



Machine learning algorithm-based risk assessment of riparian wetlands in Padma River Basin of Northwest Bangladesh

Abu Reza Md. Towfiqul Islam¹ · Swapan Talukdar² · Susanta Mahato² · Sk Ziaul² · Kutub Uddin Eibek¹ · Shumona Akhter¹ · Quoc Bao Pham^{3,4} · Babak Mohammadi⁵ · Firoozeh Karimi⁶ · Nguyen Thi Thuy Linh^{7,8}

Received: 15 November 2020 / Accepted: 1 February 2021 / Published online: 2 March 2021
© The Author(s), under exclusive licence to Springer-Verlag GmbH, DE part of Springer Nature 2021

Abstract

Wetland risk assessment is a global concern especially in developing countries like Bangladesh. The present study explored the spatiotemporal dynamics of wetlands, prediction of wetland risk assessment. The wetland risk assessment was predicted based on ten selected parameters, such as fragmentation probability, distance to road, and settlement. We used M5P, random forest (RF), reduced error pruning tree (REPTree), and support vector machine (SVM) machine learning techniques for wetland risk assessment. The results showed that wetland areas at present are declining less than one-third of those in 1988 due to the construction of the dam at Farakka, which is situated at the upstream of the Padma River. The distance to the river and built-up area are the two most contributing drivers influencing the wetland risk assessment based on information gain ratio (InGR). The prediction results of machine learning models showed 64.48% of area by M5P, 61.75% of area by RF, 62.18% of area by REPTree, and 55.74% of area by SVM have been predicted as the high and very high-risk zones. The results of accuracy assessment showed that the RF outperformed than other models (area under curve: 0.83), followed by the SVM, M5P, and REPTree. Degradation of wetlands explored in this study demonstrated the negative effects on biodiversity. Therefore, to conserve and protect the wetlands, continuous monitoring of wetlands using high resolution satellite images, feeding with the ecological flow, confining built up area and agricultural expansion towards wetlands, and new wetland creation is essential for wetland management.

Keywords Wetland dynamics · Drivers of wetland conversion · Wetland risk assessment · MNDWI · Machine learning techniques

Introduction

Wetlands are considered to be the world's most precious natural ecosystem, accounting for just 6% of the global land surface (Acreman et al. 2007; Whyte et al. 2018; Li et al. 2019). They strongly patronize global biodiversity (Savickis

et al. 2016), smoothen the water cycle (Bullock and Acreman 2003), and rein climatic change (Karim et al. 2016). Wetlands have been widely reported as the “world's kidneys,” life support system and origin of modern civilization (Wu and Lane 2017). They serve as critical functions including ecological, socioeconomic, and recreational purposes, and management

Responsible editor: Alexandros Stefanakis

✉ Nguyen Thi Thuy Linh
linhntt@tlu.edu.vn

¹ Department of Disaster management, Begum Rokeya University, Rangpur 5400, Bangladesh

² Research Scholars, Department of Geography, University of Gour Banga, Malda, India

³ Environmental Quality, Atmospheric Science and Climate Change Research Group, Ton Duc Thang University, Ho Chi Minh City, Vietnam

⁴ Faculty of Environment and Labour Safety, Ton Duc Thang University, Ho Chi Minh City, Vietnam

⁵ Department of Physical Geography and Ecosystem Science, Lund University, Sölvegatan 12, SE-223 62 Lund, Sweden

⁶ Department of Geography, environment and sustainability, University of North Carolina-Greensboro, Greensboro, NC, USA

⁷ Ho Chi Minh City University of Technology (HUTECH) 475A, Dien Bien Phu, Ward 25, Binh Thanh District, Ho Chi Minh City, Vietnam

⁸ Thuyloi University, 175 Tay Son, Dong Da, Hanoi, Vietnam

of catastrophic events (White and Kaplan 2017; Asomani-Boateng 2019; Sutton-Grier and Sandifer 2018). Moreover, wetlands act as flood prevention, stream flow maintenance, fertile farming lands (Rippon 2009), fish and wildlife safe environment (Wu and Lane 2017), sinks and sources of carbon (Kayranli et al. 2010), and multiple services to human well-being (Junk et al. 2013). Despite the benefits of wetlands for human well-being, wetlands around the world are becoming vulnerable due to abrupt changes in climate regulation and anthropological footprints, such as water control through dams, agricultural expansion, over population growth, excess groundwater withdrawal, resource over-exploitation, pollution, and rapid urbanization (Karim et al. 2016; Islam et al. 2016; Lefebvre and Laille 2019). Besides, a lack of public awareness about the advantages of the wetlands and their ecological benefits (Bai et al. 2011; Adekola and Mitchell 2011) and the decline in groundwater levels (Das and Pal 2017; Mahato and Pal 2019b; Pal et al. 2020) are important factors of wetland vulnerability, because the continuous decline of the groundwater table is categorized as one of the major factors causing the conversion of wetlands. Evidence indicates that almost half of the wetlands around the world have disappeared, 3.7 times faster than other diverse ecosystems over the past 150 years (Davidson 2014).

In such circumstances, wetlands are considered the most vulnerable ecosystems on the planet (Poff et al. 2002; Bates et al. 2008). Physical processes of wetland risk assessment threaten the magnitude of wetlands to provide these ecosystem services to humanity (Zedler and Kercher 2005). Hence, wetlands are natural ecosystems of global and local significance that need support for their effective conservation plans (Keddy 2010). Therefore, it is of paramount importance to investigate the spatiotemporal dynamics of wetland risk assessment and its underlying drivers for identifying sustainable wetland use policies.

Wetland risk assessment appears when the risk exposures of a specific wetland phenomenon approach stress the potential of wetlands to sustain the impact or the endeavors are required to reduce it (Miller and Fujii 2010). Recently, risk assessment to the integrity of wetlands including decrease in wetland area, degradation and rise in gas emissions has drawn substantial attention among the researchers (Davidson 2014; Song et al. 2014; Pal and Talukdar 2018; Debanshi and Pal 2020). Some scholars have investigated the physical processes that drive wetland risk assessment (Pal and Talukdar 2018; Saha and Pal 2019). Moreover, several studies of wetland dynamics and risk assessment have concentrated on variations in land use/cover, wetland habitat, and water level (Khaznadar et al. 2009; Pal and Talukdar 2018; Jiang et al. 2017; Malekmohammadi and Jahanishakib 2017; Debanshi and Pal 2020). The outcomes of these studies have been utilized to aid with the wetlands protection.

Recently, wetland risk assessment based on the image analysis has attracted the researchers' attention (Wu et al. 2014; Talukdar and Pal 2018). Although high-resolution images have a great advantage to show the wetland unit at the micro-level, and limited availability of such images may hinder preparing quality datasets (Sanyal et al. 2017). Landsat images have been extensively applied for this type of research in the data-scarce regions, like Bangladesh, India, instead of such high-resolution images (Talukdar and Pal 2020). In general, analysis of multi-date images can give an authentic wetland map and reliability in the flooded wetland and floodplain areas (Talukdar and Pal 2017). Various image matrices, such as normalized difference vegetation index (NDVI), normalized difference water index (NDWI), and modified normalized difference water index (MNDWI), were used in previous studies to map water bodies (Zhou et al. 2012; Mahato and Pal 2018; Mahato and Pal 2019a). However, each index has some merits and demerits, showing comparative water availability, and wet and dry soil conditions (Gao 1996). Evidence reported that these indices are not equally significant for all types of water bodies due to the heterogeneous landscape on the earth.

Machine learning algorithms have gained popularity in remote sensing-based studies because of their reliability and competency to classify wetlands (Rogan et al. 2008; Xu et al. 2018; Whyte et al. 2018). Advanced machine learning classifiers, such as random forest (RF) (Felton et al. 2019) and support vector machine (SVM), have been successfully used to classify and map wetlands (Breiman 2001; Cortes and Vapnik 1995; Maxwell et al. 2018). Prediction of wetland risk zones can be achieved by these data-driven machine-learning techniques (Miller et al. 2016; Ekberg et al. 2017; Wardrop et al. 2019; Saha and Pal 2019; Defne et al. 2020). SVM achieves higher accuracy than other common classifiers using a limited training data-set, which makes them particularly attractive for wetland classification (Zang et al. 2012; Mohammadpour et al. 2015; Liu et al. 2018). RF can handle high-dimensional datasets, which makes it attractive for processing remote sensing dataset (Pham et al. 2021; Maxwell et al. 2016; Mellor et al. 2013). There are good examples of successful application of both SVM (Petropoulos et al. 2012; Zhang and Xie 2013; Szantoi et al. 2013; Sadeghi et al. 2012; Lin et al. 2013) and RF (Pham et al. 2019; Sesnie et al. 2010; Mellor et al. 2013; Maxwell et al. 2016) in environmental studies. Furthermore, SVM has been used extensively in wetland studies (Dronova et al. 2012; Betbeder et al. 2013; Zhang and Xie 2013). RF is also frequently used in risk zoning and assessment studies (Pourghasemi and Kerle 2016; Pal and Debanshi 2021; Ghosh and Das 2020).

However, previous literature reported that very rare studies on the application of machine learning algorithms were applied in the wetland risk assessment. Therefore, based on this line of thinking, the objectives of the present study were set to

(1) analyze the spatiotemporal dynamics of wetlands for the periods of 1988–2018, (2) explore the drivers, which are responsible for wetlands degradation, and (3) predict the wetland risk assessment through the integration of machine learning algorithms and selected drivers.

The present study contains two novelties which are objective and technical aspects. In the case of objective novelty, this study is the first study in Bangladesh, which has explored the long-term wetland dynamics in the Teesta river basin. In addition, ten drivers were identified, which were responsible for wetland transformation in the present study. Based on these drivers, the risk zones of wetlands were predicted. In other words, the wetland areas having high-risk zone can be considered the most vulnerable areas of the wetlands, which can be transformed to other land use types in the near future. Therefore, probable areas in the wetlands, which are going to be transformed, have been identified precisely. This finding could be the opportunity to the authorities and planners to focus more on the high-risk zones of the wetlands to limit the wetland degradation and keep the low-risk wetlands as safe or non-convertible wetlands. In the case of technical novelty, in the present study, we used remote sensing techniques (MNDWI) to monitor the wetland dynamics for the period of 1988–2018, which is a very new work in Bangladesh. Furthermore, for the identification of drivers of wetland changes, we employed remote sensing techniques and machine learning algorithms like ANN, RF. We predicted probable wetland fragmented model, which is also a driver using ANN model. To the best of authors' knowledge, the modelling of probable wetland fragmented model is the first and novel work to present the wetland areas which are going to be fragmented. In addition, major influencing drivers for wetland transformation were identified and quantified using information gain ratio. Based on the identified drivers along with probable wetland fragmented model, the M5P, SVM, RF, and REPTree algorithms were utilized to predict the wetland risk zones, which are very novel and first in the Bangladesh. Therefore, it can be stated that present study is technically robust and sound.

Study area

The Padma river is the main downstream stretch of the Ganges river, which flows more than 2562 km originating from Gangotri glacier of the Himalaya. The basin area of Ganges river is considered one of the densely populated inhabitants on the Earth. About 407 million population of five countries, namely China, India, Nepal, Bhutan, and Bangladesh, either directly or indirectly derive benefits from the Ganges river. This river plays a vital role in the socio-ecological settings of these countries. The stretch of the Padma river runs for 108 km before the confluence with river Meghna at

Chandpur. The aggregated discharge of the Ganges river and Brahmaputra flows through the course of the Padma river, which is on an average 30,000 m³/s and it can be 75,000 m³/s during bank full stage (Hydraulics D, DHI (FAP 24) 1996). The elevation of the river course decreases by 5 cm/km (Sarker and Thorne 2006). This study has been conducted on the Padma river basin and the study area covers almost 8 districts, including Pabna, Shirajganj, Natore, Bogura, Jaypurhat, Naogaon, Rajshahi, and Chapainawabganj (Fig. 1). The study area is located between 23° 48' and 25° 18' north latitudes and 88° 27' and 89° 48' east longitudes. Annually, 900 metric tons of sediment load passes through the river out of which 60% is either silt or clay while the rest is bed load (Hossain 2010). Dewan et al. (2017) described the plan of the river as a “wandering” pattern. This region has the geo-environmental privilege to nourish the wetlands from the geological past. The monotonous plain formed by the phase after phase alluvial deposition and a small amount of slope promotes the wetland occurrences with the help of ambient monsoonal rainfall. Apart from that, the wetlands of the riparian zone regularly get the supply of water from spilling river water. As a result, this region composes favorable conditions of wetland habitat, but as mentioned earlier, the huge anthropogenic pressure would not let the wetlands to pursue ecological functions. The man-environmental conflicts are very often there in the form of draining and infilling of wetlands.

