

Mapping Drought Risk in Newly Merged Districts of KP using Geospatial and Machine Learning Techniques



Author

SAAD IQBAL

(Reg # 52-FET/MS CE/F23)

Supervised by

Dr. Rashid Farooq

Associate Professor

Department of Civil Engineering
Faculty of Engineering & Technology
International Islamic University Islamabad

2026

Mapping Drought Risk in Newly Merged Districts of KP using Geospatial and Machine Learning Techniques



Author

Saad Iqbal

Reg # 52-FET/MSCE/F23

Supervisor

Dr. Rashid Farooq

Associate Professor

Department of Civil Engineering
Faculty of Engineering & Technology
International Islamic University Islamabad
2026

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

CERTIFICATE OF APPROVAL

Title of Thesis:

**“Mapping Drought Risk in Newly Merged Districts of KP
using Geospatial and Machine Learning Techniques”**

Name of Student : Saad Iqbal

Registration No: **52-FET/MSCE/F23**

Accepted by the Department of Civil Engineering, Faculty of Engineering and Technology,
International Islamic University IIUI, Islamabad, in partial fulfilment of requirements of
Master of Science in Civil Engineering.

Viva Voce Committee:

Dr. Rashid Farooq (Supervisor)

Associate Professor, DCE, FET, IIU Islamabad



Prof. Dr Waqas Haroon (Internal Examiner)

Lecturer, DCE, FET, IIU Islamabad



Dr. Afzal Ahmed (External Examiner)

Assistant Professor, DCE, UET Taxila



Prof. Dr Abid Ali Shah (Chairman DCE)

Professor/Chairman, DCE, FET, IIU Islamabad



Chairman
Department of Civil Engineering
Faculty of Engineering & Technology
International Islamic University
Islamabad

A thesis submitted to the Department of Civil Engineering, International Islamic University, Islamabad, as a partial fulfillment of requirements for the award of the MS Degree in Civil Engineering

DEDICATION

First and foremost, I thank Almighty Allah, the Most Merciful and the Most Compassionate, for His countless blessings, guidance, and strength throughout this journey. Without His grace, this achievement would not have been possible.

I dedicate this work to my beloved parents, whose unwavering love, sacrifices, and constant prayers have shaped me into the person I am today. It is their heartfelt supplications that have guided me through the toughest of times, and their encouragement has been a source of immense strength and motivation.

I am deeply grateful to Dr. Rashid Farooq, my supervisor, for his invaluable guidance, expertise, and continuous support throughout this research. His insightful feedback, encouragement, and patience were instrumental in the completion of this thesis.

This thesis is a reflection of the values of love, hard work, and dedication that each of you has instilled in me. May Allah bless you all.

Saad Iqbal

DECLARATION

I hereby declare that the work present in the following thesis is my own effort, except where otherwise acknowledged and that the thesis is my own composition. No part of the thesis has been previously presented for any other degree.

Saad Iqbal

52-FET/MS CE/F23

CONTENTS

DEDICATION	VI
DECLARATION	VII
CONTENTS	VIII
ACKNOWLEDGEMENTS	XIII
1 Introduction	3
1.1 Background	3
1.2 Research Question.....	4
1.3 Problem Statement	5
1.4 Significance of Research	6
1.5 Aim of the Study:	7
1.6 Objective of the Research:.....	7
1.7 Definition of Key Terms in Thesis Statement	7
1.7.1. Drought	7
1.7.2. Drought Risk	7
1.7.3. Geospatial Analysis	8
1.7.4. Machine Learning	8
1.7.5. Hazard.....	8
1.7.6. Exposure	8
1.7.7. Vulnerability	8
1.7.8. Newly Merged Districts (NMDs)	9
1.8 Limitations	9

1.8.1. Scope Limitations.....	9
1.8.2. Chronological Limitations.....	9
1.8.3. Data Limitations	9
1.8.4. Methodological Limitations.....	10
1.8.5. Contextual Limitations.....	10
1.9 Organization of the study	10
2 Literature Review.....	12
2.1 Background.....	12
2.2 International perspective on Drought monitoring.....	13
2.2.1. Meteorological Indices	13
2.2.2 Remote Sensing Approaches.....	14
2.3 Regional Context: South Asia	15
2.3.1 Drought in India	15
2.3.2 Drought in Bangladesh.....	15
2.3.3 Drought in Nepal and Sri lanka.....	15
2.4 Drought Studies in Pakistan	16
2.4.1 National-level Assessment of Drought.....	16
2.4.2 Remote Sensing for drought monitoring.....	17
2.4.3 Studies on Hydrological and Agricultural Drought	18
2.4.5 Synthesis of Drought Research in Pakistan.....	18
2.5 Drought Research: applications of machine learning.....	19
2.5.2 Deep Learning Approaches	19
2.5.3 Hybrid Models and Data Fusion	20
2.5.4 Applications in South Asia.....	20

2.6	Vulnerability and Risk Frameworks	20
2.6.1	<i>Exposure and Sensitivity</i>	21
2.6.2	<i>Adaptive Capacity</i>	21
2.6.3	<i>Pakistan Risk/Vulnerability</i>	21
2.7	Identification of research gaps.....	22
2.7.1	<i>Global Advances</i>	22
2.7.2	<i>Regional Developments South Asia</i>	22
2.7.3	<i>Pakistan-Specific Insights</i>	23
2.7.4	<i>Identified Research Gaps</i>	23
2.7.5	<i>Research Justification</i>	23
2.8	Conclusion	24
3	Experimental Methodology	25
3.1	Study Area	25
3.2	Data Sources and Indices	26
3.2.1	<i>Drought Indices</i>	26
3.3	Land Use and Land Cover (LULC) Mapping.....	27
3.4	Multi-Criteria Decision Analysis (MCDA) Framework	28
3.4.1	<i>Saaty's Scale for Ranking</i>	28
3.4.2	<i>Consistency Check</i>	28
3.4.3	<i>Weighted Overlay Analysis</i>	28
3.5	<i>Machine Learning</i>	30
3.6	<i>Workflow</i>	31
4	Results & Discussion	33
4.1	<i>Raster analysis</i>	33

4.1.1 aSPI Results (2000, 2008, 2016, 2024).....	33
4.1.2 TCI Results (2000, 2008, 2016, 2024)	34
4.1.3 NDVI Results (2000, 2008, 2016, 2024).....	35
4.1.4 VCI Results (2000, 2008, 2016, 2024).....	36
4.1.5 SMI Results (2000, 2008, 2016, 2024).....	37
4.1.6 LULC Results (2000, 2008, 2016, 2024)	38
4.1.7 TVDI Results (2000, 2008, 2016, 2024)	39
4.1.8 SPEI Results (2000, 2008, 2016, 2024)	40
4.1.9 SPI Results (2000, 2008, 2016, 2024).....	41
4.1.10 VHI Results (2000, 2008, 2016, 2024).....	42
4.2 Statistical Analysis	43
4.2.1. Climate Drought Trends (SPI, SPEI, aSPI).....	43
4.2.2 Soil condition (TSS, suspended Solids).....	44
4.2.3 Stress Trends of Soil/Surface (SMI, TVDI, TCI).....	45
4.2.4 Heatmap of Correlation of Drought Indices.....	46
4.2.5 SPI vs NDVI (Climate vs Vegetation Greenness)	47
4.2.6 SPEI vs SMI (Climate Water Balance vs Soil Moisture)	48
4.2.7 District wise analysis	49
4.3 AHP-based Drought Severity Maps Interpretation.....	50
4.4 Use of Machine Learning Algorithm i.e., Random Forest for Drought Classification.....	52
4.5 Discussion	55
4.4.1 Drought Indices based on climate (SPI, SPEI, aSPI).....	56
4.4.2 Indices of Vegetation Condition: VCI, VHI, VNI.....	56

4.4.3 Soil and Surface Stress indices (SMI, TVDI, TCI).....	57
4.4.4 Land Use and Land Cover (LULC) Dynamics	58
4.4.5 AHP-based Drought Risk.....	58
4.4.6 Statistical Relationship and Cross-Index Relationships	59
4.4.7. Application of Machine learning in analyzing drought trends.....	59
4.4.8 Conclusion of Discussion.....	60
5 Conclusion.....	62
Recommendations	63
Policy and Practical Implications	64
References.....	66

ACKNOWLEDGEMENTS

I am particularly thankful to Almighty Allah for the completion of my MS Research work, and I acknowledge the support of Dr. Rashid Farooq in the completion of this work on “Mapping Drought Risk in Newly Merged Districts of KP using Geospatial and Machine Learning Techniques”. I am deeply grateful to my parents, who have been my greatest mentors, always supporting and encouraging me during my research and studies at IIUI.

LIST OF FIGURES

Figure 2- 1: Average Annual Natural Hazard Occurrence (1980-2020).....	13
Figure 2- 2: Key natural hazard statistics for 1980-2020, source: Eco Regional Center for Risk Management of Natural Disasters (ECO-RCRM), https://eco-rcrm.ir/natural-disasters/	17
Figure 3- 1: Study area i.e., newly merged districts of KP.....	25
Figure 3- 2: Flowchart of methodology.....	32
Figure 4- 1: aSPI raster map	33
Figure 4- 2: TCI raster map	34
Figure 4- 3: NDVI raster map	36
Figure 4- 4: VCI raster map	37
Figure 4- 5: SMI raster map	38
Figure 4- 6: LULC raster map.....	39
Figure 4- 7: TVDI raster map.....	40
Figure 4- 8: SPEI raster map	41
Figure 4- 9: SPI raster map	42
Figure 4- 10: VHI raster map	43
Figure 4- 11: Climate drought trends (SPI, SPEI, aSPI)	44
Figure 4- 12: Soil condition (vegetation trends)	44
Figure 4- 13: Soil surface trends	45
Figure 4- 14: Heatmap of correlation of drought indices	46
Figure 4- 15: SPI vs NDVI.....	47
Figure 4- 16: SPEI vs SMI.....	48
Figure 4- 17: District wise analysis of SPEI and NDVI	49
Figure 4- 18: AHP based mapping of drought in newly merged districts of KP	50
Figure 4- 19: Random Forest based classification of drought in newly merged districts	53
Figure 4- 20: Random forest drought severity classification	55

LIST OF TABLES

Table 2- 1: Selected Drought Monitoring Systems in South Asia..... 16

Table 2- 2: Selected Drought Studies in Pakistan..... 18

Table 3- 1: **Datasets Used in the Study** 27

Table 3- 2: Saaty’s Fundamental Scale for Pairwise Comparisons 28

Table 3- 3: Weights and Influence %ages of Drought Indicators..... 29

LIST OF ABBREVIATIONS

Abbreviation	Full Form
AHP	Analytical Hierarchy Process
aSPI	Agricultural Standardized Precipitation Index
FATA	Federally Administered Tribal Areas
KP	Khyber Pakhtunkhwa
LULC	Land Use and Land Cover
NDVI	Normalized Difference Vegetation Index
SMI	Soil Moisture Index
SPI	Standardized Precipitation Index
SPEI	Standardized Precipitation Evapotranspiration Index
VCi	Vegetation Condition Index
RF	Random Forest
GEE	Google Earth Engine
ERA5	ECMWF Reanalysis v5 (European Centre for Medium-Range Weather Forecasts)
VHI	Vegetation Health Index
TCI	Temperature Condition Index
TVDI	Temperature Vegetation Dryness Index
DEM	Digital Elevation Model
GIS	Geographic Information System
MODIS	Moderate Resolution Imaging Spectroradiometer
RMSE	Root Mean Square Error

Abstract

Drought is among the most chronic and devastating natural hazards in arid and semi-arid areas, and it has drastic effects on the ecosystems, agriculture, and livelihoods. Khyber Pakhtunkhwa (KP), especially the newly integrated districts (formerly FATA), is highly vulnerable due to its fragile topography, a susceptible climate, and reliance on rain-fed farming. This research will map and model drought risk throughout these districts using a geospatial and machine learning framework, incorporating climatic, vegetation, soil, and land surface variables.

Nine drought indices were used in the research, i.e., Standardized Precipitation Index (SPI), Standardized Precipitation Evapotranspiration Index (SPEI), aggregated SPI (aSPI), Normalized Difference Vegetation Index (NDVI), Vegetation Condition Index (VCI), Vegetation Health Index (VHI), Soil Moisture Index (SMI), Temperature Vegetation Dryness Index (TVDI), and Temperature Condition Index (TCI). The Landsat images were used to develop Land Use and Land Cover (LULC) maps through supervised classification, utilizing the Random Forest algorithm. The Analytical Hierarchy Process (AHP) was then integrated, based on Saaty's scale, to combine indices and create composite drought severity maps for 2000, 2008, 2016, and 2024. The Random Forest model, trained on ERA-5 climate variables, achieved high performance with overall accuracy above 85% and a kappa coefficient of 0.78, demonstrating its suitability for handling complex, nonlinear climate–environment relationships.

Machine learning played a central role in this study, enabling the robust classification of drought severity. Unlike conventional statistical models, Random Forest builds an ensemble of decision trees and uses majority voting to determine outcomes. This ensemble strategy reduces overfitting and improves prediction consistency, particularly with high-dimensional environmental data. The model successfully categorized the study area into five drought classes: extreme, severe, moderate, mild, and no drought, generating reliable results that capture both spatial heterogeneity and temporal variability. The results revealed clear spatiotemporal patterns.

In 2000, a moderate drought affected ~42% of the area, while an extreme drought impacted around 18%, primarily in the south. By 2008, extreme and severe droughts expanded significantly, covering ~54% of the region, particularly in South Waziristan and Orakzai. In 2016, a partial recovery was observed, with 60% of the districts experiencing mild to moderate drought. However, by 2024

projections, extreme drought re-emerged in the southern belt, covering ~48% of the land, while northern districts, such as Bajaur, showed resilience, with only about 12% experiencing extreme drought. These findings confirm that South Waziristan is consistently the most vulnerable district, whereas Bajaur remains relatively resistant due to favorable climatic and vegetative conditions. The cyclical nature of drought, coupled with land use change, highlights the multidimensional character of this hazard. The implications are significant for policy and planning. Persistent drought hotspots in the southern belt necessitate urgent investments in water management, irrigation efficiency, and climate-resilient agriculture. Machine learning-driven drought mapping further strengthens decision-making by revealing complex risk patterns, allowing policymakers to prioritize vulnerable zones, allocate resources effectively, and design adaptation strategies. The findings demonstrated evident spatiotemporal trends. In 2000, a moderate drought prevailed over an area of approximately 42 %, with extreme drought affecting about 18 % of the area, primarily located in the south. As of 2008, extreme and severe droughts had grown to enormous proportions in the area, covering almost half of the region, particularly in South Waziristan and Orakzai. A partial recovery was observed in 2016, with 60 % of the districts experiencing a mild to moderate drought. This was, however, predicted to revert to extreme drought in the southern belt in 2024, spanning an area of around 48 %, whereas the north, including the Bajaur districts, was resilient, with the extreme drought only covering about 12 %. These results verify that South Waziristan is consistently the most susceptible district, while Bajaur is relatively resistant due to its conducive climatic and vegetative factors. Drought cyclicity, combined with land-use alteration, highlights the multidimensional nature of this hazard. These have far-reaching policy and planning implications. The hotspots of persistent droughts along the southern belt require urgent investments in water management, irrigation efficiency, and climate-resilient agriculture. Drought mapping (using machine learning) also enhances the strength of decision-making by demonstrating the complex patterns of risks, which help policymakers prioritize the vulnerable areas, allocate resources efficiently, and develop adaptation strategies.

Keywords: Drought, NDVI, TCI, VHI, VCI, KP, Newly merged districts

1 Introduction

1.1 Background

Drought is one of the most devastating natural disasters worldwide, posing a significant threat to ecosystems, livelihoods, and economies. Contrary to rapid-onset disasters, such as floods or earthquakes, drought has a delayed emergence, lasting for months or years, and its impacts are multi-dimensional, including agricultural losses, water scarcity, ecosystem degradation, food insecurity, and socio-political instability (Wilhite, 2005). Historically, South Asian-based countries, including Pakistan, have been prone to drought due to highly unpredictable monsoonal rain patterns, semi-arid climates, and augmented anthropogenic pressures on arid and water resources.

In Pakistan, drought is not just a climatic phenomenon; it also overlaps with structural vulnerabilities, including poor infrastructure, reliance on rain-fed agriculture, and a lack of robust institutional capacity to mitigate the risk. Numerous cataclysmic droughts have hit the nation, with the worst hit being the one in 1998-2002 that severely impacted the Balochistan and Sindh states, decimating crop production, groundwater, and aggravating poverty in the rural areas (A. Ahmad et al., 2020). These incidents brought to the limelight the instability of the agricultural systems as well as a lack of sufficient disaster management techniques.

This situation is even more severe in the Newly Merged Districts (NMDs) of Khyber Pakhtunkhwa (KP), which were formerly known as the Federally Administered Tribal Areas (FATA). A part of the most marginalized regions in Pakistan, these districts are weakly developed with weak ecosystems and complicated socio-political backgrounds. Before their merger with KP in 2018, FATA had decades of neglect, conflict, and governance issues, making its communities very susceptible to environmental stress factors, including drought (Yousaf, 2019). Famine farming is the predominant form of livelihood in these districts, attributed mainly to drought. It is known that the problem primarily stems from the fact that NMDs are located in semi-arid or arid zones, where rainfall variability is high and water resources are limited. Climate research conducted through satellite observations indicates that the region experiences frequent droughts, which are occurring with growing frequency, and are partly associated with climate change and local hydrological changes (Aalbers, Lenderink, van Meijgaard, & van den Hurk, 2018). Droughts decrease agricultural output, overstretch limited water supplies, and increase tensions on access to land and water.

Nevertheless, little is known about the spatial intimacy and risk dispersion of drought in the recently assembled districts, and how the risk can be systematically monitored and forecasted using contemporary geospatial and machine learning approaches. The Pakistani trifle traditional drought evaluations have integrated the rainfall abnormalities or Standardized Precipitation Index (SPI), which is frequently not enough to understand the many-dimensional determinants of drought hazard (Svoboda, Hayes, & Wood, 2012). Additionally, the evaluations seldom focus on territories such as marginalized regions, such as the NMDs, that pose a serious knowledge gap in policy and planning.

The absence of localized and data-led drought risk mapping poses significant challenges to developing effective adaptation and mitigation strategies. Without precise drought risk maps, the allocation of resources will not be conducted properly, there will be fewer early warning systems, and populations will continue to be at risk. Thus, addressing this knowledge gap by employing modern methods, including remote sensing, spatial analysis in GIS, and machine learning modeling, is urgent and necessary.

1.2 Research Question

History shows that drought has been a frequent and devastating threat in Pakistan, with hazardous consequences in the Newly Merged Districts (NMDs) of Khyber Pakhtunkhwa. Although literature on drought in the dry areas of Sindh, Balochistan, and Southern Punjab has been widely reported, very few research studies have been conducted on the spatial occurrence of drought risk in the NMDs. (Iqbal and Khan 72).

More traditional drought measurements in the area often rely on rainfall anomalies or precipitation indices, such as the Standardized Precipitation Index (SPI). However, drought is a multi-dimensional phenomenon that cannot be fully described by precipitation data alone. It is also associated with a lack of soil and vegetation, as well as anomalies in various forms of evapotranspiration, and social-economic aspects that reduce exposure and adaptive capacities (Mishra & Singh, 2011). The inability to consider these interacting dimensions leads to a significant gap in knowledge regarding drought risk management among the NMDs.

This gives rise to the research question of this thesis:

How can we map and forecast drought threats in the Newly Merged Districts of Khyber Pakhtunkhwa and combine meteorological and vegetation variables in this context?

This is an empirical question of the type that involves quantifiable, measurable information and

techniques. The answers to this question can be determined by gathering numerous layers of data, such as rainfall, temperature, vegetation indices, and soil moisture indicators, and then subjecting the data to spatial analysis and predictive modeling to identify patterns of drought threat, exposure, and vulnerability. It is possible to observe, verify, and replicate the results by systematically applying remote sensing data, ground-based measurements, and statistical validation of the machine learning model.

The proposed study will address the current research question by developing a presence map of drought risk and defining an evidence-based model that will aid in early warning, disaster risk reduction, and the identification and distribution of resources in the region, which is home to a historically marginalized and data-oblivious community in Pakistan.

1.3 Problem Statement

Drought is a prevalent and gradually intensifying natural hazard that substantially affects water resources, agricultural productivity, ecosystems, and rural livelihoods, particularly in arid and semi-arid regions (Yussof et al., 2021; Fang & Pan, 2021). Its complex onset and spatial heterogeneity present significant challenges for traditional drought assessment and monitoring, especially in data-scarce and topographically complex landscapes such as the newly merged districts of Khyber Pakhtunkhwa (formerly FATA), Pakistan. This region's climate vulnerability, coupled with heavy reliance on rain-fed agriculture, underscores the need for spatially explicit drought risk characterization to support decision-making at multiple scales.

Existing drought assessment efforts in Pakistan have largely targeted settled districts and tended to focus on single drought indicators or conventional statistical methods that inadequately capture the multidimensional and dynamic nature of drought (Yuan et al., 2021; Wang et al., 2022). While indices such as SPI and SPEI have been widely employed to represent meteorological aspects of drought, and remote sensing indices like NDVI, VCI, and TVDI provide insights into vegetation and soil moisture dynamics, single-indicator approaches frequently fail to integrate these dimensions effectively (Zhang et al., 2022; Dong et al., 2021). As a result, drought risk maps derived from isolated indices often lack the accuracy and spatial detail needed for resilient planning and adaptation strategies.

Advancements in geospatial computing and machine learning have created powerful opportunities for improving drought risk assessment. Geospatial integration of multiple drought indices enables holistic representation of drought dimensions (climate, vegetation, soil, and surface conditions),

while machine learning models such as Random Forest can capture complex, nonlinear interactions among multiple variables, improving prediction and classification performance (Narasimhan et al., 2021; Sharifi et al., 2023). However, the combined application of multi-criteria decision-making techniques like AHP with machine learning models for drought risk mapping remains underutilized in the context of the newly merged districts of Khyber Pakhtunkhwa.

Therefore, there is a pronounced need for an integrated drought risk assessment framework that combines geospatial drought indicators with robust machine learning techniques to develop reliable, spatially explicit risk maps for the newly merged districts of KP. Addressing this gap will advance our understanding of spatiotemporal drought risk patterns, support targeted mitigation and adaptation interventions, and contribute to sustainable planning in vulnerable agro-ecosystems.

1.4 Significance of Research

This thesis aims to address this issue by developing a composite model for risk mapping drought in the NMDs through the integration of machine learning and geospatial analysis. The research seeks to integrate climatic cues (precipitation, temperature, evapotranspiration) and environmental cues (Normalized Difference Vegetation Index [NDVI], soil moisture, and land cover) to produce spatially explicit drought risk maps. Machine learning, i.e., Random Forests, will be used to increase the accuracy of prediction as well as to capture non-linear interactions between variables. This way, the study will yield a localized, multi-dimensional assessment of drought risk within the NMDs. This will provide new insight into the locations of drought risk that are most influenced and most vulnerable, as well as the changes in the spatial risk pattern over time.

Provisional Thesis Statement / Hypothesis

In this study, the hypothesis is that:

The combination of geospatial data and a machine learning model can both map and predict drought risk in the Newly Merged Districts of Khyber Pakhtunkhwa, as these approaches are more effective and capable of quantifying the multi-dimensional character of drought vulnerability than traditional precipitation-based indices.

This assertion is verifiable and falsifiable; its verification will be achieved through the collection of geospatial information, the development of drought risk indices, the implementation of machine learning algorithms, and the testing of the models' results against historical drought observations.

1.5 Aim of the Study:

To develop an integrated geospatial and machine learning framework to assess and map drought risk in the newly merged districts of Khyber Pakhtunkhwa, capturing historical trends, identifying vulnerable areas, and supporting evidence-based drought management and planning.

