Remote Sensing Applications: Society and Environment 33 (2024) 101113

Contents lists available at ScienceDirect

Remote Sensing Applications: Society and Environment

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

Dynamics of waterlogging and drainage congestion on the Kosi Fan, Himalayan Foreland

M. Niranjannaik, Kumar Gaurav *, Abhilash Singh, Amrit Kumar Singh

Fluvial Geomorphology and Remote Sensing Laboratory, Department of Earth and Environmental Sciences, Indian Institute of Science Education and Research Bhopal, Bypass road, Bhopal, 462066, Madhya Pradesh, India

ARTICLE INFO

Keywords:
Drainage congestion
Waterlogging
SNIC-RF
Anthropogenic interventions
Road-rail network
Sustainable cities

ABSTRACT

This study assesses the dynamics of waterlogging using time series optical satellite images from 1987 to 2021 on alluvial fan of the Kosi River in the Himalayan Foreland. We classified the satellite images to extract waterlogging patches by hybridising Simple Non-Iterative Clustering (SNIC) segmentation and Random Forest (RF) algorithms. This hybrid framework can classify the waterlogged patches from satellite images with overall accuracy ranging between 75%–90%. We observed that the waterlogging patches during the pre-monsoon period show a significantly increasing trend (4 km^2/year) from 1987 to 2021. During the post-monsoon period, this trend is not statistically significant to the 95% confidence level. We used these classified waterlogged images to understand the dynamics of waterlogging. We observed a high probability of waterlogging in areas adjacent to the Fan margin. Further, we assessed the likelihood of waterlogging in the vicinity road-rail network. The concentration of waterlogged patches is relatively high within a one-kilometer buffer of the road-rail network. This study is a step towards understanding the impact of anthropogenic intervensions on the dynamics of waterlogging and drainage congestion.

1. Introduction

Waterlogging is a condition when the root zone of a soil column is fully saturated and does not allow surface water to infiltrate further into the ground. It is generally found in regions characterised by shallow groundwater tables, topographic depressions, soil with low porosity and permeability, excessive irrigation, and drainage congestion due to natural or anthropogenic interventions (Dinka and Ndambuki, 2014; Khalil et al., 2021; Arnous and Green, 2015; Hjerdt et al., 2004). If this situation persists for a prolonged period in any region may lead to soil degradation, rendering it unsuitable for agriculture (Walne and Reddy, 2021). Globally, waterlogging consists of nearly 5%–8% of the total area of land surface (Liu et al., 2023; Bassi et al., 2014; Gardner and Finlayson, 2018; Tiner et al., 2015). About 31.8% area of waterlogging is distributed in Asia, 27.1% in North America, 15.8% in South America, 12.5% in Europe, 9.9% in Africa, and 2.9% in Oceania (Davidson et al., 2018).

In India, about 8.5 million hectares of land are affected due to waterlogging (Chowdhury et al., 2011). The majority of them are distributed on the alluvial plains of the Indus, Ganga, and Brahmaputra rivers. Waterlogging is a severe problem on the Kosi Fan of North Bihar plain (Fig. 1). Rapid urbanisation, infrastructural development, construction of flood protection structures, and excessive irrigation have further aggravated the problem. The Kosi River is highly mobile; it has migrated laterally more than 110 km in the last three centuries. In the process of channel migration, the river has deposited its sediments and built a large

* Corresponding author.
E-mail addresses: niranjannaik@iiserb.ac.in (M. Niranjannaik), kgaurav@iiserb.ac.in (K. Gaurav), sabhilash@iiserb.ac.in, abhilash.iiserb@gmail.com (A. Singh), amrit16@iiserb.ac.in (A.K. Singh).

https://doi.org/10.1016/j.rsase.2023.101113
Received 1 August 2023; Received in revised form 14 November 2023; Accepted 27 November 2023
Available online 30 November 2023

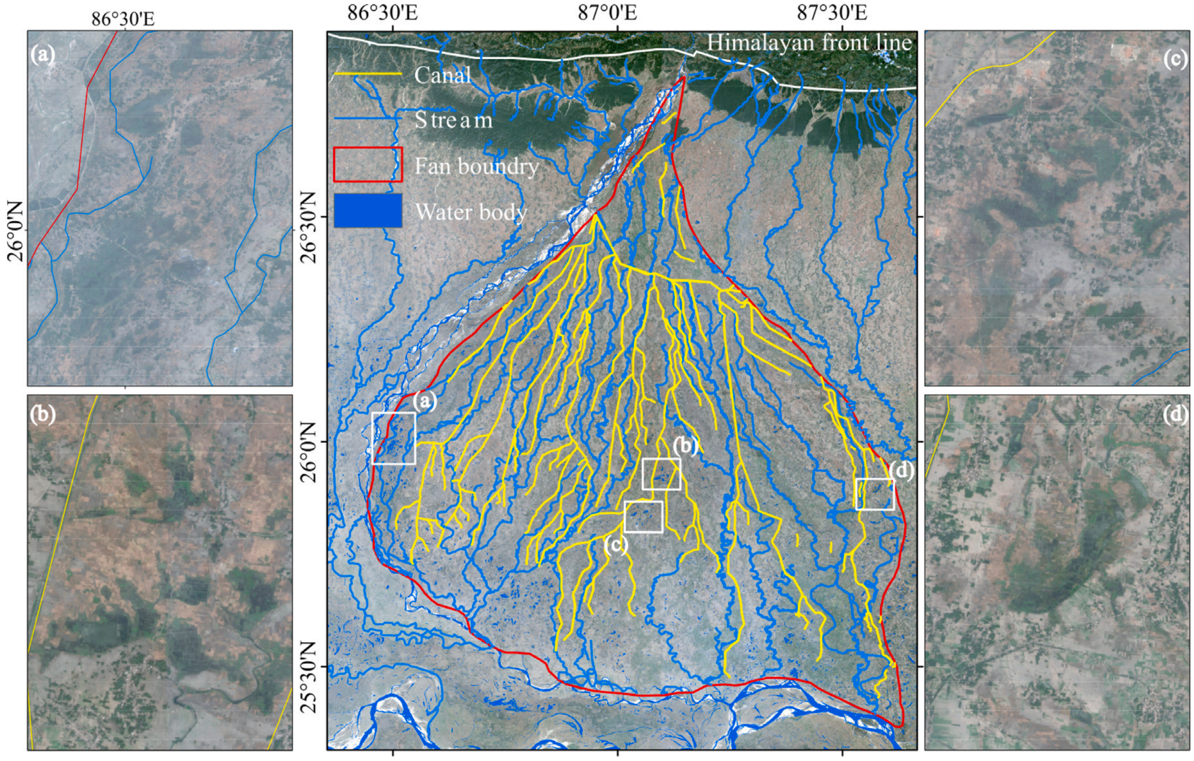

Fig. 1. Polygon in red is the boundary of the Kosi Fan. Lines in solid blue and yellow are the stream and canal network. Images (a–d) illustrate the waterlogged patches found on the Kosi Fan (square boxes in white).

conical structure (area- $10,351\ \mathrm{km}^2$; radius- 115–150 km), popularly known as the Kosi Fan (Gole and Chitale, 1966; Arogyaswamy, 1971; Wells and Dorr, 1987). To prevent channel migration, embankments were constructed in 1962 along both banks of the Kosi River. Consequently, drainage networks from the fan surface radiating towards the Kosi River are now unable to join the river. This has resulted in waterlogging at many places in the area adjacent to the embankments. Also, the lateral seepage of water from the Kosi River has caused waterlogging in the adjacent regions. Jain and Sinha (2005) identified a large portion of waterlogging patches in the lower region of the Kosi alluvial fan is groundwater-induced.

Apart from the embankments, a dense network of roads and rail, which run mostly in an E-W direction has severely affected the natural drainage of the Kosi Fan. At several places, these road and rail networks act as barriers and disconnect drainages and streams (Pandey et al., 2012; Sinha et al., 2013; Kumar et al., 2014; Singh et al., 2022). As a consequence, many locations on the Kosi Fan are characterised by severe drainage congestion and waterlogging. A significant increase in these waterlogging patches has been observed after the year 1995 (Pandey et al., 2012).

On the Kosi Fan, flooding and waterlogging problems get severe during the monsoon. The spatial distribution of the waterlogged patches appears to be highly dynamic. One can track their inter-annual variations from the time series satellite images. The spectral indices derived from the multi-spectral optical remote sensing images (*i.e.*, Landsat, and Sentinel-2) are very useful in monitoring waterlogging.

Pixel-based image classification algorithms, such as unsupervised and supervised are used to automatically classify the raw image pixels to thematic classes (Prajapati et al., 2021; Chakravorty and Jha, 2022; Amer, 2021; Neeti et al., 2022). The conventional application of classification algorithms over a fixed grid incurs several drawbacks, including high memory requirements, increased time complexity, and reduced suitability for real-time applications. These limitations stem from the inherent rigidity of fixed-grid approaches. They fail to adapt to the dynamic nature of real-world data. Therefore, they are not recommended for scenarios demanding efficient and responsive classification tasks.

To address this issue, we propose a hybrid framework that incorporates Simple Non-Iterative Clustering (SNIC) as a preprocessing step prior to applying the classification algorithm. SNIC generates superpixels of adaptive size, allowing the algorithm to dynamically adjust their dimensions based on the image content. This capability results in more accurate representations of regions within the image and also eliminates the constraints imposed by a fixed grid size. Furthermore, SNIC excels in maintaining better boundary adherence compared to alternative superpixel algorithms. This enhanced boundary adherence stems from SNIC's integration of both spatial and colour information, yielding superpixels that more closely align with object boundaries in the image. Subsequent to the

Table 1
Satellite and in-situ data used in this study.