Materials and methods

Materials

LANDSAT 4-5TM (Thematic Mapper), LANDSAT 7^{ETM+} (Enhanced Thematic Mapper Plus), and LANDSAT 8OLI (Operation Land Imager) images have been downloaded from the website of USGS Earth Explorer (United State Geological Survey). The collected satellite imageries (Level 1 Terrain Corrected (L1T) product) were pre-geo-referenced to Universal Transverse Mercator (UTM) zone 45 North projection applying the WGS-84 datum. The particulars of those satellite data, which are used in this present study, are mentioned in Table 1. Geo-referencing refers to the source of errors of multi-resolution satellite imageries, and a conventional geo-referencing tool has been applied for the present study. Therefore, geometric correction has been carried out to amend the image to the UTM-46 N projection system through image processing. The basin area has been delineated with the help of Google Earth Pro. The whole study has been carried out using WEKA (version 3.8.2), ERDAS IMAGINE (version 2014), and ArcGIS 10.5 software environment. The hydrological time series are considered inclusive responses of integrated climate situations, natural geographical conditions, and human deeds (Sang and Wang 2008). Daily water flow

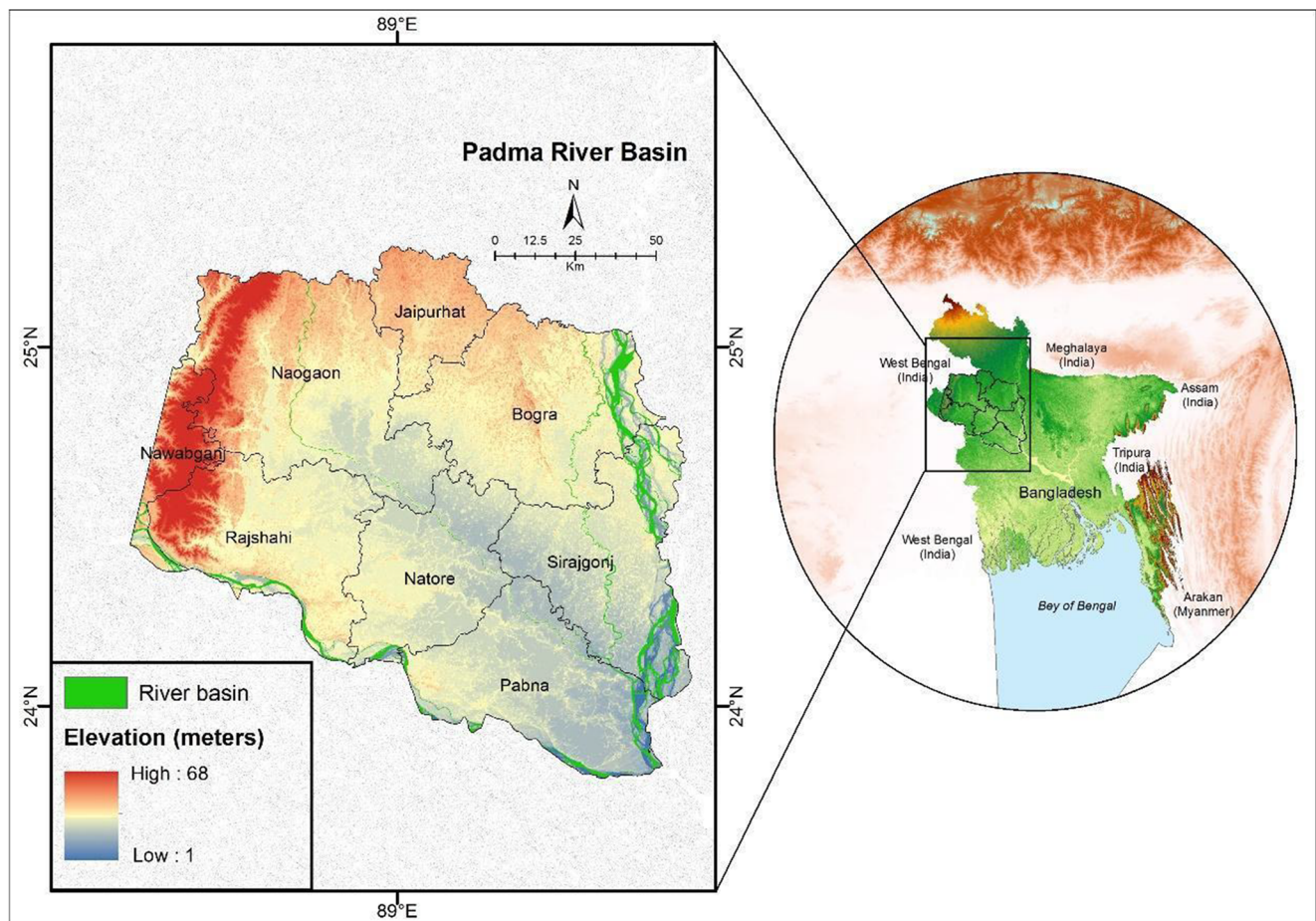


Fig. 1 Location of the study area

data (1970 to 2018) of the Padma river basin is acquired from respecting the river gauge station (Hardinge bridge station).

Methods for wetland mapping and risk assessment modelling

Wetland extraction for spatiotemporal mapping

For the identification of wetland, several satellite imageries derived indices, such as the NDWI, MNDWI, and WI, have frequently been used, while these indices are not compatible with every region for extracting accurate water bodies. The

Table 1 List of satellite imageries used for the study

| Satellite | Sensors | Path and row | Date of the imageries |
|-------------|---------|--------------|-----------------------|
| Landsat 4-5 | TM | 138/43 | 08-11-1988 |
| Landsat 4-5 | TM | | 28-11-1994 |
| Landsat 4-5 | TM | | 17-11-2000 |
| Landsat 7 | ETM+ | | 18-11-2006 |
| Landsat 7 | ETM+ | | 17-10-2012 |
| Landsat 8 | OLI | | 27-11-2018 |

distance from water feature towards other corresponding features and the spectral proximity has been considered the important parameters of satellite images based indices for extracting wetlands. The NDWI overestimates wetland areas along with impervious land and agriculture land with a comparatively higher moisture level. We found that some built-up areas are overlapped with water bodies due to complex LULC types in the study area. On the other hand, we decided to use MNDWI in the present study because it generates three types of outputs, such as (1) higher positive values in the MNDWI over water body denotes the maximum absorption of MIR (middle infrared) light compare to the NIR (near-infrared) light; (2) accordingly, the built-up area has been represented by the negative values, and (3) the negative values have been observed over the soil and vegetated areas, while the soil reflects the maximum MIR light compare to the NIR light. Similarly, the vegetated area also reflects the maximum MIR light in comparison to the green light (Jensen 2004). Therefore, the dissimilarity between the water body and the built-up area has been increased significantly rather than the output of NDWI. Consequently, the output of MNDWI generates higher values for a water body and the lower values (positive to negative) for the built-up area. The maximum

enhancement of the spectral values and brightness of the water body detected by the MNDWI have been benefited to extract the open water body, the built-up area, soil, and vegetated area more precisely, while these have been markedly concealed and removed (Xu 2006). Therefore, we used the MNDWI in the present study for resolving the flaws of NDWI. The MNDWI has been calculated using Equation 1:

$$MNDWI = \frac{\text{Green} - \text{MIR}}{\text{Green} + \text{MIR}} \quad (1)$$

where MIR represents the band 5 of Landsat 4-5TM.

Inventories

In the present study, we used wetland inventories for two purposes, such as for wetland fragmentation potentiality and for wetland risk assessment. For wetland fragmentation potentiality modeling (considered one of the important driver for wetland degradation), we, first, prepared wetland fragmented model using landscape fragmentation tool, a ArcGIS extension. For this, some steps were followed, such as (1) the wetland maps (including non-wetland areas) for all periods were integrated and prepared average wetland maps; (2) the non-wetland area was assigned as 1 and wetland area was assigned as 2 (the requirement of the landscape fragmentation tool to execute the model); (3) the landscape fragmentation maps were generated, which had six classes, such as patch, edge, perforated, small core, medium core, and large core based on the concentration of wetland pixel in a particular area; and (4) finally, we considered patch as the inventories for wetland fragmentation potentiality modeling because it is the most detached part from the concentrated or core wetlands and has the higher potentiality to be transformed or lost in near future. Then, we randomly selected 200 points from the patch as most fragmented zone, while 200 points were also collected from large core (it is considered most un-disturbed or natural wetlands) as non-fragmented zone. Then, we assigned 1 as fragmented zone and 0 as natural zone. The total 400 points were divided into training and testing datasets based 80-20 ratio. Each datasets contain both 0 and 1 value as fragmented and natural zones. Based on training datasets, the ANN model was executed to predict the fragmentation potential zones, while the testing dataset was used to validate the fragmentation potential model.

On the other hand, for wetland risk assessment, we used same landscape fragmentation map (generated using landscape fragmentation tool, an extension of ArcGIS) to prepare the inventories. In this case, the logic is that the patch and edge are the most detached parts of main or natural wetlands, which have very high risk to be converted to other land uses, while the large core of the wetlands has the very low or no risk to be converted to other land uses. Therefore, based on this logic,

we randomly selected 250 points from each fragmentation classes and assigned them as 1 and 0. Then, we divided the whole data points into training (80%) and testing (20%) datasets. The training points, which contain both 0 and 1 values, were used to extract the data from ten drivers of wetlands. Based on the training datasets, the machine learning algorithms were executed to prepare the wetland risk assessment models. Then, the validation points were used to extract the predicted data from the wetland risk assessment models, which were utilized to validate the risk assessment models.

Drivers of wetland risk assessment

Agriculture land Cropland area for all the selected years is extracted by using the normalized difference vegetation index (NDVI). The NDVI (Townshend and Justice 1986) is computed by the following equation:

$$NDVI = \frac{\text{IR} - \text{R}}{\text{IR} + \text{R}} \quad (2)$$

where the IR refers to the band 4 of Landsat 4-5TM (near-infrared), and while the band 5 is IR band of Landsat OLI. The R (red) is the 3rd and 4th band of Landsat 4-5TM and 8OLI respectively. The positive NDVI values ranging from 0 to 1 denote vegetation. The values close 1 represent the higher intensity of increasing vegetation health in the relative term.

Built-up area Wang et al. (2011) reported that built-up areas are expanding due to an unprecedented population and economic growth. Environmental problems (i.e., wetland loss) are turning to a cumulative alarming phenomenon with the expanse of impervious area (Sun and Lockaby 2012). Therefore, it is necessary to consider this parameter for modeling the wetland risk assessment. The year of 2018 was chosen for computing normalized difference built-up index (NDBI) considering the last years of post-dam phases and it was computed by using Equation 3.

$$NDBI = \frac{\text{MIR} + \text{NIR}}{\text{MIR} - \text{NIR}} \quad (3)$$

where the MIR indicates middle infrared (band 5 and 6 of Landsat 4-5TM and 8OLI respectively), NIR (near-infrared) is band 4, and 5 of Landsat 4-5 TM and 8 OLI respectively.

Soil moisture The normalized difference moisture index (NDMI) is effective to measure the moisture levels in vegetation. This index is often used in monitoring the droughts and fuel levels in fire-prone areas. It includes NIR and SWIR bands to generate a ratio aimed to mitigate brightening and atmospheric effects. NDMI is calculated by the following equation (Wilson and Sader 2002; Skakun et al. 2003)-

$$NDMI = \frac{NIR - SWIR1}{NIR + SWIR1} \quad (4)$$

where NIR indicates DN values of the near-infrared band; SWIR1 refers to DN values of the short-wave infrared 1 band.

Distance to road/railway The effect of anthropogenic activities on the wetlands is very proximate to the road, which has numerous effects such as social, economic, and ecological effects (Chomitz and Gray 1996). Therefore, the distance to the road from the wetlands has been considered the relevant causative indicator of wetland loss. The road density has been continuously increasing along the wetlands in terms of development. The Euclidean distance method was employed for preparing the distance to road and railway maps.

Population density The main source of income of the Padma river basin area is a wetland, so the population exerts pressure on the wetland. For the assessment of anthropogenic pressure on the vegetation or forest, the pixel-wise population density map was prepared based on the data of the Population Census 2001.

Slope and elevation The slope and elevation are two relevant biophysical factors for wetland risk assessment. The slope map prepared using SRTM DEM with a resolution of 30 m was split into two different slope categories. There was found 0–17.40° variation in the slope gradient over the study area. Moreover, an altitude map was also prepared from SRTM DEM data.

Fragmentation potential zones Fragmentation denotes the disintegration of the continuous large-scale landscape such as wetland into small-scale different landscape units (Huisig 2002). The extension of agricultural land, construction of roads, settlement, and germination are the main reasons for wetland fragmentation (Murungweni 2013; Pal and Talukdar 2018). The FRAGSTATS software was used for calculating different fragmentation indices. Only one landscape metric cannot describe all the perspectives of fragmentation (Davidson 1998). The number of chosen metrics can be beneficial for the explanation of landscape alteration and must be considered relative to the type of alteration (the patches) and the background matrix (the forest mosaic) carefully. In this regard, several landscape fragmentation matrices were derived to model the chances of fragmentation probability.

The metrics were selected through the literature review and the knowledge of the study area that can explain LULC alteration influence on landscape fragmentation. Therefore, in this study, several metrics were used, such as the number of patches (NP), the largest patch index (LPI), the patch density (PD), the edge density (ED), the aggregation index (AI),

landscape shape index (LSI), and perimeter area ratio (P/A ratio) for illustrating the landscape alteration between the periods of pre and post-dam. The FRAGSTAT 4.2 software was used for modeling these landscape fragmentation indices. The details of the mentioned indices are listed in Table 2.

Based on mentioned fragmentation parameters, the artificial neural network (ANN) was applied for preparing the fragmentation probability model. The ANN is an abstract mathematical and black box model. It has been applying in uncountless fields, including decision making, pattern recognition, automatic controlling systems, and robotics (Conforti et al. 2014). It can handle the complex, non-linear and unbalanced data sets. Therefore, it can imitate the functioning of the human brain and is even able to generalize and predict the output from a large number of complex inputs. For this reason, researchers across the world have widely been used to solve the different problems in different fields. The ANN model can perform like an expert, which can detect the complex predictive pattern, which is not apparent non-expert. It can act on the category, continuous and binary data without violating the assumption and characters of the data (Wang et al. 2016).

Several architectures of the neural network have widely been used (Moayedi et al. 2019; Harmouzi et al. 2019; Sevgen et al. 2019; Falah et al. 2019; Termeh et al. 2018; Zhao et al. 2019; Garosi et al. 2019; Huang and Cao 2018). In this study, the feed-forward-based multilayer perceptron (MLP) architecture was used. The standard MLP consists of three layers, such as an input, one or more hidden and output layers of non-linear activation nodes. Each layer contains many neurons or nodes, which are connected with a certain weight to every node in the next layer. Their work is to transfer the information. Thus, the neural network has been formed. The MLP uses the backpropagation algorithm for training the network until the minimum errors are achieved between the anticipated and output values of the network. Thus, the ANN model generates the results. In the present study, the following model parameters were optimized to obtain the best ANN model for train and predict the fragmentation probability model (Table 3).

