THE ASIAN BULLETIN OF BIG DATA MANAGMENT
Vol. 4. Issue 3 (2024)
https://doi.org/ 10.62019/abbdm.v4i3.190

ASIAN BULLETIN OF BIG DATA MANAGEMENT
ISSN (Print): 2959-0795
http://abbdm.com/
ISSN (online): 2959-0809

Flood Risk Analysis in Sindh, Pakistan: Predicting the Most Affected Tehsil Using Statistical and Machine Learning Models with Comprehensive Data

Amna Aziz*, Muhammad Ali, Syed Muhammad Hassan Zaidi, Faisal Nawaz, Faryal Sheikh

Chronicle

Article history

Received: July 20, 2024
Received in the revised format: July 30, 2024
Accepted: Aug 5, 2024
Available online: Aug 8, 2024

Amna Aziz is currently affiliated with Department of Computer Sciences IQRA University Karachi, Pakistan.
Email: amna.aziz26426@gmail.com

Muhammad Ali & Syed Muhammad Hassan zaidi is currently affiliated with Department of Artificial Intelligence & Mathematical Sciences Sindh-Madressatul -Islam University (SMIU) Karachi, Pakistan.
Email: mali@smiu.edu.pk
Email: m.hassan@smiu.edu.pk

Faisal Nawaz is currently affiliated with Department of Mathematics, Dawood University of Engineering and Technology (DUET), Pakistan.
Email: Faisal.Nawaz@duet.edu.pk

Faryal Sheikh is currently affiliated with Department of Computer Science Sindh- Madressatul-Islam University (SMIU) Karachi, Pakistan.
Email: faryalsheikh85@gmail.com

Corresponding Author*

Abstract

Extreme weather events pose significant risks and threats to the Sindh region, shedding light on the potential havoc they can wreak on livestock, human lives, and public infrastructure. Unusual monsoon rains in 2022 transformed the geography of Sindh, leading to devastating floods and significant damage. The current flood management strategies are effective, but there is some gap in early warning systems. Different findings suggest a need for refined approaches to improve predictive accuracy and response efficiency. The primary goal is to assess and improve existing flood response measures, identify opportunities for improvement, and strengthen resilience against flood events. In response to these challenges, this study provides actionable insights for policymakers and emergency planners to refine flood management strategies, ultimately aiming to bolster community resilience and preparedness. This paper presents a novel framework for a hybrid predictive model that combines machine learning and statistical analysis to improve the prediction of flood-induced displacement, providing timely warnings and preparing for future flood events. The selected dataset used for this research is from the 'Sindh Flood Data Analysis and Prediction 2022', which includes both demographic and geographical variables. By integrating machine learning and statistical models, the research evaluates their effectiveness in predicting flood impacts and identifying the most affected Tehsils using Python language and Seaborn library for analysis. The Gradient Boosting Classifier achieved the highest accuracy at 97.67%, followed by Logistic Regression at 95%. In contrast, Decision Trees and Lasso Regression demonstrated lower accuracies of 25% and 55%, respectively.

Keywords: Flood Risk Sindh, Extreme Weather, Machine Learning (ML), Climate Change, Disaster Management.

INTRODUCTION

As estimated, half number of people in this world are the resident of cities. Abbas and Guo (2023) research to different number of facilities and it is expected to increase rapidly by 2030. According to Aggarwal (2016), history shows that the climatic change has strong impact on the minds of people and their behaviors vary accordingly. Climate is a dynamic element of people's life. With the passage of time, change in the climate can be noticed worldwide. Ajani and van der Geest (2021) suggested intense weather, regular occurrence of heavy floods and rain, cloudburst, mudslides, at different places including rural and urban can be frequently seen in recent years. Amarnath and Rajah (2016) found that climate change, ruinous disaster and heavy floods have become the big reason for physical, social, and economic destruction worldwide. Every year flood damages societies at huge level and deviates people's lives completely. Everything including their families, behaviors and personal relationship with closest cities also badly distressed. Arshad and Shafi (2010) mentioned this disaster also led to the changes in building process. Communities deal this disaster differently with their own solutions. (Arshed et al., 2023) declared that Pakistan comes in the danger zone regarding extreme weather, and it may face many floods that can became the cause of heavy destruction in different areas of the country (Beniston, 2012). Main reasons of heavy floods and land sliding in Pakistan are Storms and occurrence of heavy monsoon rains (Bowen et al., 2021). In the view of National Disaster Management Authority (NDMA), one third of the people have been victimized by the floods since mid-June 2022. The Government of Pakistan considers 85 districts of the country as 'calamity-hit', after 17 October 2022 (Brath et al., 2002). Sindh province in Pakistan faces significant challenges that cause immense human suffering, disrupt livelihoods, and inflict substantial damage on infrastructure. Floods in Sindh are primarily caused by a confluence of factors. The first is the province's geographical location. Sindh lies in the lower Indus River basin, a vast and flat floodplain susceptible to overflowing rivers during periods of heavy monsoon rainfall (Brown et al., 2016).

Secondly, the melting of glaciers in the Himalayas, fed by the annual monsoon season, contributes significantly to the Indus River's water volume(Chai & Wu, 2023). These meltwaters, combined with heavy monsoon downpours, can overwhelm the river's capacity, leading to breaches in its embankments and widespread flooding across the floodplain. Uncontrolled deforestation in the catchment areas of the Indus River reduces the land's ability to absorb water, leading to increased surface runoff that contributes to flooding downstream (Chai & Wu, 2023). Furthermore, rapid urbanization within Sindh has resulted in the conversion of permeable land surfaces into impermeable ones, further reducing water infiltration and increasing flood risks in urban areas (Displacement Tracking Matrix, 2022). The potential impacts of floods in Sindh are far-reaching and devastating. Human life is directly threatened by floodwaters, with vulnerable populations, particularly those residing in low-lying areas, facing the greatest risk. Loss of life, injuries, and displacement are common consequences of major floods (Eleutério et al., 2013).

Livelihoods are also severely disrupted as agricultural land gets inundated, destroying crops and livestock. This can lead to food insecurity and economic hardship for affected communities (Fair et al., 2017). Infrastructure is another critical area impacted by floods. Transportation networks, communication lines, and power grids are often damaged or destroyed, hindering rescue efforts and hampering the recovery process (Gaillard & Mercer, 2013). Public health infrastructure can also be compromised, leading to outbreaks of waterborne diseases in the aftermath of floods (Ghosh et al., 2022).

Predicting future flood scenarios in Sindh requires a multifaceted approach. Hydrological modeling plays a critical role in this process. By incorporating historical data on precipitation, river discharge, and past flood events, hydrological models can simulate future flood scenarios under various climatic conditions (Haji Seyed et al., 2021). These models can be further refined by integrating data on land use changes, deforestation patterns, and potential variations in glacial meltwater volumes due to climate change (Hamidi et al., 2022). To forecast rainfall exactly, it is necessary to predict the greatest intensity, time period, and beginning time, that is not possible without the use of suitable models. To analyze natural disasters as abundant floods, prediction models have great importance. Statistical and computer-based algorithms as a part of machine learning perform major role in predicting and decision making on the basis of observed samples. Machine learning and other models are used efficiently in water resource management to predict the floods (Haq, et al., 2012), and water quality analysis (Hewitt, 2014). Analyzing flood risks in Sindh is essential for safeguarding the well-being of its population and protecting vital infrastructure. Understanding the geographical and environmental factors that contribute to flooding, coupled with the use of advanced modeling techniques and local knowledge, can help predict future flood scenarios. By assessing the potential impacts on people and infrastructure, proactive measures can be implemented to mitigate flood risks and build resilience within the province.

Therefore, this research deviates from prior studies by employing a multi-model approach, evaluating flood susceptibility through the application of distinct ML models. These models include Decision Trees, Logistic Regression, Gradient Boosting Classifier and Lasso Regression. By implementing this multi-model approach, the study seeks to achieve two primary objectives. First, it aims to visualize different features in the dataset according to most effective district and tehsil with respect to damage of people houses, infrastructure and people missing. Secondly, utilize most effective Machine Learning and Statistical Model among the mention employed to classify flood-induced building damage within the specific area. The results of the current study will be helpful to understand and predict flood damage by using hybrid techniques in the context of limited data availability. This knowledge can be instrumental in developing more efficient flood risk management strategies for vulnerable regions. For more explanation, the various studies covered in the literature review are followed by a discussion of the proposed technique, which includes applicable models. Next, the results section is written, and finally, the research conclusion section.

LITERATURE REVIEW

Floods are a recurrent natural hazard in Sindh province, Pakistan, posing a significant threat to people, infrastructure, and the overall economy. Several studies have explored various techniques for flood prediction in Sindh. Traditional methods rely on rainfall data and river discharge measurements to forecast flood events (ICHARM Report, 2009). Researchers have stressed the importance of integrating climate change projections into flood risk models to enhance preparedness and accuracy. Different works have explored the historical context of floods in Sindh, shedding light on the region's vulnerability and forming the basis for understanding current risks (Jamshed et al., 2020). Climate change is considered as a significant driver of flood dangers and its role in exacerbating floods (Khalid & Ali, 2020).

Advanced predictive modeling techniques have been employed to develop comprehensive flood risk models for Sindh, utilizing data from various sources such as hydrological data and satellite imagery to enhance prediction accuracy. Additionally, community-based adaptation strategies have been highlighting initiatives that enhance resilience and adaptive capacity (Khan et al., 2017). Hydrological modeling has been instrumental in mapping flood-prone areas and identifying vulnerable regions. The role of climate change in amplifying flood risks has been emphasizing the need for adaptive strategies and incorporating climate projections into risk assessments (Khan et al., 2017). The following subsections cover the numerous studies that have investigated the causes of severe flooding, flood prediction techniques based on statistics, computation, and artificial intelligence, and the effects of flooding on people, infrastructure, and animals in Pakistan as well as other parts of the world.