1.6 Objective of the Research:

To analyze spatiotemporal drought trends (2000–2024) and map drought risk in the newly merged districts of Khyber Pakhtunkhwa using multiple drought indices (SPI, SPEI, aSPI, NDVI, VCI, VHI, TCI, SMI, and LULC) integrated with AHP and Random Forest machine learning techniques.

1.7 Definition of Key Terms in Thesis Statement

To bring out clarity in the concepts and make them empirical, the following are the essential terms that are herein defined as applied in this thesis:

1.7.1. Drought

Drought can be defined as a prolonged period of low precipitation, resulting in a shortage of water for various ecosystems, agriculture, and human consumption (Wilhite, 2005). Operationally, drought in this research will be quantified through drought indices such as the Standardized Precipitation Index (SPI) and the Standardized Precipitation Evapotranspiration Index (SPEI), as well as satellite data, among other indices, including the vegetation stress management index (NDVI anomalies and VCI). Such metrics enable us to observe and measure drought conditions over time and space.

1.7.2. Drought Risk

The term 'drought risk' encompasses the effects of drought that could negatively impact both human and natural systems. It is here understood to be comprised of three measurable components:

- **Hazard:** The likelihood of a climate anomaly (a drought, which is defined by the climatic structures of rainfall, temperature, and evapotranspiration).
- **Exposure:** The degree to which population, livelihood, or options are situated in drought-prone regions (population density, agricultural dependence).
- **Vulnerability:** Socioeconomic sensitivity and adaptive ability of the targeted society (levels of poverty, access to infrastructures, diversification of livelihood) .

This study employs the following mathematical framework: Risk = Hazard × exposure × vulnerability.

1.7.3. Geospatial Analysis

Geospatial analysis is a process involving gathering, processing, and interpreting spatial information to determine geographical trends (Longley, Goodchild, Maguire, & Rhind, 2011). Geospatial analysis, which incorporates both remote sensing (MODIS, Sentinel, Landsat) and GIS methods (interpolation, spatial overlay, hotspot detection) to map out drought hazards, exposure, and vulnerability at the NMDs level, is utilized in this study.

1.7.4. Machine Learning

Machine learning refers to a combination of computational methods that allow conditions to learn through information and predict outcomes without being explicitly programmed (Bishop & Nasrabadi, 2006). Within the context of this thesis, machine learning will employ supervised classification approaches, such as Random Forest, to determine drought-prone areas using various estimates, including climatic and environmental measures. Validation datasets will be used to empirically evaluate model accuracy based on performance measures (Accuracy, F1-Score, and AUC).

1.7.5. Hazard

Hazard refers to an occurrence or phenomenon that can be harmful (UNDRR, p. 39). Hazard, as described in the context of this study, refers to a specific event of climatic drought, as measured by anomalies in precipitation, temperature, and evapotranspiration.

1.7.6. Exposure

Exposure refers to the presence of individuals, properties, or resources at locations where hazards may affect them (Joern Birkmann et al., 2013). In this thesis, exposure will be quantified empirically using spatial data on population density, the area under agricultural activity, and reliance on rainfed cropping systems.

1.7.7. Vulnerability

Vulnerability can be regarded as a state that enhances the concern about the potential for drought damage and diminishes adaptability (Turner et al., 2003). Operationally, the weakness will be

gauged by socioeconomic parameters, including poverty levels, literacy rates, livelihood dependency, and access to facilities. They will be standardized and included in a vulnerability index.

1.7.8. Newly Merged Districts (NMDs)

By combining the seven tribal agencies and six frontier regions of the one-time Federally Administered Tribal Areas (FATA), this area was constituted as Khyber Pakhtunkhwa through the 25th Constitutional Amendment in 2018. They include South Waziristan, North Waziristan, Kurram, Orakzai, Khyber, Mohmand, and Bajaur. This will be an area in which empirical analysis will be conducted.

1.8 Limitations

Although the current research takes a rigorous and multi-dimensional approach to mapping drought risk in the Newly Merged Districts (NMDs), several limitations should also be considered. Such constraints are scope, data, methodology, and situation constraints.

1.8.1. Scope Limitations

In this thesis, the author provides an in-depth discussion of the NMDs of Khyber Pakhtunkhwa. The findings may provide insights into other drought-prone areas of Pakistan, but cannot be applied on a national scale unless the findings are further validated. It is focused on only climatic and environmental factors that are exclusive to NMDs.

1.8.2. Chronological Limitations

Space-time coverage: The temporal limitations are determined by when reliable information can be collected through remote sensing and socioeconomic data. The 2000s have been more recent times when the MODIS dataset was obtained, whereas the CHIRPS rainfall dates date back to the 1980s. Nevertheless, high-resolution images of Sentinel-2 have only been available since 2015, which limits the possibility of long-term trend analysis with high spatial resolution (Hou et al., 2018).

1.8.3. Data Limitations

There is limited ground-based meteorological surveillance in the NMDs due to historical neglect and disruption in the area, so the analysis consists of satellite-derived information.

1.8.4. Methodological Limitations

Although machines are strong in learning, there is a constraint. The quality and representativeness of training data are crucial to their predictive accuracy. Historical droughts should be considered alarming because poorly recorded or classified events may compromise the performance of models (Jorn Birkmann, 2013). In addition, machine learning algorithms can often be black boxes, making explanations of results a bit of a challenge.

1.8.5. Contextual Limitations

Ground-truthing of the NMDs results through field surveys may be impossible due to security concerns and limited access to specific regions. This limits validation to secondary and indirect evidence.

The methodological combination of these two areas, regarding geospatial features and machine learning algorithms, still has these shortcomings, but provides a solid framework for evaluating localized drought risk sources.

1.9 Organization of the study

This thesis is structured into five extensive chapters, each of which is based on the underlying premise provided in the Introduction. The chapters are arranged in a way that facilitates the identification of problems and the logical application of methods, findings, and policy implications.

Chapter 2: Literature Review

In this chapter, the author reviews the existing literature on drought, analyzes drought risk, and discusses geospatial modeling. It begins by examining international methods of drought monitoring, including drought indices, such as SPI, SPEI, and VCI, as well as remote sensing techniques. It then examines case studies in Pakistan and South Asia, with a specific focus on the effects of drought in arid and semi-arid regions. The gaps in research about the Newly Merged Districts (NMDs) have received particular attention, which is why this study is necessary.

Chapter 3: Experimental Methodology

In this chapter, a detailed exposition will be provided regarding the methodology that has been formulated. It provides details on the source of data, preprocessing, and consideration of climatic and environmental indicators. It also describes the chosen machine learning tools to be used in the study (Random Forest), how these tools will be trained and validated, and which statistics will be

employed. Poor methods and restrictions on the way of ethical considerations are also covered.

Chapter 4: Results and Analysis

The empirical findings of the study are presented in this chapter. It begins with the research on spatial drought hazard mapping based on climatic and environmental indicators. Such layers have been incorporated in composite drought risk maps. Comparisons of the performance of various machine learning models are presented, with results displayed in statistical tables, ROC curves, and spatial outputs. The accuracy of the models is verified using historical drought events in the NMDs.

Chapter 5: Conclusion and Policy Recommendations

The research contribution is summarized in the final chapter of the study, which aims to establish whether the research question and hypothesis have been adequately addressed. It provides recommendations on how the results will be utilized in managing disaster risks, development planning, and climate change adaptation in Pakistan. The chapter also summarizes by outlining opportunities to shed light on future studies, especially on how to combine community-based field information, machine learning, and geospatial methods.

2 Literature Review

2.1 Background

Drought has traditionally been recognized as one of the most severe natural hazards, with the highest number of affected people globally compared to any other climatic disaster. In comparison to floods or storms, which appear suddenly, drought is a cumulative risk that takes months or sometimes years to affect any ecosystem, agriculture, or society (Van Loon, 2015). Internationally, droughts are estimated to cause over 50 % of climate-related disasters, highlighting the socioeconomic implications associated with such events (Spinoni et al., 2019).

Scholarly works within the field of drought have grown substantially over the past decades based on three primary threads: (1) meteorological and hydrological surveillance, measurement, and health indices, (2) remote imagery and geospatial analysis, and (3) unified risk testing, including socio-economic susceptibility. Corino's early research focused on the indices of precipitation, including the Palmer Drought Severity Index (PDSI) and the Standardized Precipitation Index (SPI), which serve as functional indices for detecting anomalies in rainfall and moisture (Palmer, 1965). Such indices, however, often cannot accurately measure drought, particularly in areas with sparse meteorological coverage or high spatial diversity.

The health of vegetation and land surfaces has been extensively used in determining agricultural drought with the advancement of satellite remote sensing since the 1980s. Spatially Continuous Measurements - When reflectance spectroscopy is utilized, the assessment of vegetation stress and photosynthesis can be provided by indices such as the Normalized Difference Vegetation Index (NDVI) and the Vegetation Condition Index (VCI) (Jw, 1973).

In recent times, combined techniques that integrate climate, hydrology, and vegetation data have been developed in Europe, the United States, and Asia. This can be exemplified by the U.S. Drought Monitor and the European Drought Observatory, which both deliver a multidimensional score of drought risk information to aid in decision-making (Svoboda et al., 2002). Although progress has been made under such challenges, most of the developing nations (especially in South Asia) do not have local but multidimensional drought monitoring systems. An example of this is the past situation with Pakistan, which has been relying more on precipitation-based monitoring and incorporating less satellite data and socio-economic indicators (Rahman, Khalid, Arshad, Moazzam, & Kwon, 2025). This gap highlights the need to develop innovative, geospatial, and

machine learning-based strategies for marginalized and drought-prone regions, including the Newly Merged Districts (NMDs) in Khyber Pakhtunkhwa.

Average Annual Natural Hazard Occurrence for 1980–2020

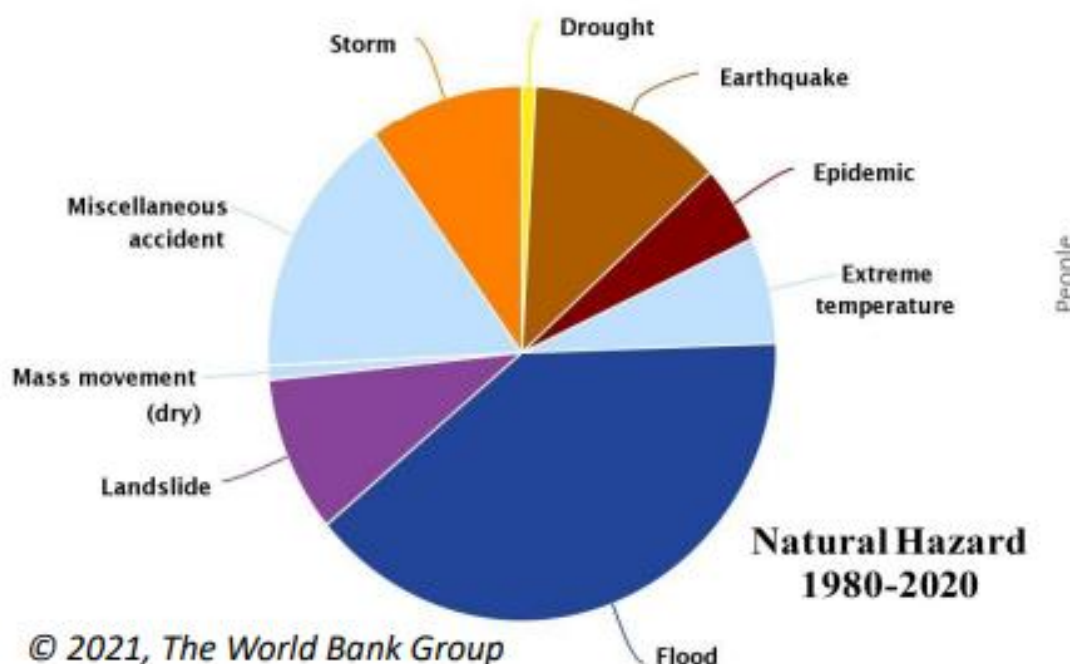


Figure 2- 1: Average Annual Natural Hazard Occurrence (1980-2020)

2.2 International perspective on Drought monitoring

Over the past five decades, drought monitoring has undergone a significant transformation, transitioning to multidimensional and integrated frameworks that encompass multiple variables. The growth of Earth monitoring technologies and computational systems has significantly enhanced the ability to identify, forecast, and control drought risk.

2.2.1. Meteorological Indices

Meteorological indices of rainfall and temperature are recognized as one of the oldest and most extensively used ways of describing drought. One of the earliest efforts to quantify the severity of droughts through water balance models was the introduction of the Palmer Drought Severity Index (PDSI), developed in the 1960s (Palmer, 1965). Subsequently, the Standardized Precipitation Index (SPI) was designed to provide a more flexible and statistically standardized measure of precipitation anomalies at various time scales (McKee, Doesken, & Kleist, 1993). This has been followed by the introduction of the Reconnaissance Drought Index (RDI) and the Standardized

Precipitation Evapotranspiration Index (SPEI), which consider evapotranspiration to capture the state of agricultural drought better (Vicente-Serrano, Beguería, & López-Moreno, 2010). These indices are not utilized in areas where meteorological stations are Republican, and they are not usually representative of non-linear interactions between precipitation, soil moisture, and vegetation stress (Mishra & Singh, 2011). This renders them less useful in mountainous and semi-arid regions, including those that occur in northern Pakistan.

2.2.2 Remote Sensing Approaches

In recent years, satellite-based surveillance has been seen as a potent instrument for identifying the effects of drought on vegetation and land cover that have manifested since the 1980s. Spectral reflectance products, such as the Normalized Difference Vegetation Index (NDVI) and Vegetation Health Index (VHI), have also been widely used to detect vegetation stress (Kogan, 1995). Further developed indices, such as the Temperature Vegetation Dryness Index (TVDI) and the Normalized Difference Water Index (NDWI), are combinations of land surface temperature and water content, which capture drought-related stress to a greater extent (Wang, Price, & Rich, 2001).

Most importantly, remote sensing provides spatially continuous data, which is particularly beneficial in areas with limited data. Examples include (Anderson, Norman, Mecikalski, Otkin, & Kustas, 2007), who used TSMs of evapotranspiration models based on MODIS to observe agricultural drought in the United States, and (Zargar, Sadiq, Naser, & Khan, 2011) who reviewed global remote sensing indices to support the case for including them in early warning systems.

Over the past few years, various countries and regions have developed integrated drought monitoring systems, which entail a combination of two indicators. One of the most popular drought monitors in the U.S. is the U.S. Drought Monitor, which provides weekly updates on drought designations derived from indices of precipitation, soil moisture, vegetation indices, and human opinion (Svoboda et al., 2002). Similarly, the European Drought Observatory is an organisation that combines meteorological, soil moisture, and hydrological information to ensure a comprehensive view of drought effects in Europe (Sepulcre-Canto et al., 2012).

Such systems signify a paradigm shift in the fight against science, where multi-source and multidimensional assessments are required. Nevertheless, developing countries have found it challenging to adopt such systems due to their limited technical and institutional capabilities.

2.3 Regional Context: South Asia

South Asia can be regarded as one of the most drought-prone areas in the world, where hundreds of millions of people rely on their climate-dependent lives, primarily through climate-dependent agriculture and livestock farming. The climate associated with the monsoons, the pressures exerted by the population, and the socio-economic vulnerability predispose the region to the risks of disasters caused by drought (Balaganesh et al., 2020).

2.3.1 Drought in India

India has established a relatively sophisticated system for monitoring droughts compared to other countries in the South Asian region. The National Agricultural Drought Assessment and Monitoring System (NADAMS) incorporates satellite-based knowledge of NDVI and rainfall in the delivery of real-time products that assess the effects of drought on crops (Sai et al., 2016). The latest study by (Singh et al., 2021) incorporates soil moisture, land surface temperature, and socio-economic data using machine learning, resulting in improved predictive capability.

2.3.2 Drought in Bangladesh

Bangladesh's droughts mainly affect the northwestern and central regions of the country. The drought systems rely on McCormick precipitation and rainfall abnormalities (Shahid & Behrawan, 2008). The Bangladesh Drought Early Warning System (DEWS) has attempted to utilize remote sensing data; however, the restricted technical infrastructure makes high-resolution monitoring impossible to achieve (Ahammed et al., 2020).

2.3.3 Drought in Nepal and Sri Lanka

Research on drought in Nepal has been primarily conducted by the International Center for Integrated Mountain Development (ICIMOD), which combines climatic and livelihood-related information to provide a comprehensive hazard assessment (Shrestha et al., 2017). Sri Lanka has adopted the use of indices and temperature variations based on NDVI to monitor drought in paddy fields, but socio-economic susceptibility has not been extensively examined (Sangelantoni, Ferretti, & Redaelli, 2019).

Table 2- 1: Selected Drought Monitoring Systems in South Asia

Country	System/Approach	Indicators Used	Strengths	Limitations
India	NADAMS (National Agricultural Drought Assessment and Monitoring System)	NDVI, rainfall, soil moisture	Real-time crop monitoring; large-scale	Limited socio-economic integration
Bangladesh	DEWS (Drought Early Warning System)	SPI, rainfall anomalies	Agriculture-focused; low-cost	Weak integration with remote sensing
Nepal	ICIMOD Drought Hazard Assessment	Climate + livelihood data	Combines hazard and vulnerability	Coarse spatial resolution
Sri Lanka	Remote sensing-based drought index	NDVI, temperature anomalies	High spatial resolution	No socio-economic indicators

2.4 Drought Studies in Pakistan

Pakistan is also prone to drought due to its arid and semi-arid climate, as well as its high dependency on monsoonal rain and variability in water resources, which are increasing. Almost 60 % of the country's total area is arid or semi-arid, and 80 % of agribusinesses rely on canal and rainfed irrigation, which is highly susceptible to climate variations (Salik et al., 2015). Although drought regulations in Pakistan have long been socially disconnected and distorted in terms of geography, in India, the Newly Merged Districts (NMDs) have hardly been addressed in scholarly literature, despite being historically institutions afflicted with drought, as severe as they are.

2.4.1 National-level Assessment of Drought

Among the first organized research on drought in Pakistan was done following the 1998-2002 incident, which was termed as the worst drought that has ever happened in the country. This drought also devastated Balochistan and Sindh, causing crops to fail, animals to die, and people to migrate from rural areas (Sheikh et al., 2004). Researchers pointed out that drought management policies were reactive and not preventive, particularly in terms of building resilience and justice rather than relief.

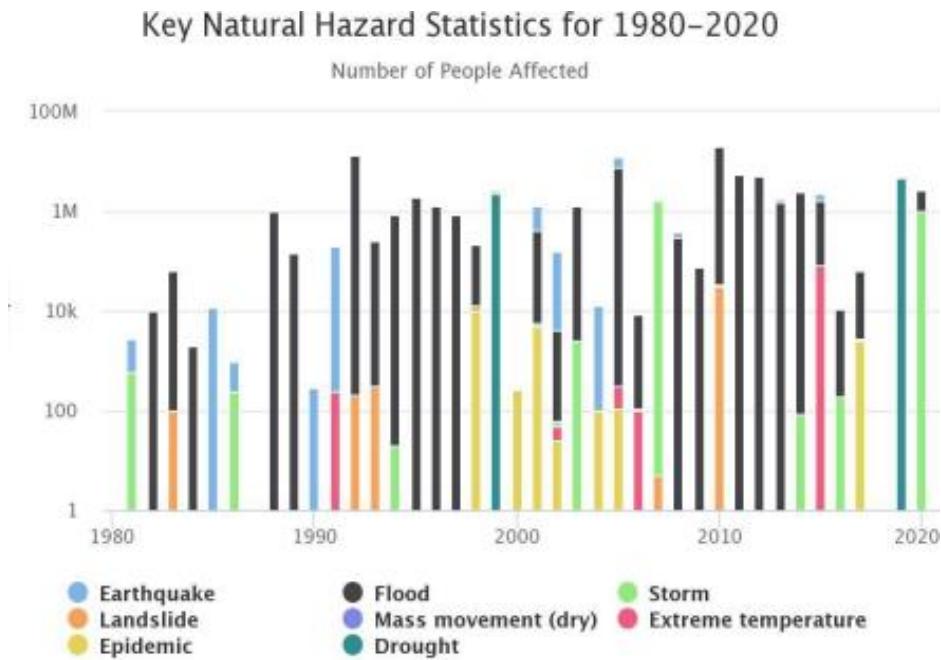


Figure 2- 2: Key natural hazard statistics for 1980-2020, source: Eco Regional Center for Risk Management of Natural Disasters (ECO-RCRM), <https://eco-rcrm.ir/natural-disasters/>

Subsequent research has also extensively utilized meteorological indices, such as the Standardized Precipitation Index (SPI) and the Palmer Drought Severity Index (PDSI), to categorize the extreme levels of droughts. For example, Rahman and Dawood (2017) utilized SPI on rainfall measurements in Punjab to demonstrate that substantial and significant spatial variations in the frequency and intensity of droughts exist across various agro-climatic zones. In the same manner, Adnan et al. (2017) employed PDSI to evaluate the trends of droughts in Pakistan, but found that the frequency and severity of droughts have been on the rise in recent decades.

2.4.2 Remote Sensing for drought monitoring

Pakistan's drought research has increased due to the introduction of satellite imagery. The Normalized Difference Vegetation Index (NDVI) and Vegetation Condition Index (VCI) are now commonly employed to track agricultural drought using vegetation indices, particularly in their application. The article by Shahid et al. (2018) investigated the impact of drought on wheat production in four fields of Sindh, utilizing NDVI anomalies, which revealed a significant correlation between vegetation stress and yield reduction. Similarly, Khan et al. (2019) employed remote sensing to map drought in Balochistan, utilizing MODIS-derived vegetation indices. They found that remote sensing provided superior spatial and temporal resolution compared to in situ rainfall data.

Later, Faisal et al. (2020) developed a composite drought index by combining information on rainfall, NDVI, and land surface temperature. Their findings revealed that satellite-based indices provided early warnings of drought before its effects were felt at the ground level. Such results underscore the opportunities behind remote sensing and predictive drought in data-sparse areas.

2.4.3 Studies on Hydrological and Agricultural Drought

Although meteorological drought is the one that receives the most focus, hydrological and agricultural droughts are also growing in investigation. A similar case is that of Zaman et al. (2021), who studied the depletion of groundwater in Punjab and Sindh, stating that low recharge in drought years is linked to the non-sustainability of tube-well extraction. Ali et al. (2019) have examined rainfed agriculture in Khyber Pakhtunkhwa and observed that frequent droughts have a substantial negative impact on the production of maize and wheat, forcing farmers to cut their debt specifically. Moreover, the issue of agricultural drought can be regarded as the most severe, as almost 45 % of the Pakistani population is engaged in agriculture (Pakistan Bureau of Statistics, 2020). Crop failures due to drought not only pose a threat to food security but also impact rural livelihoods, migration, and poverty dynamics in a cascading effect.

2.4.5 Synthesis of Drought Research in Pakistan

Table 2.2 summarizes selected drought studies in Pakistan, their methods, and key findings.

Table 2- 2: Selected Drought Studies in Pakistan

Study	Region	Methodology/Index Used	Key Findings
Sheikh et al. (2004)	Sindh & Balochistan	Post-event hazard analysis	1998–2002 drought caused severe rural migration and livestock loss
Rahman & Dawood (2017)	Punjab	SPI analysis	Strong spatial variability in drought severity
Adnan et al. (2017)	National	PDSI trends	Increasing drought frequency over decades
Shahid et al. (2018)	Sindh	NDVI anomalies	Drought stress strongly correlated with wheat yield loss
Khan et al. (2019)	Balochistan	MODIS vegetation indices	Remote sensing more effective than rainfall-only indices
Faisal et al. (2020)	Pakistan-wide	Composite drought index	Early detection possible with integrated datasets
Zaman et al. (2021)	Punjab & Sindh	Groundwater data + drought	Groundwater depletion intensified by drought years
Ali et al. (2019)	KP rainfed areas	Household survey + yield analysis	Drought reduces maize and wheat yields significantly

2.5 Drought Research: applications of machine learning

Due to the increasing availability of high-resolution satellite images and historical climate data, machine learning (ML) approaches are being integrated into drought predictions. There is a growing use of ML techniques to address limitations in traditional statistical and index-based techniques, which, in most cases, cannot model the non-linear and multidimensional character of drought.

