives, false positives, and false negatives.
The ROC curve depicts the true positive rate (sensitivity) as a function of the false positive rate (specificity). The true positive rate and true negative rate are calculated using the following formulas[25].

The true positive rate:

$$\mathrm{TPR} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{III.5}$$

The true negative rate:

$$\mathrm{TNR} = \frac{\mathrm{TN}}{\mathrm{TN} + \mathrm{FP}} \tag{III.6}$$

Such as,

- TP: True Positive (The actual value was positive and the model predicted a positive value).

- FP: False Positive (The actual value was positive and the model predicted a negative value).

- TN: True Negative (The actual value was negative and the model predicted a negative value).

- FN: False Negative (The actual value was negative and the model predicted a positive value).

The AUC parameter:

The AUC (Area Under the Curve) is a parameter associated with the ROC curve. It represents the area under the ROC curve and provides a single scalar value to measure the overall performance of the classification model. The AUC ranges from 0 to 1, where a higher value indicates better discrimination ability. An AUC of 0.5 corresponds to a random classifier, while an AUC of 1 represents a perfect classifier that achieves a perfect trade-off between sensitivity and specificity. The AUC used to evaluate the overall performance of a model in data classification.
This parameter is calculated using the following formula:

$$\mathrm{AUC} = \frac{\mathrm{TN} + \mathrm{TP}}{\mathrm{P} + \mathrm{N}} \tag{III.7}$$

Such as,

- P: Positive points (flooded).

- N: Negative points (No-flooded).

Precession

Precision, in the context of machine learning models, is a metric that measures the proportion of correctly predicted positive instances (true positives) out of the total instances predicted as positive (true positives+false positives). In other words, precision focuses on the accuracy of the positive predictions made by the model.precision is calculated as:

$$\text{Precession} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{III.8}$$

III.6 The Modeling steps

III.6.1 Data Introduction

Firstly, we gather the data, which in this case are the factors influencing flooding, in a consolidated form so that they are all contained in a single table. Then, we proceed with preprocessing the data and removing any null values... etc. Successful learning models require good data quality.

III.6.2 Correlation Matrix Extraction

Finding the relationships between the training data (factors influencing flooding) allows us to see if a given factor is directly dependent on another, expressing the degree of correlation between the different factors. Afterwards, we eliminate one of the strongly correlated factors as they will have the same influence.

III.6.3 Data Split (Training, Validation)

In this step, we divide the data into two sets:

- The training data, which will be used to train the chosen algorithms.

- The validation data, which will be used to assess the performance of the results.

In general, according to several articles in flood mapping, 80% of the data is used for training and 20% is reserved for validation [55].

III.6.4 Application of Models

In this step, we apply the previously chosen algorithms. This is done by introducing the required data in a consolidated numerical format through the use of GIS tools, so that the algorithms can learn (training phase), and then interpret the remaining results (validation phase), including the interpretation of other parameters such as the weight of the factors.

III.6.5 Validation

The second component of the data, the validation dataset, is used in this validation phase. This subset of data refines the model by introducing data that the computer has not encountered during the learning phase. This allows for the evaluation of the model's performance in flood mapping. The models used will be validated using the ROC curve and the AUC parameter. The results can serve as decision support in the development

of future urban planning documents or, more broadly, in the optimization of urban or territorial projects in the cities of this watershed[19].

III.7 Conclusion

In this chapter, we have introduced the fundamental concepts that form the basis of our work, including geographical mapping, data collection, and modern decision-making support techniques derived from artificial intelligence. These concepts serve as the starting point for our study, and a well-defined working methodology is essential for the development of a susceptibility map and the identification of key factors influencing flooding. Each factor is presented along with its spatial distribution.

Chapter IV

Results and Discussion

IV.1 Introduction

In this chapter, we will present and interpret the results obtained in the form of spatial distribution maps of various factors and the final vulnerability map. Additionally, we will discuss the reliability of our models.

IV.2 Results

IV.2.1 Factors

We will present in the following the maps of the different factors influencing floods:

Elevation:

Altitude is considered as an important factor in mapping susceptible area of flash floods, since it affects the natural flow of water. Generally, higher altitude areas are essentially safe from flash flooding and lower altitude areas have high potential for inundation during the flash flood events[21] . The range of altitude varies from 0 to 1969 m (Figure IV.1) in the present studied area.

The elevation map illustrates that the topography of the studied region can be divided into four sections. The first section (elevation < 450m) is situated in the low coastal plains in the northwestern side of the basin . The second section (elevation 780m <elevation <1055m) is located in the central area (Figure IV.1). The third section (450m< elevation< 780m) is found in the northeastern side, overlapping with the first section. Lastly, the fourth section (1056m< elevation<1969m) is situated the southwestern side of the Chellif basin.

Slope:

The slope is a parameter that indicates the level of topographic variation in a given area. It plays a significant role in determining the velocity of water flow, as a steeper slope tends to increase the speed of surface runoff. Consequently, the risk of land erosion becomes more pronounced. Using GIS, a slope map was created, revealing that the southern and northern parts of the region consist of mild slopes (<3.88°). In contrast, the central and northeastern areas of the basin exhibit high slopes (ranging from 23.29° to 89.96°). The southwestern region, on the other hand, is characterized by relatively low slope values approach of chellif wadi (Figure IV.2).

Aspect:

The aspect indicates the direction of the elevation and the flow pattern ,it depends largely on the direction of the surface. Here, the elevation aspect map was classified into five

classes (Figure IV.3) :

- Plane
- North-East
- South-East
- South-Ouest
- North-East

Rainfall:

As it is shown in Figure (IV.4), the average precipitation is distributed across five levels, with a maximum of approximately 731.3 mm in the central part of the watershed, covering certain cities such as Djelfa and Tissimsilt. The minimum precipitation of about 328.6 mm is observed in the eastern part of the watershed. The rest of the watershed experiences an average precipitation ranging from 500 to 553 mm. It is noteworthy that precipitation is significantly higher in the west compared to the east.

