ASSESSMENT OF FLASH FLOOD HAZARDS IN A SEMIARID AREA THROUGH SATELLITE AND SOCIAL MEDIA DATA MINING

by
Eidah Ghurm Alzahrani
A Dissertation
Submitted to the
Graduate Faculty
of
George Mason University
in Partial Fulfillment of
The Requirements for the Degree
of
Doctor of Philosophy
Earth Systems and Geoinformation Sciences

Committee:

______________________________ Dr. Donglian Sun, Dissertation Director

______________________________ Dr. John J. Qu, Committee Member

______________________________ Dr. Ruixin Yang, Committee Member

______________________________ Dr. Celso Ferreira, Committee Member

______________________________ Dr. Dieter Pfoser, Department Chairperson

______________________________ Dr. Donna M. Fox, Associate Dean, Office of Student Affairs & Special Programs, College of Science

______________________________ Dr. Fernando R. Miralles-Wilhelm Dean, College of Science

Date: ______________________________ Spring 2022

George Mason University Fairfax, VA

ASSESSMENT OF FLASH FLOOD HAZARDS IN A SEMIARID AREA THROUGH SATELLITE AND SOCIAL MEDIA DATA MINING

A Dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at George Mason University

by

Eidah Ghurm Alzahrani
Master of Science
Department of Geography/ Geographic Information Sciences, University of Akron, 2015
Bachelor of Social Sciences
Department of Geography and a minor degree in Education, Umm AL-Qura University, 2006

Director: Donglian Sun, Professor
Department of Geography and Geoinformation Science

Spring Semester 2022
George Mason University
Fairfax, VA

DEDICATION

This is dedicated to the soul of my father, Ghurm,
to my loving mother, Aisha,
to my wonderful brothers and sisters,
and with my merciful heart to my advisor, Dr. Donglian Sun.

ACKNOWLEDGMENTS

I would like to thank the many friends, relatives, and supporters who have made this happen, and Dr. Donglian Sun for her invaluable guidance, encouragement, and support. I thank Dr. John J. Qu, Dr. Ruixin Yang, and Dr. Celso Ferreira for their invaluable help and nice advice. I thank my brothers for their encouragement and financial support. I thank Dr. Fadi AL-Ayed for his incredible support, for giving me essential advice from his experience, and for assisting me in my research. Finally, I thank the Fenwick Library for providing a clean, quiet, and well-equipped repository in which to work.

TABLE OF CONTENTS

LIST OF TABLES

Table Page

LIST OF FIGURES

LIST OF EQUATIONS

LIST OF ABBREVIATIONS AND/OR SYMBOLS

Topographic wetness index……………………………………………………………TWI
Stream power index……………………………………………………………………. SPI
Digital elevation model…………………………………………………………….DEM
Geographic Information Systems………………………………………………………GIS
Global Man-made Impervious Surface………………………………………………GMIS
Analytical Hierarchy Process………………………………………………………….AHP
Normalized difference vegetation index……………………………………………. NDVI
Frequency ratio………………………………………………………………………. FR
Logistic regression……………………………………………………………………LR
The Global Land Survey……………………………………………………………… GLS
Convolutional Neural Networks………………………………………………………CCN

ABSTRACT

ASSESSMENT OF FLASH FLOOD HAZARDS IN A SEMIARID AREA THROUGH SATELLITE AND SOCIAL MEDIA DATA MINING

Eidah Ghurm Alzahrani, Ph.D.

George Mason University, 2022

Dissertation Director: Dr. Donglian Sun

Since 2009, flood events have shown an increasing trend in Saudi Arabia. Moreover, most floods occur in cities and may be related to rapid urbanization. Meanwhile, flooding in semiarid areas is usually short-term flash floods within small areas. Therefore, satellite-based flash flood observations are still challenging, while conventional observations are usually sparse in semiarid areas.

This study combines machine learning techniques, the statistical analysis of frequency ratio, the logistic regression, and the analytic hierarchy process (AHP) method to identify flood-prone areas in a semiarid area in southern Saudi Arabia. This study integrates thirteen flood-contributing factors such as rainfall, elevation, aspect, slope, flow accumulation, stream power index (SPI), topographic wetness index (TWI), drainage density, distance from the river, distance from roads, soil types, urban area as represented by impervious area, and normalized difference vegetation index (NDVI). Ground observations from social media, such as Twitter and YouTube, validate the prediction results. The objectives of this study include: First, analyze the impacts of the selected thirteen flood-contributing factors. Second, build a decision-tree model between a flash flood and the influencing factors.

Third, create a flood susceptibility map in southern Saudi Arabia using the AHP method. The susceptibility map shows the levels of flood risk and their respective percentages in the study area: very low 5%, low 44%, moderate 39%, high 1%, and very high 11%. The results are validated against the ground observations from social media, such as Twitter and YouTube. This research indicates 30.76% commission error and 35.71% omission error from the derived flood susceptibility map with very high and high flood risks, while the overall accuracy can reach 90.37%.

CHAPTER ONE INTRODUCTION

Section One Problem Description

Floods are a threat to human life in various parts of the world due to increasing fluctuation in the world's weather patterns. Floods have been observed in areas with adequate information about occurrence patterns. Still, predicting these floods has become a challenge (Wolman, 2001).

In addition to the unpredictability of some of these devastating floods, little research has been focused on arid areas that experience little or no rainfall. These areas pose a challenge and a risk of tragic floods due to the lack of a flood hazard map. In recent years, Saudi Arabia has been experiencing a high risk of flash floods due to extreme weather conditions with heavy rainfall. Meteorologists warned that heavy wind, thunder, and rain would affect Jazan, Baha heights, and Asir (Subyani, 2016). In addition, rapid urbanization often involves removing vegetation, soil, and depressions from the land surface (El Alfy, 2016). The permeable soil is replaced by impermeable surfaces such as roads, roofs, parking lots, and sidewalks, reducing water infiltration into the ground and increasing runoff to streams. As a result, the peak discharge, volume, and frequency of floods increase. Floods have led to property damage and loss of life. Multiple uncertainties exist in the optimal flood-control decision-making process to deal with the threats associated with climate hazards and uncontrolled urbanization (Zhu et al., 2016).

The peak discharge of a flood is influenced by many factors, including the intensity and duration of storms and rainfall, the topography of streams, the vegetation, the soil types, and the hydrologic conditions preceding storm events (Wu et al., 2010). This research will investigate the factors that may have caused floods in Saudi Arabia from 2009 to 2020. In addition, artificial intelligence and machine learning will investigate the conditions under which heavy rainfall may cause a flash flood in Saudi Arabia, especially in areas of urbanization (Abdulrazzak et al., 2019).

Floods can be mapped and monitored using remotely sensed data acquired by satellites. Also, remote sensing provides valuable data and observations that may compensate for the sparse data from field surveys and gauging stations, especially in remote areas and developing countries (de Cunha, 2012).

Section Two Flood in Saudi Arabia

Between 2009 and 2021, the Kingdom of Saudi Arabia was exposed to many floods in separate areas; some of these floods were monitored by the Civil Defense Department, which counts the number of deaths and injuries as well as property losses such as drifting cars, damaged homes, power outages, the collapse of roads, and the closure of schools (Alrehaili, 2021).

Table 1 below lists the floods recorded in Saudi Arabia that significantly impacted the country's population and economic resources. This list contains twenty floods that occurred at separate times, but most were during the rainy months of November, December, January, and May. The southern and southwestern parts of the Kingdom of Saudi Arabia are affected by the southwest monsoon, which contributes to the heavy rainfall in those parts of the Kingdom of Saudi Arabia.

Table 1 shows that floods occurred more than once in some areas, causing property and human losses. For example, floods occurred in the Makkah region and the Jeddah region in 2009, 2011, 2015, and 2017. Bahrawi et al. (2020) found that the flood characteristics of 2009-2011 showed an increased risk of flooding in watersheds, and the death rate was also high. For example, in the November 2009 flood in Jeddah, there were 122 deaths recorded, 350 people disappeared, and the number of cars swept away by the flood was estimated at 3,000 vehicles. In 2011, a flood killed ten people, and approximately 1,500 people took refuge. In contrast, the Asir region recorded the lowest deaths: only one person died, and ten people were injured. However, the business damages alone were reported to be about 1 billion Saudi Riyal (AL-Bassam et al., 2014).

In 2019, the city of Hafar Al-Batin was hit by a flood, and about seven people were lost, 11 people were injured, and 40 cars were submerged in water. In addition, about 1,100 people suffered property losses. All these floods were due to heavy rain for only 3 hours. The highest rain rate was about 115.5 mm per hour in Jeddah in 2017. Also, in 2015, the city of Jeddah was exposed to 23 mm of rain in half an hour. All these data were collected by the Civil Defense Administration, which contributed to the rescue.

Table 1. A list of flash floods in the Kingdom of Saudi Arabia from 2009 - 2021

Number	Date	Location	Material and human losses	Reason	Source
1	6 February 2021	Tabuk	Four people were rescued	heavy rainfall	General Directorate of Civil Defense (CDD)
2	6 February 2021	Hafr AL-Batin	22 people were injured.	heavy rainfall	General Directorate of Civil Defense (CDD)
3	27 April 2021	Makkah	Some streets in the Holy Capital were turned into floods	heavy rainfall	General Directorate of Civil Defense (CDD)
4	25 July to 6 August 2020	Makkah Al Madinah ‘Asir Jizan Al Bahah.	77 people were rescued and about 600 were evacuated from their homes and moved to temporary shelter. Three people died.	heavy rainfall	General Directorate of Civil Defense (CDD)
5	6 February 2021	Tabuk, Streets of Tabuk City.	14 people were rescued, and a vehicle was swept away in the flood.	heavy rainfall	General Directorate of Civil Defense (CDD)
6	6 February 2021	Hafr Al-Batin	7 people died; 11 people were injured. 16 people were rescued from vehicles. 3 buildings were damaged, along with about 40 cars. About 1,100 people were affected.	strong winds and heavy rainfall dumped 43mm of rain in just 30 minutes	General Directorate of Civil Defense (CDD)
7	22-23 May 2019	Southwest Saudi Arabia, Jazan and Najran Regions.	1 person was missing. The effect was mostly people in vehicles trapped in flood water. Volunteers searched for the missing person for 3 days.	heavy rainfall landslides	General Directorate of Civil Defense (CDD)
8	08 -10 February 2019	Madinah Region	111 people were rescued, many of them from vehicles stranded in flooding wadis. Homes were damaged for 14 families.	heavy rainfall	General Directorate of Civil Defense (CDD)

9	27 -28 January 2019	Tabuk, Jawf, Madinah, and Makkah regions.	One person died. in Madinah. About 30 people were evacuated. Schools were closed.	heavy rainfall	General Directorate of Civil Defense (CDD)
10	21 November 2017	Jeddah, Saudi Arabia.	4 people died. 481 people were rescued. 40 people in 10 families were evacuated. Streets were under water up to 50 cm deep in cities, causing major traffic problems.	heavy rainfall powerful storms 115.5 mm (4.5 inches) per hour.	NASA's Global Precipitation Measurement Mission
11	16 February 2017	Dammam, Eastern Province	The railway line Saudi Railways was damaged. 18 people were injured.	heavy rainfall	General Directorate of Civil Defense (CDD)
12	14-15 February 2017	Asir region	1 person died and at least 10 were injured. Emergency crews responded to a total of 914 calls. 280 rescued were in vehicles at the time.	heavy rainfall	General Directorate of Civil Defense (CDD)
13	27- 30 November 2016	Riyadh, Eastern Province, Asir, Jizan, Makkah and Tabuk.	8 people died. 120 people were evacuated. 120334 people Rescued 334 people	heavy rainfall	General Directorate of Civil Defense (CDD)
14	23 March 2015	Riyadh	11 were killed, and 3 were missing. Over 400 vehicles were trapped in floodwater. 300 people had to be rescued from the floods.	heavy rainfall	General Directorate of Civil Defense (CDD)
15	17 November 2015	Western, northern, and central parts of Jeddah	12 people died in Jeddah after they were electrocuted by power cables falling into floodwater. 16 people were trapped in the floods. Floodwater was up to 1 meter deep.	heavy rainfall 23 mm of rain in just 15 minutes	General Directorate of Civil Defense (CDD)
16	24-25 November 2015	Riyadh and Al-Qassim Regions.	Schools were closed; roads were blocked.	heavy rainfall	General Directorate of Civil Defense (CDD))
17	4 January 2014	(Ha'yel) Province,	3 children died. Emergency services received nearly 1000 calls. Schools were closed	heavy rainfall over the last two days,	General Directorate of Civil

		Northwest Saudi Arabia.			Defense (CDD)
18	16 November 2013	Riyadh	Four people were reported dead and 10 were missing. Schools were closed and residents were urged to stay indoors. Emergency teams rescued 1,357 people. Power to parts of the city of 5 million was knocked out.	heavy rainfall 74mm	the official SPA news agency reported.
19	13January 14 2011	Jeddah	10 people died, and 1,500 people were missing. Shelter and relief were necessary for more than 1,500 families.	11cm of heavy rainfall in 3 hours	General Directorate of Civil Defense (CDD)
20	25th November 2009.	Jeddah	122 people were killed, and more than 350 were missing. At least 3,000 vehicles were swept away or damaged. Some roads were under a meter (three feet) of water. Business losses were estimated at a billion riyals (US$270 million).	heavy rainfall More than 70 millimeters of rain fell in Jeddah in just four hours on Wednesday 25	General Directorate of Civil Defense (CDD)

Particular indentation was found that floods recurred in recent years by measuring the number of floods in Saudi Arabia using data released on the flood-list website. In addition, flood-related information has increased on social media platforms such as Twitter, YouTube, and Facebook.

Floods were more common between 2009 and 2021, as seen in Table 2. Furthermore, the Flood Frequency occurrence between 2009 and 2021 shows a rise in the number of floods, with the year 2021 recording four floods between February and April. According to the Saudi Civil Defense Directorate, most of these occurrences occurred in Tabuk, Harf Al-Batin, and Makkah.

Table 2. Flood Frequency in Saudi Arabia from 2009 – 2021

Year	Flood Frequency
2009	1
2011	1
2013	1
2014	1
2015	3
2016	1
2017	3
2019	3
2020	2
2021	4

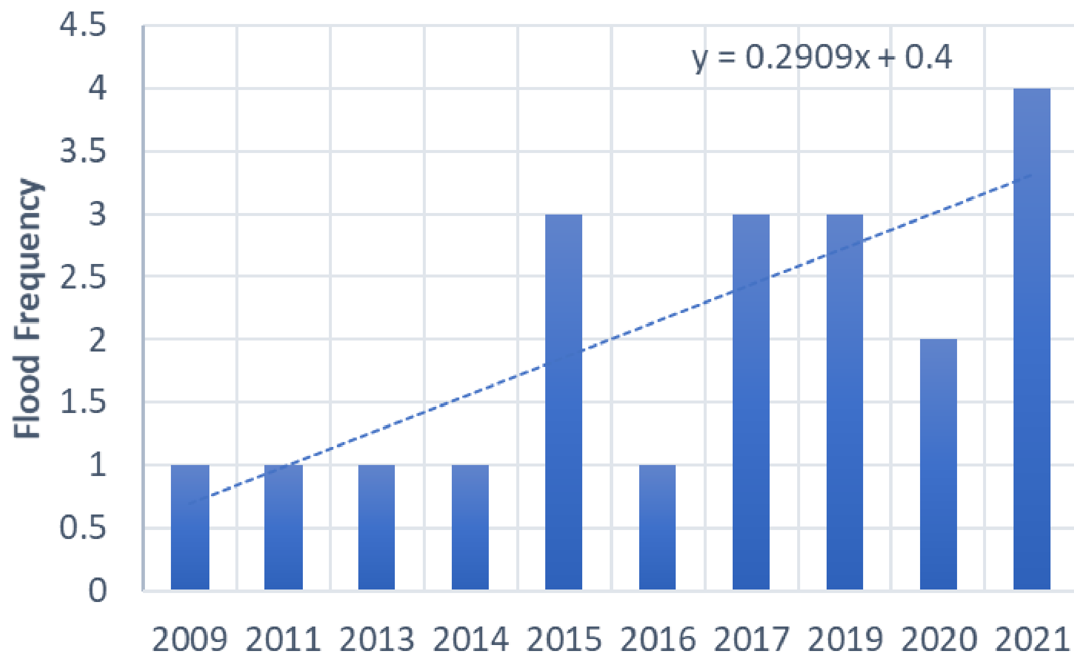

Figure 1. Flood Frequency in Saudi Arabia Between 2009 - 2021

Table 3 also shows the number of human losses that accompanied floods from 2009 to 2021, and it was found that the number of deaths was high in 2009, 2011, and 2015. Still, there was a noticeable decrease in the following years due to improved rescue efforts in Saudi Arabia. In addition, some issues were also addressed, particularly in Jeddah, by the design of water canals and dams and the removal of several slums.

As for the number of people rescued, the numbers were still high because of the change in the areas of flood occurrence. Many floods have occurred in northern areas such as Tabuk and southern and southwestern areas such as Najran, Jizan, Al Baha, and Aysar; these areas have received a high amount of rain in recent years.

Table 3. The number of Human Life Losses Due to Floods in Saudi Arabia Between 2009 – 2020

Year	People Rescued	People Died
2009	350	122
2011	1500	10
2013	10	4
2014	0	0
2015	300	23
2016	334	8
2017	531	5
2019	141	2
2020	677	0
2021	19	7

As shown by the timeline of the occurrence of floods from 2009 to 2021, there has been an increase in the number of floods, with four floods recorded between February and April in the year 2021, mainly in the Tabuk, Hafr Al-Batin, and Makkah regions, as reported by the Saudi Civil Defense Directorate. In 2018 and 2012 no reported or missing data.

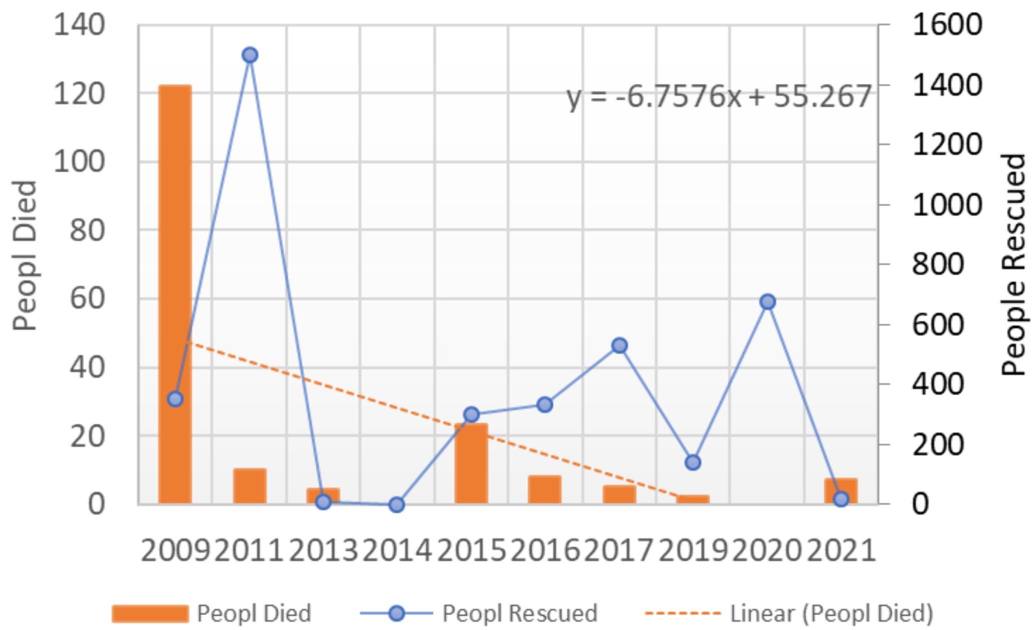

Figure 2. Human Life Losses Due to Floods in Saudi Arabia Between 2009 – 2021

<u>Section Three Research Statement</u>

It is difficult for researchers to obtain data that contribute to the study of natural disasters in Saudi Arabia. We seek through this study to provide good data sources with the lowest cost. In semiarid areas like Saudi Arabia, conventional observation is sparse.

A flash flood is usually a short-term event caused by heavy rainfall associated with clouds; however, satellite imageries still have difficulty detecting flash floods. Optical imagery cannot penetrate clouds to watch floods; radar-type sensors cannot penetrate rain clouds. After the sky becomes clear, the flash flood is also gone. Sensors like Landsat and SAR usually have a low temporal resolution or a long revisit time (6-12 days), and their limited spatial and temporal coverages may not be able to catch a short-term flash flood in semiarid areas. Therefore, there is a need to assess flash flood hazard to identify risk area and monitor flash flood in small semi-arid region like Saudi Arabia.