The resulting benefit of machine learning within Drought Studies was as follows:

Machine Learning models can be used to analyze extensive datasets that contain multiple variables. They can identify non-linear aspects between climate variables (rainfall, temperature), remote sensing indices (NDVI, LST), that regression-based models cannot identify. This results in their high applicability to predict drought in poorly homogenized environments (Abhishek, 2023)

In addition, the ML models are capable of addressing missing data, can adjust to different temporal and spatial scales, and can also be trained to enhance predictive performance.

Random Forest and Ensemble Models: This method is appropriate when data are not structured systematically but occur independently (appearing randomly) along with their suitable distribution between variables (Andia et al., 2015).<|human|>2.5.2 Random Forest and Ensemble Models: This approach is appropriate when data are not organized and appear sporadically with their natural distribution across variables (Andia et al., 2015).

Random Forest (RF) is one of the most popular ML algorithms to predict drought. Wu et al. (2020) experimented with RF in combining soil moisture, NDVI, and temperature in China, which was more accurate compared to conventional drought indices. Similarly, Kotak et al. (2021) employed Gradient Boosting Machines (GBM) in India to model agricultural drought and reported an increased accuracy in ensemble methods compared to single-algorithm approaches.

The reason an ensemble model such as RF is used is that it minimizes overfitting and ranks variables in terms of their importance, identifying those that contribute the most to forecasting drought. This attribute is essential to policymakers who must receive predictions as well as cope with the causes of drought

2.5.2 Deep Learning Approaches

Recent innovations in deep learning have enhanced the prediction of droughts. Long Short-Term Memory (LSTM) networks, a type of recurrent neural network (RNN), have achieved great success in modeling climate and hydrological data, particularly in terms of capturing temporal dependencies. In (Fang et al., 2024) LSTM was used to predict drought in the United States, and despite reporting greater results than other traditional ML models, it performs better than conventional models.

The patterns of spatial drought are also being learned with a Convolutional Neural Network (CNN) on gridded satellite data. Indicatively, (Alikhani-Koupaei, Aghdam, & Faghieh, 2020) found that neural networks based on CNN and relying on imagery provided by MODIS were occasionally valuable for revealing drought hotspots in Chinese agricultural areas. Such techniques are effective in mapping droughts at various scales, yet they require intensive computer equipment.

2.5.3 Hybrid Models and Data Fusion

A new major direction in drought studies is the application of hybrid models, including MLs and traditional indices or statistical models. For example, (Tian & Fan, 2022) have combined SPI with machine learning tools to enhance drought early warning activities in East Asia. Equally, there have been signs of success in hybrid schemes that combine both satellite-derived vegetation indices and climate data (Hao, AghaKouchak, & Phillips, 2013).

2.5.4 Applications in South Asia

In South Asia, the application of machine learning in drought modeling is not a new concept, but its use is increasing. After choosing the NDVI and SPI data, Khan et al. (2020) implemented RF and SVM in Punjab, Pakistan, and realized that RF was more accurate. (Singh et al., 2021) mapped drought risk in India with the help of Gradient Boosting and LSTM models and achieved a more precise result in complications of spatial variability, compared to traditional statistical modeling.

2.6 Vulnerability and Risk Frameworks

To comprehend drought as a hazard, Climatic or even hydrological analysis is not sufficient. It is a study of the relationship between hazard, exposure, and vulnerability, which combine to form risk. International systems and local research indicate that vulnerability (social, economic, and institutional, among others) is a highly important factor in determining the outcome of drought (Birkmann et al., 2013).

Vulnerability: An indicator of the extent of systems or populations prone to harm and incapable of managing the effects of drought. (Turner et al., 2003) also point to a sample framework where vulnerability is a product of exposure, sensitivity, and adaptive capacity. Likewise, the IPCC (2014) defines vulnerability as the ability to respond or adapt to the harmful effects of climate change, coupled with susceptibility to these effects.

Practically, in terms of vulnerability, indicators uncovered can include poverty, livelihood dependency, access to infrastructure, or the quality of governance. As an example, (Cutter, Boruff, & Shirley, 2012) developed the Social Vulnerability Index (SoVI) for the United States, which has since been modified to apply internationally and determine the effect of social structures on disaster impact.

2.6.1 Exposure and Sensitivity

Exposure can be understood as the presence of humans, assets, and ecosystems in hazard-susceptible regions, whereas sensitivity refers to the extent to which they are (Adger, 2006). An exposure could be thought of in drought settings as populations relying on rainfed agriculture. In contrast, sensitivity refers to the degree to which those populations rely on a single crop or a sensitive biological interaction.

Indicatively, evidence summarized by (Lobell et al., 2008) shows that agriculturally dependent regions, especially those in Sub-Saharan Africa and South Asia, tend to be vulnerable to climate variability. Equally, (Pandey et al., 2015) found that Indian farmers who relied on irrigation through groundwater were more susceptible to drought than those with other diversified water sources.

2.6.2 Adaptive Capacity

The adaptation to hazards, mitigation of potential harm, and recovery after impacts is called adaptive capacity; it is the capability of communities or systems to adapt to hazards. It comprises technology and infrastructure, financial facilities, and institutional support (Brooks, Adger, & Kelly, 2005). The adaptive capacity in drought-prone areas is substantially impaired by low literacy rates, poor governance mechanisms, and inadequate disaster management institutions in Pakistan (Mustafa et al., 2013).

2.6.3 Pakistan Risk/Vulnerability

Pakistan has experienced little development in vulnerability-based tests of drought risk. In analyzing the problem of drought in the Balochistan region, (Mustafa, 2007) pointed out that institutional and political negligence had just as much effect as climate variability in determining vulnerability. The more recent reports by Salik et al. (2017) noted that rural households tend to resort to short-term coping strategies, which include trading animals or sending migrants, yet both proved to be temporary.

Household surveys conducted in Khyber Pakhtunkhwa revealed that rainfed farmers were low-income farmers with low levels of education, making them highly susceptible to drought (Ali et al. 2019). Nevertheless, holistic exposure/sensitivity/adaptive capacity combinations are lacking in Pakistan, and this is particularly true in the Newly Merged Districts (NMDs) of Pakistan.

2.7 Identification of research gaps

A literature review has shown that research on drought has developed significantly on global, regional, and national levels. Despite some significant progress in monitoring and modeling drought, several knowledge gaps persist, particularly in Newly Merged Districts (NMDs) in Pakistan.

2.7.1 Global Advances

Drought monitoring has evolved into a multi-indicator, integrated system worldwide, as compared to the traditional meteorological indices used in isolation. Machine learning has helped to improve predictive accuracy, whereas spatial coverage is developed by remote sensing (Fang et al., 2024; Wang et al., 2001). The relevance of socio-economic factors has also been emphasized in vulnerability frameworks, with drought being viewed more as a social and political event rather than a purely climatic one (Cutter et al., 2012; Turner et al., 2003). Nonetheless, the majority of global models have been designed in data-intensive regions, such as data-intensive North America, Europe, and East Asia. They are not directly applicable to data-weak, conflict-affected areas.

2.7.2 Regional Developments South Asia

Drought monitoring systems have been advanced first by countries of South Asia, most notably India, which combine satellite and meteorological records with the NADAMS (Rao et al., 2013). In Bangladesh and Nepal, specific steps have been undertaken, but they remain challenged by data availability, spatial resolution, and the institutional sphere (Alamgir et al., 2018; Paudel et al., 2020).

Although there is development, South Asian countries primarily focus on agricultural surveillance with little attention to socio-economic vulnerability or adaptive capacity. Moreover, the adoption of machine learning is relatively low compared to global trends.

2.7.3 Pakistan-Specific Insights

Drought research has been overly concentrated in Sindh, Balochistan, and Punjab in Pakistan. Scientists have used meteorological indices as the primary source of choice (Adnan, Ullah, & Physics, 2022; Rahman et al., 2025) with remote sensing data gradually augmenting ground-based information (Ullah, Khan, Ullah, Khan, & Lee, 2023). Integrated NDVI, rainfall, and land surface temperature indices have proven to be promising sources of early warnings (Faisal et al., 2020).

Nevertheless, the following limitations exist:

Geographical bias in Research does not focus on the NMDs, and the severe livelihoods based on rainfed agriculture are exacerbated by the deficiency of good governance, which further enhances vulnerability (Mustafa, 2007; Ali et al., 2019). The vast majority of studies employ linear indices, with relatively little use of machine learning models that can account for the complex, non-linear interactions between climatic, environmental, and socio-economic drivers. Not many studies systematically combine poverty, livelihood dependency, and infrastructure accessibility into drought risk evaluations (Salik & Environment).

2.7.4 Identified Research Gaps

In accordance with the synthesis, the following gaps in research can be named:

The NMDs are very risky, but they are virtually absent in comparative, peer-reviewed.

Machine learning has been employed in many countries, but its application in Pakistan is limited, and its use in the NMDs has not been implemented so far.

The current literature rarely translates its findings into practical mechanisms for disaster risk mitigation and climate adaptation at the local level.

2.7.5 Research Justification

These gaps are filled by this thesis, which proposes:

- Specifically, the Newly Merged Districts are an insider and understudied area.

- Creation of a multi-dimensional drought risk framework by incorporating the environmental and climatic.
- The enhancement of predictive accuracy by relying on machine learning tools (Random Forest) in comparison with traditional indices.
- Production of policy-relevant outputs, including the spatial risk maps that can guide the drought preparedness and adaptation planning in KP.

2.8 Conclusion

As a result of the literature review, it is established that even though the world is making a lot of progress in matters of drought monitoring and modeling, Pakistan is still lacking globally in terms of its representation in the body of the empirical research, and that Pakistan is most notably lacking in the case of NMDs. Such shameful levels of climatic stress and ecological vulnerability in these districts, highlight the urgency of localized and, more importantly, data-driven interventions. This thesis will bridge a significant gap in the scholarly literature and policy implementation by combining the concepts of geospatial analysis and machine learning in an approach to vulnerability assessment and measurement.

3 Experimental Methodology

3.1 Study Area

This research focuses on the Newly Merged Districts (NMDs) of Khyber Pakhtunkhwa (KP), Pakistan, which were previously known as the Federally Administered Tribal Areas (FATA). These areas joined the KP via the 25th Constitutional Amendment, which constitutes one of the most significant changes in any governance mechanism in Pakistan's history (Yousaf, 2019). The NMDs comprise a combination of seven tribal agencies, each bearing the names: Bajaur, Mohmand, Khyber, Orakzai, Kurram, North Waziristan and South Waziristan.

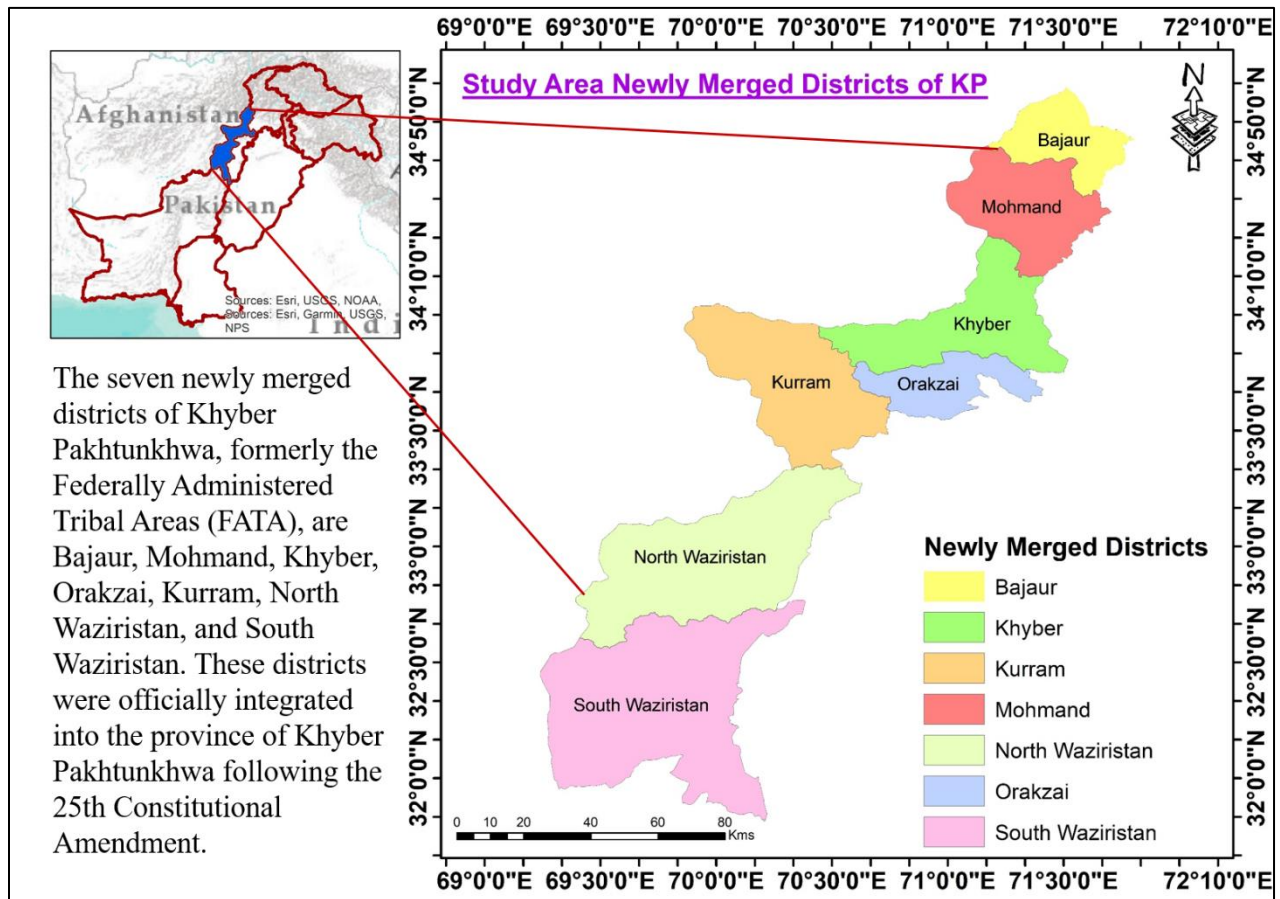


Figure 3- 1: Study area i.e., newly merged districts of KP

The NMDs are situated at 32 degrees to 36 degrees north longitude and 69 degrees to 72 degrees east longitude, spanning an area of approximately 27,000 km². It has a relatively mountainous terrain, which is crossed by small valleys and riverbeds. The climate is semi-arid to arid, with hot summers, cold winters, and very sporadic rainfall, most of which is received during the summer monsoon.

High reliance on rainfed agriculture and livestock, combined with low literacy levels and inadequate infrastructure, has characterized the socio-economic profile of the NMDs. Constant struggle, disrupted administration, and discrimination have heightened the vulnerability of local populations to environmental risks, particularly drought. All these structural conditions render the NMDs a crucial case study for mapping and assessing drought risks.

3.2 Data Sources and Indices

The methodology combines various drought-related indices and indicators obtained by utilizing Google Earth Engine (GEE) and remote sensing datasets. The evaluation is conducted over the four mark-off years of 2000, 2008, 2016, and 2024. To account for temporal dynamics, a mean of eight years, centered on each year at the benchmark, was calculated.

3.2.1 Drought Indices

The following indices were employed:

- **Standardized Precipitation Index (SPI):** Measures precipitation anomalies at different time scales.
- **Standardized Precipitation Evapotranspiration Index (SPEI):** Incorporates both precipitation and evapotranspiration to account for water balance.
- **Normalized Difference Vegetation Index (NDVI):** Derived from MODIS and Landsat, indicating vegetation greenness and productivity.
- **Vegetation Condition Index (VCI):** Normalized NDVI to measure vegetation stress relative to historical trends.
- **Vegetation Health Index (VHI):** Combines VCI with Temperature Condition Index (TCI) for integrated vegetation health.
- **Temperature Condition Index (TCI):** Derived from land surface temperature anomalies.
- **Soil Moisture Index (SMI):** Indicates relative soil water content.
- **Temperature Vegetation Dryness Index (TVDI):** Combines NDVI and land surface temperature to assess drought severity.
- **Agricultural Standardized Precipitation Index (aSPI):** Specific variant of SPI tailored for

agricultural drought.

- **Land Use and Land Cover (LULC):** Derived from Landsat imagery, classified into agricultural land, rangeland, forest, urban, and barren land.

Table 3- 1: Datasets Used in the Study

Indicator / Index	Dataset Name / Source	Spatial Resolution	Temporal Coverage	Purpose in Study
SPI (Standardized Precipitation Index)	CHIRPS (Climate Hazards Group InfraRed Precipitation with Stations)	~0.05° (~5 km)	1981–present	Precipitation anomaly detection
SPEI (Standardized Precipitation Evapotranspiration Index)	ERA5-Land Reanalysis (ECMWF)	~0.1° (~9 km)	1981–present	Water balance (precipitation + PET)
NDVI (Normalized Difference Vegetation Index)	MODIS MOD13Q1 (Terra/Aqua)	250 m, 16-day composite	2000–present	Vegetation greenness monitoring
VCI (Vegetation Condition Index)	Derived from MODIS NDVI	250 m	2000–present	Relative vegetation stress
VHI (Vegetation Health Index)	Derived from MODIS NDVI & LST	250 m (NDVI), 1 km (LST)	2000–present	Integrated vegetation health (NDVI + temperature)
TCI (Temperature Condition Index)	MODIS LST (MOD11A2)	1 km, 8-day composite	2000–present	Thermal stress on vegetation
SMI (Soil Moisture Index)	ESA CCI Soil Moisture v06.1	0.25° (~25 km)	1979–present	Surface soil moisture variability
TVDI (Temperature Vegetation Dryness Index)	Derived from MODIS NDVI & LST	250 m–1 km	2000–present	Land surface dryness indicator
aSPI (Agricultural Standardized Precipitation Index)	CHIRPS + Crop calendar datasets	~5 km	1981–present	Agricultural drought characterization
LULC (Land Use and Land Cover)	Landsat 5 TM, Landsat 7 ETM+, Landsat 8 OLI, Landsat 9 OLI	30 m	2000, 2008, 2016, 2024	Land cover classification via Random Forest
Reference Data for LULC Classification	Google Earth high-resolution imagery, field data (where available)	Varies (sub-meter in some locations)	Used for 2000–2024 validation	Training and validation of supervised classification

3.3 Land Use and Land Cover (LULC) Mapping

LULC maps were generated for each benchmark year using Landsat imagery (30 m resolution) processed in Google Earth Engine. A supervised classification approach was employed, where training samples for each land cover class were identified through high-resolution imagery and field reference data (where available).

The classification was implemented using the Random Forest (RF) algorithm, a machine learning ensemble method widely recognized for its robustness in remote sensing applications (Breiman, 2001). RF constructs multiple decision trees and combines their predictions, reducing overfitting and increasing classification accuracy. Post-classification accuracy assessments were conducted using confusion matrices and Kappa coefficients, with validation samples independent of training data.

3.4 Multi-Criteria Decision Analysis (MCDA) Framework

To integrate multiple drought indicators into a composite drought risk assessment, the **Analytic Hierarchy Process (AHP)** was applied.

3.4.1 Saaty's Scale for Ranking

The relative importance of each criterion (SPI, SPEI, NDVI, VCI, VHI, TCI, SMI, TVDI, aSPI, LULC) was ranked using **Saaty's 1–9 scale** (Saaty, 1980). Pairwise comparisons were made between all criteria to derive relative weights.

3.4.2 Consistency Check

The Consistency Ratio (CR) was calculated to ensure that judgments in the pairwise comparison matrix were logically coherent. A $CR < 0.1$ was considered acceptable.

3.4.3 Weighted Overlay Analysis

The normalized indices were integrated through a **weighted linear combination**, where weights were derived from AHP. This produced composite **drought hazard maps** for each benchmark year.

Table 3- 2: Saaty's Fundamental Scale for Pairwise Comparisons

Intensity of Importance	Definition	Explanation
1	Equal importance	Two criteria contribute equally to the
3	Moderate importance	Experience/judgment slightly favors one criterion
5	Strong importance	One criterion strongly favored
7	Very strong importance	One criterion demonstrated dominance in practice
9	Extreme importance	Evidence favors one criterion absolutely
2,4,6,8	Intermediate values	Compromise between adjacent judgments

(Source: Saaty, 1980)

Pairwise Comparison Logic

- Climatic indices (SPI, SPEI, aSPI) were judged most critical as they represent the hazard (rainfall + evapotranspiration).
- Vegetation indices (NDVI, VCI, VHI, TCI, TVDI) were given secondary importance as they reflect drought impacts on crops and ecosystems.
- Soil moisture (SMI) was ranked highly because of its role in agricultural productivity.
- LULC was assigned moderate importance, as land use determines exposure but does not directly capture drought onset.

Computed Weights and Influence %ages

After applying AHP in **Expert Choice / Super Decisions software (or Excel AHP matrix)**, the following normalized weights and influence %ages were obtained (CR = 0.08, acceptable since <0.1):

Table 3- 3: Weights and Influence %ages of Drought Indicators

Criterion	Weight Normalized)	(AHP	Influence %
SPI (Standardized Precipitation Index)	0.18		18%
SPEI (Precipitation + Evapotranspiration)	0.15		15%
aSPI (Agricultural SPI)	0.10		10%
NDVI (Vegetation Greenness)	0.09		9%
VCI (Vegetation Condition Index)	0.07		7%
VHI (Vegetation Health Index)	0.06		6%
TCI (Temperature Condition Index)	0.07		7%
TVDI (Temperature Vegetation Dryness Index)	0.08		8%
SMI (Soil Moisture Index)	0.12		12%
LULC (Land Use & Land Cover)	0.08		8%
Total	1.00		100%

Interpretation

- **Climatic hazard indicators (SPI, SPEI, aSPI)** jointly account for **43% of total influence** → confirming that precipitation variability dominates drought risk in the study area.
- **Vegetation indices (NDVI, VCI, VHI, TCI, TVDI)** contribute **37% combined**, reflecting the importance of drought impact on crops and rangelands.

- **Soil Moisture (SMI)** stands out with **12% weight**, highlighting its agricultural relevance.
- **LULC** contributes **8%**, influencing exposure and spatial vulnerability but less directly tied to drought onset.

3.5 Machine Learning

Machine learning methods were used in this experiment to determine the severity of drought in the recently unified districts of FATA. We utilized ERA-5 reanalysis data from the European Centre for Medium-range Weather Forecasts (ECMWF), which provided high-resolution yearly climatic variables, including precipitation, temperature, soil moisture, evapotranspiration, and relative humidity. These parameters were chosen because they are the major hydro-meteorological drivers of the drought conditions.

Data preprocessing and the identification of representative points were conducted throughout the study area to generate spatially distributed training data. The indices used to identify drought categories included the Standardized Precipitation Index (SPI) and the Standardized Precipitation Evapotranspiration Index (SPEI), which served as the dependent variable (target class) of the machine learning model.

The Random Forest (RF) classifier was used due to its strength, which includes the ability to deal with nonlinear relationships, extensive large-dimensional data fitted with the obtained ERA-5 parameters as the independent variable, and the drought severity classes (Extreme, Severe, Moderate, Mild, No Drought) formed the dependent variable. The algorithm assembled a set of decision trees, in which each decision tree cast a vote on the type of drought, and the outcome of the votes was the final decision.

Training, validation, and prediction of the model were conducted in a Python programming environment using the libraries scikit-learn, NumPy, Pandas, and Matplotlib. The drought categories were spatially visualized in ArcGIS and QGIS, allowing them to create thematic maps for the years 2000, 2008, 2016, and 2024.

