International Journal of Disaster Risk Reduction 121 (2025) 105442

Contents lists available at ScienceDirect

International Journal of Disaster Risk Reduction

journal homepage: www.elsevier.com/locate/ijdrr

High-resolution flood susceptibility mapping and exposure assessment in Pakistan: An integrated artificial intelligence, machine learning and geospatial framework

Mirza Waleed [a], Muhammad Sajjad [a,b,*]

[a] Department of Geography, Hong Kong Baptist University, Hong Kong Special Administrative Region of China
[b] The Division of Landscape Architecture, The University of Hong Kong, Hong Kong Special Administrative Region of China

ARTICLE INFO

Keywords:
Floods
Flood susceptibility mapping
Machine learning
Disaster risk management
Pakistan

ABSTRACT

Flood-related disasters have far-reaching impacts on infrastructure and societal well-being. Though characterizing flood susceptibilities using state-of-the-art approaches and modelling socio-economic exposure to highlight vulnerabilities is essential to assess and manage flood-associated risks, current studies are usually regional/coarser resolutions neglecting localized situations. Here we developed an integrated machine learning, artificial intelligence, and geo-spatial modelling-based framework for high-resolution flood susceptibility (30 m) and socio-economic exposure estimations at a larger scale using Pakistan as a case. To do so, the data on flooding, elevation, drainage, rainfall, Landsat-8 imagery, and gridded socio-economic layers were used. We produced the first national-scale high-resolution susceptibility maps for Pakistan, pinpointing areas at higher risk of flooding, and assessing the potential impact on the population and the economy. Our findings suggest that ∼29 % of the total area of Pakistan falls under critical flood susceptibility levels, with Sindh and Punjab being the most at-risk provinces. Notably, ∼95 million people (47 %) in Pakistan are exposed to high flood susceptibility with 74 % population of Sindh, 56 % of Punjab, and 33 % of Balochistan residing in high susceptibility areas. We further pinpoint economic hotspots in Sindh and upper Punjab as particularly vulnerable to flood risks, which calls for proactive disaster preparedness measures. Through the presented characterization of flood susceptibility and socio-economic exposure, our findings are useful to devise targeted interventions in highly exposed regions to enhance resilience and reduce the risks/impact of future floods. By addressing vulnerabilities and fostering resilience, Pakistan can effectively mitigate flood risks and safeguard its population and infrastructure.

1. Introduction

Socio-economic risks owing to substantial hazards (i.e., floods) are a central concern of public well-being, climate science, and governments at national and sub-national levels. Floods have emerged, globally, as one of the most severe hazards, bringing catastrophic damage worldwide and affecting tens of millions of people annually [1]. Approximately 23 % global population (∼1.81 billion) is exposed to higher levels of flood risk, which is further expected to alleviate to ∼2.3 billion people by 2050 [2]. Under

* Corresponding author. Office KB609, Knowles Building, The Division of Landscape Architecture, The University of Hong Kong, Pok Fu Lam, Hong Kong Special Administrative Region of China.
E-mail address: mah.sajjad@hotmail.com (M. Sajjad).

https://doi.org/10.1016/j.ijdrr.2025.105442
Received 5 July 2024; Received in revised form 28 March 2025; Accepted 28 March 2025
Available online 29 March 2025

anthropogenic global warming, an increasing trend in the frequency and duration of flooding has been observed, affecting the well-being of societies [3]. The direct asset loss caused by flooding is more significant than other hazards, surpassing hundreds of billions of United States Dollars (USD) each year. Flood exposure and impacts are more consequential in developing countries with fewer resources, limited coping capacity, and weak infrastructure; which are usually ill-equipped to timely implementation of mitigation policies and effectively adapt to the impacts of floods [4].

A detailed high-resolution exposure assessment is essential for shaping evidence-based decision-making, resource allocation, and policy formulation in disaster management and risk planning sectors. Flood susceptibility mapping (FSM) and exposure estimation are key components of disaster management. Such evaluations help identify disproportionate patterns in exposed populations and critical infrastructure to varying degrees of flood susceptibility levels [2]. Identifying these patterns further helps highlight priority intervention areas for immediate and gradual actions to reduce flood risks and mitigate potential losses to societal well-being. For exposure assessment, Nighttime Light (NTL) and population distribution function as a reasonable proxy for the socio-economic vulnerabilities of an area [5,6]. These socio-economic exposure analyses provide essential pre-disaster prevention opportunities and post-flood insights (i.e., the interconnected spatial relationship between flood-prone areas, human settlements, and economic activity hubs), which allows for a more holistic assessment of flood risks [3,5,7]. Additionally, by examining the spatial interactions between flood susceptibility, human settlement density, and economic activity, targeted interventions could be devised for effective resource allocation and implementation of risk reduction strategies to enhance flood resilience.

Pakistan, ranked among the top 10 most vulnerable countries to climate change impacts and population exposure, has experienced devastating flood events in recent decades, with major incidents occurring in 1992, 2010, 2014, 2018, 2020, and 2022 [8]. The increasing frequency and severity of these floods are driven by climate change-induced factors such as rising temperatures, irregular rainfall patterns, and glacial melting, compounded by rapid land cover changes, inadequate disaster management policies, lack of resilient infrastructure, and the country's unique topography. According to the Global Climate Risk Index, Pakistan ranked 8th among the most affected countries by climate change between 2000 and 2019 [9]. These challenges have resulted in substantial economic losses, displacement of millions, and significant fatalities, with the 2022 floods alone causing an estimated loss of USD 15.2 billion and affecting over 33 million [10].

Despite ongoing efforts to improve flood forecasting and disaster management, Pakistan faces persistent challenges in preparedness, coordination, and informed resource allocation, often leading to delayed response and recovery efforts [11–13]. Current flood risk assessments have primarily relied on coarse-resolution data (ranging from 250m to 25 km), which fail to capture localized flood-prone areas and the complex interplay of environmental and socio-economic factors necessary for effective intervention [2]. This limitation hinders policymakers and disaster management authorities from developing precise mitigation strategies and implementing timely preventive measures. To address these shortcomings, it is crucial to generate high-resolution and accurate flood-related estimates (i.e., susceptibility and socio-economic exposures) at the national scale. High-resolution assessments, such as those proposed in this study, offer finer spatial details that enable authorities to identify high-risk areas, prioritize resources effectively, and develop proactive disaster response strategies. Investing in a comprehensive, high-resolution FSM and exposure assessment framework is essential to enhance Pakistan's resilience to climate-induced flood risks and ensure more informed disaster management planning.

With the upsurge in global flood catastrophes, Remote Sensing-based FSM techniques are widely used to assess flood-prone regions in order to assist in disaster preparedness and prevention activities [14,15]. FSM involves identifying regions with varying levels of flood vulnerability based on certain physical variables. Conventional FSM techniques involve using statistical and hydrological models for identifying flood-prone regions; however, certain methods come with a downside of accuracy and thus are unreliable, particularly at larger scales [14,16–19]. Furthermore, certain hydrological models require large computational capabilities in order to generate intel – thus creating hurdle in regions with limited resources [20,21]. In this regard, Machine Learning (ML) based frameworks have emerged as valuable alternatives, which are not only able to identify flood-vulnerable regions with good accuracy but also require comparatively lesser input data for training. While existing studies have benchmarked various ML models in terms of their accuracy and transferability, they only provide partial insights on model efficiency (i.e., scalability)—representing a significant gap [22]. Highly efficient and scalable models not only ensure sustained and reliable performance but also allow ML-based frameworks to easily be extended over larger regions with comparatively lesser requirements of computational resources.

Among different ML models, tree-based are widely recognized because of their better data-handling characteristics [20]. For instance, Random Forest has been widely adopted in many studies [15,17,18,23]. However, recent studies discuss the advancement of gradient-boosting-based decision trees to be superior to simple tree-based models, especially in terms of their overall efficiency [14,24]. Among gradient boosting models, eXtreme Gradient Boosting (XGBoost) and Light Gradient Boosting Machine (LGBM) models are widely applicable; however, they have not yet been utilized for their efficiency for large-scale and high-resolution FSM. Similarly, even though the application of such ML models could be traced in various fields, the explainability of results, model outputs, and significant influencing features remain open questions. Integration of Explainable AI in ML frameworks could progressively help overcome this limitation.

On another front, it is often unwieldy working with large-scale computations (i.e., big geospatial datasets) to process and analyze information locally due to computational limitations. Google Earth Engine (GEE), a planetary-scale cloud computing platform that hosts petabytes of remote sensing data [25], helps acquire, manipulate, and preprocess large datasets and export the outputs for local utilization. Importantly, GEE also supports ML model implementation and GIS tools, which enables robust assessments with flexibility. While GEE supports some ML model-based classification, it lacks the integration of gradient boosting models, including XGBoost and LGBM, which previously have proven better for flood classification compared with conventional ML models (i.e., Random Forest) [26,27]. Therefore, instead of performing ML-based FSM modelling on GEE, this study utilized GEE primarily for data collection and pre-processing (i.e., for the preparation of flood conditioning features).

Here we develop a comprehensive ML, explainable AI, and geospatial tools integrated framework for high-resolution FSM and exposure estimation. Building on our prior research [24], where the scalability and accuracy of various machine learning models were rigorously evaluated at the watershed scale; we identified XGBoost and LGBM as the most efficient and scalable models. These models were employed in this study to identify comparative performance when applied at a national scale. The resulting outputs reflect high-resolution (30 m) flood susceptibility and exposure estimates of population (i.e., total population, children, and non-adults) and economic activity. The findings can be used to evaluate spatially relative flood risks at either a national scale (e.g., between provinces and districts) or a local scale (e.g., neighbourhood aggregation). The results can further alter our understanding regarding the geography of at-risk populations and economic activity based on the distribution of flood susceptibility. The presented integrated framework may also serve as an important complement to traditional measures of susceptibility and exposure. Similarly, spatial patterns identified in this study would potentially serve as a reference to reveal significant hotspots of extreme susceptibility and exposure for local policy-targeting purposes leading to a more resilient and prepared society.

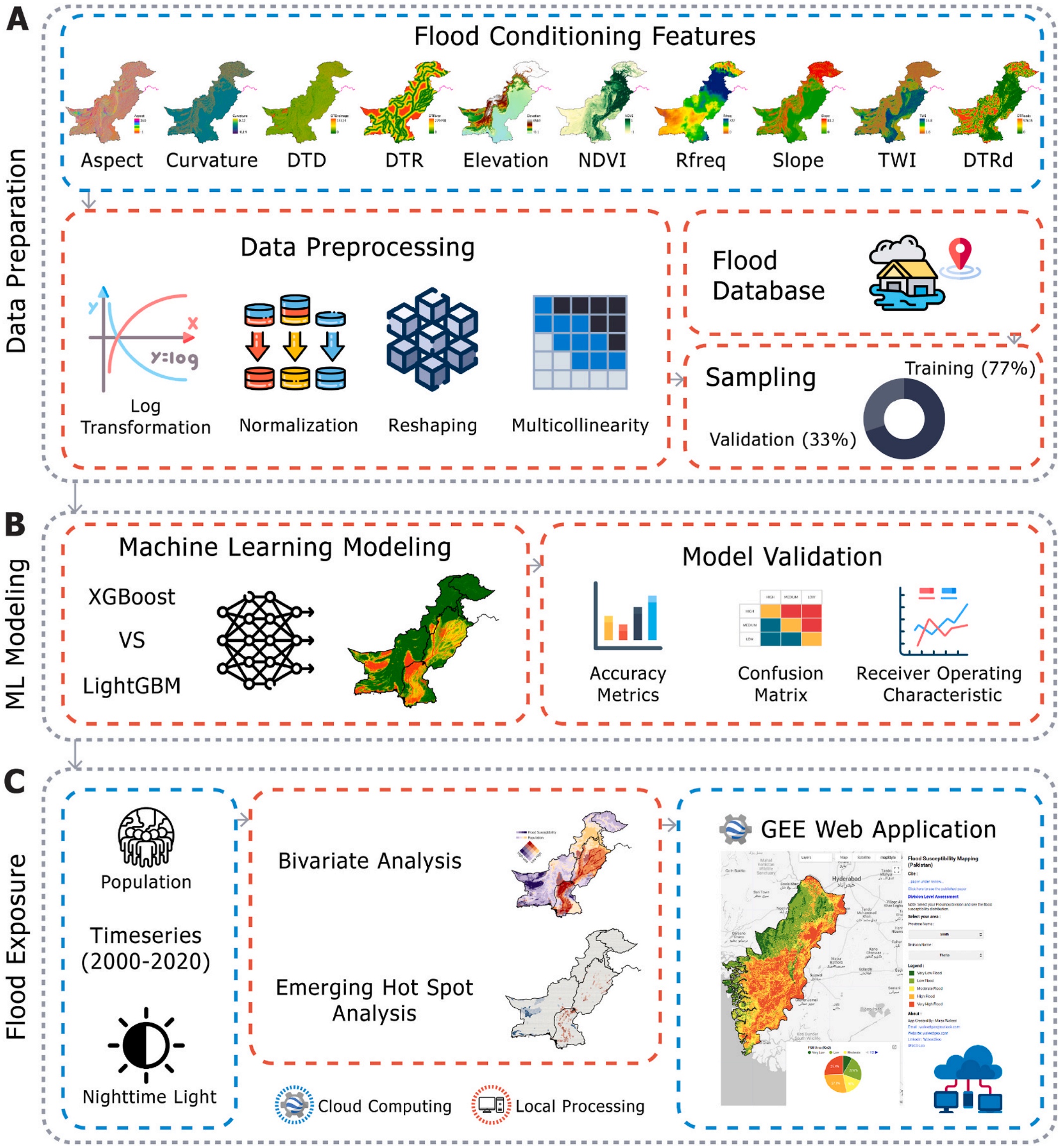

Fig. 1. Methodology flowchart. The sections in the blue show the processing steps performed using cloud computing platform (Google Earth Engine), whereas the red shows the local processing using a desktop system.