Section Four Objectives of the Study

- Study flash floods in Saudi Arabian cities using open data.
- Determine the impact weight of the selected thirteen flood-conditioning factors
- Identify risk areas and monitor flash floods in a small semi-arid region like Saudi Arabia.
- Combine the flood-conditioning factors using the weighted overlay method in ArcGIS to map the Wadi Al-Ahsbah in southern Saudi Arabia, a flood-prone area, and classify these areas based on risk.

Section Five What is Expected from the Model Outputs?

- The model can be used to understand a catchment's hydraulic behavior better and assist in developing flood control solutions.
- The model will create flood-risk maps, which are essential tools for planning and managing emergency responses in the country.
- The most important output of this assessment is to identify flood-prone areas where conventional data is unavailable.

- It also helps determine flood locations during heavy rains and provides flood warnings to residents.
- Creating a good database for flood-prone areas to avoid using them for human activities like schools and housing.
- In the end, researchers are encouraged to use open data to conduct more studies on this subject.

Section Six Impact and Benefit of Studying Flash Floods in an Arid Region

These are the impacts and benefits of studying flash floods in an arid region:

- Because of the low probability of floods in arid areas, little attention has been paid to this area in previous research, so the proposed research is essential.
- Determine risk levels for areas to help identify new construction sites and redevelop high-risk sites to avoid potential disasters in the future.
- Manage floods in arid areas, which may cause significant damage because such areas are not likely to take the required precautions.
- Help adventurers avoid the possible risk of floods (Memon et al., 2015).
- Saudi Arabia's government also focuses on developing a comprehensive disaster risk assessment, strengthening urban resilience planning, and building regional capacity for emergency response planning (GFDRR, 2017).

CHAPTER TWO LITERATURE REVIEW

One of the most dangerous forms of hazard is flash flooding. Since flash floods are difficult to track and have a rapid onset, their suddenness, rarity, small size, and peak discharge are also unexpected (Xia et al., 2011). Flash floods have severe impacts on human society, including loss of life, property damage, road and communication problems, and environmental damage. Flash floods have historically resulted in the highest number of human deaths (Jonkman and Kelman, 2005). Flooding risks are increased by the lack of hydrologic studies undertaken for most urban plans and by incomplete rainwater drainage systems and flood projects that fail to consider actual measurements and accurate pathways of the main wadis (Ashraf and Ahmed, 2019). The planning process and long-term urban growth must identify flood-prone areas and develop urban flooding maps based on geomatics and hydrological and hydraulic modeling (Marco, 2019).

Flood mapping is an integral part of flood risk management since it involves identifying geographical areas that could be prone to flooding (Salamon et al., 2014). Flood mapping includes developing two different maps that aim to show the flood danger levels in each region (Alfieri et al., 2017). First, a flood hazard map determines water depth in flooded areas (Alfieri et al., 2017). It solves this by categorizing the flood extent into three scenarios: low, medium, and high likelihood. A flood risk map is a map of possible flood areas that considers the region's population, economic activities, and, ultimately, areas with a higher potential risk of flooding (Rosser et al., 2017). To minimize losses in lives, property, and facilities, identifying the dynamics of floods in dry environments and forecasting a reliable flood hazard map while considering different factors and conflicting objectives are essential in Saudi Arabia's vision for 2030. A recent study (Essel et al., 2017) proposed that hydrologic assessment using geospatial approaches could identify different hydrologic components, plan hydrologic designs, and construct possible scenarios to reduce the risk of flash flooding.

<u>Section One Flood Modeling</u>

Ashraf and Al-Alola (2020) used the analytic hierarchy process (AHP) to extract weighted averages of eight parameters that influence flood-prone areas, including flow accumulation, distance from the wadi network, slope, and rainfall density, drainage density, and rainfall speed. They discovered that 22.12% of urban areas and 46.39 % of agricultural areas are vulnerable to high or very high flooding hazards (Ahmed and Ashraf, 2019). They used the hydraulic modeling program (HEC-RAS) to implement the risk matrix model while designing a two-dimensional model to measure the flood's speed, depth, and spread. Flooding was classified by Opolot (2013) based on the characteristics of the flood, the size of the affected area as a spatial element, and the triggering precipitation event's duration as a temporal element, expanding on Bronster's classification (Al-Ghamdi, 2012) that investigated the effects of urban development on flood hazards in Makkah, Saudi Arabia, using the NRCS curve number method. According to the findings, the residential areas of Makkah city have grown by 197 percent, while overall flood volumes have grown by 248 percent. The Convolutional Neural Networks (CNN) -based methods technique uses flood susceptibility mapping to two separate CNN classifications (Wang et al., 2020). In the CNN architecture, three data presentation methods are designed.

Eleven flood-triggering factors related to past flood events were used to construct the proposed CNN-based methods. In addition to the current support vector machine (SVM) classifier, the output of these CNN-based methods was evaluated using several objective parameters. The experiments showed that all CNN-based approaches could generate more accurate and valuable flood susceptibility maps.

Sarkar and Mondal (2020) used a frequency ratio model to demarcate flood vulnerability areas in the Kulik river basin. Parameters such as slope, elevation, rainfall, drainage density, land use–land cover, TWI, population density, road density, and household density were recommended to help understand flood mechanisms. Flood locations were obtained from the flood inventory map. The flood vulnerability zone map was divided into very low, moderately low, highly vulnerable, and highly vulnerable. The flood-prone map built with the FR model is highly accurate, with an AUC value of 0.901 for success rate.

Another study evaluated the Wi method's efficacy and compared its findings to frequency ratio (FR) and logistic regression (LR) methods. Thirteen variables were used, including elevation height, slope, aspect, curvature, geology, and soil. The area under the curve (AUC) and the Kappa index can be used to compare model results. The AUC prediction rates for LR, Wi, and FR were 79.45 percent, 78.18 percent, and 67.33 percent, respectively. The work suggested that the Wi system conducts flood susceptibility analysis effectively (Tehrany and Kumar, 2018).

Section Two Flood Mapping Techniques and Social Media

People and organizations use microblogging services like Twitter as communication channels during mass emergencies to provide status updates, provide aid, request help, and search for actionable information (Alharbi et al.,2019). (Petrovic et al., 2013) Twitter, it was discovered, frequently breaks incoming news about disaster-related incidents faster than traditional news channels. Detecting disaster-related information early allows decision-makers to respond quickly and successfully during disasters. Besides GIS, recent studies have focused on using social media in reporting flooding and event mapping (Smith et al., 2017). This advent of social media flood reporting can be attributed to a worldwide increase in the Internet and smartphones (Munasinghe et al., 2018). Social media has been used in various parts of the world because flood calamities may not be predicted, thus leading to the use of the information that citizens provide, either by their description or their video footage (Munasinghe et al., 2018). Their research on the use of social media for mapping (Rosser et al., 2017). Twitter, for example, provides a rich source of real-time information regarding emergencies from which meaningful information can assist situational awareness (Alharbi et al.,2019), revealing the use of geo-referenced reports related to floods. This method was described as a way of determining the credibility of information given by the citizens. Therefore, the research by Rosser, Leibovici, and Jackson (2017) has been used to describe the use of citizens' information in various formats, including video footage, where the information is ranked and stored for future events in response to emergencies. They also suggest using video footage to determine the hydraulics of flash floods. Aburizaiza (2019) used social media data to determine the locations of floods in Jeddah in 2009 and compared this data to the Topographic wetness index (TWI). She found that YouTube was more useful than Twitter because Twitter data was filtered and obtained using only English keywords. Finally, Rosser et al. (2017) also propose an interpolation approach that uses citizens' or social media users' photographs to determine flood levels.

In addition, research conducted by Herfort and colleagues in 2014 focused on assessing whether the geospatial distribution of tweets represents the geospatial distribution of floods. The research used hydrological information and digital elevation data to assess whether the tweets' geographical positioning represents the floods' geospatial information (Herfort, 2014). This research aimed to extract data relevant for mapping River Elbe floods with the aid of social media. The study found that only a small amount of information is georeferenced by Twitter users, thus hinting at the possibility of the use of geospatial distribution of tweets to determine the distribution of floods.

Another study conducted on flood mapping in Pakistan suggested using topographic information (Memon et al., 2015). This research differed from other research that used topographic information; this study incorporated a moderate-resolution imaging spectro-radiometer, which helps provide images that researchers can view and easily interpret. The use of social media and GIS for flood mapping, as found in various sources consulted by this research, clearly shows that using the two techniques is crucial in mapping data in arid areas, where there is a likelihood of unpredicted floods due to the unpredictable nature of rainfall (Middleton et al., 2013).

CHAPTER THREE STUDY AREA

Section One Location

Saudi Arabia is the largest country in the Middle East, covering an area of 2.253 x 10 km2. Al-Zahrani (2008). This study, however, will focus on the 2017 flood in Wadi Al-Ahsbah. Wadi Al-Ahsbah flows into the Red Sea from the Kingdom of Saudi Arabia’s central-western region. Flooding has occurred in several coastal areas. (Al-Zahrani, 2005)

The Wadi Al-Ahsbah valley is the longest in southern Saudi Arabia. The valley is located in the Tihama region and flows from east to west into the Red Sea. It is 50 kilometers long, and the total area is 1177 square kilometers. Many cities, villages, and residential areas are located on its banks; the most notable is the city of Al-Makhwah, which is situated on Wadi Rush and is one of the most important tributaries of Wadi Al-Hisbah. Estimates of the size of the floods in this valley indicate that they present a hazard to everything in the valley’s lower water basin and do not benefit from the amount of surface water wasted each year, which ends up in the sea. These floods typically cause property damage, loss of life, and damage to facilities. Furthermore, the torrential water pools eventually create swamps for diseases harmful to human health (Khimi, 2003). However, this study will focus on the 2017 flood in the Wadi Al-Ahsbah region (Figure 3).

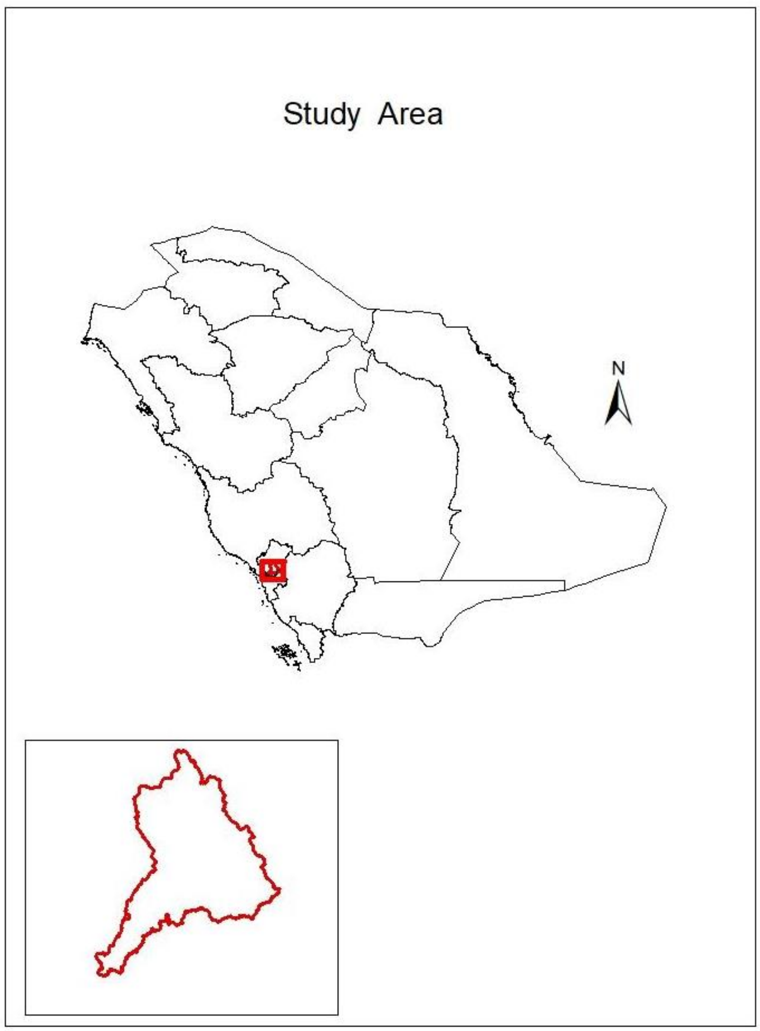

Figure 3. The study Area

Section Two Climate

The climate in Saudi Arabia is predominantly desert, and the country is extremely hot in summer (Peter,2008). Almost everywhere, rainfall is limited, with a peak from November to April. The current study looked at rainfall patterns and extremes in Saudi Arabia for the 42 years 1978 to 2019 (Figure 4) (Almazroui, 2020). Saudi Arabia is at risk of several natural hazards, including floods, sand and dust storm, and drought. Heavy rainfall in Saudi Arabia sometimes results in flash floods. The country receives intense rainfall, especially in the mountainous southwestern region, which floods seasonal water courses (Sharif et al., 2016). In the study of climatic variables, the rainfall quantities at each station show that the central and southern parts of Saudi Arabia receive the most substantial rainfall (Almazroui, 2020). Due to the high latitudinal range, Saudi Arabia's climate is influenced by various weather patterns. According to seasonal rainfall data, most rain falls during the winter and spring seasons Figure 5 (Almazroui,2020).

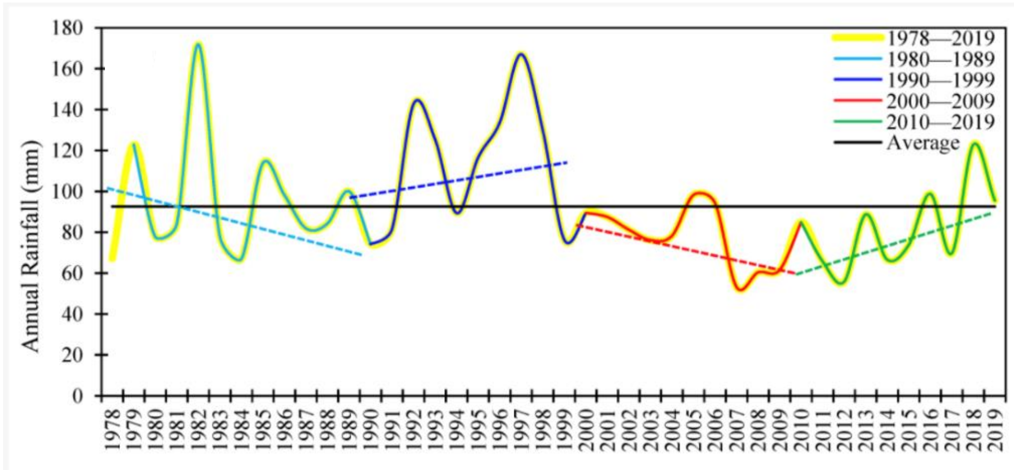

Figure 4. A Time Series of Country-Averaged Annual Rainfall (mm)

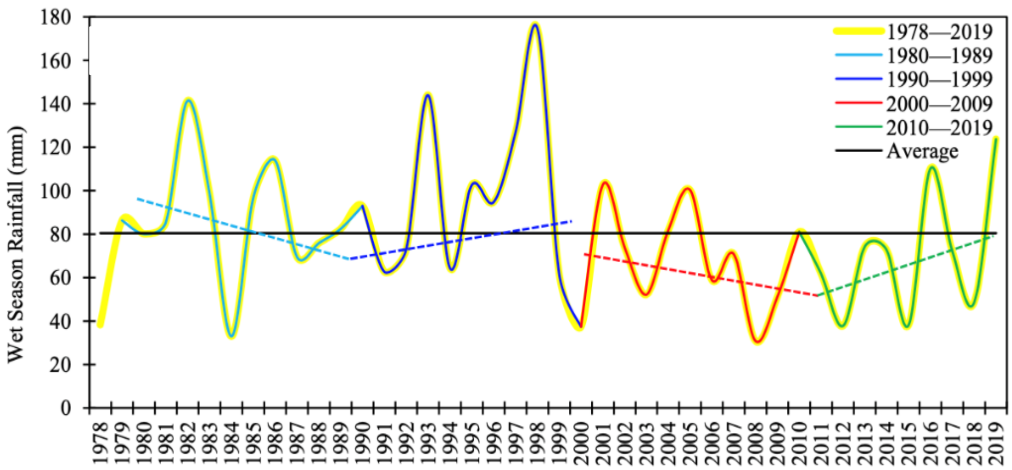

Figure 5. A Time Series of Saudi Arabia's Wet Seasonal Rainfall

Wadi Al-Ahsbah is in the AL-Baha region, for which the temperature and rainfall averages from 1991 to 2020 are shown in Figure 6.

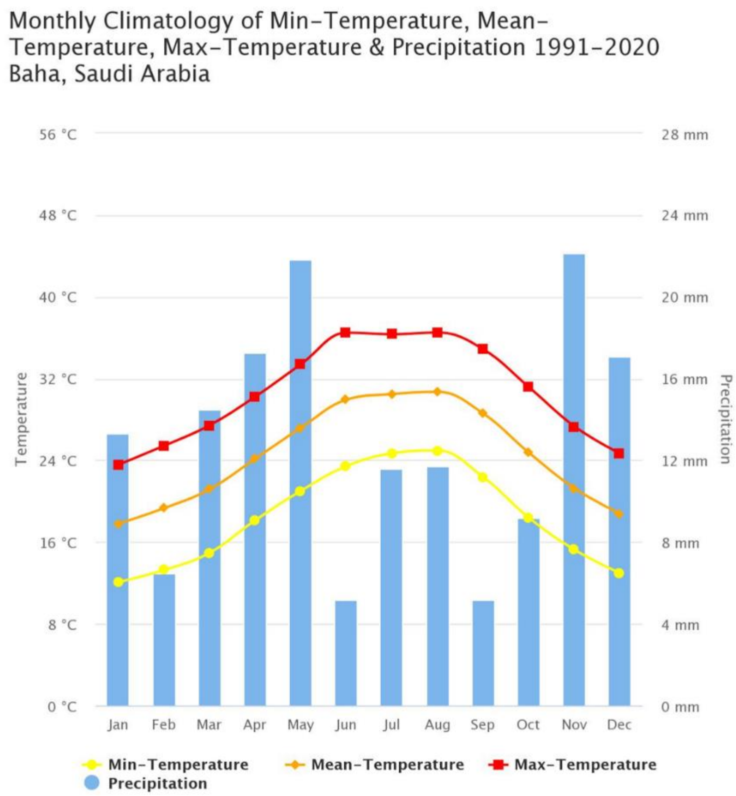

Figure 6. Average monthly temperature and rainfall from 1991 to 2020 for the AL-Baha region

Section Three Urban Development in Saudi Arabia

Worldwide, throughout the last century, there has been a gradual increase in urbanization (Newbold and Scott, 2013). In the United States, urbanized areas are defined as “areas having 50,000 or more population” (United Nations Statistics Division 2007 Demographic Yearbook: Table 6). In recent years, rapid urbanization has been observed in the Kingdom of Saudi Arabia, and large numbers of people have moved from rural areas to urban areas that provide all services (Moustapha et al., 1985).

According to the Municipal and Rural Affairs Ministry, Saudi Arabia has significantly increased urbanization since the 1950s. Saudi Arabia’s urban population had grown from 21% in 1950 to 80% in 2015 (Alahmadi and Atkinson, 2019). Table 4 shows the increase in population from 1970 to 2020. The population of Saudi Arabia reached about 34 million in 2020. The resulting growth in urban housing was much faster than the development plans available.

Table 4. Saudi Arabia Population Between 1970 and 2020

Population	
Year	Million
1970	5.8
2000	20.5
2018	33.7
2020	34.81

The rapid expansion of urbanization in Saudi cities has resulted in a lack of technological and administrative capacity to plan for it properly. In addition, there is a lack of cooperation between government departments and executive authorities, and there is also a lack of skills to manage urban growth (Bahrawi et al., 2020). This has resulted in increased built-up areas at the expense of the natural environment, valleys, and vegetation cover. The random urban expansion contributes to the waste of vegetation cover and the destruction of the ecosystem. In addition, unplanned urbanization leads to high human and economic losses (Al-Zamil and Al-Qarni, 2019). Many difficulties were encountered when attempting to study urban growth because of the lack of data or difficulties in obtaining it. Therefore, satellite data was used for mapping urban areas, such as the Operational Line-Scan System (OLS) of stable night-time (SNT) light images (Alahmadi and Atkinson, 2019).

Section Four The Effect of Urban Growth on Waterways

Several Saudi cities, including Riyadh, Jeddah, Mecca, Jizan, and Najran, were built on the sides of valleys. During the season of heavy rain, these cities may be prone to flooding (Al Zahrani et al., 2017). In some agricultural areas, factors such as soil fertility and groundwater availability have contributed to the spread of urban growth.