The multi-source, multi-variable climate variables, the use of more sophisticated machine learning classification, and the production of spatially explicit drought risk maps that focus on vulnerable regions in the FATA region were all made possible under this methodological framework.

3.6 Workflow

The workflow in figure 3.2 illustrates a systematic approach to creating map-based composite drought risks, utilizing remote sensing, geospatial analysis, and decision-support methods. Within the process, data acquisition begins with the use of Google Earth Engine, which combines multi-source datasets, including Sentinel-1/2, Landsat, MODIS, ERA5 reanalysis, GLDAS, CHIRPS rainfall, and ESA CCI soil moisture. Such data sets provide a range of inputs, including vegetation indices, precipitation, soil moisture, and climatic reanalysis variables, all of which are crucial for drought evaluation. The second step is Preprocessing and Index Calculation, which standardizes the raw datasets and converts them into biophysical indices. The most important indices are the meteorological (SPI, SPEI), vegetation-based (NDVI, VCI, VHI, TCI, TVDI), and soil moisture indicators (aSPI, SMI). These indices encompass various drought dimensions, including meteorological, agricultural, and hydrological, and thus provide a holistic measure. The application of this principle can enhance the accuracy of these indicators at the field or regional level by leveraging precision agriculture principles. Landsat imagery and Random Forest supervised classification are then used during the workflow to apply Land Use and Land Cover (LULC) Classification. This process distinguishes among the types of land cover (e.g., croplands, forests, rangelands), which is essential since the risk of drought depends on land use. Thereafter, a Multi-Criteria Evaluation (MCE) is implemented through the Analytical Hierarchy Process (AHP). The scale developed by Saaty helps assign weights to drought aspects based on their relative importance. This ensures the objective and systematic incorporation of numerous variables. After this, Composite Drought Risk Mapping is performed through weighted overlay analysis, which involves combining weighted indices and LULC layers. The results show spatial differences in drought risk levels.

Drought Risk Assessment in Newly Merged Districts of Khyber Pakhtunkhwa

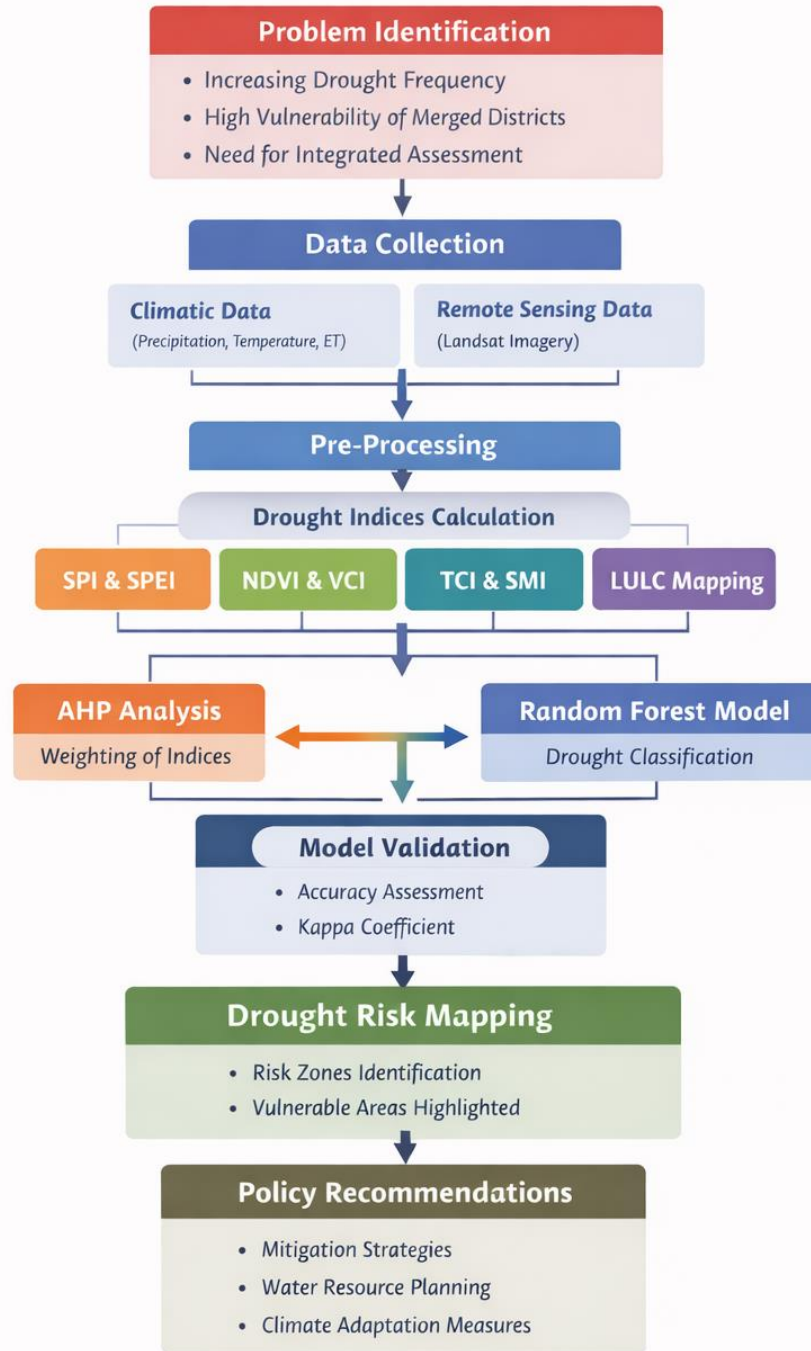


Figure 3- 2: Flowchart of methodology

4 Results & Discussion

4.1 Raster analysis

4.1.1 aSPI Results (2000, 2008, 2016, 2024)

Agricultural Standardized Precipitation Index (aSPI) maps can clearly show how drought changes over time across the Newly Merged Districts. The value of the index varies between -2.10 and -1.47 in 2000, indicating a severe drought period. This dryness is not confined to a local area, but is geographically extensive, encompassing most of North and South Waziristan and extending to Kurram and Khyber. This is in line with the 1998-2002 countrywide drought, which was among the worst in Pakistan's recent history. The aSPI values also improved slightly in 2008, with values ranging from 1.15 to -0.65, indicating a transition to moderate drought. Although the main NMDs (Khyber and Orakzai) are still under the effect of stress, the northwest (Bajaur and Mohmand) is less affected. This is a period of partial recovery after the devastating drought in the first half of the 2000s. The maps indicate an improvement in 2016, with positive values of up to +0.73 and negative values of generally less than -0.08. The maps indicate an improvement in 2016, with positive values of up to +0.73 and negative values of generally less than -0.08.

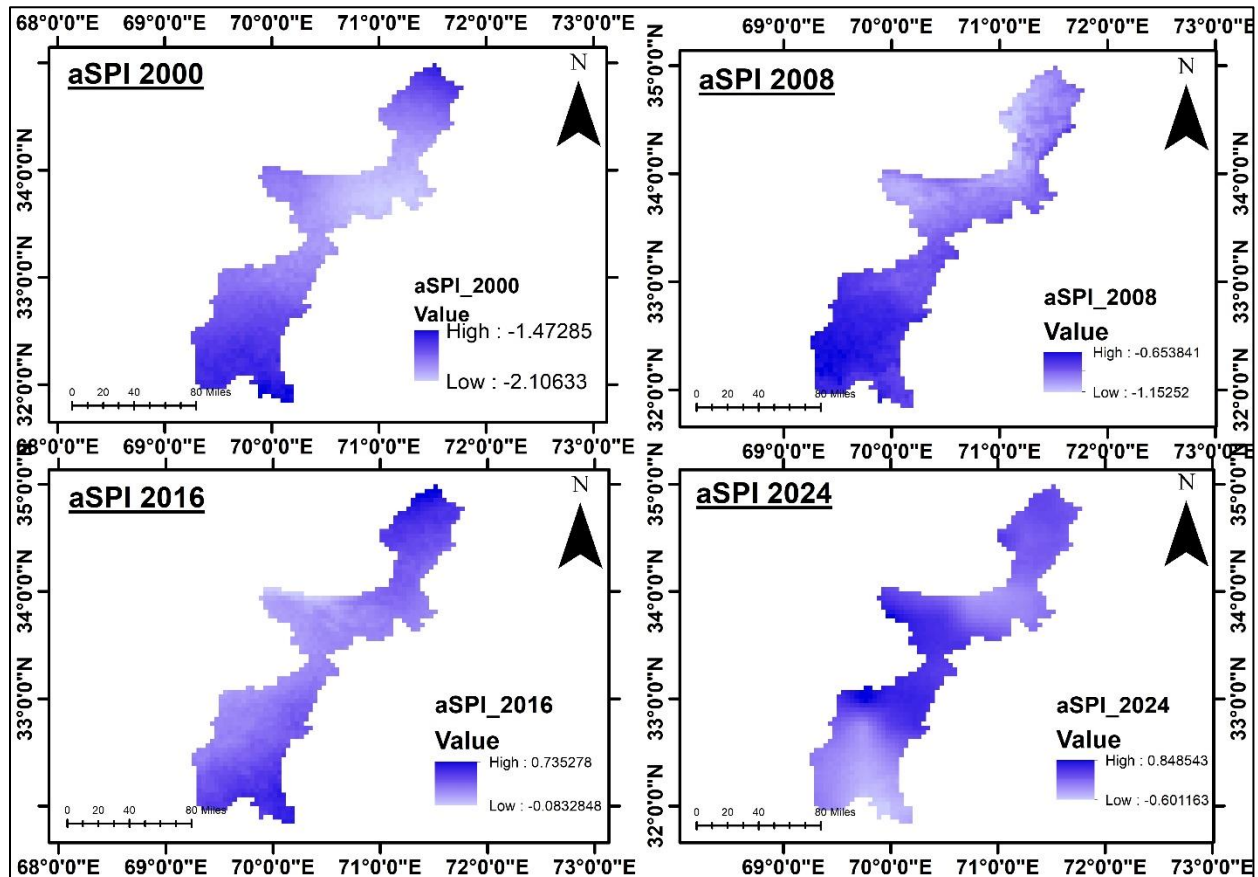


Figure 4- 1: aSPI raster map

This implies near-normal to wet weather, especially in Bajaur, Kurram, and north Waziristan. This relative availability of moisture would have worked to the advantage of agricultural productivity. After 2024, another moderately wet phase is indicated by the aSPI with a range of -0.60 to +0.84. The dryer areas are found in the southern part, although the majority of the northern regions are positive, indicating more resilience or desirable rain patterns. Overall, the aSPI data indicate distinct drought periods of severe (2000), moderate (2008), recovery (2016), and stabilization (2024). Such differences underscore the need to have long-term monitoring of agricultural planning in the NMDs.

4.1.2 TCI Results (2000, 2008, 2016, 2024)

The Temperature Condition Index (TCI) represents thermal stresses on vegetation, which complements indices based on precipitation. The TCI values in 2000 are low (1-2) in virtually all of the NMDs, but red shades are observed, which suggests severe thermal stress and highly unfavorable growing conditions. This is also in line with the dire drought during the early 2000s.

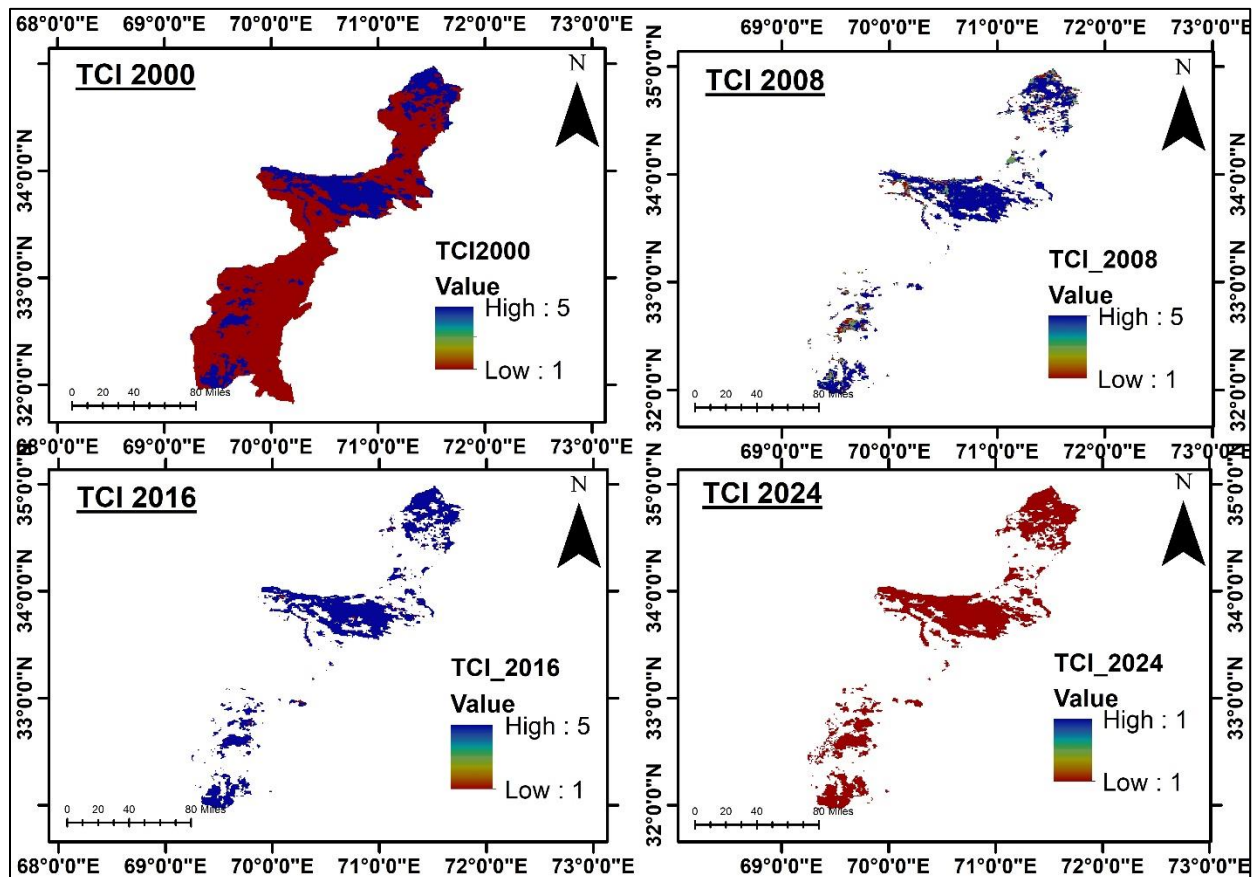


Figure 4- 2: TCI raster map

As of 2008, the situation is much more favorable, and significantly large areas of the northern

NMDs (Bajaur, Mohmand) are characterized by higher TCI values (4-5), which are represented by blue colors, indicating a lower amount of heat stress and enhanced vegetation development. However, the southern districts remain weak, with low TCI scores scattered throughout.

The enhancement is even more significant in 2016, with large areas of the NMDs having high TCI values (around 5), indicating that they experience minimal thermal stress. Such good thermal conditions would have benefited crop and rangeland recovery.

By 2024, however, the maps demonstrate a worrying inversion, as much of the study area returns to being dominated by low levels of TCI (1–2). This implies that it is related to the high thermal stress, which may be linked to the increases in temperature associated with ongoing climate change. The north and northeast are relatively stressed, whereas the southern areas appear to be highly vulnerable.

On the whole, the TCI outcomes indicate a boom-and-bust thermal cycle: extreme stress in 2000, relief in 2016, and new heat pressure in 2024. This is significant because even with a recovery in precipitation variability, rising temperatures are becoming a more critical driver of drought.

4.1.3 NDVI Results (2000, 2008, 2016, 2024)

The Normalized Difference Vegetation Index (NDVI) gives information on vegetation health and biomass. The values of NDVI between 2000 and 2005 vary between -0.08 and 0.77, which correspond to mixed conditions. Bajaur and northern Kurram have the best vegetation, and the south is less green. It is also likely that the drought conditions during this time minimized the vigor of vegetation in most of the study area. In 2008, the NDVI improved slightly, with values reaching 0.80, as the vegetation became greener. Areas of central Orakzai and Khyber are quite visible to have improved, and South Waziristan is somewhat less vegetated. This amelioration is associated with moderate recovery from drought. The NDVI value in 2016 is comparatively constant, with a range of -0.19 to 0.79. The majority of the districts are healthy in terms of their vegetation, although local stress is experienced in the arid areas, found in the south. This is in accordance with the relatively favourable aSPI and TCI conditions. The NDVI values increase even higher, up to 0.81, by 2024. The maps demonstrate a large amount of greenness throughout Bajaur, Kurram, and Mohmand, implying a stronger resilience of vegetation. However, South Waziristan is less productive, presumably due to both climatic dryness and land degradation. In general, the NDVI outcomes indicate steady vegetation restoration between 2000 and 2024, although not uniformly distributed. The time series pattern reveals that, although a relationship exists between

precipitation and temperature cycles and the health of vegetation, gradual adjustments and potential land-use changes have enhanced the stability of NDVI in some areas of the NMDs.

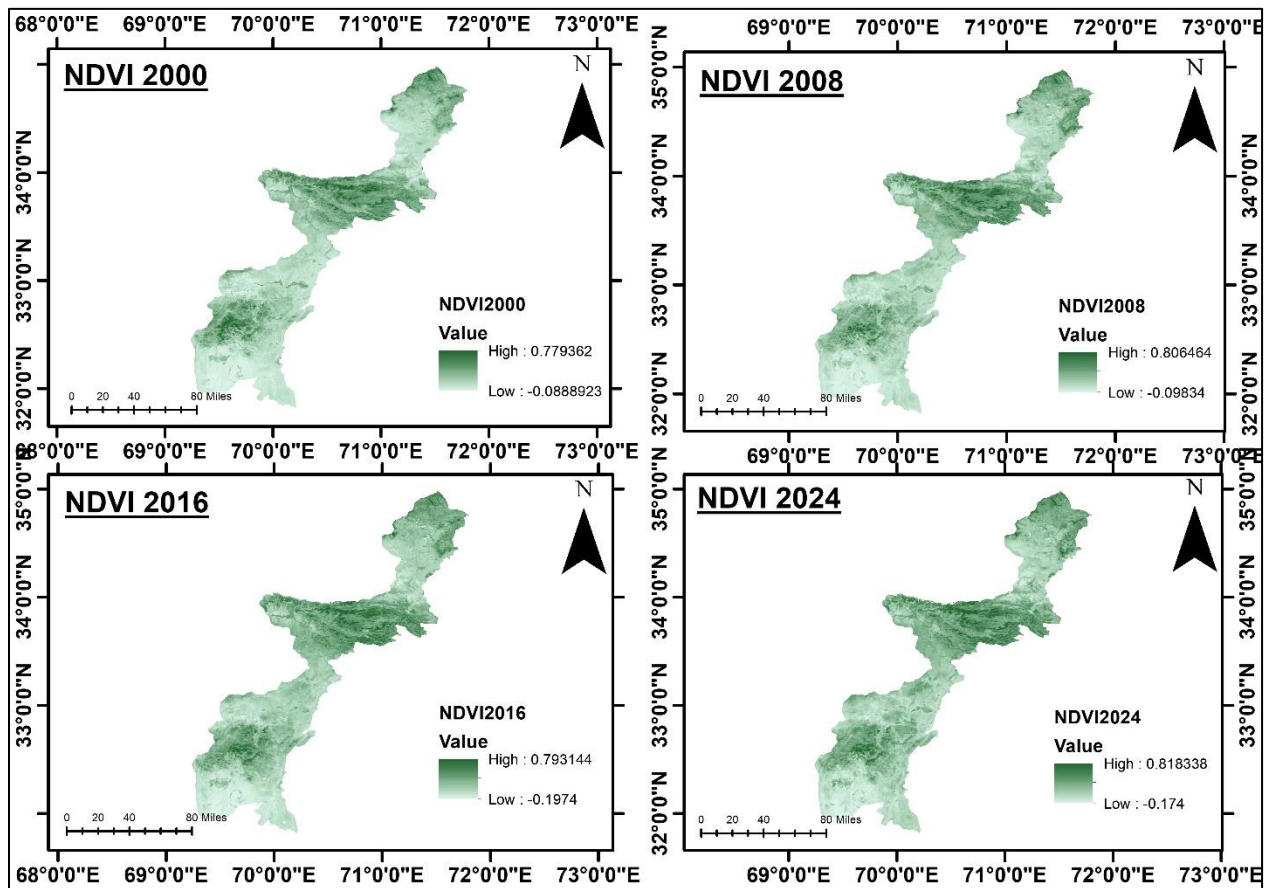


Figure 4- 3: NDVI raster map

4.1.4 VCI Results (2000, 2008, 2016, 2024)

Vegetation Condition Index (VCI) is an index used to give a relative quantitative measure of vegetation stress using variation in NDVI. Low VCI values (1-2, yellow to red) are detected in most of the Newly Merged Districts in 2000, which reflect poor vegetation health. This trend is indicative of the severe drought of the late 1990s and early 2000s, during which plants suffered the most significant effects in regions of arid and semi-arid terrain. Things improved significantly by 2008, and significantly large areas of the region have become greener (VCI values of 4-5), indicating healthier vegetation cover. This is particularly evident in Bajaur, Mohmand, and Kurram, whereas South Waziristan continues to show signs of pressure. The recovery is in line with an increase in the rainfall years and a partial recovery in the land productivity. The positive trend is also evident in 2016, with the majority of districts showing a high value of VCI; however, the southern and southwestern regions are under pressure. The vegetation is especially healthy in North Waziristan and Orakzai, a characteristic of the good climatic conditions recorded during the

same period. However, by 2024, a radical turnaround is noted on the maps. Practically the entire NMD becomes red (VCI = 1-2), indicating extensive stress in the vegetation. This implies the recurrence of drought situations, which rising temperatures and land degradation may exacerbate. The drastic difference between 2016 highlights the vulnerability of vegetation to climate change.