In-situ					
Type of the data		Instrument			Time period
Field visit		RTK Survey and Garmin-84			15–19 March, 2019 and 11–20 December, 2019
Satellite data					
Type of the data	**Instrument/satellite**	**Time period**	**Images**	**Spatial resolution**	**Temporal resolution**
Optical	Landsat-5	1987–2011	284	30 m	16 days
	Landsat-7	2000–2003	27	30 m	16 days
	Landsat-8	2013–2016	28	30 m	16 days
	Sentinel-2	2016–2021	259	10 m	10 days
	High-resolution Google images	2002–2021	–	0.15–15 m	–
DEM	SRTM	2007–2013	1	30 m	Single

superpixel conversion, we apply the RF classifier on optical satellite images from 1987–2021 (Landsat and Sentinel-2) to accurately extract the waterlogging patches on the Kosi Fan. We performed the trend analysis to quantify the seasonal dynamics of these waterlogging patches, separately for the pre- and post-monsoon periods. Based on this, we compute the probability of a given pixel on the image to be classified as waterlogging. Finally, we evaluate the spatial association of waterlogging patches in the proximity of anthropogenic interventions such as road-rail networks. The findings of this study can be used to identify drainage congestion and potential locations where waterlogging can initiate or become severe due to anthropogenic interventions on the Kosi Fan.

2. Materials and methods

2.1. Dataset

2.1.1. Satellite data

We used a combination of multi-temporal satellite (optical) images and a Digital Elevation Model (DEM) to study the dynamics of waterlogging on the Kosi Fan. We have processed 598 individual images of Landsat and Sentinel-2 satellite missions from a period between 1987 to 2021 (Tables 1 & A.1). To perform the topographic analysis, we have used the Shuttle Radar Topography Mission (SRTM) DEM version 3 (spatial resolution 30 m). These datasets were imported into the Google Earth Engine (GEE) code editor environment (https://code.earthengine.google.com/) for further processing and analysis.

This study uses specific bands from the optical images, including Blue (0.45–0.51 μm), Green (0.53–0.59 μm), Red (0.64–0.67 μm), Near Infrared (NIR) (0.77–0.90 μm), and Shortwave Infrared (SWIR) (1.55–1.75 & 2.09–2.35 μm). The spatial resolution of Landsat images is 30 m and the revisit time is 16 days. In addition, we employed Sentinel-2 (A & B) multi-spectral images from 2016 to 2021. Sentinel-2 (A & B) together enables a temporal resolution of 5 days. The Blue, Green, Red, and NIR bands of Sentinel 2 have a spatial resolution of 10 m, and 20 m for the SWIR. We re-sampled all the Sentinel-2 bands at 30 m grid size using the bi-linear interpolation to ensure consistent spatial representation across the dataset. We have manually extracted the road, rail, embankments, and canal networks on the Kosi Fan from the Google Earth images (Table 1).

2.1.2. In-situ

We conducted field surveys in the pre- (5–19 March) and post-monsoon (11–20 December) of 2019 to record the location of waterlogged patches and their spatial extent (Fig. 2). We recorded the position (latitude and longitude) of 119 waterlogged patches distributed throughout the Kosi Fan by using a handheld Garmin-84 GPS. For some of the waterlogging patches, we have measured their boundary using a real-time kinematic GPS (GeoMax Zenith35 Pro model) in differential mode. This enables us to obtain the periphery of waterlogged patches. At each measurement location, we first established a reference station and set up a base receiver in static mode. We then set up a rover receiver and configured it with the base receiver. We walk around the outer periphery of a waterlogged patch with the rover and record the coordinate and elevation every five seconds. This procedure yields an uncertainty of less than 10%.

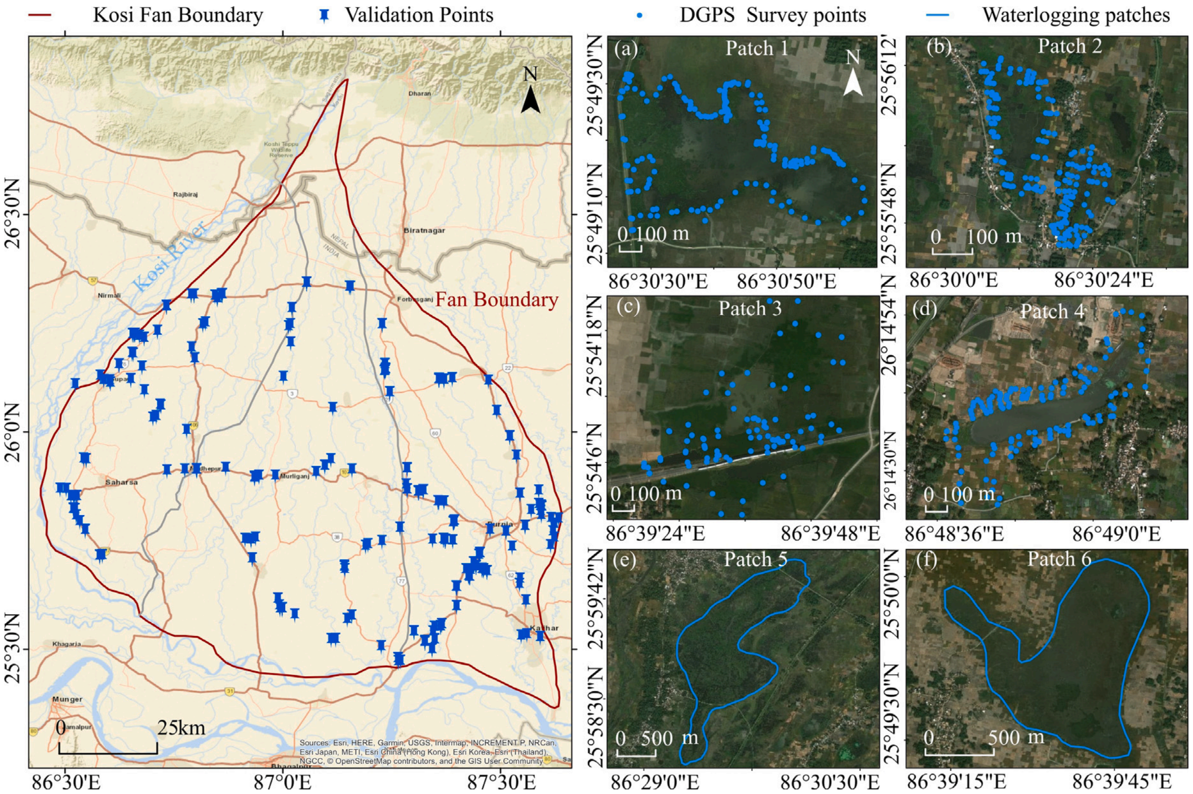

Fig. 2. Locations of waterlogging patches surveyed in the field. The placemark in blue are the locations of waterlogged patches surveyed in the field. Their coordinates were recorded by using handheld and RTK GPS. The Google Earth images shown on the right panel show the waterlogging patches surveyed (dots) in the field using a differential GPS in RTK mode (a–d). The periphery of the waterlogged patches is manually digitised from Google Earth images.

2.2. Image processing

We categorised the satellite images based on their acquisition dates into pre- (March-May), and post-monsoon (October–December) periods. We have selected the blue, green, red, NIR, and SWIR bands to compute indices such as the Normalised Difference Vegetation Index (NDVI), Soil Adjusted Vegetation Index (SAVI), and Optimised Soil Adjusted Vegetation Index (OSAVI), Normalised Difference Water Index (NDWI), and Modified Normalised Difference Water Index (MNDWI) (Fig. 3). The equations to compute these indices are provided in Appendix A (Table A.2). The SAVI and OSAVI are effective in detecting waterlogging patches with low vegetation coverage and exposed soil (Poulin et al., 2010; Taddeo et al., 2019). As compared to NDVI, the MNDWI can accurately detect waterlogging patches covered with grass reeds. This is due to the fact that SWIR bands have more light absorbance capabilities as compared to NIR band (Singh et al., 2015). We now take the median of different indices (NDVI, SAVI, OSAVI, MNDWI, and NDWI), and image bands (green, blue, red, NIR, SWIR) of Landsat and Sentinel-2 satellites separately for the pre-and post-monsoon periods from 1987–2021. Finally, we use these layers to automatically classify the waterlogged patches.

2.3. Extraction of waterlogging

We applied image segmentation and RF algorithms to extract waterlogging patches from the satellite images (Fig. 4). We used a SNIC algorithm within the GEE environment. The SNIC algorithm generates object-based segments corresponding to different features from the satellite image. It requires four input parameters: cluster size, compactness factor, connectivity, and neighbourhood size. The size of a cluster determines the spacing of seed locations for superpixels, which are small clusters of connected pixels with defined geometry. The compactness factor controls the shape of the clusters, with higher values leading to more compact clusters. The connectivity parameter manages the contiguity and merging of adjacent clusters, while the neighbourhood size parameter helps to prevent artifacts at the tile boundaries. We experimented with different cluster sizes and found a seed spacing value of 5 yields optimal results and accordingly selected parameter values of 1, 8, and 10 for compactness, connectivity, and neighbourhood size factor, respectively (Shafizadeh-Moghadam et al., 2021; Tassi and Vizzari, 2020). We then overlay the segmented features on stacked layers created from the individual median images. Within the boundary of each cluster, we take the mean pixel values of