Information gain ratio to assess the factors influencing the wetland risk assessment

Before start modelling process, it is important to primarily appraise the relevance of the parameters for influencing flood (Unler and Murat 2010). The importance of each collected parameter enumerates by using their statistical characteristics in this feature process. The information gain ratio (InGR) technique (Quinlan 1986) is a feature selection model, which is capable of recognizing the high-ranking parameters for explaining the flood susceptibility prediction. This model provides an InGR value to each wetland transforming drivers to evaluate its relevance. The higher the InGR value indicates

Table 2 Details specification of the landscape matrices generated using the FRAGSTAT 4.2 software

| Acronym | Name (units) | Description | Justification |
|-----------|----------------------------|---|---------------|
| NP | Number of patch | The landscape has been covered by the total numbers of patches | Fragmentation |
| PD | Patch density (per 100 ha) | The presence of the particular numbers of patches in the particular area of the landscape. | Fragmentation |
| ED | Edge density | McGarigal and Marks (1995) defined ED as the numbers of edges in respect to the landscape area that are likely to boost in the primary stages of habitat fragmentation. | Fragmentation |
| LPI | The largest patch index | The particular patch in the class has been covering the largest area, which is considered LPI (expressed as the percentage to the total landscape area). | Dominance |
| AI | Aggregation index | Bregt and Wopereis (1990) defined the AI as the ratio between the actual edge and the total numbers of possible edges. | Fragmentation |
| P/A ratio | Perimeter area ratio | The PA ratio is considered the measure of the magnitude to which the perimeter of the particular object's floor is exposed. | Fragmentation |

more influence of the drivers to wetland degradation. One of the main reasons to consider the InGR model for evaluation of the parameters is its simplicity and efficacy. However, we followed Equation 5 for executing the model.

Gain ratio(y, M)

$$= \frac{\text{Entropy}(M) - \sum_{i=1}^n \frac{|M_i|}{|M|} \text{Entropy}(M_i)}{-\sum_{i=1}^n \frac{|M_i|}{|M|} \log \frac{|M_i|}{|M|}} \quad (5)$$

where the attribute y belongs to a training point M with subsets $M_i = 1, 2, 3, \dots, n$

Wetland risk assessment using machine learning algorithms

M5P model M5P consists of a conventional decision tree having the option of linear regression functions at the nodes. The divergence metric is known as the standard deviation reduction (SDR), which is used to generate the decision tree. Moreover, a linear regression function is also applied to develop tree models. The process works through pruning, evacuation, and substitution of trees. Finally, a final tree model is constructed (Suthar and Aggarwal 2019). A tree model is generally used to predict the output of several input values after analyzing the provided data sets. Table 4 shows the

Table 3 Calculated parameters for ANN algorithms used in the study

| Parameters | ANN |
|--|------------------|
| Number of hidden layers | 6 |
| Activation function in the hidden unit | Logistic |
| Number of iteration | 1000 |
| Learning algorithm | Back propagation |
| Learning rate | 0.2 |
| Momentum | 0.3 |
| Seed | 5 |

optimized model parameters of different algorithms used in the study.

A linear regression model exists at all the leaf of the tree model for casting the non-existent value of the input data that arrive in this leaf. This model has two parts, such as in the first part, a decision-making tree has been constructed, whereas in the second part, pruning of non-essential branches and discarding of these subtrees with linear regression functions have been carried out. The construction of a model associates the output values of the training data to the input values. They explicitly analyze the patterns and relationships implicit in data based on rules and regression equations. On the contrary, the other artificial intelligent models, such as ANN and SVR, keep them as hidden (Nahm-Chung et al. 2010; Etemad-Shahidi and Ghaemi 2011; Etemad-Shahidi and Bonakdar 2009; Azadi et al. 2016). A predicted data set was created by the training data set.

Random forest Random forest (RF) is known as the modified bagging supervised machine learning method that is mainly applied for the prediction as well as the classification (Polikar 2012). Recently, the RF has been applied for time series forecasting (Qiu et al. 2017; Tyralis and Papacharalampous 2017). In this part of the study, the RF has been used for assessing the risk assessment of the wetlands. The application of ensemble machine learning for predicting wetland risk assessment is very new. The RF algorithm is a nonparametric ensemble classifier technique that works by using the algorithm of the flexible decision tree of Breiman (Breiman 2001). The RF builds decision trees, in which each tree is constructed by utilizing the bootstrap training samples (Breiman 2001). To build a better model, it is important to grow a large tree (Breiman 2001). Accordingly, it is necessary to have a requisite number of selected predictor variables at all the nodes of the trees. The number of observations at the terminal nodes of the trees would be minimum. In this method, randomly selected training data from the actual dataset through the algorithm were applied to generate the model (Catani et al. 2013a; Breiman 2001; Costache and Bui 2019; Youssef et al.

Table 4 Optimized model’s parameters of different algorithms used in the study

| Parameters | M5P | SVM | RF | REPTree |
|--------------------------------|-------|-------------------------|-----|---------|
| Batch size | 100 | 100 | | 100 |
| Minimum number of instance | 4 | | | |
| Unpruned | False | | | |
| Bag size percent | - | | 100 | |
| Number of iteration | 0.2 | | 100 | |
| Threads | - | | 1 | |
| Seed | | | 5 | 4 |
| Max depth | - | | 3 | 3 |
| Number of fold | | | | 3 |
| Minimum proportion of variance | | | | 0.001 |
| Filter type | | Normalize training data | | |
| Kernel | | Radial basis function | | |
| Gamma | | 0.18 | | |

2016). The performance of different models is dependent significantly on the optimization of the model’s parameters (Table 4). However, classification errors have been minimized by expanding each tree, whereas output has been influenced by random selection (Zabihi et al. 2016). The main objective of the RF algorithm is to observe how much error in prediction increases with the shifting of the output of data for a certain variable. Therefore, it can compute the significance of the variable, while the other variables remain unchanged (Liaw and Wiener 2002; Catani et al. 2013b; Jog et al. 2017). A strong correlation between the training data and the predicted data model and minimum error has been observed in this model of the present study.

REPTree Generally, REPTree is considered speedy decision tree learning, which generates a decision tree by using information gain or minimizing the variance. REPTree incorporates the regression tree logic and prepares multiple trees in various iterations (Table 4). Thereafter, it picks unsuitable ones from all constructed trees. That will be taken into account as the representative. The mean square error has been used in pruning the tree on the forecasting made by the tree. The REPTree, a speedy decision tree learner, constructs a decision or regression tree based on the information gain as the splitting criterion and prunes it based on decreased error pruning. It only selects values for numeric attributes once (Srinivasan and Mekala 2014; Dhakate et al. 2014; Kalmegh 2015). The correlation between training data and predicted model was correlation coefficient 0.925, MAE was 0.01, and RMSE was 0.017 (Table 5).

Support vector machine One of the widely used soft computing machines learning techniques is support vector machine (SVM), which commonly applied for solving the problems of classification, prediction, pattern recognition, and regression. Moreover, the SVM has been performed to forecast the time

series analysis and gives good performance among the artificial intelligence models. Therefore, the many branches, such as statistics, finance, and environment along with hydrology, have been started to use SVM (Adnan et al. 2017; Garsole and Rajurkar 2015; Kisi 2015; Deo et al. 2017; Costache 2019; Gong et al. 2016). The structural risk minimization and statistical machine learning process have been considered the philosophy behind the SVM development, which are two different fundamental bases of the SVM (Hamidi et al. 2015). The risk minimization decreases the upper-bound generalization error, rather than the traditional local training error. The model can be expressed by Equation 6:

$$z = f(y) = \sum_{i=1}^p w_i \phi_i = w\phi(y) \tag{6}$$

where the output of the model refers to the part of linear $P,()$ The converter is presented by the nonlinear model. The SVM model is represented by the Equation 7:

$$z = f(y) = \left\{ \sum_{i=1}^l w_i K(Y_i, Y) \right\} - c \tag{7}$$

where K is the kernel function, w_i and c represent the parameters of the model. L denotes the number of learning patterns,

Table 5 Temporal change of wetland area

| Year | Wetland area (km ²) |
|------|---------------------------------|
| 1988 | 2830.077 |
| 1994 | 3203.102 |
| 2000 | 2594.32 |
| 2006 | 2692.139 |
| 2012 | 2184.107 |
| 2018 | 771.49 |

accordingly, while the Y_i and Y represent the data vector for network learning and independent vector, respectively. The parameters of the model are ascertained with the maximizing the objective of the function.

Based on the previous literature, it can be stated that no machine learning algorithms are perfect to resolve the problems. In the present study, the applied machine learning algorithms also have some flaws, such as random forest, REPTree, and MSP have been suffered from over-fitting and less modeling speed, while the SVM has been suffered from misclassification error. Although previous literature stated that the mentioned machine learning algorithms have been applied successfully to solve the various environmental problems. However, several flaws of standalone machine learning algorithms could be overcome by employing the ensemble machine learning algorithms.

Validation

In the present study, the validation was conducted for two purposes. The first purpose was to validate the wetland maps, while the second purpose was to validate the risk assessment models. However, to validate the wetland maps, the collection of ground truth was done in two ways. First, we visited field in 2018 for collecting the wetlands location using global positioning system (GPS). For field survey, we took help from topographical map, local people's perception, and expert opinion. Therefore, based on the collected ground truth of the wetlands location, we validated the wetland maps of 2018. On the other hand, for the past wetland maps (2012 and 2006), we collected ground truth or wetland locations from very high resolution image of Google Earth. Based on the collected ground truth or reality, we employed the Kappa coefficient to compare between the satellite image-based wetland maps and ground data. Thus, the wetland validation was performed for the wetlands of 2006, 2012, and 2018. Furthermore, the wetland risk assessment models were evaluated and validated using ROC curve. To construct ROC curve, we used validation datasets (20%).

The ROC curve has been used to validate the performance of the ensemble machine learning algorithms used for modeling the wetland risk assessment. The sensitivity as the x-axis and the specificity as the y-axis are plotted to construct the ROC curve. The number of positive pixels (the pixels considered to a particularly vulnerable class was truly predicted or identified), which are correctly predicted, is considered the sensitivity of the AI model, while the number of negative pixels, which are correctly predicted, is considered the specificity (the pixels not considered to a particularly vulnerable class was truly predicted or identified). The sensitivity and specificity were calculated following Equations (8) and (9) (Costache and Bui 2020; Saha et al. 2021):

$$\text{Sensitivity} = \frac{a}{a + c} \quad (8)$$

$$\text{Specificity} = \frac{d}{b + d} \quad (9)$$

where a refers to the true positive, d represents as the true negative, b refers to the false positive, and c indicates the false negative.

The AUC of the ROC curve indicates the magnitude of the performance of the ensemble machine learning algorithms for predicting the wetland risk assessment (Costache et al. 2020; Yariyan et al. 2020). In addition, the AUC varies from 0 to 1, while the values adjacent to 1 indicate the high magnitude or satisfactory model performance (Arabameri et al. 2021).

The Pearson's correlation coefficient, mean absolute error (MAE), and root mean square error (RMSE) were utilized for evaluating the performances of machine learning algorithms during training the models of wetland risk assessment.

Figure 2 displays the overall methodology flowchart of the study which is adopted for this research.

Result and discussion

Spatiotemporal analysis of wetland

The spatiotemporal analysis of wetland has been shown in Fig. 3. Among 1988 to 2018, six figures have been analyzed by using MNDWI for assessing the change of the wetland area. The index score of MNDWI was more than 0 for water bodies (including water covered cropland) where some vegetation areas were mixed with it. The higher values show the dominance of wetlands.

Based on ground truth data and high resolution google earth images, we validated the satellite images derived wetlands by using kappa coefficient. The overall accuracy for 2006, 2012, and 2018 were 86.26%, 84.38%, and 87.6% respectively which indicate very agreement between ground reality and satellite-derived images. The images of previous years (before 2006) were not available. Therefore, based on the findings of the accuracy assessments for 2006–2018, we concluded that the extracted wetlands for 1988–2000 would have agreements with ground reality as like last years because we followed the same methods and procedures for extracting.

Figure 3 reveals that the wetland area is changing with time. Visually, the maximum wetland area has been seen in 1994. MNDWI value was also higher in 1994. In the other 5 years, the value is decreasing gradually. In 2018, the dominance of the water body is less than the previous years, so at present, the poor situation of the wetland area is a major concern for us. In 2018, the wetlands are lost in most of the area of

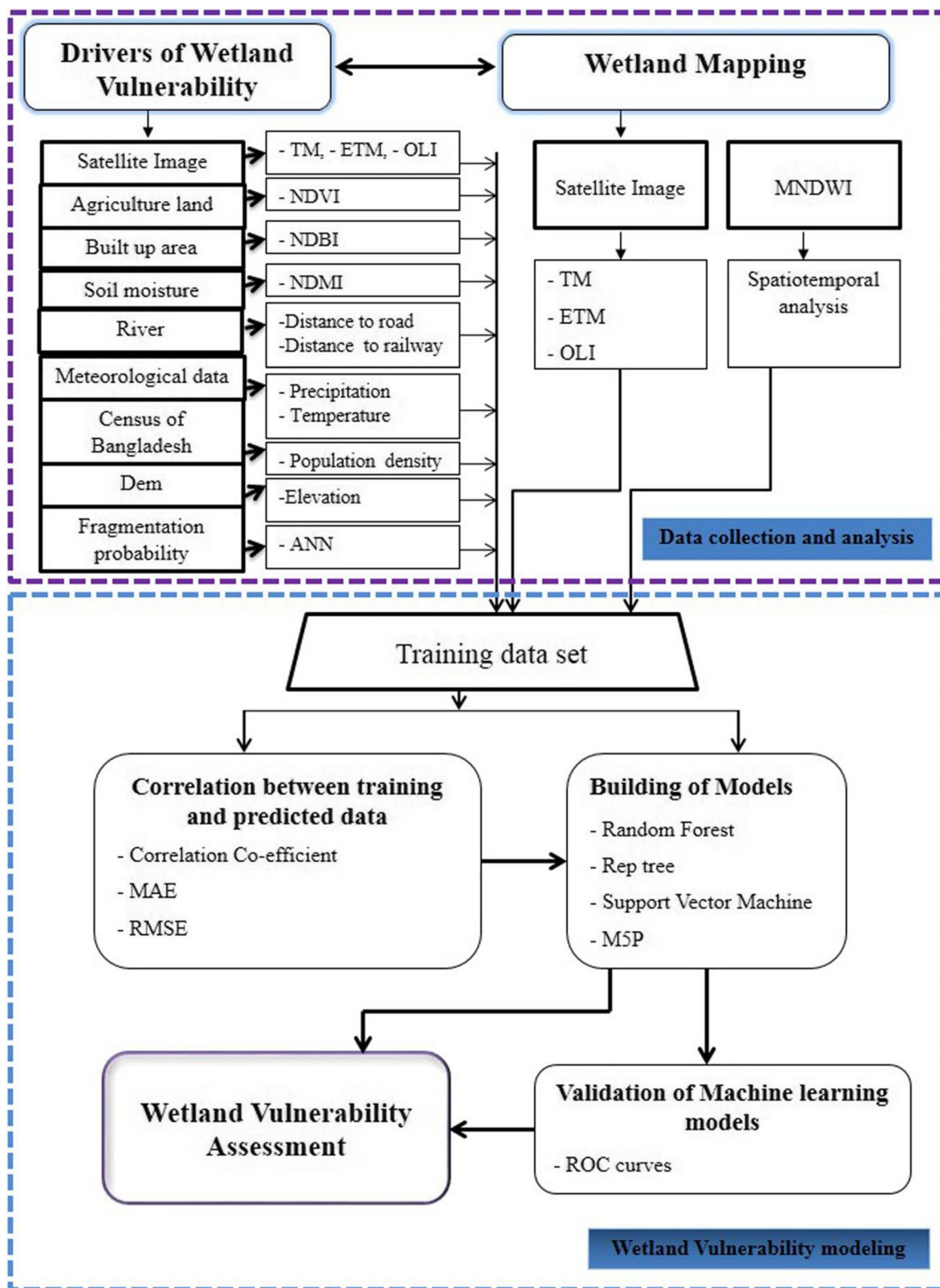


Fig. 2 Flow chart of the methods adopted for this research

the basin. Only a small portion exists in the northwestern part of the Padma basin.