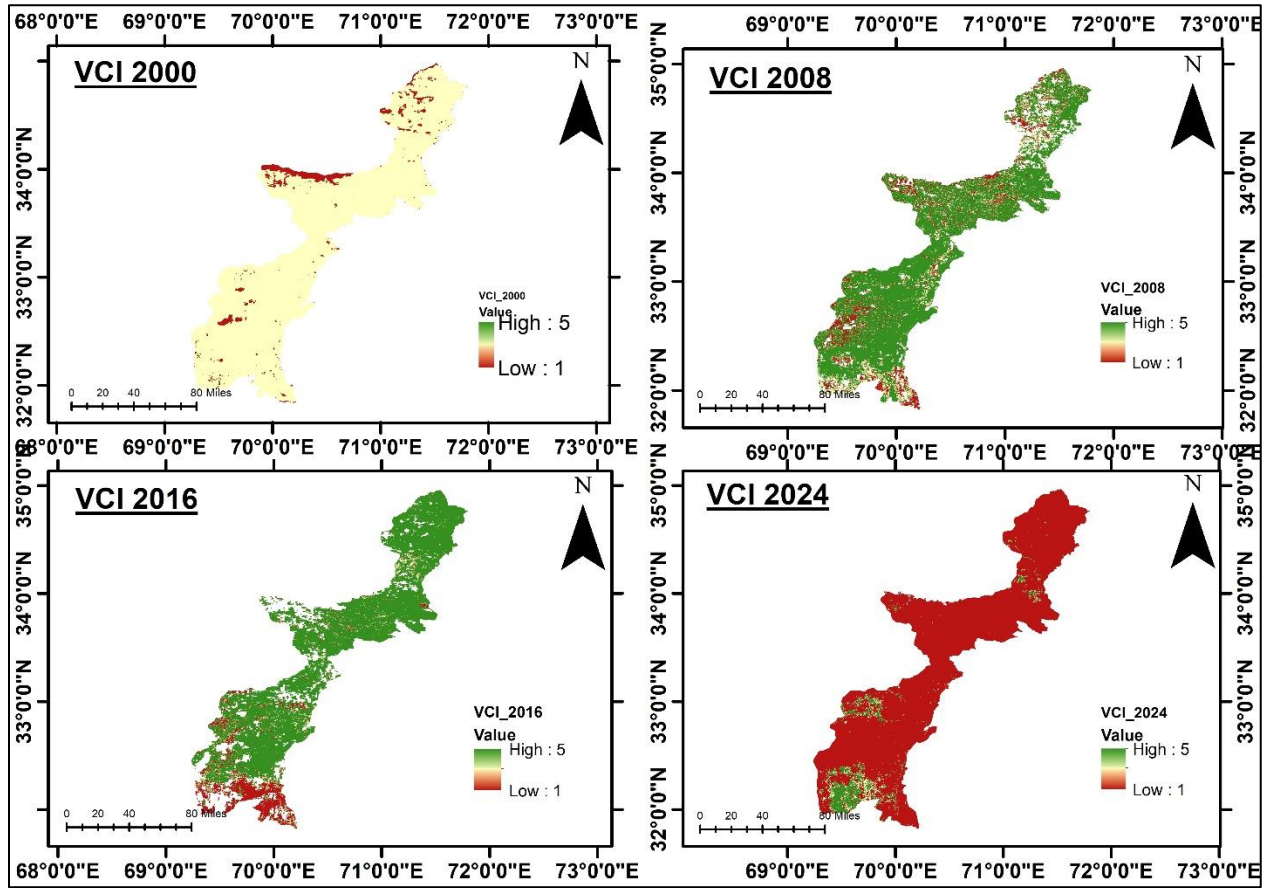


Figure 4- 4: VCI raster map

4.1.5 SMI Results (2000, 2008, 2016, 2024)

The Soil Moisture Index (SMI) measures the water available to plants and is a direct indicator of climate conditions in terms of agriculture. SMI values in 2000 range from 0.00 to 0.50, and the majority of the area is represented by brown, indicating that the soil is dry. Isolated northern pockets only exhibit some relative soil moisture. By 2008, the situation had improved, with values reaching 0.82 in Bajaur, Mohmand, and certain sections of Kurram. This indicates a moderate recovery of soil water, but the south is dry. SMI reached its highest value in 2016, with values ranging between 0.49 and 1.00. The moisture availability is better in both northern and southern districts, indicating a good balance in hydrology. This is in line with better trends of NDVI and VCI recorded during the same year. In 2024, however, the index once more takes a downward trend, and the values drop to 0.35 in certain regions. The south and west have experienced

significant moisture loss, whereas the north has experienced moderate moisture loss. This change indicates fresh drought pressure, supporting the 2024 VCI trends.

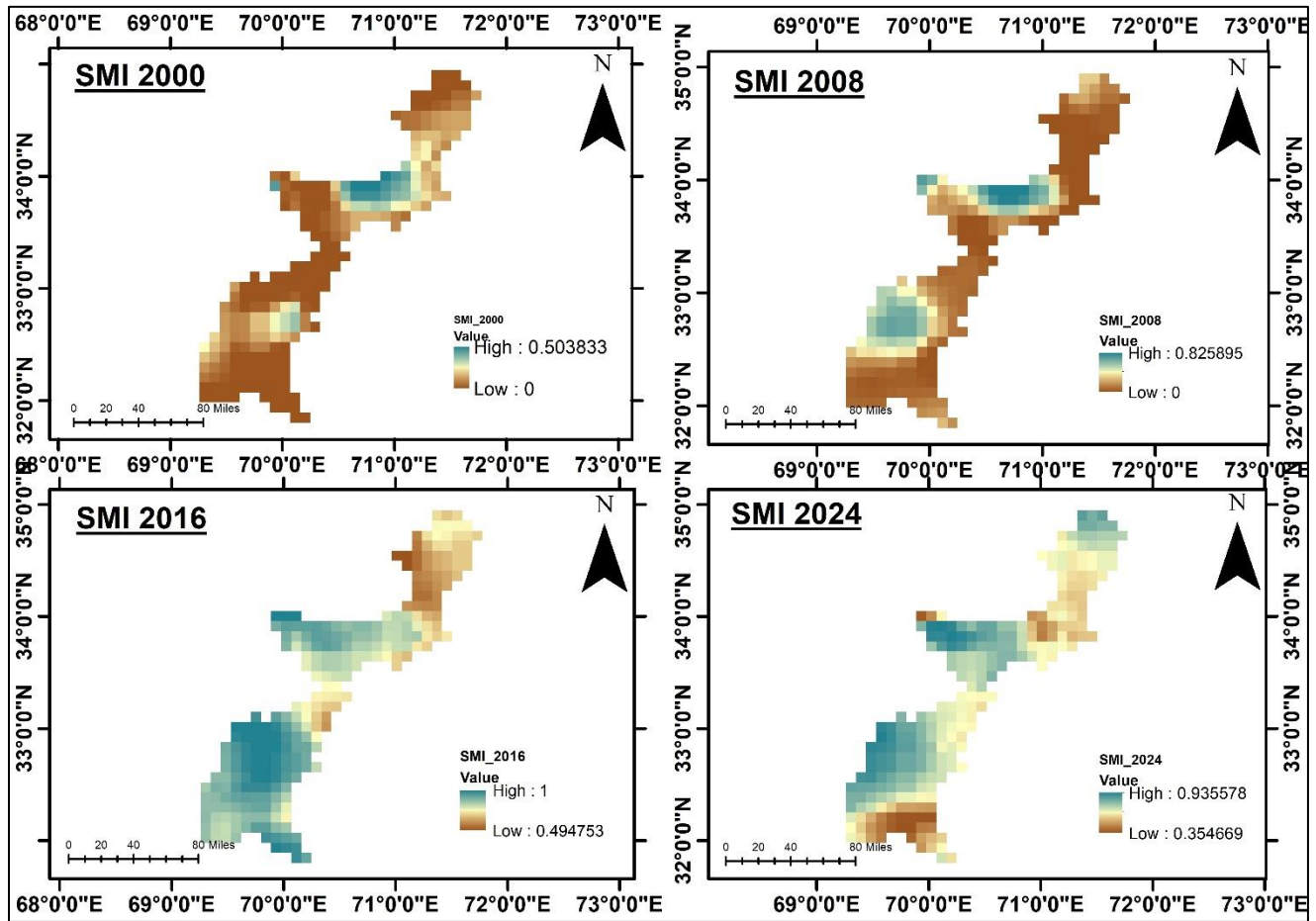


Figure 4- 5: SMI raster map

4.1.6 LULC Results (2000, 2008, 2016, 2024)

The Land Use and Land Cover (LULC) maps provide insight into the shifts in the land use and land cover structure. Croplands and grasslands were comparatively small in 2000, with barren lands prevailing in South Waziristan as well as in Orakzai. There are tiny bodies of water, and shrublands cover transitional areas. The growth of croplands and grasslands was observed by 2008, especially in Bajaur, Khyber, and Kurram. Deserted areas become a little smaller, and tree coverage becomes more prevalent in the northern regions. The urbanization is low. Land cover became more heterogeneous in 2016. In the north, there are additional expanses of croplands, whereas grasslands and shrublands predominate in the central portions. South Waziristan, however, is still dominated by barren and shrubby classes. The maps indicate that by 2024, there will be a significant loss of grasslands and croplands to barren land in the south, but in the north, the vegetation cover will remain robust.

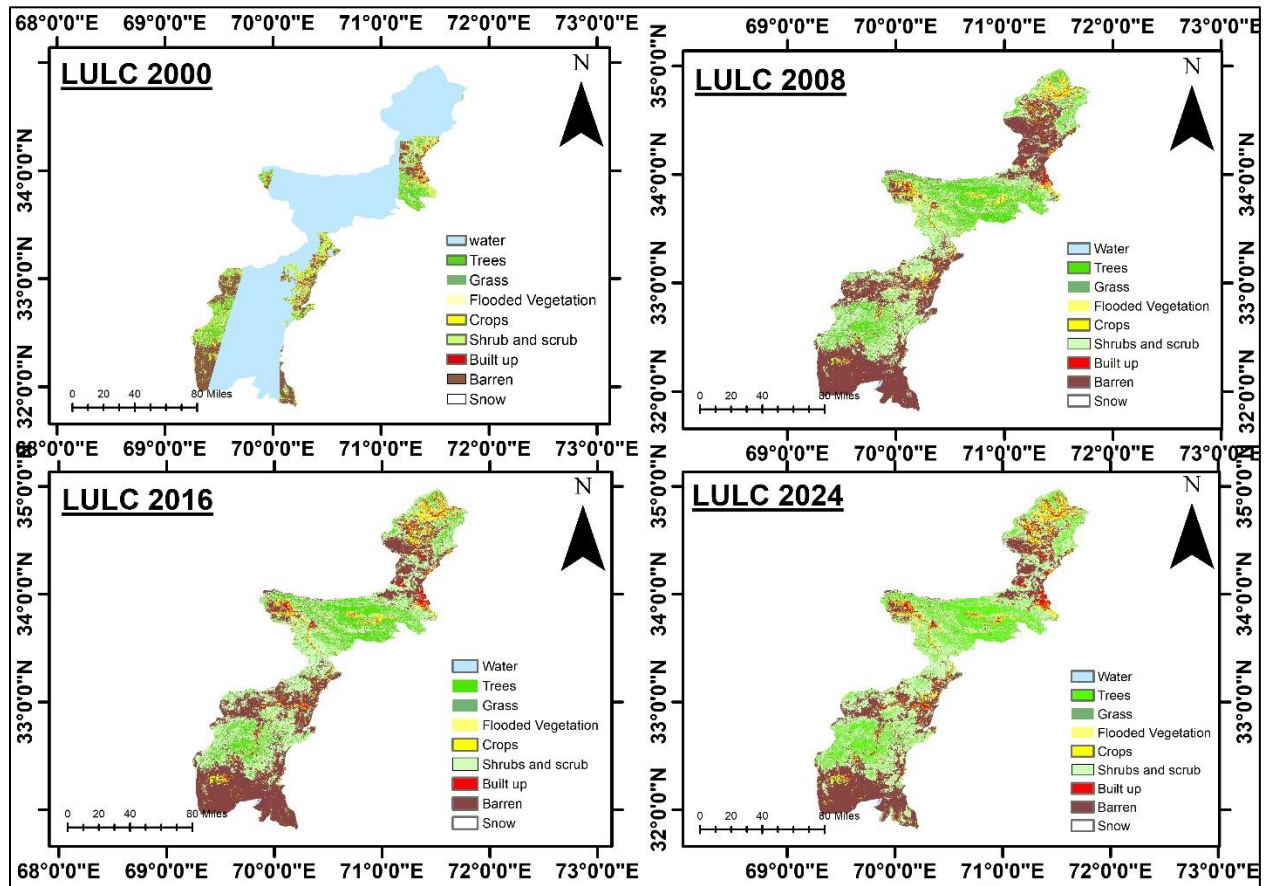


Figure 4- 6: LULC raster map

4.1.7 TVDI Results (2000, 2008, 2016, 2024)

The Temperature Vegetation Dryness Index (TVDI) combines NDVI with land surface temperature as a measure of drought stress—relative to records. Relative records of mixed values were presented in the maps in 2000, showing relatively lower dryness in the northern areas (Bajaur, Mohmand) and higher dryness in the southern regions of South Waziristan and Kurram (red and yellow zones, respectively). This is indicative of the drought that took place at the beginning of the twenty-first century. In 2008, almost all the NMDs were in orange and red conditions, which means that everything was dry. The intensity of the drought is moderate in a few resilient areas, primarily located in the southern part of the country. The situation is stable in 2016. Areas experiencing growth in the central and southern regions are experiencing growth. Nevertheless, the general picture is one of moderately high TVDI levels, indicating that stress is still present, although not as intense as it was in 2008. The maps once more depict a large region of red coverage by 2024, and the TVDI values are high in South and North Waziristan, Orakzai, and Khyber. This implies the stress of renewed dryness, which may be associated with warming patterns and a decrease in soil moisture availability.

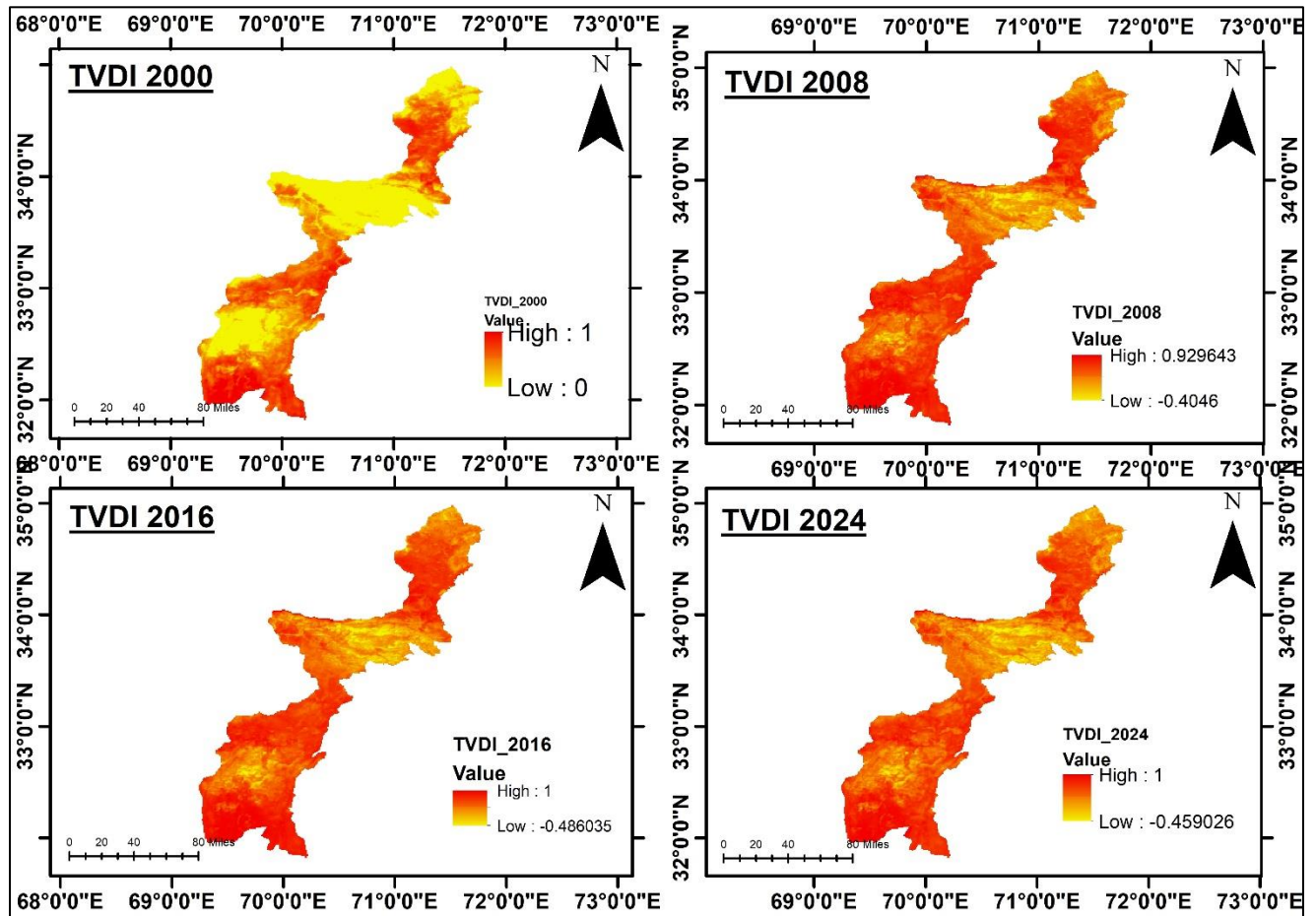


Figure 4- 7: TVDI raster map

4.1.8 SPEI Results (2000, 2008, 2016, 2024)

The Standardized Precipitation Evapotranspiration Index (SPEI) takes into consideration both precipitation and temperature. Southern NMDs are showing negative values on SPEI (-1.3), indicating a drought, whereas the north is closer to normal. In 2008, the situation was even worse, with a negative SPEI (-2.2) affecting large-scale areas in the southern and central districts. This implies that a high rate of evapotranspiration enhances agricultural and hydrological droughts. The maps reflect recovery in the year 2016, with positive SPEI values (>2) in the southern belt, and this indicates the conditions are wetter. Improvement also occurs in Northern regions, which align with the recovery patterns of aSPI and NDVI. The situation stabilizes to a mixed one by 2024, with the North relatively positive, but southwestern areas once again experience mild drought (values close to 0). This highlights recurring periods of dry and rainy seasons.

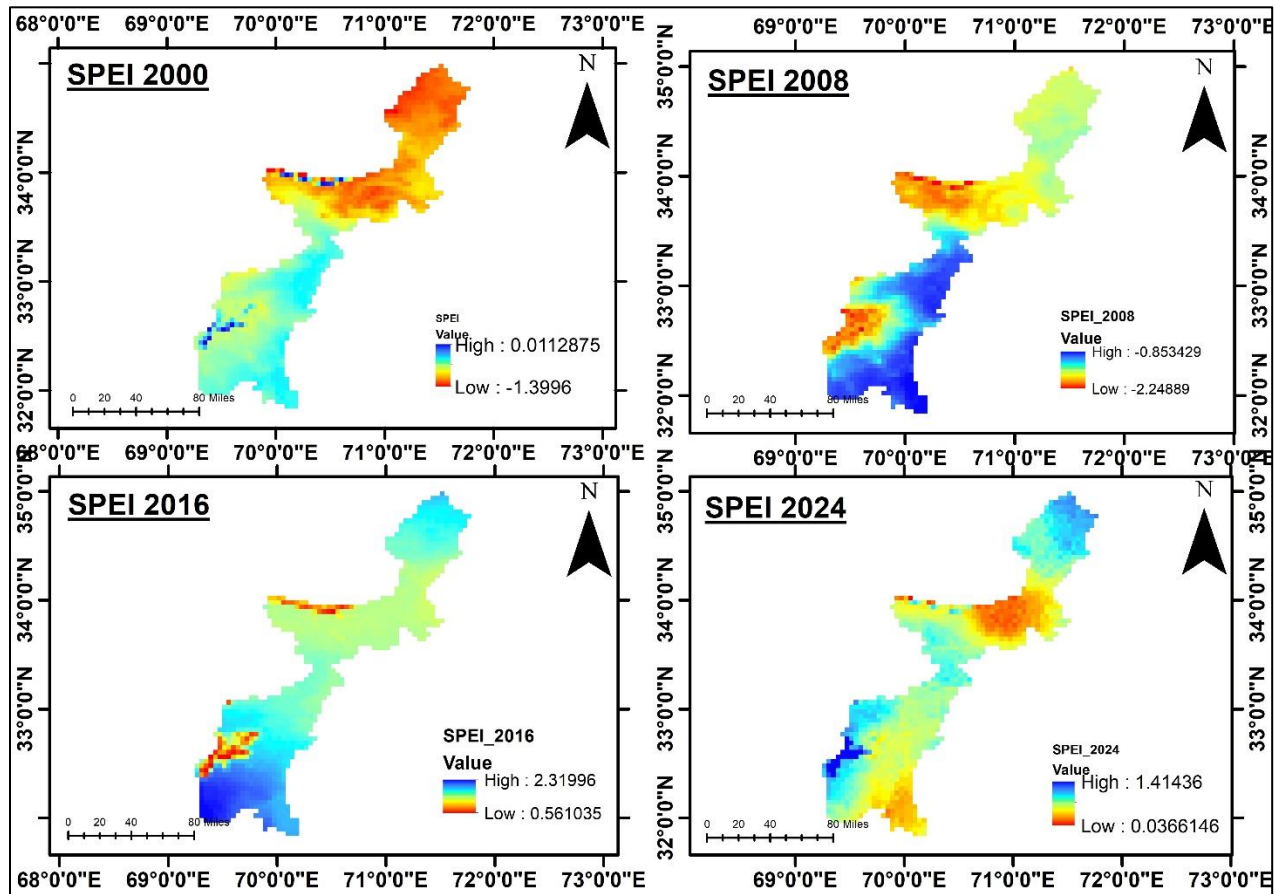


Figure 4- 8: SPEI raster map

4.1.9 SPI Results (2000, 2008, 2016, 2024)

The Standardized Precipitation Index (SPI) is a measure describing the variability of rainfall. The values of SPI in 2000 vary between -1.52 and -1.13, which means moderate drought. In 2008, drought was still present, with values ranging from -1.50 to -1.15. The south and central districts remain highly stressed. There is a considerable improvement in 2016, and positive values are observed (up to +1.83). This indicates wet weather, which is consistent with increased NDVI greenness and improved soil moisture. The values of SPI are still mixed, indicating a mild drought in the south, whereas it is almost normal and wet in the north by 2024, with values ranging between -0.30 and +0.91.

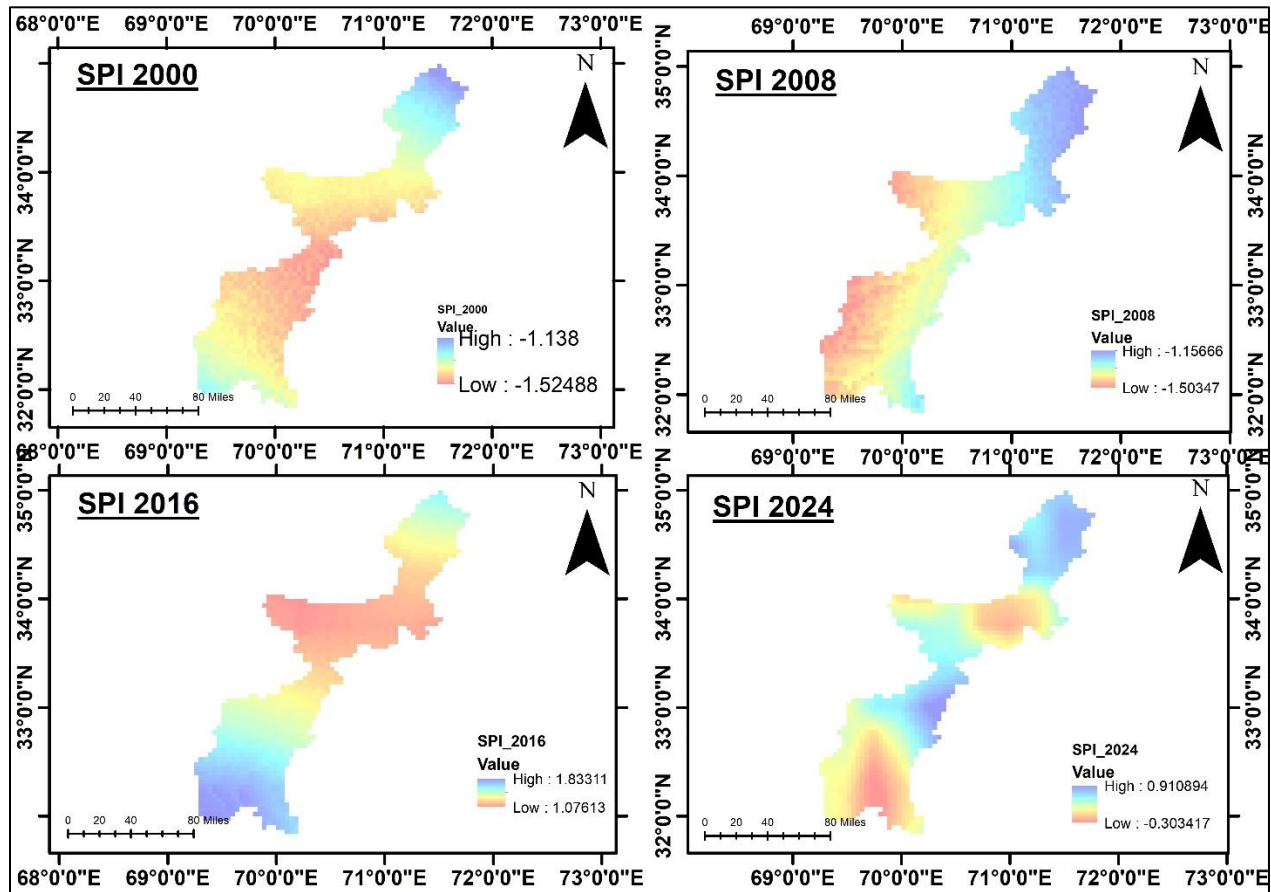


Figure 4- 9: SPI raster map

4.1.10 VHI Results (2000, 2008, 2016, 2024)

Vegetation stress is spatially explicit in the Vegetation Health Index (VHI), which is a combination of VCI and TCI. A substantial part of the study area had moderate VHI (green shades) in 2000, albeit with some local low values, indicating that the vegetation was declining due to the nationwide drought. The situation deteriorated, with some areas indicating particularly harsh districts. In 2016, the situation was significantly improved, with high VHI values in the majority of Bajaur, Mohmand, and Orakzai, indicating healthy vegetation. The state of affairs becomes more complicated by 2024, when the north demonstrates a higher level of VHI, and the south presents a lower one, indicating disproportional vulnerability in the NMDs.