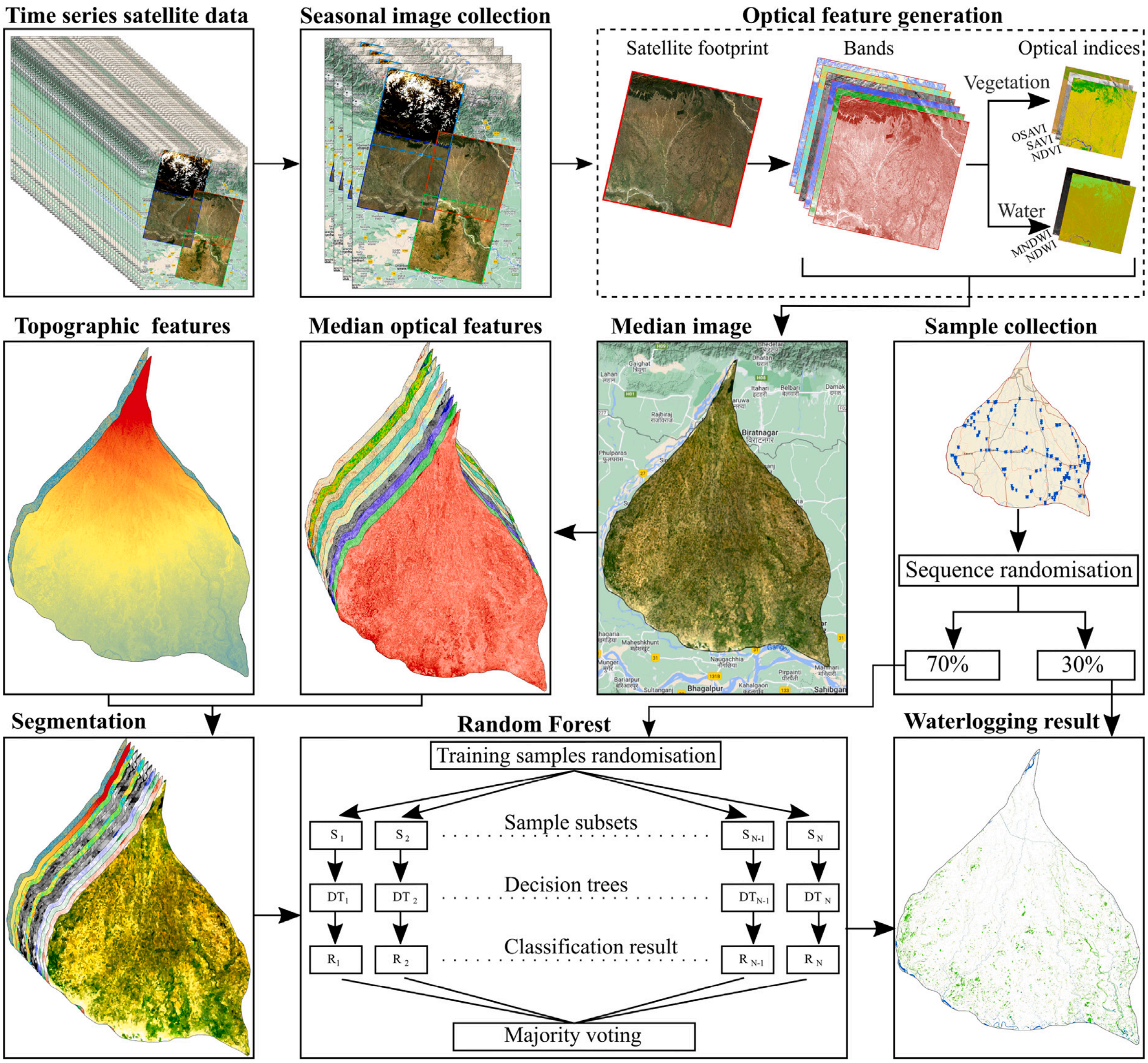

Fig. 3. Flow chart illustrates the methodology adopted in this study.

the individual bands of the stack image. Finally, we use them as training samples to classify waterlogging patches from satellite images by using the RF algorithm.

We use 70% of the training samples with their corresponding class level to train the RF classifier. It randomly selects N subsets of training samples and their corresponding classes to create independent decision trees for each subset. The best split of nodes is determined by minimising the correlation between trees (Ao et al., 2019). Each decision tree then predicts the classification result. Subsequently, the RF algorithm performs majority voting by combining the results of each decision tree to assign a final decision. Executing RF on the GEE environment requires three input parameters to be specified; the number of trees, variables per split, and seed. We set the values for decision trees, variables per split, and seed as 100, 3, and 0, respectively by trial and error method. We now test the RF model by using the remaining 30% of the samples to compute the classification accuracy. Once the model is trained and validated, we apply it to classify the satellite image.

The classification results in disconnected boundaries of the waterlogged patches and some pixels within it are classified as non-waterlogged. To connect the disconnected boundaries of the waterlogging patches, we applied the connected component algorithm (ee.Image.connectedComponents). Simultaneously, we applied morphological operations (erosion and dilation) on classified images. They merge the pixels that were initially classified as non-waterlogged inside the waterlogged. Once the artifacts from the classified images are rectified, we compute the area of individual patches of waterlogged and surface water bodies. Finally, we use time series images to compute the probability of a given pixel on the Kosi Fan being classified as waterlogging (P_w) by using Eq. (1);

$$P_w = \frac{\text{Number of times waterlogging observed at the same pixel}}{\text{Total number of classified images}}, \tag{1}$$

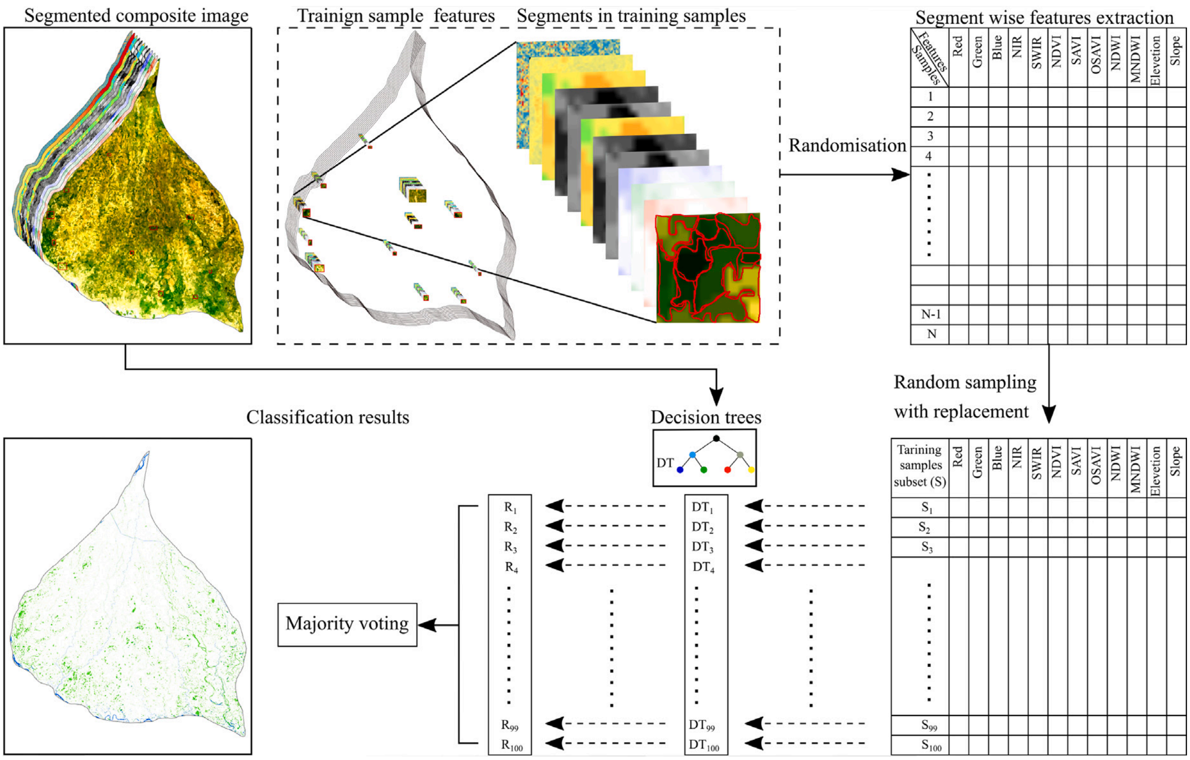

Fig. 4. Classification scheme adopted to classify the satellite images by using the RF algorithm.

2.4. Trend analysis

We used a non-parametric Mann–Kendall test to assess the trend of waterlogging for the pre- and post-monsoon periods separately from 1987–2021. It is useful for testing the hypothesis about the presence or absence of a trend within a time series (Mann, 1945; Hofmann et al., 2023). We formulated the null hypothesis (H_0) that the total area of the waterlogging patches on the Kosi Fan is constant. Eq. (2) estimates the sign for each pair of observations in a time series ($x_j - x_i$);

$$sign(x_j - x_i) = \begin{cases} +1, & \text{if } (x_j - x_i) > 0 \\ 0, & \text{if } (x_j - x_i) = 0 \\ -1, & \text{if } (x_j - x_i) < 0 \end{cases}, \tag{2}$$

where x_i and x_j are observations are ranked from i = 1 to n−1 and j = i+1 to n, respectively.

If the value of τ (Eq. (3)) deviates significantly from 0 in either direction (positive or negative), it signifies an existing trend in a times series. The concordant pairs are the observations of x_i and x_j (where $i < j$) that have the same direction (increasing or decreasing). The discordant pairs have different directions (increasing and decreasing). The total number of pairs includes the combination of concordant and discordant pairs.

$$\tau = \frac{\text{Number of concordant pairs} - \text{Number of discordant pairs}}{\text{Total number of pairs}}. \tag{3}$$

We estimated the S-statistics by using Eq. (4). If the sum of its signs is positive, this indicates an increasing trend in a time series and vice-versa. If the sum of signs approaches zero, indicates no trend.

$$S = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} sign(x_j - x_i). \tag{4}$$

Subsequently, the variance of S-statistics is estimated using the following expression in Eq. (5);

$$var(S) = \frac{n(n-1)(2n+5)}{18}, \tag{5}$$

where n is the number of samples present in a time series data.

Table 2
Mann–Kendall test and Sen's slope statistics.

Mann–Kendall test		
Parameters	Post-monsoon	Pre-monsoon
n	25	31
Z-values	0.93	2.03
p-value	0.35	0.041
S	41	121
VarS	1832.33	3461.66
τ	0.13	0.26
Sen's slope		
Confidence level range within 95%	−8.13 18.24	0.25 7.68
β	5.61	4.21

We have performed the $Z-\text{test}$ to estimate the probability that the trend occurred by chance, considering the sample size and the distribution of time series data (Eq. (6)) . If $Z > 0$, it indicates an increasing trend, and vice versa.

$$Z = \begin{cases} \frac{S-1}{\sqrt{var(S)}}, & \text{if } S > 0 \\ 0, & \text{if } S = 0 \\ \frac{S+1}{\sqrt{var(S)}}, & \text{if } S < 0 \end{cases}. \tag{6}$$

We applied the Sen's slope estimate (β) to determine the magnitude of a trend in time series data according to Eq. (7);

$$\beta = Median\left(\frac{x_j - x_i}{j - i}\right); j > i, \tag{7}$$

where β is Sen's slope estimate.

2.5. Impact of structural interventions

We assess the impact of anthropogenic interventions (roads, rail, embankments, canal network) on waterlogging on the Kosi Fan. To do this we compute the conditional probability (Eq. (8)) of waterlogging within a 1 km buffer of the road and rail network;

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{\frac{N(A\cap B)}{N}}{\frac{N(B)}{N}}, \tag{8}$$

where A, B, and N represent waterlogging occurrence, the road-rail network, and the number of observations, respectively.