The water discharge is decreasing over time due to the Farakka dam. Various anthropogenic activities are responsible

for changing climate and the rainfall is decreasing over time. The groundwater layer is depleting due to excessive pumping of water and the wetlands become vulnerable. The table below shows the area of wetland in different years.

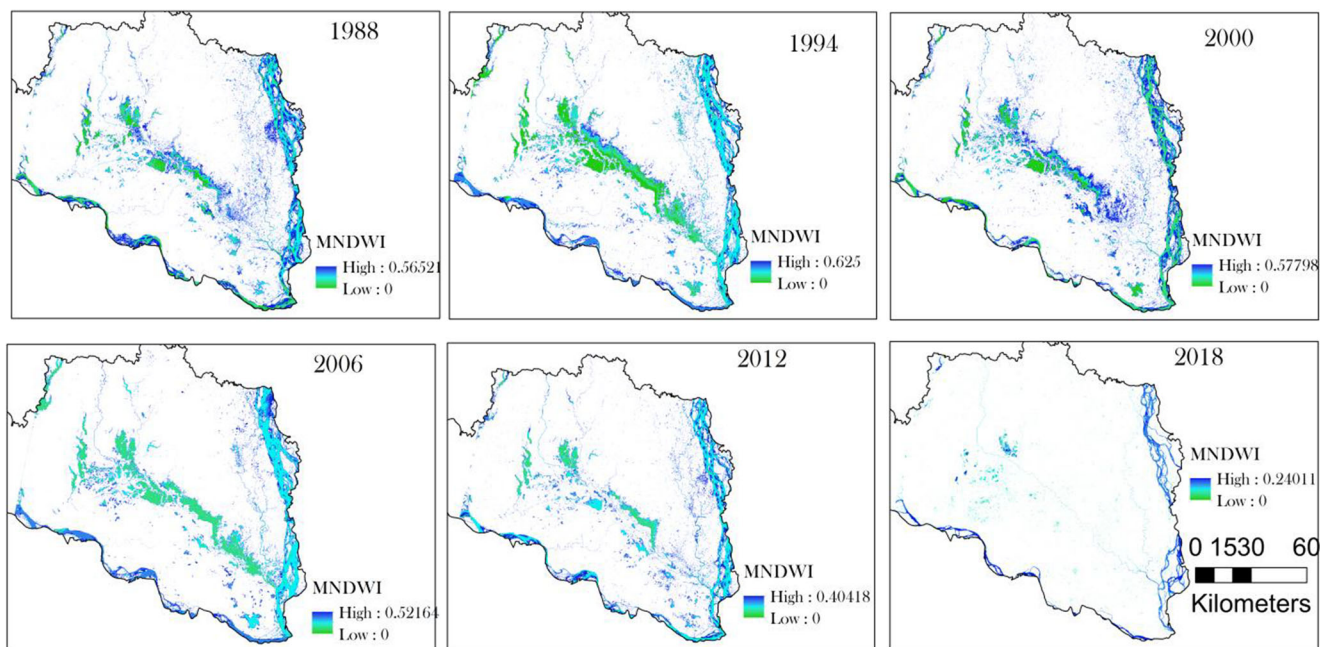


Fig. 3 Spatiotemporal dynamics of wetland

In 1988, 1994, 2000, 2006, 2012, and 2018, the total area of the wetland is 2830.07, 3203.10, 2594.32, 2692.13, 2184.10, and 771.49 km² respectively (Table 5). The area coverage of the wetland is almost one-third of the wetland of 2012 and one-fourth percent of the wetland of 1988. The table confirms how rapidly the wetlands are lost with time.

Figure 3 shows that the areas of wetlands have been changed since 1988 to 2018, but wetlands area was increased in 1994 and 2006 (Table 5). Therefore, it cannot be stated that wetlands decreased continuously; we tried to correlate the wetlands area with rainfall for similar dates in Rajshahi Upazila. It can be found that the rainfall was 60, 130, 85, 102, 36, and 55.7 mm in 1988, 1994, 2000, 2006, 2012, and 2018 respectively. Based on the rainfall data, it can be concluded that rainfall played a major role for inundation of wetlands. However, other factors also played a major role in decreasing wetlands as can be observed in 2018, because rainfall increased, but the wetlands area did not increase accordingly.

Drivers of wetland risk assessment

Wetland risk assessment has been triggered by some drivers. Some are natural and some are anthropogenic.

Elevation

The presence of wetlands is less in the high elevated area. The altitude map shows that the north-western part of the basin is a high elevated area. Figure 3 depicts that most of the wetland is in the center of the basin area where the elevation is very low. The elevation of the basin is increasing as the flow is reduced

and the sediment is deposited in the area mainly after the flood.

Slope

The slope is an important biophysical triggering factor for the wetland. Natural wetland occurs along the lower slope angle. The slope gradient varied between 0° and 17.40°. The wetlands are concentrated mainly in the lower slope area. The effect of surface slope on wetlands shows the surface area with steeper slope makes the wetland squeeze very fast than the surface having a gentle slope.

Distance to road network

The wetland closer to the road network is more exposed to human disturbances and faces several economic, social, and ecological impacts. Therefore, the proximity of wetland towards the road network could be considered an important driver of wetland risk assessment. Figure 6f shows many road networks are formed in the wetland area which triggers the wetland risk assessment of this basin. This increasing number of roads disconnected the main river from the wetlands and the water supply from the river to the wetland is decreasing.

Fragmentation probability zones

We considered fragmentation of wetlands played a major role for wetland degradation. Therefore, to assess the wetland risk zones, the fragmentation should be considered for modelling. In the present study, we prepared fragmentation probability

zones which indicate the chances to be fragmentation of wetlands. This parameter can be considered highly important for analyzing wetland risk, because high fragmented probability zones will have higher chances to be converted to other land uses. Therefore, to model the fragmentation probability zones, we considered several fragmentation related parameters, such as NP, PD, ED, PARA, LPI, and AI.

NP analyzes the frequency of segments of a distinct land-use type. Patches, the smallest unit, have the high feasibility of transformation, whereas the large core wetlands seem to be quite safe. The total number of patches was calculated in the GIS environment for 2018. Increasing the patch number indicates the high probability of fragmentation.

The PD estimates the patch frequency of the particular unit area (Du et al. 2016; Li et al. 2019; Alhamad and Alrababah 2018). The wetlands, which have experienced the higher patch density, can be considered the most vulnerable to wetland's alteration, as the wetlands have lost their integrity and spatial continuance, while the PD increases when the number of patches increases. Therefore, it can be considered an indication of the extent to which a landscape has been fragmented (Gullström et al. 2008).

The ED equalizes the aggregate of the lengths of every edge present in the landscape. It has been calculated by dividing the ED with the area of the total landscape and multiplying by 10,000 (for conversion to hectares). Increasing the edge density shows a high fragmentation probability. Gullström et al. (2008) suggested that the ED is the scientific measurement of the complicacy of the conformation of patches having the exponent of the spatial diversification in the landscape.

The perimeter-area ratio (PARA) is one of the significant measures of shape index. A shape index metric has a problem that is its variation with the size of the patch. For instance, if the shape remains unchanged, the enhancement in the patch size will reduce the perimeter-area ratio. The lower perimeter area shows lower fragmentation probability.

The LPI indicates the higher area covered by the particular patch in the landscape. Notable reduction in LPI and an increase in PD represent the highly fragmented types of land cover. The range of the LPI is 1 to 100.

The patch in the landscape has several structures. If the patch forms the single and compact structure, except for square, then the maximum aggregation can be achieved. The AI ranges from $0 < AI < 100$. The higher the value of AI indicates a lower fragmentation probability and vice-versa.

Figure 4 shows the parameter of the fragmentation probability of wetlands. We have prepared the chances of fragmentation probability based on the mentioned fragmentation indices using ANN. The output of the chances of fragmentation probability was considered one of the major inputs or parameters for wetland risk assessment modeling.

Landscape fragmentation analysis of forest areas exhibited that the core parts have faced the highest disturbances due to the major transformation taken place in this region. The change of fragmentation probability has been divided into five classes based on the fragmentation probability parameters.

Figure 6e clearly depicts that a significant portion of the core area has a moderate, high, and very high probability of fragmentation. The very high rate of fragmentation is identified in the core area of the wetland. The higher the fragmentation probability, the higher the wetland risk assessment.

The ROC curve shows the validation of wetland fragmentation method (Fig. 5). Here, the area under the curve (AUC) is 0.809, which indicates the fragmentation method is considered to be highly accurate and we can use this for an indicator of wetland risk assessment.

Built-up area

Built-up area is an important indicator of anthropogenic consequences on wetland and its deterioration. This also raises the chance of wetland loss as the settlement expands with an increase in population size. It has been evident from several studies that well access to wetland areas and closeness to the population centers increase the wetland risk assessment rate. It has been seen quite often that wetland adjacent to human settlements always remains under certain pressure. Figure 7 shows the built-up area is increasing near the wetland area. Expansion of agricultural lands and human settlement (Fig. 7i) breach the connecting channels between the main river and wetlands resulting in a curtail in the water supply to the wetland. Therefore, a large wetland area remains inundated which eventually increases wetland risk assessment.

Population distribution

As the built-up area or settlements are increasing near the wetland area, the population density of the area is also increasing people directly or indirectly increases the risk assessment of the wetlands. Figure 6c shows the population density range from low (341.516 person/sq²) to high (3816.23 person/km²). There are moderate to high population density where the wetlands are identified. The anthropogenic activity is also responsible for increasing wetland risk assessment.

Agricultural expansion

Huge pressure on wetland has emerged from the agricultural practice of the rural inhabitants. Figure 6d shows the density of agricultural land. The agricultural lands are surrounding the wetland area when the wetlands are dried up then the people

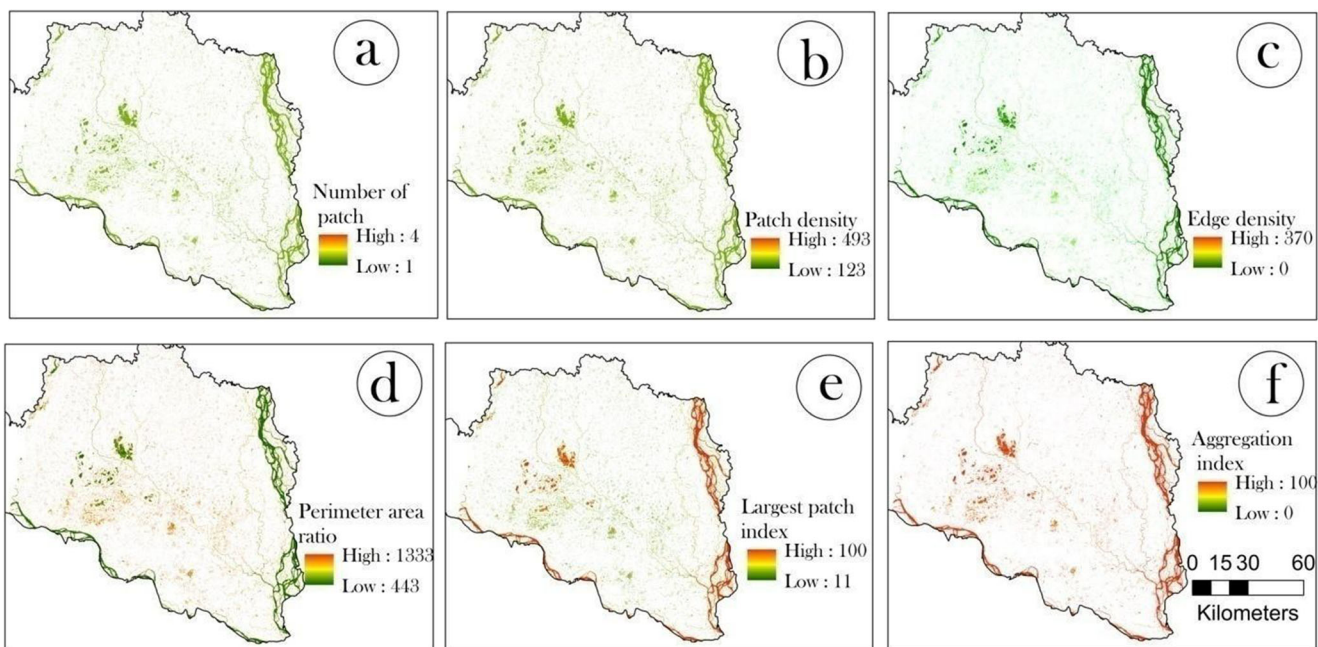


Fig. 4 Fragmentation probability parameter of wetland risk assessment: **a** number of patches, **b** patch density, **c** edge density, **d** perimeter area ratio, **e** the largest patch index, and **f** aggregation index

use the land as agricultural land. This activity damages the natural condition of the wetland.

