



Machine learning models for wetland habitat vulnerability in mature Ganges delta

Swades Pal¹ · Sandipta Debanshi¹

Received: 22 May 2020 / Accepted: 26 October 2020 / Published online: 4 January 2021
© Springer-Verlag GmbH Germany, part of Springer Nature 2021

Abstract

The present study attempts to measure wetland habitat vulnerability (WHV) in the Indian part of mature Ganges delta. Predictive algorithms belonging to bivariate statistics and machine learning (ML) algorithms were applied for fulfilling the data mining and generating the models. Results show that 60% of the wetland areas are covered by moderate to very high WHV, out of which $> 300 \text{ km}^2$ belong to very high WHV followed by a high vulnerability in almost 150 km^2 . This areal coverage increases by 10–15% from phase II to phase III. On the other hand, a relatively safe situation is confined to $< 200 \text{ km}^2$. The receiver operating characteristic curve, root-mean-square error, and correlation coefficient are used to assess the accuracy of these models and categorization of habitat vulnerability. Ensemble modeling is done using the individual models having a greater accuracy level in order to increase accuracy. A field-based model of the same is prepared by gathering information directly from the field which also exhibits similar results with the algorithm-based models. Analysis of residuals in standard regression strongly supports the relevance of the selected parameters and multi-parametric models.

Keywords Wetland habitat vulnerability (WHV) · Bivariate statistics · Machine learning (ML) algorithms · Ensemble modeling and field-based modeling

Introduction

Wetlands are one of the utmost important habitat components of the environment and provider of a wide range of ecosystem services to support the socio-ecological well-being of human (Balvanera et al. 2017; Cao et al. 2018), characterized by surface depression with the appearance of water perennially, seasonally, or periodically (Gerbeaux et al. 2016; Finlayson et al. 2016). Besides some basic social, economic, and recreational services, wetlands conduct a life support system for human habitation (White et al. 2017; De Groot et al. 2018; Asomani-Boateng 2019), by regulating climate (McInnes 2016), purifying and storing water (Almuktar et al. 2018), providing safeguard to the soil erosion (Meng and Dong

2019), protecting aquatic biodiversity (Leon et al. 2018), securing foods and providing health benefits (Yikii et al. 2017; Sutton-Grier and Sandifer 2018), and reducing the risk of hazard and disaster (Belle et al. 2018). All these services make wetland to be considered as the most productive environmental component (Wondie 2018; Wu et al. 2018) and utilized by the people all across the world (Lamsal et al. 2017). Wetland can be said a booster of ecosystem service as it generates two-fifth of the global ecosystem services occupying just 6% of its surface area (Morganti et al. 2019). UNEP (2011) has recognized wetland as a highly potential natural capital for developing a sustainable green economy. In spite of having such a vital essence, wetlands are counted as one of the very fast decaying environmental components (Davidson 2018). According to Warns new report (2018), the global wetland disappearance rate has now become three times faster than any other diverse ecosystem. In their first-ever global wetland outlook, Ramsar Convention on Wetlands (2018) has reminded that 35% of global wetlands have been completely disappeared after 1975, and this rate has been increased since 2000. This report emphasizes climate change, increasing density of population, urban expansion, and changing pattern of consumption responsible for such hasten wetland loss.

Responsible Editor: Alexandros Stefanakis

✉ Sandipta Debanshi
debanshi.sandipta93@gmail.com

Swades Pal
swadespal2017@gmail.com

¹ Department of Geography, University of Gour Banga, Malda, India

However, the health of wetland habitat is increasingly becoming under threat from environmental pressure which may be termed as wetland habitat vulnerability (WHV). It can be defined as the exposure of wetland ecological function toward any particular risk phenomenon that harm the habitat condition of wetland and special endeavor is needed to resilience the natural wetland effect (Miller et al. 2010). Fang et al. (2019), Rojas et al. (2019), and Malik and Pai (2019) have marked increasing human footprint responsible for drying, infilling, and transforming of wetland area mainly into built-ups and cropland. Some other anthropogenic activities and demands like water extraction for agriculture and industry, draining of industrial and household wastewater into wetland, and harvesting wetland vegetation for fuel and livestock for food are also responsible for putting pressure on wetland habitat. The effect of such activities is very high in third world countries where the protection of wetland habitat is often overlooked and neglected by people and policy-makers (Malekmohammadi and Jahanishakib 2017). In the deltaic floodplain region, excessive groundwater extraction, river flow regulation through damming, and loss of connection with the main channel are some of the frequently identified causes behind putting pressure on wetland habitat (Pal 2016). Moreover, the nature-centric society of third world countries enhances the dependency of people over wetland (Wondie 2018), but they lack the consciousness about the potentially of wetland and proper technique of utility; therefore, sometimes, the wetland areas are considered as wasteland (Pal 2011). In this regard, the state of India does not ignite any hope as the wetlands of this country are currently facing huge population pressure (Bassi et al. 2014) as well as losing 2–3% of the wetland area each year (Prasher 2018).