As a result of economic growth, most cities grew and expanded at the expense of areas for water run-off. Moreover, at the end of the last century, Saudi Arabia experienced periods of drought, which prompted people to build houses and farms randomly near waterways without considering the run-off areas, the terrain, and the slope of the land. Since Saudi Arabia experienced a significant rise in rainfall between 2009 and 2020, these residential activities exposed these new residential areas to flooding during the rainy seasons (Almazroui et al., 2012; World Bank, 2020).

The Geological Survey has confirmed that the climate of Saudi Arabia is desert. It causes heavy rains in short bursts, resulting in significant floods and damage to infrastructures such as roads and bridges, as seen during the Jeddah floods. As posted on the flood-list website, a flood struck Jeddah on November 21, 2017. As a result, four people were killed, 481 people were rescued, and 40 people from ten families were evacuated. In addition, streets in cities were under water up to 50 cm deep, causing severe traffic problems due to heavy rain of 115.5 mm (4.5 inches) per hour.

CHAPTER FOUR METHODOLOGY

This paper proposes a machine learning methodology to assess flash flood hazards with multi-satellite images automatically. The goal is to develop a system that can help automatically identify the factors contributing more to the flood areas in massive amounts of data. The statistical analysis method will use a decision tree and combine the frequency ratio (FR) and the application of logistic regression (LR).

Section One GIS

A geographic information system (GIS) allows for considering several factors that affect floods' frequency (Dewan, 2007). In this part, data mining and GIS will be used to analyze the factors that contribute to flooding, using Spatial Analyst and other tools.

4.1.1. Pre-Processing

1) Terrain data was collected using the Digital Elevation Model (DEM):

The elevation data from the Shuttle Radar Topography Mission (SRTM) at 30-m resolution in a latitude/longitude projection (EPSG:4326), obtained from NASA (EPSG:20437, Ain el Abd / UTM zone 37N), was used to re-project the DEM

2) The Define Projection by Data Management tool in Arc GIS is used to transform a raster dataset from one coordinate system to another in Arc GIS.

3) The Resample tool in Arc GIS is used to change the spatial resolution of a raster dataset and set rules for aggregating or interpolating values across the new pixel sizes.

4) (Data Management) > Clip Raster tool in Arc GIS is used to clip a raster with the study area.

5) (Conversion tool) > was used to convert point to raster.

6) (Extraction tool) > sample was used to create a table that can be worked with in Weka software.

7) Spatial Analyst tool in the ArcGIS is used to transform the DEM into a slope map (Spatial Analyst > Surface > Slope). This map describes the slope for each raster cell in degrees, based on the elevation at each point.

8) Another tool was used to calculate the aspect within the GIS spatial analysis (Spatial Analyst > Surface > Aspect).

9) Performed calculation of the flow accumulation to concentrate flow areas, which can identify stream channels using the GIS spatial analysis tool (Spatial Analyst Tools > Hydrology > Flow Accumulation).

10) The Stream Power Index (SPI) was calculated by using this formula (Danielson, 2013):

SPI = Ln ((Flow Accumulation Raster) + 0.001) * ((Slope Raster/100) + 0.001)) Arc Map software was used to produce the SPI and TWI from the DEM (Jaafari et al., 2014).

11) The Topographic Wetness Index (TWI) measures topographic influences on basic hydrological processes (Schillaci et al., 2015). TWI was calculated using interactions between fine-scale landforms and the up-gradient contributing land surface area, as follows (Beven et al., 1979):

TWI can be calculated from Equation (1)

$$\mathbf{TWI = ln\,[CA/Slope]} \tag{1}$$

Where CA is the local upslope catchment area that drains through a grid cell.

Slope is the steepest outward slope measured in drop/distance for each grid cell, i.e., the tangent of the slope angle for each grid cell (Tarboton, 1997).

12) To find (Distance from the river), we used (Spatial analyst tool) Hydrology > stream link and stream to feature.
13) In ArcGIS, we used Euclidean distance, a standard tool used mainly in multiple criteria analysis. This tool helps create a raster from a vector layer or another raster that visually and colorfully represents the current distances from that river or roads to the remainder of the field.

15) To transform a raster into an integer raster, reclassify tool in the ArcGIS is used to classify all factors to five classifications using the Equal interval method.This tool converted all raster images before using the weighted overlay.

16) All flood contributing factors were classified into five classes using the Equal Interval method for FR modeling (Ayalew and Yamagishi, 2005). Each conditioning factor has a spatial resolution of 30-m. Thus, for each class of conditioning factors, determine the total number of flood pixels (Mojaddadi and Pradhan, 2017). Figure 7 shows the flow chart of the processing.

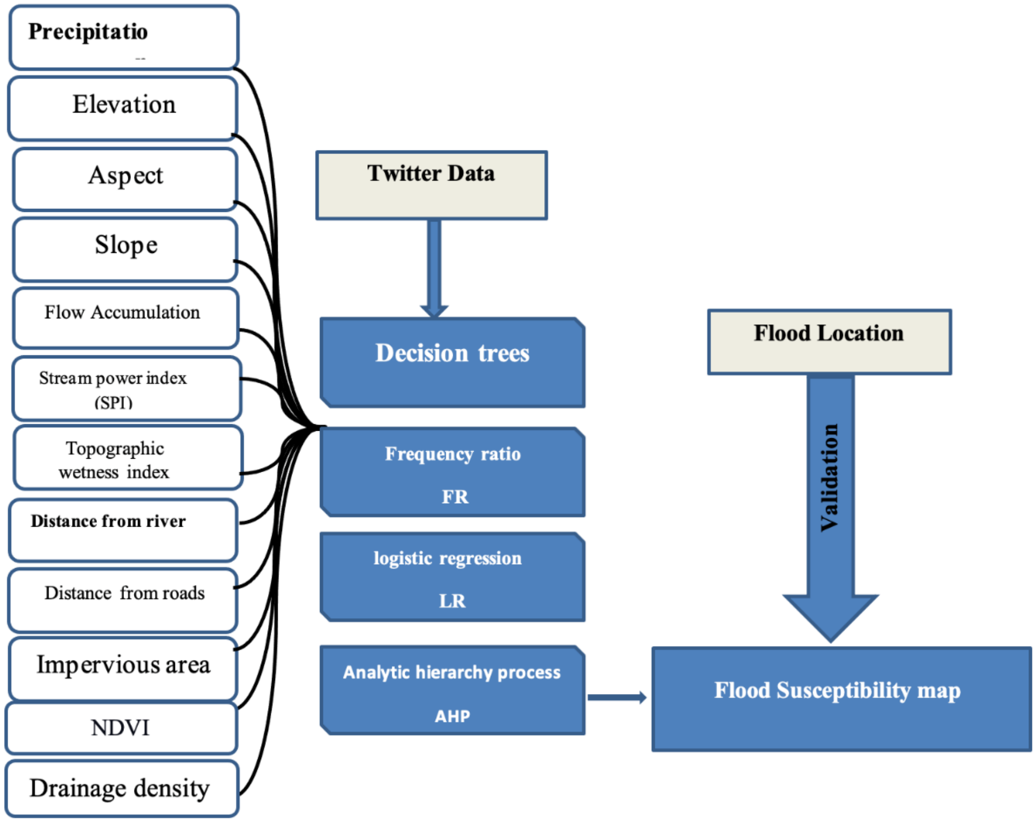

Figure 7. Flow chart of modeling flood susceptibility map.

Section Two Statistical Analysis Method

The decision trees, the frequency ratio (FR), and logistic regression (LR) were primarily applied to models. The flood Twitter Point data X Y from 2017 was collected and used as the models were applied using the statistical application Arc GIS and the data mining application Weka. The related factors were collected or measured from the DEM, soil, stream power index (SPI), topographic wetness index (TWI), distance from the river, precipitation (rainfall), and (GMIS) maps to examine the association between the factors and flood susceptibility.

After choosing the research area, the study's dependent variables were the flood-occurrence locations, and the independent variables were the other factors that affect the occurrence of floods.

The spatial datasets contain a total of ten susceptibility factors. In addition, the Twitter flood-location Point data (X, Y) will be used for validation purposes.

When the data from the Arc GIS file was finished with classification using the five equal intervals method, it was converted into the STATISTICA format, then the decision trees, frequency ratio (FR), and the logistic regression (LR), which were primarily applied to models, were applied in the program. The resampling nearest neighbor assignment result for flood conditions is shown in Table 5.

Table 5. The resampling result for flood conditions (sample)

DEM	SLOPE	ASPECT	FLOW DIRECTION	ACCUMULATION	DISTANCE FROM RIVER (m)	RAIN (mm/hr)	SOIL	SPI	TWI	Urban classification	FLOOD Type
137	6.85	292.17	32	98	1263	2.27	1	353.15	10.11	5	Flood
124	7.28	200.77	8	327	532	1.83	1	1253.6	11.24	3	Flood
144	7.97	298.07	16	0	400	1.83	1	0	5.36	2	Flood
148	9.04	338.75	64	13	400	1.83	1	62.07	7.87	2	Flood
153	1.34	225	16	1	400	1.82	1	0.7	7.81	2	Flood
139	4.73	264.29	16	3	800	1.83	1	7.45	7.27	4	Flood

4.2.1. ***Decision Trees DT***

A decision tree (DT) classifier represents a hierarchical model composed of internal nodes, leaf nodes, and branches (Bhaduri et al., 2008). A decision tree approach can help determine the threshold values of the predictors. Moreover, it can integrate all the possible candidate predictors. The basic strategy of DT is to select an attribute that will best separate the samples into individual classes (such as flood and no-flood in this study) by the measurement Information Gain Ratio. The main advantage of decision trees is that they are easy to construct, and the resulting trees are readily interpretable. From the set of input variables, decision rules will be generated through precise analysis. Without strict assumptions, this method can model relationships between variables regarding data distribution (Myles et al., 2004). Also, no specific rules are needed for the data format, as it can be nominal or scalar. It represents the relationship between a dependent variable and predicting factors. A root node, a set of internal nodes, and a set of terminal nodes construct the tree. Each node of the tree produces a binary decision that separates the classes. This analysis continues, and the tree moves down until the terminal node is reached.

4.2.2. Frequency Ratio FR

Lee et al. (2012) used the FR method for the flood-prone areas of South Korea. The researchers stated that the FR method could be easily applied to areas with limited map data at a low cost. Since the FR value represents the relationship between each class of impact factors and flood location, weights will be assigned to each class under each factor with accuracy (Neshat and Pradhan, 2015) using the ratio of the probabilities of the 'flooded and 'not flooded' areas (Bonham-Carter, 1994). Flood susceptibility mapping by the FR model is made simple with RS and GIS-like advanced techniques. The following formula was used to calculate FR values:

Frequency ratio is calculated by Equation (2)

$$\mathbf{FR = (E/F) / (M/L)} \tag{2}$$

where E is the total number of pixels with flash flooding hazards in the study area,

F is the total number of pixels with flash flooding hazards in each class of conditions,

M is the total number of pixels in the study area, and

N is the number of pixels in each class of the condition.

A higher FR indicates a stronger connection between flood incidence and conditioning factors. If the FR value is greater than 1, it shows a good correlation; if FR is less than 1, it shows a weak correlation (Regmi et al., 2013).

4.2.3. ***Logistic Regression LR***

Flood-susceptibility mapping requires knowledge of flooding processes and the relevant conditioning factors, such as rainfall, soil, and elevation (Ayalew and Yamagishi, 2005). In this section, "logistic regression" is used to examine a dependent variable of a specific event (flood) and the relationship between that event and multiple independent variables (conditioning factors) that may influence the frequency of the event (Tehrany, 2018). The "success" and "fail" categories in logistic regression are represented by the term y = 1 for the "Flood" category and y = 0 for the "non-flood" category.

Probability index is calculated by Equation (3)

$$P = 1/(1 + e^{-z}) \tag{3}$$

P is the expected flood probability, which ranges from 0 to 1 on an S-shaped curve, and

Z is a linear combination that is represented in the following equation:

$$Z = b_0 + b_1x_1 + b_2x_2 + b_3x3 + b_nx_n \tag{4}$$

CHAPTER FIVE DATA USED

The data used in this study include multiple remote sensing and social media data; the sources and descriptions are listed in Table 6.

Table 6. Descriptions and Sources of Data Used in This Study

Data	Data Description	Source
GPM	Precipitation data with 0.1◦ spatial resolution and half-hourly temporal resolution	NASA https://giovanni.gsfc.nasa.gov/giovanni/
Soil type	Vector Data with 27 soil types	Digital Soil Map of the World Maps & Layers, owned by Alison Hillegeist http://worldmap.harvard.edu/data/geonode:DSMW_RdY
DEM	Digital elevation model data at 30-meter resolution	NASA Shuttle Radar Topography Mission (SRTM)
GMIS	Global Man-made Impervious Surface percentage data at 30- meter resolution	NASA Socioeconomic Data and Application Center (SEDAC) https://sedac.ciesin.columbia.edu/data/set/sdei-viirs-dmsp-dlight/data-download
Twitter	Flood location data	Keyhole https://keyhole.co/index-1/?home_force=true
NDVI	Level 3 Global MODIS/Terra 16-day composite vegetation Indices at 250 m resolution	NASA Earth Data https://ladsweb.modaps.eosdis.nasa.gov/missions-and-measurements/products/MOD13Q1#product-information

Section One Flood Contributing Factors

Researchers have proposed many spatial methods for mapping flood hazard zones and flood risk zones to distinguish flood-prone areas spatially. However, many flood-related variables are needed to build a flood hazard assessment model (Tehrany et al., 2014). There are nine factors in the data set used in this study. Each factor is described below.

5.1.1. Precipitation Intensity

Precipitation intensity is a critical factor since it significantly affects the total of floods over a wide range of time and locations (Souissi et al., 2018). Therefore, rainfall is a significant water source; the data were collected from the Earth data Giovanni website and selected using a time-averaged map from 2017. That source was NASA Global Precipitation Measurement (GPM) half-hourly data with a spatial resolution of 0.1∘ and a time scale of half-hourly. The (GPM) represents the period 2017. In addition, we have obtained 14 images with half-hourly temporal resolution. Figure 8 shows the Global Precipitation Measurement (GPM) data at half-hourly temporal resolution for the Wadi Al-Ahsbah region in southern Saudi Arabia.

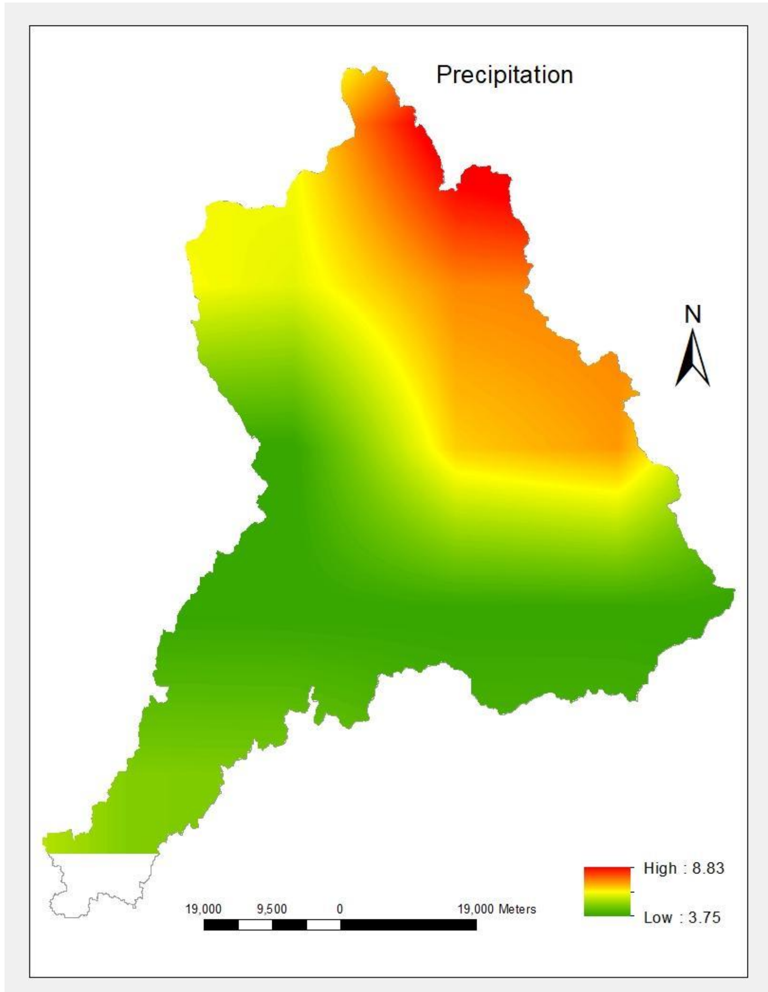

Figure 8. The Precipitation Distribution Map

5.1.2. Surface Elevation

Digital Elevation Model (DEM) data show the physical land surface in flood models. A DEM's spatial resolution determines the quantity of land covered by a single grid cell. DEM resolutions typically range from 1000 m to 2m or less (Saksena et al., 2015). We used 30-meter resolution elevation data from NASA's Shuttle Radar Topography Mission (SRTM) with an arcsecond resolution (3601x3601pixels) in a latitude/longitude projection

(EPSG:4326) in this study. The digital elevation model generated additional rasters that depicted slopes, aspects, flow directions, and flow accumulations in the research region, then complemented the original raster. These results were used with the rest of the data to determine the flood's risk map. Figure 9 shows a Digital Elevation Model (DEM) map in the study area.

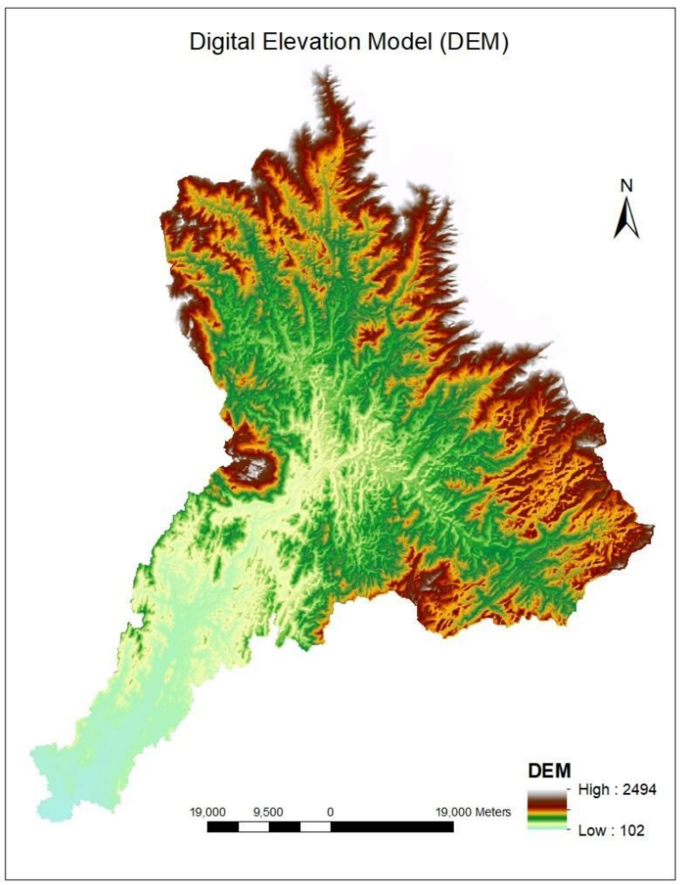

Figure 9. The Digital Elevation Model Map

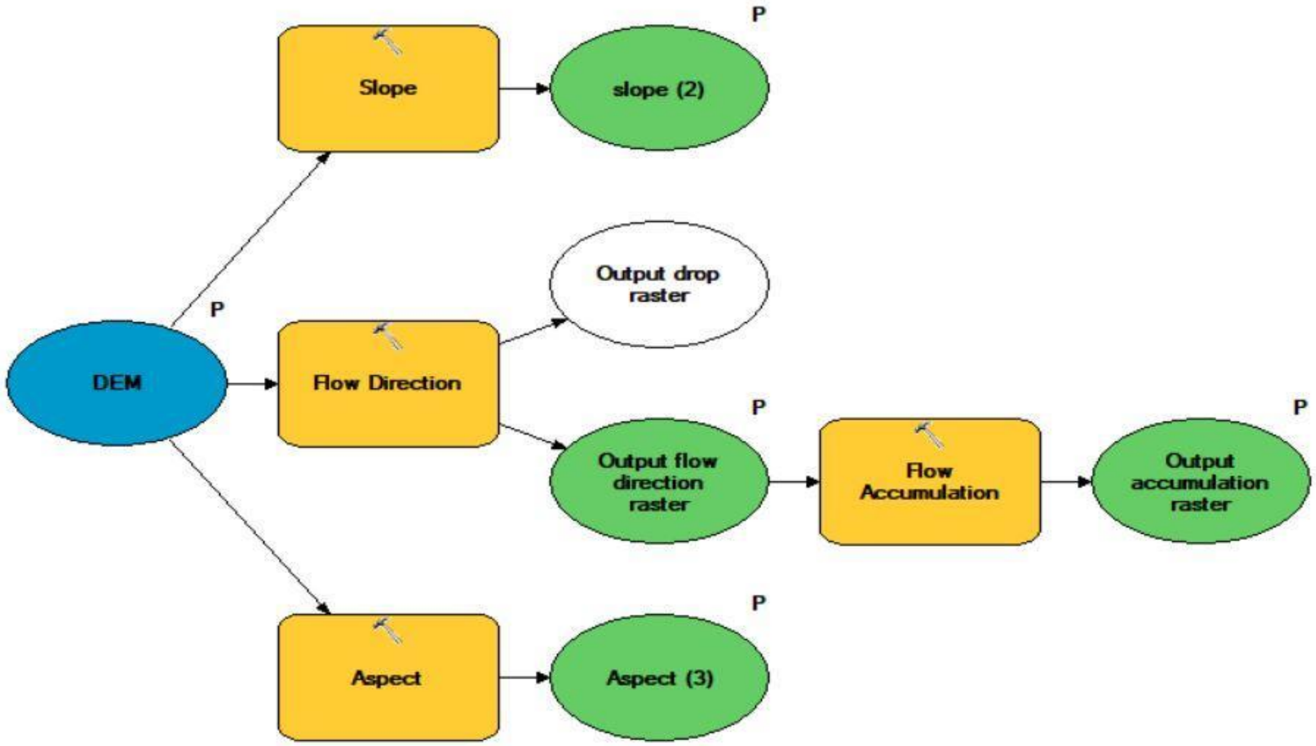

Figure 10. Model Builder for Slope, Flow direction, and Aspect

The geoprocessing model was generated in ArcMap using the Model Builder tool. This process allows us to enter, process, and create good outcomes in the shortest amount of time. First, the Slope, Aspect, and Flow Direction were calculated using the model in Figure 10.