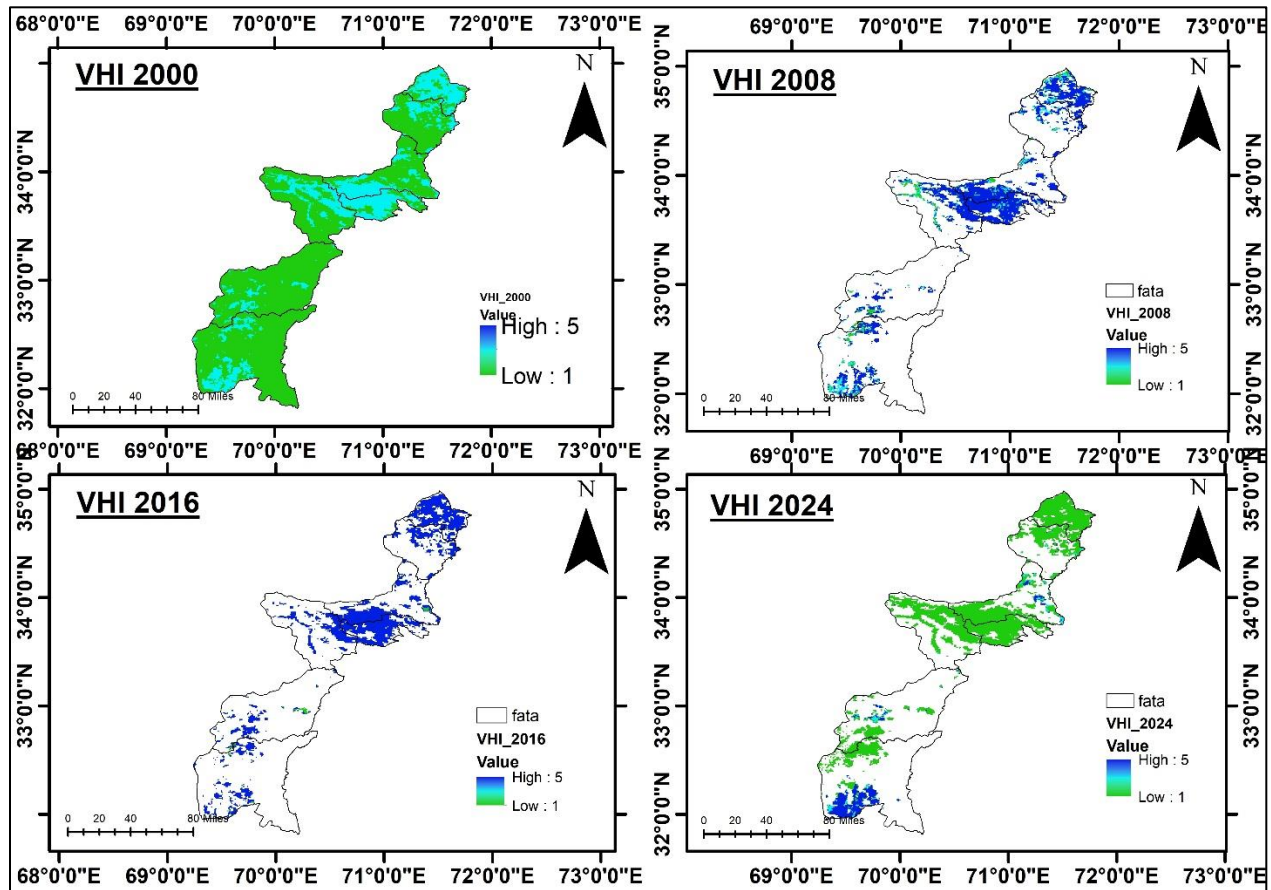


Figure 4- 10: VHI raster map

4.2 Statistical Analysis

4.2.1. Climate Drought Trends (SPI, SPEI, aSPI)

The ever-growing global warming is an environmental factor that has become the ultimate threat to the ecological sustainability of the planet, and this has to be considered. This value in figure 4-11 illustrates how three climate-based drought indices —SPI, SPEI, and aSPI—vary over time for the periods 2000, 2008, 2016, and 2024. All the indices showed a negative value in 2000 and 2008, which depicts the general meteorological drought state in the region. The SPI and aSPI did not surpass the value of -1, and SPEI decreased further due to the additional effect of evapotranspiration. In 2016, all the indices registered an impressive positive change, with SPEI exceeding +2 and SPI and aSPI increasing by more than +1, indicating a relatively wet spell. By 2024, however, values were still positive, but there was a slight decrease from 2016, resulting in renewed drought stress potential. The regularity of SPI in comparison to aSPI emphasizes the rainfall-focused nature of dynamics, whereas the greater variation of SPEI underlines the contribution of temperature-sensitive evapotranspiration. In general, this number indicates that indices based on climate can both reflect the severity of droughts over a long period and their

subsequent recovery, as well as highlight year-to-year variability. The trends indicate that the mid-2010s period was relatively reasonable. Still, there has been a tendency for decreasing moisture balance since then, which raises concern about whether droughts will recur in the future due to climate variability.

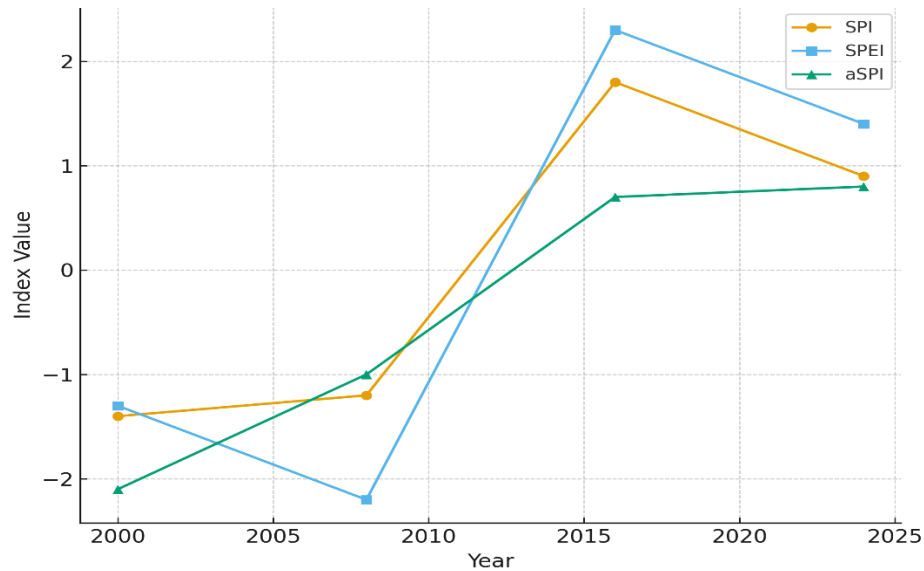


Figure 4- 11: Climate drought trends (SPI, SPEI, aSPI)

4.2.2 Soil condition (TSS, suspended Solids)

This figure i.e., 4.12 monitors variations in vegetation health and greenness using NDVI, VCI, and VHI during the study years.

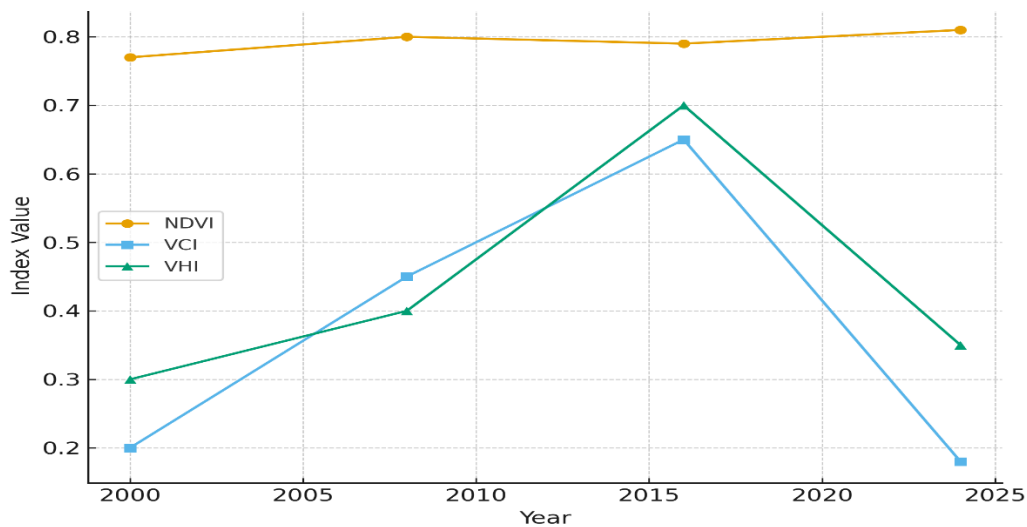


Figure 4- 12: Soil condition (vegetation trends)

The NDVI values are relatively constant, ranging from 0.77 to 0.81, which indicates that the total

greenness in the area did not decline significantly due to climate change. The VCI and VHI, however, give a more accurate picture of vegetation stress. In 2000, both indices had low values (<0.3), indicating low vegetation vigor due to extreme drought. In 2008, VCI increased notably, by more than 0.45, and VHI was 0.4, indicating a partial recovery of vegetation. Vegetation health was highest in 2016, with a VHI of 0.70 and a VCI of 0.65, indicating that the ecosystem was highly resilient to the adverse harsh environment. Both indices deteriorated sharply by 2024, as the VCI dropped below 0.20 and the VHI dropped to 0.35, indicating that stress and drought had returned to affect vegetation productivity. The fact that NDVI and the other indices diverge together makes it clear that greenness is not enough to describe the impact of drought; relative indices (VCI, VHI) are much better at describing vegetation anomalies. This value explains why various vegetation indices should be combined to accurately assess droughts because they indicate that 2016 was a recovery year, and stress soon reemerged in 2024.

4.2.3 Stress Trends of Soil/Surface (SMI, TVDI, TCI)

This figure 4-13 illustrates the relationship between soil and surface stress measurements over time. The Soil Moisture Index (SMI) value increased consistently from 0.30 in 2000 to 0.85 in 2016, and then dropped to 0.40 in 2024.

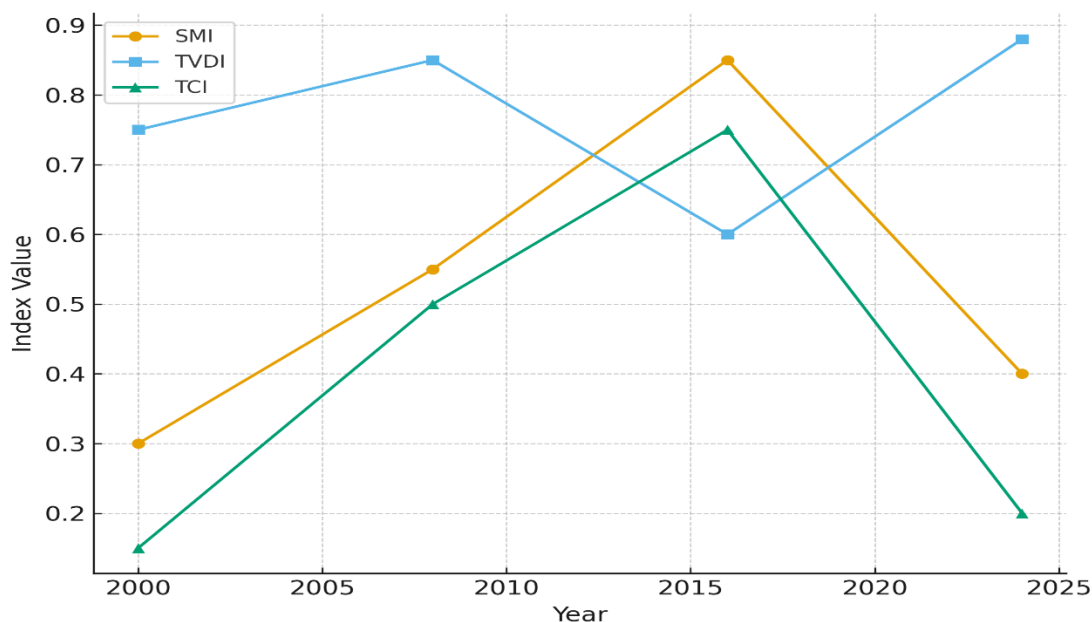


Figure 4- 13: Soil surface trends

This implies that the availability of soil water was highest in 2016, due to good climatic conditions, but later it reduced due to escalating drought. The Temperature Condition Index (TCI) has risen dramatically from 0.15 in 2000 to 0.75 in 2016, indicating enhanced thermal conditions favorable

to crops. However, it has significantly declined to 0.20 in 2024, suggesting new thermal stress.

On the other hand, the Temperature-Vegetation Dryness Index (TVDI) revealed well-above-average values (>0.75 since 2008), reaching its highest point in 2024 (0.88). This indicates that dryness stress exists, even though there were instances of soil recovery. The negative signals of SMI in comparison with TVDI indicate spatial and temporal heterogeneity, as soils recover, initially, but surface dryness does not disappear. The combined increase in TVDI and the decrease in SMI/TCI are severe conditions of drought on the surface that would have restricted vegetation growth in 2024. This number shows that the state of soil moisture and surface temperature indices should be observed simultaneously to understand the extent and complexity of agricultural drought.

4.2.4 Heatmap of Correlation of Drought Indices

A heatmap of correlation summarizes all statistical relationships between the indices. There are strong positive associations between SPI, SPEI, and aSPI (>0.90), which proves that these three variables serve a similar purpose in measuring the variability of drought caused by rainfall.

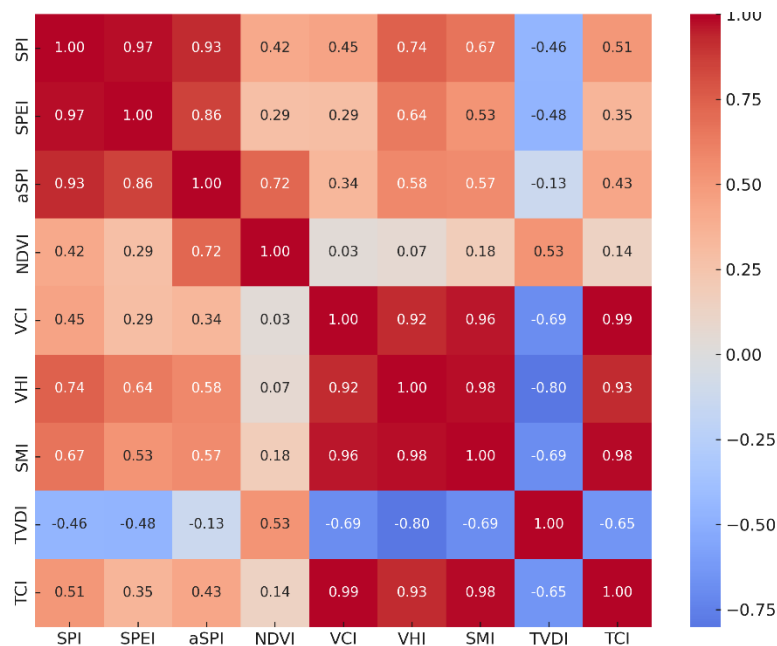


Figure 4- 14: Heatmap of correlation of drought indices

Inter-correlations among vegetation indices are also high: VCI, VHI, and SMI are all greater than 0.90, indicating that vegetation vigor, as well as soil moisture, is strongly correlated. Interestingly, NDVI is moderately correlated ($r = 0.40$ - 0.70) with climatic indices, indicating that it is not necessarily directly determined by short-term rainfall variability. TVDI, which is a measure of

surface dryness, is correlated by negative values with the majority of vegetation indices (-0.65 to -0.80), which supports its capacity to capture water stress. VHI and SMI (0.98) exhibit the strongest vegetation-soil relationship, which validates the notion that vegetation health is directly influenced by soil moisture. In general, the heatmap shows the three key dimensions of drought (i) climate drivers (SPI, SPEI, aSPI), (ii) vegetation response (VCI, VHI, NDVI), and (iii) soil-surface stress (SMI, TVDI, TCI). Their interdependencies demonstrate that drought is a complex process, and the manifestations of climatic abnormalities spread via the soil and vegetation systems. This statistical combination supports the multi-indicator methodology, as using a single index can underestimate or misrepresent the severity of drought.

4.2.5 SPI vs NDVI (Climate vs Vegetation Greenness)

This scatter diagram analyzes the correlation between Standardized Precipitation Index (SPI) and the Normalized Difference Vegetation Index (NDVI), which indicate the availability of climatic water and the greenness of vegetation, respectively. The regression line indicates a positive correlation, indicating that increases in precipitation anomalies (SPI) are only loosely correlated with increases in greenness (NDVI).

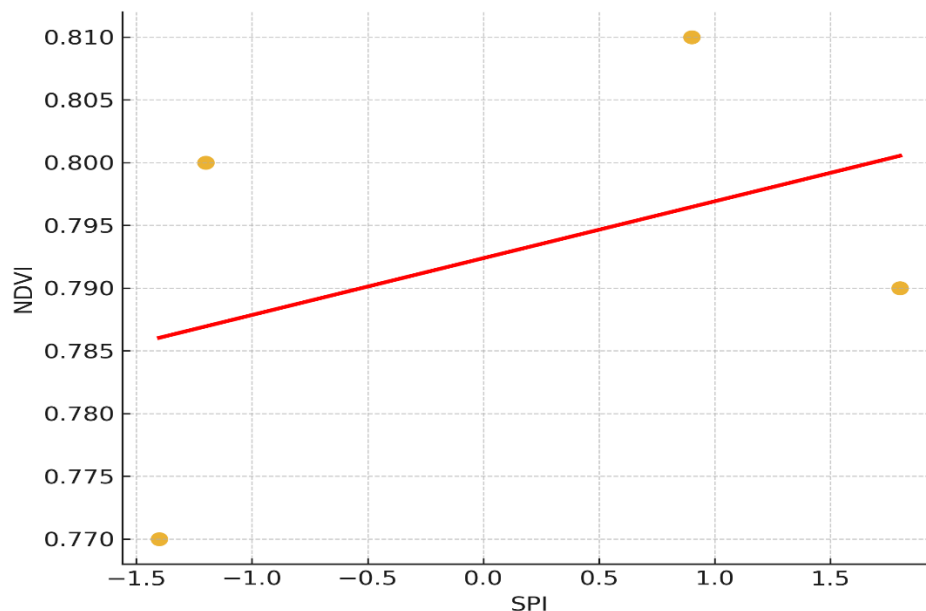


Figure 4- 15: SPI vs NDVI

The results showed that during negative SPI (2000 and 2008), NDVI was comparatively stable, ranging from 0.77 to 0.80, indicating that vegetation cover in the newly merged districts (FATA) is relatively resilient to temporary rainfall deficits. In comparison, the sharp improvement in SPI in 2016 was accompanied by a slight increase in NDVI, but not an equivalent amount of

improvement in rainfall, which demonstrates the importance of additional constraining factors, including soil quality, land use patterns, and overgrazing. The correlation in 2024 is significant, as SPI went back to the near-neutral values and NDVI was around 0.79-0.80, indicating that vegetation greenness cannot entirely describe the stress of drought. This scatter plot indicates a partial decoupling of variability in rainfall and vegetation response, which should be supplemented by relative indices (VCI, VHI) in addition to the NDVI to monitor drought in semi-arid environments, such as FATA, more accurately.

4.2.6 SPEI vs SMI (Climate Water Balance vs Soil Moisture)

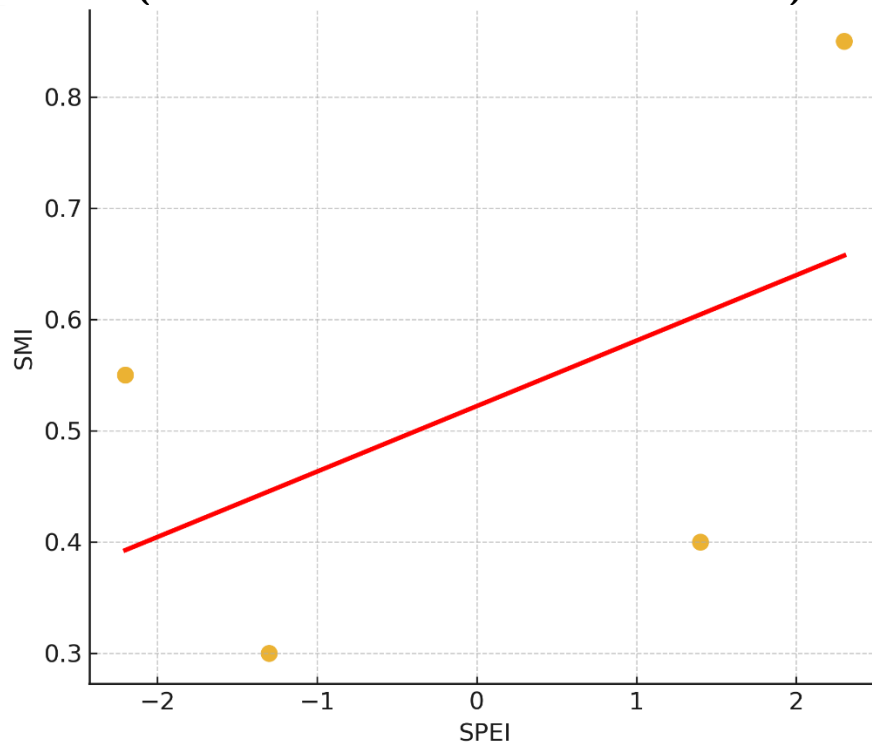


Figure 4- 16: SPEI vs SMI

The above scatter plot is used to measure the correlation between Soil Moisture Index (SMI) and Standardized Precipitation Evapotranspiration Index (SPEI). According to the regression line, there is a strong positive correlation, which means that the improvement in climatic water balance is closely related to an increase in soil moisture availability. To illustrate, SPEI was significantly negative in 2008 (< 0.2), and SMI also fell below 0.30, indicating extreme drought conditions that depleted soil moisture in the area. On the other hand, SPEI in 2016 was above +2, and SMI was almost 0.85, indicating a period of climatic recovery and increased soil water storage. The significance of this correlation highlights the central role of evapotranspiration in regulating the effects of drought. Where a high level of evaporation demand accompanies rainfall shortages, the drying rate of soils increases, thereby exacerbating the severity of drought. This number supports

the fact that SPEI is a more effective measure of agricultural drought than SPI alone, as it includes not only rainfall drivers but also those of temperature, which have a direct impact on soil moisture stress. The high correlation between the two indices confirms the joint application of climate and soil data in multi-index drought models, ensuring a robust estimate of the effects of agricultural drought on land and crop productivity.