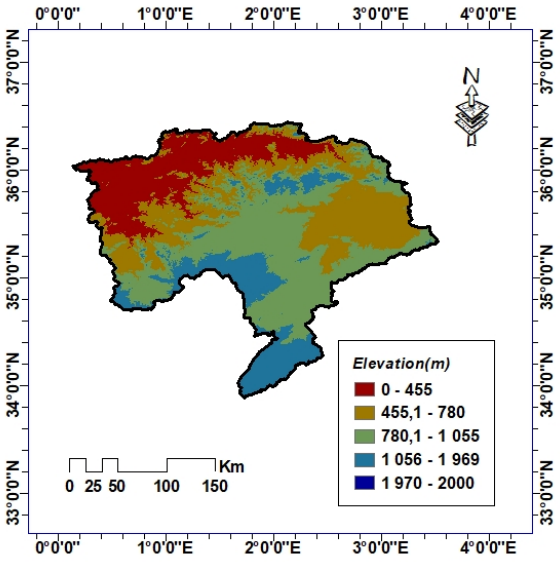


Figure IV.1: Elevation Map

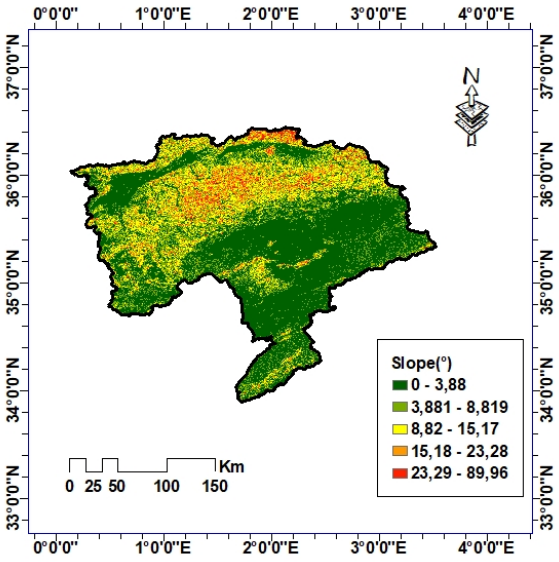


Figure IV.2: Slope Map

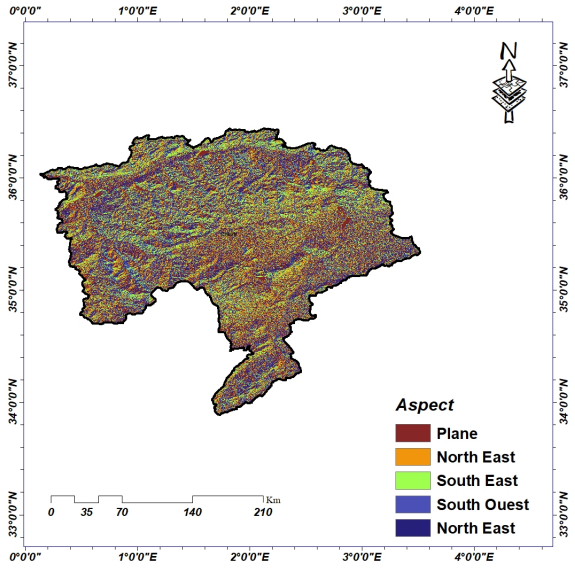


Figure IV.3: Aspect Map

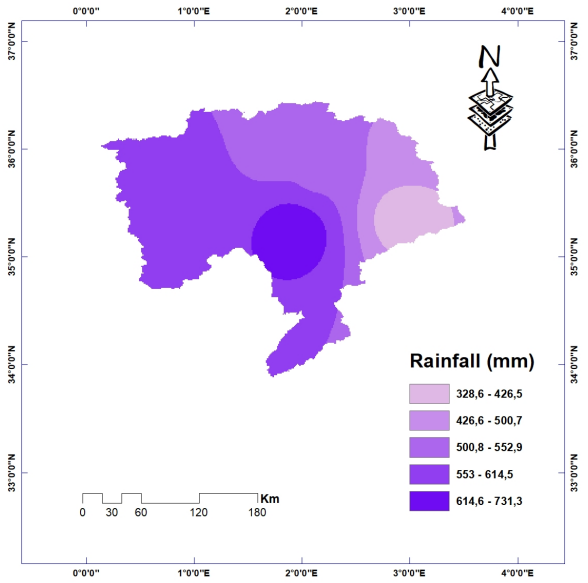


Figure IV.4: Rainfall Map

LULC:

land use/ landcover have been identified as factors that could impact floods Changes in land use/land cover may have a significant impact on flashfloods prone areas, as they affect various hydrological processes such as infiltration, evaporation, evapotranspiration, and runoff. This can either accelerate or decelerate the volume and velocity of water flowing over the surface during heavy rainfall events. In this study, the LULC map consists of several thematic maps, including water, trees and crops, built areas, bare ground, and rangeland as shown in figure(IV.5).

Hillshade:

hillshade illustrates slope shading. Just like the up north and central south regions are characterized by high shading values, indicating a convergence of water flow. The remaining watershed area is characterized by low to moderate shading values, resulting in a weak convergence of water flow in this part Figure (IV.6).

Profile curvature/Plane curvature:

Topographic characteristics of an area were basically understood by plan and profile curvature. Here, the plane and profile curvature value ranged from (−10.32 to 10.73) and (−13.33 to 12.14) (Figure IV.7).

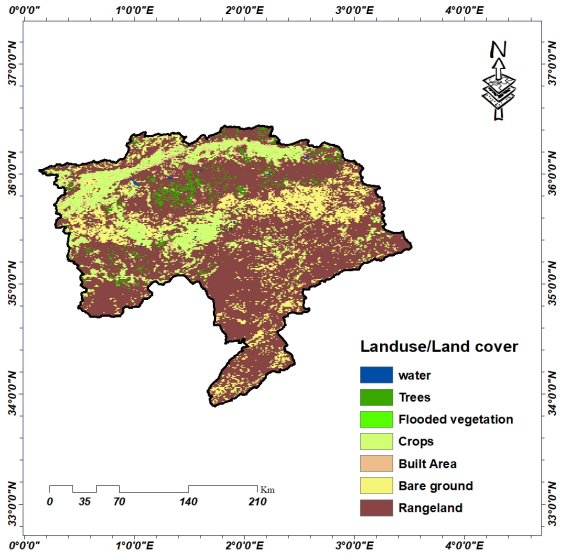


Figure IV.5: (Landuse/Landcover) Map

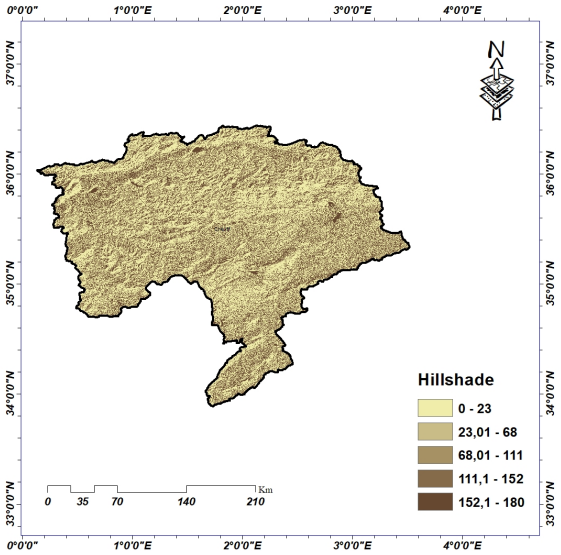
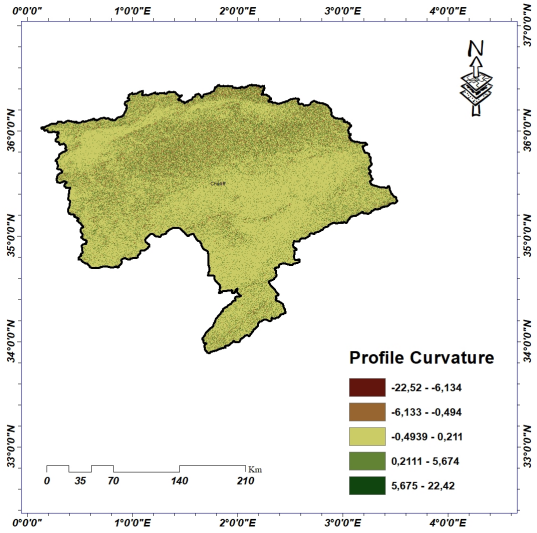


Figure IV.6: Hillshade Map



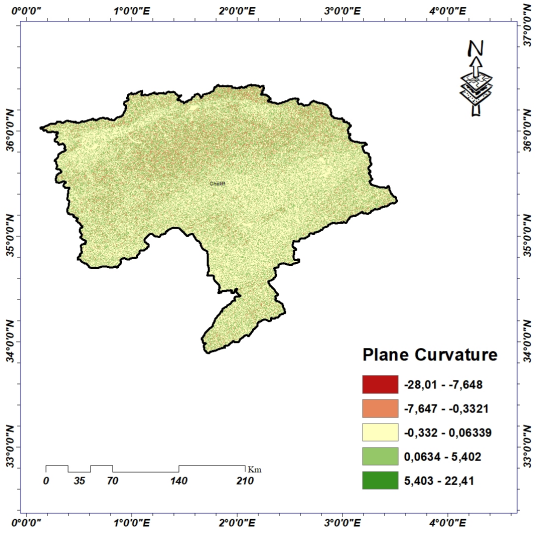


Figure IV.7: Profile curvature/Plan curavture Map

Stream density:

It appears that (Figure IV.8) illustrates that stream density values are high in the path of the main river Chellif wadi. Additionally, it shows that the density of the rivers is exceptionally high at the outlet, which is the point where the two main rivers of the catchment area converge. As one moves away from the rivers, the stream density gradually decreases until it reaches nearly zero values. This suggests that the concentration of streams is the greatest near the main river and diminishes as you move further away from it.

Distance from river:

Can significantly affect the severity and extent of water-induced flash flood. Many researchers consider the distance from the river as a key factor in assessing the flood risk. When a location is situated close to a river, it becomes more vulnerable to flood due to the increased water flow volume and velocity, which accelerate the process of flash floods. As a result, areas located in close proximity to a river are at a higher risk of experiencing flash flood in our region the distance from river values varies between (0 and 51 km) (see figure(IV.9))

Flow accumulation:

The topography of the land can contribute to flow accumulation, with low-lying areas or channels posing a higher risk of flooding. The range of flow accumulation varied from 0 m to 54 966000 in this study area (Figure IV.10).