We further examined the dynamics of waterlogged patches at the nodes where road and rail networks intersect rivers or drainage networks. Such conditions can create barriers that impede the downstream fluxes, leading to waterlogging. We identified a total of 243 nodes, which correspond to the intersections of rivers and road-rail networks across the Kosi Fan (Fig. 6).

3. Results

3.1. Accuracy assessment

To assess the classification accuracy of satellite images, we employed ground truth data consisting of 119 waterlogged patches, gathered both in the field and through visual interpretation of high-resolution Google Earth images. Transition matrices were computed for the pre- and post-monsoon periods for each year from 1987 to 2021. The overall accuracy of waterlogged patches classified from Landsat and Sentinel-2 images ranges between 75% and 90%. The transition matrix reports the classification accuracy (Fig. 5). Other accuracy matrices such as recall, precision, and F1-score are provided in Appendix B (Table B.2).

3.2. Dynamics of waterlogging

To comprehensively investigate the dynamics of waterlogging patches, we conducted a thorough analysis of their spatial and temporal variations spanning from 1987 to 2021, with a specific focus on both the pre-monsoon and post-monsoon periods (Fig. 6-Fig. 7). The results indicate that the extent of waterlogging is consistently high during the post-monsoon period. Seasonal fluctuations in waterlogging are discernible on the Kosi Fan in both the pre- and post-monsoon periods. Notably, a spike in waterlogging patches was observed during the post-monsoon season of 2008. This surge can be attributed to the catastrophic Kosi flood that occurred in August 2008.

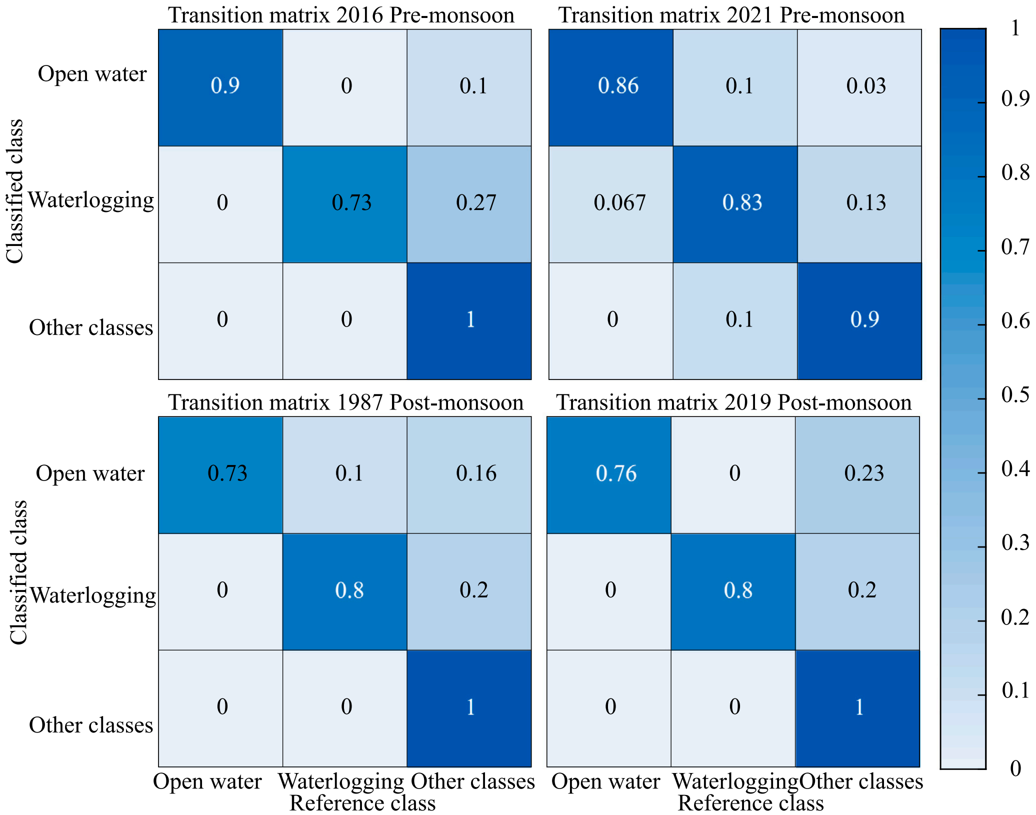

Fig. 5. The transition matrix illustrates the overall accuracy of the classification results obtained for the pre-monsoon (2016 and 2021) and post-monsoon (1987 and 2019) periods.

The statistical tests conducted on the time series data for pre- and post-monsoon waterlogging yield Z-values of 2.03 and 0.93, with corresponding p-values of 0.04 and 0.35, respectively (Table 2). These results highlight a statistically significant increasing trend in waterlogging patches during the pre-monsoon period (Fig. 8). However, to the 95% confidence level, this trend is not significant for the post-monsoon period. Table 2 reports the Sen's slope (β) value. The rate of increase in waterlogging during the pre-monsoon period is approximately $4\ \text{km}^2/\text{yr}$.

3.3. Anthropogenic impacts on waterlogging

We observed a higher probability of waterlogging occurrence at the western and eastern margins of the Kosi Fan, particularly in the close proximity to streams and areas where structural interventions, such as road-rail networks, embankments, and canals, impede natural drainage (Fig. 9). To assess drainage congestion, we plot the number of waterlogging patches in relation to their distance from the road-rail network (Fig. 10). Notably, the majority of waterlogging patches are concentrated within a one-kilometer radius of the road-rail network, their frequency gradually decreasing as we move farther away. We examined the distribution of waterlogging patches based on their sizes (Fig. 10). We found that approximately 90% of these patches have an area less than $0.05\ \text{km}^2$, while the remaining 10% vary in size between $0.05–1\ \text{km}^2$.

Now we compute the conditional probability of waterlogging occurrence within a one-kilometer buffer created around the road-rail network (Fig. 11). We observed that the waterlogging patches are frequent in regions where the road-rail network intersects with the drainage network. We systematically identified intersections between the road-rail and river networks, represented as nodes, and scrutinised waterlogging patches in their vicinity. These intersection points were categorised into three distinct groups: fully connected, disconnected, and partially connected nodes (Fig. 12). To provide a dynamic perspective, we generated a time series plot depicting the number of nodes experiencing waterlogging conditions (Fig. 13).

The Kosi Fan has witnessed a rapid surge in road-rail network development work after the year 2012 (Kumar et al., 2014; Goswami, 2019; Nhai, 2022; Morth, 2022). These developmental activities appear to have aggravated waterlogging in the study area.

4. Discussion

The hybrid framework for image classification we proposed can be used to monitor the seasonal variability of waterlogging patches from optical satellite images. The overall classification accuracy of waterlogged patches extracted from Landsat and Sentinel-2 images varies between 75% and 90%. The total area of waterlogging on the Kosi Fan is relatively high in the post-monsoon period.

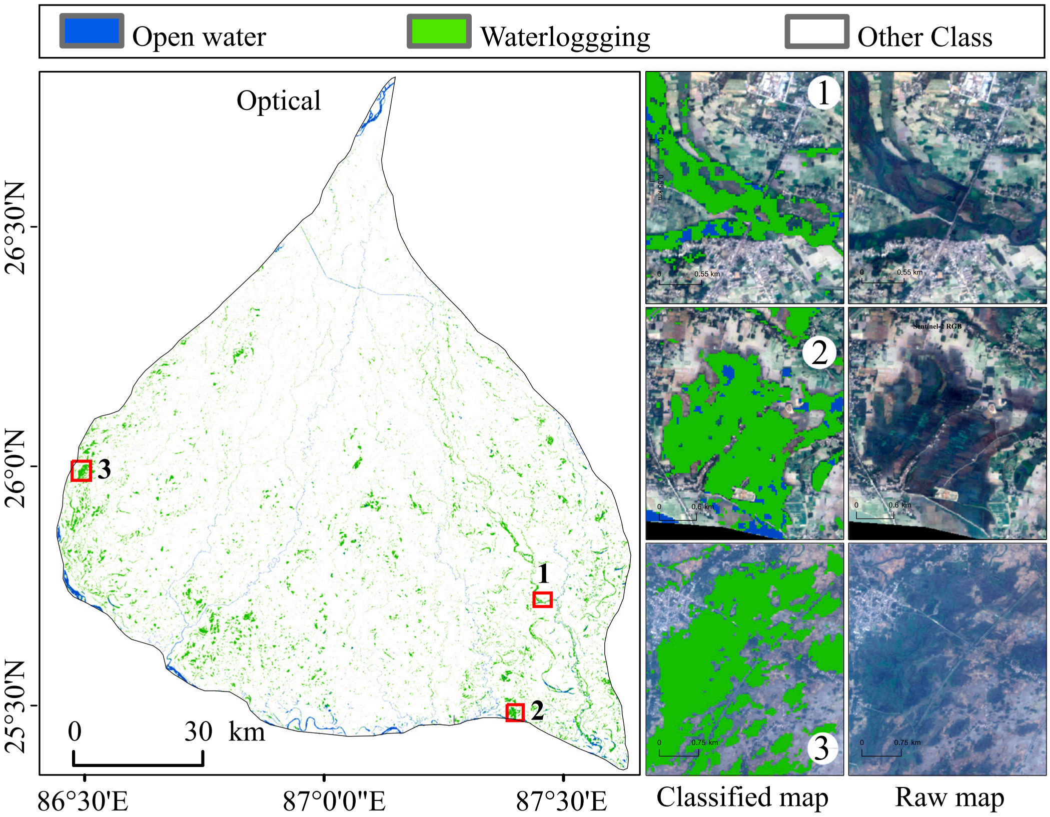

Fig. 6. Classification of waterlogging pixels from the satellite images of the post-monsoon period of 2019. The images on the right panel show the zoomed area waterlogging patches.