Causes of heavy flood in Pakistan

Monsoon system gives exigency situation to Pakistan every year and it becomes challenging to overcome this disaster's aftereffects. Over the past few years, this situation has become more challenging because of the lack of resources to store the existing water and their capability of absorption and to alleviate the high floods. As far as Sindh is concerned there is limited data on causes of flooding. However, several studies, particularly those following the 2010 floods in Pakistan, investigated the broader causes of these events (Kundzewicz et al., 2014). Haq, Mateeul, et al. concluded by their finding that weather events triggered, but weren't the sole cause, of the disaster (Louw et al., 2019). They used remote sensing and topographical data and found most damage results from factors like backwater effects from dams and barrages, decrease in water capacity sediment passage, and lack of water provision.

Sindh has two major river avulsions. Sukkur Barrage, a breach in the Tori Bund produced an avulsion, flooding around 8,000 square kilometers of cultivated area and diverting area of the Indus River 50-100 kilometers west of its usual passage. This break arose before two days of first flood at the Sukkur Barrage (Majeed et al., 2023). (Memon et al., n.d.) considered Global warming, the initial cause of flooding in Pakistan. Every year Pakistan faces dangerous results of climate change as heavy floods appearance. Furthermore, they indicated that elevated temperatures, glacial melting, intense monsoon precipitation, lack of government focus, and inadequate governance are the primary factors contributing to this significant flooding event.

High potential floods are considered the most threatening natural hazard to people and their socio-economic development. (Miyamoto et al., 2014) reported that from 1900- 2006 floods are accountable for more than 19% of mortality rate and effected people above 48%. Furthermore, 26% out of 72% economical damage from natural disasters is due to floods. (Mosavi et al., 2018) discuss the impact and consequence of global warming as severe flood in Pakistan and considered in the list of top ten affected countries. The research details the catastrophic consequences of the floods, affecting over 33 million people, destroying millions of homes, and causing billions in damages in terms of infrastructure.

Flood Prediction Methods

Flooding is the most frequent natural disaster that causes huge damage to agriculture, people life stock and infrastructure. (Mukherjee et al., 2014) highlighted the importance of utilizing satellite information as: topographic data, land use and rainfall to insufficiently gauged river basin. Authors demonstrated the Cagayan River basin in Philippines to predict flood effectively using IFAS (Integrated Flood Analysis System) model that is a distributed Hydrological model, developed by ICHARM (International Centre for Water Hazard and Risk Management). They also utilized GSMaP (Global Satellite Mapping of Precipitation) model to forecast the flood and found replicate river discharge with great accuracy (ICHARM Report, 2009). A predicted floods in district Jhelum, Punjab by utilizing geospatial model, Frequency Ratio Model, and Analytical Hierarchy process. Areas having extreme risk to moderate risk were classified by the usage of these methods (Munir et al., 2022). Hydrological models offer a more comprehensive approach by simulating the movement of water through the catchment area. These models consider factors like precipitation, soil moisture, and vegetation cover to predict flood peaks and inundation extent (Munir et al., 2022). By using approaches of Remote sensing, satellite images and Light Detection and Ranging (LiDAR) data valuable information is provided for flood mapping and real-time monitoring (Musarat et al., 2021). (Nanditha et al., 2023) explored the usefulness of Frequency Ratio (FR) model to find the vulnerability and susceptibility of flood using the frequency ratio (FR) model along the Mahananda River basin. Through the utilization of diverse parameters and machine learning algorithms, areas prone to flooding in the Mahananda River basin have been pinpointed. Furthermore, validation of the flood susceptibility model was conducted using the receiver operating characteristic curve (ROC). For identifying the flood's susceptibility used Frequency Ratio (FR) Model to analyze the flood-prone zones in Sindh and Punjab and exposed highly suspected areas that can be affected by potential flood by using flood inventory mapping. (Noji & Lee, 2005) considered Remote Sensing (RS) and Geographical Information System (GIS) useful to monitor and manage the flood disaster to obtained Moderate Resolution Imaging Spectroradiometer (MODIS) Aqua and Terra images during flooding incidents in Sindh province, Pakistan to alleviate the impact of cloud cover. They further analyzed floods in Pakistan in forthcoming years by using the Integrated Flood Analysis system (IFAS) model. This model is capable of simulating floods at any point in the basin.

(Otto et al., 2023) adopted an advanced digital machine learning approach i.e. automatic autoregressive method and provided the statistical data regarding any change in the Kabul River Sawat from 2011- 2030. Support Vector Machine (SVM) and Deep Learning method (Convolutional Neural Network CNN) applied (Qamer et al., 2023), to evaluate flood susceptibility in Egypt after 2016 flood. The results clarified that CNN provided high prediction accuracy i.e. 86% than SVM model that provided average performance on prediction i.e. 71.6%. To identify and predict the places of heavy flood occurrence is very difficult process because of unexpected change in climate and human activities. Machine learning techniques and other prediction models have simplified the process of predicting flood-prone areas that is necessary to manage flood dangers on or before time (Rizwan et al., 2023).

Assessing Impacts of Floods on People, livestock, and Infrastructure

Pakistan experiences increasingly frequent and intense extreme weather events. These events damage ecosystems, displace people, and harm settlements and infrastructure (Saeed et al., 2024). Floods, tropical cyclones, droughts, landslides, and earthquakes are just some of the natural hazards the country faces. The ND-Gain Index highlights this vulnerability, reported Pakistan as the country faces more impacts of climatic harshness and it also lacks in preparedness (Sardar, et al, 2008). Unfortunately, the poorest populations are hit the hardest. Their reliance on already strained resources like agriculture, livestock, and water makes them especially vulnerable to both catastrophe and the concequences of variation in climate. Floods have devastating social and economic consequences in Sindh. These events displace communities, disrupt livelihoods, and cause damage to critical infrastructure like roads, bridges, and communication networks.

Floods can also contaminate water sources and increase the risk of waterborne diseases. The recent floods are expected to significantly impact Pakistan's economic well-being. Real Gross Domestic Product (GDP) growth projections were at 4.5% for the year, but these figures will likely be revised downward due to the floods. Inflation rates, the size of the fiscal deficit, and the balance of payments are all expected to be negatively affected. (Shafi, 2010) discussed how people lost their of cattles, destruction of their property and infrastructure, and destruction of agricultural crops and land in Pakistan. (Shah et al., 2020) reported 70,238 km per square area was flooded occur in 2010 Pakistan. It affected 5.88 million people, in which the region comprises a road network spanning 1500 kilometers, railway tracks extending over 382 kilometers, forests covering an area of 498 square kilometers, and agricultural land totaling 16,440 square kilometers. Author considered Remote Sensing model efficient to analyze effect of flood damage. (Shehzad, 2023) conducted case study in Pakistan to evaluate the vulnerability of interconnected sectors and assess the repercussions of a disaster. They utilized the inoperability input-output model (IIM) to provide an approximate estimate of losses. The IIM enables the quantification of production impacts and the assessment of how disturbances affect the interconnected system by determining economic losses and inoperability levels. Additionally, PROMETHEE (preference ranking organization method for enrichment evaluations) with IIM, considering criteria such as economic system connections, financial losses, and percentage of inoperability.

This integration aids in analyzing the impact of disasters and ranking sectors based on predefined criteria. Most severe flood hit Pakistan in 2022, the reason behind this flood was heavy monsoon rain. Approximately 33 million individuals were affected by this situation. The agricultural losses in the highly productive Indus plains have made the country riskier. (Tompkins & Adger, 2004) used a loss and damage assessment method (L&D) to assess damage regarding crops. They adopted multi-sensor satellite data extracted from flood photos for flood areas, crops damage areas, and GPM (measurement of rainfall intensity. They reported that 18% of Sindh was flooded and 1.1 million of this was cropland. Crops were further damaged by heavy rainfall, floods, and due to less resources. Almost 57% of the area of crops was damaged. They also mentioned the loss in the production that is 88% of cotton, 80% of rice, and 61% in sugarcane. (Triana et al., 2013) record real-time images of assets, cattle, societies, fields and other damages occur during flood in Pakistan August 2022 in the province of Punjab. They used a machine application for flood survey. This application gave them opportunity to analyze the real condition of damage in people's possessions, fields and cattle because of flood.

The random forest algorithm was used to predict flooded- boundary through machine learning. (Urlainis et al., 2022) considered networks as a collection of elements with unique purposes to propose a new way to assess how floods might damage critical infrastructure networks like electricity, water, and transportation. Authors focus is to analyze that how different components are damaged and lead dysfunction of entire network and area along with considering the resources and actions needed to get the network back up and running after a flood, making it valuable for planning resilience strategies.

Table1.
Techniques and strategies in flood management

Reference	Data	Developed System
Mukherjee et al.	Cagayan River basin	Predict Flood using IFAS
Munir et al.	Jhelum, Punjab	Classify extreme to moderate risk district flood
Otto et al.	Kabul River Sawat	Autoregressive ML model for any change
Qamer et al.	Egypt Flood	SVM and CNN evaluate flood susceptibility

METHODOLOGY

The study focuses on the Sindh province of Pakistan, a region highly susceptible to flooding due to its geographic, climatic, and socio-economic characteristics.