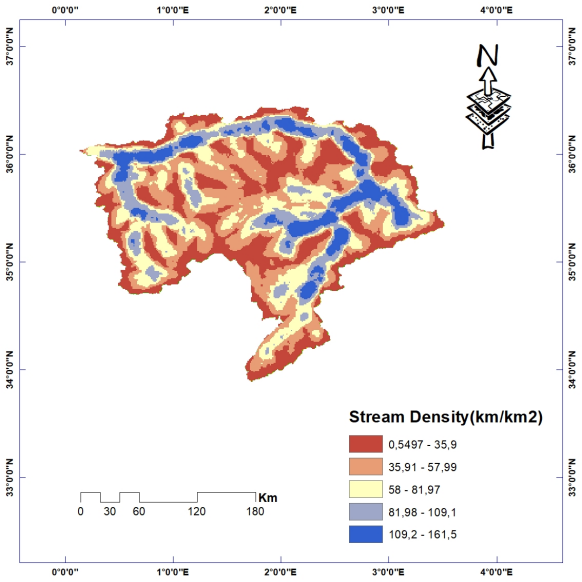


Figure IV.8: Stream density Map

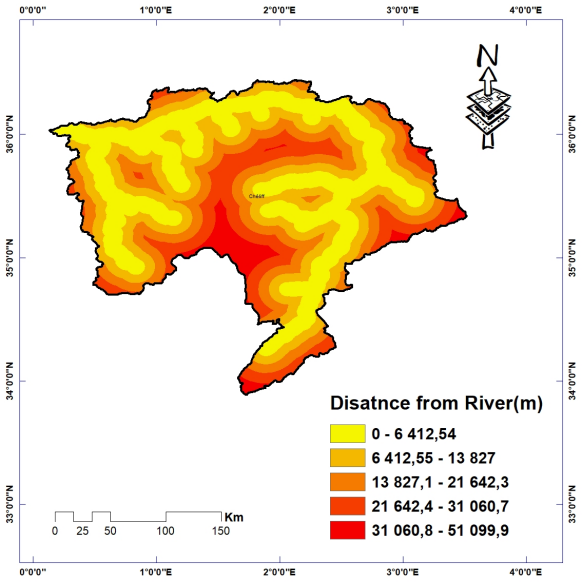


Figure IV.9: Distance from river Map

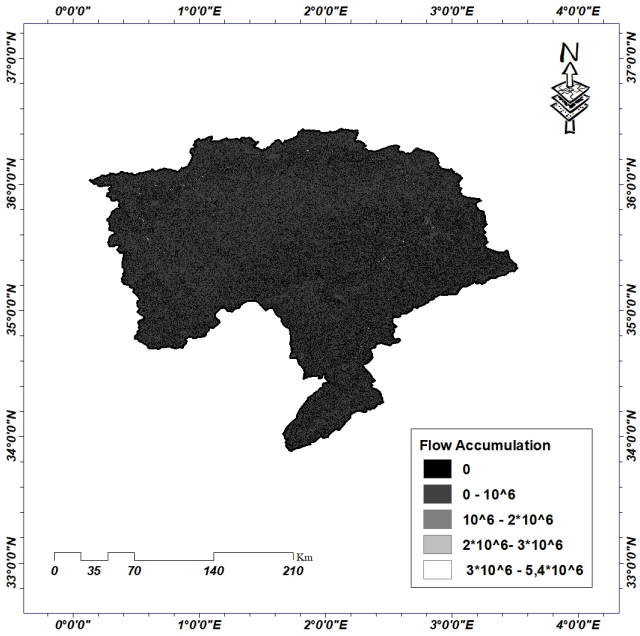


Figure IV.10: Flow accumulation Map

Stream power index (SPI):

stream power index represents a quantitative measure of the erosive power of flowing water in a river system which valuable for flood risk assesment. As can be observed, the values of this index are significant only at the river courses, while they are extremely low or even absent elsewhere(Figure (IV.11)).

Topographic wetness index (TWI):

It is used to evaluate the impact of the topography on the hydrological process. The TWI map shows that low TWI values (-15.4 ;-9.54) are present in the center and South regions of the basin; therefore, a weak humidity is present in the two regions. The moderate TWI values ranging from (-4.84 to -2.35) are present in North, South and South-east of the basin indicating the presence of an average humidity (Figure IV.12). A high humidity is detected in the outlet and the course of waterways due to the presence of runoff, the TWI values at this region varied from 1.54 to 16.5 in north and east.

NDVI:

According to Figure (IV.13), the study area can be divided into three main parts based on the Normalized Difference Vegetation Index (NDVI):

1. The first part showing values ranging from (-0.55 to 0.11), representing water bodies, urbanized areas, rocks, and bare soils, primarily located in the east and south

regions of the area where there is high human settlement.

2. The second part displays values ranging from (0.11 to 0.26), located primarily in the center of the watershed and representing various cultivated lands.

3. The third part demonstrates values rangin from(0.27to 0.79) generally indicating forested areas, public gardens, and agricultural zones, primarily situated in the north regions.

NDWI:

According to Figure (IV.14), the study area can be divided into three main parts based on the the Normalized Difference Water Index (NDWI):

1. The first part exhibits values ranging from (-0.79 to -0.22),which indicatesthe absence of water or the presence of non-water surfaces like land or built-up areas primarily located in the southern half of the region where there is high human settlement.

2. The second part displays values ranging from (-0.22 to 0.0018),represents urbanized areas and bare soils located in the north and in the east of the zone.

3.The third part demonstrates values ranging from (0.0018 to 0. 32) which indicate water bodies and areas where water stagnates .

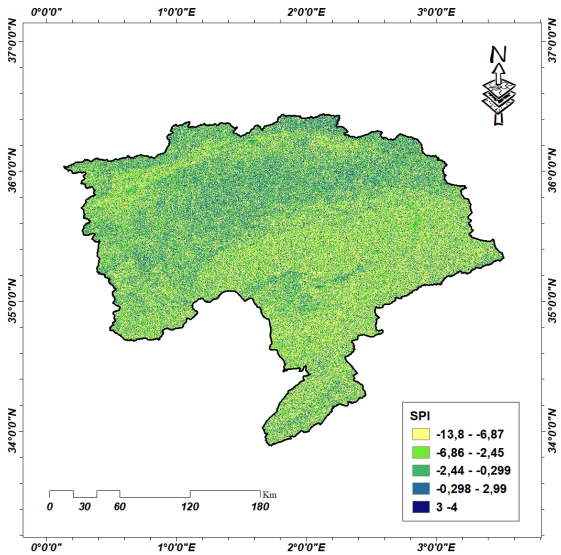


Figure IV.11: SPI Map

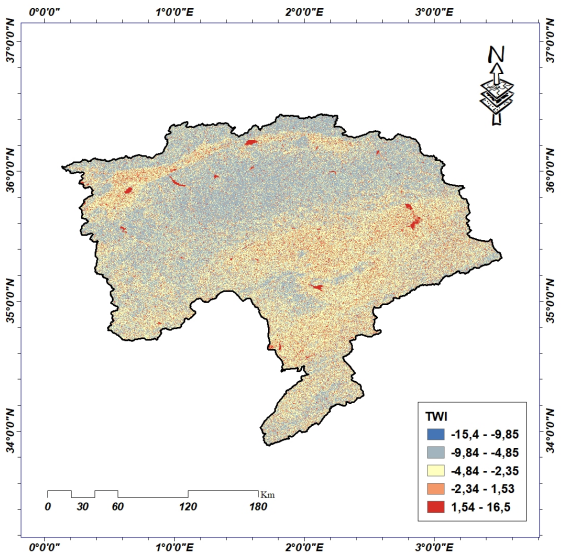


Figure IV.12: TWI Map

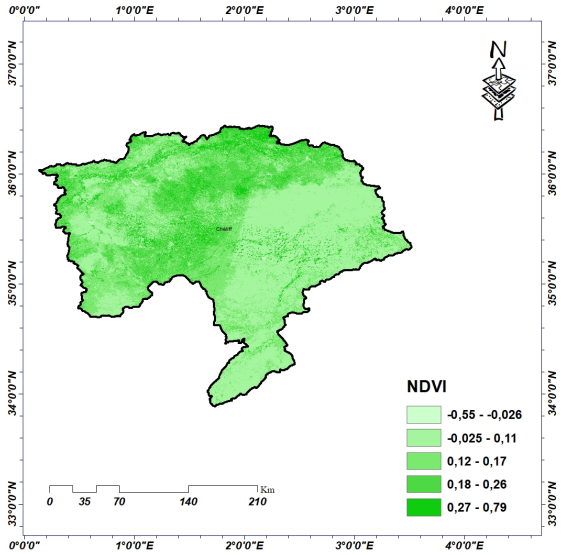


Figure IV.13: NDVI Map

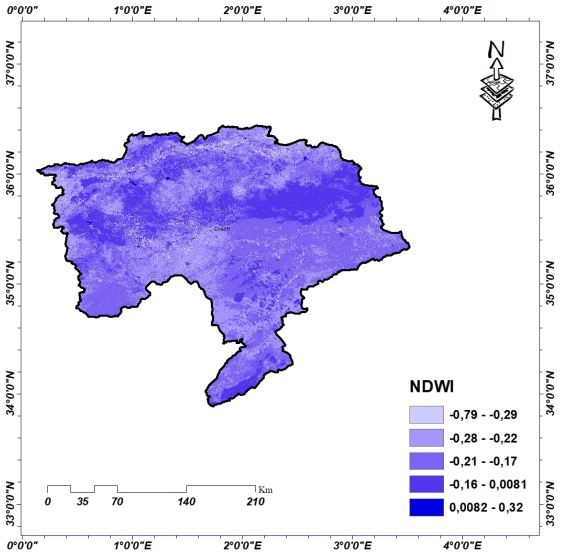


Figure IV.14: NDWI Map

IV.2.2 The importance of factors

Feature selection:

In order to select the most relevant variables for the classification model, a feature selection process was carried out considering 15 variables (elevation ,slope, aspect of elevation, Rainfall ,Hillshade ,Land use/Land cover(LULC),Profile curvature ,Plan curvature, Stream density ,Distance from river ,Flow accumulation, Stream power index (SPI),Topographic wetness index (TWI),vegetation cover (NDVI),Water Index(NDWI)).