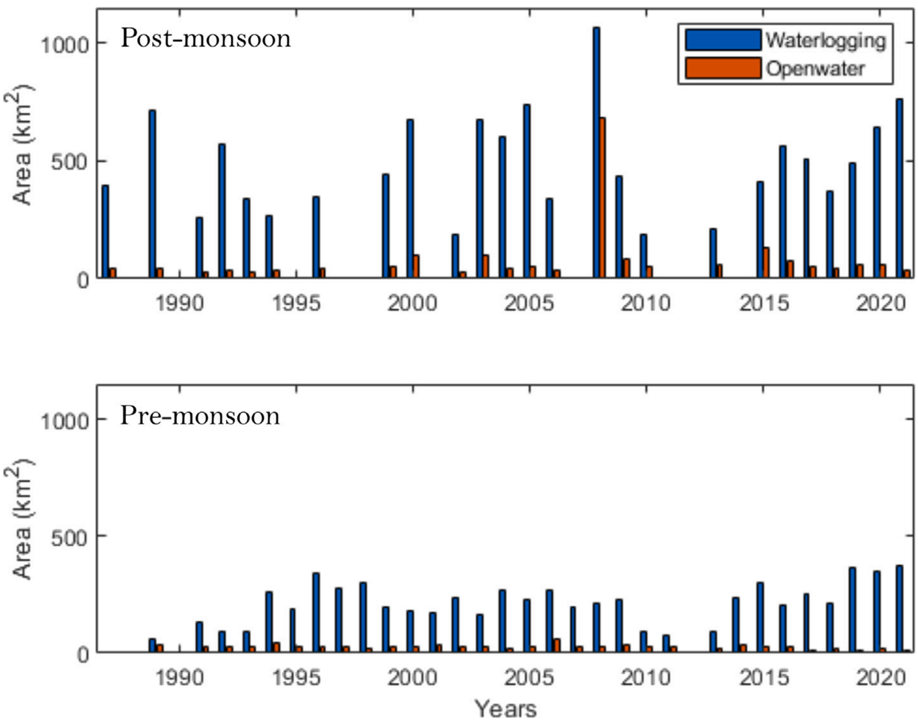

Fig. 7. Variation in the area of waterlogging patches and openwater extracted from satellite area from 1987 to 2021 for the pre and post-monsoon.

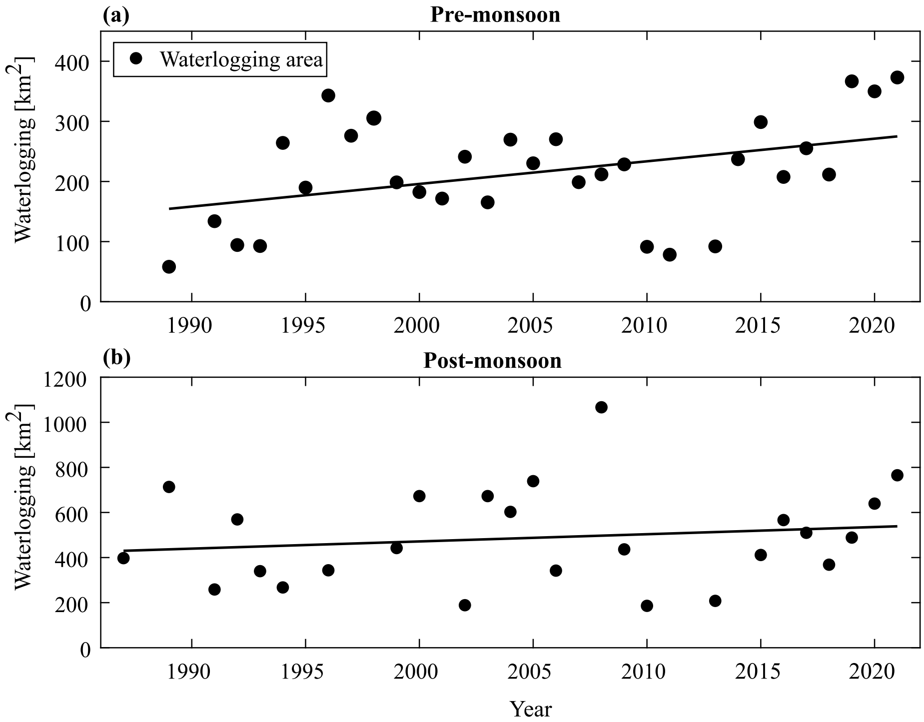

Fig. 8. Trend of the waterlogging area from 1987 to 2021 during (**a**) pre-monsoon and (**b**) post-monsoon periods, respectively. Circles in black are the total area of waterlogging on the Kosi Fan in different years in pre- and post-monsoon. The solid line is the linear fitted to the data.

This is because of the fact that a large portion of the Kosi Fan gets inundated during the monsoon (June-September) period. The flood water starts to recede after the monsoon. The high residence time of flood water stored in the topographic depression, low-lying areas, ditches, and near drainage congestion due to structural barriers (i.e.; road, rail, embankments) results in waterlogging for a longer duration. These temporary storages dry due to intense solar radiation received during the pre-monsoon. This results in seasonal dynamics in waterlogging on the Kosi Fan. Time series analysis of waterlogged patches from 1987 to 2021 reveals a significant increasing trend for the pre-monsoon period. However, this trend is very mild for the post-monsoon period. On the Kosi Fan, we observe waterlogging patches show a strong spatial association in the proximity of road and rail networks. These structural interventions disconnect the flux movement in the downward direction, leading to localised accumulation of water in the rainy season. Prevailing such a condition for a longer duration may cause permanent waterlogging in the region. Fig. 14 shows the waterlogging condition in pre-monsoon on the Kosi Fan in the past two decades. We observed areal expansion of existing and the formation of new patches of waterlogging in the study area from 1987–2010 to 2011–2021. This observation correlates with the infrastructural development activities on the Kosi Fan (Nhai, 2022; Morth, 2022). The waterlogging conditions are likely to intensify in the future under the projected infrastructural development in the study area. This condition can have adverse impacts in terms of soil degradation, loss of agricultural productivity, and increasing risk of flooding.

4.1. Topographic and anthropogenic impact on waterlogging

The topography plays a pivotal role in shaping the waterlogging dynamics of the Kosi Fan. Characterised by a distinctive convex-up transverse profile and a concave upward radial profile, the Kosi Fan demonstrates unique geomorphic features (Fig. 15). The highest concentration of waterlogged patches occurs prominently along the fan margin. Notably, divergent drainage networks become apparent, radiating in the South-West and South-East directions from the fan axis. Our observations are in accordance with the waterlogging in the low-lying areas, particularly in the proximal and medial regions of the Kosi Fan, with comparatively lower occurrences nearer the fan axis (Fig. 9).

Furthermore, to assess the impact of structural interventions on waterlogging, we conducted an analysis of road and rail network density (expressed in km/km^2) within a 10×10 km grid (Fig. 16). Grids encompassing major urban centres such as Saharsa, Purnea, and Madhepura have the highest road-rail density (0.7 to 0.8). Regions adjacent to these cities have a network density ranging between 0.3 to 0.7. We have computed the total area of waterlogging patches in the proximal ($4\ km^2$), medial ($80\ km^2$), and distal ($304\ km^2$) parts of the Kosi Fan. Further, we have computed the area of waterlogging in different lobes of the Kosi Fan (Fig. 16). The total area of waterlogging patches in lobes 1, 2, and 3 are respectively $118\ km^2$, $88\ km^2$, and $181\ km^2$.

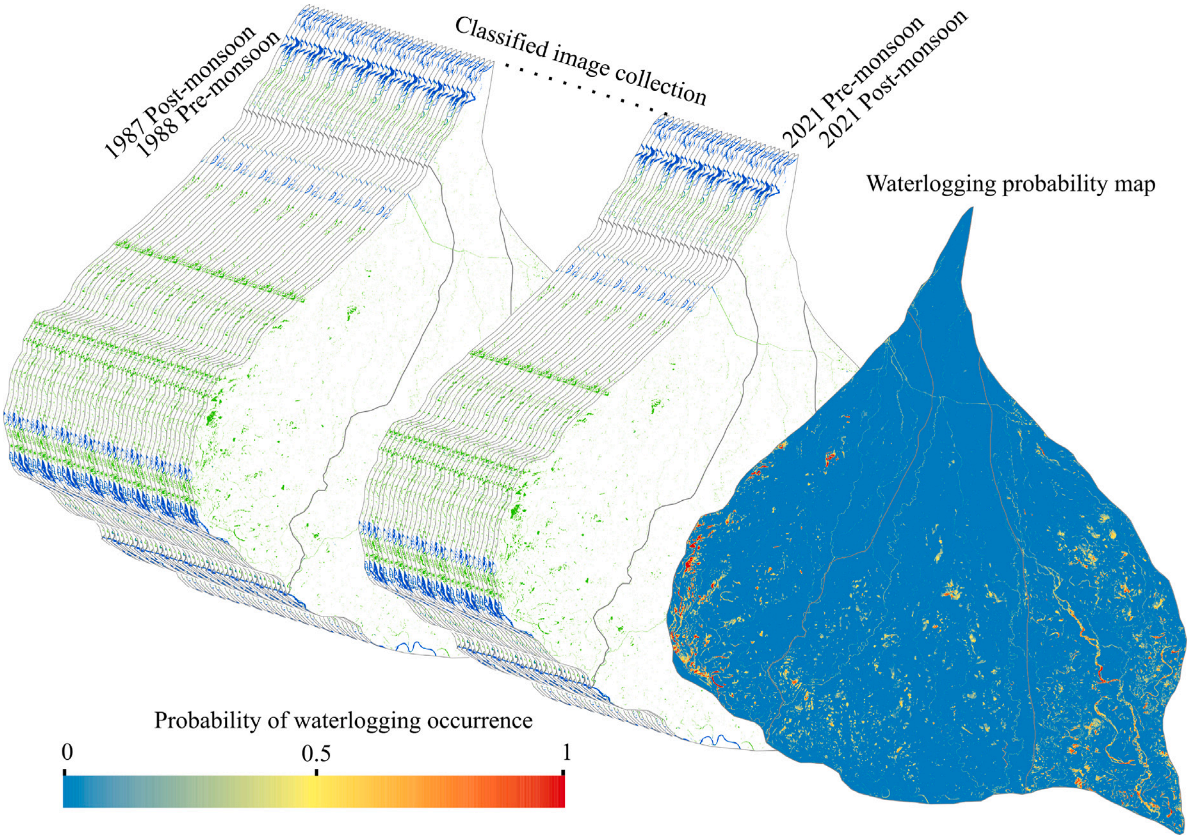

Fig. 9. Probability of waterlogging occurrence. The stack of classified image collections from 1987 to 2021 is shown in the background. The probability of waterlogging occurrence is shown in a single image on the right front side of the figure. In comparison, the red, yellow, and blue waterlogging patch areas show a high, medium, and low probability of waterlogging occurrence, respectively.