2. Materials and methods

2.1. Flood susceptibility and exposure estimation methodology

The methodological approach in the proposed framework includes several stages, such as data preparation, flood conditioning features, data preprocessing, training samples, machine learning modelling, model validation, flood exposure analysis, and web application development (Fig. 1). We start with multi-sources data acquisition including topographical, hydrological, meteorological data, and satellite images. These data are then processed using a number of preprocessing methods, including log transformation, normalization, reshaping array, and multicollinearity assessment. This step is important to ensure that the data are cleaned, easily understandable by ML models, and have the least effect on the model's performance [22]. Once the data are pre-processed, ~20,000 samples are generated through an equal random stratified sampling approach among which 77 % of data are used to train the ML model and the rest 23 % are used for model validation. After sample preparation, the ML models are trained, and FSM results are generated. Besides, an accuracy assessment is performed to ensure the sustained performance and reliability of the model outputs. Once the model output fulfills reasonable accuracy, the FSM results are used for further exposure analysis. Two variables namely population and NTL are used, representing the socio-economic situation, to evaluate spatial-temporal patterns. Lastly, the FSM and population exposure results are made available online as a stand-alone GEE web-based application (available at https://github.com/waleedgeo/FSM-PK).

2.2. Flood conditioning features (FCFs) preparation

For FSM, ten flood conditioning features (FCFs) are finalized based on comprehensive literature review [14,15,18,19] including surface aspect, curvature, distance to drainage (DTDrainage), distance to river (DTRiver), surface elevation, normalized difference vegetation index (NDVI), rainfall frequency (Rfreq), slope, and topographic wetness index (TWI). The source details and description of each feature are provided in Table 1. These features were chosen not only for their alignment with global flood susceptibility modeling practices but also for their relevance to Pakistan's specific hydrological, topographical, and environmental context, ensuring both applicability and contextual accuracy. Besides, the FCFs selection was also informed by recent literature (e.g., Ref. [24]), which evaluated the influence of FCFs at a local scale in Pakistan, highlighting their suitability in flood susceptibility modeling.

The elevation feature is first prepared using a digital elevation model (DEM). Conventionally, DEM is used to represent an elevation feature, which is available from various sources, including the Copernicus DEM, with global coverage and a spatial resolution between 90 and 30 m. However, most of the available DEMs contain forest and building artifacts, which hinders the effective application of the dataset in various applications including flood modelling. Recently a forest and buildings removed DEM dataset was proposed, which provides DEM information with greater accuracy [28]. Therefore, this study utilized forest and buildings removed DEM to avoid any data-induced errors. Elevation values range from −8.1 to 8569 m in Pakistan, where lower elevations in floodplain areas such as Sindh and Punjab are highly susceptible to floods, while higher elevations generally experience localized flash floods. Based on this DEM, four more variables including aspect, curvature, slope, and TWI are estimated. Aspect refers to the direction that a slope faces, which is calculated by the orientation of the slope. Aspect values range from 0 to 360°, with lower values (north-facing slopes) retaining more moisture and contributing to higher flood susceptibility, especially in the northern regions of Pakistan. Curvature is the representation of how the earth's surface deviates from being flat or planer and thus quantifies the degree of surface bending as well as bending rate at a fixed point. Curvature values range from 0 to 0.12, where higher values in steep mountainous regions like northern Pakistan increase runoff potential, while lower values in flat areas such as Punjab and Sindh lead to water accumulation and prolonged flooding. The slope is defined as a degree of surface elevating or descending over a given horizontal distance and is usually calculated as either a ratio or percentage. Slope values range from 0 to 83.2°, with steep slopes in northern Pakistan causing rapid runoff and flash floods, while gentle slopes in floodplains slow down water movement, increasing susceptibility to stagnant flooding. Lastly, TWI is a derived terrain index, which identifies the relative wetness or water saturation of a landscape, and is calculated using slope, flow accumulation, and contributing area. TWI values range from 2.6 to 35.8, with higher values in low-lying flood-prone areas such as Punjab indicating greater water retention capacity and flood susceptibility, while lower values suggest efficient drainage in elevated regions. The model used to derive TWI is as follows:

$$TWI = \log \frac{\sigma}{\tan \beta} \tag{Eq.1}$$

where σ is the upslope contributing area (m^2) and β is the slope (radiance).

Next, NDVI feature (proposed by Tarpley et al. [29]) is prepared, for which Landsat-8 images between 2021 and 2023 are taken, with less than 10 % cloud coverage. NDVI values range from −1 to 1, with lower values in urban and barren areas increasing runoff and flood risk, while higher values in vegetated areas improve water infiltration, reducing flood susceptibility. An average over the period is then taken to prepare a single composite image, which is then used to compute NDVI using the following model:

$$NDVI = \frac{NIR - Red}{NIR + Red} \tag{Eq.2}$$

where NIR and Red represent bands 5 and 4 in Landsat 8 imagery, respectively.

Lastly, the distance to roads, rivers, and drainage features are prepared using the Euclidean distance approach [30]. For this

Table 1
Datasets, their source and description.

Dataset Name	Resolution (year)/Type	Description	References	Source
Drainage	Vector shapefile, Level 08 (~150 Km^2 average basin size)	Hierarchical sub-basin breakdown derived from HydroSHEDS at 15 arc-second (~450m) resolution. Used to generate distance-to-drainage raster at 30m.	[44]	https://data.hydrosheds.org/file/hydrobasins/standard/hybas_as_lev08_v1c.zip
Elevation	30m raster (2022)	Forest And Buildings Removed Copernicus Digital Elevation Model. Files used: "N30E060-N40E070_FABDEM_V1-0.zip" and "N20E060-N30E070_FABDEM_V1-0.zip"	[28]	https://data.bris.ac.uk/datasets/25wfy0f9ukoge2gs7a5mqpq2j7/N30E060-N40E070_FABDEM_V1-0.zip https://data.bris.ac.uk/datasets/25wfy0f9ukoge2gs7a5mqpq2j7/N20E060-N30E070_FABDEM_V1-0.zip
Flood Layers (2010 and 2014)	Vector shapefiles	Flood extent polygons derived from UNOSAT's flood portal.	https://unosat.org	Link for the portal: https://unosat.org/products Direct link to filtered data: https://unosat.org/products?date_from=2009-01-01&date_to=2014-12-31&region=Pakistan&activation_type=Floods&title=&is_charter=null
Landsat-8	30m raster (2021–2023)	Orthorectified and Terrain Precision Corrected surface reflectance data. Flood extent polygon derived from satellite imagery analysis of the 2022 floods.	[45]	https://developers.google.com/earth-engine/datasets/catalog/LANDSAT_LC08_C02_T1_L2
Flood Layer 2022	Vector Shapefile	Flood extent polygon derived from satellite imagery analysis of the 2022 floods.	[46]	Derived from analysis described in Akhtar et al. [46]. Dataset obtained through personal communication with the authors. No direct download link is available.
CHIRPS Rainfall	0.05° (~5 km) raster (2010–2022)	Climate Hazards Group InfraRed Precipitation with Station data. Daily rainfall estimates	[47]	https://developers.google.com/earth-engine/datasets/catalog/UCSB-CHG_CHIRPS_DAILY
River Boundary	Vector shapefile derived from 15 arc-second (~450m) resolution data	Global river network dataset from HydroSHEDS. Used to generate distance-to-river raster at 30m.	[44]	https://data.hydrosheds.org/file/HydroRIVERS/HydroRIVERS_v10_as_shp.zip

purpose, vector shapefiles of rivers and drainage are extracted from DEM, and roads are obtained from the Global Roads Inventory Project (GRIP) dataset (see Table 1). The Euclidean Distance tool in ArcGIS Pro. is employed to calculate the proximity of each cell to the given feature. DTDrainage values range from ∼0 to 15,524 m, where areas closer to drainage channels, especially in river basins, face higher flood risks due to increased surface runoff. DTRiver values range from ∼0 to 279,498 m, with proximity to major rivers such as the Indus and Chenab increasing flood exposure, whereas distant areas experience lesser flood susceptibility. DTRoads values range from ∼0 to 97,615 m, where areas closer to roads, particularly in urban centers, face higher flood risks due to increased impervious surfaces that reduce infiltration. For Rfreq feature, data from the Climate Hazards Group InfraRed Precipitation with Station data (CHIRPS) satellite are filtered between 2010 and 2022 and all rainfall events accounting above 10 mm of rainfall are

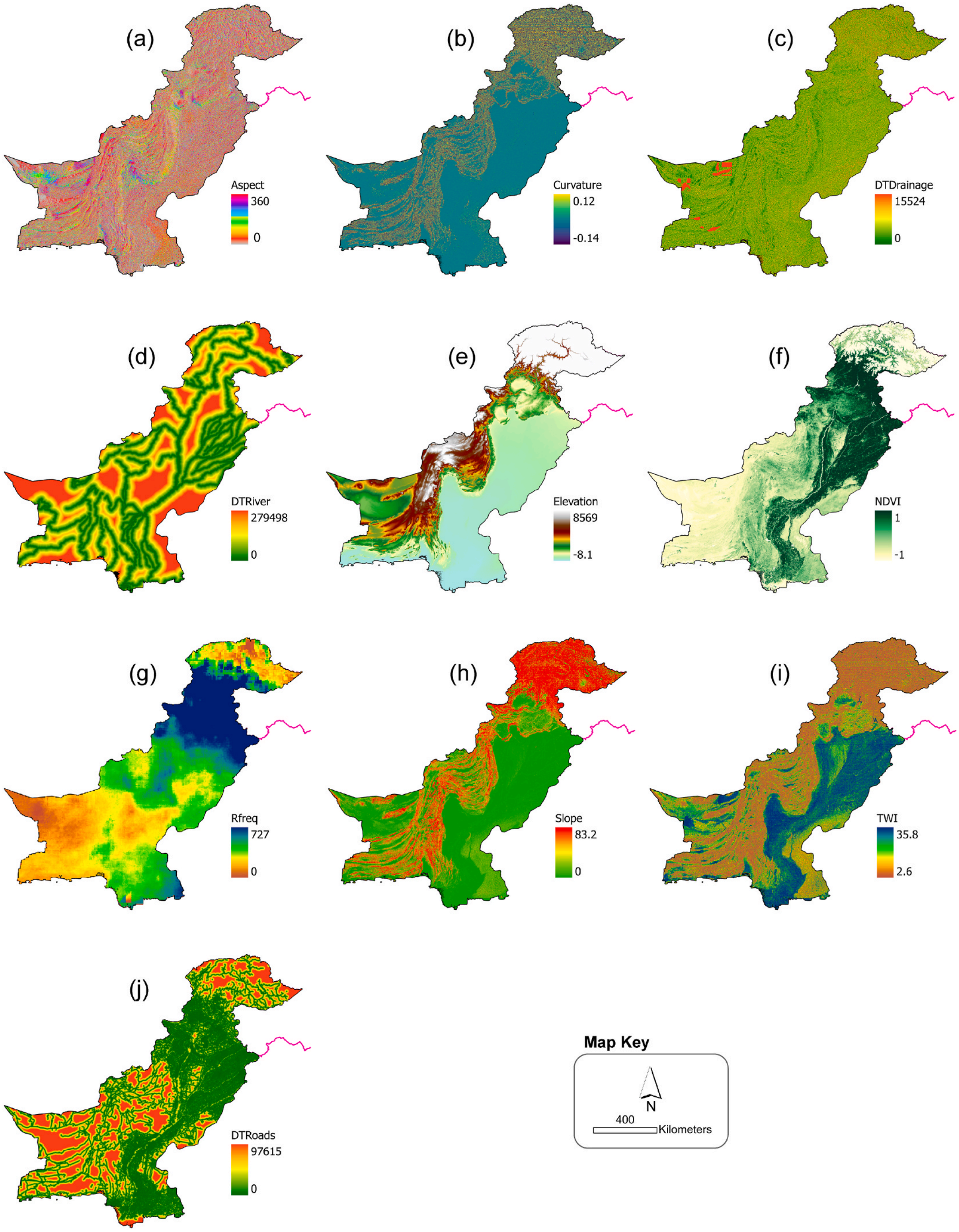

Fig. 2. Flood Conditioning Features (FCFs) including (a) aspect, (b) curvature, (c) dist. to drainage, (d) dist. to river, (e) elevation, (f) NDVI, (g) rainfall freq., (h) slope, (i) TWI, and (j) dist. to roads.

summed up, thus resulting in rainfall frequency [22]. Rainfall frequency varies from ~0 to 727 events, with higher values in monsoon-affected areas like Punjab and Khyber Pakhtunkhwa increasing flood susceptibility, while arid regions face less frequent but intense rainfall events. All these FCFs are initially processed in GEE (i.e., filtering, cloud removal, and composite) and are then exported to a local workflow where ArcGIS Pro. software is used to further process them (i.e., Euclidean distance function, TWI, etc.). The spatial distribution of all FCFs in the study area is presented in Fig. 2.

2.3. FCFs preprocessing

To make the FCFs unbiased, and more understandable to the ML model, a few preprocessing steps are essential including log transformation, normalization, reshaping array, and multicollinearity analysis. In ML modelling of RS data, often the end model is influenced by extreme values of input data and therefore induces skewness-based errors. Hence, log transformation is employed to reduce the skewness of input data to an extent where it does not affect the modelling process [31]. Moreover, with log transformation, a scaling factor is added to avoid negative values in the data. The model used to apply log transformation is given as:

$$\textit{Log Transformed} = \log\{\delta + \gamma\} \tag{Eq.3}$$

where δ is the single FCF, and γ is the factor added to the respective FSC for scaling and to avoid negative values in the data.