5.1.3. Aspect

Figure 10 displays the aspect of each raster cell grouped into compass directions (north, northwest). The Reclassify tool was used to classify the aspect map based on elevation to five classifications (Amar Sitabi, 2015).

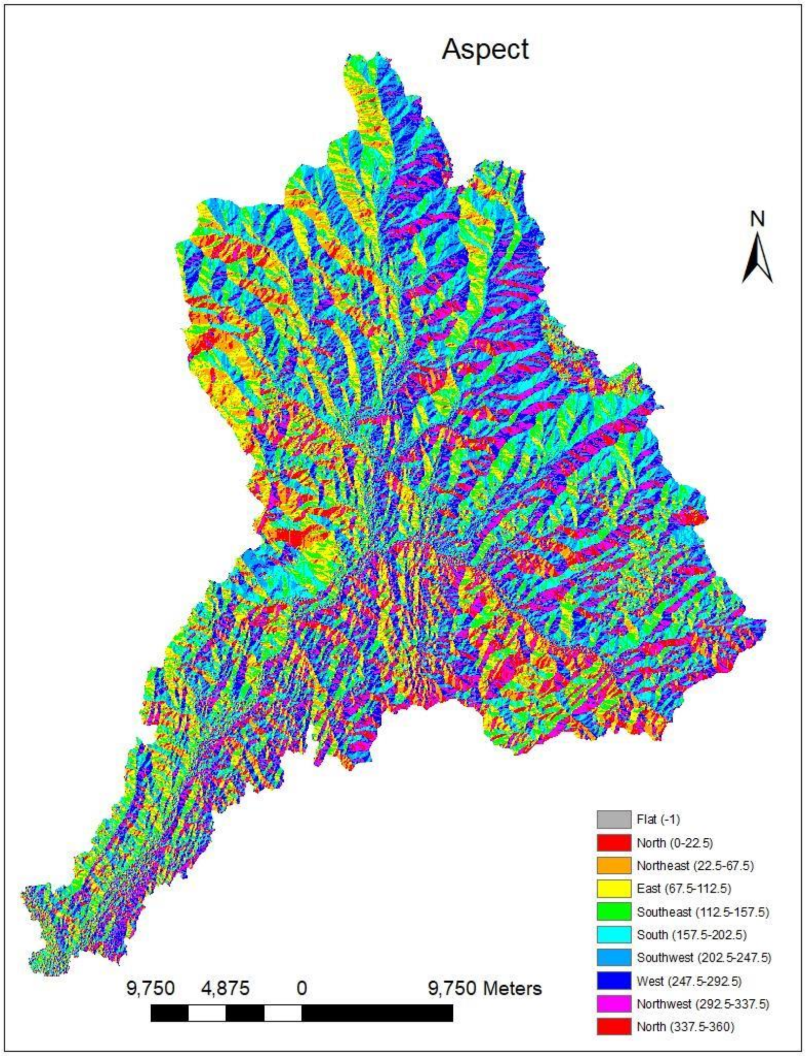

Figure 11. The Aspect of each cell of a raster surface is organized by direction Map

5.1.4. Slope

The slope is a topographical factor regarded as a vital hydrology parameter (Tehrany et al., 2013). Furthermore, slopes play a significant role in determining flood areas since they control the speed of the water on the neighboring slopes, where the amount of rushing water is much more than in flat areas. Therefore, the slopes were classified into five categories in this study: very low, low, medium, high, and extremely high. Figure 12 shows the slope classification map in the study area.

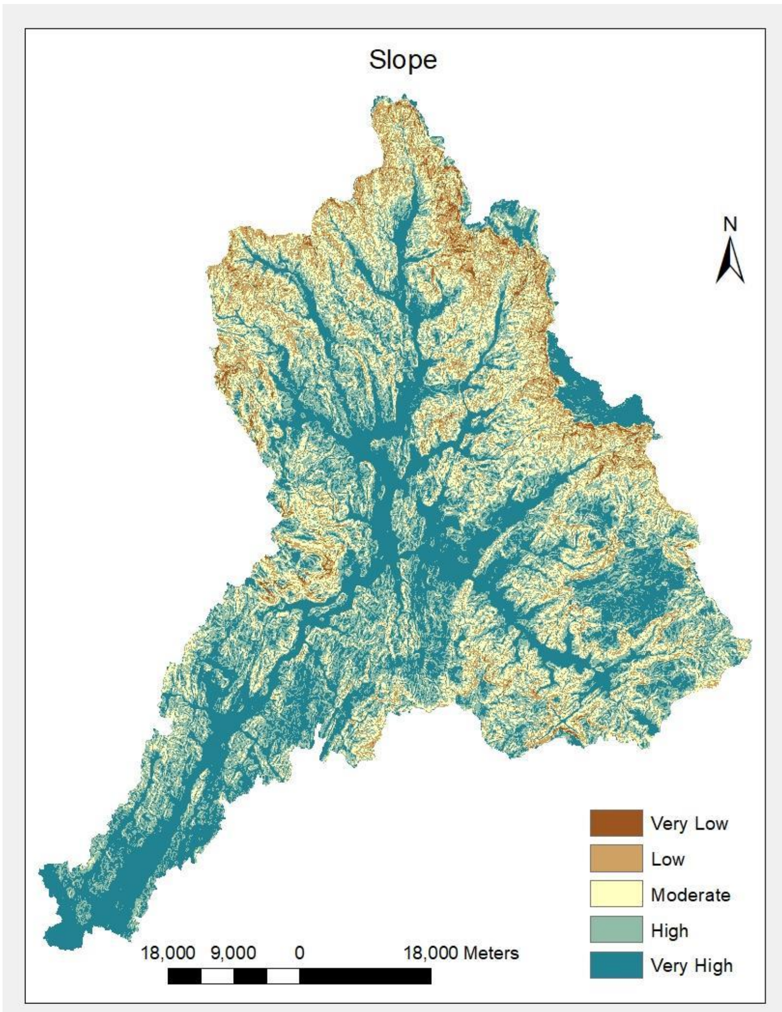

Figure 12. The Slope Classification Map

5.1.5. Flow Accumulation

Flow Accumulation produces a raster of cumulative flow to each cell, as determined by adding the weights of all cells that flow into each downslope cell. High-flow accumulation output cells are places of the concentrated flow used to identify stream channels. Per relevant studies, the most crucial factor has been identified as flow accumulation (Kazakis et al., 2015). In Figure 13, the Flow Accumulation raster was classified into five classifications.

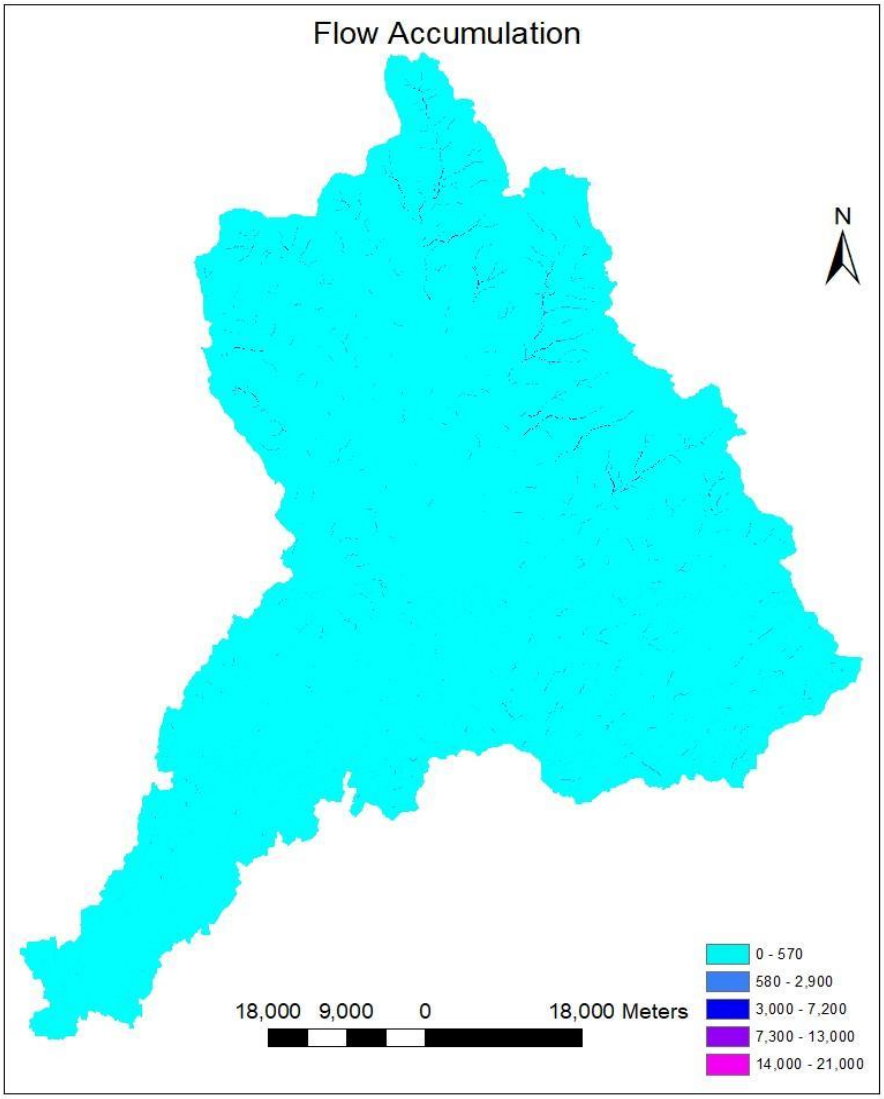

Figure 13. Classifications of Flow Accumulation Map

5.1.6. ***Stream Power Index SPI***

The SPI is used to measure the erosive power of overland flow at a particular location on a topographic surface (Moore and Grayson, 1991). SPI is a significant component affecting channel widening and hence floods (Righini et al., 2017). Therefore, a more significant SPI number should indicate a greater probability of erosion in the area. Figure 14 shows the SPI distribution calculated from Equation 1.

SPI can be expressed by Equation (5)

$$SPI = A \tan \beta \tag{5}$$

where A is the upstream contributing area (m2), and

β is the slope in each cell (degrees).

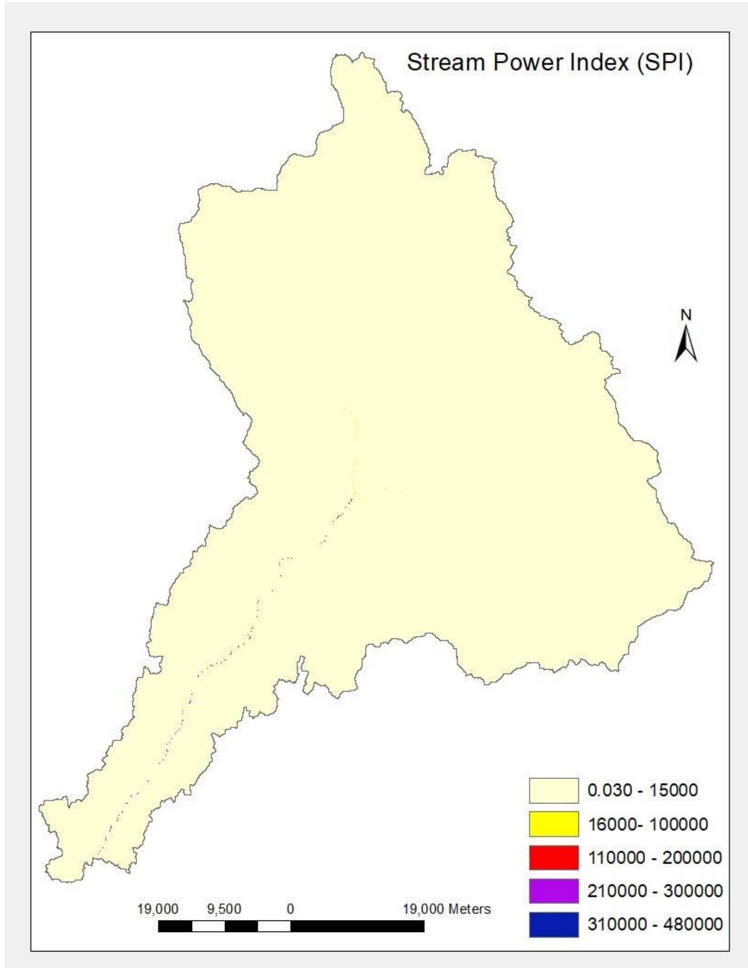

Figure 14. Classifications of Stream Power Index (SPI) Map

5.1.7. Topographic Wetness Index TWI

The TWI identifies and measures the saturated region exposed to overland flow (Wilson and Gallant, 2000). Beven and Kirkby (1979) developed the TWI as a component of TOPMODEL's runoff model. The topographic wetness index (TWI) can be used instead of the conventional method of identifying flood-prone areas just by contours. The TWI is a more cost-effective method of determining floods than standard hydrodynamic models

(Pourali et al., 2014). A tool has been developed through the ArcGIS software from the toolbox and model-builder tool to obtain the topographic wetness index (TWI). Figure 16 shows the topographic wetness index using the digital elevation model as shown in Figure 15 for the model builder for the topographic wetness index (TWI) and the stream power index (SPI).

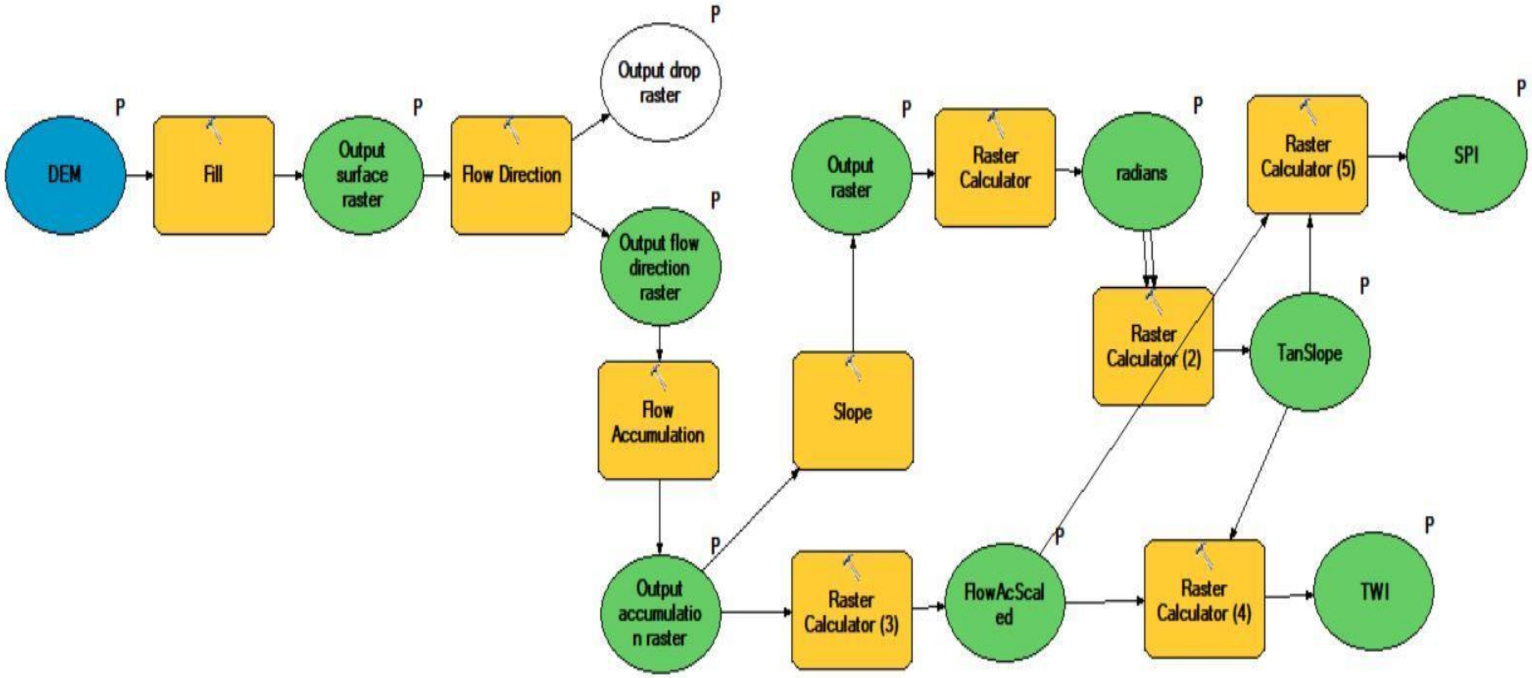

Figure 15. ArcGIS Model Builder for (TWI) and (SPI

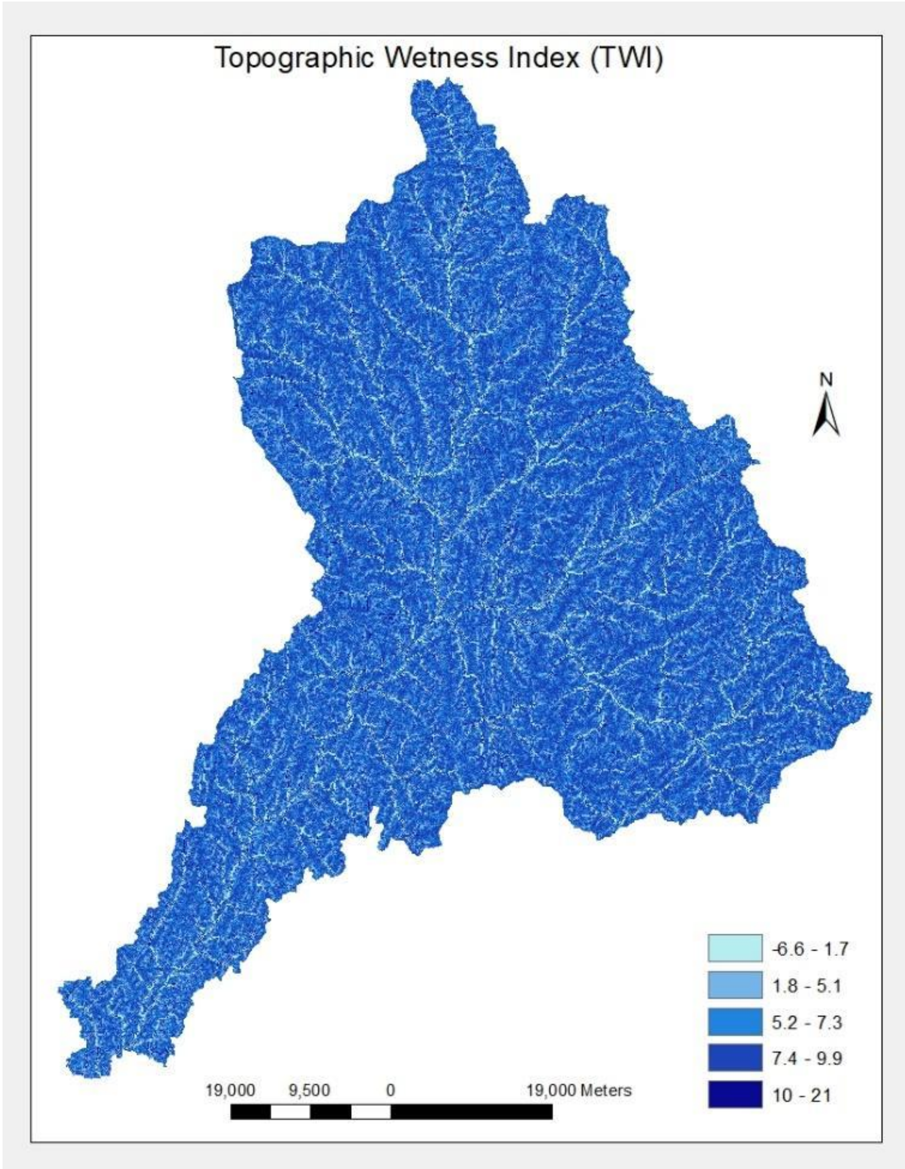

Figure 16. The topographic wetness index (TWI) Map

5.1.8. Drainage Density

When dividing the total length of all stream lengths in the basin by the total area of the basin, we get a measure of the amount of dissection of the watershed, which is called the drainage density (Bhattacharjee, 2016) Drainage densities in semiarid to humid environments range from 2 to 12 km^2, with most of the variation reflecting changes in rainfall (Abrahams 1972). Figure 17 shows the Drainage density map with a classification of areas according to their density.