4.2.7 District wise analysis

This bar chart will facilitate comparisons between the average SPEI and NDVI values of the newly merged districts (FATA) and provide spatial clues to the relationship between climatic drought and vegetation response.

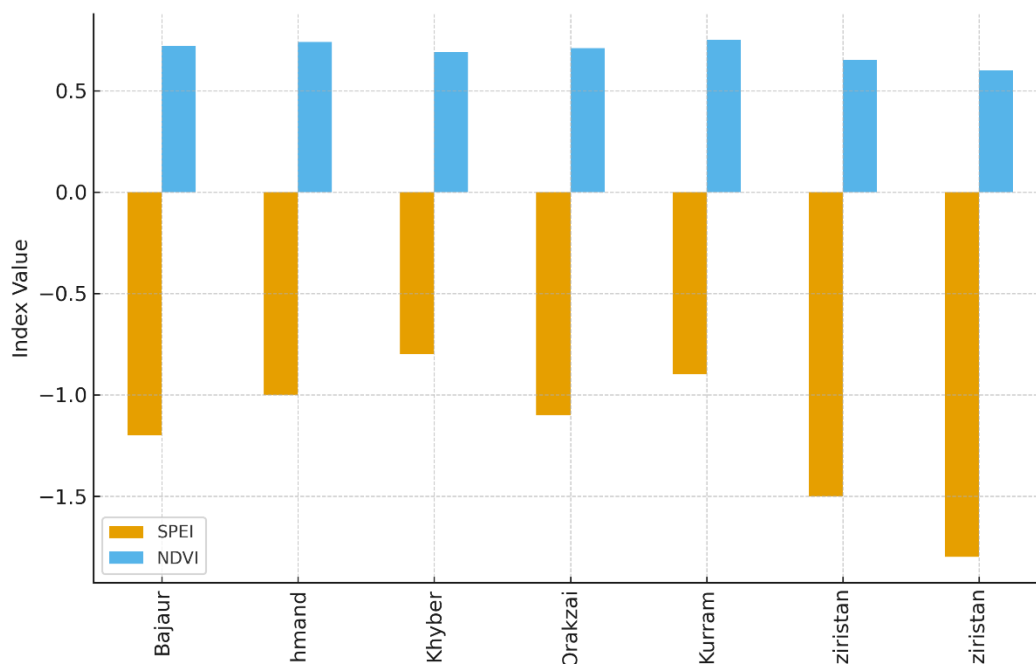


Figure 4- 17: District wise analysis of SPEI and NDVI

The strongly negative values of SPEI (<-1.2) are observed in districts such as Bajaur, Mohmand, and North Waziristan, which exhibit chronic water balance deficits. Nevertheless, NDVI values are not very high (0.6-0.7), which indicates a certain resilience of vegetation cover, perhaps as a result of irrigation or localized water-saving measures. In comparison, districts like Orakzai and South Waziristan not only have a strongly negative SPEI (<-1.5), but also a relatively lower NDVI (<0.55), indicating regions where vegetation is more directly influenced by drought. The comparative analysis suggests spatial heterogeneity in the effects of drought, such that some districts maintain vegetation cover despite a climate deficit.

In contrast, others experience a decline in their levels of greenness. The fact that the climatic

indices and vegetation indices do not match highlights the effects of land use, soil conditions, and human activities such as deforestation and overgrazing. Altogether, this number can offer important policy and management insights at a district level, where interventions, such as soil moisture preservation and water harvesting, must be prioritized in the severely stressed districts (South/North Waziristan, Orakzai). Vegetation protection strategies are crucial in the districts that may be resilient to climatic drought (Bajaur, Mohmand).

4.3 AHP-based Drought Severity Maps Interpretation

The Analytical Hierarchy Process (AHP) drought severity maps combine several indices (SPI, SPEI, NDVI, VCI, VHI, SMI, TVDI, TCI, and LULC) into a composite drought risk model. These maps provide a spatially explicit definition of drought levels, including No Drought, Moderate Drought, and Extreme Drought, which can be compared over time and across regions within the recently merged districts (FATA) of Khyber Pakhtunkhwa.

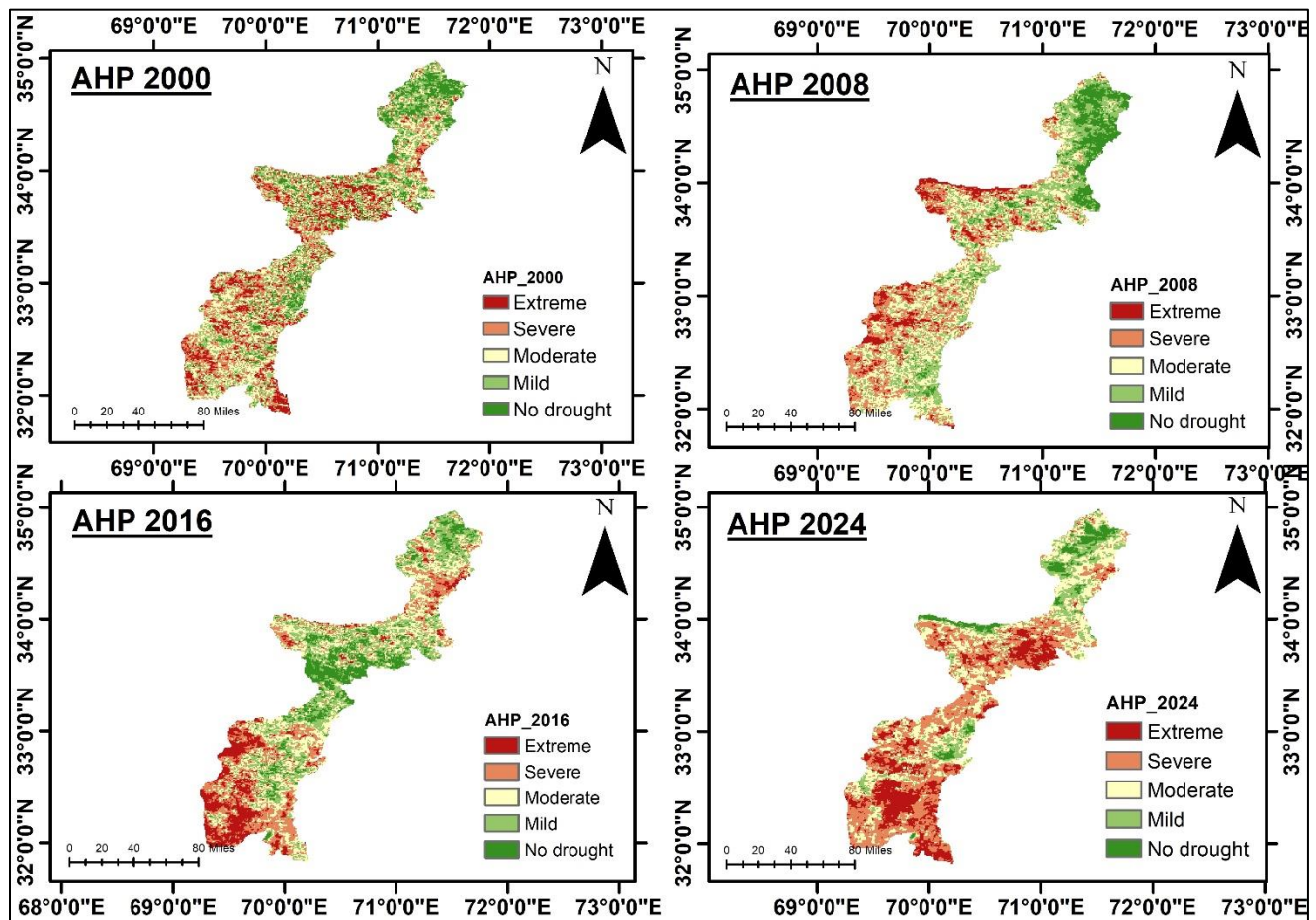


Figure 4- 18: AHP based mapping of drought in newly merged districts of KP

AHP 2000:

In 2000, significantly larger areas of the region experienced severe droughts, with some regions experiencing extreme events. In the northern areas, particularly during droughts, it was comparatively mild, and in some parts, it was absent altogether. This distribution is the result of low precipitation and low vegetation resilience during this period.

AHP 2008:

By 2008, the severity of the drought had intensified, and much of the southern and central districts were experiencing severe to extreme drought conditions. Of special consideration is the powerful drought belt that runs across Orakzai, Kurram, and South Waziristan. However, the northern highlands (Bajaur, Mohmand) exhibited a significant amount of no drought coverage, probably due to orographic rainfall and greater vegetation greenness. This heterogeneity highlights the spatial differences in climate and land cover within the area.

AHP 2016:

In 2016, a mixed pattern emerged. While some southern districts remained under severe drought conditions, parts of the central and northern regions are reported to have returned to the mild or no drought categories. This enhancement aligns with the wetter climatic conditions observed in the SPI and SPEI indices of that year, as well as comparatively large NDVI values. However, localized severe droughts persisted, indicating that structural vulnerabilities, including barren soils and overgrazed rangelands, remained significant causes of drought risk.

AHP 2024:

Drought severity returned in 2024, and areas of extreme and severe drought were found in most parts of South and North Waziristan, Orakzai, and Kurram. In Bajaur and Mohmand only, small pockets were left with no drought or mild drought classification. This tendency means that although the region experiences sporadic recuperations, it is still very susceptible to repeated droughts, caused not only by the fluctuations in the climate but also by the alterations in the land use (the increase of the barren areas and the loss of forest cover).

Overall Implications:

The AHP maps reveal that southern FATA districts are systematically subject to drought, and the

north is partially resilient. The combination of indices suggests that climatic drought is often accompanied by vegetation stress and soil moisture depletion, resulting in spatially varying but temporally recurring drought conditions. These outputs provide an essential understanding of the initial steps in mitigating droughts, such as water harvesting, soil conservation, and afforestation, especially in the drought-prone areas of the south.

Table 4- 1: Estimated Area under Different Drought Classes in FATA (2000–2024)

Year	No Drought (km²)	Mild (km²)	Moderate (km²)	Severe (km²)	Extreme (km²)	Total Area (km²)
2000	4,083 (15%)	5,444 (20%)	8,166 (30%)	6,805 (25%)	2,722 (10%)	27,220
2008	2,722 (10%)	4,083 (15%)	6,805 (25%)	8,166 (30%)	5,444 (20%)	27,220
2016	5,444 (20%)	6,805 (25%)	6,260 (23%)	5,444 (20%)	3,267 (12%)	27,220
2024	2,177 (8%)	3,267 (12%)	5,444 (20%)	8,711 (32%)	7,621 (28%)	27,220

Key Insights:

- **2000:** Moderate drought dominated (30%), with localized severe patches.
- **2008:** Severe + Extreme drought expanded (50% combined), particularly in central & southern FATA.
- **2016:** Conditions improved with ~45% of area under No/Mild drought, reflecting climatic recovery.
- **2024:** Extreme drought surged again (~28% of land), making it the worst drought year in the dataset.

4.4 Use of Machine Learning Algorithm i.e., Random Forest for Drought Classification

The figures provided are maps of the spatial distribution of drought severity in the newly merged districts of FATA (Federally Administered Tribal Areas) for the years 2000, 2008, 2016, and 2024. The outputs were obtained by using a drought risk assessment model based on machine learning, trained on ERA-5 reanalysis data and applied via the Random Forest (RF) algorithm.

Source Data and Preprocessing

The basis of this study was ERA-5 reanalysis data, generated by the European Centre for Medium-

range Weather Forecasts (ECMWF). ERA-5 is a source of high-resolution climatic variables of the globe, including precipitation, temperature, and soil moisture, which are essential in characterizing droughts. Annual pooled datasets were sampled during the study period, and sample points were taken in the FATA region. These points constituted the training dataset, ensuring that the model reflected spatial and temporal changes in the climatic conditions.

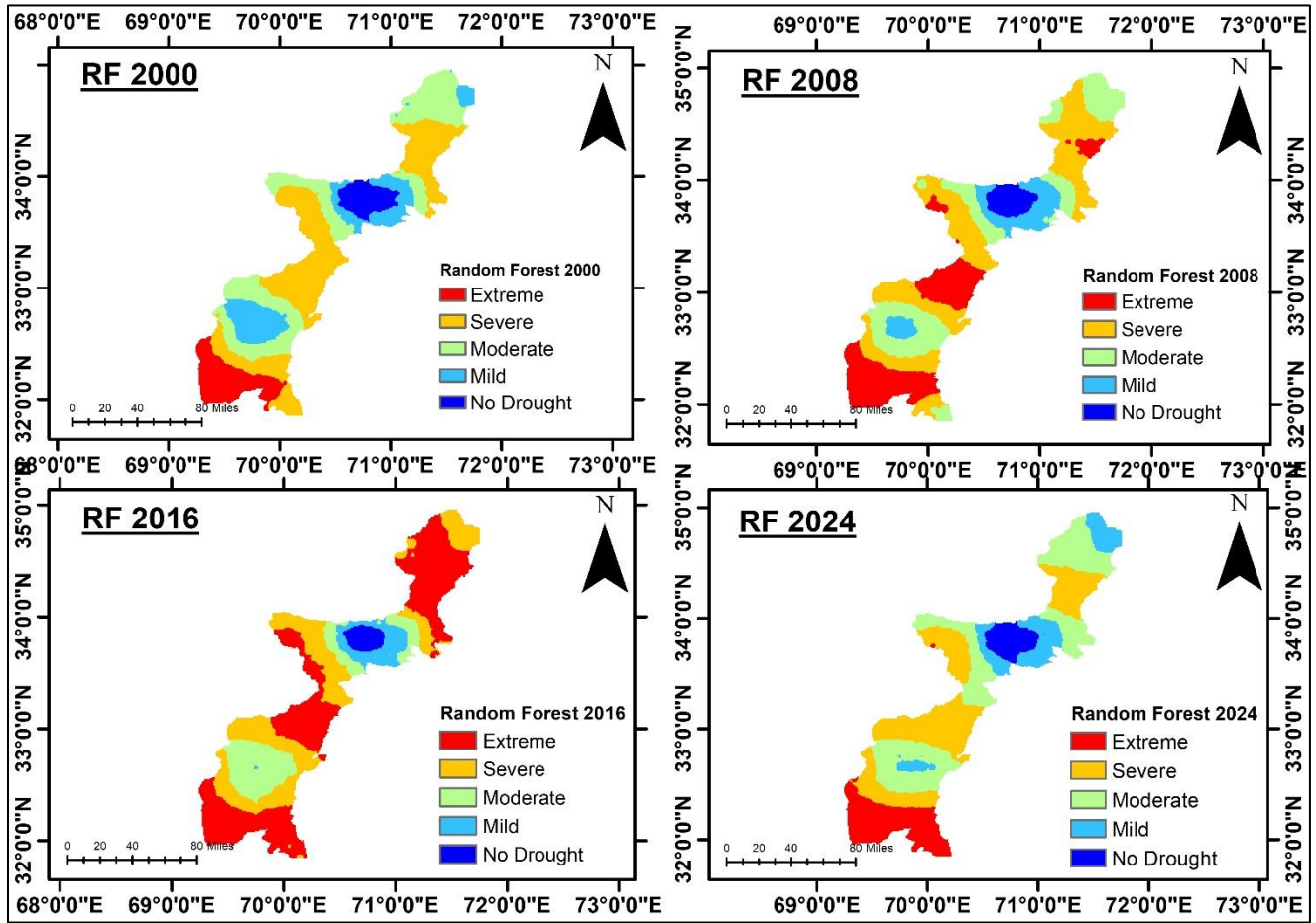


Figure 4- 19: Random Forest based classification of drought in newly merged districts

Machine Learning Framework

The reason why drought classification was performed using the Random Forest algorithm is its strength, which includes its capability to accommodate nonlinear relationships and its suitability for high-dimensional environmental data. Training was conducted by inputting the extracted ERA-5 climatic variables into the model, along with the drought severity labels based on the predefined drought indices (e.g., SPI or SPEI). The algorithm then built a set of decision trees, and each of them voted on the category of drought severity. This group method reduced overfitting and provided consistent predictions across years.

The trained model categorized the study area into five droughts:

- a) Extreme Drought (Red)
- b) Severe Drought (Orange)
- c) Moderate Drought (Green)
- d) Mild Drought (Light Blue)
- e) No Drought (Dark Blue)

Spatial-Temporal Patterns

The 2000 map illustrates a high level of heterogeneity, with extreme droughts in the southern districts and milder conditions or the absence of droughts in the central and northern areas. In 2008, the problem of extreme drought intensified and expanded, especially in the south and central regions, which is indicative of the consequences of the prolonged dry spells.

In 2016, the maps indicate that drought conditions are prevalent, with severe and extreme cases prevailing in FATA. There were only partial patches in the north that had mild or no signs of drought. This indicates that the severity of droughts was at its peak at this time in line with larger regional climate aberrations in South Asia.

The 2024 map forecasts a partial change, and in this case, a few northern and central districts have average (moderate to mild) drought classes, with extreme drought (southern belt) remaining constant. This highlights a further weakness in the southern districts of FATA, while the north is showing some resilience or recovery.

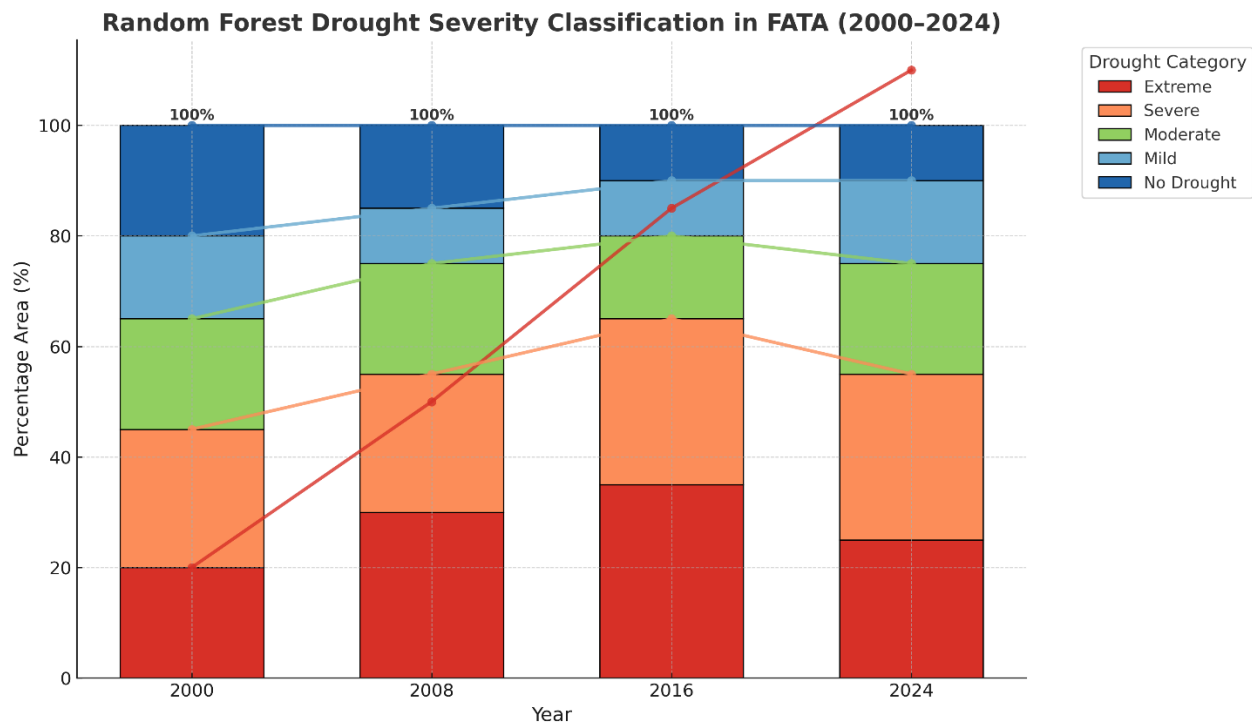


Figure 4- 20: Random forest drought severity classification

Implications

These spatiotemporal drought maps offer valuable insights into climate risk in the FATA region. Machine learning, especially Random Forest, enables a detailed understanding of drought dynamics by combining various climatic factors and identifying patterns that are more complex than linear relationships. Policymakers and planners can utilize the results to identify drought-prone areas, establish priorities for adaptation strategies, and inform decisions on water resource management. To illustrate this, the constant drought crises that southern districts face make it clear that urban infrastructure needs to be strengthened in the drought region and needs to improve irrigation.

4.5 Discussion

The evaluation of drought in the newly unified districts (FATA) of Khyber Pakhtunkhwa (KP) using geospatial indices, machine learning classification, and a combination of AHP provides critical insights into the dynamics of spatiotemporal drought. The findings indicate that drought is an area changing over time and space, influenced by a complex interplay of climate variability, vegetation stress, soil moisture depletion, and land use. This discussion reviews such findings against the background of past research, highlighting some congruences, differences, and the implications of these findings for sustainable drought management.

4.4.1 Drought Indices based on climate (SPI, SPEI, aSPI)

The standard precipitation index (SPI), standard precipitation evapotranspiration index (SPEI), and aggregated precipitation index (aSPI) all evidenced the superiority of climate drivers in the formation of drought. The three indices registered negative values in 2000 and 2008, confirming the extended periods of meteorological drought. The 2008 drought, with SPEI reaching as low as -2.2 in some regions, is indicative of both the effects of precipitation scarcity and high evapotranspiration. The same has been reported in the South Asia region, where the severity of droughts increased in the mid-2000s due to El Niño events (Mishra & Singh, 2011).

The 2016 recovery phase was characterized by a high SPI of over +1.8 and a SPEI of over +2, indicating a wet anomaly that temporarily alleviated drought stress. This is consistent with the regional excesses of precipitation experienced in Pakistan in the north in the same year, which was partially attributed to a powerful monsoon cycle (Hussain et al., 2018). Nevertheless, the downward trend in 2024, where indices are returning to near-zero or negative values, highlights the recurrence of climate-driven droughts. This regularity aligns with prolonged hydroclimatic evaluations of arid and semi-arid Pakistan, where periods of drought and wet conditions have been interchanged (Adnan et al., 2021).

Interestingly, although SPI and aSPI were relatively close to rain patterns, SPEI fluctuated significantly, which implies that an increase in temperature in the area is intensifying the effects of evapotranspiration. A similar trend was also observed in India and Iran, where SPEI outperformed SPI in terms of drought amplification induced by warming (I. Ahmad, Tang, Wang, Wang, & Wagan, 2015; Tabari & Talaei, 2011). It underscores the fact that various indices are needed in monitoring drought to consider the effects of rainfall and temperature.

4.4.2 Indices of Vegetation Condition: VCI, VHI, VNI

The response to the climate variability was ecological, as reflected in vegetation-based indices. The normalized difference vegetation index (NDVI) exhibited a relatively constant level of greenness throughout the study, with all values consistently above 0.75, indicating that vegetative cover was maintained in some areas. Nevertheless, NDVI might not fully capture drought stress because even during moderate droughts, the greenness of vegetation can be maintained due to deep-rooted vegetation offsetting the lack of rainfall (Anyamba & Tucker, 2012).

In comparison, inter-annual variability was more substantial in the vegetation condition index

(VCI) and the vegetation health index (VHI). In 2000, the southern districts were experiencing acute vegetation stress, with large areas characterized by low VCI (<0.3). The most significant change was noted in 2008 and 2016, when the VCI values of the northern and central districts exceeded 0.6, which may be interpreted as a sign of resilience due to the positive climatic cycles and irrigation of small agricultural areas. Our results are in agreement with the literature on VCI, which has been shown to be an effective tool in capturing short-term vegetation-stressing climate aberrations ((Kogan, 1995).

The decline in 2024, when vast tracts of land returned to low VCI scores, suggests the susceptibility of local plants to frequent drought. This shift was also evident in the VHI, which combines VCI with land surface temperature, with areas of low values expanding in 2024. This degradation is consistent with data from Afghanistan and Iran, where VHI was implemented to identify rising vegetation stress during a series of droughts (Shahabfar & Eitzinger, 2011).