After reducing the skewness of data, it is crucial to normalize the data in a defined range so that the ML model can comprehend the data easily. For instance, in the case of FSM, the values range needs to be between 0 and 1, (i.e., values closer to 0 indicate the least flood potential, whereas values closer to 1 indicate the highest flood potential). Therefore, all the log-transformed FCFs are then normalized using the minimum-maximum method [31]. Features inversely related to flooding, such as NDVI and elevation, were normalized using negative normalization (Eq. (4)), while those directly related, such as rainfall frequency, were normalized using positive normalization (Eq. (5)). This differentiation preserves the directionality of the relationship, ensuring accurate interpretation by the ML models.

$$\textit{Negative Normalizion} = \frac{\max(\delta) - \delta}{\max(\delta) - \min(\delta)} \tag{Eq.4}$$

$$\textit{Positive Normalization} = \frac{\delta - \min(\delta)}{\max(\delta) - \min(\delta)} \tag{Eq.5}$$

where min (δ) and max (δ) represents minimum and maximum statistics of the log-transformed FCF distribution. Once the normalization is performed, the data needs to be reshaped to feed it to the ML model. While this step is optional, some ML models are prone to shape (dimensions) of data. Therefore, it is recommended that the data be either in 1D or 2D, depending on the end model architecture. In the case of LightGBM and XGBoost, both work efficiently with 2D datasets [32]. Hence, we reshaped our existing 1D data into 2D using NumPy's '*reshape*' command. Moreover, the resulting NumPy's arrays are stored into compressed *.NPZ* format, which ensures data compression and results in efficient data storage along with faster access to multi-dimensional arrays for ML modelling. These preprocessing steps not only ensure unbiased and comprehensible data for ML models but also standardize the workflow, making it scalable and applicable to diverse geographic regions and datasets.

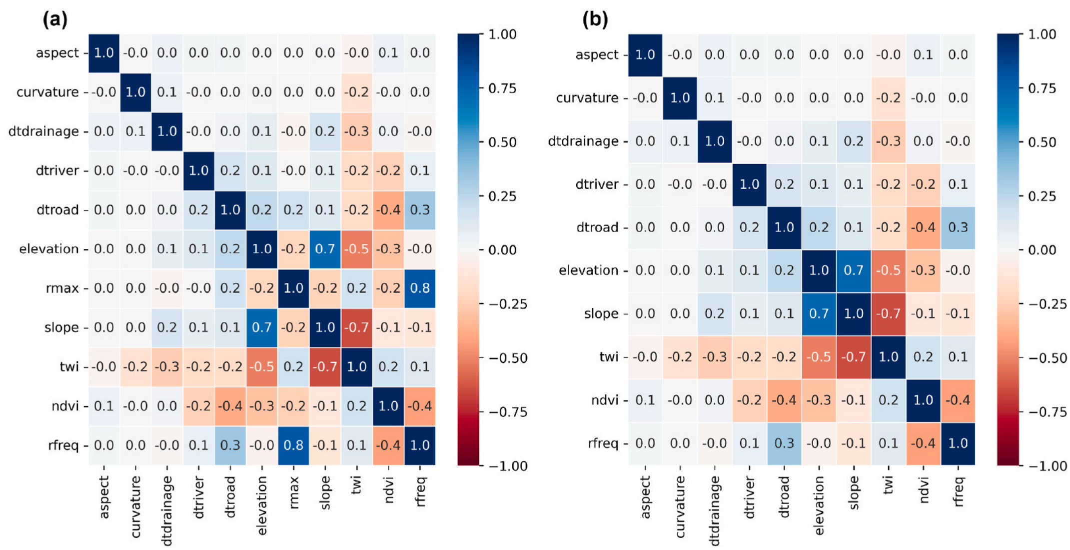

Fig. 3. (a) Initial correlation matrix including Rainfall maximum (Rmax), and (b) final correlation matrix without Rmax feature.

2.4. Multicollinearity between FCFs

Multicollinearity is a phenomenon in which the predictor variables (FCFs in our case) show high correlation among each other, and thus can induces bias in the final ML model [33]. For this study, initially 11 FCFs were prepared including 10 in Fig. 2 and maximum rainfall (Rmax). The maximum rainfall variable was prepared using the maximum observed rainfall between 2010 and 2022 duration. However, during multicollinearity test, it was observed that the feature Rmax shows a very high correlation (0.8) with Rfreq variable (see Fig. 3a). Therefore, we removed the highly correlated Rmax variable, and re-performed multicollinearity analysis, in which no variable exceeds the correlation threshold of greater than 0.75 (see Fig. 3b).

2.5. Machine learning (ML)-based flood susceptibility modelling

Data scare regions such as Pakistan, have faced many hurdles in on-time disaster management in the past due to unavailability of data to support rapid disaster relief measures ([12,34]; World Bank, 2022). In terms of flooding, there is no explicitly existing flood database provided by the government or international body for flood relief in Pakistan. Therefore, the initial step in this study was to prepare a flood database consisting of historical flood locations at national level, which could not only aid in this study, but also benefit upcoming initiatives for flood management. Flood layers are prepared for the 2010, 2014, and 2022 flood events (Table 1). Next, all three layers were overlayed, and common flood pixels were identified, which reflect the reoccurring flooding for different regions highlighting higher flood risks. To ensure balanced representation of flood and non-flood locations, a stratified sampling approach was employed, selecting an equal number of samples from both classes. This strategy mitigates class imbalance, which could introduce prediction biases, and enhances the reliability of the ML models in detecting flood-prone areas. Based on the reoccurrence of flood hazards in different areas, a total of 20,000 stratified random points were taken: 10,000 for flooded locations and 10,000 for non-flooded locations. These sample points are then stored in a geospatial database, to be later utilized by ML modelling.

In the past, statistical, and hydrological flood models were employed to study flood hazards, which were challenging due to large input data requirements, and high computational cost. On the contrary, ML has emerged as a domain that provides robust and accurate results with lesser input data and computational requirements [35]. In recent years, among ML models, gradient boosting models have gained significant recognition due to their ability to identify complex patterns without compromising accuracy [36]. Recent studies suggest LGBM and XGBoost to be the top models in flood modelling, with highest accuracy and least computational requirement [14]. Both models work by iteratively training an ensemble of weak decision tree models, where each subsequent model focuses on the misclassified instances by the previous models. This process allows them to effectively capture complex patterns and dependencies in the flood classification data. These models are capable of handling large datasets with high-dimensional features and exhibit impressive generalization capabilities [32]. As flood susceptibility is a probability-based estimation of flood occurrence per pixel, both XGBoost and LGBM support generalization and probability estimate from binary classification output, thus suitable for FSM at larger scales.

2.6. Validation of ML models

The most crucial step in ML-based FSM is the accuracy assessment which ensures the reliability of predicting flood prone areas in real world scenarios. In accuracy assessment, relying on a single accuracy metric is not reliable as it does not consider the under-fitting, nor over-fitting of the model [33]. If the input contains imbalanced data (unequal samples of flooded and non-flooded locations) it can induce bias. To address this, existing studies have used multiple accuracy metrics including Precision, F1 Score, Recall, Overall Accuracy, ROC (Receiver Operating Characteristics) and AUC (Area under the ROC Curve), and Jaccard Score [35,37]. Moreover, some recent studies have used adjusted accuracy to compare different statistical and ML models using multiple accuracy metrics [31].

To enhance generalizability and mitigate class imbalance effects, we employed stratified k-fold cross-validation, where the dataset was randomly partitioned into k subsets (where k = 10) while maintaining the proportion of flooded and non-flooded samples in each

Table 2
Details of accuracy metrics used in this study.

Name	Equation	References
Accuracy	$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$	[48]
Adjusted Accuracy	$Adj.\ Acc = \frac{Norm^{\mp}(\delta^1, \delta^2, \ldots \delta^{\infty})}{\sum N}$	[31]
F1-Score	$F1s = \frac{Precision \times Recall}{Precision + Recall}$	[49]
Jaccard Score	$Js = \frac{TP}{TP + FN}$	[50]
Precision	$Precision = \frac{TP}{TP + FP}$	[51]
Recall	$Recall = \frac{TP}{TP + FN}$	[52]
ROC and AUC	The ROC curve is constructed by plotting the true positive rate (TPR) against the false positive rate (FPR) at different classification thresholds.	[22]

fold. This approach prevents overfitting and ensures that the model is trained and tested on diverse data splits. Additionally, our validation framework integrates multiple accuracy metrics, as mentioned earlier, to provide a holistic evaluation of model performance. The incorporation of adjusted accuracy, which normalizes multiple metrics into a single performance measure, further strengthens the reliability of flood susceptibility assessment. The details of accuracy assessment metrics, and their equations are provided in Table 2.

where TP, TN, FP, and FN represent True Positive, True Negative, False Positive, and False Negative respectively. Furthermore, ROC and AUC represent Receiver Operating Characteristics and Area under the ROC Curve respectively. In adjusted accuracy δ is the accuracy metrics used, whereas N represents the total number of accuracy metrics used.

2.7. Feature importance through explainable-AI

For ML model training and prediction, different FCFs will have different effects on model output and thus it is crucial to understand the underlying working of the ML model to effectively understand the model output decision. However, the main difficulty in this process is the ML model black box issue (i.e., the lack of model transparency and interpretability for its output). Therefore to overcome this, we employed recently introduced explainable-AI (EAI) technique [21]. To employ this technique SHAP model available in Python is used and is applied to each model (LGBM and XGBoost) individually. The SHAP technique utilizes the concept of Shapley values from cooperative game theory to assign an "importance" score to each feature for a given prediction. The values are calculated to ensure fairness and consistency, with each feature receiving its fair share of contribution to the prediction. This guarantees a balanced distribution of influence among the features, accurately reflecting their impact on the model's output [22]. While SHAP provides insights into model behavior, it comes with certain drawbacks, one of which is its dependence on the quality and representativeness of training data. For instance, if the training data is imbalanced or has biases, the SHAP resulting interpretations will not be the true representation of model [38]. To overcome this issue, we adopted a stratified equal random sampling approach, as discussed earlier, which ensured training data randomness and balance between flood and non-flooded classes.

To understand the influence of FCFs on flood susceptibility modelling using different models, two SHAP methods namely summary plot (feature influence plot) and force plots (prediction force plot) are used. Summary plots provide a global view of feature importance by compiling SHAP values across all instances, allowing for an assessment on the relative impact of different features on model training. They also assist in identifying trends, such as whether higher or lower values of a feature increase the model's training capability. On the other hand, force plots offer a local perspective by illustrating the contribution of each feature to a specific prediction. This detailed view can explain individual predictions by displaying how each feature's value affects the model's output, which is crucial for case-by-case analysis in FSM.

2.8. Flood exposure estimation and its spatial disparities

For flood exposure analysis, average yearly image collections of NTL and population are prepared using GEE and are then exported into local ArcGIS Pro. workflow for further exposure analyses. For NTL, a harmonized image collection of NTL is filtered between 2000 and 2020 (representing two decades). The harmonized collection of NTL data from the Defense Meteorological Satellite Program (DMSP)/Operational Linescan System (OLS) and the Visible Infrared Imaging Radiometer Suite (VIIRS) are first merged into an image collection of yearly mean images ranging from 2000 to 2020 [39]. Similarly, the population dataset provided by the Worldpop [40] is used for each year between 2000 and 2020. For the exposure estimation, the most recent population layer is employed. To identify the spatial-temporal patterns, all 20 layers are utilized in the Emerging Hotspot Analysis tool (detailed below). For detailed profiling of the exposed population, three sets of populations including the total population, non-adult (0–14 years old) population, and elderly population (+65 years) are considered. The reason to incorporate non-adults and the elderly population is to represent the vulnerable groups within different levels of flood susceptibility. This assessment helps us in the detailed profiling of population exposure to flood risks.

To use time-series datasets for pattern mining, the entire study area is divided into several hexagons, with a 5 Km^2 radius. The time-series datasets are sampled for each location based on the sum and each year. This resulted in each hexagon containing the total population and NTL, between 2000 and 2020. Once the data are sampled, the columns are converted into rows, using Python and GeoPandas module function named "melt". Lastly, the finalized hexagon shapefile is imported into ArcGIS Pro. workflow, where the year column is assigned the data type as "*datetime*" using ArcGIS Pro. tool "*Covert Time Field*". The newly formed shapefile is then used in the tool "Create Space Time Cube" to create a Network Common Data Form (NetCDF) cube.

Flood exposure analysis is divided into two parts. First, we use the bivariate analysis to signify flood exposure in four groups including flood susceptibility *vs* NTL, flood susceptibility *vs* population, flood susceptibility *vs* elderly, and flood susceptibility *vs* non-adults. To provide geographical references, bivariate thematic maps are produced. For flood exposure patterns identification, the Emerging Hotspot Analysis tool available in the Space Time Pattern Mining toolbox in ArcGIS Pro. is employed. This tool helps us to identify spatial-temporal emerging hotspots of exposure in the study area. In the context of the flood, space-time analysis involves analyzing emergence and changes over time in the input variables (i.e., population and NTL) within different high flood susceptible areas.