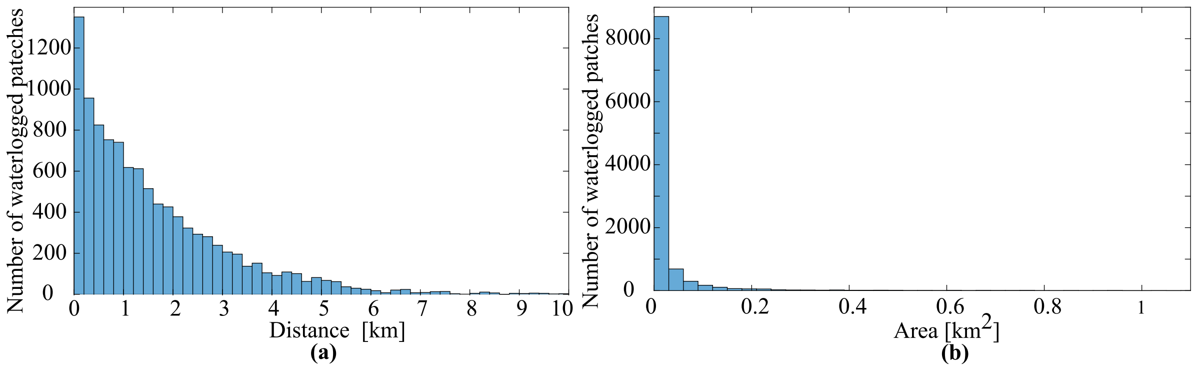

Fig. 10. Statistics of the waterlogging patches on the Kosi Fan. (**a**) Histogram of the waterlogging patch minimum distance from the road-rail network (bin size 200 m). (**b**) Histogram of the waterlogged patches area in the Kosi Fan (bin size $0.025\ \mathrm{km}^2$ or $25000\ \mathrm{m}^2$).

4.2. Limitations and future directions

This study provides a robust approach of assessing the dynamics of waterlogging under the influence of drainage congestion due to structural interventions on the Kosi Fan. The hybrid image classification algorithm is able to automatically identify the waterlogged pixels from optical satellite images with an overall accuracy greater than 75%. Time series analysis of waterlogged patches extracted from satellite images (1987–2021), separately for the pre- and post-monsoon reveals increasing trends. However, to the 95% confidence level, this trend is not significant in the post-monsoon period. Spatial analysis of the waterlogging on the Kosi Fan suggests a significant control of topography and drainage congestion due to the construction of road and rail networks.

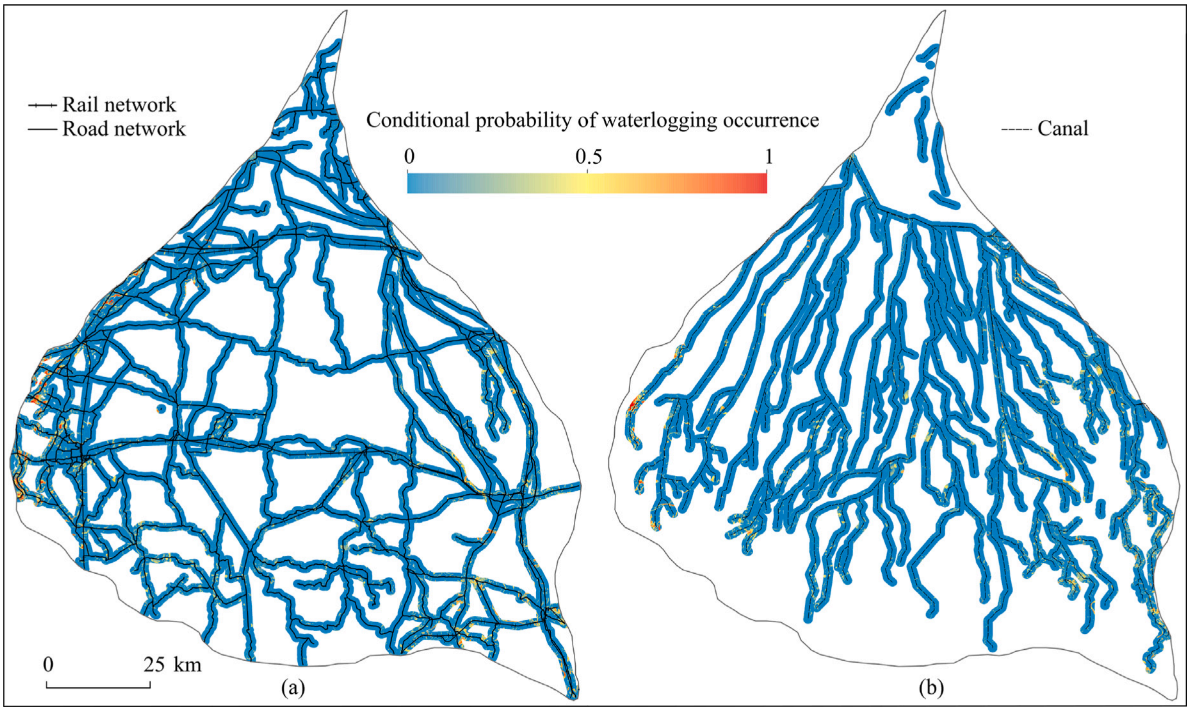

Fig. 11. Conditional probability of occurrence of waterlogging within the 1 km proximity of (**a**) road-rail network and (**b**) canal network on the Kosi Fan.

Despite its merit, this study has certain limitations. For example, the limitation of optical satellite sensors to acquire images in cloudy weather and heavy rainfall. This limits our ability to extract waterlogged patches using optical satellite images during the monsoon period. These limitations can be largely addressed by employing a combination of techniques, such as the fusion of multi-sensor data, the application of sophisticated machine learning algorithms for cloud detection and removal, and the exploration of deep learning models for image inpainting. Further, this study does not make any distinction between the natural and anthropogenic waterlogged patches. A finer level of image classification to automatically detect such waterlogged patches from satellite images would help to assess and compare their trends separately.

This study provides valuable insights for devising strategies to alleviate waterlogging challenges adjacent to road and rail networks. This becomes especially critical for addressing drainage congestion. A recent upswing in waterlogging strongly correlates with drainage congestion. This poses a considerable risk of inducing severe flooding on the Kosi Fan. This holistic approach will contribute to a more critical understanding and effective management of the complex interplay between waterlogging and the broader flood dynamics in the Kosi Fan.

5. Conclusions

We have performed a time series analysis using optical satellite images to assess the dynamics of waterlogging on the Kosi Fan during the pre- and post-monsoon periods. We have used the inherent topography and drainage congestion due to the road and rail network to explain the potential cause for the development of waterlogging on the Kosi Fan. Based on the finding of this study, following conclusions can be drawn.

- The hybrid algorithm based on SNIC-RF can classify the waterlogging pixels from Landsat and Sentinel-2 images with an overall accuracy between 75%–90%. It can detect the waterlogged patches covered with grass reeds.
- Time series analysis (1987–2021) of the Landsat and Sentinel-2 images reveals an increasing trend in the area of waterlogged patches. This trend is about $4\ \mathrm{km}^2/\mathrm{year}$ for the pre-monsoon period. To the 95% confidence level, it is not significant for the waterlogging patches observed during the post-monsoon period.
- The topographic depression at the fan margin and divergent drainage networks in the SE and SW from the Fan axis provide an accommodation of rainwater leading to waterlogging.
- The waterlogging patches are widely found on the Kosi Fan in the proximity of the road and rail networks. These networks act as barriers that obstruct the movement of fluxes further downstream during the rainy season, leading to inundation and eventually waterlogging in their proximity. The waterlogging at the intersection of drainage and road network on the Kosi Fan has significantly increased (about 15–20%) both in the pre- and post-monsoon after the year 2014.
- About $118\ \mathrm{km}^2$ and $181\ \mathrm{km}^2$ area of lobes−1 and 3 of the Kosi Fan is waterlogged. The lobe-2 is the least affected by waterlogging, only $88\ \mathrm{km}^2$ of its area is waterlogged.

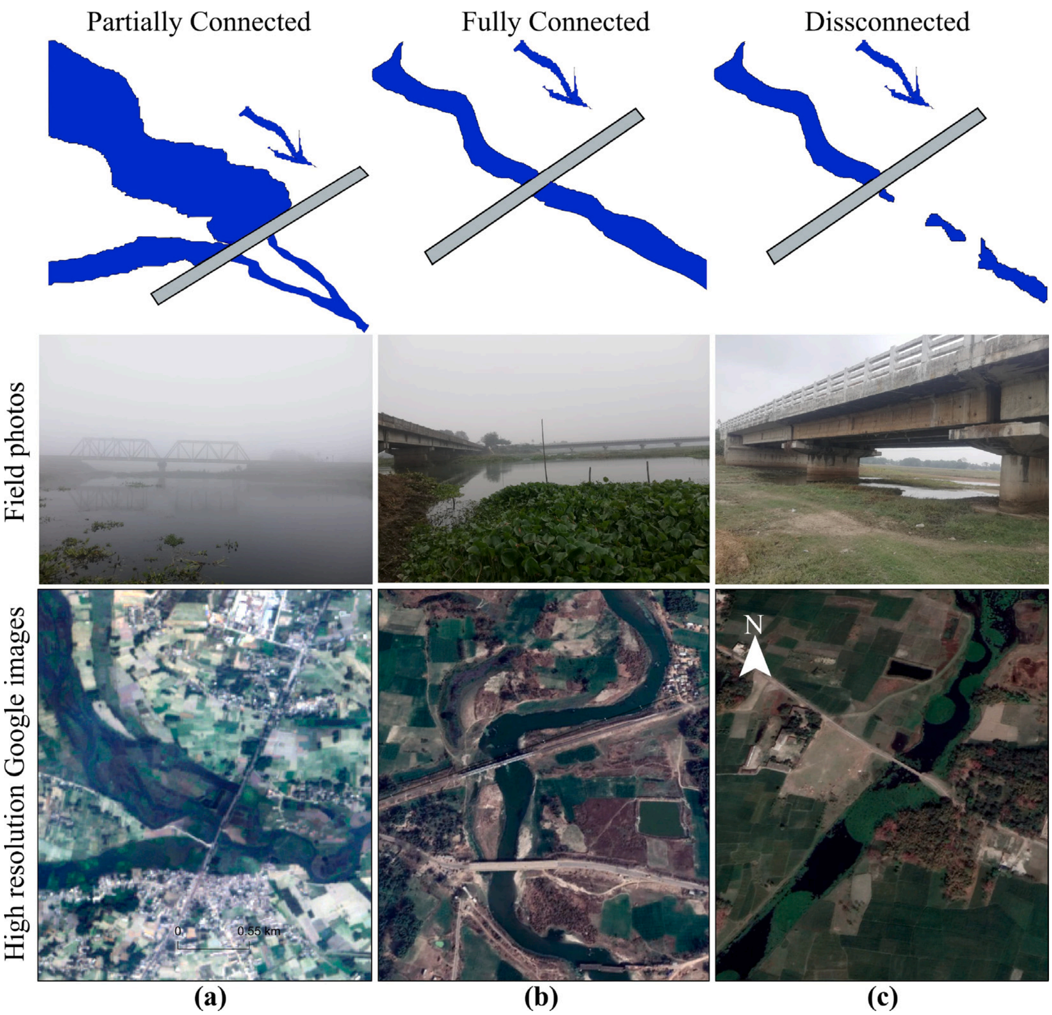

Fig. 12. Schemetic to illustrate the channel dis-connectivity due to structural interventions (road, rail, canal, embankments, etc.). The field photographs and corresponding locations on the satellite images are shown for visualisation.