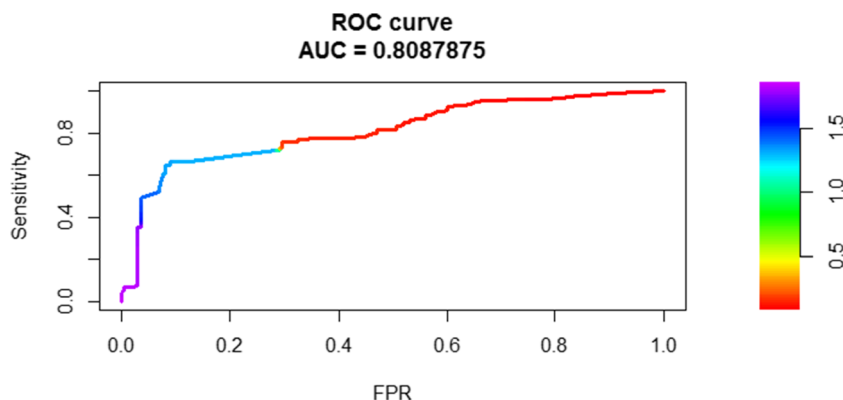
Distance to river

The wetlands which have a high distance from the river are almost in a highly vulnerable condition. The wetland areas which are situated near the river have a low chance of loss as the water supply does not disturb by distance. Figure 6 shows that the wetland which exists in 2018 has a lower distance from the river.

Distance to the railway station

The distance to the railway station is high. The wetland which was near to the rail line has been lost, as the elevation increases in those areas.

Fig. 5 The ROC curve for validating the model of chances of fragmentation probability



Wetland change rate

The wetland change rate is high in the present condition. The area of the wetland decreases with time. According to the drivers of wetland change, it is found that the existing wetland in the Padma river basin is in a vulnerable condition. The changing rate of these wetlands is very high. The drivers of wetland risk assessment clearly depict that the anthropogenic activity is mainly responsible for this high changing rate.

Risk assessment using M5P, RF, REPTree, and SVM

For machine learning algorithms viz. M5P, RF, REPTree, and SVM are used for wetland risk assessment of Padma river basin. All these models are classified into five wetland vulnerable zones (Fig. 7). In the M5P model-based map, the high and very high vulnerable area covers 356.899 km² and

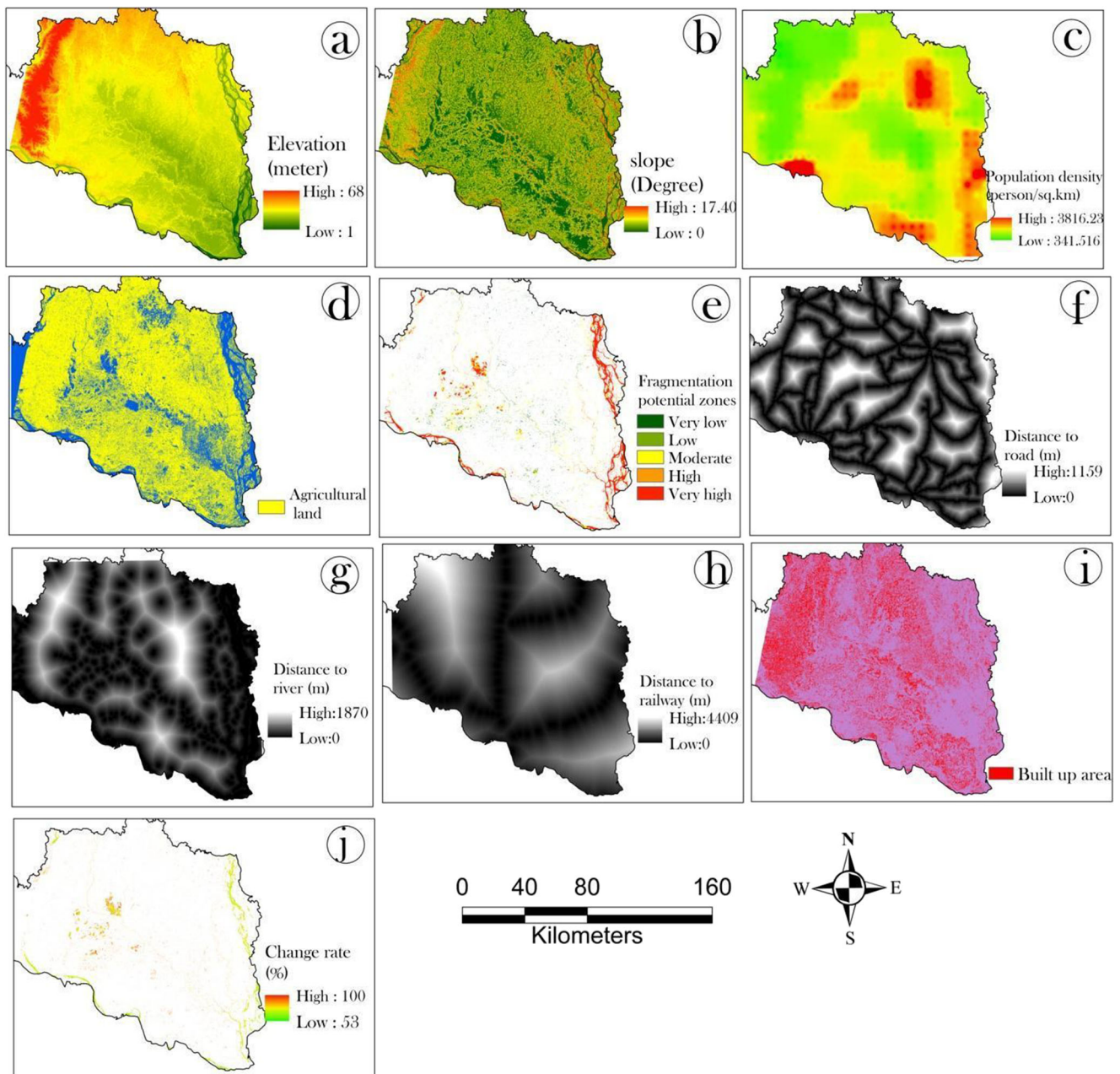


Fig. 6 The wetland risk assessment conditioning parameters, **a** elevation, **b** slope, **c** population density, **d** encroachment of agricultural land, **e** wetland fragmentation zones, **f** distance to road, **g** distance to river, **h** distance to railway, **i** encroachment of built-up areas, and **j** wetland change rate

140.588 km² respectively and the lowest area (82.722 km²) is found over the very low vulnerable zone (Table 6). RF model reveals that 99.49km², 98.50 km², 97.07km², 308.61 km², and 167.82 km² areas come under very low, low, moderate, high, and very high vulnerable zones respectively (Table 6). In the REPTree-based vulnerability model, 85.70 km², 103.81 km², 102.25km², 345.80km², and 133.94 km² areas come under very low, low, moderate, high, and very high vulnerable zones respectively (Table 6). Accordingly, area under high and very high vulnerable zone is computed as 322.98 km² and 107.105 km² respectively in case of SVM based model. In these four

models, the maximum area found over the high vulnerable zone which is more than 40% of the total area. Very high vulnerable wetlands are mostly situated in the western and northeastern part of the basin.

Validation of risk assessment model

We validated wetland risk assessment models using machine learning algorithms for training and testing data (Figs. 8a and 8b). For doing so, the pixel rank of the wetland and sensitivity was plotted in descending order, respectively on the x- and y-

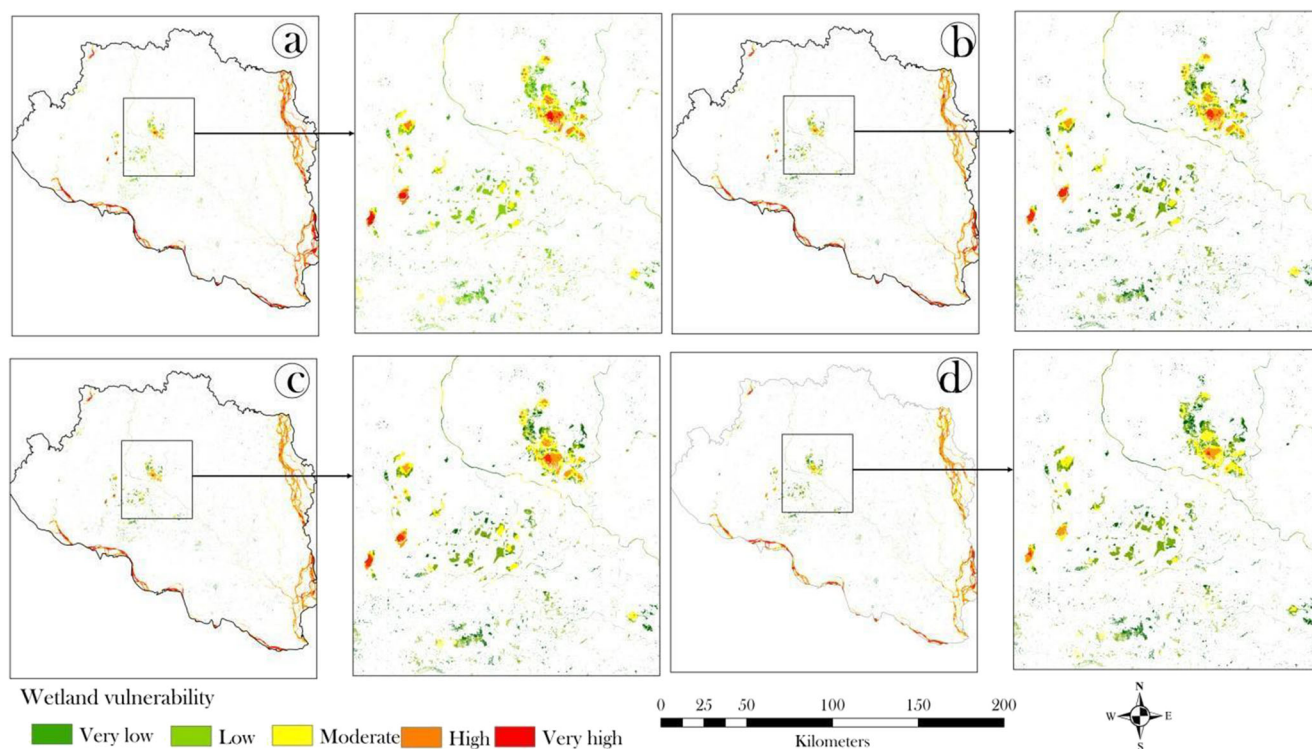
Table 6 Vulnerable wetlands of various categories in 2018 by using machine learning techniques

| Wetland risk assessment | | Very low | Low | Moderate | High | Very high |
|-------------------------|-------------------------|----------|---------|----------|----------|-----------|
| SVM | Area (%) | 13.17 | 14.46 | 16.62 | 41.86 | 13.88 |
| | Area (km ²) | 101.637 | 111.524 | 128.259 | 322.989 | 107.105 |
| | Pixel count | 112931 | 123916 | 142510 | 358877 | 119006 |
| REPTree | Area (%) | 11.11 | 13.46 | 13.25 | 44.82 | 17.36 |
| | Area (km ²) | 85.704 | 103.814 | 102.253 | 345.80 | 133.94 |
| | Pixel count | 95227 | 115349 | 113615 | 384223 | 148823 |
| RF | Area (%) | 12.9 | 12.77 | 12.58 | 40 | 21.75 |
| | Area (km ²) | 99.49 | 98.505 | 97.074 | 308.613 | 167.823 |
| | Pixel count | 110554 | 109451 | 107860 | 342904 | 186471 |
| M5P | Area (%) | 10.72 | 12.91 | 11.89 | 46.26 | 18.22 |
| | Area (km ²) | 82.7226 | 99.6093 | 91.6965 | 356.8995 | 140.5881 |
| | Pixel count | 91914 | 110677 | 101885 | 396555 | 156209 |

axis according to the accumulated 1% interval (Fig. 8). The AUC of wetland risk assessment for training data by using M5P, RF, REPTree, and SVM is respectively 0.81, 0.83, 0.77, and 0.83 (Fig. 8a), while the AUC for testing data of M5P, RF, REPTree, and SVM is 0.805, 0.808, 0.813, and 0.773 respectively (Fig. 8b). Raysid et al. (2016) suggest the accuracy category of the model based on its ROC AUC values. These categories are excellent; AUC value lies between 0.90 and 1.0, good value lies between 0.8 and 0.9, fair value lies between 0.7 and 0.8, poor value lies between 0.6 and 0.7, and the model fails if the value is less than 0.6. From

the above discussion and findings of the AUC values, it can be inferred that the results are accurate enough and have good agreement with ground truth reality while assessing and demarcating wetland risk assessment. Among all of these methods, SVM and RF are considered the best methods. The four methods are used for identifying and assessing the accuracy of the result. One method supports the other method for validation.