3. Results

3.1. Flood susceptibility modelling and accuracy assessment

The accuracy assessment of the LGBM and XGBoost models is presented in Fig. 4, highlighting their comparative performance through multiple accuracy metrics. The results indicate that LGBM outperforms XGBoost across all assessed metrics (Fig. 4a), achieving higher precision (0.86 vs. 0.83), recall (0.86 vs. 0.84), and F1-score (0.86 vs. 0.84), signifying its superior capability in correctly identifying flood-prone areas while maintaining a balance between false positives and false negatives. The adjusted accuracy metric, which combines multiple accuracy indices, further reinforces LGBM's advantage with a value of 0.85 compared to 0.82 for XGBoost. The confusion matrices (Fig. 4b) provide additional insights, showing that LGBM correctly classified 43.72 % of flood-prone areas, while XGBoost achieved 42.84 %. Additionally, LGBM demonstrated lower misclassification rates with false positives at 6.98 % and false negatives at 7.00 %, compared to XGBoost's 8.49 % and 7.88 %, respectively. These findings indicate that LGBM is better at minimizing errors and ensuring higher predictive reliability. The ROC and AUC curves (Fig. 4c) reveal that both models achieved a mean AUC of 0.81, demonstrating strong discriminatory power. However, LGBM exhibited a narrower standard deviation (±0.12) compared to XGBoost (±0.13), indicating more consistent performance across different cross-validation folds. The ROC curves illustrate that LGBM maintains a higher true positive rate across varying false positive thresholds, making it a robust choice for large-scale flood susceptibility assessments.

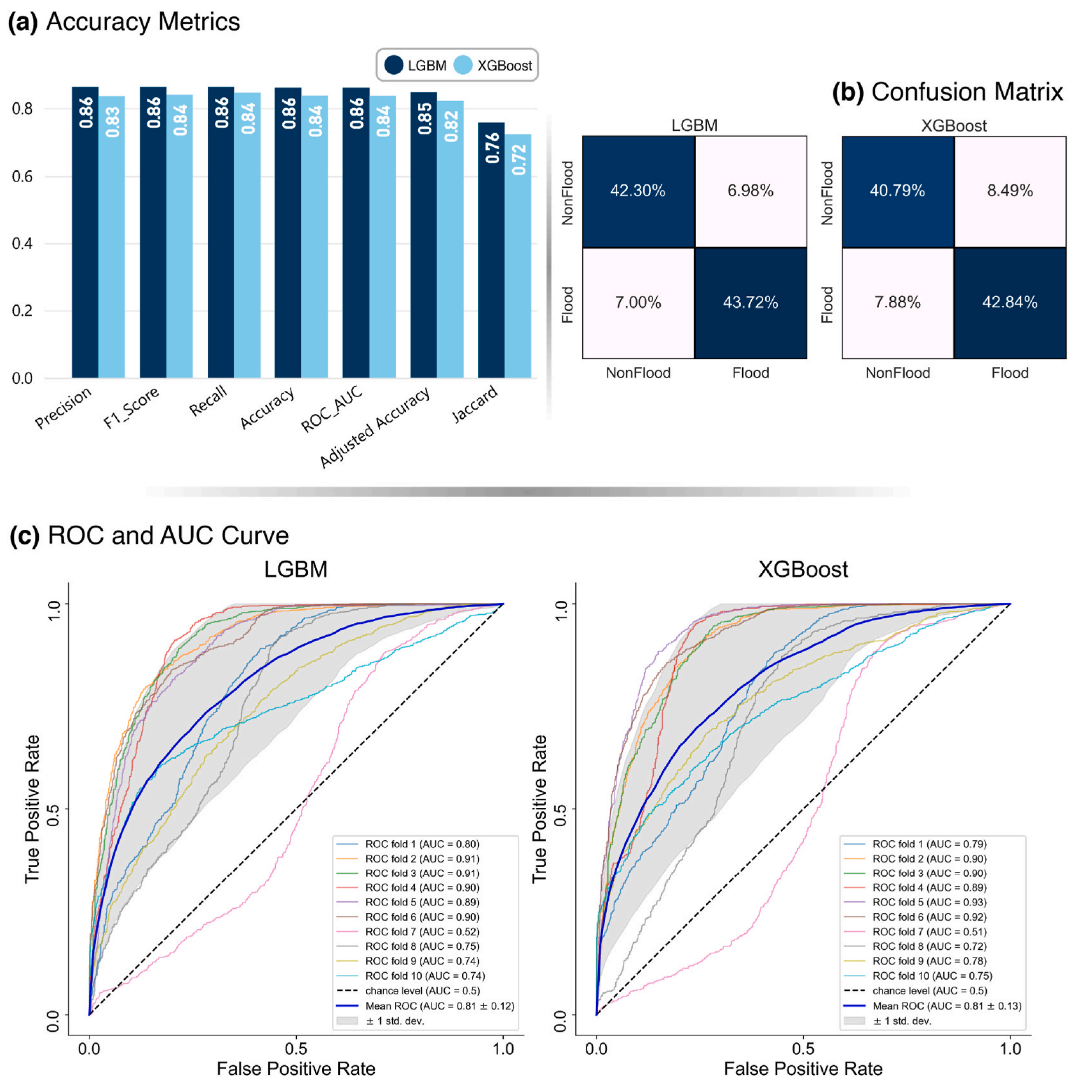

Fig. 4. Accuracy Assessment metrics (a). confusion matrix (b), and Receiver Operating Characteristics (ROC) and Area Under Curve (AUC) (c).

3.2. Flood susceptibility disparities in Pakistan

The results of FSM are shown in Fig. 5, in which Fig. 5a shows LGBM based flood susceptibility map, Fig. 5b shows XGBoost based flood susceptibility map, and lastly, Fig. 5c shows flood susceptibility distribution area in each province based on (highly accurate) LGBM model results. We categorize the overall distribution of flood susceptibility into five groups based on quantiles. These categories include “very low”, “low”, “moderate”, “high”, and “very high”, which are based on five quantiles ranging from the lowest 20 % (i.e., 0–20 %) to the upper 20 % (i.e., 80–100 %) values of the distribution [2]. In Fig. 5a, LGBM based FSM show 8 % of the area of Pakistan in the very high flood susceptible class, whereas 9 % for high susceptible classes, and 12 % for moderate flood susceptibility class.

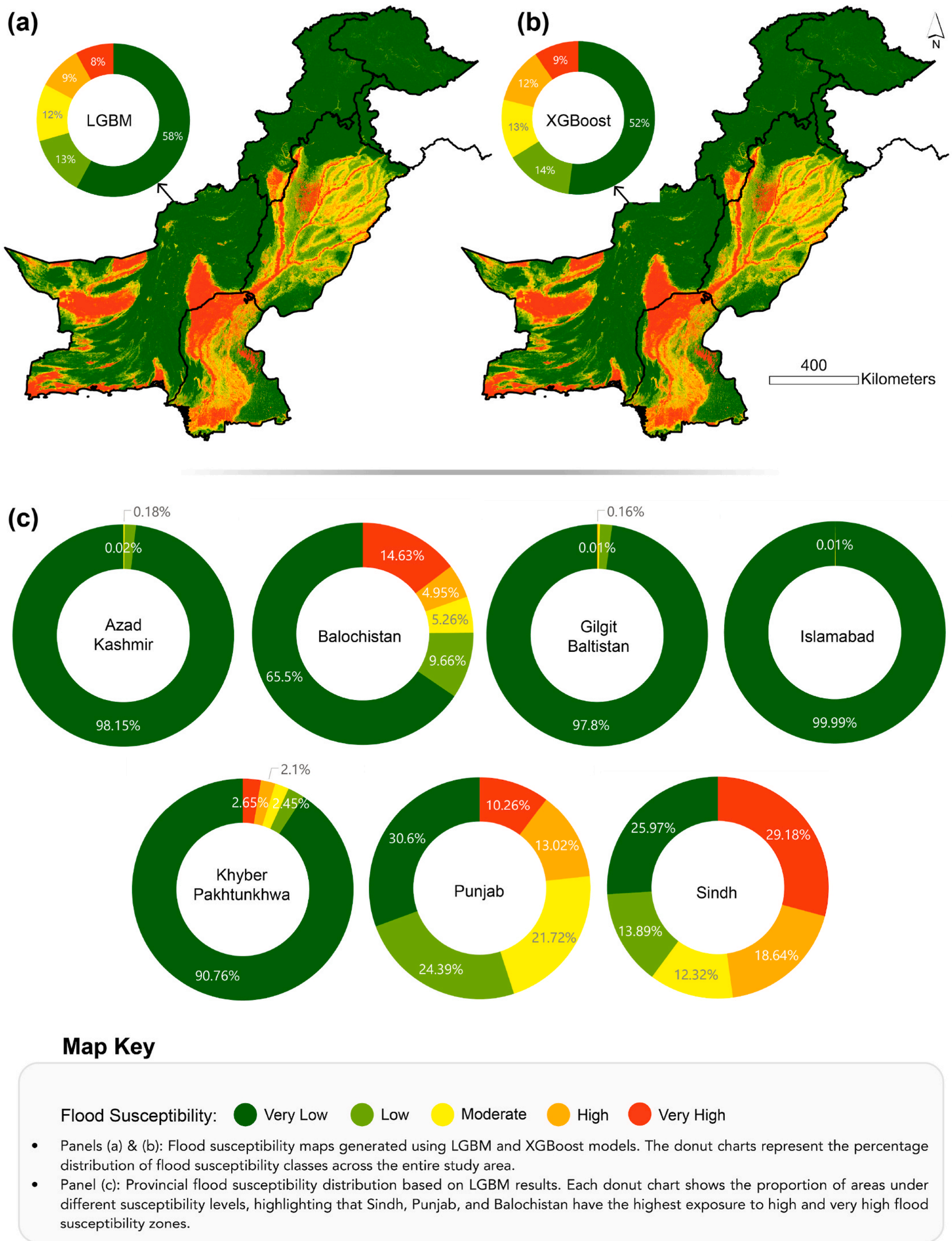

Fig. 5. Flood susceptibility maps for LGBM (a), XGBoost (b), and area distribution of flood susceptibility in each province based on LGBM (c).

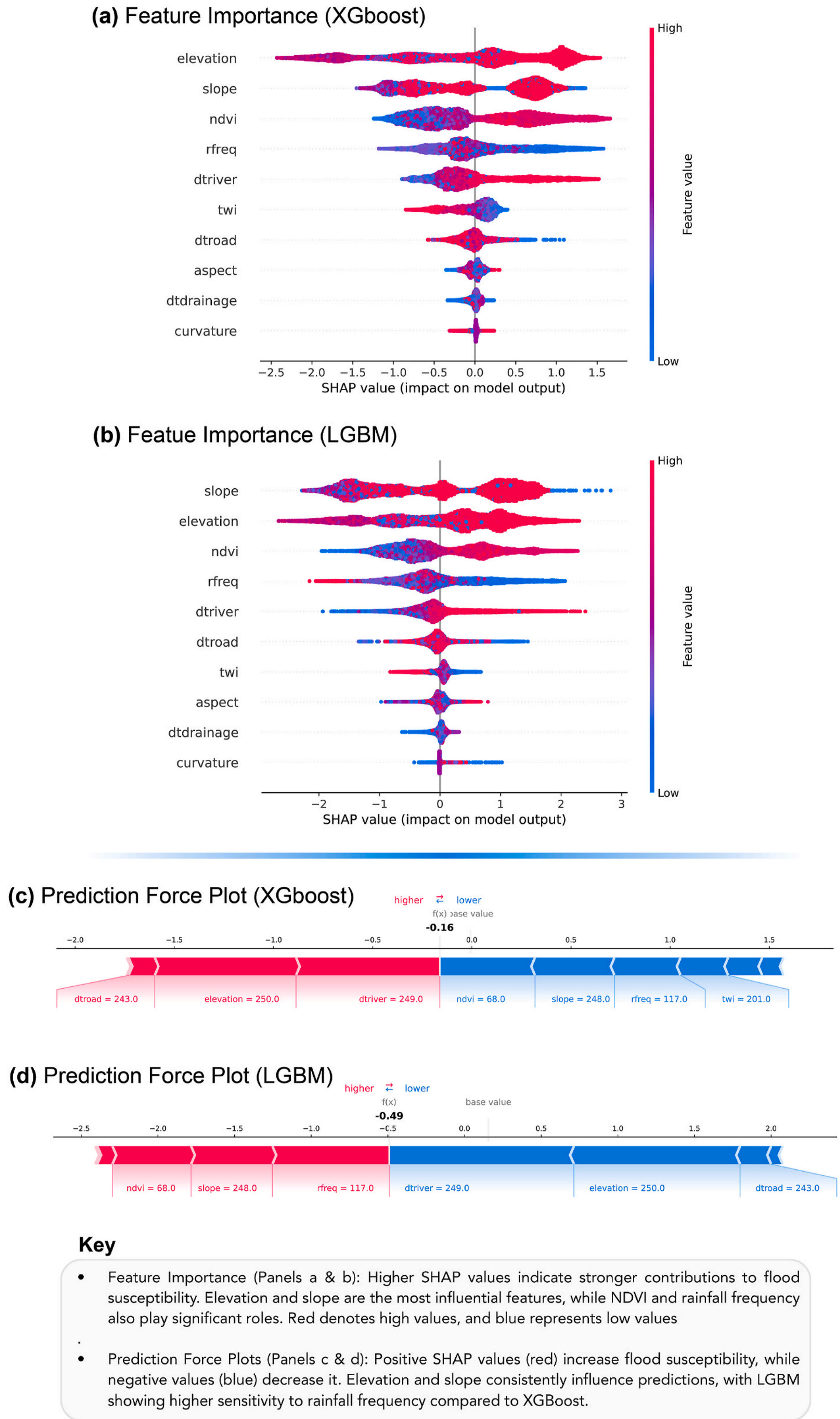

Fig. 6. Flood Conditioning Factor (FCF) importance ranking using (a) feature importance (XGBoost), (b) feature importance (LGBM), (c) prediction force plot (XGBoost), and (d) prediction force plot (LGBM).