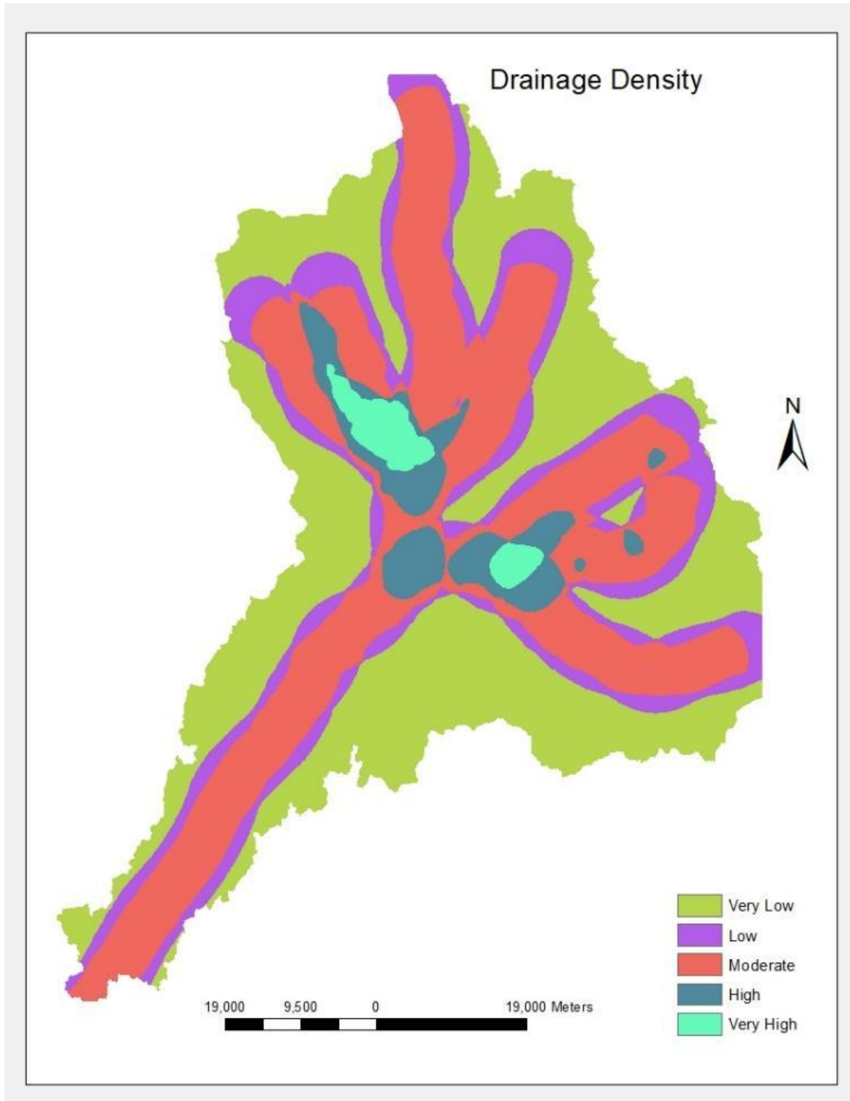

Figure 17. The Classification and Distribution of Drainage Density

5.1.9. Distance from River

The distance from the river is critical in identifying flood-prone areas and evaluating the flood hazard index. (Hernandez and Lutz, 2010) demonstrate that regions next to river networks are particularly prone to floods. Like flow accumulation, when there is much rain, the water levels in rivers rise, and the water overflows into the areas nearest to the river, causing floods. The map of distance from the river was generated using the Euclidean distance tool in the spatial analyst tool of ArcGIS. The Arc Map 10.1.8 thematic map in Figure 18 was classified into five categories: very high, high, moderate, low, and very low. A description of the distance from the river is shown in Table 7.

Table 7. Description of Distance from River

Distance Values	Class Values (Ranking)
400 -800 (M)	Very High Risk
800 -1200 (M)	High Risk
1200 -1600 (M)	Moderate Risk
1600 – 12000 (M)	Low Risk
>12000 (M)	Very Low Risk

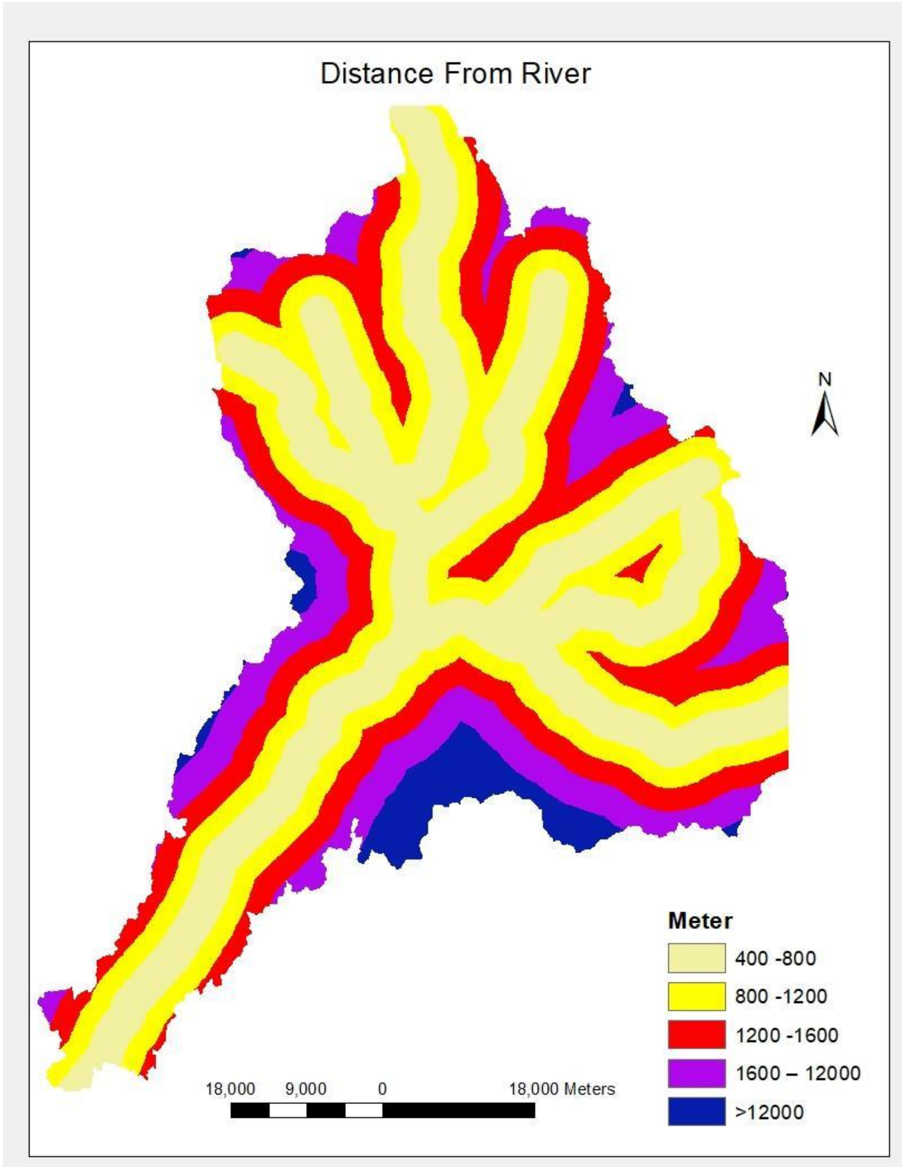

Figure 18. The Classification map of distance from the river

5.1.10. Distance from Roads

Surface-water runoff is increased on impervious surfaces such as roads, pavements, and parking lots because rainwater cannot filter through the soil. Therefore, it is considered one of the most critical factors that help us identify flood-prone locations, based on the study that looked at the distance from roads (Tehrany et al. 2017). Figure 19 shows how to use ArcGIS' spatial analyst tool, the Euclidean distance tool, to map how far this place is from a road. This map was divided into five categories: very high, high, moderate, low, and very low.

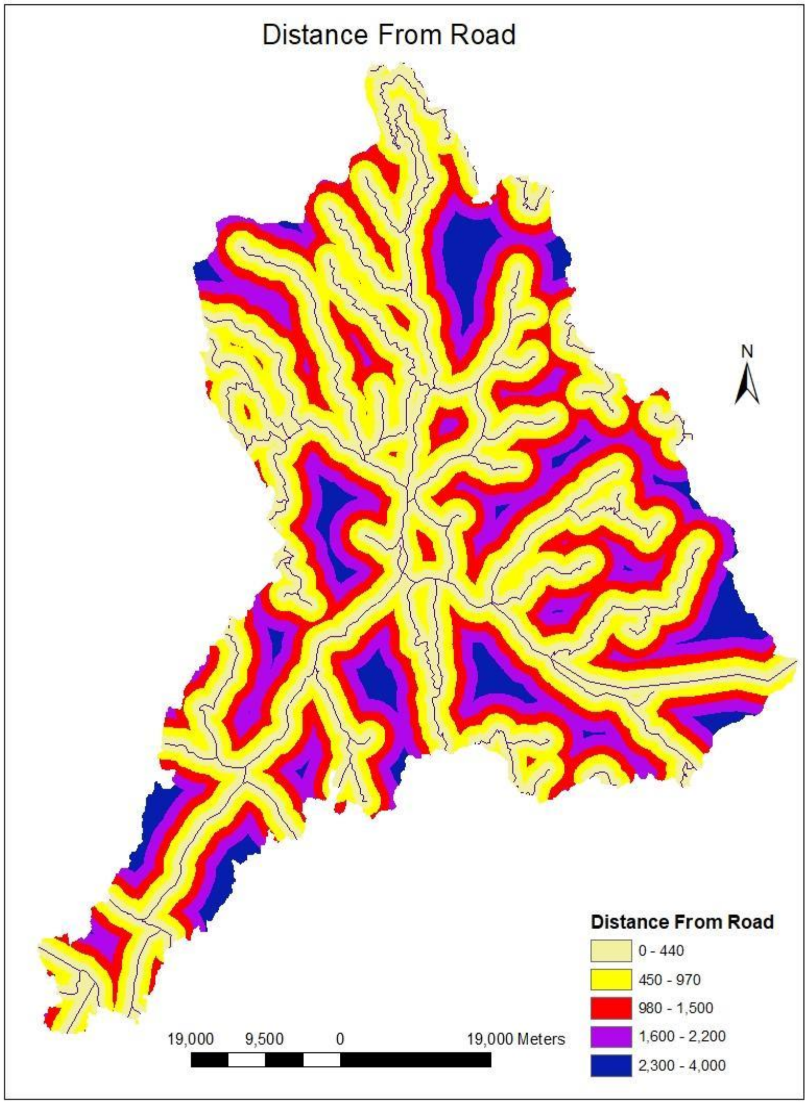

Figure 19. The Classification of distance from Roads Map

5.1.11. Soil Types

The soil map is critical in identifying flood-prone areas since soil type directly impacts drainage. Also, soil impacts water storage and permeability. (Mojaddadi et al. 2017; Tehrany et al. 2017). The soil was taken into consideration in the study. Twenty-seven soil types are found in Saudi Arabia; our study area included three types of soil (Mojaddadi and Pradhan, 2017). Figure 20 shows the three types of soil: silty clay, which does not drain well at all; clay, which does not drain well; and loamy, which covers most of the study area (including high regions) and is well-drained. The Digital Soil Map data was obtained from World Map as Vector Data. Alison Hillegeist owns this data. A description of soil types in the study area is shown in Table 8.

Table 8. A description of soil types

Code	Type	Meaning
1	Silty Clay	Does not drain well at all
2	Clay	Does not drain well
3	Loamy	Well-drained

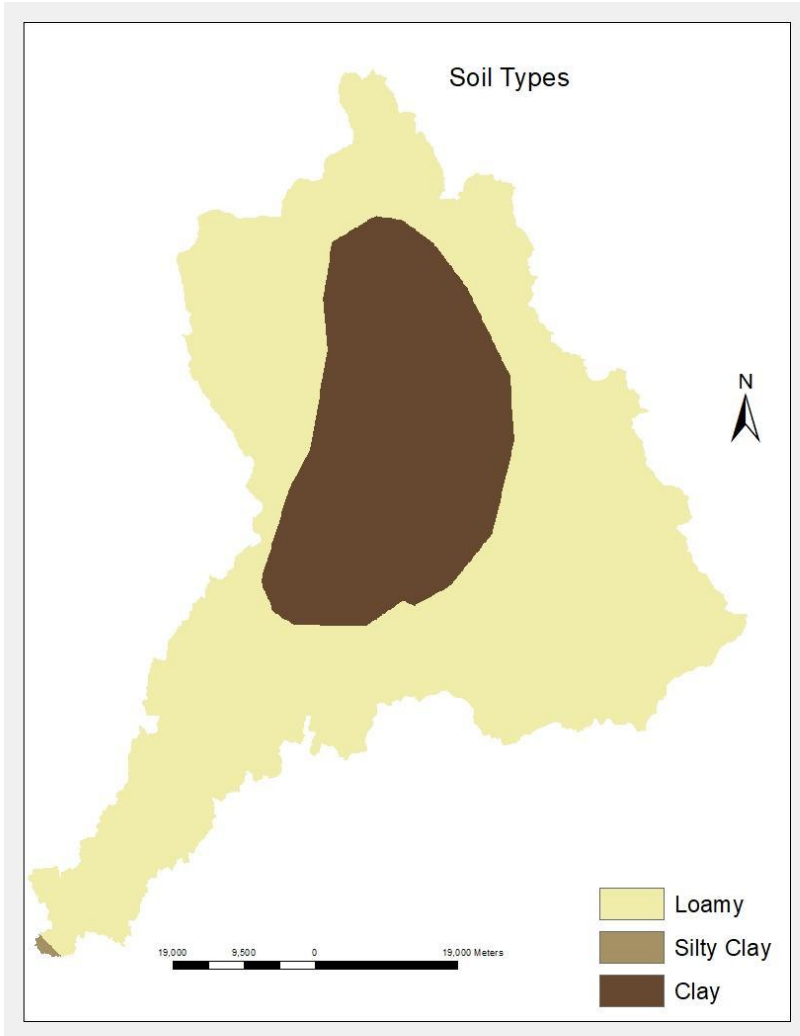

Figure 20. Soil classification map

5.1.12. Urban Area as Represented Impervious Areas

Urbanization is one of the factors that cause change on the Earth's surface, with most of the world's population residing in urban areas (Nigussie, 2019). As a result, to fully understand the changes that occur during this activity, it is required to obtain urbanization maps, which are critical in creating flood sites. In addition, compared to vegetated regions, urbanized areas create significantly greater surface runoff (Tehrany et al., 2015). Because of this, land use is a significant contributor to flooding (Beckers et al., 2013). However, for a large country like Saudi Arabia, which has a land area of over 2 million km^2, processing such high spatial resolution is challenging (Alahmadi and Atkinson, 2019)

Due to a shortage of data, it was challenging to create a map with a recent update in the study area. Therefore, global High-Resolution Urban Data from Landsat was used. The (GMIS) Dataset contains global estimates of fractional impervious cover generated from the Global Land Survey (GLS) Landsat dataset for the target year 2010. The GMIS dataset is made up of two parts:

1) global percent of impervious cover

2) per-pixel-related impervious cover uncertainty

These layers are co-registered to the same spatial extent and serve as a companion dataset to the Global Human Built-up and Settlement Extent (HBASE) dataset (Wang et al., 2017). These data can be used in local modeling studies of urban influences on energy and water and country-level analysis. Lower values of the (GMIS) index, which runs from 0-100 (%), are associated with places with no impervious surfaces (vegetation, no human build-up). Buildings, roads, other human-made things, and rocks have higher values (Colstoun et al., 2017). Figure 21 shows the impervious area distribution map in the study area from the Global Man-made Impervious Surface (GMIS) Dataset from Landsat (2010). It has a 30m special resolution produced from Landsat satellite data.

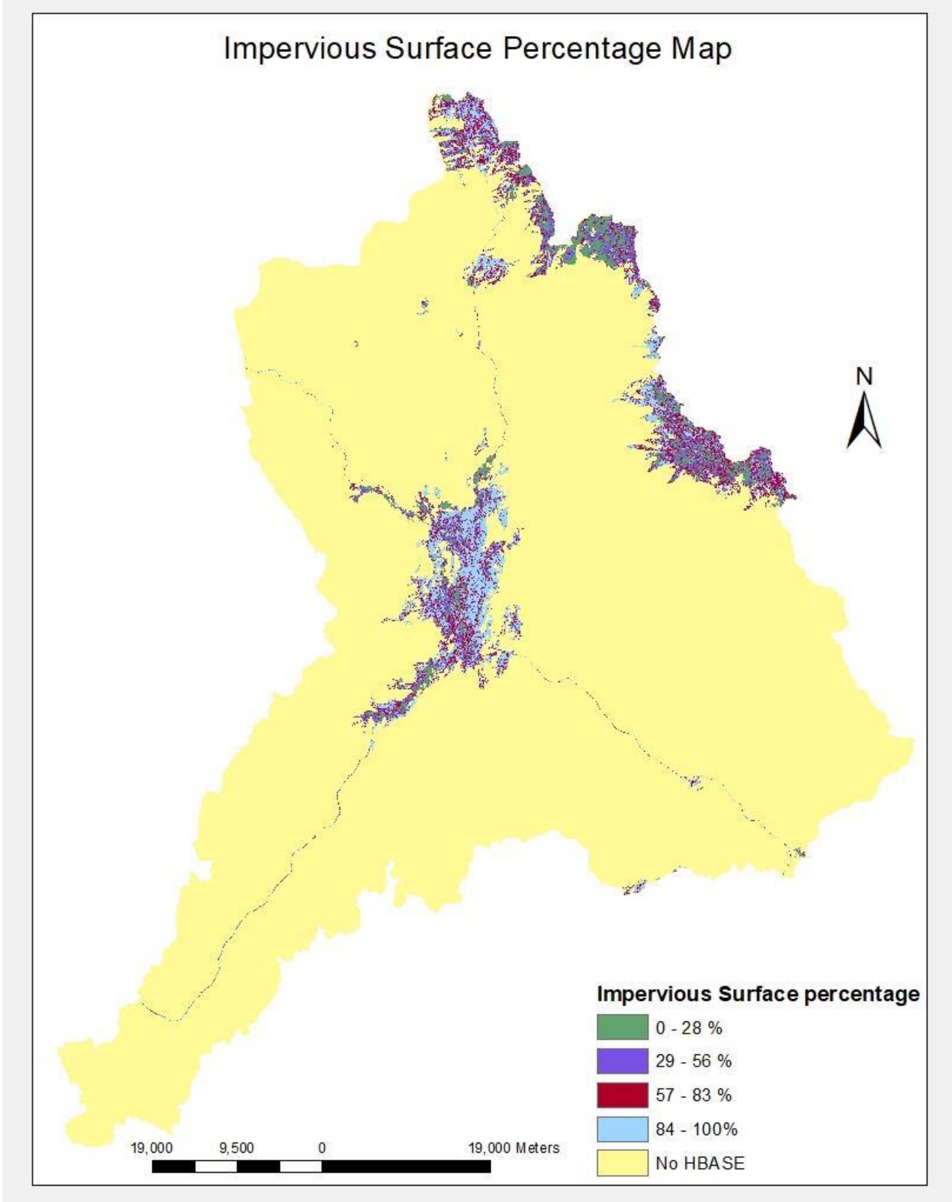

Figure 21. Classification of Impervious Surfaces map

5.1.13. Normalized Difference Vegetation Index NDVI

The Normalized Difference Vegetation Index (NDVI) is a graphical indicator used to assess whether a target includes live green vegetation (Pettorelli, 2013). As previously stated, vegetation has a negative relationship with flooding because it reduces the amount of water that runs off. According to the results of this study, in Figure 22, the NDVI was classified into five classes: very low vegetation, low vegetation; medium vegetation; high vegetation; and very high vegetation.