The correlation coefficient was 0.882, while the MAE was 0.013 and the RMSE was 0.022 for the REPTree model based on training data (Table 7), while SVM outperformed for both

**Fig. 7** The wetland risk assessment modelling using a M5P, b RF, c REPTree, and d SVM

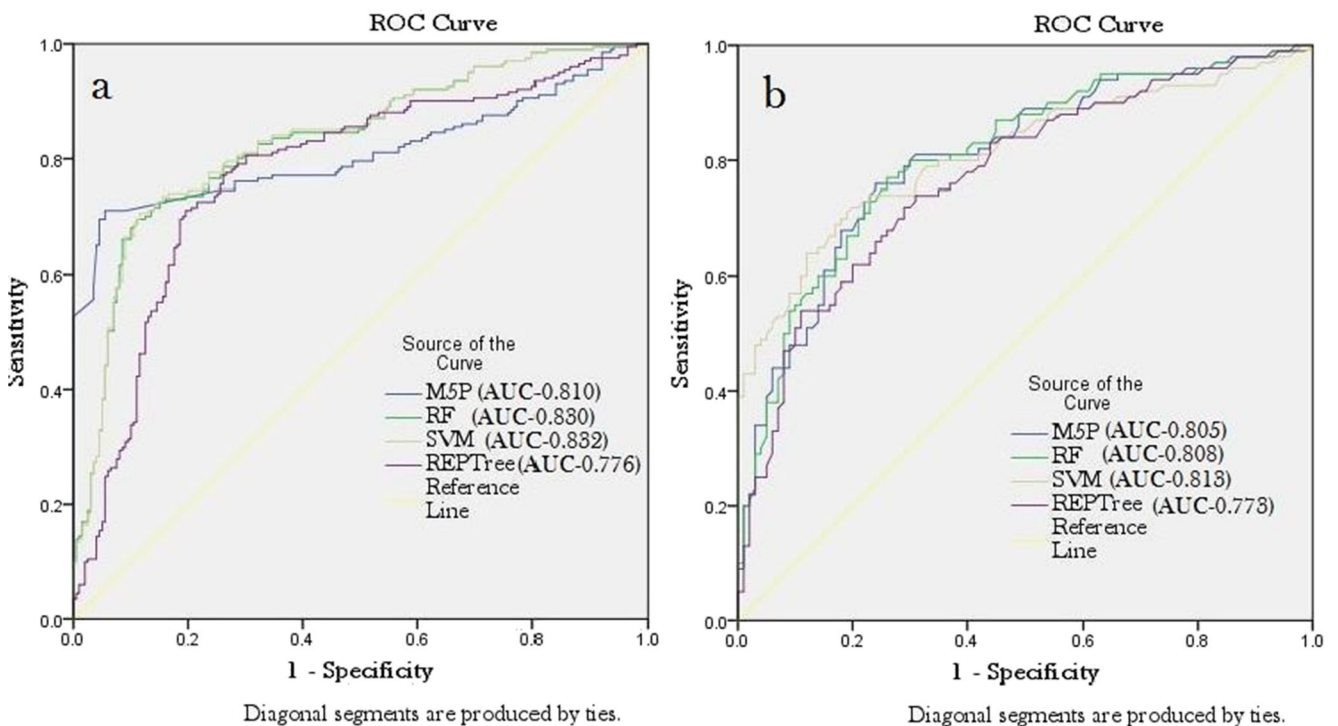


Fig. 8 ROC curve of a training data and b testing data

training and testing based model, followed by RF and M5P (Table 7)

The findings of the accuracy assessment of the wetland risk assessment models by machine learning algorithms showed that SVM outperformed modeling, followed by RF algorithms.

Discussion

The study tries to find out the present vulnerable condition of the wetlands in the Padma river basin. Khan et al. (2014) remarked that risk assessment can help us to devise strategies to reduce damages and fatalities; additionally, it helps to recognize the susceptibility of landscape people or property to damage.

Table 7 Correlation coefficient, MAE, and RMSE of M5P, RF, REPTree, and SVM

| Model | Correlation coefficient | | MAE | | RMSE | |
|---------|-------------------------|---------|----------|---------|----------|---------|
| | Training | Testing | Training | Testing | Training | Testing |
| M5P | 0.905 | 0.846 | 0.011 | 0.017 | 0.02 | 0.06 |
| SVM | 0.95 | 0.88 | 0.003 | 0.009 | 0.006 | 0.01 |
| RF | 0.925 | 0.91 | 0.01 | 0.023 | 0.017 | 0.015 |
| REPTree | 0.882 | 0.854 | 0.013 | 0.026 | 0.022 | 0.029 |

In the result section, the spatiotemporal dynamics of the wetland has been evaluated by MNDWI. The analysis shows that the area of the wetland is decreasing with time. In 1994, the wetland area was high. In 2018, the area decreased unfavorably. It has become almost one-third of the wetland area of 2012. The dynamic of wetlands makes clear that the wetlands are in a vulnerable condition. Pal and Talukdar (2018) found that 21.42% permanent and 29.34% non-perennial wetland area has been decreased after dam construction in the Punarbhaba river. The statistical analysis evaluated that the Farakka dam is largely responsible for the alteration of the inundation pattern of the river resulting in the squeezing of the wetland areas in riparian zones with time. The present situation of the wetland is very poor in this basin (Fig. 5).

Various drivers are analyzed for assessing wetland risk assessment. The wetlands of the Padma river basin are strongly impacted by agricultural extension, decreasing the river flow, population expansion, decreasing rainfall, increasing temperature, building of road and rail networks, and increasing the elevation of the area. These drivers hinder the flow of the river water to the wetlands. Expansion of agricultural land is one of the major causes of the rising physical habitat risk assessment of the wetlands (Saha and Pal 2019; Das and Pal 2017). In the Gangetic floodplain, more than 50% of the wetland area has disappeared in the last century (Panigrahy et al. 2012). The study reveals that not always agricultural invasion itself deteriorates wetland directly, more often the lean season, areal squeezing and lowering of water depth make the wetland area lucrative for agricultural encroachment. Regulations of

the river regime and water flow in riparian zones through dam condition promote this process (Talukdar and Pal 2017). Pal and Akoma (2009), Yin et al. (2012), Dronova et al. (2012), Das and Pal (2017), and Saha and Pal (2019) identified agricultural interferences as a major cause behind the crisis and loss of wetland habitat in a riparian floodplain environment. Mondal and Pal (2018) stated that infrastructural development, construction of communication network, and expansion of built-up areas are some other causes that result in fragmentation and connectivity loss of wetland areas. Due to this fragmentation, the ecosystem of inner wetland parts becomes more exposed to anthropogenic interventions (Abdu-Raheem 2014; Nindi et al. 2014; Pal and Saha 2017). From the fragmentation parameters which were used for analysis, it becomes clear that the chance of wetland fragmentation probability has been increased. Dewan et al. (2012), Kamusoko and Aniya (2007), and Hanson et al. (2007) worked on fragmentation and changes of the land use and land cover, in which fragmentation has been used for identifying wetland risk assessment. Successive fragmentation of wetlands increases the number of patches, smaller cores, and path density, whereas decreases large core area, large patch index, and aggression index which play a vital role behind the risk assessment of the wetlands. The changing rate of the wetland is very high. Due to the construction of the dam, the water scarcity has been seen in the wetland. It may disturb the morphological and breeding habits of the wetland species. Here it should be mentioned that the construction of the dam is not solely capable of reducing flood, but the embankment alongside the channel also stops the flood water to enter in the floodplain area (Pal 2015). Therefore, a significant curtail can be seen in the water supply to the wetlands (Day et al. 2011; Lukina et al. 2016) which increases the stress on wetlands. The capturing of tie channels by agricultural lands or fragmentation into smaller stretches due to rigid embankment also causes water scarcity in the wetlands in many cases (Woodruff et al. 2013). Apart from these reasons for modification of wetland hydrology, insisting on historical drought, over-extraction of wetland water also has the capability to pose severe damage to the wetland habitat condition.

The drivers which have been discussed above are responsible for the wetland risk assessment. Four machine learning techniques have been used in this study for identifying and quantifying wetland risk assessment. Some recent studies advocated for the successful application of soft computing techniques, viz. M5P, RF, REPTree, and SVM, can be accurately identified the wetland vulnerable areas (Sihag et al. 2019; Angelaki et al. 2018; Singh et al. 2017; Choubin et al. 2019). M5P, RF, REPTree, and SVM are used here for estimating the wetland area which is a vulnerable condition. All four methods showed almost similar results. The wetland was divided into five categories and among these categories, a large portion of the wetland is in high to very high vulnerable

conditions. The ROC curve shows that among the four methods, SVM and RF are considered the best methods. More than half of the total area of the wetland is at present in critical condition. The four alternative methods are used; the result of one method can help to validate others. If the four used methods assess the results almost in the same rhythm, the results may be treated as acceptable. As a result, the four methods have given results in the same rhythm which makes this study valid.

Despite the successful application of machine learning modeling of wetland vulnerability, few limitations of this research should be stated. First, wetland expansion occurs mostly during the rainy season caused by heavy precipitation, but we used only post-monsoon images due to dense cloud cover images of the rainy monsoon season. Second, the spatial resolution is an obstacle for good spatial precision, so better resolution of elevation and image data might have generated a more accurate outcome. Third, the flow modification upstream of the Padma river, lack of water supply from the river, encroachments of agricultural land, roads, settlements, and lowering of groundwater level are the major cause of wetland degradation and losses in various forms. Apart from these causes of wetland risk assessment, the hidden causes such as the absence of deep people's perception of the productivity of wetland are one of the key vital reasons. These causes will have immense negative effects on the ecological stability of wetland habitat and local inhabitant's livelihood patterns. The current research has produced good outcomes, although some limitations of this research can be overcome. Machine learning models have clearly demonstrated that a larger part of the wetland area is dominated by high to very high vulnerabilities. This study will shed light on some scientific basis for conserving and restoring wetlands. Future work should concentrate on forecasting the wetland area for a particular period and the effect of wetland risk assessment in the Padma river basin. In addition to this, field data-based risk assessment modeling can improve the precision of wetland risk assessment results. It is mentioned that we could not represent acid deposition effects on wetlands, a probably imperative indicator of wetland health aspect. Further studies are also required that include this driver in wetland vulnerability indicators. This study gives a consistent, defensible policy implication to insight into the distribution and protection of wetlands over the Padma River basin area. Our analysis also detects wetland high vulnerability and gets useful information essential to building efficient plans for conserving wetlands. Respective authorities managing the protection of these vulnerable wetlands should be cautious in monitoring their state for signs of loss and consider these regions for restoring or remediating. A

robust tool to wetland conservation by ecological managers and conservation agencies within the Padma River basin wetlands could lessen this potential risk.

Conclusion

This research, to authors' knowledge, is the first study about assessing the spatiotemporal dynamics of wetland risk assessment by the integration of machine learning and remote sensing over the Padma River basin in Bangladesh. The following major conclusions are summarized from the aforementioned findings.

1. The spatial-temporal dynamics of the wetland indicates that the wetland decreased by one-fourth of the total area.
2. Multiple drivers influence the wetland transformation. Of these, the distance to the river and built-up areas are the two most influential drivers affecting the wetland risk assessment based on the InGR.
3. The fragmentation probability is increasing due to increasing of the built-up area, agricultural activity, and overpopulation in the wetland area over time. 4. More than half of the total wetland area was categorized as very high to highly vulnerable conditions. Findings showed that the RF model appears as the best model followed by M5P and REPTree models.

From the abovementioned findings and analysis, it is evident that a large portion of the wetlands is suffering from high-risk assessment which is further supported by machine learning modeling approaches. Therefore, policymakers and corresponding authorities can easily prioritize the area that requires wetland conservation. Before taking any long-term wetland conservation strategies, special attention should be focused on the awareness built-up among inhabitants and how to be a wetland economy associated with the market economy. Our study provides a basic and vital information on the ecological and wetland conservation perspective and offers planning tools that will allow policy-makers to assign limited resources more efficiently to keep and restore the resident's wetlands.

Acknowledgements The authors would like to thank USGS for providing the satellite images.

Author contribution Abu Reza Md. Towfiqul Islam and Swapan Talukdar conceptualized the work. Abu Reza Md. Towfiqul Islam, Swapan Talukdar, Quoc Bao Pham, Babak Mohammadi, Firoozeh Karimi, and Nguyen Thi Thuy Linh provided overall guidance and continuous examination of the work. Susanta Mahato, Sk Ziaul, Kutub Uddin Eibek, and Shumona Akhter prepared datasets and carried out the analysis. Swapan Talukdar, Susanta Mahato, Sk Ziaul, Kutub Uddin Eibek, Shumona Akhter, Quoc Bao Pham, Babak Mohammadi, Firoozeh Karimi, and Nguyen Thi Thuy Linh initiated manuscript writing. All authors discussed the results and contributed to the final manuscript.

Data availability All the data and materials related to the manuscript are published with the paper, and available from the corresponding author upon request.

Declarations

Ethics approval and consent to participate Not applicable.

Consent for publication All the co-authors agreed to publish the manuscript.

Competing interests The authors declare no competing interests.

References

- Abdu-Raheem BO (2014) Improvisation of instructional materials for teaching and learning in secondary schools as predictor of high academic standard. *Nigerian J Soc Stud* 17(1):131–143
- Acreman MC, Fisher J, Stratford CJ, Mould DJ, Mountford JO (2007) Hydrological science and wetland restoration: some case studies from Europe. *Hydrol Earth Syst Sci* 11(1):158–169
- Adekola O, Mitchell G (2011) The Niger Delta wetlands: threats to ecosystem services, their importance to dependent communities and possible management measures. *Int J Biodivers Sci Ecosyst Serv Manag* 7(1):50–68
- Adnan RM, Yuan X, Kisi O, Anam R (2017) Improving accuracy of river flow forecasting using LSSVR with gravitational search algorithm. *Adv Meteorol* 2017:1–23
- Alhamad MN and Alrababah M, 2018. Quantify spatial heterogeneity using patch indices based on remote sensing data. In *EGU General Assembly Conference Abstracts* (Vol. 20, p. 321).
- Angelaki, A., Singh Nain, S., Singh, V. and Sihag, P., 2018. Estimation of models for cumulative infiltration of soil using machine learning methods. *ISH J Hydraul Eng* pp.1–8.
- Arabameri A, Pal SC, Costache R, Saha A, Rezaie F, Danesh AS, Pradhan B, Lee S, Hoang N-D (2021) Perdition of gully erosion susceptibility mapping using novel ensemble machine learning algorithms. *Geomat Nat Haz Risk* 12(1):469–498
- Asomani-Boateng R (2019) Urban wetland planning and management in Ghana: a disappointing implementation. *Wetlands* 39(2):251–261
- Azadi S, Amiri H, Rakhshandehroo GR (2016) Evaluating the ability of artificial neural network and PCA-M5P models in predicting leachate COD load in landfills. *Waste Manag* 55:220–230
- Bai J, Huang L, Yan D, Wang Q, Gao H, Xiao R, Huang C (2011) Contamination characteristics of heavy metals in wetland soils along a tidal ditch of the Yellow River Estuary, China. *Stoch Env Res Risk* A 25(5):671–676
- Bates, B., Kundzewicz, Z. and Wu, S., 2008. *Climate change and water*. Intergovernmental panel on climate change secretariat.
- Betbeder J, Gond V, Frappart F, Baghdadi NN, Briant G, Bartholomé E (2013) Mapping of Central Africa forested wetlands using remote sensing. *IEEE J Sel Top Appl Earth Obs Remote Sens* 7(2):531–542
- Bregt AK, Wopereis MCS (1990) Comparison of complexity measures for choropleth maps. *Cartogr J* 27(2):85–91
- Breiman L (2001) Random Forests. *Mach Learn* 45(1):5–32
- Bullock, A. and Acreman, M., 2003. The role of wetlands in the hydrological cycle.
- Catani F, Lagomarsino D, Segoni S, Tofani V (2013a) Landslide susceptibility estimation by random forests technique: sensitivity and scaling issues. *Natural Hazards and Earth System. Sciences* 13(11):2815
- Catani F, Lagomarsino D, Segoni S, Tofani V (2013b) Landslide susceptibility estimation by random forests technique: sensitivity and scaling issues. *Nat Hazards Earth Syst Sci* 13(11):2815
- Chomitz K, Gray D (1996) Roads, lands, markets, and deforestation: a model of land use in Belize. *World Bank Econ Rev* 10:487–512