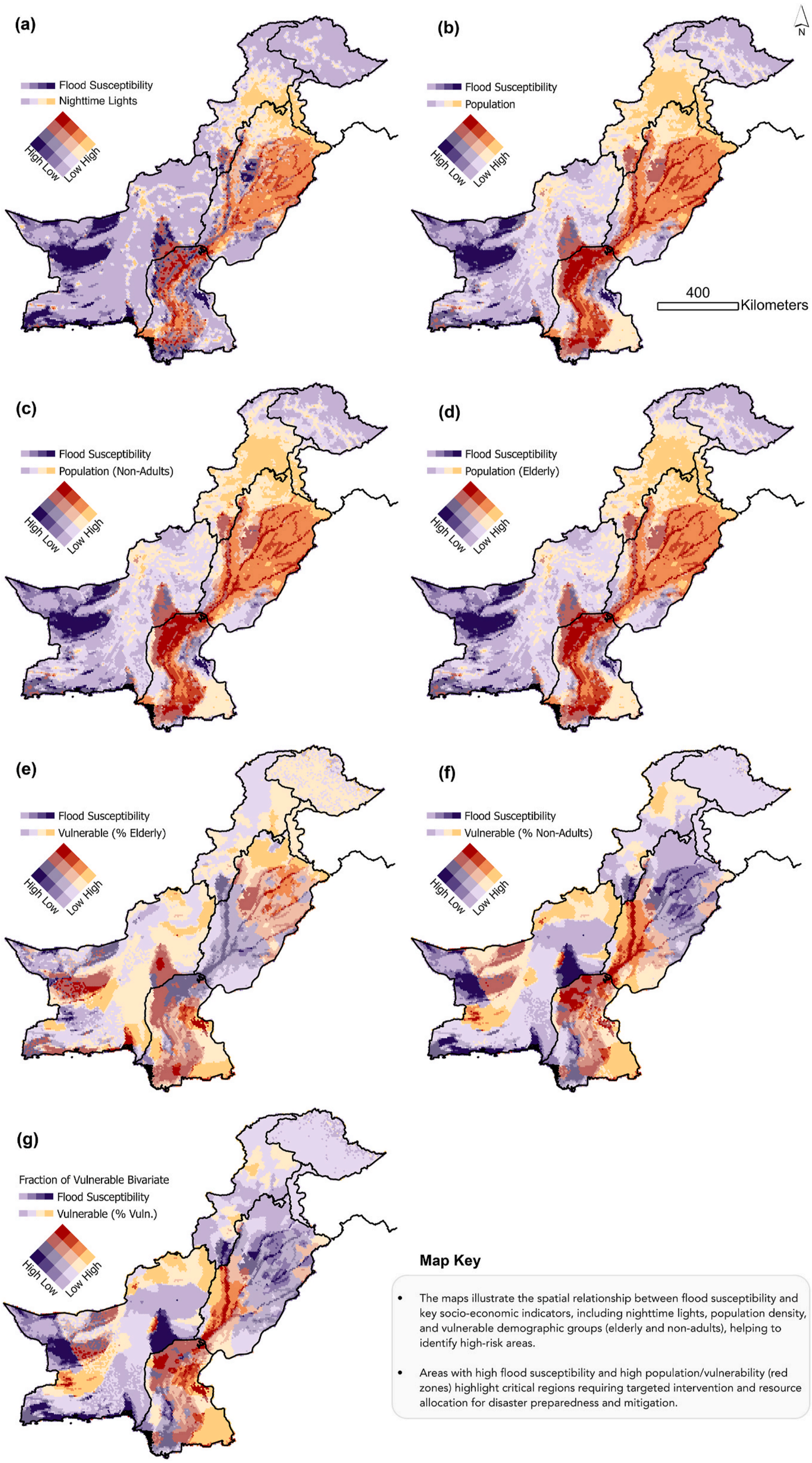

Fig. 7. Bivariate visualization showing flood exposure in four groups including (a) flood susceptibility with Nighttime Light (NTL), (b) flood susceptibility with Population, (c) flood susceptibility with Population (non-adults), and (d) flood susceptibility with Population (elderly).

Cumulatively, ~29 % of the total area of Pakistan falls under critical flood susceptibility levels. Similarly, Fig. 5b shows XGBoost based FSM, in which 9 %, 12 %, and 13 % of total area of Pakistan lies under very high, high, and moderate flood susceptibility class. This results into ~34 % of the total area of Pakistan experiencing critical flood susceptibility based on XGBoost. Though both models represent varied percentage of total susceptible area, they show similar geographical susceptibility distribution in Pakistan. It is evident that higher susceptibility is predicted for the areas lying in northern Sindh province, while moderate to high flood susceptibility areas are prominent in Punjab province along the river network.

Moreover, LGBM and XGBoost indicated 42 % and 48 % of the total area of Pakistan to be between low to very high flood susceptibility zones, respectively. This situation indicates that nearly half of the regions of Pakistan are prone to varying levels of flood events. For a regional comparison, Fig. 5c visualizes the area of each flood susceptibility zone in different provinces in Pakistan based on the LGBM model. It is observed that a very high flood susceptible zone lies in Sindh (~29 %) followed by Balochistan (~15 %) and Punjab (~10 %). Similarly, the high flood susceptibility zones are in the same three provinces with Sindh at ~19 %, Punjab at ~13 %, and Balochistan at ~5 %. In terms of critical flood susceptibility levels at provincial scales, Sindh ranks top with ~62 %, Punjab at second with ~45 %, and Balochistan at third with ~25 % of their total area under critical flood susceptibility levels.

Lastly, a comparison of the LGBM and XGBoost models reveals that LGBM estimates a slightly lower percentage of very high flood susceptibility areas (8 %) compared to XGBoost (9 %), suggesting that XGBoost may have a higher sensitivity in detecting high-risk zones. However, LGBM demonstrates greater spatial consistency and robustness in identifying flood-prone areas with lower misclassification rates. At the provincial level, Sindh exhibits the highest proportion of very high flood susceptibility (~29 %), which corresponds with historical flood patterns and its low-lying terrain. Punjab, with its extensive river networks, shows significant susceptibility, particularly along the Indus River, while Balochistan, despite its rugged topography, faces critical flood risks due to arid conditions and flash flood potential. These findings align with previous studies (e.g., Ref. [2]) but provide enhanced spatial detail at a 30m resolution, offering more localized insights for flood management.

3.3. Influence of flood conditioning features on susceptibility

To assess the contribution of each FCF to susceptibility, and to address the conventional black box issue of ML-based modelling, Explainable-AI is employed [21]. The result of different factors influencing susceptibility at both the training and prediction stages of ML modelling are shown in Fig. 6. The summary plots (based on model training) for XGBoost and LGBM models, respectively, are presented in Fig. 6a and b. Whereas Fig. 6c, and d shows Force plots (based on model prediction) for XGBoost and LGBM, respectively. The summary plot of XGBoost (Fig. 6a) shows elevation, slope, and NDVI as the top three contributing factors with higher values of elevation, slope, and NDVI, having a high impact on model output. On the other hand, lower values of NDVI represent minimal effect on model output. Similarly, in the summary plot of LGBM (Fig. 6b), the top three contributing FCFs are slope, elevation, and NDVI. However, in this case, the slope stood at the top with the highest impact on the model compared to other FCFs. The distribution of values of slope, elevation, and NDVI follows the same pattern as of XGBoost, with higher values of slope and elevation positively impacting the model, whereas the majority of negative NDVI values impact the susceptibility negatively.

In the force plot of XGBoost (Fig. 6c), DTRiver and elevation impact the susceptibility negatively (i.e., their negative values contribute towards model output prediction), whereas NDVI and slope contribute positively towards model prediction (positive values of NDVI and slope contribute towards model output prediction). In the force plot of LGBM (Fig. 6d), Rfreq and Slope appear to be the highly negative contributing FCFs, whereas DTRiver and elevation appear to be highly positive contributing FCFs.

3.4. Exposure estimation and its geographical disparities

Among seven analyzed exposure sets, the first group (Fig. 7a) shows the distribution of flood susceptibility and economic activity (represented by larger NTL intensities). It is observed majority of the areas experiencing high flood susceptibility and high economic activity are situated in Sindh and Punjab provinces, mostly around riverine basins. To further illustrate, the primary areas of interest are the regions where both susceptibility and economic activity are higher (i.e., regions in red shades), as these areas should be prioritized for flood risk mitigation and response measures in pre- and post-disaster situations, respectively. The second group (Fig. 7b) represents the distribution of flood susceptibility with the population in which the red areas (High flood susceptibility and high population) significantly clustered around northern Sindh and riverine basins in Punjab. Moreover, most of Punjab's population lies under moderate to high flood exposure (i.e., the population susceptible to flooding). Fig. 7c shows the geographical distribution of flood susceptibility and non-adult population illustrating that the highly exposed non-adults show a similar distribution as of group 2 (Fig. 7b); with highly concentrated in Sindh and Punjab regions. Similarly, group 4 (Fig. 7d) shows the distribution of flood susceptibility with the elderly population, following a similar geographical pattern as of groups 3 and 4 (Fig. 7b and c); with highly exposed elderly population located in Sindh and Punjab. Groups 5–7 (Fig. 7e–g) show the percentage of vulnerable populations in parallel to susceptibility levels.

In Fig. 7e, the bivariate relationship of the percentage of elderly with flood susceptibility is presented. This analysis presents patterns of high-density elderly populations falling under higher flood susceptibility levels. For instance, Fig. 7e shows the higher percentage of exposed elderly lies in upper Punjab along river basins, in upper and lower Sindh, and around the western regions of Balochistan province. Similarly, Fig. 7f shows the distribution of percentage of non-adults with flood susceptibility. The spatial distribution is somehow opposite to that of elderly; with higher density of exposed non-adult population located in lower Punjab alongside river basins, and upper Sindh. Lastly, Fig. 7g shows the relationship between percentage of total vulnerable population (including both elderly and non-adults) and flood susceptibility. It is observed that the patterns of exposed vulnerable population follow similar distribution to those of Fig. 7f; with higher proportions of exposed vulnerable population located in lower Punjab and upper Sindh.

Based on the flood susceptibility levels (LGBM Model) and population (total population), we further estimate the exposed population (Fig. 8). For a two-tier assessment, donut charts are categorized into national and provincial level. At the national level, our estimates show that approximately 27.6 million people (~13.8 % of total population) in Pakistan is exposed to very high flood susceptibility level, 28 million (~14 % of total population) resides in regions with high flood susceptibility level, and 39.1 million (~19.5 % of total population) falls in the areas with a moderate flood susceptibility level. Considering the climate change induced intensification of flood events in the future, which could transform the moderate susceptibility zones into higher category, nearly 94.7 million

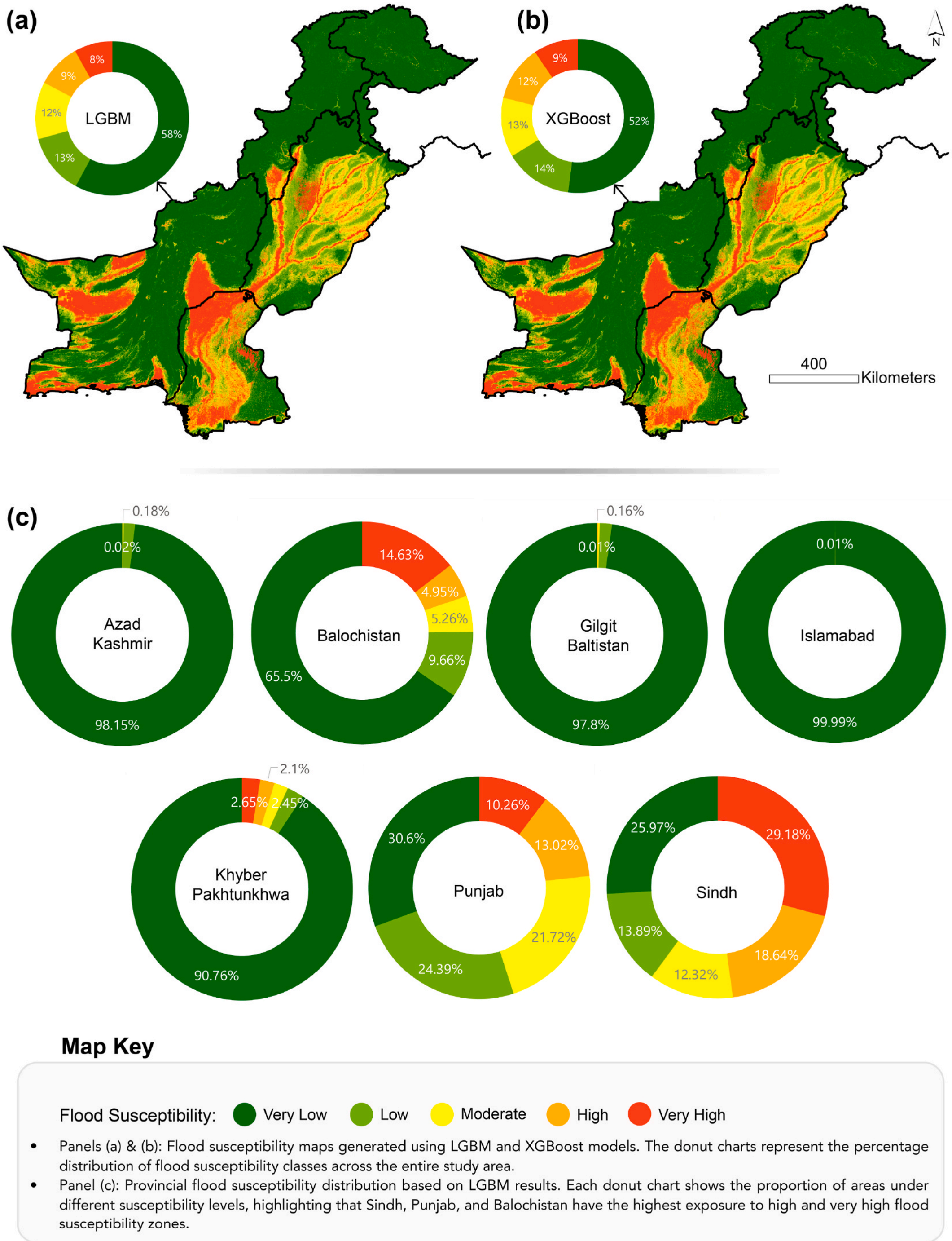

Fig. 8. Millions of peoples exposed to each flood susceptibility class in Pakistan and in each province.