Geographically, Sindh is in the southeastern part of Pakistan and is bordered by the Arabian Sea to the south. The province is characterized by a diverse landscape, including river systems, deltas, and arid plains. Climatically, Sindh experiences a range of weather conditions, including monsoon rains and extreme temperatures. Figure 1 shows the complete working flow of our proposed model. First, we analyzed the selected dataset with the help of exploratory data analysis and data processing, and after that, different methods were applied for predicting Tehsil.

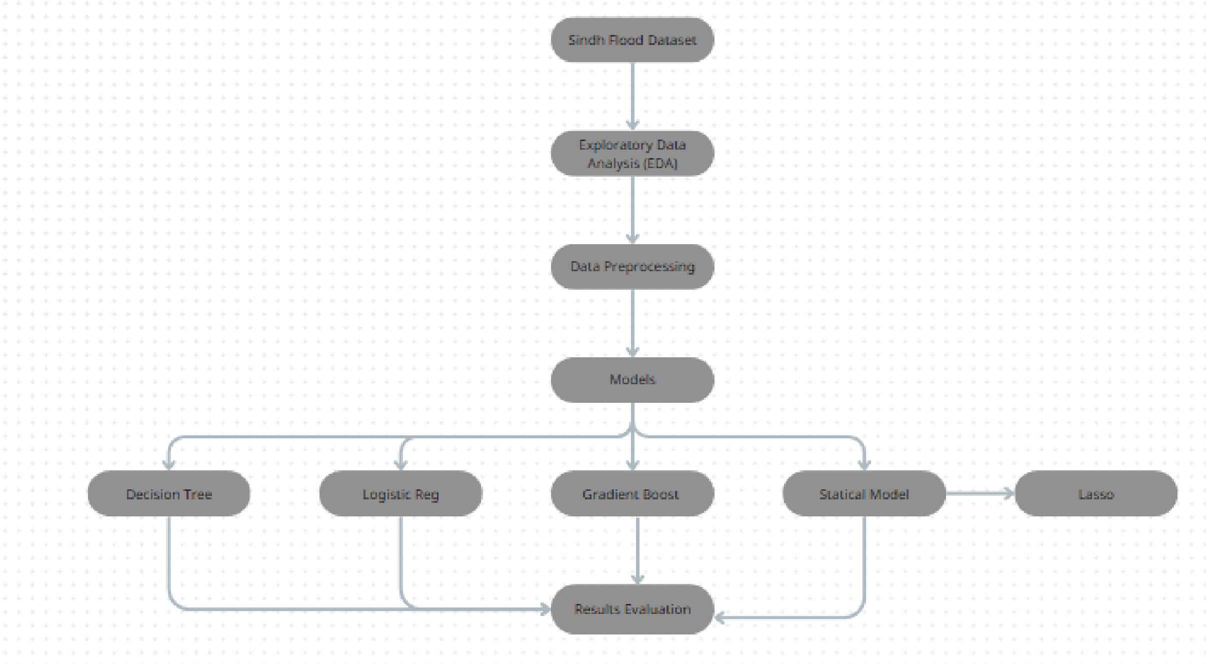

Figure1.
Proposed Model Flow

About Dataset

The dataset was chosen from the 'Sindh Flood Data Analysis & Prediction' Kaggle project. Following a strategic and detailed approach that prioritized the inclusion of demographic and geographical variables. This dataset offers a comprehensive view of the impacts and consequences of floods on individuals and regions within Pakistan's Sindh province. The demographic information like age, gender, and family structures shed light on the human aspects affected by floods. Additionally, geographical variables such as divisions, districts, tehsils, and villages offer spatial context, enriching the dataset with detailed regional information. The deliberate selection of this dataset stems from its thoroughness and relevance to the research goal of predicting displacement outcomes. Python programming language is for analysis of complete data pattern i.e. missing values, size of gathered data of district and tehsil. Further machine learning algorithms such as Decision Trees, Gradient Boosting Classifiers, Logistic Regression, and Lasso Regression were selected based on their ability to handle diverse data types and complex relationships within the dataset.

Challenges and Limitation of Dataset

The sets of data are arranged into three subsets: training data, testing data, and validation data, each containing 15 features. The training dataset, with dimensions (2021, 15), is used for training machine learning models, while the testing dataset, with dimensions (253, 15), evaluates the model's performance independently. The validation dataset, also with dimensions (253, 15), aids in fine-tuning model parameters and preventing overfitting during development. The dataset is not in proper normalized form there are several missing values were imputed with 'NONE' to maintain dataset integrity. There is no labeling of columns, neither categorical nor numerical formats. To address these challenges missing values is adjusted through strategic imputation to avoid bias. Imbalanced distributions were managed by adjusting model parameters and employing various evaluation metrics to ensure a robust analysis. Additionally, some data quality issues were resolved by cross-referencing with supplementary sources and conducting thorough data validation. Futher the validity and reliability of the data is tested with different tests with model performance measured using Mean Squared Error, Recall, F1 Score and accuracy. Due to limited availability of historical flood data, particularly at the tehsil level, Inconsistent data collection methods, missing values, and errors in recorded data can significantly impact the model's performance.

```
Index(['Province', 'Division', 'District', 'Tehsil', 'Village',
       'What resources and assets were lost and/or damaged during displacement by the majority of the TDPs?',
       'Number of shelters in this current village Number of shelters in TDPs? place of origin that are partially damaged?',
       'Number of TDP households living in buildings ? formal relief sites (i.e. schools, public buildings)',
       'Number of TDP returnees (individuals) and (households)',
       'breakdown_TDPs_infants_children_male',
       'breakdown_TDPs_youth_adults_male', 'breakdown_TDPs_elder_male',
       'breakdown_TDPs_infants_children_female',
       'breakdown_TDPs_youth_adults_female', 'breakdown_TDPs_elder_female'],
      dtype='object')
```

Figure2.
Selected Columns of Dataset

Exploratory Data Analysis (EDA)

Exploratory Data Analysis (EDA) is the first step in grasping the complexities of the dataset. By employing various visualizations and statistical evaluations, we garnered valuable insights into the socio-economic and demographic factors linked to flood-induced displacement in the Sindh region. Below, we outline the key findings from our EDA analysis.

Categorical Variables Analysis

Figures 3, 4, 5, and 6 are generated histograms for categorical variables like 'Division,' 'District,' 'Tehsil,' and 'Village,' offering insights into their distributions. These visualizations helped in the geographical and administrative spread of the dataset.