Normalized Difference Vegetation Index (NDVI) can be calculated by Equation (6)

$$\text{NDVI} = \frac{NIR - Red}{NIR + Red} \tag{6}$$

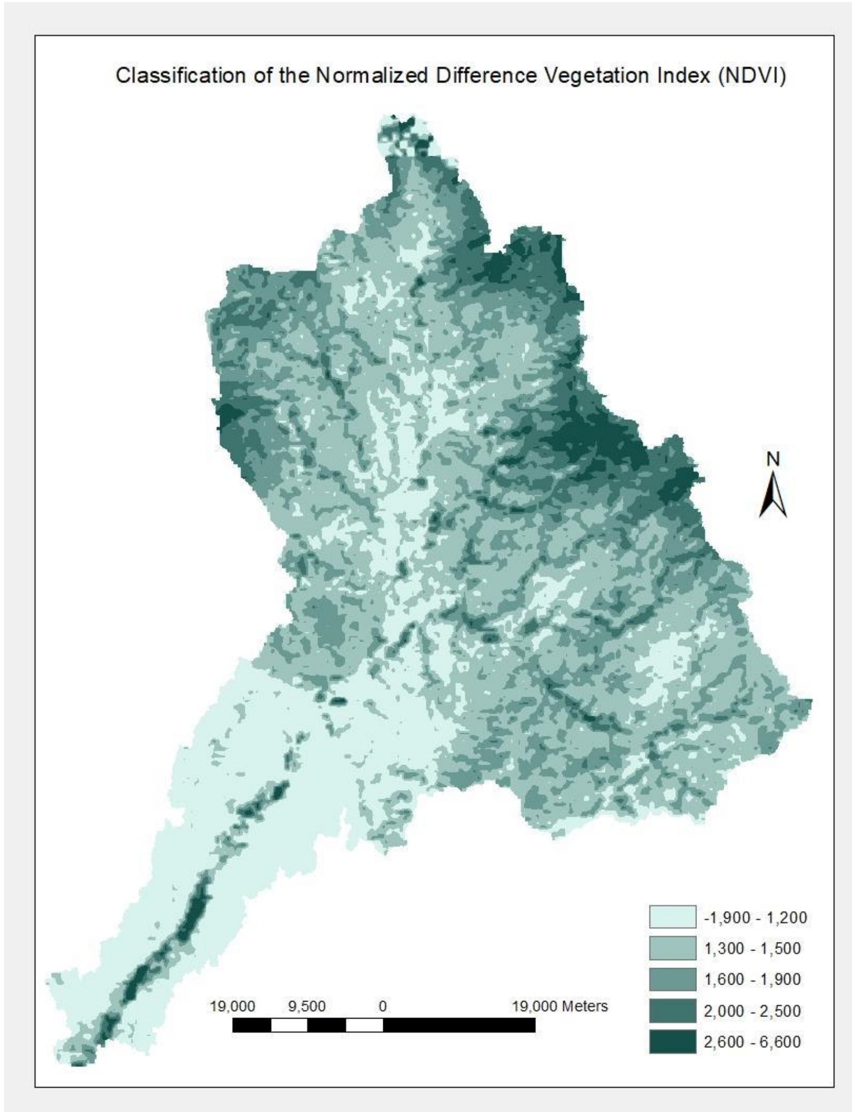

Figure 22. Classification of the Normalized Difference Vegetation Index (NDVI)

5.1.14. Social Media Data

Social media data was used as a validation tool in this study. This data contained additional information about the flood areas and was gathered by the public via social media platforms such as Twitter and YouTube for November 2016, as listed on the flood list website. The data was collected by subscribing to the Keyhole website, and the study area's boundaries sorted the points in Figure 23. This data set contains two categories. The first was assigned to represent the flooded area; it has 114 locations. The second category, signified by the number 2, represents 548 locations that are not prone to flooding. These points were found at high elevations with minimal possibility of flooding. In addition, we collected social media data reported in both Arabic and English.

The word used to search for social media data were as follows:

المخواه سيول#, #الباحه, #تهامة_الباحة, #Saudi, #Flooding, #Flood #SaudiArabia, #AlBaha, Saudi Arabia al Baha flood, Heavy rainfall in al Baha Saudi Arabia, JCB rescue mission in Saudi Arabia Al Baha.

Most of the data contain multiple images. A few of them describe the extent of the disaster, such as the collapse of electricity poles, the shoveling of cars, the deaths of people, and the disruption of traffic. On the other hand, some tweets got many adventurers to walk around the floodwaters as places experience a shortage of precipitation throughout the year.

Table 9. Description of flood location data.

Code	Description
1	This one was assigned to represent the flooded area reported in social media in both Arabic and English
2	Represents locations that are not prone to flooding. It was included for data mining using the decision tree algorithm. These locations were found at high elevations with very little possibility of flooding

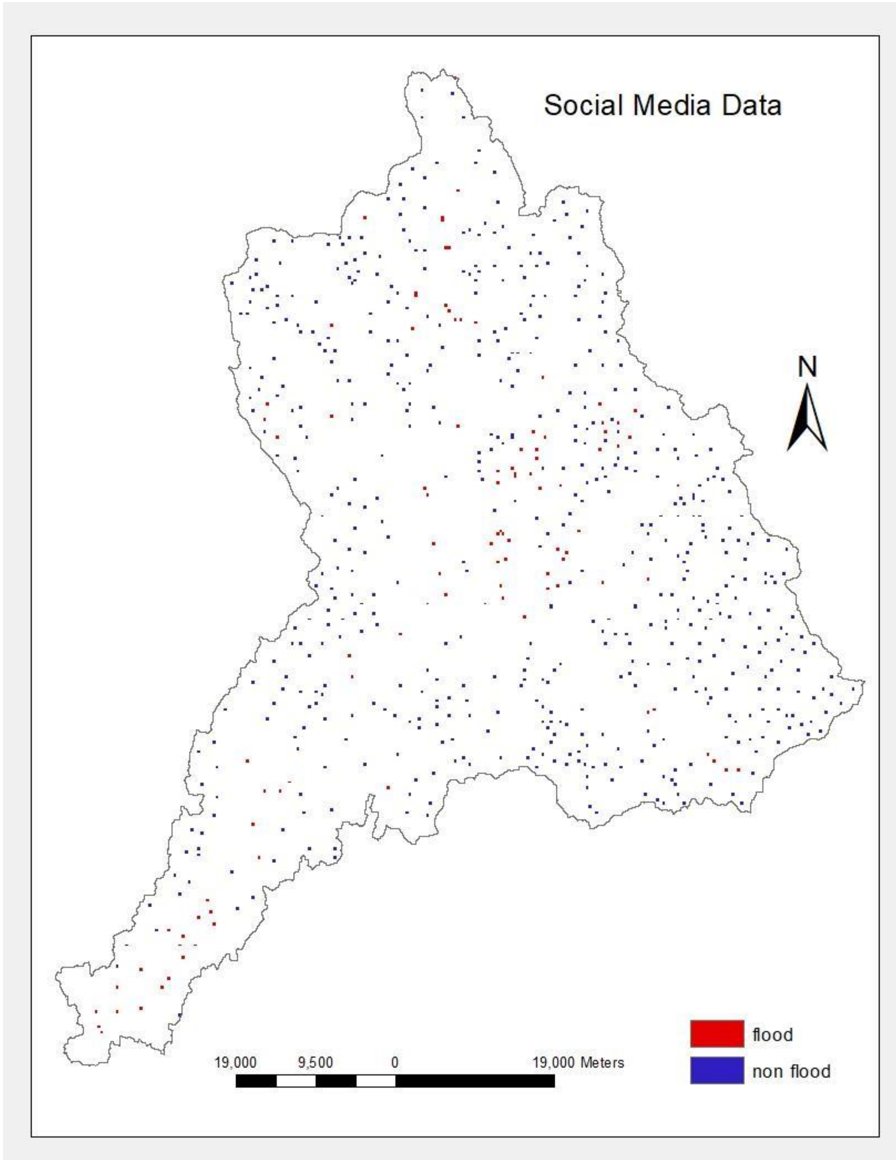

Figure 23. Social media flood location data

CHAPTER SIX RESULTS

Section One Results from the Decision Tree

The multiple factors contribute to the generation of flood-susceptibility mapping; in this section, the study focused on ten factors: elevation, aspect, slope, flow accumulation, Stream Power Index, Topographic Wetness Index, Drainage Density, Distance from River, Distance from Roads, Soil Type, Urban Classes as Represented by Impervious Area Percentage (%) in Table (11), Normalized Difference Vegetation Index (NDVI). The extracted data from the resampling nearest neighbor assignment result table was processed using the “Treej48” method, and the result is shown below in Table 10.

Table 10. Class of Impervious Surface Percentage

Class	Percentage of Impervious Area
1	**0 – 6.3**
2	**6.3 – 12.6**
3	**12.6 – 18.9**
4	**18.9 – 25.5**
5	**25.2 – 31.5**
6	**31.5 – 37.8**
7	**37.8 – 44.1**
8	**44.1 – 50.4**
9	**50.4 – 56.7**
10	**56.7 – 63**

Table 11. Resampling nearest neighbor assignment result using Decision Tree

```
Number of Leaves  :     16

Size of the tree :      31


Time taken to build model: 0.01 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         522               91.2587 %
Incorrectly Classified Instances        50                8.7413 %
Kappa statistic                          0.6848
Mean absolute error                      0.1119
Root mean squared error                  0.2831
Relative absolute error                 38.0651 %
Root relative squared error             73.9676 %
Total Number of Instances              572

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.686    0.038    0.795      0.686    0.737      0.687    0.826     0.649     Flood
                 0.962    0.314    0.934      0.962    0.948      0.687    0.826     0.915     Non Flood
Weighted Avg.    0.913    0.265    0.909      0.913    0.910      0.687    0.826     0.867

=== Confusion Matrix ===

   a   b   <-- classified as
  70  32 |   a = Flood
  18 452 |   b = Non Flood
```

The overall extent of the flood was this: 662 locations were identified, of which 548 were chosen at random to be validated in the study region, and 114 were chosen based on Twitter data. A DEM was used to validate the chosen positions, proving that they had been extracted correctly. The classifier had an accuracy of 91.25%.

The correctly classified instances and incorrectly classified instances show the percentage of test instances that were correctly classified and the percentage of test instances that were incorrectly classified. The raw numbers are shown in the confusion matrix, with a and b representing the class labels. Here, there were 608 instances, so the percentages and raw numbers add up: aa + bb = 70+ 452= 522, ab + ba = 18+ 32 = 50.

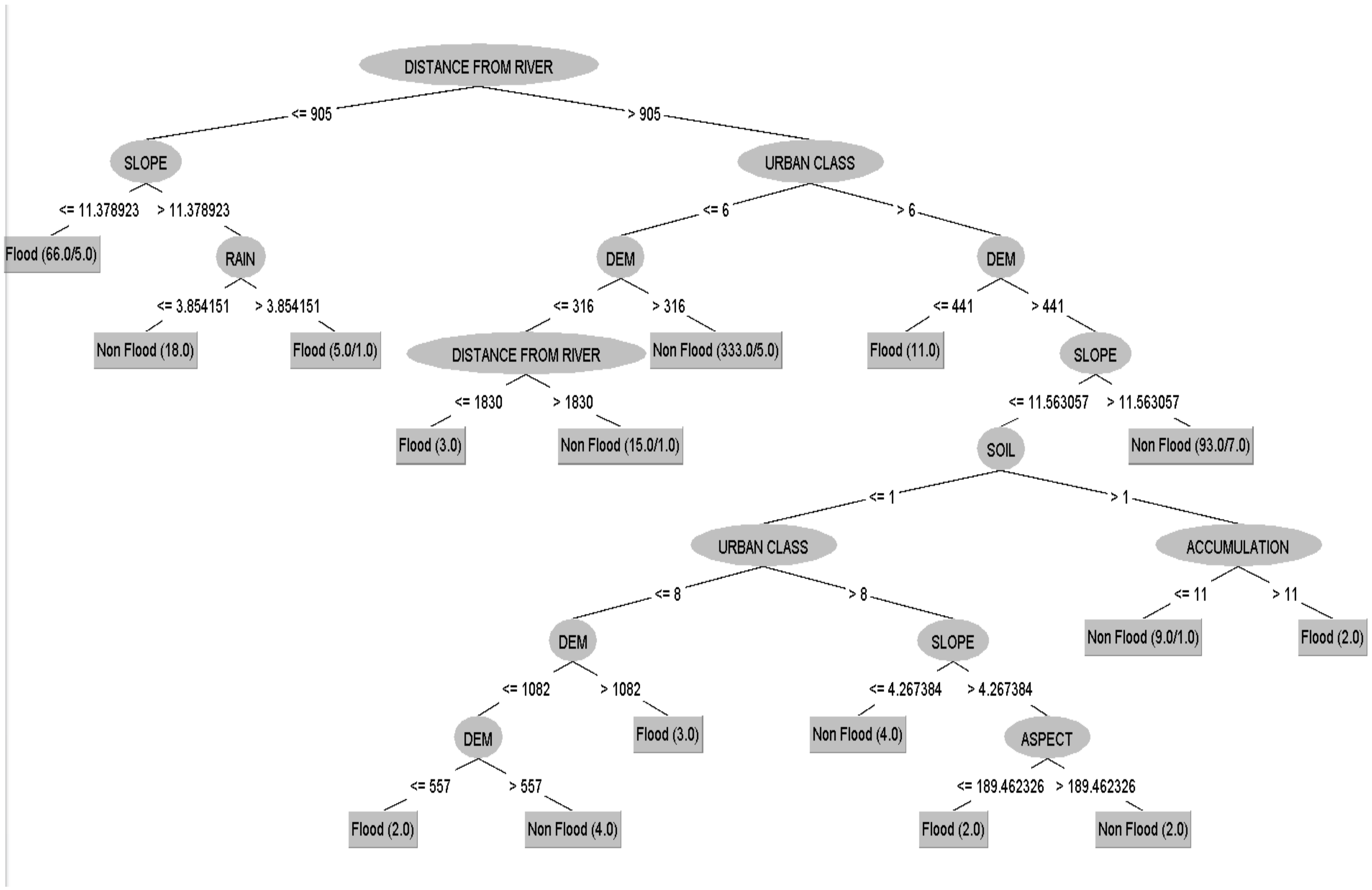

Figure 24. Decision tree showing rules for flood prediction

In the Decisions model, the Tree created and displayed rules for the dataset as a structure in a tree. A more critical effect on floods is the conditioning mechanisms that are in the order of the structure of the trees. The dependent data consisted of flood pixels and non-flood samples for the decision tree model. The most important attribute or predictor is "distance from the river."

1) The results from the Decision Tree indicate flash floods likely happen at these locations:
    - close to the river (distance from the river is less than 905 (m).
    - with a small slope (less than 11.38).
    - Under heavy rainfall (>3.85 mm/hr.).
2) Flash Food is also related to urbanization:
    - If an urban class is less than 6 (Impervious area <37.8%), then low places (DEM<316 m) and close to the river (distance to river <=1830 m) may experience a flash flood.
    - If urbanization is greater than 6 (Impervious area >37.8%), low places (DEM<431m) with high stream power (SPI) (>223.25) may experience flash flood.
    - If urbanization class is greater than 8 (Impervious area >44.1%), flash flood may occur at the areas with large slope (>4.26).
3) Flash Food is also related to soil type:
    - For soil types >1, high flow accumulation can result in a flash flood.

Section Two Assess the flash flood susceptibility

The analytic hierarchy process (AHP) is a mathematically organized approach for organizing and evaluating complicated choices. It was established in the 1970s by Thomas L. Saaty (Forman and Gass, 2001). The most basic methods evaluate the criteria based on their relevance and direct weight evaluation when the total of the evaluations equals one (100 percent) (Vojtek et al., 2019). AHP allows us to give a numerical value to the assessment criteria to assess their relative importance. AHP can assign a higher number to a more significant criterion (Indeed, 2021). AHP is used to evaluate the factors causing flash floods. AHP is a helpful technique for handling quantitative decision-making analysis. The current study used the AHP approach to create a flash flood assessment model based on the factors generating flash floods in the study region (Dano, 2020). AHP assists decision-makers in finding the one that best meets their objectives and knowledge of the problem (Forman and Gass, 2001). AHP was applied by (Al-shabeeb, 2016) to identify the most critical aspects of problematic rainwater catchment locations.

The Wadi Al-Ahsbah flood is identified using the AHP method, which uses Digital Elevation Models (DEM) and slope degree, which are the most critical factors, followed by river density, distance from rivers, flow accumulation, distance from roads, rainfall, stream power index (SPI), topographic wetness index (TWI), soil type data, urban area, and Normalized difference vegetation index (NDVI).

Using the weighted linear tables method and Pairwise comparisons, it is reasonable to estimate which areas will be affected by the floods. The AHP approach compares criteria using PCM matrices, allowing it to identify which criteria are of the highest importance. There are 9 PCMs, each ranging from 1 to 9. As an example, consider the following scale table (Table 12) (Al-shabeeb,2016). Table 13 shows the relative importance of each component, with the most important one being given the most weight. In this study, several factor weights were examined to identify the one that produced an excellent correlation with flood risk maps. AHP is a multiple-criteria decision-analytic (MCDA) technique that is integrated into GIS (Al-shabeeb, 2016)

Table 12. Scales for the pairwise comparisons method, adapted from (Al-shabeeb, 2016)

Intensity of importance	Definition	Explanation
1	Equal importance	Two elements contribute equally to the objective
3	Moderate importance	Experience and judgment slightly favor one element over another
5	Strong Importance	Experience and judgment strongly favor one element over another
7	Very strong importance	One element is favored very strongly over another, its dominance is demonstrated in practice
9	Extreme importance	The evidence favoring one element over another is of the highest possible order of affirmation
2,4,6,8 can be used to express intermediate values		

Table 13. A sample from the questionnaire used to determine the relative importance of criteria

Factor weighting score																		
criteria	Less importance than								Equal	More importance than								criteria
TWI	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	TWI
Elevation																		Elevation
Aspect																		Aspect
Flow Accumulation																		Flow Accumulation
Slope																		Slope
Precipitation																		Precipitation
Built Up																		Built Up
NDVI																		NDVI
SPI																		SPI
Distance from river																		Distance from river
Distance from road																		Distance from road
Drainage density																		Drainage density
Soil																		Soil

The next step is to normalize the matrix by adding all the numbers in each row and column. The opinion pairwise comparison matrix is shown below in Table 14.

A normalized score was derived for each column by multiplying the entries by the total sum of each column equals one.

Table 14. The opinion pairwise comparison matrix

Weight Matrix		TWI	DEM	Aspect	Flow Accumulation	Slope	Precipitation	Built Up	NDVI	SPI	Distance from river	Distance from road	Drainage density	soil	
	Number	1	2	3	4	5	6	7	8	9	10	11	12	13	SUM
TWI	1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	13.00
DEM	2	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	3.00	1.00	1.00	1.00	15.00
Aspect	3	1.00	1.00	1.00	1.00	1.00	2.00	1.00	1.00	2.00	1.00	2.00	1.00	1.00	16.00
Flow Accumulation	4	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	3.00	1.00	1.00	1.00	1.00	15.00
Slope	5	1.00	0.50	1.00	1.00	1.00	1.00	1.00	1.00	1.00	2.00	1.00	1.00	1.00	13.50
Precipitation	6	1.00	1.00	0.50	1.00	1.00	1.00	1.00	1.00	5.00	1.00	1.00	1.00	1.00	16.50
Built Up	7	1.00	1.00	1.00	1.00	1.00	1.00	1.00	3.00	5.00	5.00	5.00	5.00	1.00	31.00
NDVI	8	1.00	1.00	1.00	1.00	1.00	1.00	0.33	1.00	3.00	3.00	1.00	1.00	1.00	16.33
SPI	9	1.00	1.00	0.50	0.33	1.00	0.20	0.20	0.33	1.00	1.00	1.00	1.00	1.00	9.57
Distance from river	10	1.00	0.33	1.00	1.00	0.50	0.33	0.20	0.33	1.00	1.00	1.00	1.00	1.00	9.70
Distance from road	11	1.00	1.00	0.50	1.00	1.00	1.00	0.20	1.00	0.25	1.00	1.00	5.00	1.00	14.95
Drainage density	12	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	13.00
Soil	13	1.00	1.00	1.00	1.00	1.00	1.00	0.20	1.00	1.00	1.00	0.20	1.00	1.00	11.40
Sum		13.00	11.83	11.50	12.33	12.50	12.53	9.13	13.67	25.25	22.00	17.20	21.00	13.00	194.95

There are three parts involved in determining the consistency ratio:

1. Calculate the consistency measure in a spreadsheet in Excel software.
2. Calculate the consistency index (CI), also known as the consistency index of the data.