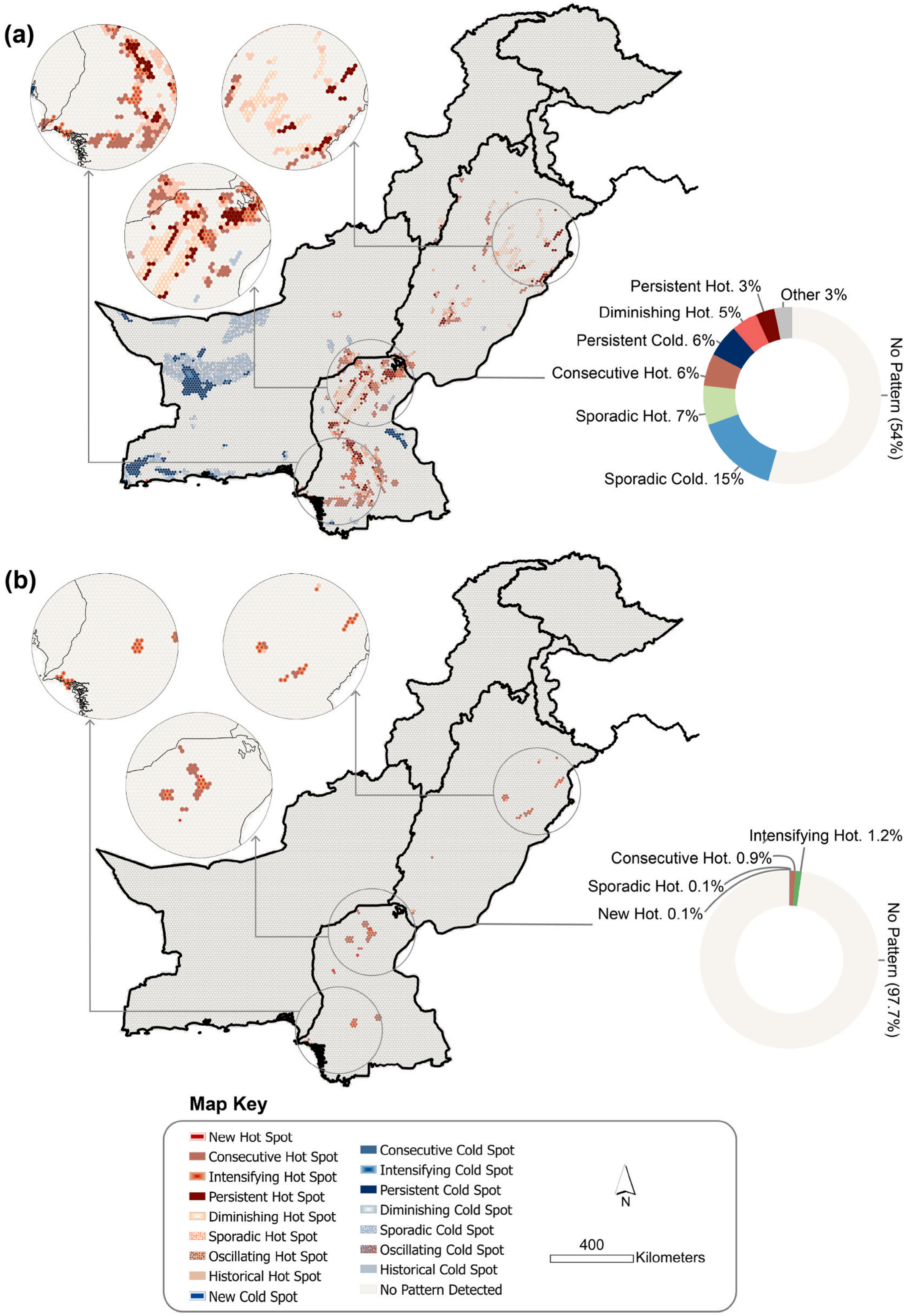

Fig. 9. Emerging hotspot analysis for very high flood class and (a) NTL and (b)total population.

people in Pakistan (~47 % via combining very high, high, and moderate susceptibility zones) are exposed to critical flood susceptibility levels, which is nearly half of the total population of in the country. Lastly, low and very low flood susceptibility classes consist of 37.3 million (~18.6 %) and 67.5 million (~33.7 %) people respectively.

From a regional (sub-national) perspective, Balochistan, Punjab, and Sindh show the highest population exposure to higher flood susceptibility levels, where Khyber Pakhtunkhwa (KPK), Gilgit (Gilgit Baltistan), Islamabad, and Kashmir are the least exposed. In Balochistan, approximately 1.9 million people are exposed to very high flood susceptibility, 0.5 million are residing in high susceptibility, and 0.5 million in moderate susceptibility. Combining all three account for ~3 million (~34 % in the province) people exposed to critical flood susceptibility levels. For Punjab, around 9.9 million people reside in the very high susceptibility zone, 17.4 million in high susceptibility zone, and 31.8 million in moderate susceptibility zone. Thus, combined estimation accounts for an overall of 59.1 million people (~54 % in the province) exposed to critical flood susceptibility levels. Sindh on the other hand, consists of the highest of 15.4 million people living in the very high, 9.6 million people in high, and 6.3 million people living in moderate flood susceptibility zones, respectively. Combined, ~31.3 million people (~74 % in the province) in Sindh fall under critical flood susceptibility levels. For the other four provinces including KPK, Gilgit, Islamabad, and Kashmir, the exposure of population to very high, high, and moderate is found to be < 1 million people.

3.5. Population and economic activity exposure profiles of flood susceptibility

The population and economic activity exposure profiles are estimated using the emerging hotspot identification. This spatial-temporal hotspot identification procedure identifies 17 different emerging classes in the data (Fig. 9). Fig. 9a drills down for the emerging hotspot patterns based on the very high flood susceptibility class and NTL, where Fig. 9b illustrates the emerging hotspot patterns based on the very high flood susceptibility class and total population. Several hot and cold spots are evident for the economic activity (Fig. 9a), with hotspots primarily located in Sindh and Punjab, and cold spots in Balochistan. Overall, the pattern mining approach identify Sporadic (7 %), Consecutive (6 %), Diminishing (5 %), and persistent (3 %) hot spots in Pakistan. These hot spots show that in the last two decades, a large proportion of high economic activity has been experienced in the regions with very high flood susceptibility. Comparatively, the pattern mining approach in Fig. 9a identifies cold spots including Persistent cold (6 %), and Sporadic cold spots (15 %). These cold spots show that during the last two decades, some areas (primarily in Balochistan) exist with very high flood (~>95 % confidence) but with lower economic activity throughout the time span. Moreover, Fig. 9a shows the majority of Sindh in Persistent and Consecutive hot spot, implying that the high economic activity areas in Sindh lies under very high flood susceptibility with 95 % confidence. In Punjab, Persistent, Diminishing, and Sporadic hot spots are identified, especially around major cities like Lahore and Faisalabad, the cities well known for their economic and industrial status. These hot spots show that major cities in Punjab sustaining economic activity falls within the very high flood susceptibility zones, with over 95 % confidence in the last 20 years. In Fig. 9b, only a few clusters of patterns are identified, located in Sindh and Punjab due to the fact that only a very high flood susceptibility class is used to assess the predominant patterns in population. This also implies that, in the very high flood zones, northern regions of Sindh and Punjab constitute major proportions of their populations. In terms of emerging patterns, only hot spots are identified, namely Intensifying (1.2 %), Consecutive (0.9 %), Sporadic (0.1 %), and New (0.1 %) hot spots.

4. Discussion

In recent years, the need for data addressing DRM has emerged, and a special focus is given to high resolution and highly accurate data products. For floods, the data regarding flood susceptibility and exposure are crucial to assist in on-time disaster preparedness, monitoring, and prevention activities [6,7]. However, the development of models able to identify flood prone regions efficiently and accurately is a challenge, especially when focusing on higher-resolution and larger scales. Where existing studies focus on providing flood susceptibility or exposure information in small scale regions (e.g., Ref. [17,22,41]), this study developed a ML-based framework capable of performing large scale (national) flood susceptibility, with relatively higher efficiency and accuracy documented in earlier studies. To demonstrate that, we use Pakistan as a case study, a country that has faced alarming consequences as a result of rapid flood cases in recent decades due to its geographical location, weak infrastructure, lack of resilience and resources to support DRM activities, and lastly, unavailability of existing accurate data on flood management at the national level [13,34].

Our framework integrates scalable machine learning models, specifically XGBoost and LGBM, offering significant advancements over traditional statistical and hydrological FSM methods, which often face challenges with computational demands and accuracy at larger scales. By leveraging the efficiency of these gradient-boosting models, the framework provides high-resolution FSM at a national scale, ensuring consistent and reliable performance. The integration of Explainable AI tools further enhances the transparency and interpretability of model outputs, addressing key limitations associated with both conventional methods and simpler ML approaches like Random Forest. Building on prior work [24], which assessed the scalability and accuracy of various models, this study applies XGBoost and LGBM at a national scale and evaluates their performance. Both models exhibit robust and sustained efficiency, with LGBM demonstrating a slight edge over XGBoost in terms of adjusted accuracy (0.85 vs. 0.82) and correctly classified flood samples, as shown in the confusion matrix. The superior performance of the LGBM over XGBoost can be attributed to its unique leaf-wise splitting strategy, which allows it to model complex patterns more efficiently than the level-wise approach used in XGBoost. Additionally, LGBM is optimized for handling sparse data and categorical features, enabling it to achieve faster convergence with reduced overfitting risks [32]. These advantages make LGBM particularly suitable for flood susceptibility modeling, where high-resolution datasets and diverse terrain conditions require precise and scalable solutions. The resulting findings highlight the adaptability of the proposed ML-based framework for large-scale applications, enabling accurate, high-resolution insights into flood-prone areas and exposure disparities across diverse geographic contexts, while maintaining computational efficiency.

The high-resolution national scale FSM findings (Fig. 5) in Pakistan showed concerning outcomes, with larger population zones falling under very high, high, and moderate flood susceptibility zones. Overall, results from this study based on the LGBM model (Fig. 5a) showed that 29 % of the total area of Pakistan falls under critical flood susceptibility levels (here, critical flood susceptibility levels consist of combining very high, high, and moderate levels). In comparison, XGBoost predicted 34 % of the area falling withing critical flood susceptibility levels (Fig. 5b). At the provincial levels (Fig. 5c), Sindh showed the highest distribution of flood susceptibility levels, with nearly 29 % of the area very high, ∼19 % of the area high, and ∼12 % of the area in moderate flood zone. Overall, ∼60 % of the Sindh area falls under critical flood levels, which calls for appropriate measures (i.e., building and enhancing resilience) to avoid future impacts of flooding. Punjab stood second, with ∼10 % area as very high, ∼13 % area as high, and ∼22 % as moderate flood susceptibility levels, combined ∼45 % of the area in Punjab falls under critical flood susceptibility levels. For Balochistan, ∼15 % of the area falls under very high, ∼5 % of area in high, and ∼5 % in moderate, resulting in a total of 25 % of the area of the province falling under critical flood susceptibility levels. Given these findings, these three provinces namely Sindh, Punjab, and Balochistan are of utmost importance and among them, the critical flood susceptible areas are of highest priority when addressing any disaster management activity in these locations.

The bivariate visualizations (Fig. 7) show the overlapping of various groups with flood susceptibility levels, thus reflecting the core regions affected. Two main groups (i.e., NTL and Population) overlap with FSM layer, where the NTL shows the core economic zones falling under flood levels, whereas population shows the exposed population. From Fig. 7, most of the areas in Sindh and Punjab lie under high intensity flood susceptibility levels, showing highly exposed economic zones as well as exposed population. Moreover, the vulnerable groups (% Elderly, % Non-Adults, and % Vulnerable) showed higher concentration levels in lower Punjab and upper Sindh regions.

The results of population exposure (Fig. 8) indicate that in Pakistan, ∼14 % (∼28 millions) of the people are exposed to the highest (very high) level of flood susceptibility, ∼14 % (∼28 millions) to the high level of flood susceptibility, whereas ∼20 % (∼39 millions) of the population falls under moderate flood susceptibility class. Combined, this study revealed that ∼47 % (∼95 millions) of the total population in Pakistan are exposed to critical levels of flood susceptibility class. Furthermore, at provincial level, three provinces namely Balochistan, Punjab and Sindh have the highest population exposure levels, compared to KPK, Gilgit Baltistan, Islamabad, and Kashmir. In terms of population exposed to critical flood levels, Sindh stood at top with ∼74 %, Punjab at second with 54 %, and Balochistan at third with 33 % of the total population exposed to critical flood susceptibility levels.

With that, the population exposure at provincial scale also emphasizes the crucial need for a targeted approach in disaster management. These findings have significant implications for disaster preparedness and response activities in core affected areas. For instance, critical level of population exposure in Sindh, Punjab, and Balochistan accentuate the need for improved infrastructure, development of flood emergency protocols as well as early warning systems, and better policies for long-term flood resilience. Besides, it is imperative to establish a strong flood management system, with strategies focusing on key elements like flood resilient buildings, better evacuation plans, and resource allocation in flood emergencies [34]. Lastly, our findings further signify the importance of social flood management awareness especially in the population of critically susceptible areas, thus promoting community resilience through effective training and education programs.