Moreover, the top 10 variables were selected using feature importance scores obtained from LGBM. The feature importance scores were then calculated by evaluating the impact of each feature on the model's accuracy. The 10 variables with the highest feature importance scores were retained for the final classification model. This approach ensures that only the most informative variables are included in the model for predicting soil Flashflood in the Cheliff basin.

The selected factors are defined as follows: elevation ,slope, Rainfall , Profile curvature ,Stream density ,Distance from river, Stream power index (SPI),Topographic wetness index (TWI),vegetation cover (NDVI),Water Index(NDWI).

Table IV.1: Selection of variable

Ranking	**Features**	**Importances (%)**
1	TWI	15,58
2	SPI	14,76
3	Rainfall	11,78
4	Slope	10,90
5	Profile curvature	10,31
6	Distance from river	9,84
7	NDVI	8,73
8	Stream density	5,92
9	NDWI	4,45
10	Elevation	3,46
11	LULC	2,46
12	Aspect	1,11
13	Plan curvature	0,35
14	hillshade	0,35
15	Flow accumulation	0,00

Correlation Matrix:

A correlation matrix is a table containing correlation coefficients between variables, in our case these variables are the flood conditioning factors.

This matrix is used to assess the dependence between different conditioning factors at the same time, such that each cell represents a correlation value which is the degree of dependence between two conditioning factors, this value is between -1 and 1.To avoid the double effect of the same factors on the modelling, the Topographic wetness index(TWI) and Distance from river variables were eliminated based on their high correlation with Stream density and Stream Power Index (SPI) respectively.

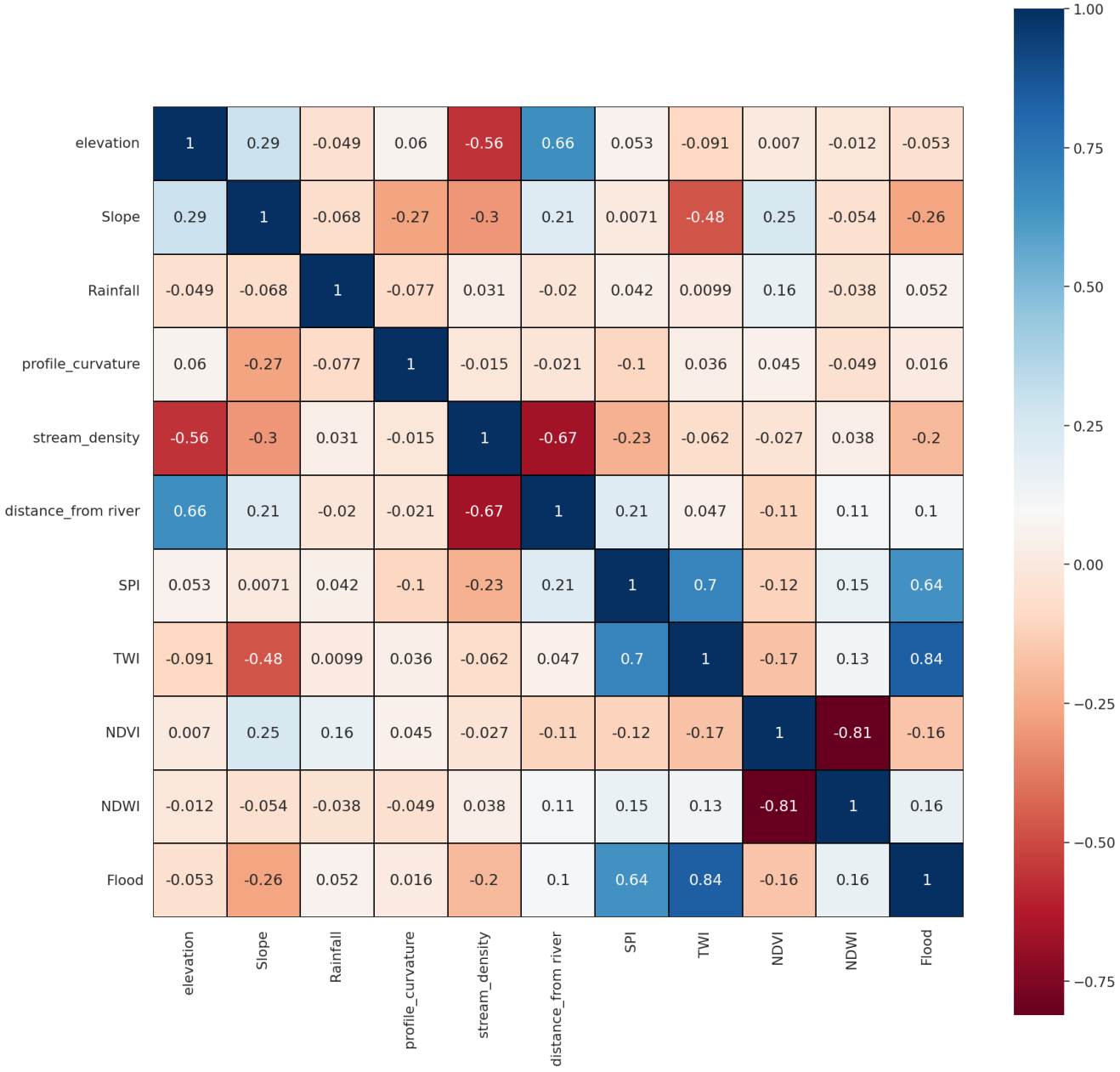


Figure IV.15: Correlation Matrix

The machine learning models have the ability to provide estimates of the importance of factors, which is one of their most useful features. In our case, we opted for the LGBM, AdaBoost and Stacking (LGBM-AdaBoost) models.

- For the LightGBM model, the most critical parameters were :SPI,Stream density,Rainfall,Profile curvature,NDWI,Slope,Elevation,NDVI

- For the AdaBoost model, the most critical parameters were :SPI, Rainfall,Stream density , Slope , Elevation, NDVI, NDWI,Profile curvature

• For the Stacking model, the most critical parameters were:SPI,Rainfall,Stream density, Profile curvature,NDWI,Slope,Elevation,NDVI

In summary, significant similarities could be observed in the rankings importance for the three models.

Table IV.2: Ranking of importance of factors

Factor Class	**LGBM**	**AdaBoost**	**Stacking**
1	SPI	SPI	SPI
2	Stream density	Rainfall	Rainfall
3	Rainfall	Stream density	Stream density
4	Profile curvature	Slope	Profile curvature
5	NDWI	Elevation	NDWI
6	Slope	NDVI	Slope
7	Elevation	NDWI	Elevation
8	NDVI	Profile curvature	NDVI

IV.2.3 The susceptibility map

According to the LGBM model, the susceptibility map presented in(figure IV.16) was generated using the quantile method for classification. It includes five classes of vulnerability, the spatial coverage of areas susceptible to very high and high flooding is respectively 10.26% and 23.97% of the total area of the zone. The rest of the zone is associated with regions of moderate, low and very low vulnerability. The percentages of area covered by these zones are respectively 29.71%, 24.40% and 11.63%