With the aim of reducing the damages of wetland habitat health caused by physical and anthropogenic pressure and some other unpredicted events, researchers have been paying attention from the last decade of the previous century (Yan et al. 2016). The method of characterizing such vulnerability has constantly been debated (Metzger and Schröter 2006; Xiao et al. 2015; Malekmohammadi and Jahanishakib 2017) and nowadays, wetland vulnerability corresponds to the encompassment of the wetland characteristics by exposure and susceptibility in such a way when coping is difficult in the natural process (Miller et al. 2010; Saha and Pal 2019). Vulnerability assessment of wetland habitat refers to measuring the degree to which the wetland attends susceptibility or incapability to cope with the unfavorable climate change and other inconsistent changes like land use transformation, changes in flow regime, or overutilization of wetland resource (Pal and Talukdar 2018a). Few types of research have tried to address the issue of wetland vulnerability in terms of ecosystem health, whereas few types of research have considered the matter of wetland vulnerability in terms of its physical vulnerability (Erwin 2009; Saha and Pal 2019). Wetland

vulnerability from the perspective of ecosystem health is more connotative with the habitat condition of the wetland and signifies the possible productivity level of the wetland and its ecosystem health. Thus, the assessment of wetland habitat vulnerability involves the integration of ecological factors which promote healthy habitat condition and distressing factors which degrade the habitat condition (Trevisan et al. 2020; Bourgoin et al. 2020). In the present study, area (more than 6000 km²) such as tusk is quite difficult and time-consuming by monitoring individual wetland which leads to modeling approach with the help of space-borne data and geographical information system (GIS) technique (Debanshi and Pal 2020a). Researchers have been trying to improve the robustness of the modeling approach in vulnerability study since the development of remote sensing-based GIS modeling (El-Akmary et al. 2019). In order to do so, the bivariate statistical approach is frequently used and considered to be providing a realistic model in the field of environmental vulnerability study (Dou et al. 2019; Nohani et al. 2019). Some of these methods have already been applied while studying wetland vulnerability and have given satisfactory result (Pal and Talukdar 2018b; Saha and Pal 2019). The present study adopts a bunch of bivariate data-driven approach namely frequency ratio (FR), Shannon entropy (SE), and weight of evidence (WoE) to measure the vulnerability of wetland habitat. In recent years, machine learning (ML) or artificial intelligence (AI) (Lary et al. 2018) is considered to be able to extract information from spatial big data with a higher accuracy level (VoPham et al. 2018) especially when nonlinearity of relationship and complexity of feature space prevail in the input dataset (Xie et al. 2018; Rodriguez-Galiano et al. 2015; Jung and Lee 2019). Therefore, in the present context, five widely used ML algorithms: logistic regression (LR), reduced error pruning (REP) tree, random forest (RF), artificial neural network (ANN), and support vector machine (SVM), have been chosen to integrate habitat vulnerability driving data layers and generate spatial models of wetland habitat vulnerability.

In deltaic floodplain areas, this vulnerability assessment is much needed as the area of these wetlands frequently changes. In the present study area, the Indian part of mature Ganges delta exhibits well distinguishable floodplain characteristics. This region carries an exceptional hydro-geological setting of deltaic floodplain and has a history of nourishing the wetland habitat since the geological past (Islam 2016). Since the development of Kolkata Township during the British colonial period, infrastructural and agronomical development of latter phases, this region faces immense pressure of population which may worse wetland habitat vulnerability. Previous researches have made an attempt to investigate and explore physical wetland vulnerability in the adjacent study area having quite a similar geo-environment (Chakraborty et al. 2018; Pal and Talukdar 2018a, b; Saha and Pal 2019). Wetlands of Gangetic moribund delta have recently been explored to some extent by Paul and Pal