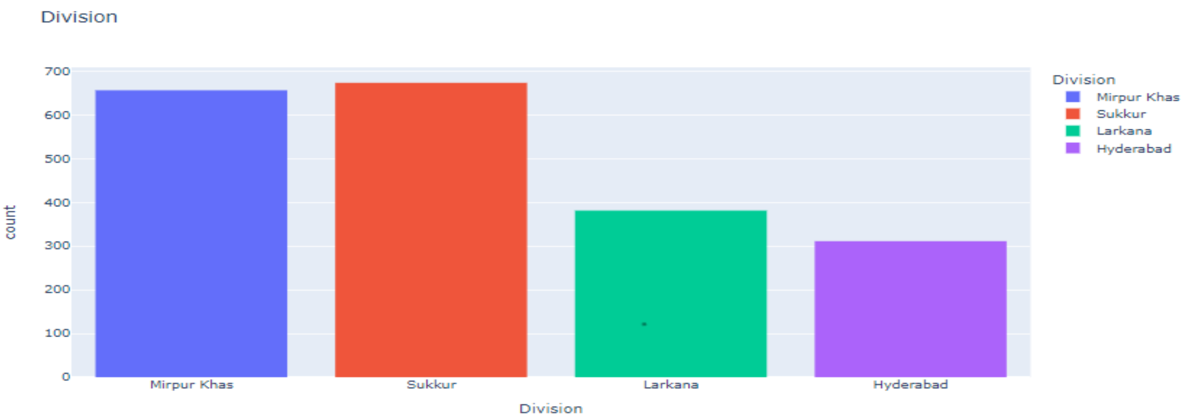

Figure3.
Categorical Vriable Analysis Division Based

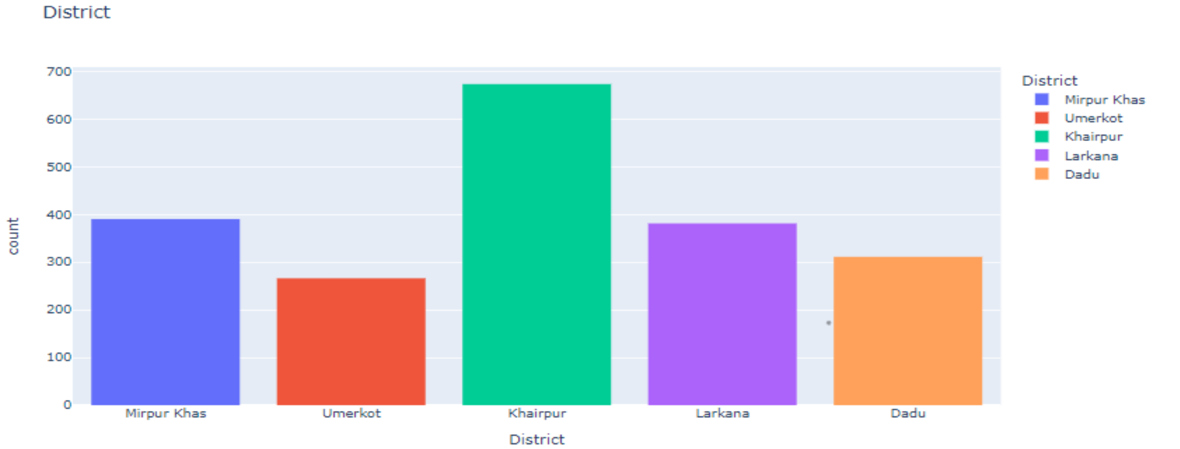

Figure4.
Categorical Vriable Analysis District Based

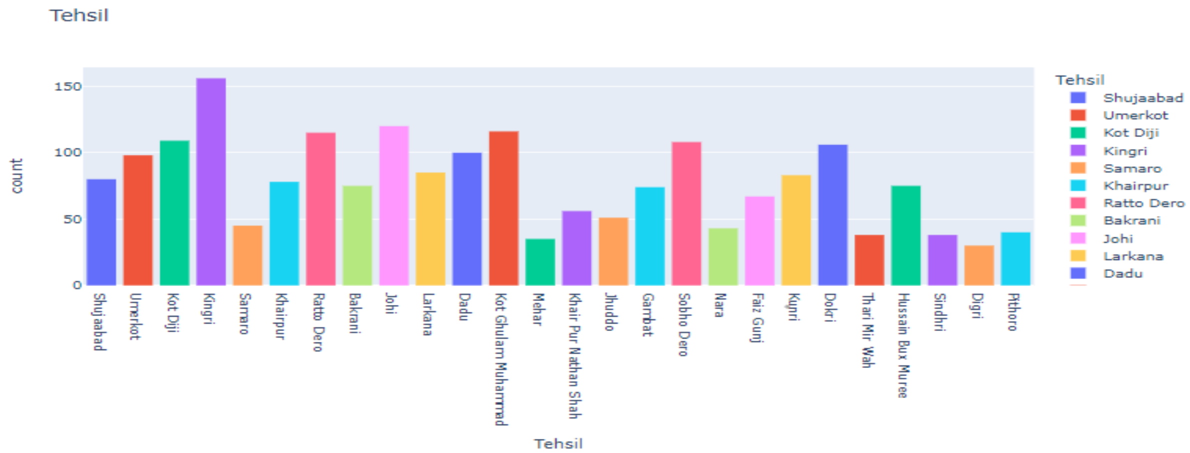

Figure5.
Categorical Vriable Analysis Tesil Based

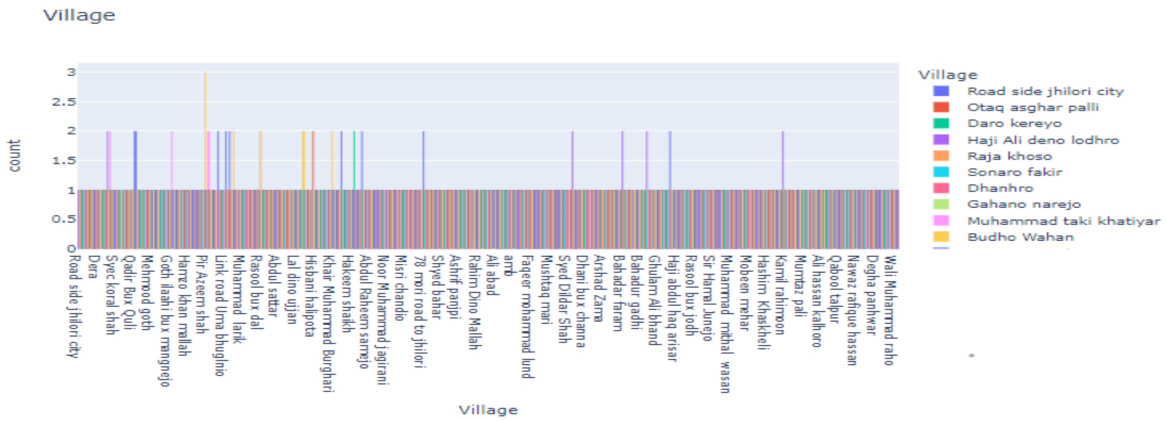

Figure6.
Categorical Vriable Analysis Village Based

Frequency Analysis of Displacement Factors

Histograms were utilized to examine the frequency of factors contributing to displacement, such as resources and assets lost during most displacements. This frequency analysis shed light on common themes in the impact of floods on individuals and communities Figure7.

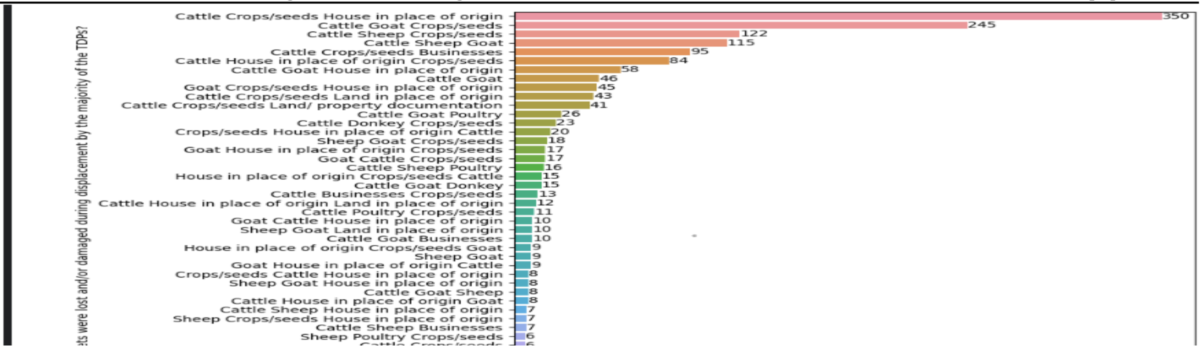

Figure7.
Frequency Analysis of Displacement

Data Preprocessing

During the data preprocessing phase, we employed a careful blend of techniques and methods, each meticulously tailored to tackle specific challenges and improve the overall quality of the dataset. This section provides a more in-depth exploration of the techniques utilized.

Handling Missing Values

Figure8, and 9 addressing missing values poses a common challenge in real-world datasets, and our approach was both systematic and strategic. We conducted a thorough analysis to pinpoint columns with a notable percentage of missing values. The chosen method is strategically imputing missing values with the label 'NONE.' This approach was selected to maintain the dataset's integrity by distinguishing cases where certain information was genuinely absent or not applicable. Such strategic imputation ensures that subsequent analyses are not disproportionately influenced by the absence of data points, thus guarding against potential bias in the results.

```
Province                                                                                          0
Division                                                                                          0
District                                                                                          0
Tehsil                                                                                            0
Village                                                                                           0
What resources and assets were lost and/or damaged during displacement by the majority of the TDPs?        141
Number of shelters in this current village Number of shelters in TDPs? place of origin that are partially damaged?     0
Number of TDP households living in buildings ? formal relief sites (i.e. schools, public buildings)     1231
Number of TDP returnees (individuals) and (households)                                         1203
breakdown_TDPs_infants_children_male                                                            956
breakdown_TDPs_youth_adults_male                                                                956
breakdown_TDPs_elder_male                                                                       956
breakdown_TDPs_infants_children_female                                                          956
breakdown_TDPs_youth_adults_female                                                              956
breakdown_TDPs_elder_female                                                                     956
dtype: int64
```

Figure 8.
Missing Values from Dataset

```
Province                                                                                          0
Division                                                                                          0
District                                                                                          0
Tehsil                                                                                            0
Village                                                                                           0
What resources and assets were lost and/or damaged during displacement by the majority of the TDPs?        0
Number of shelters in this current village Number of shelters in TDPs? place of origin that are partially damaged?     0
Number of TDP households living in buildings ? formal relief sites (i.e. schools, public buildings)     0
Number of TDP returnees (individuals) and (households)                                            0
breakdown_TDPs_infants_children_male                                                              0
breakdown_TDPs_youth_adults_male                                                                  0
breakdown_TDPs_elder_male                                                                         0
breakdown_TDPs_infants_children_female                                                            0
breakdown_TDPs_youth_adults_female                                                                0
breakdown_TDPs_elder_female                                                                       0
dtype: int64
```

Figure 9.
Handling NULL Values from Dataset

Label Encoding for Categorical Features

Transforming categorical features, which are qualitative, into numerical representations was necessary for their smooth integration into machine learning models. We utilized the technique of Label Encoding, that assign a numerical code for all the special categories within a categorical variable. This transformation not only ensured compatibility with various machine learning algorithms but also maintained the inherent ordinal relationships between categories, where relevant. This careful encoding process ensures that the qualitative details captured by categorical features are accurately translated into a format suitable for rigorous quantitative analysis Figure 10.

	Province	Division	District	Tehsil	Village	What resources and assets were lost and/or damaged during displacement by the majority of the TDPs?	Number of shelters in this current village Number of shelters in TDPs? place of origin that are partially damaged?	Number of TDP households living in buildings ? formal relief sites (i.e. schools, public buildings)	Number of TDP returnees (individuals) and (households)	breakdown_TDPs_infants_children_male	breakdown_TDPs_y
0	0	2	3	21	1643	30	20	408	330	202	
1	0	2	4	25	1464	30	160	408	71	202	
2	0	3	1	12	388	27	13	133	330	164	
3	0	3	1	11	674	57	10	408	330	47	
4	0	2	4	20	1587	191	32	408	226	18	
...	...	...	...	...	...	...	...	...	...	...	
2016	0	3	1	5	1412	28	60	314	142	202	
2017	0	3	1	11	901	57	0	408	251	12	
2018	0	3	1	4	1348	148	50	408	330	202	
2019	0	2	3	13	358	57	17	339	215	35	
2020	0	3	1	5	777	29	160	408	80	201	

2021 rows × 15 columns

Figure 10.
Shows Label Encoding On Categorical Data

Selected Models Description

The Model Development phase stands as the core of the research effort, wherein advanced machine learning algorithms and statistical models are utilized to extract patterns, relationships, and predictive insights from the carefully preprocessed dataset. This section offers a detailed examination of the array of models employed, elucidating their unique strengths, intricacies, and implications.