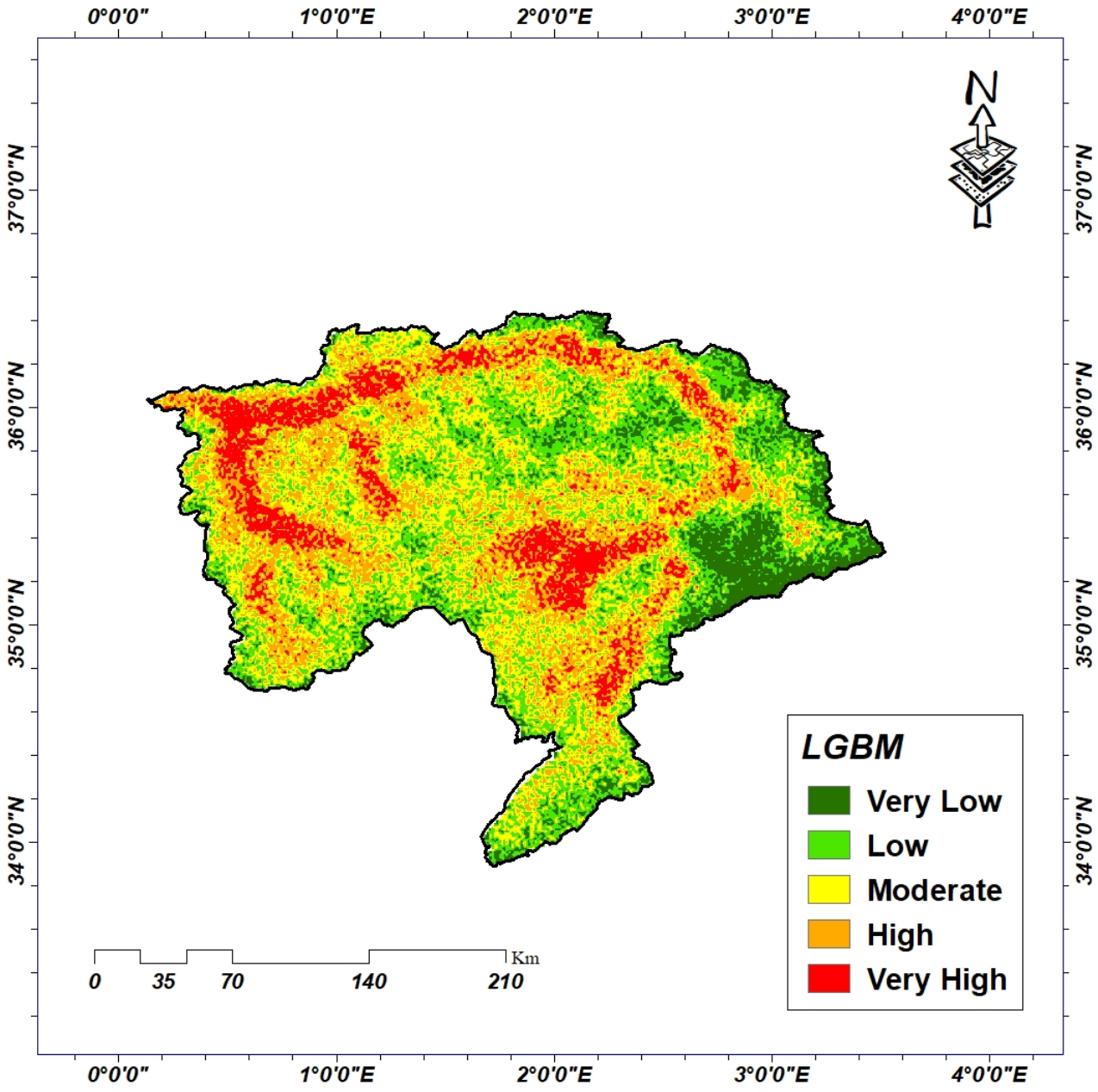


Figure IV.16: The flood susceptibility map with LGBM

According to the AdaBoost model, the susceptibility map presented in Figure (IV.17) below, was generated using, for classification, the method of quantile . The spatial coverage of areas susceptible to very high and high flooding is respectively 7.23% and 29.45% of the total area. These areas are mainly located in the region near the wadis, mainly the Chellif wadi . The rest of the zone is associated with regions of moderate, low and very low susceptibility. The percentages of the areas covered by these zones are respectively 27.38%, 18.97% and 16.94%.

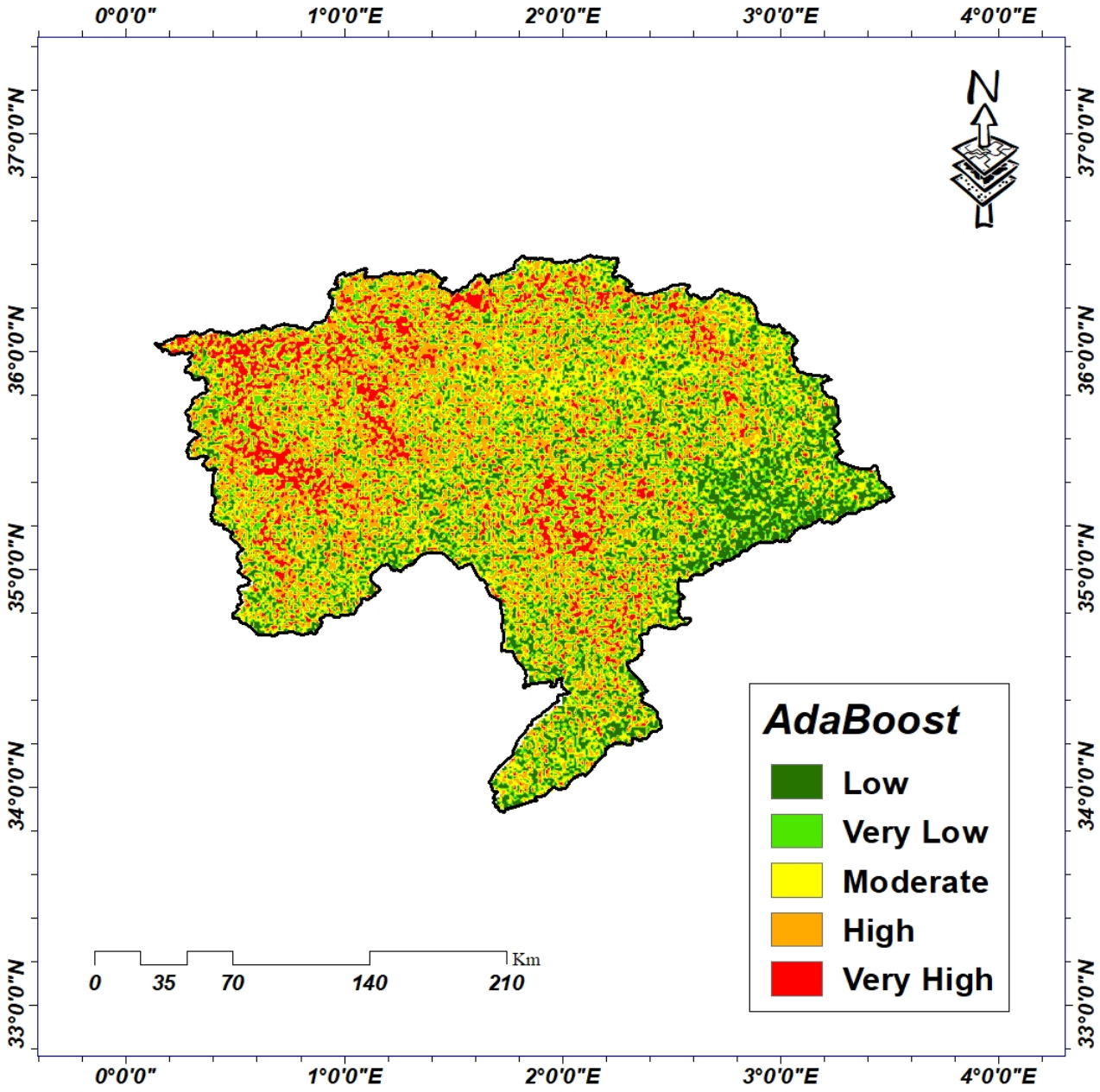


Figure IV.17: The flood susceptibility map with AdaBoost

According to the Stacking model, the susceptibility map presented by Figure(IV.18) below. It includes five classes of susceptibility.where the spatial coverage of areas susceptible to very high and high flooding are respectively 9.73% and 25.44% of the total area. The rest of the zone is associated with regions of moderate, low and very low susceptibility. The percentages of the areas covered by these zones are respectively 28.60%, 21.02% and 15.18%