- Choubin B, Borji M, Mosavi A, Sajedi-Hosseini F, Singh VP, Shamshirband S (2019) Snow avalanche hazard prediction using machine learning methods. *J Hydrol* 577:123929
- Conforti M, Pascale S, Robustelli G, Sdao F (2014) Evaluation of prediction capability of the artificial neural networks for mapping landslide susceptibility in the Turbolo River catchment (northern Calabria, Italy). *Catena* 113:236–250
- Cortes C, Vapnik V (1995) Support-vector networks. *Mach Learn* 20(3):273–297
- Costache R (2019) Flash-Flood Potential assessment in the upper and middle sector of Prahova river catchment (Romania). A comparative approach between four hybrid models. *Sci Total Environ* 659:1115–1134
- Costache T, Bui DT (2019) Spatial prediction of flood potential using new ensembles of bivariate statistics and artificial intelligence: A case study at the Putna river catchment of Romania. *Sci Total Environ* 691:1098–1118
- Costache R, Bui DT (2020) Identification of areas prone to flash-flood phenomena using multiple-criteria decision-making, bivariate statistics, machine learning and their ensembles. *Sci Total Environ* 712:136492
- Costache R, Pham QB, Avand M, Linh NTT, Vojtek M, Vojteková J, Lee S, Khoi DN, Nhi PTT, Dung TD (2020) Novel hybrid models between bivariate statistics, artificial neural networks and boosting algorithms for flood susceptibility assessment. *J Environ Manage* 265:110485
- Das RT, Pal S (2017) Exploring geospatial changes of wetland in different hydrological paradigms using water presence frequency approach in Barind Tract of West Bengal. *Spat Inf Res* 25(3):467–479
- Davidson C (1998) Issues in measuring landscape fragmentation. *Wildlife Soc Bull (1973-2006)* 26(1):32–37
- Davidson NC (2014) How much wetland has the world lost? Long-term and recent trends in global wetland area. *Mar Freshw Res* 65(10):934–941
- Day J, Ibáñez C, Scarton F, Pont D, Hensel P, Day J, Lane R (2011) Sustainability of Mediterranean deltaic and lagoon wetlands with sea-level rise: the importance of river input. *Estuar Coasts* 34(3):483–493
- Debanshi S, Pal S (2020) Wetland delineation simulation and prediction in deltaic landscape. *Ecol Indic* 108:105757
- Defne Z, Aretxabaleta AL, Ganju NK, Kalra TS, Jones DK, Smith KE (2020) A geospatially resolved wetland risk assessment index: synthesis of physical drivers. *PLoS One* 15(1):e0228504
- Deo RC, Kisi O, Singh VP (2017) Drought forecasting in eastern Australia using multivariate adaptive regression spline, least square support vector machine and M5Tree model. *Atmos Res* 184:149–175
- Dewan AM, Yamaguchi Y, Rahman MZ (2012) Dynamics of land use/cover changes and the analysis of landscape fragmentation in Dhaka Metropolitan, Bangladesh. *GeoJournal* 77(3):315–330
- Dewan A, Corner R, Saleem A, Rahman MM, Haider MR, Rahman MM, Sarker MH (2017) Assessing channel changes of the Ganges-Padma River system in Bangladesh using Landsat and hydrological data. *Geomorphology* 276:257–279
- Dhakate PP, Patil S, Rajeswari K, Abin D (2014) Preprocessing and classification in WEKA using different classifiers. *Int J Eng Res Appl* 4(8):91–93
- Dronova I, Gong P, Clinton NE, Wang L, Fu W, Qi S, Liu Y (2012) Landscape analysis of wetland plant functional types: the effects of image segmentation scale, vegetation classes and classification methods. *Remote Sens Environ* 127:357–369
- Du S, Xiong Z, Wang YC, Guo L (2016) Quantifying the multilevel effects of landscape composition and configuration on land surface temperature. *Remote Sens Environ* 178:84–92
- Ekberg MLC, Raposa KB, Ferguson WS, Ruddock K, Watson EB (2017) Development and application of a method to identify salt marsh risk assessment to sea level rise. *Estuar Coasts* 40(3):694–710
- Etemad-Shahidi A, Bonakdar L (2009) Design of rubble-mound breakwaters using M5/machine learning method. *Appl Ocean Res* 31:197–201
- Etemad-Shahidi A, Ghaemi N (2011) Model tree approach for prediction of pile groups scour due to waves. *Ocean Eng* 38:1522–1527
- Falah F, Rahmati O, Rostami M, Ahmadisharaf E, Daliakopoulos IN and Pourghasemi HR, 2019. Artificial neural networks for flood susceptibility mapping in data-scarce urban areas. In *Spatial modeling in GIS and R for Earth and Environmental Sciences* (pp. 323–336). Elsevier.
- Felton BR, O'Neil GL, Robertson MM, Fitch GM, Goodall JL (2019) Using random forest classification and nationally available geospatial data to screen for wetlands over large geographic regions. *Water* 11(6):1158
- Gao BC (1996) NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sens Environ* 58(3):257–266
- Garosi Y, Shekhabadi M, Conoscenti C, Pourghasemi HR, Van Oost K (2019) Assessing the performance of GIS-based machine learning models with different accuracy measures for determining susceptibility to gully erosion. *Sci Total Environ* 664:1117–1132
- Garsole P and Rajurkar M, 2015. Streamflow forecasting by using support vector regression. In *Proc., 20th Int. Conf. of Hydraulics, Water Resources and River Engineering*.
- Ghosh S, Das A (2020) Wetland conversion risk assessment of East Kolkata Wetland: a Ramsar site using random forest and support vector machine model. *J Clean Prod* 275:123475
- Gong Y, Zhang Y, Lan S, Wang H (2016) A comparative study of artificial neural networks, support vector machines and adaptive neuro fuzzy inference system for forecasting groundwater levels near Lake Okeechobee, Florida. *Water Resour Manag* 30(1):375–391
- Gullström M, Bodin M, Nilsson PG, Öhman MC (2008) Seagrass structural complexity and landscape configuration as determinants of tropical fish assemblage composition. *Mar Ecol Prog Ser* 363:241–255
- Hamidi O, Poorolajal J, Sadeghifar M, Abbasi H, Maryanaji Z, Faridi HR, Tapak L (2015) A comparative study of support vector machines and artificial neural networks for predicting precipitation in Iran. *Theor Appl Climatol* 119(3–4):723–731
- Hanson T, Brunsfeld S, Finegan B, Waits L (2007) Conventional and genetic measures of seed dispersal for *Dipteryx panamensis* (Fabaceae) in continuous and fragmented Costa Rican rain forest. *J Trop Ecol* 23(6):635–642
- Harmouzi H, Nefeslioglu HA, Rouai M, Sezer EA, Dekayir A, Gokceoglu C (2019) Landslide susceptibility mapping of the Mediterranean coastal zone of Morocco between OuedLaou and El Jebha using artificial neural networks (ANN). *Arab J Geosci* 12(22):696
- Hossain MY (2010) Morphometric relationships of length-weight and length-length of four Cyprinid small indigenous fish species from the Padma River (NW Bangladesh). *Turk J Fish Aquat Sci* 10(1):131–134
- Huang C, Cao J (2018) Impact of leakage delay on bifurcation in high-order fractional BAM neural networks. *Neural Netw* 98:223–235
- Huising EJ (2002) Wetland monitoring in Uganda. *Int Arch Photogr Remote Sens Spatial Inform Sci* 36:127–135
- Hydraulics D, *DHI (FAP 24)*, (1996). *Bed Material Sampling in Ganges, Padma, Old Brahmaputra and Jamuna* (No. 8). Special Report.
- Islam S, Cenacchi N, Sulser TB, Gbegbelegbe S, Hareau G, Kleinwechter U, Mason-D'Croz D, Nedumaran S, Robertson R, Robinson S, Wiebe K (2016) Structural approaches to modeling the impact of climate change and adaptation technologies on crop yields and food security. *Global Food Sec* 10:63–70

- Jensen JR (2004) Introductory digital image processing: a remote sensing perspective, 3rd edn. Prentice Hall, Toronto, Canada
- Jiang W, Lv J, Wang C, Chen Z, Liu Y (2017) Marsh wetland degradation risk assessment and change analysis: a case study in the Zoige Plateau, China. *Ecol Indic* 82:316–326
- Jog A, Carass A, Roy S, Pham DL, Prince JL (2017) Random forest regression for magnetic resonance image synthesis. *Med Image Anal* 35:475–488
- Junk WJ, An S, Finlayson CM, Gopal B, Květ J, Mitchell SA, Mitsch WJ, Roberts RD (2013) Current state of knowledge regarding the world's wetlands and their future under global climate change: a synthesis. *Aquat Sci* 75(1):151–167
- Kalmegh S (2015) Analysis of weka data mining algorithm REPTree, simple cart and randomtree for classification of indian news. *Int J Innov Sci Eng Technol* 2(2):438–446
- Kamusoko C, Aniya M (2007) Land use/cover change and landscape fragmentation analysis in the Bindura District, Zimbabwe. *Land Degrad Dev* 18(2):221–233
- Karim F, Petheram C, Marvanek S, Ticehurst C, Wallace J, Hasan M (2016) Impact of climate change on floodplain inundation and hydrological connectivity between wetlands and rivers in a tropical river catchment. *Hydrol Process* 30(10):1574–1593
- Kayranli B, Scholz M, Mustafa A, Hedmark Å (2010) Carbon storage and fluxes within freshwater wetlands: a critical review. *Wetlands* 30(1):111–124
- Keddy PA (2010) *Wetland ecology: principles and conservation*. Cambridge University Press, Cambridge
- Khan A, Khan HH, Umar R, Khan MH (2014) An integrated approach for aquifer risk assessment mapping using GIS and rough sets: study from an alluvial aquifer in North India. *Hydrogeol J* 22(7):1561–1572
- Khaznadar M, Vogiatzakis IN, Griffiths GH (2009) Land degradation and vegetation distribution in Chott El Beida wetland, Algeria. *J Arid Environ* 73(3):369–377
- Kisi O (2015) Pan evaporation modeling using least square support vector machine, multivariate adaptive regression splines and M5 model tree. *J Hydrol* 528:312–320
- Lefebvre M and Laille P, 2019. Citizens, local politicians' and urban green space managers' trade-offs in the transition towards pesticide-free urban green spaces: a discrete choice experiment. In *Comité JEVI, Plan Ecophyto II, Ministère de la transition écologique et solidaire*.
- Li WQ, Wang D, Jiao JL, Yang KJ (2019) Effects of vegetation patch density on flow velocity characteristics in an open channel. *J Hydrodyn* 31(5):1052–1059
- Liaw A, Wiener M (2002) Classification and regression by randomForest. *R News* 2(3):18–22
- Lin Y, Shen M, Liu B, Ye Q (2013) Remote sensing classification method of wetland based on an improved SVM. *nt Arch Photogramm Remote Sens Spat Inf Sci* 1(1):179–183
- Liu T, Abd-Elrahman A, Morton J, Wilhelm VL (2018) Comparing fully convolutional networks, random forest, support vector machine, and patch-based deep convolutional neural networks for object-based wetland mapping using images from small unmanned aircraft system. *GI Sci Remote Sens* 55(2):243–264
- Lukina AO, Boutin C, Rowland O, Carpenter DJ (2016) Evaluating trivalent chromium toxicity on wild terrestrial and wetland plants. *Chemosphere* 162:355–364
- Mahato S, Pal S (2018) Changing land surface temperature of a rural Rarh tract river basin of India. *Remote Sens Appl: Society and Environment* 10:209–223
- Mahato S, Pal S (2019a) Groundwater potential mapping in a rural river basin by union (OR) and intersection (AND) of four multi-criteria decision-making models. *Nat Resour Res* 28(2):523–545
- Mahato S, Pal S (2019b) Influence of land surface parameters on the spatio-seasonal land surface temperature regime in rural West Bengal, India. *Adv Space Res* 63(1):172–189
- Malekmohammadi B, Jahanishakib F (2017) Risk assessment assessment of wetland landscape ecosystem services using driver-pressure-state-impact-response (DPSIR) model. *Ecol Indic* 82:293–303
- Maxwell AE, Warner TA, Fang F (2018) Implementation of machine-learning classification in remote sensing: an applied review. *Int J Remote Sens* 39(9):2784–2817
- Maxwell AE, Warner TA, Strager MP (2016) Predicting palustrine wetland probability using random forest machine learning and digital elevation data-derived terrain variables. *Photogrammetric Engineering & Remote Sensing* 82(6):437–447
- McGarigal K and Marks BJ, (1995). FRAGSTATS: spatial analysis program for quantifying landscape structure. USDA Forest Service General Technical Report PNW-GTR-351.
- Mellor A, Haywood A, Stone C, Jones S (2013) The performance of random forests in an operational setting for large area sclerophyll forest classification. *Remote Sens* 5(6):2838–2856
- Miller RL, Fujii R (2010) Plant community, primary productivity, and environmental conditions following wetland re-establishment in the Sacramento-San Joaquin Delta, California. *Wetl Ecol Manag* 18(1):1–16
- Miller KM, Mitchell BR, McGill BJ (2016) Constructing multimetric indices and testing ability of landscape metrics to assess condition of freshwater wetlands in the Northeastern US. *Ecol Indic* 66:143–152
- Moayedi H, Mehrabi M, Mosallanezhad M, Rashid ASA, Pradhan B (2019) Modification of landslide susceptibility mapping using optimized PSO-ANN technique. *Eng Comput* 35(3):967–984
- Mohammadpour R, Shaharuddin S, Chang CK, Zakaria NA, Ab Ghani A, Chan NW (2015) Prediction of water quality index in constructed wetlands using support vector machine. *Environ Sci Pollut Res* 22(8):6208–6219
- Mondal D, Pal S (2018) Monitoring dual-season hydrological dynamics of seasonally flooded wetlands in the lower reach of Mayurakshi River, Eastern India. *Geocarto Int* 33(3):225–239
- Murungweni FM, 2013. Effect of land use change on quality of urban wetlands: a case of Monavale wetland in Harare. *GeoinforGeostat: An Overview* S1. of, 5, p.2.
- Nahm-Chung J, Popescu I, Kelderman P, Solomatine DP, Price RK (2010) Application of model trees and other machine learning techniques for algal growth prediction in Yong dam reservoir, Republic of Korea. *J Hydroinfr* 12:262–274
- Nindi SJ, Maliti H, Bakari S, Kija H and Machoke M, 2014. Conflicts over land and water resources in the Kilombero Valley floodplain, Tanzania.
- Pal R (2015) Channel Avulsion Archives and Morphological Readjustment near the Bhagirathi-Mayurakshi Confluence in the Lower Gangatic Plain, West Bengal, India. *J Environ Earth Sci* 5(3):2224–3216
- Pal, S. and Akoma, O.C., 2009. Water scarcity in wetland area within Kandi Block of West Bengal: a hydro-ecological assessment. *Ethiop J Environ Stud Manag* 2(3).
- Pal S, Saha TK (2017) Exploring drainage/relief-scape sub-units in Atrayee river basin of India and Bangladesh. *Spat Inf Res* 25(5):685–692
- Pal S, Talukdar S (2018) Application of frequency ratio and logistic regression models for assessing physical wetland risk assessment in Punarbhaba river basin of Indo-Bangladesh. *Hum Ecol Risk Assess: An International Journal* 24(5):1291–1311
- Pal S, Debanshi S (2021) Machine learning models for wetland habitat vulnerability in mature Ganges delta. *Environ Sci Pollut Res*. <https://doi.org/10.1007/s11356-020-11413-8>