Machine Learning Models

Machine Learning Models have effective role in enabling computers to make decisions and predictions on the basis of data sets. These models are built by training algorithms on data to recognize patterns and relationships. Once trained, the models can be used to predict outcomes on new, unseen data. The description of selected Machine Learning algorithms are mentioned below.

Decision Tree Model

The Decision Tree model, revered for its interpretability and capacity to capture, and A Strategic choice in the model ensemble. Trained on the preprocessed dataset, this model embarked on a journey through the data's intricate branching points, making decisions based on feature thresholds to predict displacement outcomes. The model's interpretability is invaluable, offering insights into the factors deemed most influential in predicting 'Tehsil' outcomes. The calculated accuracies for 'Tehsil' provided a baseline understanding of the model's performance, serving as a reference point for more complex models in the ensemble. Following are the mathamatical formula of algorithm.

$I_G(p) = 1 - \sum_{i=1}^{J} p_i^2$ Equation (1)

$H(p) = -\sum_{i=1}^{J} p_i \log_2(p_i)$ Equation (2)

$IG(T, a) = H(T) - \sum_{v \in Values(a)} \frac{|T_v|}{|T|} H(T_v)$ Equation (3)

Where:

- (IG(T, a)) is the information gain by splitting tree (T) on attribute (a).
- (H(T)) is the entropy of the entire set (T).
- (Values(a)) is the set of all possible values for attribute (a).
- (T_v) is the subset of (T) for which attribute (a) has value (v).
- (|T_v|/|T|) is the weighted proportion of the number of elements in (T_v) to the number of elements in (T).

Gradient Boosting Classifier

Gradient Boosting Classifier is a machine learning technique for regression and classification problems, which produces a prediction model in the form of an ensemble of weak prediction models, typically decision trees. Following is the mathamatical formula of algorithm.

$F_m(x) = F_{m-1}(x) + \gamma_m h_m(x)$ Equation (1)

Where:

- (F_m(x)) is the boosted model at step (m).
- (F_{m-1}(x)) is the boosted model from the previous step.
- (\gamma_m) is the learning rate, which scales the contribution of each tree.
- (h_m(x)) is the weak learner (decision tree) fitted on the negative gradient.

Logistic Regression

Logistic Regression in the context of machine learning is used for classification problems, particularly binary classification. It works by estimating the probability that a given input point belongs to a certain class. Following are the mathamatical formula of algorithm.

The logistic function, also known as the sigmoid function, is defined as:

$\sigma(z) = \frac{1}{1+e^{-z}}$ Equation (1)

Where:

(z) is the weighted sum of the input features (z = w_0 + w_1x_1 + w_2x_2 + … + w_nx_n), and (w_i) are the model parameters.

The probability that an instance belongs to the positive class is then modeled as:

$P(Y = 1 \mid X = x) = \sigma(w_0 + w_1x_1 + w_2x_2 + \ldots + w_nx_n)$

Equation (2)

If this probability is greater than or equal to 0.5, we predict that the instance belongs to the positive class; otherwise, it belongs to the negative class.

Statistical Models

Statistical models in machine learning are algorithms that are based on statistical theories to discover patterns and make predictions from data. They often involve estimating parameters from the data that can be used to predict outcomes on new, unseen data. Common statistical models include linear regression, logistic regression, and various types of regularized regression such as Lasso Regression.

Lasso Regression

Lasso Regression, is a kind of linear regression that shrinks the values of data towards its mean. This process stimulates the models which have fewer parameters. Following is the mathamatical formula of algorithm.

$$\min_{w} \{\frac{1}{2n} \| Xw - y \|_2^2 + \alpha \| w \|_1\} \quad \text{Equation (1)}$$

Where:

(w) represents the coefficient vector.

(X) is the matrix of input features.

(y) is the vector of output values.

(||Xw - y||_2^2) is the sum of squared residuals.

(||w||_1) is the L1-norm of the coefficient vector (sum of absolute values).

(\alpha) is a tuning parameter that controls the strength of the penalty.

RESULTS AND DISCUSSION

The evaluation of our predictive models for 'Tehsil' outcomes in the context of flood-induced displacement in the Sindh region yielded diverse and insightful results. Here, we present a comprehensive overview of the key findings derived from each modeling methodology.

Decision Tree Model

The Decision Tree model demonstrated a moderate level of accuracy, successfully capturing essential patterns in 'Tehsil' predictions. Its transparent decision-making paths provided valuable insights into influential factors, making it a suitable tool for initial exploration. After generating a training dataset with 1,000 samples and a testing dataset with 200 samples, the Decision Tree Classifier is trained on features such as `Province`, `Division`, `District`, `Village`, and various demographic breakdowns of TDPs (Temporarily Displaced Persons). Predictions made on the test set were evaluated using precision, recall, and F1 score metrics. The precision score was 0.1946, indicating a relatively low rate of correctly predicted positive observations out of the total predicted positives. The recall score was 0.2016, suggesting a modest ability to capture actual positive observations. The F1 score, which balances precision and recall, was 0.1525, reflecting an overall moderate performance of the model.

Additionally, a confusion matrix was generated and visualized using a heatmap, providing a detailed view of the true vs. predicted classifications. The results highlight the Decision Tree model's current limitations and provide a basis for further model optimization.

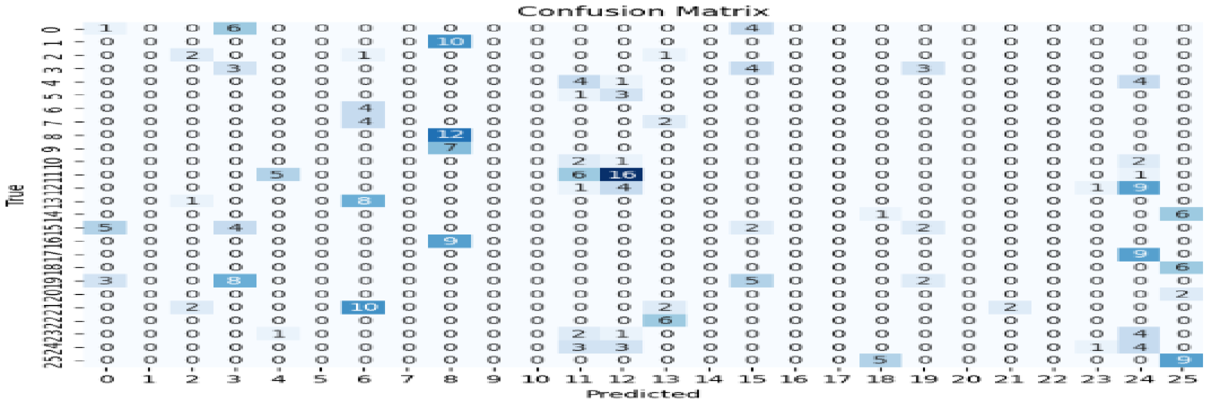

Figure11.
Confusion Matrix of Decsion Tree

Precision: 0.474876896794292
Recall: 0.475
F1 Score: 0.4743560862056019

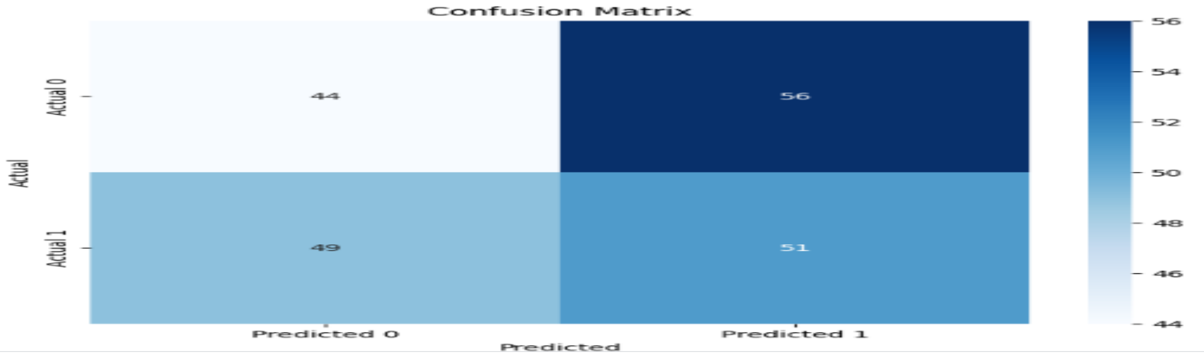

Figure12.
Confusion Matrix of Unique Test

Gradient Boosting Classifier

The Gradient Boosting Classifier emerged as the standout performer, showcasing exceptional accuracy on both the training and test datasets. While the model's iterative refinement captured intricate patterns, its interpretability was limited, emphasizing its strength in predictive accuracy. With a staggering 97.67% accuracy on the training data and a flawless 100% accuracy on the test data, this model emerged as a frontrunner in the predictive modeling ensemble.