Consistency index is presented by Equation (7)

$$CI = \frac{\lambda max - n}{n - 1} \tag{7}$$

n = The order of the 13 flood factors.

RI = 1.56

Table 15. number of criteria

Number of criteria	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
R.I.	0.0	0.0	0.52	0.89	1.11	1.25	1.35	1.40	1.45	1.49	1.52	1.54	1.56	1.58	1.59

Consistency index (CI) = $\frac{14.701}{13-1}$

Consistency index (CI) = 0.1418

3. Calculate the consistency ratio (CI/RI), where RI is a random index. (Table 14)

Consistency Ratio = CI /RI > = 1.56

Consistency Ratio is calculated by Equation (8)

$$CR=CI/RI \tag{8}$$

Consistency Ratio (CR) = $\frac{CI}{MEAN\ RANDOME\ ci}$

= $\frac{0.1418}{1.56}$ = 0.090

The CR value is calculated by dividing the CI value by the Random Consistency Index (RCI), specified in Table 16.

Table 16. The estimated weights (priority vector), CI, RI, and CR for expert opinions

Weight Matrix	Number	SUM	Geometric Mean	Weighs	A3 Matrix = A1*A2	A4= A3/A2
TWI	1	13.00	1.00	0.08	0.98	13.00
Elevation	2	15.00	1.09	0.08	1.14	13.87
Aspect	3	16.00	1.17	0.09	1.24	13.93
Flow Accumulation	4	15.00	1.09	0.08	1.14	13.87
Slope	5	13.50	1.00	0.08	1.02	13.50
Precipitation	6	16.50	1.07	0.08	1.26	15.45
Built Up	7	31.00	1.79	0.14	2.41	17.80
NDVI	8	16.33	1.09	0.08	1.24	15.09
SPI	9	9.57	0.63	0.05	0.72	15.19
Distance from river	10	9.70	0.65	0.05	0.73	14.74
Distance from road	11	14.95	0.85	0.06	1.12	17.37
Drainage density	12	13.00	1.00	0.08	0.98	13.00
Soil	13	11.40	0.78	0.06	0.85	14.32
Sum		194.95	13.20	1.00		

Average	**14.701**
CI	0.1418
RI	1.56
C. Ratio	0.0909

The consistency ratio tells us how consistent the judgment matrix is. A higher number means less consistent, and a lower number means more consistency. For example, if the consistency ratio is 0.10 or less, the decision-makers answers are relatively consistent. Therefore, the researcher should seriously consider re-evaluating responses during pairwise comparisons for consistency ratios greater than 0.10 (Elsheikh et al., 2015).

This study showed a Consistency Ratio (CR) of 0.09, which was well below the threshold level of 0.1 and indicated high consistency. As a result, the weights are acceptable.

The flood susceptibility map was processed using ArcGIS software, using these steps: classify the factor data into five categories (very low, low, moderate, high, and very high); reclassify using the Spatial Analyst tool, and use the Spatial Analyst overlay tool. Thirteen flood contributing factors were used to collect the necessary data.

The analysis was conducted using the ArcGIS 10.8 software. First, all thirteen maps obtained from the pre-analysis processes were added, resulting in a map with five classifications: very low, low, moderate, high, and very high. The Spatial Analyst tool was then applied using weighted overlay and weighted sum tools. We added all elements to this tool, placing the weight of each factor and taking into account the classification scale from 1 to 5. Additionally, we ensured that the sum of all influence weights was equal to 1.

The order of the classifications in terms of importance must be determined before the process can be completed. For instance, although the Topographic wetness index TWI was arranged from low to high, Elevation and Build Up were sorted from high to low (Table 17). The final map of flood susceptibility was created by adding the weighted total of all conditioning factors to the flood map. After applying the sum of the weights of the conditioning factors, the flood susceptibility map was obtained. The final map of flood susceptibility was classified using the Jenks grading method. The Jenks Natural Breaks Classification (or Optimization) system is a data classification method that tries to put a set of values into "natural" groups. It is one of the best types of classification in a data set. Each group includes items that share characteristics and make sense together (George, 1967). Figure 25 shows five different classifications of flood-prone areas.

Table 17. Susceptibility class ranges and rating

Flood Causative Criterion	Unit	Class	Susceptibility Class Range and Rating	Susceptibility Class Rating	Weight (%)
Topographic wetness index	Index	2 – 5.3	Very Low	1	0.08
		5.4 – 7	Low	2	
		7.1 – 9.5	Moderate	3	
		9.6 – 13	High	4	
		14 – 24	Very High	5	
Elevation	M	100 – 430	Very High	5	0.08
		440 – 730	High	4	
		740 – 1,100	Moderate	3	
		1,200 – 1,700	Low	2	
		1,800 – 2,500	Very Low	1	
Aspect	Degree	72	Very Low	1	0.09
		72- 140	Low	2	
		150-220	Moderate	3	
		230-290	High	4	
		300- 360	Very High	5	
Flow Accumulation	m	1,200,000 – 1,200,000	Very High	5	0.08
		510,000 – 1,100,000	High	**4**	
		260,000 – 500,000	Moderate	3	
		58,000 – 250,000	Low	2	
		0 – 57,000	Very Low	1	
Slope	%	0 – 8.86	Very Low	1	0.08
		8.87 – 18.34	Low	2	

		18.35 – 27.51	Moderate	3	
		27.52 – 38.2	High	4	
		38.21 – 77.93	Very High	5	
Precipitation	MM/Huer	0.0 – 1.1	Very Low	1	0.08
		1.2 – 2.8	Low	2	
		2.9 – 4.6	Moderate	3	
		4.7 – 6.4	High	4	
		6.5 – 8.8	Very High	5	
Built Up	M	0 – 42	Very High	5	0.14
		43 – 76	High	4	
		77 – 120	Moderate	3	
		130 – 170	Low	2	
		180 – 200	Very Low	1	
NDVI		-1,900 – 1,200	Very High	5	0.08
		1,300 – 1,500	High	4	
		1,600 – 1,900	Moderate	3	
		2,000 – 2,500	Low	2	
		2,600 – 6,600	Very Low	1	
SPI	M	310000 – 480000	Very High	5	0.05
		210000 – 300000	High	**4**	
		110000 – 200000	Moderate	3	
		16000-100000	Low	2	
		0.030 – 15000	Very Low	1	
Distance from Roads	M	0 – 80	Very High	5	0.06
		81 – 160	High	4	
		170 – 240	Moderate	3	
		250 – 320	Low	2	

		330 – 400	Very Low	1	
Distance from River	M	0 – 39	Very High	5	0.05
		40 – 79	High	4	
		80 – 120	Moderate	3	
		130 – 160	Low	2	
		170 – 200	Very Low	1	
Drainage Density	Level	1	Very Low	1	0.08
		1.1 – 2	Low	2	
		2.1 – 3	Moderate	3	
		3.1 – 4	High	4	
		4.1 – 5	Very High	5	
Soil type	Loamy	1	Low	1	0.06
	Silty Clay	2	High	3	
	Clay	3	Moderate	2	

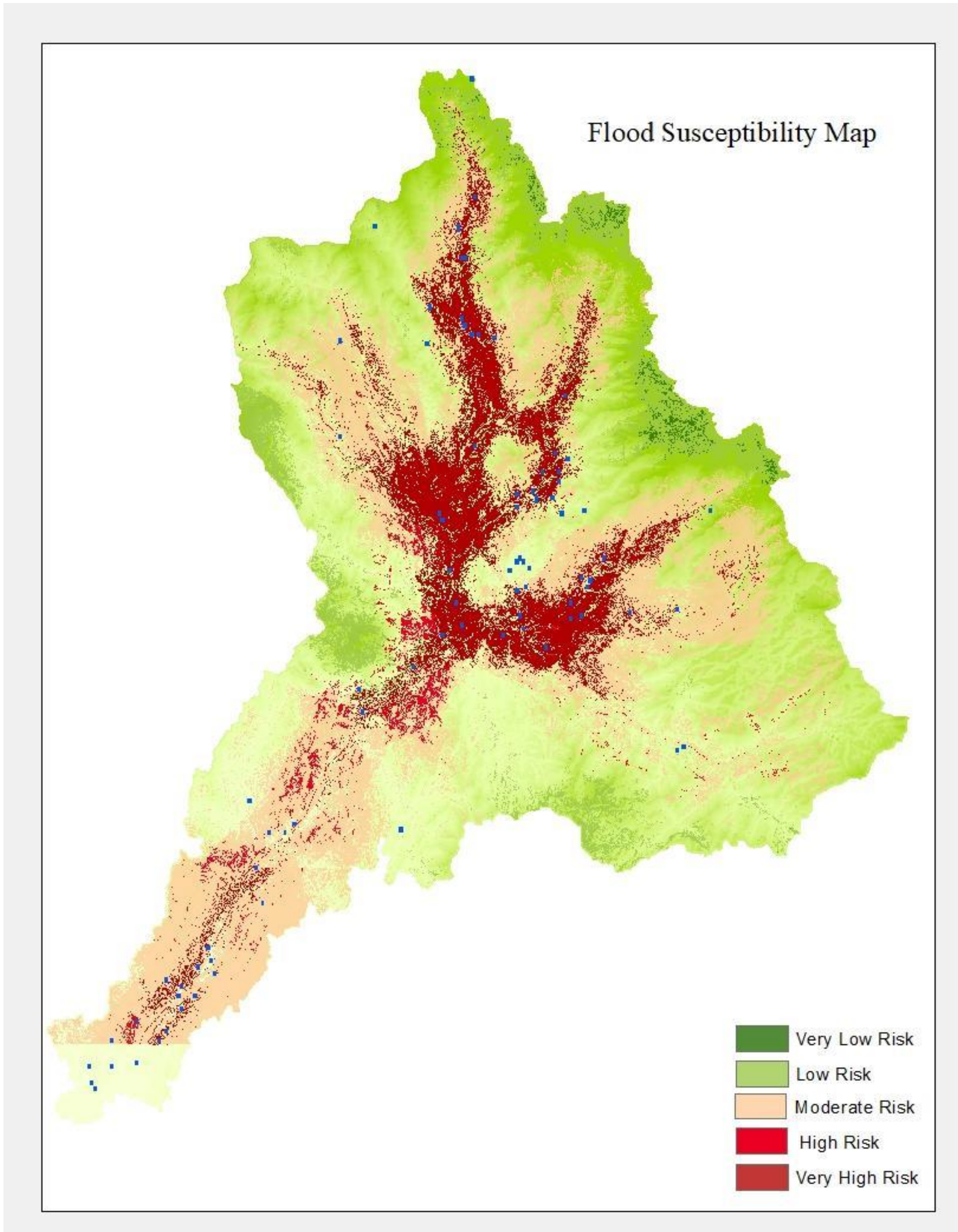

Figure 25. Flood susceptibility map

In the second stage, we measured the area of each category and divided it by the total study area. The result is shown in Table 18.

Table 18. Area of flood susceptibility map as a percentage of total area

Flood Risk	% of the Total Area
Very Low	5 %
Low	44 %
Moderate Risk	39 %
High	1 %
Very High	11 %
Total	100%

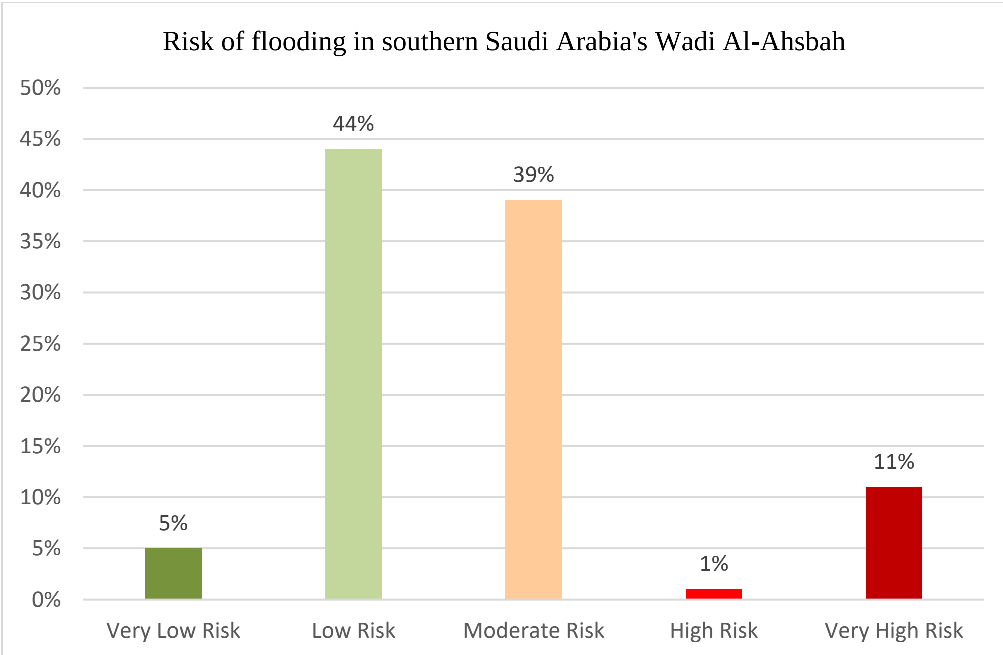

Figure 26. Risk of flooding in southern Saudi Arabia's Wadi Al-Ahsbah

The results we obtained from previous analyses found that 5% of the area is considered to have a very low risk of flooding, and floods are excluded in this area due to its height above ground level.

Also, 44% of the area was classified as having a low risk of flooding. Areas with a medium risk of flooding constitute 39% of the study area. It was found that 1% of the study area has a high risk of flooding, and 11%, has a very high risk of flooding, exposed to the dangers of flooding in rainy seasons. Table 18 and Figure 26 show the levels of risk in Wadi Al-Ahsbah region.

To understand this classification well, it is necessary to review the table of weights for each factor; the highest weight percentage was recorded for the built-up factor at a weight of 0.14; Elevation Models DEM, slope degree, and topographic wetness index TWI were weighted at .08 in Table 16.

Section Three Validate the Derived Flood Susceptibility Map Against Observation from Social Media

It is possible to determine flood-prone areas using analytical models, and we can also find some data through social media on the Internet to see if it confirms the flood-prone areas. Alharbi and Lee (2019) stated that social media platforms are a rich source of real-time information on crises. The study focused on high-risk floods via social media in the Arabic language. The classifiers used were classical machine learning (ML) and deep neural networks (DNN). The study found that deep learning was more effective than ML at identifying flood-related posts.

Furthermore, Sun et al. (2015) compared the findings to geotagged Flickr postings about floods and discovered that 95% of Flickr contributions happened within the ATMS-derived flood area.

In this study, ground observations from social media, including Twitter and YouTube, were used to validate the flood sites and learn the extent to which the danger areas match the flood sites recorded on social networking sites. The intersect tool was used in ArcGIS software to match each category and the flood points from social media. Next, we compared the locations of flood-prone areas and those not exposed to repeated flooding with all five classifications to determine how many sites were affected by the flood and how many were not affected. Table 19 shows the results of repeating this step ten times using the intersect tool. It was found that 47 social media sites match high-risk areas.

Table 19. Flood susceptibility classifications with flood location points

Risk Category	Number of Flood Locations from social media	Number of Non-Flood Locations from social media
Very Low Risk	2	30
Low Risk	10	134
Moderate Risk	23	364
High Risk	16	16
Very High Risk	47	12

Table 20. Confusion Matrix

Flood (High to very high risk from Prediction)	Number of Flood Locations (Observation)	Number of Non-Flood Locations (Observation)
TRUE	True positive TP	False positive FP
False	False negative FN	True negative TN

Table 21. Confusion Matrix2

Flood from Prediction (High to very high risk)	Number of Flood Locations (Observation)	Number of Non-Flood Locations (Observation)
True	**63**	**28**
False	**35**	**528**

FP: The false positive non-flood locations

TP: The true positive flood locations

FN: The false negative flood locations

TN: The true negative non-flood locations

Commission error is calculated from the following equation (9)

$$\textbf{Commission Error} = \frac{False\ Positive}{False\ Positive + True\ Positive} = \frac{False\ Positive}{total\ predicted} \tag{9}$$

$$= \frac{28}{28+63} = \frac{28}{91} = 30.76\%$$

Commission error was calculated as the number of flood locations incorrectly classified as a percentage of all non-flood areas in each non-flood location category. In addition, we found that 30% of the flood locations from social media sites were incorrect because they are in very low flood risk areas.

Omission Error is calculated by Equation (10)

$$Omission\ Error = \frac{False\ negative}{False\ negative + True\ Positive} = \frac{False\ negative}{Total\ reference} \tag{10}$$

$$= \frac{35}{35+63} = \frac{35}{98} = 35.71\%$$

We calculated the omission error as the number of true-flood locations incorrectly classified as non-flood as a percentage of all flood locations. We found that 35.71% of the identified sites with flooding may be predicted as non-flood in the future.

Overall Accuracy can be calculated by the following Equation (11)

Overall Accuracy $= \frac{True\ Positive + True\ negative}{True\ Positive + False\ Positive + True\ negative + False\ \ negative}$

$$= \frac{63 + 528}{63 + 35 + 528 + 28} = \frac{591}{654} = 90.37\% \qquad (11)$$

Validation of the flood susceptibility map results against observations from social media data showed a good result, reaching 90.37% overall accuracy (where 100% indicates perfect prediction). This percentage is not far from that result of using the analytic hierarchy process (AHP) with flood condition data.

There is a significant correlation between the flood locations identified by social media data and the areas determined by the AHP in this study, with roughly 63 sites occurring in very high and high-risk areas. In addition, 35 locations were found to have been shared by users outside the area of flood risk due to repeated share data from different locations not within the area of flood risk.

CHAPTER SEVENTH CONCLUSIONS AND DISCUSSIONS

In this study, multi-satellite data are collected to derive thirteen factors that may affect the incidence of flash floods in semiarid areas, like the Wadi Al-Hisbah region in southwestern Saudi Arabia.

There are limitations and challenges to accessing urban and population data. The Global Man-made Impervious Surface (GMIS) data was used to identify populated urban areas.

Classification accuracy of 91.25 % can be obtained from the Decision Tree algorithm. The Decision Tree results indicate that low-lying urban areas close to rivers with high impervious percentages and high accumulation flow may experience a flash flood in a semiarid region.

A flood susceptibility map in the Wadi Al-Hisbah region in southwestern Saudi Arabia is created with the practical application of the AHP technique. According to the flood susceptibility map, 1% of the region is classified as having a high probability of flooding, and 11% of the region has a very high probability of flooding. These studies show that specific locations are at a high risk of flooding during the rainy seasons. Additionally, the study discovered that 10% of the very high-risk area is located in a residential area, with some flooding incidents impacting agricultural activity at the bottom of the valley.

The derived flood susceptibility map results show a decrease in the flood area in the flatlands below the valley, while the vulnerable area increases in the sloping areas towards the middle, where the population is concentrated. The reason for this is the spread of agricultural lands below the valley. Also, the slope and construction in the water flow areas are the dominant factors.

To investigate the urbanization impacts on flash flood, impervious areas were given the highest weight during the creation of the flood susceptibility map. When comparing the impervious map with the social media observations, it was found that most of these flood locations from social media are concentrated in impervious urban areas.

When validated against the social media observations, 30.76% commission error and 35.71% omission error are found from the derived flood susceptibility map, while the overall accuracy can reach 90.37%.

It may be possible to improve the results of this risk map with the AHP technique by adding more data, such as the rainfall data from weather stations and soil moisture data from satellites.

Current satellite-based flood products may not capture flash floods, while conventional observations are usually sparse over the semiarid area. It is not easy for a researcher to obtain the data; a lot of time and effort may be needed. The cost is high even for social media data, making it difficult to obtain data for more events. To address the limited data availability in Saudi Arabia, some alternative approaches are needed to assess the spatial variability of flood risks.

As part of future works, our plans are in three areas: analyze flood-related messages from social media to evaluate our models; search for high-resolution data with a resolution of higher than 30 meters; and gather rainfall data from local monitoring stations to combine it with other variables that influence the likelihood of flooding.

REFERENCES

Aburizaiza, M. (2019). Determining Flood Prone Areas for An Ungauged Basin Using Topographic Wetness Index and Soft Data. Earth Systems and Geoinformation Sciences, George Mason University.

Alahmadi, M. & Atkinson, P. M. (2019). Three-Fold Urban Expansion in Saudi Arabia from 1992 to 2013 Observed Using Calibrated DMSP-OLS Night-Time Lights Imagery. *Remote Sensing*, 11 (19), p.2266.