- Pal S, Kundu S, Mahato S (2020) Groundwater potential zones for sustainable management plans in a river basin of India and Bangladesh. *J Clean Prod* 257:120311
- Panigrahy S, Murthy TVR, Patel JG, Singh TS (2012) Wetlands of India: inventory and assessment at 1:50000 scale using geospatial techniques. *Current Science* 102:852–856
- Petropoulos GP, Kalaitzidis C, Vardoulakis KP (2012) Support vector machines and object-based classification for obtaining land-use/cover cartography from Hyperion hyperspectral imagery. *Comput Geosci* 41:99–107
- Pham QB, Yang TC, Kuo CM, Tseng HW, Yu PS (2019) Combining random forest and least square support vector regression for improving extreme rainfall downscaling. *Water* 11(3):451
- Pham QB, Yang TC, Kuo CM, Tseng HW, Yu PS (2021) Coupling Singular Spectrum Analysis with Least Square Support Vector Machine to Improve Accuracy of SPI Drought Forecasting. *Water Resour Manag* 1–22
- Poff NL, Brinson MM and Day JW, 2002. *Aquatic ecosystems & global climate change: potential impacts on inland freshwater and coastal wetland ecosystems in the United States*. Pew Center on Global Climate Change.
- Polikar R (2012) Ensemble learning. In: *Ensemble machine learning*. Springer, Boston, pp 1–34
- Population Census (2001) Preliminary report. Dhaka: Bangladesh Bureau of Statistics 2009:43
- Pourghasemi HR, Kerle N (2016) Random forests and evidential belief function-based landslide susceptibility assessment in Western Mazandaran Province, Iran. *Environ Earth Sci* 75(3):185
- Qiu X, Zhang L, Sugathan PN, Amarantunga GA (2017) Oblique random forest ensemble via Least Square Estimation for time series forecasting. *Inf Sci* 420:249–262
- Quinlan JR (1986) *Induction of Decision Trees*. Kluwer Academic Publishers, Dordrecht, pp 81–106
- Rasyid AR, Bhandary NP, Yatabe R (2016) Performance of frequency ratio and logistic regression model in creating GIS based landslides susceptibility map at Lompobattang Mountain, Indonesia. *Geoenvironmental Disasters* 3(1):1–16
- Rippon S (2009) ‘Uncommonly rich and fertile’ or ‘not very salubrious’? The perception and value of wetland landscapes. *Landscapes* 10(1): 39–60
- Rogan J, Franklin J, Stow D, Miller J, Woodcock C, Roberts D (2008) Mapping land-cover modifications over large areas: a comparison of machine learning algorithms. *Remote Sens Environ* 112(5): 2272–2283
- Sadeghi R, Zarkami R, Sabetrafar K, Van Damme P (2012) Use of support vector machines (SVMs) to predict distribution of an invasive water fern *Azolla filiculoides* (Lam.) in Anzali wetland, southern Caspian Sea, Iran. *Ecol Model* 244:117–126
- Saha TK, Pal S (2019) Exploring physical wetland risk assessment of Atrayee river basin in India and Bangladesh using logistic regression and fuzzy logic approaches. *Ecol Indic* 98:251–265
- Saha A, Pal SC, Arabameri A, Blaschke T, Panahi S, Chowdhuri I, Chakraborty R, Costache R, Arora A (2021) Flood Susceptibility Assessment Using Novel Ensemble of Hyperpipes and Support Vector Regression Algorithms. *Water* 13(2):241
- Sang YF, Wang D (2008) Wavelets selection method in hydrologic series wavelet analysis. *J Hydraul Eng* 39(3):295–300
- Sanyal T, Kaviraj A, Saha S (2017) Toxicity and bioaccumulation of chromium in some freshwater fish. *Hum Ecol Risk Assess: An International Journal* 23(7):1655–1667
- Sarker MH, Thorne CR (2006) Morphological response of the Brahmaputra–Padma–Lower Meghna river system to the Assam earthquake of 1950. *Braided Rivers: process, deposits, ecology and management* 21:289–310
- Savickis J, Bottacin-Busolin A, Zaramella M, Sabokrouhiyeh N, Marion A (2016) Effect of a meandering channel on wetland performance. *J Hydrol* 535:204–210
- Sesnie SE, Finegan B, Gessler PE, Thessler S, Ramos Bendana Z, Smith AM (2010) The multispectral separability of Costa Rican rainforest types with support vector machines and Random Forest decision trees. *Int J Remote Sens* 31(11):2885–2909
- Sevgen E, Kocaman S, Nefeslioglu HA, Gokceoglu C (2019) A novel performance assessment approach using photogrammetric techniques for landslide susceptibility mapping with logistic regression, ANN and random forest. *Sensors* 19(18):3940
- Sihag P, Tiwari NK, Ranjan S (2019) Prediction of unsaturated hydraulic conductivity using adaptive neuro-fuzzy inference system (ANFIS). *ISH J Hydraul Eng* 25(2):132–142
- Singh AP, Medida S, Duraisamy K (2017) Machine-learning-augmented predictive modeling of turbulent separated flows over airfoils. *AIAA J* 55:2215–2227
- Skakun RS, Wulder MA, Franklin SE (2003) Sensitivity of the thematic mapper enhanced wetness difference index to detect mountain pine beetle red-attack damage. *Remote Sens Environ* 86(4):433–443
- Song K, Wang Z, Du J, Liu L, Zeng L, Ren C (2014) Wetland degradation: its driving forces and environmental impacts in the Sanjiang Plain, China. *Environ Manag* 54(2):255–271
- Srinivasan DB, Mekala P (2014) Mining social networking data for classification using REPTree. *Int J Adv Res Comp Sci Manag Stud* 2(10):155–160
- Sun G and Lockaby BG, (2012). Water quantity and quality at the urban–rural interface. *Urban–Rural Interfaces: linking people and nature, (urbanruralinter)*, pp.29–48.
- Suthar M and Aggarwal P, (2019). Modeling CBR Value using RF and MSP Techniques. In *MENDEL* (Vol. 25, No. 1, pp. 73–78).
- Sutton-Grier AE, Sandifer PA (2018) Conservation of wetlands and other coastal ecosystems: a commentary on their value to protect biodiversity, reduce disaster impacts, and promote human health and well-being. *Wetlands* 39:1295–1130 21–8
- Szantoi Z, Escobedo F, Abd-Elrahman A, Smith S, Pearlstone L (2013) Analyzing fine-scale wetland composition using high resolution imagery and texture features. *Int J Appl Earth Obs Geoinf* 23:204–212
- Talukdar S, Pal S (2017) Impact of dam on inundation regime of flood plain wetland of punarbhaha river basin of barind tract of Indo-Bangladesh. *Int Soi Water Conserv Res* 5(2):109–121
- Talukdar S, Pal S (2018) Impact of dam on flow regime and flood plain modification in Punarbhaha River Basin of Indo-Bangladesh Barind tract. *Water Conservation Science and Engineering* 3(2):59–77
- Talukdar S, Pal S (2020) Modeling flood plain wetland transformation in consequences of flow alteration in Punarbhaha river in India and Bangladesh. *J Clean Prod* 261:120767
- Termeh SVR, Kornejady A, Pourghasemi HR, Keesstra S (2018) Flood susceptibility mapping using novel ensembles of adaptive neuro fuzzy inference system and metaheuristic algorithms. *Sci Total Environ* 615:438–451
- Townshend JRG, Justice CO (1986) Analysis of the dynamics of African vegetation using the normalized difference vegetation index. *Int J Remote Sens* 7(11):1435–1445
- Tyralis H, Papacharalampous G (2017) Variable selection in time series forecasting using random forests. *Algorithms* 10(4):114
- Unler A, Murat A (2010) A discrete particle swarm optimization method for feature selection in binary classification problems. *Eur J Oper Res* 206(3):528–539
- Wang S, Wang Y, Feng X, Zhai L, Zhu G (2011) Quantitative analyses of ammonia-oxidizing Archaea and bacteria in the sediments of four nitrogen-rich wetlands in China. *Appl Microbiol Biotechnol* 90(2): 779–787
- Wang Q, Xie H, Ngo HH, Guo W, Zhang J, Liu C, Liang S, Hu Z, Yang Z, Zhao C (2016) Microbial abundance and community in

- subsurface flow constructed wetland microcosms: role of plant presence. *Environ Sci Pollut Res* 23(5):4036–4045
- Wardrop DH, Hamilton AT, Nassry MQ, West JM, Britson AJ (2019) Assessing the relative vulnerabilities of Mid-Atlantic freshwater wetlands to projected hydrologic changes. *Ecosphere* 10(2):e02561
- White E, Kaplan D (2017) Restore or retreat? Saltwater intrusion and water management in coastal wetlands. *Ecosyst Health Sustain* 3(1):e01258
- Whyte A, Ferentinos KP, Petropoulos GP (2018) A new synergistic approach for monitoring wetlands using Sentinels-1 and 2 data with object-based machine learning algorithms. *Environ Model Softw* 104:40–54
- Wilson EH, Sader SA (2002) Detection of forest harvest type using multiple dates of Landsat TM imagery. *Remote Sens Environ* 80(3):385–396
- Woodruff JD, Martini AP, Elzidani EZ, Naughton TJ, Kekacs DJ, MacDonald DG (2013) Off-river waterbodies on tidal rivers: Human impact on rates of infilling and the accumulation of pollutants. *Geomorphology* 184:38–50
- Wu Q, Lane CR (2017) Delineating wetland catchments and modeling hydrologic connectivity using lidar data and aerial imagery. *Hydrol Earth Syst Sci* 21(7):3579
- Wu S, Kusch P, Brix H, Vymazal J, Dong R (2014) Development of constructed wetlands in performance intensifications for wastewater treatment: a nitrogen and organic matter targeted review. *Water Res* 57:40–55
- Xu H (2006) Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery. *Int J Remote Sens* 27(14):3025–3033
- Xu T, Guo Z, Liu S, He X, Meng Y, Xu Z, Xia Y, Xiao J, Zhang Y, Ma Y, Song L (2018) Evaluating different machine learning methods for upscaling evapotranspiration from flux towers to the regional scale. *J Geophys Res-Atmos* 123(16):8674–8690
- Yariyan P, Janizadeh S, Phong TV, Nguyen HD, Costache R, Le HV, Pham BT, Pradhan B, Tiefenbacher JP, (2020) Improvement of Best First Decision Trees Using Bagging and Dagging Ensembles for Flood Probability Mapping. *Water Resour Managt* 34 (9):3037–3053
- Yin L, Colman BP, McGill BM, Wright JP, Bernhardt ES (2012) Effects of silver nanoparticle exposure on germination and early growth of eleven wetland plants. *PLoS One* 7(10):e47674
- Youssef AM, Pourghasemi HR, Pourtaghi ZS, Al-Katheeri MM (2016) Landslide susceptibility mapping using random forest, boosted regression tree, classification and regression tree, and general linear models and comparison of their performance at WadiTayyah Basin, Asir Region, Saudi Arabia. *Landslides* 13(5):839–856
- Zabihi M, Pourghasemi HR, Pourtaghi ZS, Behzadfar M (2016) GIS-based multivariate adaptive regression spline and random forest models for groundwater potential mapping in Iran. *Environ Earth Sci* 75(8):665
- Zang SY, Zhang C, Zhang LJ, ZHANG YH (2012) Wetland remote sensing classification using support vector machine optimized with genetic algorithm: a case study in Honghe Nature National Reserve. *SciGeogr Sin* 32(4):434–441
- Zedler JB, Kercher S (2005) Wetland resources: status, trends, ecosystem services, and restorability. *Annu Rev Environ Resour* 30:39–74
- Zhang C, Xie Z (2013) Object-based vegetation mapping in the Kissimmee River watershed using HyMap data and machine learning techniques. *Wetlands* 33(2):233–244
- Zhao Z, Lou Y, Chen Y, Lin H, Li R, Yu G (2019) Prediction of interfacial interactions related with membrane fouling in a membrane bioreactor based on radial basis function artificial neural network (ANN). *Bioresour Technol* 282:262–268
- Zhou SS, Zhang SS, Wang JJ, Zheng X, Huang F, Li WD, Xu X, Zhang HW (2012) Spatial correlation between malaria cases and waterbodies in Anopheles sinensis dominated areas of Huang-Huai plain, China. *Parasit Vectors* 5(1):106

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.