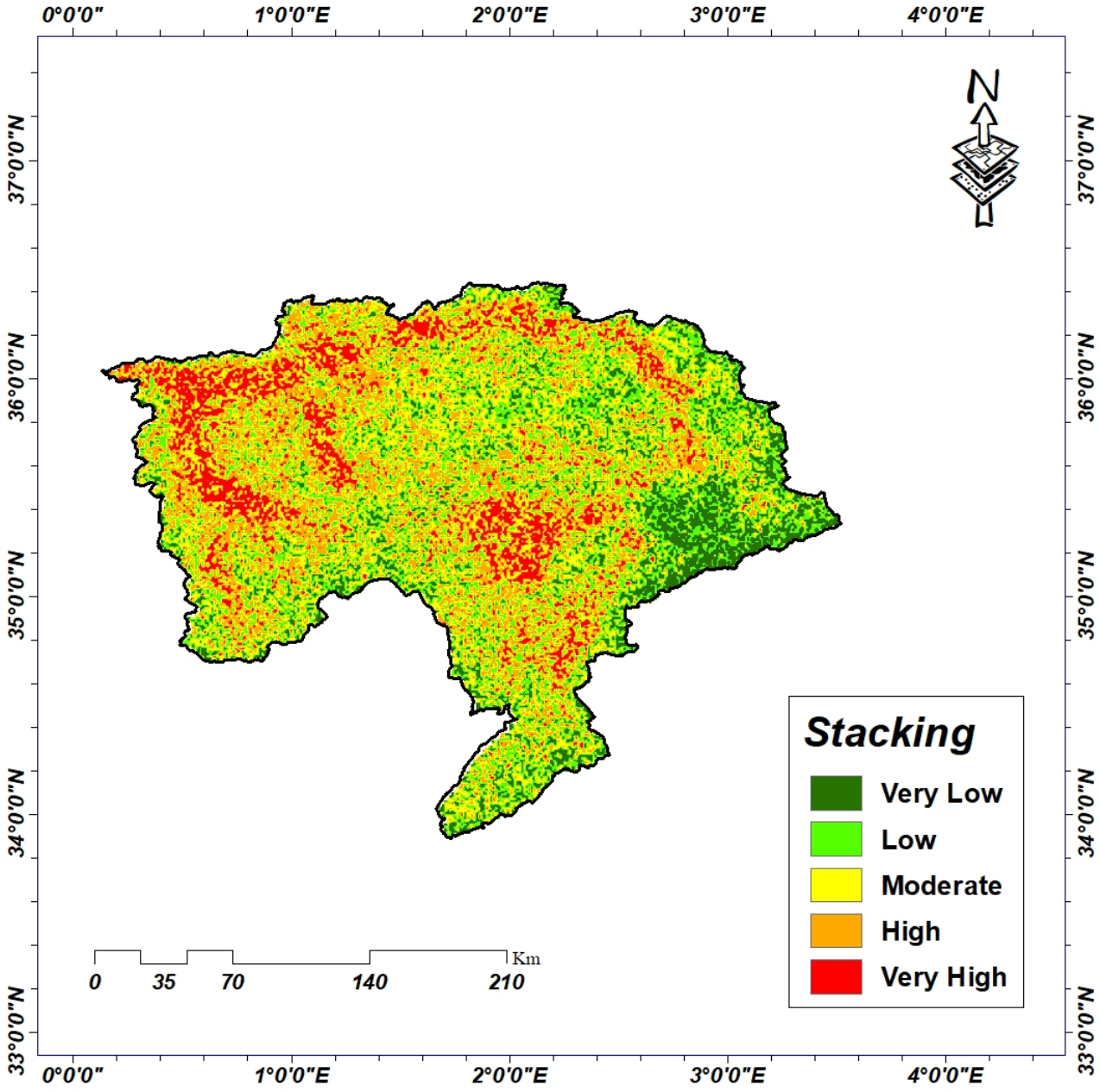


Figure IV.18: The flood susceptibility map with Stacking

Finally, the results are shown in the following table

Table IV.3: Flood Susceptibility Zone Classes

Model	Surface	Susceptibility class				
		Very High	High	Moderate	Low	Very Low
LGBM	%	10,26	23,97	29,71	24,40	11,63
AdaBoost	%	7,23	29,45	27,38	18,97	16,94
Stacking	%	9,73	25,44	28,60	21,02	15,18

IV.2.4 Validation

In the flood susceptibility analysis, it is important to locate areas likely to be affected by future floods. Whatever the method used for validation, it is important to validate susceptibility maps to unknown future floods[4]. Validation of model predictions was performed using the AUC value of the ROC (receiver operating characteristics) curve.

The ROC curve evaluates the prediction accuracy of each model by plotting the observed and predicted values. The ROC method calculates flood susceptibility map prediction success rates based on data of the historical flood events. The ROC AUC illustrates the accuracy of a prediction model by determining the ability of the technique to calculate the occurrence and non-occurrence of a flood from historical data in the predefined area [47].

AUC is also used to qualitatively assess the accuracy of flood susceptibility prediction[48]. The ROC curve results shown in Figures (IV.19), (IV.20) and (IV.21) indicated that the Stacking model (LGBM-AdaBoost) had the highest prediction accuracy (AUC = 0.9898) (Figure IV.16), followed by the LighGBM(AUC = 0.9857) (Figure IV.17), and finally the AdaBoost model (AUC = 0.9615) (Figure IV.18). The analysis of the AUC values showed that the Stacking model was more efficient than the others. was more efficient than the others. The relationship between the AUC and the prediction accuracies of the flood susceptibility map can be described in the following table[4]:

Table IV.4: Classification of performance with AUC values

AUC value	Model Performance
0,6-0,7	Average
0,7-0,8	Good
0,8-0,9	very good
0,9-1,0	excellent

Therefore the accuracy of the prediction of the Stacking(LGBM-AdaBoost),LGBM and AdaBoost models is excellent.

Table IV.5: Performance of the three models

	LGBM	**AdaBoost**	**Stacking**
Sensitivity(TPR)	0,980	1	0,968
Specificity(TNR)	0,942	0,92	0,990
precession	0,943	0,927	0,989
Accuracy	0,961	0,961	0,979

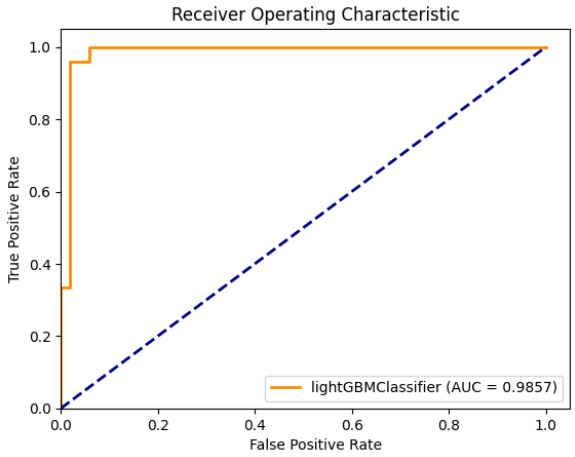


Figure IV.19: The ROC Curve with LGBM model

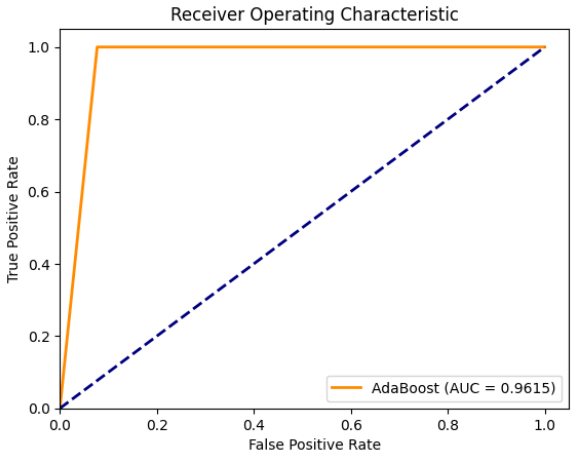


Figure IV.20: The ROC Curve with AdaBoost model

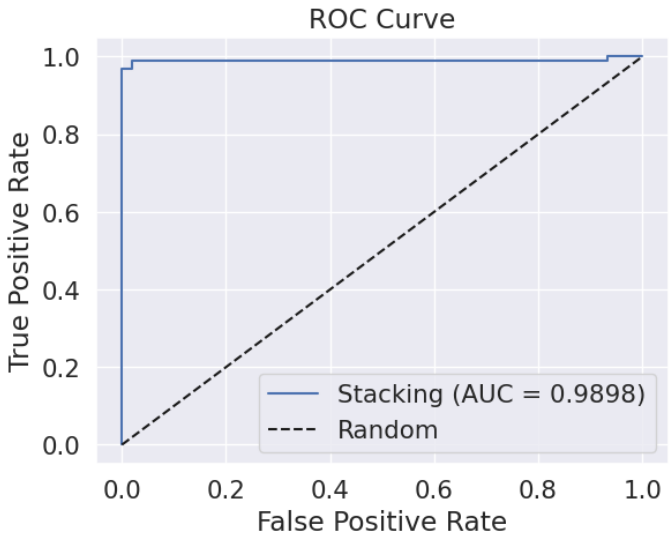


Figure IV.21: The ROC Curve with Stacking model

IV.3 Discussion

While it is impossible to completely avoid sudden floods, the most crucial measures that can be implemented focus on predicting future floods. Areas prone to flooding can be identified, and suitable structural and non-structural measures can be applied to mitigate flood losses. The models implemented for flood mapping in this study are LightGBM, AdaBoost and Stacking . In this study, the flood susceptibility map that we produced was based on three different models that were evaluated and compared.

Furthermore,the most influencial factors are: SPI, Stream density, Rainfall, Profile curvature, NDWI, Slope,elevation, NDVI.Flood mapping identifies areas that are prone to flooding based on considering factors that affect floods, such as geological, hydrological, topographical, and morphological conditions at the local level[55]. Based on the obtained maps, it is evident that the area's most susceptible to flooding, as indicated by the three models used, are typically found in regions characterized by a high stream density and substantial rainfall .The flood susceptibility maps generated by our study reveal that the areas likely to be affected by a very high level of risk of flooding are areas closer to wadi chlef. According to the LGBM model, the flood-prone areas represent 10.26% of the total basin area. The AdaBoost model indicates that these areas represent 7.23% of the total surface. Lastly, the stacking ensemble model shows that these areas account for 9.73% of the total surface.

Validation using the area under the curve (AUC) method yielded performance scores of 98% and 96% for LGBM and AdaBoost models, Moreover, it is important to consider that the mapping of flashfloods-susceptible areas using AI techniques reveals variations in the machine's understanding of the phenomena[32]. To further improve the accuracy, a stacking method was employed, which combines the previous machine learning techniques in a hybrid model. This stacking approach yielded exceptionally high results, with an accuracy reaching 99%. By leveraging the strengths of multiple models, the stacking method enhances the predictive power and reliability of the flashflood susceptibility mapping process[11].

Overall, the integration of AI techniques and the application of the stacking method have proven to be effective in accurately identifying flashflood-prone areas and can reduce the impact of flooding and risk of financial and economic losses in the future.

IV.4 Conclusion

In this chapter, to generate a flood risk map displaying different susceptibility classes and serving as a crucial predictive tool for managing this phenomenon, we employed the LGBM algorithm, AdaBoost and Stacking algorithms .Based on the AUC values, the Stacking model demonstrated the best predictive performance, achieving an AUC of 0.99.