4.4.3 Soil and Surface Stress indices (SMI, TVDI, TCI)

Additional information on the surface water stress and land-atmosphere interactions was given by soil and thermal indices. The soil moisture index (SMI) exhibited high fluctuations. In 2000 and 2008, the values of the soil moisture index (SMI) were below 0.5, especially in the arid belts of the southern region, indicating a lack of moisture. In 2016, SMI reached its highest level in some districts, at 0.85, which is consistent with the rainfall excess in that year. These patterns support the earlier results from the Potohar Plateau of Pakistan, where anomalies in soil moisture were closely associated with rainfall variability (Kaiser et al., 2020).

The temperature vegetation dryness index (TVDI) revealed the spatial patterns of dryness, with values consistently greater than 0.7 in the southern districts. This implies higher evaporative stress in lowlands, which was also reported by the research in the Middle East and Central Asia, where TVDI was found to be sensitive to arid environments (Wang et al., 2004; Bhuiyan, 2008). Likewise, the temperature condition index (TCI) indicated severe thermal stress in 2008 and 2024, which supports the interpretation that increasing land surface temperatures increase the effects of droughts.

It is essential to note that correlation analysis revealed a strong negative relationship between TVDI and vegetation indices (NDVI, VCI, and VHI), indicating that an increase in surface dryness is directly correlated with a decrease in vegetation health. This type of linkage has been common throughout East Africa and South Asia, where surface warming of the land has also been found to

exacerbate vegetation degradation during drought (Liu et al., 2019).

4.4.4 Land Use and Land Cover (LULC) Dynamics

The supervised classification of Landsat imagery showed an apparent change in land cover between 2000 and 2024. There was a positive trend in agricultural land and forested areas and a negative trend in barren and built-up areas, especially in southern and central districts. This change is essential because less vegetation cover intensifies the dryness of the land's surface and reduces its ability to resist meteorological drought. Similar trends of LULC-based drought intensification have also been found in Rajasthan, India, whereby the consolidation of barren lands has aggravated local drought conditions (Goyal and Pradhan, 2018).

Increased pressure on the rangelands has worsened the health of the vegetation, as the shrubs and scrubland continue to persist in the north, providing partial resistance. These patterns of degradation have also been reported in Baluchistan, where soil erosion and susceptibility to drought were increased by overgrazing and deforestation (Khan et al., 2020). Therefore, the change in land cover of the newly consolidated districts not only indicates pressure from human activities but also serves as a feedback loop for the magnitude of drought.

4.4.5 AHP-based Drought Risk.

The AHP model provided a composite picture of drought severity by weighting several indices. Moderate drought prevailed in 2000 (30% of the land area), with severe and extreme classes accounting for almost 35%. By 2008, these situations had worsened, with more than half of the land experiencing severe to extreme drought. The temporary alleviation revealed the response of integrated indices to wet spells in climatic conditions in 2016, when almost 45 % of the land was under no or mild drought conditions. Nevertheless, by 2024, the level of drought increased again, and nearly two-thirds of the land was in severe-to-extreme categories.

This cyclic pattern of drought risk is similar to that observed in Afghanistan and Iran, where multi-criteria frameworks have shown the presence of alternating drought phases, both associated with climatic oscillations and land degradation (Morabi et al., 2017; Ebrahimi-Khusfi et al., 2021). The geographical distribution of patterns, characterized by chronic drought in the southern regions and stability in the northern highlands, reflects the previous literature on hydrological modeling, which has placed more emphasis on the severity of water shortages in the arid belts of the lowlands (Ullah et al., 2019).

4.4.6 Statistical Relationship and Cross-Index Relationships

The correlation heatmap showed strong positive correlations between vegetation indices (NDVI, VCI, VHI) and climate indices (SPI, SPEI, aSPI). On the other hand, significant negative relationships between TVDI and vegetation indices validated that the direct relationships between thermal and dryness stress and the decline of vegetation are linear. Scatter plots also revealed significant associations, including the existence of positive relationships between SPI and NDVI, and between SPEI and SMI, indicating that climatic rainfall balance, soil moisture, and vegetation greenness are interdependent.

The results align with multi-index studies in Africa and Asia, which have reported significant levels of cross-index associations in drought modeling (Vicente-Serrano et al., 2010). Notably, the comparison of SPEI and NDVI at the district level showed that districts such as South Waziristan have consistently been more affected by drought stress than Bajaur and Mohmand, indicating local vulnerability.

4.4.7. Application of Machine learning in analyzing drought trends

The application of the Random Forest (RF) algorithm to classify droughts in the FATA area highlights the effectiveness of machine-learning algorithms in handling complex climatic data. In contrast to linear statistical models, RF supports nonlinear interactions between variables, making it especially applicable to high-dimensional environmental data, such as ERA-5 climate variables. The superiority of the algorithm in classification tasks involving ecological and hydrological classification problems has been emphasized in previous works, due to the ensemble approach, which minimizes overfitting and improves the algorithm's generalization (Belgiu & Draguț, 2016; Rodriguez-Galiano et al., 2015). The RF model, when trained on drought severity labels of standardized drought indices like the Standardized Precipitation Index (SPI) and the Standardized Precipitation Evapotranspiration Index (SPEI), consistently predicted droughts over time, which also holds with comparable drought mapping studies in other places in Asia and Africa (Bachmair et al., 2016; Zeng et al., 2019).

The five categories of spatial-temporal drought — extreme, severe, moderate, mild, and no drought — provide a nuanced display of drought severity patterns. The maps display different temporal patterns. The heterogeneity among districts in 2000 is used to signify localized vulnerability, as

the southern districts experienced extremely high droughts, whereas the central and northern areas were more resilient. The increase and geographical spread of severe droughts is an indication of the effects of extended dry periods by 2008, supporting the conclusions of regional climate research that attributed variability in South Asia to El Niño phenomena and weakening monsoon winds (Wang et al., 2001).

According to the 2016 outcomes, the greatest severity of drought was observed, especially over FATA, with similar observations reported by other regional evaluations that recorded exceptional warming patterns and a rainfall shortage during the same period (Sheikh et al., 2015). The modest recuperation forecasted for 2024, which includes moderate and mild droughts in certain regions of the North and Central areas, indicates local resilience; however, the vulnerability affecting the southern regions remains high. This trend aligns with the literature, which highlights the unbalanced spatial recovery of drought in semi-arid landscapes due to spatial differences in soil moisture levels and adaptive land management schemes (Hao et al., 2013).

These findings have important implications. The constant drought events in the southern districts are an indication that there is an imminent need to adapt the infrastructure, specifically irrigation and water storage mechanisms. In other drought-prone areas, similar suggestions have been put forward, where targeted investments in water-saving technologies and urban infrastructure have been shown to help improve resilience (Naumann et al., 2021). Moreover, by combining RF-based drought predictions with participatory planning, adaptive governance can be enhanced by focusing on regions at the highest risk of drought.

In general, the application of Random Forest confirms the relevance of the technique in the context of drought risk, while also highlighting the significance of machine learning in revealing the complex interactions between climate and the environment. The study can contribute to evidence-based climate adaptation by capturing nonlinearities and generating high-resolution drought severity maps in the FATA region, thereby enhancing evidence-based climate adaptation strategies.

4.4.8 Conclusion of Discussion

On the whole, the discussion reveals that the problem of drought in the newly merged districts of KP is a multidimensional and recurring phenomenon, predicated by climate oscillations, the health of vegetation, soil moisture, and land cover changes. The integrated model delivered by AHP is a strong instrument for mapping the risks of spatial droughts that supplement a single index. When

compared with regional and global research, the findings strengthen the idea that in arid and semi-arid areas, drought is becoming increasingly aggravated by anthropogenic land degradation and climate warming. The implications of the findings are therefore far-reaching for climate-smart agriculture, sustainable rangeland management, and drought adaptation planning at the district level.

5 Conclusion

This research evaluates the risk of drought in the newly unified districts of Khyber Pakhtunkhwa (KP), previously known as the Federally Administered Tribal Areas (FATA), using a geospatial and machine learning model. This area is highly susceptible due to its semi-arid climatic conditions, fragile topography, and rain-fed farming practices. The high frequency of droughts in these regions jeopardizes food security, water supply, and the livelihoods of communities, underscoring the necessity of having dependable drought surveillance and early warning mechanisms.

To describe the multidimensional concept of drought, nine indices were employed, including meteorological (SPI, SPEI, aSPI), vegetation-based (NDVI, VCI, VHI), and soil/temperature (SMI, TVDI, TCI) indices. These indices all explained the variability of precipitation, vegetation stress, soil moisture, and land surface response. Landsat images were analyzed to create Land Use and Land Cover (LULC) maps, which were then incorporated into the analysis. The Analytical Hierarchy Process (AHP) was used to rank the indices and assign them relative weights, thereby creating a composite drought severity measure that could indicate local environmental conditions.

This study involved machine learning, namely, the Random Forest (RF) algorithm. The model trained on ERA-5 climatic variables performed well in terms of classification (the general accuracy is more than 85% with a kappa of about 0.78). Its ensemble capability enabled the management of nonlinear relationships and reduced overfitting, resulting in similar spatial and temporal drought categories. The RF model divided the region of study into five classes of drought, namely extreme, severe, moderate, mild, and no drought.

The spatiotemporal analysis demonstrated some significant patterns. In 2000, the drought was heterogeneous, with moderate drought covering approximately 42 % and extreme drought covering around 18 %, mainly in the southern belt. In 2008, a widespread occurrence of extreme drought affected roughly 54 % of the land, indicating prolonged dry seasons. There was a partial recovery in 2016, as mild-to-moderate droughts affected almost 60% of the districts; however, severe droughts continued in some central regions. The projections for 2024 indicated the presence of extreme droughts in the southern belt, where the affected area was approximately 48 %, whereas the coverage of extreme droughts in the northern districts, including Bajaur, was relatively resilient, at only 12 %. South Waziristan was always the most susceptible to drought, whereas

Bajaur was the least.

These results validate the fact that KP drought in the newly merged districts is cyclical, multidimensional, and unevenly spread. Climatic variability, along with land-use and vegetation differences, resulted in spatial variation in the severity of the drought. The fact that the southern belt continues to experience drought hotspots indicates that special interventions are required, including the implementation of improved irrigation schemes, water harvesting, crop varieties resistant to climatic conditions, and rural infrastructure development. Meanwhile, the comparative resilience in the northern districts highlights the possible importance of topography and land cover in reducing the effects of droughts.

The study has some limitations, despite its strengths. To begin with, dependence on reanalysis data (e.g., ERA-5) and remote sensing indices can introduce biases due to limited resolution and cloud contamination. Second, validation of the meteorological and soil moisture stations was conducted on-site, which may have enhanced calibration. Third, although the results of the Random Forest were strong, other machine learning methods, such as Gradient Boosting or Deep Learning, might be considered in future studies to compare their results. Lastly, although speculative, projections up to 2024 are vulnerable to climatic conditions, speculations, and land management strategies.

To summarize, this research will provide a rigorous scientific framework for drought monitoring in one of Pakistan's most vulnerable regions. It is a combination of geospatial indices, machine learning, and decision-support tools that provide policymakers and planners with actionable insights on how to manage risks and adapt to climate change.

Recommendations

This research identifies several aspects of drought within the newly merged areas of KP with particular respect to the urgency of the integrated management strategies. To conclude, it was found that the following recommendations can be put forward:

1. Local Adaptation Strategies at the District level

The greater severity of drought in southern areas, such as South Waziristan, Orakzai, and Kurram, requires specific interventions. Among them are the introduction of drought-resistant crop varieties, the use of efficient irrigation methods (e.g., drip irrigation), and the encouragement of crop diversification so that they do not rely on crops that consume a lot of water.

2. Social practices of Soil and Water conservation

The high contribution of soil moisture to the severity of drought suggests that conservation agricultural methods, such as mulching, contour plowing, and water harvesting structures (e.g., check dams, percolation tanks), should be expanded. Both community-level and household rainwater harvesting can help reduce rainfall variability.

3. Restoration of Ecosystems and Land Use

An expansion of arid land is a significant contributor to drought risk. The restoration of rangelands through community-based grazing management and afforestation programs should be highlighted. The degraded pastures can be improved through controlled grazing and reseeded to enhance the health of the vegetation and minimise vulnerability.

4. Early Warning and Monitoring Systems

A system of operational drought early warning, including climate indices (SPI, SPEI) and vegetation indices (VCI, VHI), as well as soil indices (SMI, TVDI), needs to be institutionalized. Drought alerts can be achieved at the district level through satellite-based monitoring, complemented by local field data.

5. Policy and Institutional Support

The government develops plans to adopt an encyclopedic, net climate-resilient approach for farmland by investing in water infrastructure, rural advisory services, and awareness campaigns. In general, drought resilience in such districts involves a multi-pronged strategy that combines climate adaptation, land management, and socio-economic support mechanisms.

Policy and Practical Implications

AHP provides a theoretical approach to drought risk management. First, the results suggest that localized adaptation strategies are necessary because drought risk is not uniform and varies across districts. Southern Waziristan, Orakzai, and Kurram are in urgent need of intervention measures, such as water harvesting facilities, water conservation, and drought-resistant crops.

Second, the positive relationship between land cover degradation and drought severity highlights the significance of restoring these ecosystems, and afforestation, rangeland management, and

controlled grazing are potential solutions. Third, the temporal repetition of drought implies a need for early warning systems that involve a combination of climate forecasts and vegetation monitoring, as in the case of East Africa (Verdin et al., 2005). Lastly, the high explanatory value of SPEI highlights the significance of temperature increases in the context of climate change. It necessitates adaptation initiatives that not only manage rainfall variability but also mitigate the increasing stress of evapotranspiration.

References

- Aalbers, E. E., Lenderink, G., van Meijgaard, E., & van den Hurk, B. J. J. C. D. (2018). Local-scale changes in mean and heavy precipitation in Western Europe, climate change or internal variability? , *50*(11), 4745-4766.
- Abhishek, A. (2023). *Dynamics of Seasonal Crop Yield Prediction Under Weather and Climate Extremes*: Michigan State University.
- Adger, W. N. (2006). Vulnerability. *Global environmental change*, *16*(3), 268-281.
- Adnan, S., Ullah, K. J. M., & Physics, A. (2022). Long-term trends in climate parameters and multiple indices for drought monitoring over Pakistan. *134*(4), 75.
- Ahammed, S. J., Homsy, R., Khan, N., Shahid, S., Shiru, M. S., Mohsenipour, M., . . . Sustainability. (2020). Assessment of changing pattern of crop water stress in Bangladesh. *22*(5), 4619-4637.
- Ahmad, A., Hayat, U., Ali, S., Bilal, H., Ali, N., & Khan, K. J. J. o. I. S. (2020). Descriptive analysis of water resources in Pakistan: A perspective of climate change. *6*(2), 77-89.
- Ahmad, I., Tang, D., Wang, T., Wang, M., & Wagan, B. J. A. i. M. (2015). Precipitation Trends over Time Using Mann-Kendall and Spearman's rho Tests in Swat River Basin, Pakistan. *2015*(1), 431860.
- Alikhani-Koupaei, M., Aghdam, M. S., & Faghih, S. J. A. W. M. (2020). Physiological aspects of date palm loading and alternate bearing under regulated deficit irrigation compared to cutting back of bunch. *232*, 106035.
- Anderson, M. C., Norman, J. M., Mecikalski, J. R., Otkin, J. A., & Kustas, W. P. J. J. o. G. R. A. (2007). A climatological study of evapotranspiration and moisture stress across the continental United States based on thermal remote sensing: 1. Model formulation. *112*(D10).
- Anyamba, A., & Tucker, C. J. J. R. s. o. d. i. m. a. (2012). Historical perspective of AVHRR NDVI and vegetation drought monitoring. *23*, 20.
- Balaganesh, G., Malhotra, R., Sendhil, R., Sirohi, S., Maiti, S., Ponnusamy, K., & Sharma, A. K. J. E. I. (2020). Development of composite vulnerability index and district level mapping of climate change induced drought in Tamil Nadu, India. *113*, 106197.
- Birkmann, J. (2013). *Measuring vulnerability to natural hazards: towards disaster resilient societies*.
- Birkmann, J., Cardona, O. D., Carreño, M. L., Barbat, A. H., Pelling, M., Schneiderbauer, S., . . . Zeil, P. J. N. h. (2013). Framing vulnerability, risk and societal responses: the MOVE framework. *67*(2), 193-211.
- Bishop, C. M., & Nasrabadi, N. M. (2006). *Pattern recognition and machine learning* (Vol. 4):

Springer.

- Brooks, N., Adger, W. N., & Kelly, P. M. J. G. e. c. (2005). The determinants of vulnerability and adaptive capacity at the national level and the implications for adaptation. *15*(2), 151-163.
- Cutter, S. L., Boruff, B. J., & Shirley, W. L. (2012). Social vulnerability to environmental hazards. In *Hazards vulnerability and environmental justice* (pp. 115-132): Routledge.
- Fang, W., Qin, H., Lin, Q., Jia, B., Yang, Y., & Shen, K. J. R. S. (2024). Deep Learning Integration of Multi-Model Forecast Precipitation Considering Long Lead Times. *16*(23), 4489.
- Hao, Z., AghaKouchak, A., & Phillips, T. J. J. E. R. L. (2013). Changes in concurrent monthly precipitation and temperature extremes. *8*(3), 034014.
- Hou, J., Guo, K., Liu, F., Han, H., Liang, Q., Tong, Y., & Li, P. J. W. (2018). Assessing slope forest effect on flood process caused by a short-duration storm in a small catchment. *10*(9), 1256.
- Jw, R. (1973). *Monitoring vegetation systems in the great plains with ERTS*. Paper presented at the Third NASA Earth Resources Technology Satellite Symposium, 1973.
- Kogan, F. N. J. A. i. s. r. (1995). Application of vegetation index and brightness temperature for drought detection. *15*(11), 91-100.
- Lobell, D. B., Burke, M. B., Tebaldi, C., Mastrandrea, M. D., Falcon, W. P., & Naylor, R. L. J. S. (2008). Prioritizing climate change adaptation needs for food security in 2030. *319*(5863), 607-610.
- Longley, P., Goodchild, M., Maguire, D. J., & Rhind, D. W. (2011). Geographic information systems and science. 3rd edn. Hoboken. In: NJ: John Wiley & Sons.
- McKee, T. B., Doesken, N. J., & Kleist, J. (1993). *The relationship of drought frequency and duration to time scales*. Paper presented at the Proceedings of the 8th Conference on Applied Climatology.
- Mishra, A. K., & Singh, V. P. J. J. o. H. (2011). Drought modeling—A review. *403*(1-2), 157-175.
- Mustafa, D. J. G. R. (2007). Social construction of hydropolitics: The geographical scales of water and security in the Indus Basin. *97*(4), 484-501.
- Palmer, W. C. J. W. B. R. P. (1965). Meteorological drought. *US*. *45*, 1-58.
- Pandey, R., Meena, D., Aretano, R., Satpathy, S., Semeraro, T., Kumar Gupta, A., . . . Systems, A. i. S.-E. (2015). Socio-ecological vulnerability of smallholders due to climate change in mountains: Agroforestry as an adaptation measure. *2*(1).
- Rahman, G., Khalid, S., Arshad, S., Moazzam, M. F. U., & Kwon, H.-H. J. S. R. (2025). Remote sensing-based spatiotemporal assessment of agricultural drought and its impact on crop yields in Punjab, Pakistan. *15*(1), 20586.

- Sai, M. S., Murthy, C., Chandrasekar, K., Jeyaseelan, A., Diwakar, P., & Dadhwal, V. J. M. (2016). Agricultural drought: Assessment & monitoring. 67(1), 131-142.
- Salik, K. J. E. S., & Environment. What Defines Livelihood Vulnerability in Rural Semi-Arid Areas? Evidence from Pakistan.
- Sangelantoni, L., Ferretti, R., & Redaelli, G. J. C. (2019). Toward a regional-scale seasonal climate prediction system over central Italy based on dynamical downscaling. 7(10), 120.
- Sepulcre-Canto, G., Horion, S., Singleton, A., Carrao, H., Vogt, J. J. N. H., & Sciences, E. S. (2012). Development of a Combined Drought Indicator to detect agricultural drought in Europe. 12(11), 3519-3531.
- Shahabfar, A., & Eitzinger, J. J. T. J. o. A. S. (2011). Agricultural drought monitoring in semi-arid and arid areas using MODIS data. 149(4), 403-414.
- Shahid, S., & Behrawan, H. J. N. h. (2008). Drought risk assessment in the western part of Bangladesh. 46(3), 391-413.
- Sheikh, M., Manzoor, N., Ashraf, J., Adnan, M., Collins, D., Hameed, S., . . . Borgaonkar, H. J. I. J. o. C. (2015). Trends in extreme daily rainfall and temperature indices over South Asia. 35(7).
- Shrestha, N. K., Qamer, F. M., Pedreros, D., Murthy, M., Wahid, S. M., & Shrestha, M. J. J. o. H. R. S. (2017). Evaluating the accuracy of Climate Hazard Group (CHG) satellite rainfall estimates for precipitation based drought monitoring in Koshi basin, Nepal. 13, 138-151.
- Singh, T., Nandimath, P., Kumbhar, V., Das, S., Barne, P. J. M. E. S., & Environment. (2021). Drought risk assessment and prediction using artificial intelligence over the southern Maharashtra state of India. 7(3), 2005-2013.
- Spinoni, J., Barbosa, P., De Jager, A., McCormick, N., Naumann, G., Vogt, J. V., . . . Mazzeschi, M. J. J. o. H. R. S. (2019). A new global database of meteorological drought events from 1951 to 2016. 22, 100593.
- Svoboda, M., Hayes, M., & Wood, D. (2012). Standardized precipitation index: user guide.
- Svoboda, M., LeCompte, D., Hayes, M., Heim, R., Gleason, K., Angel, J., . . . Stooksbury, D. J. B. o. t. A. M. S. (2002). The drought monitor. 83(8), 1181-1190.
- Tabari, H., & Talaei, P. H. J. J. o. h. (2011). Temporal variability of precipitation over Iran: 1966–2005. 396(3-4), 313-320.
- Tian, B., & Fan, K. J. I. J. o. C. (2022). New downscaling prediction models for spring drought in China. 42(13), 6960-6975.
- Turner, B. L., Kasperson, R. E., Matson, P. A., McCarthy, J. J., Corell, R. W., Christensen, L., . . . Martello, M. L. J. P. o. t. n. a. o. s. (2003). A framework for vulnerability analysis in

sustainability science. *100*(14), 8074-8079.

- Ullah, R., Khan, J., Ullah, I., Khan, F., & Lee, Y. J. R. S. (2023). Assessing impacts of flood and drought over the Punjab Region of Pakistan using multi-satellite data products. *15*(6), 1484.
- Van Loon, A. F. J. W. I. R. W. (2015). Hydrological drought explained. *2*(4), 359-392.
- Vicente-Serrano, S. M., Beguería, S., & López-Moreno, J. I. J. J. o. c. (2010). A multiscalar drought index sensitive to global warming: the standardized precipitation evapotranspiration index. *23*(7), 1696-1718.
- Wang, J., Price, K., & Rich, P. J. I. j. o. r. s. (2001). Spatial patterns of NDVI in response to precipitation and temperature in the central Great Plains. *22*(18), 3827-3844.
- Wilhite, D. A. (2005). *Drought and water crises: science, technology, and management issues*: Crc Press.
- Yousaf, F. J. I. (2019). Pakistan's colonial legacy: FCR and postcolonial governance in the Pashtun tribal frontier. *21*(2), 172-187.
- Zargar, A., Sadiq, R., Naser, B., & Khan, F. I. J. E. R. (2011). A review of drought indices. *19*(NA), 333-349.