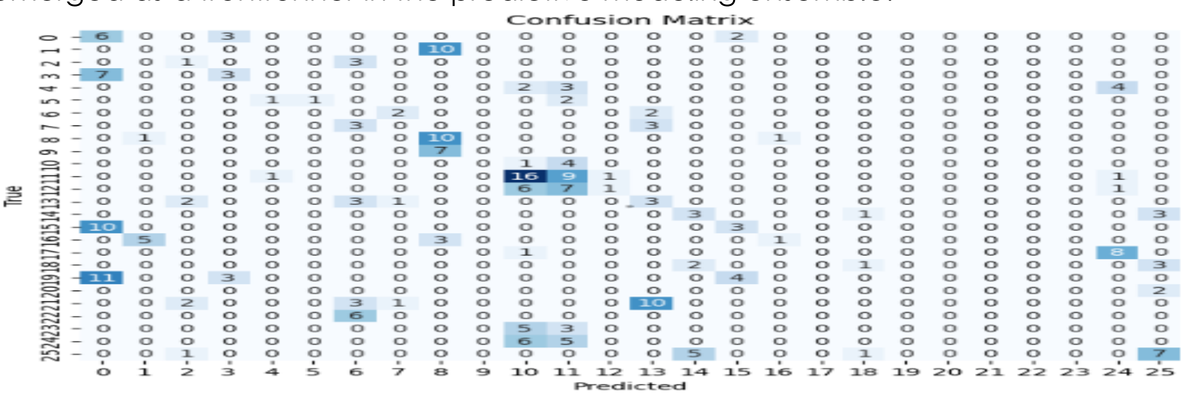

Figure13.
Confusion Matrix of Gradient Boosting Classifier

The Gradient Boosting Classifier (GBC) algorithm trained on random data demonstrated superior predictive accuracy, yielding precision, recall, and F1 scores of 0.549, 0.55, and 0.548, respectively. The confusion matrix depicted a well-balanced classification performance, with comparable distributions of true positives and true negatives. This suggests that the GBC algorithm effectively leverages the random data to make accurate predictions, outperforming both the pretrained Lasso model and the Lasso model trained on random data in figure14.

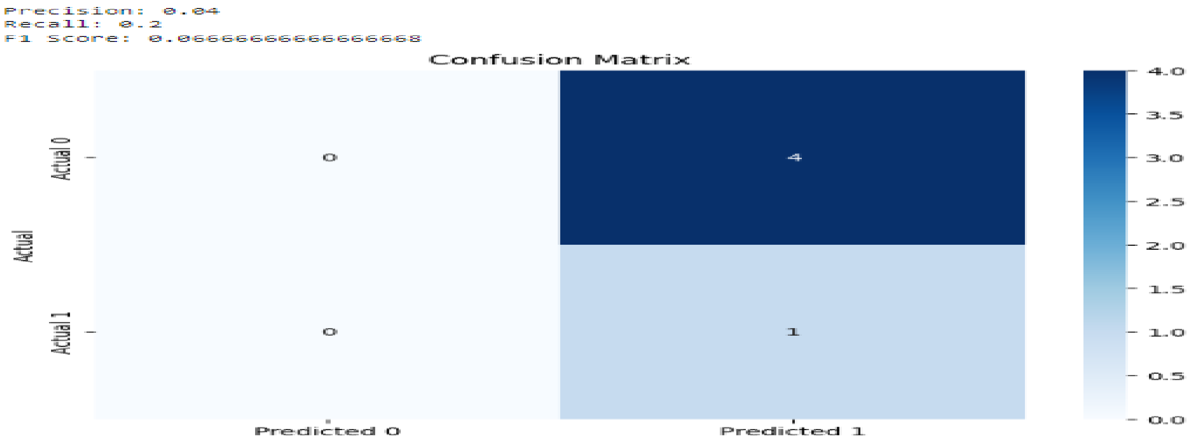

Figure14.
Confusion Matrix of Unique Test

Logistic Regression Model

The accuracy scores revealed a nuanced performance, with a 16.60% accuracy on the test data, 20.78% on the training data, and a perfect 90% accuracy when predictions were applied to the test data. The Logistic Regression model's unique contribution lies in its ability to provide a baseline understanding of the dataset's linear relationships and offer a simplified interpretation of feature importance. The Heatmap visualization of the confusion matrix, where the x-axis represents the predicted labels and the y-axis represents the true labels. Each cell in the heatmap represents the number of predictions made for each combination of true and predicted labels. Figure 15, Logistic Regression model showed lower accuracy, highlighting its limitations in capturing non-linear relationships inherent in 'Tehsil' predictions. Its simplicity and interpretability made it a valuable tool for diagnosing linear relationships within the dataset shows in figure16.

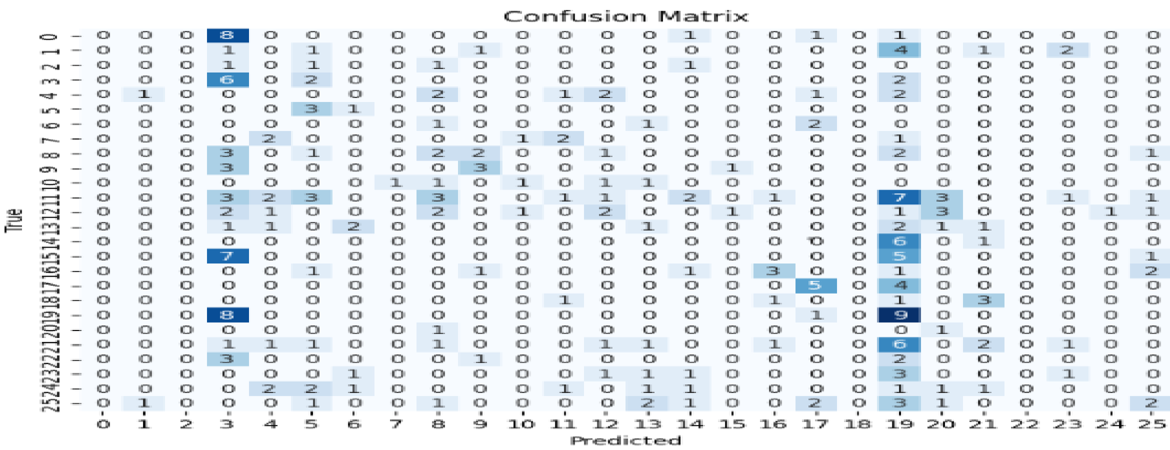

Figure15.
Confusion Matrix Logistic Regression Model

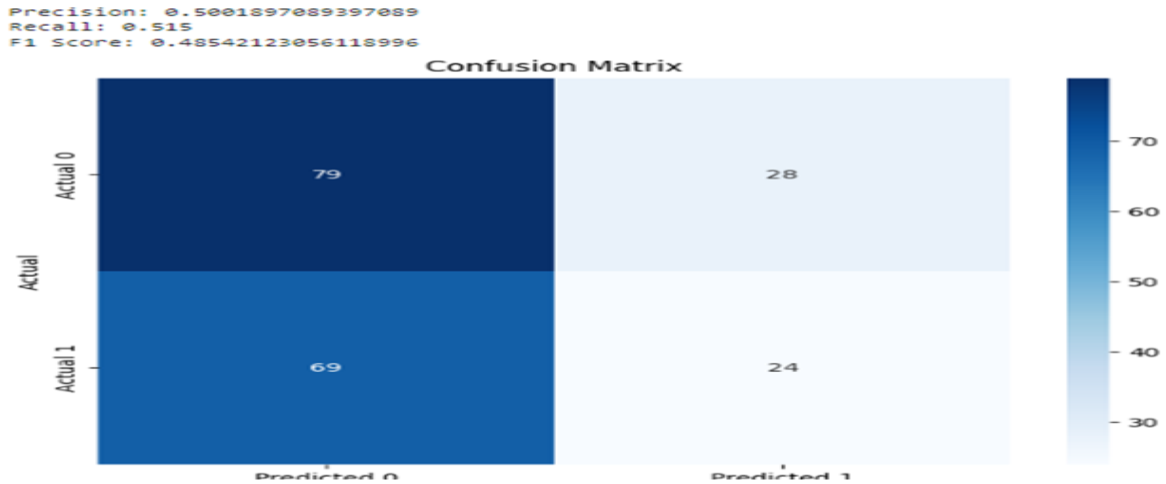

Figure16.
Confusion Matrix of unique test Logistic Regression Model

Lasso Regression

The fitting of the Lasso Regression model to the training data involves a meticulous examination of the relationship between the selected features and the target variable, 'Tehsil.' The model's fit_regularized method, incorporating L1 regularization, encapsulates the essence of the Lasso technique. As the regularization process unfolds, certain coefficients are strategically shrunk to zero, highlighting features deemed less influential in predicting 'Tehsil' outcomes. The coefficients derived from the Lasso Regression analysis serve as a compass for interpreting feature importance. Each coefficient quantifies the impact of a respective feature on the predicted 'Tehsil' values. By inspecting these coefficients, researchers gain valuable insights into the socio-economic and demographic variables that play a pivotal role in shaping displacement predictions. Figure 17 shows the Lasso Regression analysis yielded modest accuracy metrics on both the validation and test datasets. The feature selection capability enhanced interpretability by quantifying the impact of influential factors. The Heatmap visualization of the confusion matrix, where the x-axis represents the predicted labels and the y-axis represents the true labels. Each cell in the heatmap represents the number of predictions made for each combination of true and predicted labels.