This study will enable us to make informed decisions in the future regarding the development and planning within the Cheliff watershed

GENERAL CONCLUSION

This research highlights the significant potential of machine learning and tools, as well as meta-models, in enhancing the identification, visualization, and interpretation of flashfloods susceptible areas. Focusing on the Cheliff basin in Algeria, this study successfully generated flashfloods susceptibility maps using tree distinct machine learning algorithms: LightGBM, AdaBoost, and stacking ensemble methods. The findings shed light on the factors influencing flashfloods-prone areas in the Cheliff basin, key parameters are : SPI, Stream density ,Rainfall ,Profile curvature ,NDWI, Slope ,Elevation, NDVI.

Furthermore, the utilization of stacking ensemble methods exhibited exceptional accuracy and significantly improved the prediction and mapping of flashfloods-susceptible areas. By combining predictions from multiple base models, the stacking ensemble approach provided a more robust and reliable estimation of flashfloods susceptibility. This hybrid model effectively leveraged the strengths of individual machine learning algorithms while mitigating their weaknesses, resulting in a highly accurate and comprehensive assessment of flashfloods risks. The reliable flashfloods susceptibility maps generated in this study serve as invaluable tools for decision-makers and government officials involved in flashfloods risk management. The integration of machine learning techniques, along with the stacking ensemble method, offers a promising approach to better delineate, visualize, and interpret flashfloods-prone areas.

Future research work can focus on further testing the performance of the developed models by incorporating more varied features, such as climate change,river morphology and antecedents conditions, into the input datasets. This will enhance the models' capability to predict and understand floods and flash floods. Additionally, evaluating the efficacy of the models in diverse locations across the country can provide insights into their generalizability and guide necessary modifications to adapt to different scenarios. To streamline the model development process, researchers can develop programming tools like widgets and scripts that simplify the collection of crucial input datasets. Furthermore, it is important to investigate both the direct and indirect costs associated with floods and flash floods, as they often result in significant economic losses. Research studies focusing on economic cost analysis can provide valuable information for stakeholders and support informed decision-making. Collaboration among researchers, stakeholders, and funding agencies is crucial to support these research efforts and advance our understanding of flood modeling, leading to effective flood risk management and sustainable practices.

Bibliography

[1] M Ahmadlou, M Karimi, S Alizadeh, A Shirzadi, D Parvinnejhad, H Shahabi, and M Panahi. Flood susceptibility assessment using integration of adaptive network-based fuzzy inference system (anfis) and biogeography-based optimization (bbo) and bat algorithms (ba). *Geocarto International*, 34(11):1252–1272, 2019.

[2] Mohammad Ahmadlou, A'kif Al-Fugara, Abdel Rahman Al-Shabeeb, Aman Arora, Rida Al-Adamat, Quoc Bao Pham, Nadhir Al-Ansari, Nguyen Thi Thuy Linh, and Hedieh Sajedi. Flood susceptibility mapping and assessment using a novel deep learning model combining multilayer perceptron and autoencoder neural networks. *Journal of Flood Risk Management*, 14(1):e12683, 2021.

[3] Kazi Rifat Ahmed and Simu Akter. Analysis of landcover change in southwest bengal delta due to floods by ndvi, ndwi and k-means cluster with landsat multi-spectral surface reflectance satellite data. *Remote Sensing Applications: Society and Environment*, 8:168–181, 2017.

[4] Ahmed EM Al-Juaidi, Ayman M Nassar, and Omar EM Al-Juaidi. Evaluation of flood susceptibility mapping using logistic regression and gis conditioning factors. *Arabian Journal of Geosciences*, 11:1–10, 2018.

[5] Raffaele Albano, Aurelia Sole, Domenica Mirauda, and Jan Adamowski. Modelling large floating bodies in urban area flash-floods via a smoothed particle hydrodynamics model. *Journal of Hydrology*, 541:344–358, 2016.

[6] Ethem Alpaydin. *Introduction to machine learning*. MIT press, 2020.

[7] Fatimah Alzamzami, Mohamad Hoda, and Abdulmotaleb El Saddik. Light gradient boosting machine for general sentiment classification on short texts: a comparative evaluation. *IEEE access*, 8:101840–101858, 2020.

[8] Richard E Bellman and Lotfi Asker Zadeh. Decision-making in a fuzzy environment. *Management science*, 17(4):B–141, 1970.

[9] Roberto Bentivoglio, Elvin Isufi, Sebastian Nicolaas Jonkman, and Riccardo Taormina. Deep learning methods for flood mapping: a review of existing applications and future research directions. *Hydrology and Earth System Sciences*, 26(16):4345–4378, 2022.

[10] Ali Bouamrane, Oussama Derdous, Noura Dahri, Salah-Eddine Tachi, Khereddine Boutebba, and Mohamed T Bouziane. A comparison of the analytical hierarchy

process and the fuzzy logic approach for flood susceptibility mapping in a semi-arid ungauged basin (biskra basin: Algeria). *International Journal of River Basin Management*, 20(2):203–213, 2022.

[11] Hamza Bouguerra, Salah Eddine Tachi, Hamza Bouchehed, Gordon Gilja, Nadir Aloui, Yacine Hasnaoui, Abdelmalek Aliche, Saâdia Benmamar, and Jose Navarro-Pedreño. Integration of high-accuracy geospatial data and machine learning approaches for soil erosion susceptibility mapping in the mediterranean region: A case study of the macta basin, algeria. *Sustainability*, 15(13):10388, 2023.

[12] Hamid Bourenane, Youcef Bouhadad, and Mohamed Said Guettouche. Flood hazard mapping in urban area using the hydrogeomorphological approach: case study of the boumerzoug and rhumel alluvial plains (constantine city, ne algeria). *Journal of African Earth Sciences*, 160:103602, 2019.

[13] Isabelle Braud, P-A Ayral, Christophe Bouvier, F Branger, G Delrieu, J Le Coz, G Nord, J-P Vandervaere, S Anquetin, M Adamovic, et al. Multi-scale hydrometeorological observation and modelling for flash flood understanding. *Hydrology and Earth System Sciences*, 18(9):3733–3761, 2014.

[14] Dieu Tien Bui, Phuong-Thao Thi Ngo, Tien Dat Pham, Abolfazl Jaafari, Nguyen Quang Minh, Pham Viet Hoa, and Pijush Samui. A novel hybrid approach based on a swarm intelligence optimized extreme learning machine for flash flood susceptibility mapping. *Catena*, 179:184–196, 2019.

[15] Wentao Cai, Ruihua Wei, Lihong Xu, and Xiaotao Ding. A method for modelling greenhouse temperature using gradient boost decision tree. *Information Processing in Agriculture*, 9(3):343–354, 2022.

[16] Kamran Chapi, Vijay P Singh, Ataollah Shirzadi, Himan Shahabi, Dieu Tien Bui, Binh Thai Pham, and Khabat Khosravi. A novel hybrid artificial intelligence approach for flood susceptibility assessment. *Environmental modelling & software*, 95:229–245, 2017.

[17] Rosemary Charlton, Rowan Fealy, Sonja Moore, John Sweeney, and Conor Murphy. Assessing the impact of climate change on water supply and flood hazard in ireland using statistical downscaling and hydrological modelling techniques. *Climatic change*, 74(4):475–491, 2006.

[18] Bahram Choubin, Ehsan Moradi, Mohammad Golshan, Jan Adamowski, Farzaneh Sajedi-Hosseini, and Amir Mosavi. An ensemble prediction of flood susceptibility using multivariate discriminant analysis, classification and regression trees, and support vector machines. *Science of the Total Environment*, 651:2087–2096, 2019.

[19] Romulus Costache. 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, 2019.

[20] Kishor Dandapat and Gopal Krishna Panda. Flood vulnerability analysis and risk assessment using analytical hierarchy process. *Modeling Earth Systems and Environment*, 3:1627–1646, 2017.

[21] Subhasish Das, S-H Lee, Pawan Kumar, Ki-Hyun Kim, Sang Soo Lee, and Satya Sundar Bhattacharya. Solid waste management: Scope and the challenge of sustainability. *Journal of cleaner production*, 228:658–678, 2019.

[22] Sumit Das. Geographic information system and ahp-based flood hazard zonation of vaitarna basin, maharashtra, india. *Arabian Journal of Geosciences*, 11(19):576, 2018.

[23] Esmaeel Dodangeh, Bahram Choubin, Ahmad Najafi Eigdir, Narjes Nabipour, Mehdi Panahi, Shahaboddin Shamshirband, and Amir Mosavi. Integrated machine learning methods with resampling algorithms for flood susceptibility prediction. *Science of the Total Environment*, 705:135983, 2020.

[24] Zhice Fang, Yi Wang, Ling Peng, and Haoyuan Hong. Integration of convolutional neural network and conventional machine learning classifiers for landslide susceptibility mapping. *Computers & Geosciences*, 139:104470, 2020.