Al-Bassam A. M, Zaidi F. K., & Hussein M.T. (2014). Natural Hazards in Saudi Arabia. Extreme Natural Hazards, Disaster Risks and Societal Implications (pp.243-251)

Alfieri, L., Salamon, P., Bianchi, A., Neal, J., Bates, P., & Feyen, L. (2014). Advances in pan-European flood hazard mapping. *Hydrological processes*, 28(13), pp.4067-4077.

Al-Ghamdi, K.A., Elzahrany, R.A., Mirza M.N., & Dawod G. M. (2012). Impacts of urban growth on flood hazards in Makkah City, Saudi Arabia. *International Journal Water Resource Environment Engineering*, 4, p.2334

Alharbi, A. & Lee M., 2019. Crisis Detection from Arabic Tweets. In Proceedings of the 3rd Workshop on Arabic Corpus Linguistics. pages 72–79, Cardiff, United Kingdom. Association for Computational Linguistics.

AL-Mazroui, M. (2020). Rainfall Trends and Extremes in Saudi Arabia in Recent Decades. *Atmosphere*, 11(9), p.964

Almazroui, M., Nazrul Islam, M., Athar, H., Jones, P. D., & Rahman, M. A. (2012). Recent climate change in the Arabian Peninsula: annual rainfall and temperature analysis of Saudi Arabia for 1978–2009. *International Journal of Climatology*, 32(6), pp.953-966.

Alrehaili, N. R. (2021). A systematic review of the emergency planning for flash floods response in the kingdom of Saudi Arabia. *The Australian Journal of Emergency Management*, 36(4), pp.82-88.

Al-Zahrani, M.I., Al-Harthy, S.G., Hawsawi, H.M., Al-Amawi, F.A., Al-Theban, M.S., Khemy, H.A., & Shen, Z. (2005). Saudi Threats to Floods in Wadi Qanounah and Wadi Al Ahsbah in the Southwestern Part of the Kingdom of Saudi Arabia, *The Saudi Geological Survey (SGS*), Thchnical report SGS-TR-2005-3

Al-Zarah, A.I., 2008. Chemistry of groundwater of Al-Ahsa Oasis eastern region Saudi Arabia and its predictive effects on soil properties. *Pakistan Journal of Biological Sciences*, 11(3), pp.332-341.

Ashraf, A. G. & Ahmed, F.D. (2019). Flood Risk Assessment of the Wadi Nu'man Basin. Mecca, Saudi Arabia (During the Period, 1988–2019) Based on the Integration of Geomatics and Hydraulic Modeling: A Case Study. Water, 11(9), p.1887.

Ashraf, A., Al-Alola, S. S., Alogayell, H.M., Mohamed,.H.M., Alkadi S.A. & Ismail I.Y. (2020). Integration of GIS-Based Multicriteria Decision Analysis and Analytic Hierarchy Process to Assess Flood Hazard on the Al-Shamal Train Pathway in Al-Qurayyat Region, Kingdom of Saudi Arabia. *Water*, 12(6), p.1702.

Ashraf, A.G., Ahmed, F.D., Youssef, A. M., & Pradhan, B.(2019). Flood Hazard Assessment of the Urban Area of Tabuk City, Kingdom of Saudi Arabia by Integrating Spatial-Based Hydrologic and Hydrodynamic Modeling. *Sensors*, 19 (5), p.1024.

Ayalew, L, Yamagishi, H. (2005). The application of GIS-based logistic regression for landslide susceptibility mapping in the Kakuda-Yahiko Mountains. *Central Japan. Geomorphology*. 65, pp.15–31.

Abrahams, A.D., (1972). Drainage densities and sediment yields in eastern australia, Australian Geographical Studies, *Wiley online library*. 10, pp.19-41.

Bahrawi, J., Ewea, H., Kamis, A, & Elhag, M. (2020). Potential flood risk due to urbanization expansion in arid environments, Saudi Arabia. *Natural Hazards*. 104, pp.795–809.

Beven, K.J., Kirkby, M. J., & Seibert J. (1979). A physically based, variable contributing area model of basin hydrology. *Hydrological Science Bulletin*, 24, pp.43-69.

Bhaduri, K., Wolff, R., Giannella, C., & Kargupta, H. (2008). Distributed decision-tree induction in peer-to-peer systems. Statistical Analysis and Data Mining: *The ASA Data Science Journal*, 1 (2), pp.85-103.

Bhattacharjee, N. (2016). Links among Drainage Density, Frequency and Flooding in Nanoi River Basin, Assam. *International Journal of Interdisciplinary Research in Science Society and Culture*, 2(1), pp.2395-4345.

Bonham-Carter, G.F. (1994). Geographic information systems for geoscientists: modeling with GIS. In Bonham-Carter F (ed) Computer methods in the geosciences. Pergamon, Oxford.

Brown de Colstoun, E. C., Huang C., Wang P., Tilton J. C., Tan B. J., Phillips, S., Niemczura, Ling P.Y., & Wolfe R. E. (2017). Global Man-made Impervious Surface (GMIS) Dataset from Landsat. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC).

Chen, Q., Lin, X., Lin, C., & Ji, R. (2013). Impact analysis of different spatial resolution DEM on object-oriented landslide extraction from high resolution remote sensing images. Ninth International Conference on Natural Computation (ICNC), 2013, pp. 940-945, doi: 10.1109/ICNC.2013.6818111.

da Cunha, L. K. (2012). Exploring the benefits of satellite remote sensing for flood prediction across scales (Doctoral dissertation, The University of Iowa).

Danielson, T. (2013). Utilizing a High-Resolution Digital Elevation Model (DEM) to Apply Stream Power Index (SPI) to the Gilmore Creek Watershed in Winona County, Minnesota. *Resource Analysis*, 15, p.11.

Dano, U.L. (2020). An AHP-based assessment of flood triggering factors to enhance resiliency in Dammam, Saudi Arabia. *GeoJournal*, DOI:10.1007/s10708-020-10363-5.

Dewan, A.M, Islam, M.M, Kumamoto, T., & Nishigaki, M. (2007). Evaluating flood hazard for land-use planning in greater Dhaka of Bangladesh using remote sensing and GIS techniques. *Water Resource Management*, 21(9), pp.1601–1612.

Didan, K., Huete, A., & SIPS, M. (2015). MOD13Q1 MODIS/Terra Vegetation Indices 16-Day L3 Global 250m SIN Grid. NASA LP DAAC.

El Alfy, M. (2016). Assessing the impact of arid area urbanization on flash floods using GIS, remote sensing, and HEC-HMS rainfall–runoff modeling. *Hydrology Research*, 47(6), pp.1142-1160.

Elfeki, A, Kamis, M.A, Kassab, A, Alamri, M, Chaabani, N.A., & Noor, K. (2019). Flash flood risk assessment in urban arid environment: case study of Taibah and Islamic universities' campuses. Medina, Kingdom of Saudi Arabia. Geomatics, *Natural Hazards, and Risk*, 10(1), pp.780-796.

Essel, B. (2019). The application of GIS in mapping of flood hazard areas and assessing of risk in Kumasi, Ghana. Evaluating the application of the statistical index method in flood susceptibility mapping and its comparison with frequency ratio and logistic regression methods. Geomatics, *Natural Hazards and Risk*, 10(1), pp.79-101 .

Fernández D.S, & Lutz M.A. (2010). Urban flood hazard zoning in Tucumán Province, Argentina, using GIS and multicriteria decision analysis. *Engineering Gelotology,* 111(1–4), pp.90–98.

Forman E. H., & Gass, S. I. (2001). The Analytic Hierarchy Process-An Exposition. Operations Research, 49(4), pp. 469-486.

Herfort, B., Albuquerque d.e, Schelhorn, J. P., S. J., and Zipf, A. (2014). Does the spatiotemporal distribution of tweets match the spatiotemporal distribution of flood phenomena? A study about the River Elbe Flood in June 2013. Proceedings of the 11th International ISCRAM Conference, January 2014.

Jenks, George F. (1967) The Data Model Concept in Statistical Mapping, *International Yearbook of Cartography* ,7, pp.186–190.

Khimi, H.A.H, Al-Balkhi, A.B., Al-Harthy, S.G., Al-Zahrani, M.I., & Al-Dahlawi, A.M. (2003). Possible Flood Risks in Wadi Besh in the southwestern part of the Kingdom of Saudi Arabia, Open-file report SGS-2003-2.

AL-azamil, W., & Al-Qarni, A. (2019). Urban sprawl on the natural environment in the Al-Baha region in the Kingdom of Saudi Arabia, the Geographical Environment Forum with Vision 2030, Princess Noura University.

Lee M-J, Kang J.e, & Jeon, S. (2012). Application of frequency ratio model and validation for predictive flooded area susceptibility mapping using GIS. IEEE International Proceedings of the Geoscience and Remote Sensing Symposium (IGARSS), p.4. Jul 22–27, Munich, Germany.

Lee, S. (2005). Application of logistic regression model and its validation for landslide susceptibility mapping using GIS and remote sensing data. *International Journal of Remote Sensing*, 26(7), pp.1477-1491.

Marco, C., Antonio, M., Javier, S., & Fernando, S. (2019). Flood Risk Evaluation in Urban Spaces: The Study Case of Tormes River (Salamanca, Spain). International *Journal Environmental Research Public Health*, 16(1), p.5, doi:10.3390/ijerph16010005.

Memon, A. A., Muhammad, S., Rahman, S., & Haq, M. (2015). Flood monitoring and damage assessment using water indices: A case study of Pakistan flood-2012. *The Egyptian Journal of Remote Sensing and Space Science*, 18(1), pp.99-106.

Middleton, S. E., Middleton, L., & Modafferi, S. (2013). Real-time crisis mapping of natural disasters using social media. *IEEE Intelligent Systems*, 29(2), pp.9-17.

Mojaddadi, H., Pradhan, B., Nampak, H., Ahmad, N. & bin Ghazali, A.H. (2017). Ensemble machine-learning-based geospatial approach for flood risk assessment using multi-sensor remote-sensing data and GIS. Geomatics, *Natural Hazards and Risk*, 8 (2), pp.1080-1102.

Moustapha, A.F., Costa, F.J., & Noble, A.G. (1985). Urban development in Saudi Arabia: Building and subdivision codes. *Cities*, 2(2), pp.140-148.

Munasinghe, D., Cohen S., Huang, Y.F., Tsang, Y.P., Zhang J., & Fang Z. (2018). Intercomparison of Satellite Remote Sensing-Based Flood Inundation Mapping Techniques. *JAWRA Journal of the American Water Resources Association*, 54(4), pp.834-846.

Myles, A. J., Feudale, R. N., Liu, Y., Woody, N. A., & Brown, S. D. (2004). An introduction to decision tree modeling. *Journal of Chemometrics*, 18(6), pp.275-285.

Neshat A. & Pradhan B. (2015). An integrated DRASTIC model using frequency ratio and two new hybrid methods for groundwater vulnerability assessment. *Natural Hazards*, 76, pp.543–563.

Opolot, E. (2013). Application of remote sensing and geographical information systems in flood management: a review. *Research Journal Of Applied Sciences Engineering And Technology*, 6(10), pp.1884-1894.

Ouerghi, R.F.A.S. & Elhag, A.R. (2015). Flood Risk Map Based on GIS, and Multi Criteria Techniques (Case Study Terengganu Malaysia), *Journal of Geographic Information System*, 7, pp. 348-357.

Peter, V. (2008). Saudi Arabia: an environmental overview. Taylor & Francis. p.141. ISBN 978-0-415-41387-9.

Petrovic, S., Osborne, M., McCreadie, R., Macdonald, C., Ounis, I., & Shrimpton, L. (2013). Can twitter replace newswire for breaking news? Proceedings of the International AAAI Conference on Web and Social Media, 7(1), pp.713-716.

Pettorelli, N. (2013). The normalized difference vegetation index. Oxford University Press.

Porio, E. (2011). Vulnerability, adaptation, and resilience to floods and climate change-related risks among marginal, riverine communities in Metro Manila. *Asian Journal of Social Science*, 39 (4), pp.425-445.

Pourali, S.H., Arrowsmith, C., Chrisman, N., Matkan, A. A., & Mitchell, D. (2016). Topography Wetness Index Application in Flood-Risk-Based Land Use Planning. *Applied Spatial Analysis and policy*, 9, pp.39–54.

Rahman, S. & Rahman, M. A. (2015). Climate extremes and challenges to infrastructure development in coastal cities in Bangladesh. *Weather and Climate Extremes*, 7, pp.96- 108.

Regmi A.D., Devkota K.C., Yoshida, K., Pradhan, B, Pourghasemi, H.R, Kumamoto, T, & Akgun, A. (2013). Application of frequency ratio, statistical index, and weights-of-evidence models and their comparison in landslide susceptibility mapping in Central Nepal Himalaya. *Arabian Journal Geosciences*, 7(2), pp.725–742.

Righini, M., Surian, N., Wohl, E., Marchi, L., Comiti, F., Amponsah, W. & Borga, M. (2017). Geomorphic response to an extreme flood in two Mediterranean rivers (northeastern Sardinia, Italy): Analysis of controlling factors. *Geomorphology*, 290, pp.184-199.

Saaty, T.L, & Vargas, L.G. (1991). Prediction, Projection and Forecasting. Kluwer Academic Publishers, Dordrecht, 251 pp., ISBN: 978-94-015-7954-4.

Saksena, S., & Merwade, V. (2015). Incorporating the Effect of DEM Resolution and Accuracy for Improved Flood Inundation Mapping. *Journal of Hydrology*, 530, pp.180-194.

Sarkar, D., & Mondal, P. (2020). Flood vulnerability mapping using frequency ratio (FR) model: a case study on Kulik river basin, Indo-Bangladesh Barind region. Applied Water Science, 10, p.17.

Şen, Z., Al-Harithy, S., As-Sefry, S., & Almazroui, M. (2017). Aridity and Risk Calculations in Saudi Arabian Wadis: Wadi Fatimah Case. *Earth Systems and Environment*, 1(2), pp.1–11.

Sharif, H. O., Al-Juaidi, F. H., Al-Othman, A., Al-Dousary, I., Fadda, E., Jamal-Uddeen, S., & Elhassan, A. (2016). Flood hazards in an urbanizing watershed in Riyadh, Saudi Arabia. Geomatics, *Natural Hazards and Risk*, 7(2), pp.702-720.

Sitabi, A. (2015). DEM Analysis – The many uses and derivatives of a Digital Elevation Model. Earth Data Analysis Center, University of New Mexico.

Small, C., Sousa, D., Yetman, G., Elvidge, C., & MacManus, K. (2018). Decades of urban growth and development on the Asian megadeltas. *Global and Planetary Change*, 165, pp.62-89.

Smith, L., Liang, Q., James, P., & Lin, W. (2017). Assessing the utility of social media as a data source for flood risk management using a real-time modelling framework. *Journal of Flood Risk Management*, 10(3), pp.370-380.

Souissi, D., Msaddek, H.M., Zouhri, L., Chenini, I., EL May, M., & Dlala, M. (2018). Mapping groundwater recharge potential zones in arid region using GIS and Landsat approaches, southeast Tunisia. *Hydrological Sciences Journal*, 63(2), pp.251-268.

Subyani, A. M., & Al-Modayan, A. A. (2011). Flood analysis in western Saudi Arabia. Journal of King Abdulaziz University, *Earth Sciences*, 22(2), pp.17–36.

Sun, D., Li, S., Zheng, W., Croitoru, A., Stefanidis, A., & Goldberg, M. (2015). Mapping floods due to Hurricane Sandy using NPP VIIRS and ATMS data and geotagged Flickr imagery. *International Journal of Digital Earth*, 9 (5), pp.427-441.

Tarboton, D.G. (1997). A New Method for the Determination of Flow Directions and Contributing Areas in Grid Digital Elevation Models. Water Resources Research, 33(2), pp.309-319

Tehrany M.S, Pradhan B, & Jebur M.N. (2013). Spatial prediction of flood susceptible areas using rule-based decision tree (DT) and a novel ensemble bivariate and multivariate statistical model in GIS. *Journal of Hydrology*, 504, pp.69–79.

Tehrany, M.S., Kumar, L. (2018). The application of a Dempster–Shafer-based evidential belief function in flood susceptibilitymapping and comparison with frequency ratio and logistic regression methods. *Environmental Earth Sciences*, 77, p.490.

Tehrany, M.S., Kumar, L., Jebur, M.N., & Shabani, F (2018). Evaluating the application of the statistical index method in flood susceptibility mapping and its comparison with frequency ratio and logistic regression methods. Geomatics, *Natural Hazards and Risk*, 10(1), pp.79-101.

Tehrany, S., Shabani, F., Jebur, M., Hong, H., Chen, W., & Xie, X. (2017). GIS-based spatial prediction of flood prone areas using standalone frequency ratio, logistic regression, weight of evidence and their ensemble techniques. Geomatics, *Natural Hazards and Risk*, 8(2), pp.1538–1561.

Vojtek, M. & Vojteková, J. (2019). Flood Susceptibility Mapping on a National Scale in Slovakia Using the Analytical Hierarchy Process. *Water*, 11(2), p.364.

Wang, P., Huang, C., Brown de Colstoun, E. C., Tilton, J. C., & Tan, B. (2017). Global Human Built-up And Settlement Extent (HBASE) Dataset From Landsat. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC).

Wang, Y., Fang, Z., Hong, H., & Peng L. (2020). Flood susceptibility mapping using convolutional neural network frameworks. *Journal of Hydrology*, 582, p.124482, doi: 10.1016/j.jhydrol.2019.124482.

Webster, P. J., Toma, V. E., & Kim, H. M. (2011). Were the 2010 Pakistan floods predictable? Geophysical Research Letters, 38, L04806, doi:10.1029/2010GL046346.

Wolman, M. G. (2001). Evaluating alternative techniques floodplain mapping. *Water Resources Research*, 7(6), pp.1383-1392.

Wu, S. J., Lien, H. C., & Chang, C. H. (2010). Modeling risk analysis for forecasting peak discharge during flooding prevention and warning operation. *Stochastic Environmental Research and Risk Assessment*, 24(8), pp.1175-1191.

Xia, J., Falconer, R. A., Lin, B., & Tan, G. (2011). Numerical assessment of flood hazard risk to people and vehicles in flash floods. *Environmental Modelling & Software*, 26(8), pp.987-998.

Youssef, A. M., Sefry, S. A., Pradhan, B., & Alfadail, E. A. (2016). Analysis of causes of a flash flood in Jeddah city (Kingdom of Saudi Arabia) of 2009 and 2011 using multi-sensor remote sensing data and GIS. Geomatics, *Natural Hazards, and Risk*, 7(3), pp.1018–1042.

Zhu, F., Zhong, P. A., Sun, Y., & Yeh, W. W. G. (2017). Real-time optimal flood control decision making and risk propagation under multiple uncertainties. *Water Resources Research*, 53(12), pp.10635-10654

Websites:

Alison, H., (2020). Digital Soil Map of the World. Retrieved from Harvard website: http://worldmap.harvard.edu/data/geonode:

Al-Zamil, W.S., (2019). Urban sprawl on the natural environment in Al-Baha region in the Kingdom of Saudi Arabia. Retrieved from King Saud University website:https://faculty.ksu.edu.sa/ar/waalzamil/publication/288056.

DSMW_RdY NASA, (2021). 30-Meter SRTM Tile Downloader. Retrieved from https://dwtkns.com/srtm30m/

Global Facility for Disaster Reduction and Recovery (GFDRR), (2017). Middle East and North Africa. Retrieved from https://www.gfdrr.org/en/region/middle-east-and-north-africa-mena

NASA. Giovanni. Retrieved from NASA website: https://giovanni.gsfc.nasa.gov/giovanni/#service=AcMp&starttime=&endtime=

Population Pyramid.net. (2020). PopulationPyramid.net Population pyramid of the world from 1950 to 2100, made available under a Creative Commons license CC BY 3.0 IGO: http://creativecommons.org/licenses/by/3.0/igo/, Retrieved from https://www.populationpyramid.net/ar/%D8%A7%D9%84%D8%B3%D8%B9%D9%88%D8%AF%D9%8A%D8%A9/2020/

Wikipedia, (2022). Saudi Arabia. Retrieved from Wikipedia website: https://en.wikipedia.org/w/index.php?title=Saudi_Arabia&oldid=1012027486

BIBLIOGRAPHY

Eidah Ghurm Alzahrani graduated from Seventh High School, Makkah, Saudi Arabia, in 2002. She received her Bachelor of Social Sciences degree in the Department of Geography and a minor degree in Education at Umm AL-Qura University in 2006. She was employed as a teacher in Umm AL-Qura University for two years and received her Master of Science degree in the Department of Geography/ Geographic Information Sciences at the University of Akron in 2015.