(2019), but the habitat vulnerability of the wetland of the Indian mature Gangetic delta has not been well explored yet on a regional scale except some of the case studies on some individual wetlands like East Kolkata Wetland (Mondal et al. 2017; Everard et al. 2019) or some Baors of North 24 Pargana district (Mandal and Kaviraj 2009). For comprehending the vulnerability problem and devising a management strategy, the first step is to assess the vulnerability of the wetland at a regional scale. But, studies regarding the wetlands of the Gangetic delta or other adjacent regions have paid lesser attention to this aspect. But, regional level assessment of wetlands habitat vulnerability is essential with advanced methods with the capability to derive the accurate result. It will be helpful for demarcating safe wetlands and vulnerable wetlands and on a regional scale, it will be resulting in the measurement and zonation of that vulnerability. Apart from that, the third world countries still lack any active mechanism to precisely map wetland vulnerability. Once this vulnerability assessment is completed, it may provide a baseline to the policy-makers while formulating a priority basis viable management strategy.

Therefore, the main objective of this study is to measure, monitor, and model the habitat vulnerability of the wetlands in the Indian part of mature Ganges delta in order to support management purposes. Measurement and monitoring of the vulnerability state on a regional scale will be helpful for identifying and demarcating the wetlands' need in management support and will lead to decision-making regarding the development of the management framework. Previous studies like Kadavi et al. (2018), Mosavi et al. (2018), Hong et al. (2018), Liu et al. (2018), Pham et al. (2019b), and Jamali (2019) have suggested, not to use just one or two models, rather compare the results of different models in order to robust the prediction of any environmental characteristics or process. Keeping this in mind, the results of both bivariate statistics and ML algorithms are compared in this study. It also tries to ensemble prediction for improving model result in order to obtain a greater accuracy level. Finally, it is attempted further to build a field data-based wetland vulnerability model for assessing the performances of the other models built. The residuals of the standard regression of field-based model are also calculated using the representative factors as a deterministic variable in order to assess the relevance of those factors.

Study area

The areal extension of this study area (Fig. 1), the Indian part of Mature Ganges delta, is from $23^{\circ} 4' 19''$ N/ $88^{\circ} 29' 27''$ E to $22^{\circ} 12' 23''$ N/ $89^{\circ} 3' 42''$ E which corresponds to a part of Ganges-Brahmaputra delta and covers 6358.21 km^2 . This region approaches the international boundary between India and Bangladesh in the east, surrounding by the Indian part of the moribund delta in the north, active Ganges delta in the south,

and Bhagirathi-Hooghly River in the west. The characteristic geological feature of this region is its almost completion of the delta formation process here (Bagchi and Mukerjee 1983). Landsat Operational Land Imagery (OLI)-based water body extraction index delineated 1008.653 km^2 of water bodies in 2018. Lesser elevation and very little slope variation allow the river to form meander and shift their channel. With time, this kind of left channel and ox-bow lakes endorse the wetland ambiance which is found very often here. Apart from that, low topographic variation of depositional plain, innumerable tie channel, and riparian environment promotes wetland occurrences. Wetlands get a supply of water mainly from precipitation and inundated water from the river. Last hundred years, rainfall data show that the average annual rainfall is 1340.48 mm , out of which 82% held in the monsoon season (June to September). According to the census of India in 2011, the percentage of urban population is 57.27 with 1.82% annual growth.

Materials and methods

Materials and data pre-processing

Google Earth imagery available for the most recent date and the Survey of India (SOI) toposheets (no. 79A/4, 79A/8, 79A/12, 79A/16, complete series of 79B, 79E/3, and 79E/4) were used as materials for preparing the base map of this study. The demarcation of mature Ganges delta and its Indian part was done based on the scheme of Bagchi (1944). Wetlands of the study area were delineated from Landsat imageries, available in the archive of Landsat imagery of the United State Geological Survey (USGS). Noise-free representative images of both seasons (pre- and post-monsoon) of the last 30 years had been downloaded from USGS (<https://earthexplorer.usgs.gov>). Prior to use those satellite images, it is highly recommended to run some pre-processing techniques in order to extract high-quality results (Krig 2016). Keeping this in mind, radiometric corrections were done following Landsat user handbook (Landsat project science office 2002) by firstly converting a pixel-wise digital number (DN) of the images to top of atmosphere (TAO) radiance using the following Eq. (1).

$$\text{TOA} = a * \text{DN} + b \quad (1)$$

where a and b are the gain and offset factors available at image metadata file.