The identified hotspots (in Fig. 9) are predominantly located in upper and lower Sindh, and upper Punjab, and are classified as Sporadic, Consecutive, Diminishing, and Persistent hotspots, thus, signifying the different levels of flood susceptibility over the past two decades. Fig. 9a also reveals the presence of cold spots, primarily in Balochistan, where high flood susceptibility persists with lower economic activity. Notably, a significant portion of Sindh is identified as Persistent and Consecutive hotspots, suggesting that areas with high economic activity in Sindh are consistently exposed to a high risk of flooding. Similarly, in Punjab, the hotspots indicate areas with a high concentration of population in the very high flood zones. Fig. 9b, focuses on the relationship between the very high flood class and total population, identifies only a few clusters of patterns primarily in Sindh and Punjab. This suggests that the upper regions of Sindh and Punjab have a significant proportion of their populations residing in areas with a high susceptibility to flooding. The emerging patterns in this analysis mostly consist of hotspots, including Intensifying, Consecutive, Sporadic, and New hotspots, indicating the dynamic nature of flood exposure in these areas. Overall, these findings highlight the critical need for targeted flood mitigation and adaptation strategies in the identified hotspots to minimize the potential socio-economic impacts of flooding in Pakistan.

In terms of population exposure estimates, we contextualize our findings recent estimates by other studies. For instance, Rentschler et al. [2] recently presented population exposed to global floods and ranks Pakistan as the fifth country with nearly 71.8 million (31.1 %) people exposed to high levels of flood. Whereas our study identified ∼94.7 million (47 %) people exposed to critical level of floods. The difference in the estimated population exposure might be attributed to the fact that our study generated high resolution FSM data product at 30 m spatial resolution while their estimates are based on a coarser assessment. Furthermore, our study used state of the art ML modelling approach and multiple environmental factors acting as FCFs, resulting in finer mapping of susceptibility in Pakistan. Furthermore, the study by Rentschler et al. [2] showed Sindh and Punjab provinces in Pakistan with the highest population exposure, which aligns with our findings of Sindh with ∼72 % and Punjab with 54 % of the population exposed to critical flood susceptibility levels.

4.1. Fostering flood susceptibility-informed policy implications

By providing detailed and accurate information regarding areas at risk of flooding, the data can aid policymakers, disaster management authorities, and local governments in identifying vulnerable populations and critical infrastructures that require immediate attention and intervention. This data is essential in the strategic planning of mitigation measures, targeted early warning systems, and effective evacuation plans, ultimately reducing potential human and economic losses. The integration of this data into national disaster risk management policies and provincial frameworks can enable targeted interventions that address both immediate and long-term flood risks. Additionally, this national scale high resolution FSM and exposure data can support the efficient allocation of resources to ensure that flood defenses, infrastructure resilience, and community preparedness investments are directed to where they are most needed. Incorporating high-resolution flood susceptibility and exposure data into disaster risk management strategies can significantly enhance proactive flood risk management in Pakistan. This includes supporting urban planning by enforcing zoning regulations and guiding flood-resistant infrastructure development, as well as aiding climate adaptation initiatives to mitigate vulnerabilities in flood-prone regions. Besides, this research contributes to the broader field of disaster risk management by enhancing our understanding of the complex interactions between geographical factors, climate change, and flood hazards. This knowledge can inform the development of robust early warning systems, evacuation plans, and infrastructure resilience strategies, ultimately saving lives and minimizing economic losses in the face of future flood events in Pakistan.

The findings of this study provide actionable insights for policymakers and disaster management authorities. The identification of critical flood susceptibility zones, such as the ~62 % of Sindh, ~45 % of Punjab, and ~25 % of Balochistan under critical flood risk, underscores the need for targeted interventions in these regions. For instance, resources can be prioritized for developing flood-resilient infrastructure, including levees and drainage systems, in areas with very high susceptibility. Moreover, early warning systems tailored to local vulnerabilities can be established to ensure timely evacuation and risk communication. In addition to physical infrastructure improvements, these results also highlight the importance of region-specific zoning regulations. By restricting development in high-risk areas and encouraging sustainable land-use practices, policymakers can reduce exposure to floods and associated economic losses. Furthermore, understanding the spatial distribution of flood risks can guide long-term climate adaptation strategies, such as reforestation and the restoration of natural floodplains, particularly in areas like Sindh and Punjab with significant agricultural and economic activity.

The methodology and insights developed in this study can also serve as a template for other flood-prone regions globally, enabling scalable and data-driven policy solutions. Pakistan can address the increased risk of floods by operationalizing and empowering climate change institutions, adopting appropriate tools and frameworks, and benchmarking global best practices. Additionally, these findings have practical utility in financial risk assessments, allowing insurers to refine flood coverage policies and better allocate resources to mitigate economic losses.

4.2. Limitations and the way forward

While this study performed the first national scale high resolution FSM based on ML models, it is equally imperative to discuss the shortcomings and future directions. For instance, it is important to document the sensitivity of the results presented in this study, particularly in relation to how flood susceptibility is classified, and exposure is estimated. The study uses five flood susceptibility thresholds to categorize the level of flood, and then analyses exposure risk for exposed populations and economic areas using population and NTL data respectively. It is worth noting that the results are prepared with great attention to sensitivity. For example, this study uses a quantile-based distribution to classify the continuous flood susceptibility into five classes, each representing 20 % of the total FSM distribution. The categories depict varying levels of flood intensity, with the very high FSM class representing FSM levels between 80 and 100 % flood probabilities. Changing the distribution function could potentially impact the output FSM, which is sensitive to the distribution method.

While the incorporation of EAI significantly enhances model transparency and fosters understanding of the factors influencing predictions, it is crucial to recognize that EAI techniques are not free from context-dependent biases. For instance, the spatial variability of input data and differences in regional characteristics may influence the reliability of feature importance rankings. Moreover, the assumptions inherent in EAI algorithms can affect the consistency of interpretations, particularly when applied to heterogeneous study areas. These limitations necessitate a cautious approach to interpreting EAI outputs, emphasizing the importance of complementing EAI-driven insights with domain knowledge to mitigate potential biases and misinterpretations.

This study integrates socioeconomic indicators such as population data from WorldPop and NTL data as proxies for economic activity to provide a nuanced assessment of flood exposure. These indicators enable us to identify vulnerable population groups, including non-adults, the elderly, and other categories, as well as regions with significant economic activity at risk. Through such data, our framework advances beyond traditional physical and climatic determinants, offering valuable insights into the human and economic dimensions of flood exposure. Nonetheless, we recognize that additional environmental and social variables, such as land-use practices, governance quality, and ecosystem integrity, could further enhance the assessment's comprehensiveness. For instance, land-use changes, such as deforestation and urban sprawl, can exacerbate flood risks, while governance factors like disaster preparedness and resource allocation influence adaptive capacities. Future research could integrate such variables to develop a more holistic framework for flood susceptibility and exposure assessment. Therefore, through combining physical, social, and environmental factors, multi-disciplinary approaches could improve the predictive accuracy of models and their relevance for policy interventions. Moreover, indicators such as official income statistics, industrial output, or census-based data—can be integrated to cross-validate and refine NTL measurements, yielding a more robust representation of economic exposure in flood-prone regions.

Vulnerability to flooding is a multi-faceted concept that extends beyond mere exposure to flood-prone areas. According to frameworks proposed by Adger [42] and IPCC [43], vulnerability is determined by the interplay of exposure, sensitivity, and adaptive capacity. In this study, vulnerability was primarily assessed through exposure-based indicators such as population groups and economic activity. While non-adults and elderly populations were considered highly vulnerable due to their increased dependency and potential mobility constraints, it is acknowledged that factors such as socio-economic conditions, health infrastructure, and community resilience also play critical roles in determining vulnerability. Future studies should consider integrating additional sensitivity and adaptive capacity factors to provide a more holistic vulnerability assessment. Furthermore, NTL data was utilized as a proxy for economic activity to identify highly exposed economic zones. While NTL provides valuable insights into economic distribution and infrastructure density, its limitations must be acknowledged, particularly in rural or underdeveloped areas where economic activity may not correlate directly with illumination. Future studies could incorporate other socio-economic datasets, such as income levels, employment rates, or industrial zones, to enhance the accuracy of exposure assessments.

This study also acknowledges several ethical and practical considerations. High-resolution flood susceptibility data, while valuable for disaster management, carries risks of potential misuse, such as inequitable resource allocation or discriminatory practices. To address this, it is crucial to promote transparent and inclusive decision-making processes to ensure equitable benefits for all, particularly vulnerable populations. Additionally, relying on historical flood data poses limitations as it may not capture evolving patterns driven by land-use changes and climate change, underscoring the need for integrating real-time monitoring and updated datasets. Moreover, the presented framework is highly adaptable and can be applied to other developing countries with similar vulnerabilities. Many nations face challenges like data scarcity, limited computational resources, and diverse topographies, making this scalable and resource-efficient methodology suitable for global replication. By leveraging open-source tools and globally accessible datasets, the framework ensures accessibility for regions with limited resources. Although this study does not explicitly incorporate climate change scenarios, their integration is a crucial step for enhancing long-term flood risk assessments. Future research can combine our methodology with projections from climate models, such as CMIP6 or ERA5 datasets, to analyze changes in precipitation patterns, river discharges, and extreme weather events. Incorporating such scenarios would provide policymakers with actionable insights to prepare for the anticipated intensification of flood risks under global warming.

Lastly, this study initially utilized GEE for variable preparation, and data preprocessing. For this case, the GEE done a great job, and the platform easily managed to process data within minutes. However, due to GEE platform limitations on ML models availability, as well as the pixel processing allowance per query, this study developed a localized framework to manually export all variables locally, again converting them into numerical format readable by ML models, and lastly by using localized ML models for training and prediction. This situation, however, can be improved in Future with further optimizing the existing FSM framework with GEE available ML models (i.e. Random Forest) or either by incorporating the more efficient models (LGBM or XGBoost) in GEE. Furthermore, the GEE also allows users to purchase additional computational units based on their needs, which can easily overcome computational deficiency needs.

5. Conclusions

Large scale FSM allows valuable insights, which are crucial for effective disaster preparedness and response activities. However, due to scarcity of data in developing countries and lack of availability of large computational clusters, existing studies relied on smaller scale FSM assessments. This study developed a ML-based FSM framework, capable of performing national scale FSM with highest possible efficiency. Pakistan being one of the countries affected by floods, lack data to support flood management. In this regard, this study used the newly developed FSM framework to generate flood susceptibility maps for Pakistan. Besides, population and economic exposure analysis were also carried out to highlight core affected regions in the country. The results of this study provide valuable insights into the population exposure to flood risks in different provinces of the country. With an adjusted accuracy of 0.85, LGBM model surpasses XGBoost. The FSM maps in Pakistan shows 8 % of the country falling into very high flood susceptibility level, while 9 % falling into high susceptibility level, with majority of FSM patterns around northern Sindh, and alongside river network in Punjab. The exposure analysis showed alarming results, with over 47 % of the total population of Pakistan falling under critical flood susceptibility levels. For provinces, Sindh ranks at top with 74 % of the population under critical flood susceptibility levels, whereas Punjab and Balochistan stood second and third with 54 % and 33 % respectively. Lastly, the results of space-time pattern mining analysis also rank Sindh and Punjab as mostly exposed, with several economic hotspots located in Sindh and upper Punjab, highlighting a larger portion of economic activity in those areas falling under very high flood susceptibility levels. On the contrary, the population vs flood susceptibility identified very few patterns, mostly around upper Sindh and Punjab.

These FSM and flood exposure patterns imply that major proportions of population and economic activity areas in these regions fall within very high flood susceptibility levels. The study directly addresses the research gap by offering a scalable, high-resolution flood susceptibility framework tailored to the unique geographic and socio-economic conditions of Pakistan. By leveraging machine learning and explainable AI, the framework provides a comprehensive and data-driven approach to flood risk management, surpassing the limitations of existing low-resolution assessments. The findings emphasize the critical need for policymakers to prioritize interventions in high-risk areas, such as Sindh and Punjab, by investing in flood-resilient infrastructure, enhancing early warning systems, and enforcing land-use regulations to mitigate risks effectively. Additionally, the integration of high-resolution FSM data into national disaster management policies can facilitate better planning, resource allocation, and proactive measures to build long-term resilience against future flood events. The findings emphasize the urgency of addressing the vulnerabilities in highly exposed regions, such as Sindh and Punjab, while also highlighting the importance of maintaining preparedness in areas with lower exposure. By implementing appropriate measures and fostering a resilience-based disaster planning, Pakistan can enhance its capacity to respond to floods effectively and minimize the adverse impacts in the future.

CRediT authorship contribution statement

Mirza Waleed: Writing – review & editing, Writing – original draft, Visualization, Software, Formal analysis, Data curation, Conceptualization. **Muhammad Sajjad:** Writing – review & editing, Visualization, Supervision, Project administration, Methodology, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Sajjad M. is funded by the HKBU Research Grants Committee (Start-up Grant-Tier 1, RC-STARTUP/21–22/12), Hong Kong SAR. Waleed M. is supported by a postgraduate studentship from the HKBU Research Grant Committee (PhD studentship, 2022–2026). We are thankful to all the institutes (mentioned within the text) for the provisioning of relevant data to carry out this valuable study. The research is conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest. All the data used for several analyses are freely available and the resources are mentioned within the paper. The data resulting from this research will be made available through the project GitHub repository: https://github.com/waleedgeo/FSM-PK.

Data availability

All the data are available freely and the sources are mentioned in the paper.