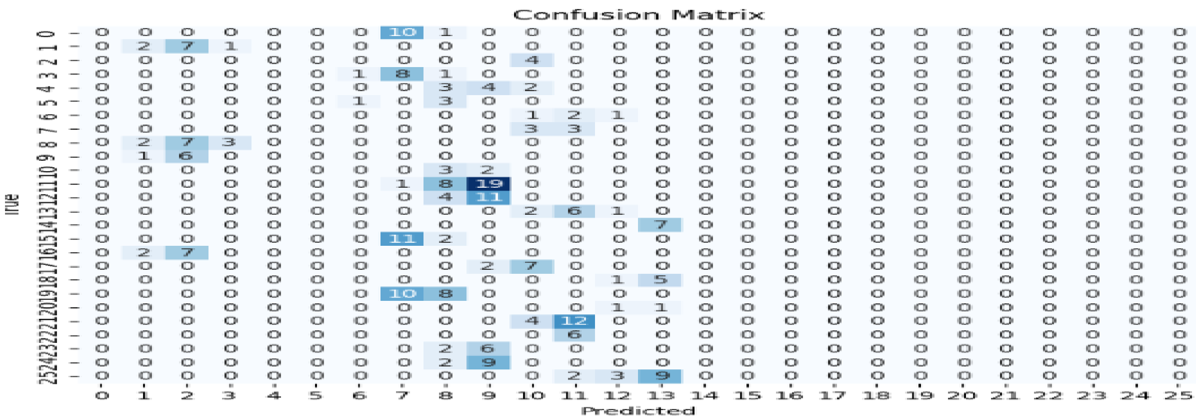

Figure17.
Confusion Matrix Lasso Regression

Training a Lasso model on randomly generated data showcased slightly improved predictive performance, with precision, recall, and F1 scores of 0.50, 0.515, and 0.485, respectively. The confusion matrix revealed a more balanced distribution of true positives and true negatives, indicating a moderate level of classification accuracy. While the model showed an enhancement in predictive capabilities compared to the pretrained Lasso model with dummy data, there is still room for improvement in achieving higher precision and recall scores shown in figure18.

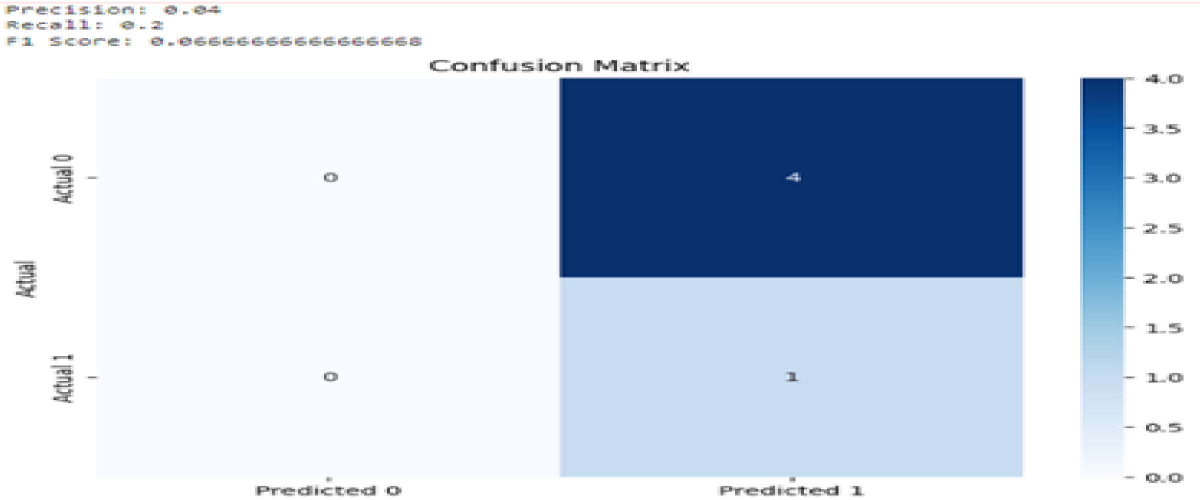

Figure18.
Confusion Matrix of Unique test Lasso Regression

RESULTS ANALYSIS

Gradient Boosting Dominance: The unparalleled accuracy of the Gradient Boosting Classifier positions it as a robust tool for predicting 'Tehsil' outcomes. Its adaptability to complex relationships makes it a pivotal asset for real-world applications. The Decision Tree model: while not achieving the highest accuracy, contributed foundational insights through its transparent decision-making process. It serves as an excellent starting point for understanding the dataset's nuances. The recall 0.03 and on other the F1 score is 0.39. Logistic Regression Diagnostics: The Logistic Regression model, with its simplicity and interpretability, played a crucial role in diagnosing linear relationships. While not a top performer in accuracy, it provided valuable insights into simpler patterns. The recall score is 0.169 and on other side F1 score is 0.133. Lasso Regression: though modest in accuracy, enriched our understanding by quantifying feature impact. Its feature selection capability added granularity to the interpretative landscape. And the recall score is 0.102 and on other side F1 score is 0.079.

Table2.
Models Performance

ALGORITHM	MSE	RECALL	F1 SCORE	Accuracy Score
Decision Tree Model	78.25	0.245	0.1600	0.25
Gradient Boosting Classifier	82.20	0.20	0.167	0.97.67
Logistic Regression	81.27	0.169	0.133	0.95
Lasso Regression (Statsmodel)	63.01	0.102	0.079	0.55

The findings of this study have significant implications for local and regional disaster management agencies, particularly concerning the enhancement of early warning systems and overall flood preparedness:

- **Enhanced Early Warning Systems**: The high accuracy of the Gradient Boosting Classifier in predicting flood impacts demonstrates the value of advanced predictive models in improving early warning systems. By integrating such models, disaster management agencies can provide more accurate and timely warnings, enabling communities to take proactive measures and reduce flood-related damages.

- **Improved Predictive Accuracy**: The study highlights the need for incorporating machine learning models that account for complex patterns in flood data. Agencies should consider adopting advanced predictive tools to refine their forecasting capabilities, thus enhancing their ability to anticipate and respond to flood events effectively.

- **Targeted Disaster Response**: The study's findings on the effectiveness of different models for identifying affected areas (Tehsils) can help agencies prioritize resource allocation and response efforts. Understanding which areas are most vulnerable enables better-targeted interventions, reducing the overall impact of floods on communities.

CONCLUSION

This study has examined various aspects of flood risk evaluation in Sindh, Pakistan. Improved prediction methods are crucial for issuing timely warnings and preparing for future flood events. By integrating machine learning algorithms with statistical models, the framework offers a comprehensive tool that not only predicts potential impacts with high accuracy but also provides actionable insights for mitigation strategies. The model's success in identifying vulnerable areas within the Sindh region underscores its potential for application in other regions prone to extreme weather events. In future, government and non-governmental organizations, including flood management cells, should prioritize the enhancement of early warning systems by developing and integrating advanced predictive technologies over real-time data and advanced analytics.

DECLARATIONS

Acknowledgement: We appreciate the generous support from all the contributor of research and their different affiliations.

Funding: No funding body in the public, private, or nonprofit sectors provided a particular grant for this research.

Availability of data and material: In the approach, the data sources for the variables are stated.

Authors' contributions: Each author participated equally to the creation of this work.

Conflicts of Interests: The authors declare no conflict of interest.

Consent to Participate: Yes

Consent for publication and Ethical approval: Because this study does not include human or animal data, ethical approval is not required for publication. All authors have given their consent.

REFERENCES

Abbas, H. W., & Guo, X. (2023). Climate-related vulnerability assessment toward disaster risk reduction: Insight from Pakistan. Journal of Homeland Security and Emergency Management, 20(3), 307-350. doi: 10.1515/jhsem-2022-0092

Aggarwal, A. (2016). Exposure, hazard and risk mapping during a flood event using open source geospatial technology. Geomatics, Natural Hazards and Risk, 7(4), 1426-1441. doi: 10.1080/19438183.2015.1053534

Ajani, A., & van der Geest, K. (2021). Climate change in rural Pakistan: Evidence and experiences from a people-centered perspective. Sustainability Science, 16(6), 1999-

2011. doi: 10.1007/s11663-021-00983-z
Amarnath, G., & Rajah, A. (2016). An evaluation of flood inundation mapping from MODIS and ALOS satellites for Pakistan. Geomatics, Natural Hazards and Risk, 7(5), 1526-1537. doi: 10.1080/19438183.2015.1075302
Arshad, R. R., & Shafi, S. (2010). Pakistan floods 2010: Preliminary damage and needs assessment. Asian Development Bank and World Bank.
Arshed, N., Saeed, M. I., Salem, S., Hanif, U., & Abbas, M. (2023). National strategy for climate change adaptability: A case study of extreme climate-vulnerable countries. Environment, Development and Sustainability. doi: 10.1007/s10668-023-09988-6
Beniston, M. (2012). Impacts of climatic change on water and associated economic activities in the Swiss Alps. Journal of Hydrology, 412, 291-296. doi: 10.1016/j.jhydrol.2011.11.075
Bowen, S., Elliott, S., & Hardison-Moody, A. (2021). The structural roots of food insecurity: How racism is a fundamental cause of food insecurity. Sociology Compass, 15(7), e12846. doi: 10.1111/socc.12846
Brath, A., Montanari, A., & Toth, E. (2002). Neural networks and non-parametric methods for improving real-time flood forecasting through conceptual hydrological models. Hydrology and Earth System Sciences, 6(4), 627-639. doi: 10.5194/hess-6-627-2002
Brown, K. M., Hambidge, C. H., & Brownett, J. M. (2016). Progress in operational flood mapping using satellite synthetic aperture radar (SAR) and airborne light detection and ranging (LiDAR) data. Progress in Physical Geography, 40(2), 196-214. doi:10.1177/0309133315595021
Chai, J., & Wu, H.-Z. (2023). Prevention/mitigation of natural disasters in urban areas. Smart Construction and Sustainable Cities, 1(1), 4. doi: 10.1007/s42091-023-00005-9
Chapman, S., Watson, J. E. M., Salazar, A., Thatcher, M., & McAlpine, C. A. (2017). The impact of urbanization and climate change on urban temperatures: A systematic review. Landscape Ecology, 32, 19
Chapman, S., Watson, J. E. M., Salazar, A., Thatcher, M., & McAlpine, C. A. (2017). The impact of urbanization and climate change on urban temperatures: A systematic review. Landscape Ecology, 32, 1921-1935. doi: 10.1007/s10980-017-0578-y
Eleutério, J., Hattemer, C., & Rozan, A. (2013). A systemic method for evaluating the potential impacts of floods on network infrastructures. Natural Hazards and Earth System Sciences, 13(4), 983-998. doi: 10.5194/nhess-13-983-2013
Fair, C. C., Kuhn, P., Malhotra, N. A., & Shapiro, J. (2017). Natural disasters and political engagement: Evidence from the 2010–11 Pakistani floods.
Gaillard, J. C., & Mercer, J. (2013). From knowledge to action: Bridging gaps in disaster risk reduction. Progress in human geography, 37(1), 93-114. doi: 10.1177/0309132512459828
Ghosh, A., Biswas Roy, M., & Kumar Roy, P. (2022). Flood susceptibility mapping using the frequency ratio (FR) model in the Mahananda River Basin, West Bengal, India. In India II: Climate change impacts, mitigation, and adaptation in developing countries (pp. 73-96). Springer International Publishing.
Haji Seyed, S. B., Sharafati, A., Motta, D., & Yaseen, Z. M. (2021). River water quality index prediction and uncertainty analysis: A comparative study of machine learning models. Journal of environmental chemical engineering, 9(1), 104599. doi: 10.1016/j.jece.2020.104599
Hamidi, A. R., Jing, L., Shahab, M., Azam, K., Tariq, M. A. U. R., & Ng, A. W. M. (2022). Flood exposure and social vulnerability analysis in rural areas of developing countries: An empirical study of Charsadda District, Pakistan. Water, 14(7), 1176. doi: 10.3390/w14071176
Haq, M., Akhtar, M., Muhammad, S., Siddiqi, P., & Jillani, R. (2012). Techniques of remote sensing and GIS for flood monitoring and damage assessment: A case study of Sindh province, Pakistan. The Egyptian Journal of Remote Sensing and Space Science, 15(2), 135-141.
Hewitt, K. (2014). Hazards to water resources development in high mountain regions. The Himalayan sources of the Indus. In Hydrology of Disasters (pp. 294-312). Routledge.
ICHARM Report. (2009). Global trends in water related disasters: An insight for policymakers. International Centre for Water Hazard and Risk Management (UNESCO).