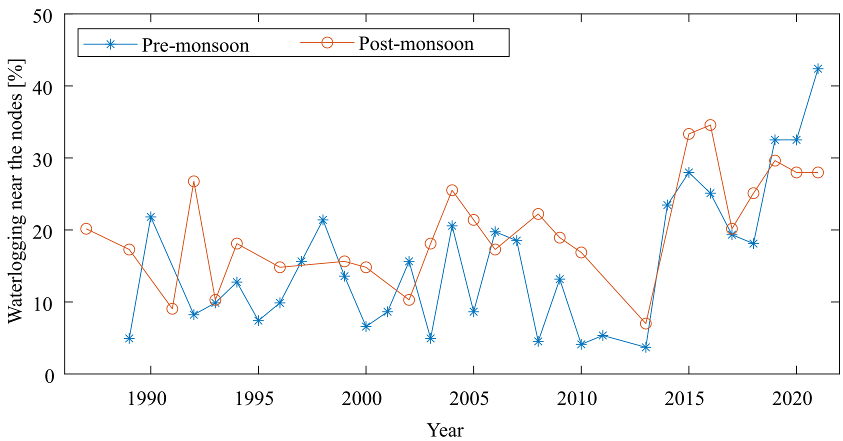

Fig. 13. Waterlogged near the nodes (%) within the proximity of road-rail and river network intersection on the Kosi Fan from 1987–2021.

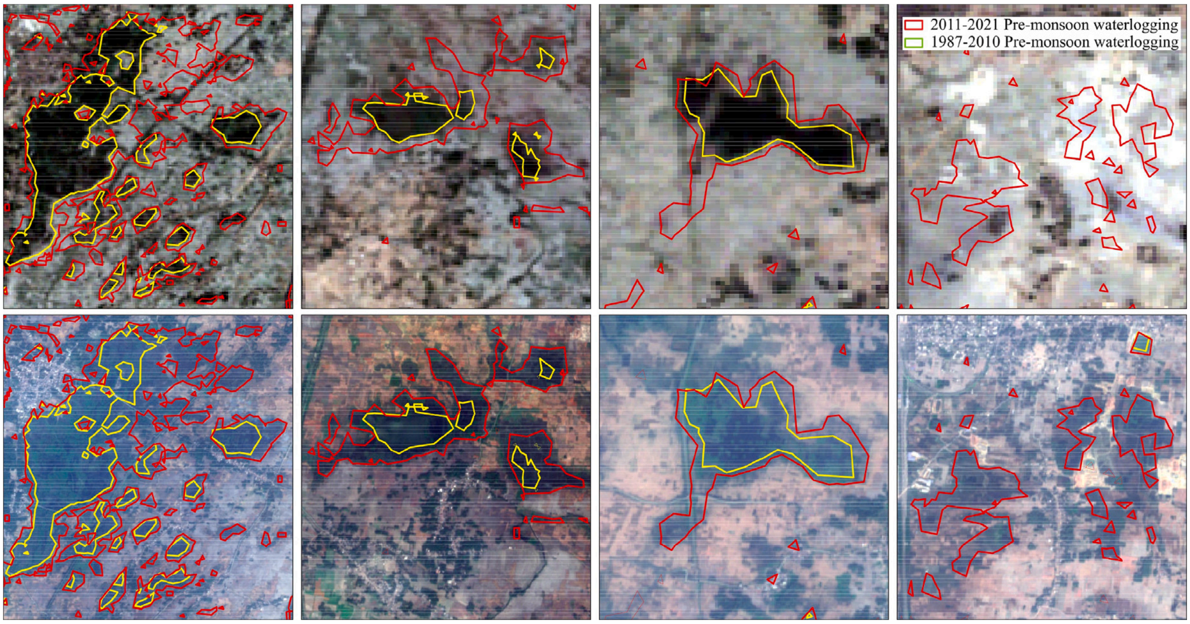

Fig. 14. Decadal variation in the spatial extent of pre-monsoon waterlogging areas during 1987–2010 and 2011–2021 at four different locations (from left to right). The top four images depict the 1987–2010 period, while the bottom four images represent 2011–2021. Waterlogged areas for the 1987–2010 and 2011–2021 periods are delineated with yellow and red polygons, respectively.

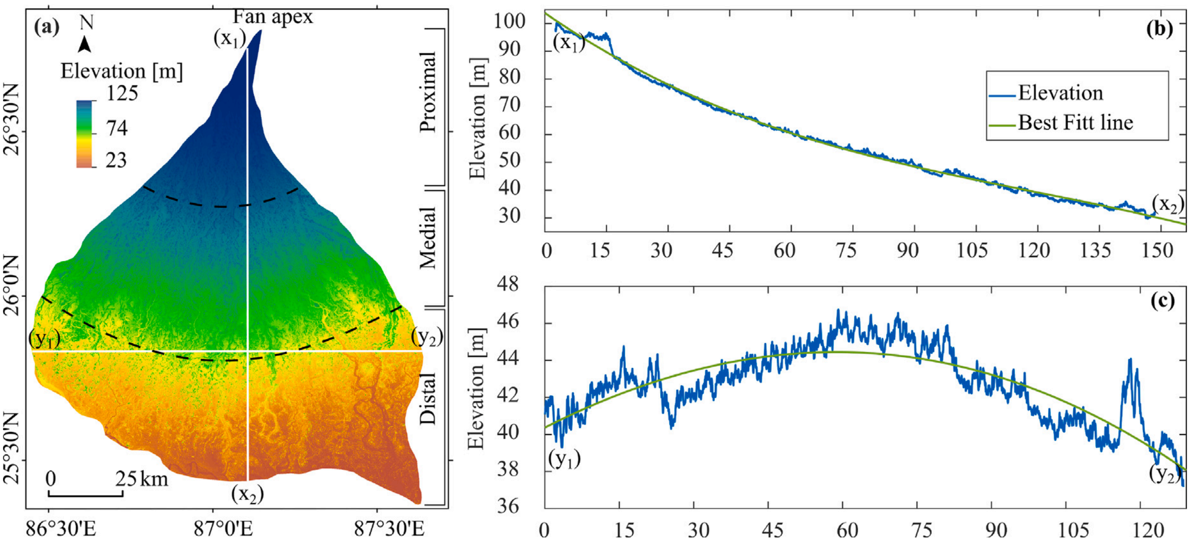

Fig. 15. (**a**) Digital elevation model of the Kosi Fan. (**b**) Lateral elevation profile. (**c**) Longitudinal elevation profile of the Kosi Fan.

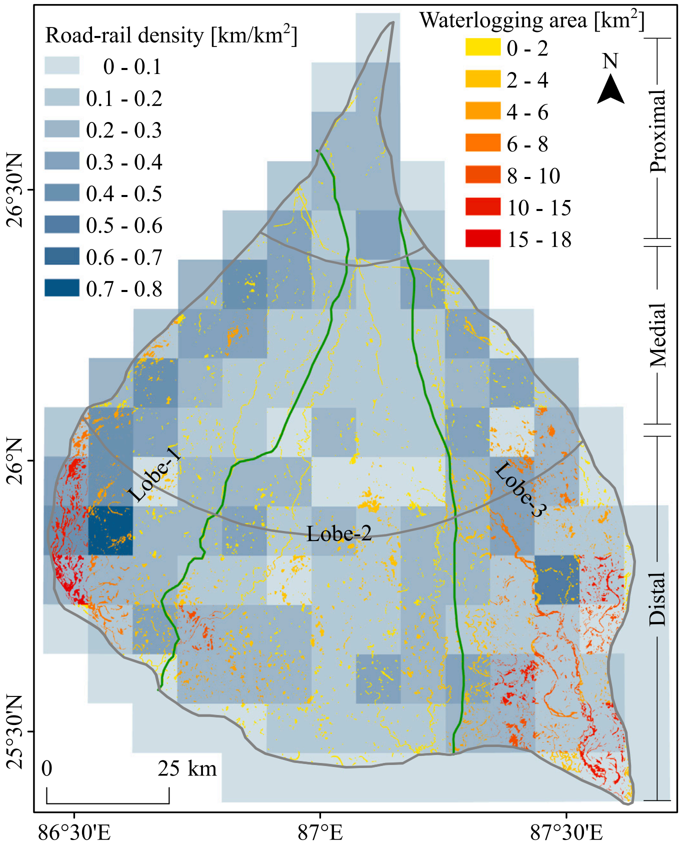

Fig. 16. The road-rail network density and waterlogging area within each grid on the Kosi Fan.

Funding

This research was funded to Dr. Kumar Gaurav by the Science and Engineering Research Board (SERB), Govt. of India through grant ECR/2017/001154.

Code availability

The code developed in this study will be made available on request.

CRediT authorship contribution statement

M. Niranjannaik: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Writing – original draft, Writing – review & editing, Visualization. **Kumar Gaurav:** Investigation, Writing – original draft, Writing – review & editing, Visualization, Supervision. **Abhilash Singh:** Methodology, Writing – original draft, Writing – review & editing, Visualization. **Amrit Kumar Singh:** Software, Methodology, Data curation, Visualization.