References

[1] B. Merz, G. Blöschl, S. Vorogushyn, F. Dottori, J.C.J.H. Aerts, P. Bates, M. Bertola, M. Kemter, H. Kreibich, U. Lall, E. Macdonald, Causes, impacts and patterns of disastrous river floods, Nat. Rev. Earth Environ. 2 (2021) 592–609, https://doi.org/10.1038/s43017-021-00195-3.
[2] J. Rentschler, M. Salhab, B.A. Jafino, Flood exposure and poverty in 188 countries, Nat. Commun. 13 (2022) 3527, https://doi.org/10.1038/s41467-022-30727-4.
[3] L. Devitt, J. Neal, G. Coxon, J. Savage, T. Wagener, Flood hazard potential reveals global floodplain settlement patterns, Nat. Commun. 14 (2023) 2801, https://doi.org/10.1038/s41467-023-38297-9.
[4] M. Sajjad, Disaster resilience in Pakistan: a comprehensive multi-dimensional spatial profiling, Appl. Geogr. 126 (2021) 102367, https://doi.org/10.1016/j.apgeog.2020.102367.
[5] Jian Fang, C. Zhang, Jiayi Fang, M. Liu, Y. Luan, Increasing exposure to floods in China revealed by nighttime light data and flood susceptibility mapping, Environ. Res. Lett. 16 (2021) 104044, https://doi.org/10.1088/1748-9326/ac263e.
[6] A. Smith, P.D. Bates, O. Wing, C. Sampson, N. Quinn, J. Neal, New estimates of flood exposure in developing countries using high-resolution population data, Nat. Commun. 10 (2019) 1814, https://doi.org/10.1038/s41467-019-09282-y.
[7] T.K.J. McDermott, Global exposure to flood risk and poverty, Nat. Commun. 13 (2022) 3529, https://doi.org/10.1038/s41467-022-30725-6.
[8] World Bank, Pakistan: Flood Damages and Economic Losses over USD 30 Billion and Reconstruction Needs over USD 16 Billion - New Assessment, 2022.
[9] D. Eckstein, V. Künzel, L. Schäfer, The Global Climate Risk Index 2021, Germanwatch, Bonn, 2021.
[10] P. Waidelich, F. Batibeniz, J. Rising, J.S. Kikstra, S.I. Seneviratne, Climate damage projections beyond annual temperature, Nat. Clim. Change 1–8 (2024), https://doi.org/10.1038/s41558-024-01990-8.
[11] M. Sajjad, Z. Ali, M. Waleed, Has Pakistan learned from disasters over the decades? Dynamic resilience insights based on catastrophe progression and geo-information models, Nat. Hazards (2023), https://doi.org/10.1007/s11069-023-05976-1.
[12] M.A.U.R. Tariq, N. van de Giesen, Floods and flood management in Pakistan, Phys. Chem. Earth (2012) 11–20, https://doi.org/10.1016/j.pce.2011.08.014. Parts A/B/C 47–48.
[13] H.B. Waseem, I.A. Rana, Floods in Pakistan: a state-of-the-art review, Natural Hazards Research (2023), https://doi.org/10.1016/j.nhres.2023.06.005.
[14] S.T. Seydi, Y. Kanani-Sadat, M. Hasanlou, R. Sahraei, J. Chanussot, M. Amani, Comparison of machine learning algorithms for flood susceptibility mapping, Remote Sens. 15 (2023) 192, https://doi.org/10.3390/rs15010192.
[15] G. Zhao, B. Pang, Z. Xu, D. Peng, L. Xu, Assessment of urban flood susceptibility using semi-supervised machine learning model, Sci. Total Environ. 659 (2019) 940–949, https://doi.org/10.1016/j.scitotenv.2018.12.217.
[16] T. Gudiyangada Nachappa, S. Tavakkoli Piralilou, K. Gholamnia, O. Ghorbanzadeh, O. Rahmati, T. Blaschke, Flood susceptibility mapping with machine learning, multi-criteria decision analysis and ensemble using Dempster Shafer Theory, J. Hydrol. 590 (2020) 125275, https://doi.org/10.1016/j.jhydrol.2020.125275.
[17] S. Saravanan, D. Abijith, Flood susceptibility mapping of Northeast coastal districts of Tamil Nadu India using Multi-source Geospatial data and Machine Learning techniques, Geocarto Int. (2022) 1–30, https://doi.org/10.1080/10106049.2022.2096702.
[18] M.Z. Serdar, S.B. Ajjur, S.G. Al-Ghamdi, Flood susceptibility assessment in arid areas: a case study of Qatar, Sustainability 14 (2022) 9792, https://doi.org/10.3390/su14159792.
[19] I.A.R.M. Towfiqul, S. Talukdar, S. Mahato, S. Kundu, K.U. Eibek, Q.B. Pham, A. Kuriqi, N.T.T. Linh, Flood susceptibility modelling using advanced ensemble machine learning models, Geosci. Front. 12 (2021) 101075, https://doi.org/10.1016/j.gsf.2020.09.006.
[20] B.T. Pham, C. Luu, T.V. Phong, P.T. Trinh, A. Shirzadi, S. Renoud, S. Asadi, H.V. Le, J. von Meding, J.J. Clague, Can deep learning algorithms outperform benchmark machine learning algorithms in flood susceptibility modeling? J. Hydrol. 592 (2021) 125615 https://doi.org/10.1016/j.jhydrol.2020.125615.
[21] B. Pradhan, S. Lee, A. Dikshit, H. Kim, Spatial flood susceptibility mapping using an explainable artificial intelligence (XAI) model, Geosci. Front. 14 (2023) 101625.
[22] O. Seleem, G. Ayzel, A.C.T. de Souza, A. Bronstert, M. Heistermann, Towards urban flood susceptibility mapping using data-driven models in Berlin, Germany, Geomatics Nat. Hazards Risk 13 (2022) 1640–1662, https://doi.org/10.1080/19475705.2022.2097131.
[23] C. Hao, A.P. Yunus, S. Siva Subramanian, R. Avtar, Basin-wide flood depth and exposure mapping from SAR images and machine learning models, J. Environ. Manag. 297 (2021) 113367, https://doi.org/10.1016/j.jenvman.2021.113367.
[24] M. Waleed, M. Sajjad, Advancing flood susceptibility prediction: a comparative assessment and scalability analysis of machine learning algorithms via artificial intelligence in high-risk regions of Pakistan, Journal of Flood Risk Management 18 (2025) e13047, https://doi.org/10.1111/jfr3.13047.

[25] J.A. Cardille, M.A. Crowley, D. Saah, N.E. Clinton (Eds.), Cloud-Based Remote Sensing with Google Earth Engine: Fundamentals and Applications, Springer International Publishing, Cham, 2024, https://doi.org/10.1007/978-3-031-26588-4.
[26] N. Gorelick, Google earth engine, in: EGU General Assembly Conference Abstracts, American Geophysical Union Vienna, Austria, 2013 11997.
[27] L. Yang, J. Driscol, S. Sarigai, Q. Wu, H. Chen, C.D. Lippitt, Google earth engine and artificial intelligence (AI): a comprehensive review, Remote Sens. 14 (2022) 3253, https://doi.org/10.3390/rs14143253.
[28] L. Hawker, P. Uhe, L. Paulo, J. Sosa, J. Savage, C. Sampson, J. Neal, A 30 m global map of elevation with forests and buildings removed, Environ. Res. Lett. 17 (2022) 024016, https://doi.org/10.1088/1748-9326/ac4d4f.
[29] J.D. Tarpley, S.R. Schneider, R.L. Money, Global vegetation indices from the NOAA-7 meteorological satellite, J. Clim. Appl. Meteorol. 23 (1984) 491–494.
[30] P.-E. Danielsson, Euclidean distance mapping, Comput. Graph. Image Process. 14 (1980) 227–248.
[31] G. Ali, M. Sajjad, S. Kanwal, T. Xiao, S. Khalid, F. Shoaib, H.N. Gul, Spatial–temporal characterization of rainfall in Pakistan during the past half-century (1961–2020), Sci. Rep. 11 (2021) 6935, https://doi.org/10.1038/s41598-021-86412-x.
[32] E. Al Daoud, Comparison between XGBoost, LightGBM and CatBoost using a home credit dataset, International Journal of Computer and Information Engineering 13 (2019) 6–10.
[33] M. Waleed, M. Sajjad, M.S. Shazil, Urbanization-led land cover change impacts terrestrial carbon storage capacity: a high-resolution remote sensing-based nation-wide assessment in Pakistan (1990–2020), Environ. Impact Assess. Rev. 105 (2024) 107396, https://doi.org/10.1016/j.eiar.2023.107396.
[34] J.S. Nanditha, A.P. Kushwaha, R. Singh, I. Malik, H. Solanki, D.S. Chuphal, S. Dangar, S.S. Mahto, U. Vegad, V. Mishra, The Pakistan flood of august 2022: causes and implications, Earths Future 11 (2023) e2022EF003230, https://doi.org/10.1029/2022EF003230.
[35] E. Dodangeh, B. Choubin, A.N. Eigdir, N. Nabipour, M. Panahi, S. Shamshirband, A. Mosavi, Integrated machine learning methods with resampling algorithms for flood susceptibility prediction, Sci. Total Environ. 705 (2020) 135983, https://doi.org/10.1016/j.scitotenv.2019.135983.
[36] M. Nhangumbe, A. Nascetti, S. Georganos, Y. Ban, Supervised and unsupervised machine learning approaches using Sentinel data for flood mapping and damage assessment in Mozambique, Remote Sens. Appl.: Society and Environment (2023) 101015, https://doi.org/10.1016/j.rsase.2023.101015.
[37] S. Hitouri, A. Varasano, M. Mohajane, S. Ijlil, N. Essahlaoui, S.A. Ali, A. Essahlaoui, Q.B. Pham, M. Waleed, S.K. Palateerdham, A.C. Teodoro, Hybrid machine learning approach for gully erosion mapping susceptibility at a watershed scale, IJGI 11 (2022) 401, https://doi.org/10.3390/ijgi11070401.
[38] Y. Xu, K. Lin, C. Hu, X. Chen, J. Zhang, M. Xiao, C.-Y. Xu, Uncovering the dynamic drivers of floods through interpretable deep learning, Earths Future 12 (2024) e2024EF004751, https://doi.org/10.1029/2024EF004751.
[39] X. Li, Y. Zhou, M. Zhao, X. Zhao, A harmonized global nighttime light dataset 1992–2018, Sci. Data 7 (2020) 168, https://doi.org/10.1038/s41597-020-0510-y.
[40] A.J. Tatem, WorldPop, open data for spatial demography, Sci. Data 4 (2017) 1–4.
[41] O. Rahmati, H. Darabi, M. Panahi, Z. Kalantari, S.A. Naghibi, C.S.S. Ferreira, A. Kornejady, Z. Karimidastenaei, F. Mohammadi, S. Stefanidis, D. Tien Bui, A. T. Haghighi, Development of novel hybridized models for urban flood susceptibility mapping, Sci. Rep. 10 (2020) 12937, https://doi.org/10.1038/s41598-020-69703-7.
[42] W.N. Adger, Vulnerability. Global environmental change, resilience, vulnerability, and adaptation: a cross-cutting theme of the international human dimensions programme on, Glob. Environ. Change 16 (2006) 268–281, https://doi.org/10.1016/j.gloenvcha.2006.02.006.
[43] IPCC, AR4 Climate Change 2007: Impacts, Adaptation, and Vulnerability — IPCC, 2007.
[44] B. Lehner, G. Grill, Global river hydrography and network routing: baseline data and new approaches to study the world's large river systems, Hydrol. Process. 27 (2013) 2171–2186, https://doi.org/10.1002/hyp.9740.
[45] M. Cook, J.R. Schott, J. Mandel, N. Raqueno, Development of an operational calibration methodology for the Landsat thermal data archive and initial testing of the atmospheric compensation component of a land surface temperature (LST) product from the archive, Remote Sens. 6 (2014) 11244–11266, https://doi.org/10.3390/rs61111244.
[46] Z. Akhtar, U. Qazi, R. Sadiq, A. El-Sakka, M. Sajjad, F. Ofli, M. Imran, Mapping flood exposure, damage, and population needs using remote and social sensing: a case study of 2022 Pakistan floods. https://doi.org/10.1145/3543507.3583881, 2023.
[47] C. Funk, P. Peterson, M. Landsfeld, D. Pedreros, J. Verdin, S. Shukla, G. Husak, J. Rowland, L. Harrison, A. Hoell, J. Michaelsen, The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes, Sci. Data 2 (2015) 150066, https://doi.org/10.1038/sdata.2015.66.
[48] P. Olofsson, G.M. Foody, M. Herold, S.V. Stehman, C.E. Woodcock, M.A. Wulder, Good practices for estimating area and assessing accuracy of land change, Rem. Sens. Environ. 148 (2014) 42–57, https://doi.org/10.1016/j.rse.2014.02.015.
[49] M. Waleed, M. Sajjad, Leveraging cloud-based computing and spatial modeling approaches for land surface temperature disparities in response to land cover change: evidence from Pakistan, Remote Sens. Appl.: Society and Environment 25 (2022) 100665, https://doi.org/10.1016/j.rsase.2021.100665.
[50] A.E. Maxwell, T.A. Warner, L.A. Guillén, Accuracy assessment in convolutional neural network-based deep learning remote sensing studies—part 1: literature review, Remote Sens. 13 (2021) 2450.
[51] C. Chen, B. Hu, Y. Li, Easy-to-use spatial random-forest-based downscaling-calibration method for producing precipitation data with high resolution and high accuracy, Hydrol. Earth Syst. Sci. 25 (2021) 5667–5682, https://doi.org/10.5194/hess-25-5667-2021.
[52] S. Moradian, A. AghaKouchak, S. Gharbia, C. Broderick, A.I. Olbert, Forecasting of compound ocean-fluvial floods using machine learning, J. Environ. Manag. 364 (2024) 121295, https://doi.org/10.1016/j.jenvman.2024.121295.