Jamshed, A., Birkmann, J., McMillan, J. M., Rana, I. A., & Lauer, H. (2020). The impact of extreme floods on rural communities: Evidence from Pakistan. In Climate Change, Hazards and Adaptation Options: Handling the Impacts of a Changing Climate (pp. 585-613).

Khalid, M. A., & Ali, Y. (2020). Economic impact assessment of natural disaster with multi-criteria decision making for interdependent infrastructures. Environment, development and sustainability, 22, 7287-7311. doi: 10.1007/s10668-019-00361-6

Khan, I., Lei, H., Shah, A. A., Khan, I., & Muhammad, I. (2021). Climate change impact assessment, flood management, and mitigation strategies in Pakistan for sustainable future. Environmental Science and Pollution Research, 28, 29720-29731. doi: 10.1007/s11356-021-12080-8

Khan, S. N., et al. (2017). Hydrological Modeling of Catchment Areas Using HEC-HMS Model. Journal of Engineering Science and Technology, 12(4), 1422-1430.

Kundzewicz, Z. W., Kanae, S., Seneviratne, S. I., Handmer, J., Nicholls, N., Peduzzi, P., Mechler, R., et al. (2014). Flood risk and climate change: Global and regional perspectives. Hydrological Sciences Journal, 59(1), 1-28. doi: 10.1080/02626667.2013.831152

Louw, E., Olanrewaju, C. C., Olanrewaju, O. A., & Chitakira, M. (2019). Impacts of flood disasters in Nigeria: A critical evaluation of health implications and management. Jàmbá: Journal of Disaster Risk Studies, 11(1), 1-9. doi: 10.4314/jamba.v11i1.3

Majeed, M., Lu, L., Anwar, M. M., Tariq, A., Qin, S., El-Hefnawy, M. E., El-Sharnouby, M., Li, Q., & Alasmari, A. (2023). Prediction of flash flood susceptibility using integrating analytic hierarchy process (AHP) and frequency ratio (FR) algorithms. Frontiers in Environmental Science, 10, 1037547. doi: 10.3389/fenvs.2023.1037547

Memon, M. H., Iqbal, M. A., Ahmed, N., & Idris, I. (n.d.). Gender and Social Vulnerability to Climate Change: A Study of Disaster Prone Areas in Sindh.

Miyamoto, M., Ono, M., Okazumi, T., Nabesaka, S., & Iwami, Y. (2014). Applicability of a flood forecasting method utilizing global satellite information to an insufficiently-gauged river basin: A case of a river basin in the Philippines.

Mosavi, A., Ozturk, P., & Chau, K.-w. (2018). Flood prediction using machine learning models: Literature review. Water, 10(11), 1536. doi: 10.3390/w10111536

Mukherjee, B., Habib, M. F., & Dikbiyik, F. (2014). Network adaptability from disaster disruptions and cascading failures. IEEE Communications Magazine, 52(5), 230-238. doi: 10.1109/MCOM.2014.6815213

Munir, A., Ghufran, M. A., Ali, S. M., Majeed, A., Batool, A., Khan, M. B. A. S., & Abbasi, G. H. (2022). Flood susceptibility assessment using frequency ratio modelling approach in Northern Sindh and Southern Punjab, Pakistan. Polish Journal of Environmental Studies, 31(4), 3249-3261.

Musarat, M. A., Alaloul, W. S., Rabbani, M. B. A., Ali, M., Altaf, M., Fediuk, R., Vatin, N., et al. (2021). Kabul river flow prediction using automated ARIMA forecasting: A machine learning approach. Sustainability, 13(19), 10720. doi: 10.3390/su131910720

Nanditha, J. S., Kushwaha, A. P., Singh, R., Malik, I., Solanki, H., Chuphal, D. S., Dangar, S., Mahto, S. S., Vegad, U., & Mishra, V. (2023). The Pakistan flood of August 2022: Causes and implications. Earth's Future, 11(3).

Noji, E. K., & Lee, C. Y. (2005). Disaster preparedness. In H. Frumkin (Ed.), Environmental health: From global to local (1st ed., pp. 745-780). Jossey-Bass Publishers.

Otto, F. E. L., Zachariah, M., Saeed, F., Siddiqi, A., Kamil, S., Mushtaq, H., Arulalan, T., et al. (2023). Climate change increased extreme monsoon rainfall, flooding highly vulnerable communities in Pakistan. Environmental Research: Climate, 2(2), 025001. doi: 10.1088/2752-5295/ac956a

Pakistan – Pakistan Flood Response Baseline Assessment - Sindh Province (October 2022) | Displacement Tracking Matrix. (n.d.). Retrieved from https://dtm.iom.int/reports/pakistan-pakistan-flood-response-baseline-assessment-sindh-province-october-2022

Qamer, F. M., Abbas, S., Ahmad, B., Hussain, A., Salman, A., Muhammad, S., Nawaz, M., Shrestha, S., Iqbal, B., & Thapa, S. (2023). A framework for multi-sensor satellite data to evaluate crop production losses: The case study of 2022 Pakistan floods. Scientific Reports, 13(1), 4240. doi: 10.1038/s41598-023-31741-5

Rizwan, M., Li, X., Chen, Y., Anjum, L., Hamid, S., Yamin, M., Chauhdary, J. N., Shahid, M. A., & Mehmood, Q. (2023). Simulating future flood risks under climate change in the source region of the Indus River. Journal of Flood Risk Management, 16(1), e12857. doi: 10.1111/jfr3.12857

Saeed, U., Hussain, M., Mukhtar, R., Younas, I., Ali, F., Akmal, F., & Khan, S. (2024). Development of a spatial framework for flash flood damage assessment and mitigation by coupling analytics of machine learning and household level survey data–A case study of rapid collaborative assessments and disbursement of public funds to the affectees of floods 2022, Punjab Pakistan. International Journal of Disaster Risk Reduction, 108, 104463. doi: 10.1016/j.ijdrr.2023.104463

Shafi, M. S. (2010). Pakistan: Pakistan Floods (2010) Damage and Needs Assessment.

Shah, S. M. H., Mustaffa, Z., Teo, F. Y., Imam, M. A. H., Yusof, K. W., & Al-Qadami, E. H. H. (2020). A review of the flood hazard and risk management in the South Asian Region, particularly Pakistan. Scientific African, 10, e00651. doi: 10.1016/j.sciaf.2020.e00651

Shehzad, K. (2023). Extreme flood in Pakistan: Is Pakistan paying the cost of climate change? A short communication. Science of The Total Environment, 880, 162973. doi: 10.1016/j.scitotenv.2023.162973

Tompkins, E. L., & Adger, W. N. (2004). Does adaptive management of natural resources enhance resilience to climate change?. Ecology and society, 9(2).

Triana, E. S., O'Donnell, A., Afzal, J., Enriquez, S., Dezfuli, G., Larsen, B., & Strukova, E. (2013). Social Resilience in Sindh, Pakistan: Addressing Sustainable Development Priorities.

Urlainis, A., Ornai, D., Levy, R., Vilnay, O., & Shohet, I. M. (2022). Loss and damage assessment in critical infrastructures due to extreme events. Safety science, 147, 105587. doi: 10.1016/j.ssci.2021.105587

Zafar, S., & Zaidi, A. (2019). Impact of urbanization on basin hydrology: A case study of the Malir Basin, Karachi, Pakistan. Regional Environmental Change, 19(6), 1815-1827. doi: 10.1007/s00477-019-01618-7