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.

Data availability

Data will be made available on request.

Acknowledgements

We acknowledge IISER Bhopal for providing institutional support. We thank the Science and Engineering Research Board (SERB) for funding this research.

Appendix A. Datasets and features

The appendix consists of Tables A.1 and A.2.

Appendix B. Accuracy assessment

See Table B.1
TN = True positive
TP = True positive
FN = False negative
FP = False positive

TN = True positive
TP = True positive
FN = False negative
FP = False positive

$$\text{Overall accuracy} = \frac{\text{TN} + \text{TP}}{\text{TN} + \text{FP} + \text{FN} + \text{TP}}, \tag{B.1}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \tag{B.2}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \tag{B.3}$$

$$\text{F1 Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}. \tag{B.4}$$

Table A.1
Details of the number of optical satellite images used in this study.

Satellite	Period	Number of images		Total images
		Pre-monsoon	Post-monsoon	
Landsat-5	1987–2001, 2004–2011	176	108	284
Landsat-7	2000–2003	14	13	27
Landsat-8	2013–2015	18	10	28
Sentinel-2	2016–2021	105	154	259
Total Optical images	1987–2021	313	285	598

Table A.2
Indices used in this study.

Index	Definition
Normalised Difference Vegetation Index (NDVI)	$\frac{\text{NIR - Red}}{\text{NIR + Red}}$
Soil Adjusted Vegetation Index (SAVI)	$\frac{\text{NIR - Red}}{\text{NIR + Red + 0.5}} \times (1 + 0.5)$
Optimised Soil Adjusted Vegetation Index (OSAVI)	$\frac{\text{NIR - Red}}{\text{NIR + Red + 0.16}}$
Normalised Difference Water Index (NDWI)	$\frac{\text{Green - NIR}}{\text{Green + NIR}}$
Modified Normalised Difference Water Index	$\frac{\text{Green - SWIR}}{\text{Green + SWIR}}$

Table B.1
Example of confusion matrix of the actual and predicted class values.

		Actual values	
		Positive	Negative
Predicted values	Positive	TP	FP
	Negative	FN	TN

Table B.2
Comparison of waterlogging classification accuracies statistics such as overall accuracy, kappa statistics, precision, recall, and F1-score.

2016 Pre-monsoon	Precision	Recall	F1-score	
Water	0.9	0.98	0.95	Overall accuracy = 87.78%
waterlogging	0.73	0.98	0.85	Kappa statistics = 0.82
Other class	0.97	0.73	0.85	
2021 Pre-monsoon				
Water	0.87	0.93	0.90	Overall accuracy = 86.667%
waterlogging	0.81	0.81	0.81	Kappa statistics = 0.8
Other class	0.90	0.84	0.87	
1987 Post-monsoon				
Water	0.73	0.97	0.85	Overall accuracy = 83.33%
waterlogging	0.80	0.89	0.84	Kappa statistics = 0.75
Other class	0.98	0.73	0.85	
2019 Post-monsoon				
Water	0.77	0.98	0.87	Overall accuracy = 85.56%
waterlogging	0.80	0.97	0.89	Kappa statistics = 0.78
Other class	0.98	0.70	0.82	

References

Amer, R., 2021. Spatial relationship between irrigation water salinity, waterlogging, and cropland degradation in the arid and semi-arid environments. Remote Sens. 13 (6), 1047.
Ao, Y., Li, H., Zhu, L., Ali, S., Yang, Z., 2019. The linear random forest algorithm and its advantages in machine learning assisted logging regression modeling. J. Pet. Sci. Eng. 174, 776–789.
Arnous, M.O., Green, D.R., 2015. Monitoring and assessing waterlogged and salt-affected areas in the eastern Nile Delta region, Egypt, using remotely sensed multi-temporal data and GIS. J. Coast. Conserv. 19 (3), 369–391.
Arogyaswamy, R., 1971. Some geological factors influencing the behaviour of the Kosi. Rec. Geol. Surv. India 96 (2), 42–52.
Bassi, N., Kumar, M.D., Sharma, A., Pardha-Saradhi, P., 2014. Status of wetlands in India: A review of extent, ecosystem benefits, threats and management strategies. J. Hydrol. Reg. Stud. 2, 1–19.
Chakravorty, B., Jha, R., 2022. Assessment of spatio-temporal variation of waterlogged areas lying between lower Gandak and Burhi Gandak basins, India. J. Indian Soc. Remote Sens. 50 (3), 583–596.
Chowdhury, S.R., Kumar, A., Brahmanand, P., Ghosh, S., Mohanty, R., Jena, S., Sahoo, N., Panda, G., 2011. Application of Bio-Drainage for Reclamation of Waterlogged Situations in Deltaic Orissa. Directorate of water Management.
Davidson, N.C., Fluet-Chouinard, E., Finlayson, C.M., 2018. Global extent and distribution of wetlands: trends and issues. J. Marine Freshwater Res. 69 (4), 620–627.
Dinka, M.O., Ndambuki, J.M., 2014. Identifying the potential causes of waterlogging in irrigated agriculture: the case of the wonji-shoa sugar cane plantation (Ethiopia). Irrigation Drainage 63 (1), 80–92.
Gardner, R.C., Finlayson, C., 2018. Global wetland outlook: state of the world's wetlands and their services to people. In: Ramsar Convention Secretariat. pp. 2020–2025.
Gole, C.V., Chitale, S.V., 1966. Inland delta building activity of Kosi river. J. Hydraul. Div. 92 (2), 111–126.
Goswami, S., 2019. Geomorphology, Hydrology and Human-Environment Interactions in the Kosi Megafan, India (Ph.D. thesis). University of Oregon.
Hjerdt, K., McDonnell, J., Seibert, J., Rodhe, A., 2004. A new topographic index to quantify downslope controls on local drainage. Water Resour. Res. 40 (5).
Hofmann, G., Silva, R., Weber, E., Barbosa, A., Oliveira, L., Alves, R., Hasenack, H., Schossler, V., Aquino, F., Cardoso, M., 2023. Changes in atmospheric circulation and evapotranspiration are reducing rainfall in the Brazilian Cerrado. Sci. Rep. 13 (1), 11236.
Jain, V., Sinha, R., 2005. Response of active tectonics on the alluvial Baghmati River, Himalayan foreland basin, eastern India. Geomorphology 70 (3–4), 339–356.
Khalil, M.M., Abotalib, A.Z., Farag, M.H., Rabei, M., Abdelhady, A.A., Pichler, T., 2021. Poor drainage-induced waterlogging in Saharan groundwater-irrigated lands: Integration of geospatial, geophysical, and hydrogeological techniques. Catena 207, 105615.
Kumar, R., Jain, V., Babu, G.P., Sinha, R., 2014. Connectivity structure of the Kosi megafan and role of rail-road transport network. Geomorphology 227, 73–86.
Liu, K., Harrison, M.T., Yan, H., Liu, D.L., Meinke, H., Hoogenboom, G., Wang, B., Peng, B., Guan, K., Jaegermeyr, J., et al., 2023. Silver lining to a climate crisis in multiple prospects for alleviating crop waterlogging under future climates. Nature Commun. 14 (1), 765.
Mann, H.B., 1945. Nonparametric tests against trend. Econometrica 245–259.
Morth, 2022. Ministry of Road Transport and Highways, Annual Report 2021–2022. Government of India Ministry of road transport and highways.
Neeti, N., Pandey, A., Chowdary, V., 2022. Delineation of waterlogged areas using geospatial technologies and google earth engine cloud platform. In: Geospatial Technologies for Land and Water Resources Management. Springer, pp. 125–135.
Nhai, 2022. National Highway Authority of India, Annual Report 2021–2022. Government of India Ministry of road transport and highways.

Pandey, A., Singh, S.K., Nathawat, M., 2012. Analysing the impact of anthropogenic activities on waterlogging dynamics in Indo-Gangetic plains, Northern Bihar, India. Int. J. Remote Sens. 33 (1), 135–149.

Poulin, B., Davranche, A., Lefebvre, G., 2010. Ecological assessment of Phragmites australis wetlands using multi-season SPOT-5 scenes. Remote Sens. Environ. 114 (7), 1602–1609.

Prajapati, G.S., Rai, P.K., Mishra, V.N., Singh, P., Shahi, A.P., 2021. Remote sensing-based assessment of waterlogging and soil salinity: a case study from Kerala, India. Results Geophys. Sci. 7, 100024.

Shafizadeh-Moghadam, H., Khazaei, M., Alavipanah, S.K., Weng, Q., 2021. Google earth engine for large-scale land use and land cover mapping: An object-based classification approach using spectral, textural and topographical factors. GIScience Remote Sens. 58 (6), 914–928.

Singh, A., Naik, M.N., Gaurav, K., 2022. Drainage congestion due to road network on the Kosi alluvial Fan, Himalayan Foreland. Int. J. Appl. Earth Obs. Geoinf. 112, 102892.

Singh, K.V., Setia, R., Sahoo, S., Prasad, A., Pateriya, B., 2015. Evaluation of NDWI and MNDWI for assessment of waterlogging by integrating digital elevation model and groundwater level. Geocarto Int. 30 (6), 650–661.

Sinha, R., Gaurav, K., Chandra, S., Tandon, S., 2013. Exploring the channel connectivity structure of the August 2008 avulsion belt of the Kosi river, India: Application to flood risk assessment. Geology 41 (10), 1099–1102.

Taddeo, S., Dronova, I., Depsky, N., 2019. Spectral vegetation indices of wetland greenness: Responses to vegetation structure, composition, and spatial distribution. Remote Sens. Environ. 234, 111467.

Tassi, A., Vizzari, M., 2020. Object-oriented lulc classification in google earth engine combining snic, glcm, and machine learning algorithms. Remote Sens. 12 (22), 3776.

Tiner, R.W., Lang, M.W., Klemas, V.V., 2015. Remote Sensing of Wetlands: Applications and Advances. CRC Press.

Walne, C.H., Reddy, K.R., 2021. Developing functional relationships between soil waterlogging and corn shoot and root growth and development. Plants 10 (10), 2095.

Wells, N.A., Dorr, Jr., J.A., 1987. Shifting of the Kosi river, northern India. Geology 15 (3), 204–207.