[25] Tom Fawcett. An introduction to roc analysis. *Pattern recognition letters*, 27(8):861–874, 2006.

[26] Yoav Freund, Robert Schapire, and Naoki Abe. A short introduction to boosting. *Journal-Japanese Society For Artificial Intelligence*, 14(771-780):1612, 1999.

[27] Yoav Freund and Robert E Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of computer and system sciences*, 55(1):119–139, 1997.

[28] Eric Gaume, Valerie Bain, Pietro Bernardara, Olivier Newinger, Mihai Barbuc, Allen Bateman, Lotta Blaškovičová, Günter Blöschl, Marco Borga, Alexandru Dumitrescu, et al. A compilation of data on european flash floods. *Journal of Hydrology*, 367(1-2):70–78, 2009.

[29] Hang Ha, Chinh Luu, Quynh Duy Bui, Duy-Hoa Pham, Tung Hoang, Viet-Phuong Nguyen, Minh Tuan Vu, and Binh Thai Pham. Flash flood susceptibility prediction mapping for a road network using hybrid machine learning models. *Natural hazards*, 109(1):1247–1270, 2021.

[30] Amna M Handhal, Shaymaa M Jawad, and Alaa M Al-Abadi. Gis-based machine learning models for mapping tar mat zones in upper part (dj unit) of zubair formation in north rumaila supergiant oil field, southern iraq. *Journal of Petroleum Science and Engineering*, 178:559–574, 2019.

[31] HAP Hapuarachchi, QJ Wang, and TC Pagano. A review of advances in flash flood forecasting. *Hydrological processes*, 25(18):2771–2784, 2011.

[32] Ioanna Ilia, Paraskevas Tsangaratos, Ploutarchos Tzampoglou, Wei Chen, and Haoyuan Hong. Flash flood susceptibility mapping using stacking ensemble machine learning models. *Geocarto International*, 37(27):15010–15036, 2022.

[33] Abu Reza Md Towfiqul Islam, Swapan Talukdar, Susanta Mahato, Sonali Kundu, Kutub Uddin Eibek, Quoc Bao Pham, Alban Kuriqi, and Nguyen Thi Thuy Linh. Flood susceptibility modelling using advanced ensemble machine learning models. *Geoscience Frontiers*, 12(3):101075, 2021.

[34] Khabat Khosravi, Himan Shahabi, Binh Thai Pham, Jan Adamowski, Ataollah Shirzadi, Biswajeet Pradhan, Jie Dou, Hai-Bang Ly, Gyula Gróf, Huu Loc Ho, et al. A comparative assessment of flood susceptibility modeling using multi-criteria decision-making analysis and machine learning methods. *Journal of Hydrology*, 573:311–323, 2019.

[35] Masoud Bakhtyari Kia, Saied Pirasteh, Biswajeet Pradhan, Ahmad Rodzi Mahmud, Wan Nor Azmin Sulaiman, and Abbas Moradi. An artificial neural network model for flood simulation using gis: Johor river basin, malaysia. *Environmental earth sciences*, 67:251–264, 2012.

[36] Guetter Elmekkedem Korieb Hamza. Etude des transports sédimentaires de l'oued chéliff et leur évolution au fil des dernières décennies, sous l'influence de la variabilité climatique. 2018.

[37] Wenjing Li, Kairong Lin, Tongtiegang Zhao, Tian Lan, Xiaohong Chen, Hongwei Du, and Haiyan Chen. Risk assessment and sensitivity analysis of flash floods in ungauged basins using coupled hydrologic and hydrodynamic models. *Journal of Hydrology*, 572:108–120, 2019.

[38] Shakeel Mahmood and Atta-ur Rahman. Flash flood susceptibility modeling using geo-morphometric and hydrological approaches in panjkora basin, eastern hindu kush, pakistan. *Environmental earth sciences*, 78:1–16, 2019.

[39] Lorenzo Marchi, Marco Borga, Emanuele Preciso, and Eric Gaume. Characterisation of selected extreme flash floods in europe and implications for flood risk management. *Journal of Hydrology*, 394(1-2):118–133, 2010.

[40] Mohamed Massaoudi, Shady S Refaat, Ines Chihi, Mohamed Trabelsi, Fakhreddine S Oueslati, and Haitham Abu-Rub. A novel stacked generalization ensemble-based hybrid lgbm-xgb-mlp model for short-term load forecasting. *Energy*, 214:118874, 2021.

[41] Stuart K McFeeters. The use of the normalized difference water index (ndwi) in the delineation of open water features. *International journal of remote sensing*, 17(7):1425–1432, 1996.

[42] Mahdi Panahi, Esmaeel Dodangeh, Fatemeh Rezaie, Khabat Khosravi, Hiep Van Le, Moung-Jin Lee, Saro Lee, and Binh Thai Pham. Flood spatial prediction modeling using a hybrid of meta-optimization and support vector regression modeling. *Catena*, 199:105114, 2021.

[43] Bohdan Pavlyshenko. Using stacking approaches for machine learning models. In *2018 IEEE second international conference on data stream mining & processing (DSMP)*, pages 255–258. IEEE, 2018.

[44] Margaret S Petersen. Impacts of flash floods. *Coping with flash floods*, pages 11–13, 2001.

[45] Binh Thai Pham, Mohammadtaghi Avand, Saeid Janizadeh, Tran Van Phong, Nadhir Al-Ansari, Lanh Si Ho, Sumit Das, Hiep Van Le, Ata Amini, Saeid Khosrobeigi Bozchaloei, et al. Gis based hybrid computational approaches for flash flood susceptibility assessment. *Water*, 12(3):683, 2020.

[46] Binh Thai Pham, Chinh Luu, Tran Van Phong, Phan Trong Trinh, Ataollah Shirzadi, Somayeh Renoud, Shahrokh Asadi, Hiep Van Le, Jason von Meding, and John J Clague. Can deep learning algorithms outperform benchmark machine learning algorithms in flood susceptibility modeling? *Journal of hydrology*, 592:125615, 2021.

[47] Zohre Sadat Pourtaghi and Hamid Reza Pourghasemi. Gis-based groundwater spring potential assessment and mapping in the birjand township, southern khorasan province, iran. *Hydrogeol J*, 22(3):643–662, 2014.

[48] Biswajeet Pradhan. Landslide susceptibility mapping of a catchment area using frequency ratio, fuzzy logic and multivariate logistic regression approaches. *Journal of the Indian Society of Remote Sensing*, 38:301–320, 2010.

[49] Himan Shahabi, Ataollah Shirzadi, Somayeh Ronoud, Shahrokh Asadi, Binh Thai Pham, Fatemeh Mansouripour, Marten Geertsema, John J Clague, and Dieu Tien Bui. Flash flood susceptibility mapping using a novel deep learning model based on deep belief network, back propagation and genetic algorithm. *Geoscience Frontiers*, 12(3):101100, 2021.

[50] Hariklia D Skilodimou, George D Bathrellos, Konstantinos Chousianitis, Ahmed M Youssef, and Biswajeet Pradhan. Multi-hazard assessment modeling via multi-criteria analysis and gis: a case study. *Environmental Earth Sciences*, 78:1–21, 2019.

[51] BELATTAR Souhir et al. *Quantification du transport solide dans le bassin versant de Cheliff Nord-Ouest de l'Algérie.* PhD thesis, university center of abdalhafid boussouf-MILA, 2021.

[52] Jose Candido Stevaux, Hudson de Azevedo Macedo, Mario Luis Assine, and Aguinaldo Silva. Changing fluvial styles and backwater flooding along the upper paraguay river plains in the brazilian pantanal wetland. *Geomorphology*, 350:106906, 2020.

[53] Swapan Talukdar, Bonosri Ghose, Shahfahad, Roquia Salam, Susanta Mahato, Quoc Bao Pham, Nguyen Thi Thuy Linh, Romulus Costache, and Mohammadtaghi Avand. Flood susceptibility modeling in teesta river basin, bangladesh using novel ensembles of bagging algorithms. *Stochastic Environmental Research and Risk Assessment*, 34:2277–2300, 2020.

[54] Mahyat Shafapour Tehrany, Biswajeet Pradhan, and Mustafa Neamah Jebur. Spatial prediction of flood susceptible areas using rule based decision tree (dt) and a novel ensemble bivariate and multivariate statistical models in gis. *Journal of hydrology*, 504:69–79, 2013.

[55] Mahyat Shafapour Tehrany, Biswajeet Pradhan, Shattri Mansor, and Noordin Ahmad. Flood susceptibility assessment using gis-based support vector machine model with different kernel types. *Catena*, 125:91–101, 2015.

[56] Kashif Ullah and Jiquan Zhang. Gis-based flood hazard mapping using relative frequency ratio method: A case study of panjkora river basin, eastern hindu kush, pakistan. *Plos one*, 15(3):e0229153, 2020.

[57] Yun Xing, Qiuhua Liang, Gang Wang, Xiaodong Ming, and Xilin Xia. City-scale hydrodynamic modelling of urban flash floods: the issues of scale and resolution. *Natural Hazards*, 96:473–496, 2019.