Secondly, TOA was converted to TOA reflectance (TOAr) using Eq. (2) to remove the errors due to earth-sun distance and exo-atmospheric solar irradiance.

$$\text{TOAr} = \pi * \text{TOA} * d^2 / \text{ESUN}_\lambda * \cos\theta_s \quad (2)$$

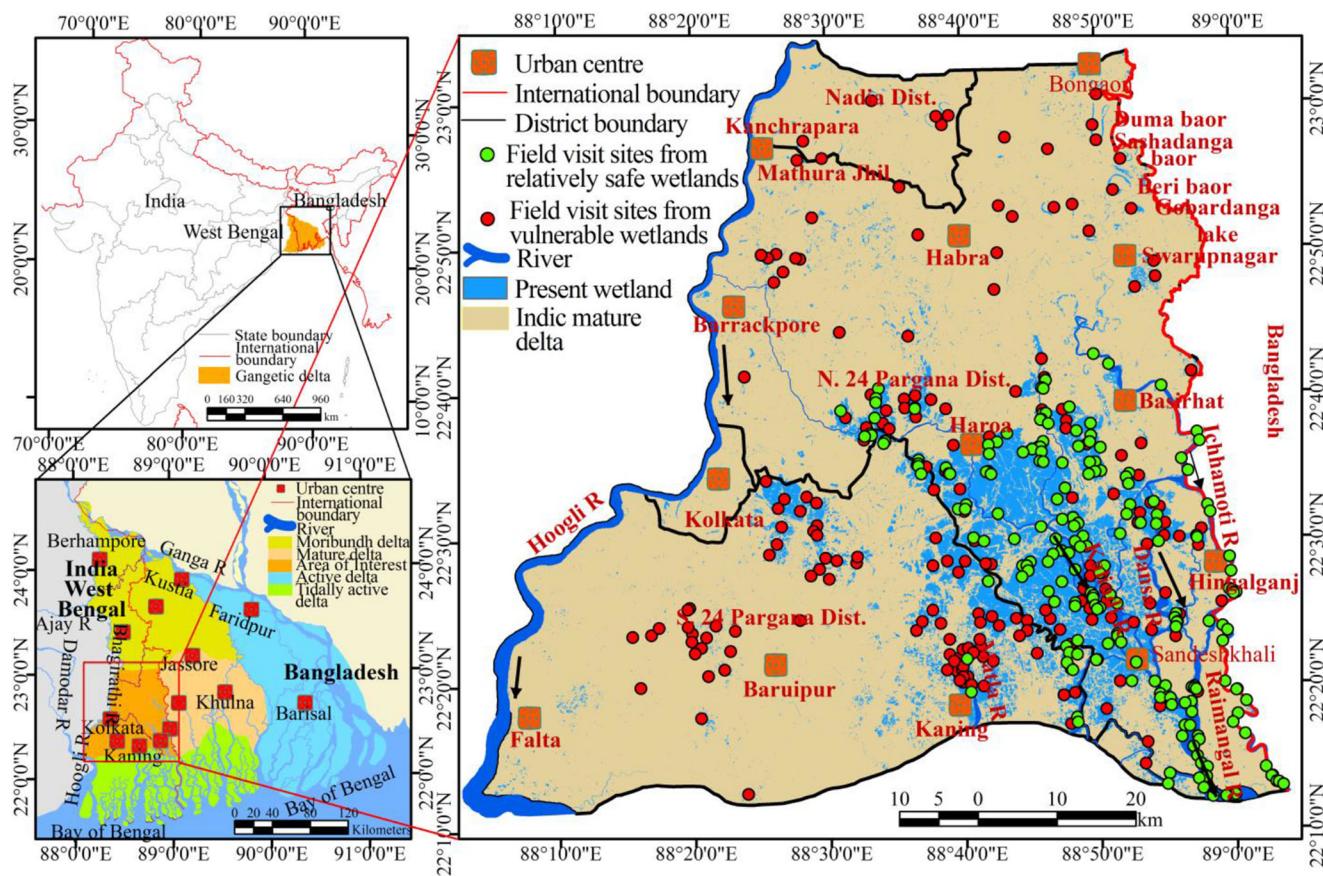


Fig. 1 Location of the study area

where d is the distance between earth and sun, $ESUN_{\lambda}$ is the mean solar exo-atmospheric irradiance, and θ_s is the solar zenith angle

Apart from the pre-processing of the satellite images, the coordinates of ground control points (GCPs) and sample sites were also converted to point features to ease the further progression.

Methods

Rationale of parameter selection and data layer preparation

Ten parameters related to wetland water phenomena and its habitat condition were selected in this study in order to determine the habitat vulnerability. These are water presence frequency (WPF), inter-phase change in WPF, water depth, and inter-phase change in depth, fragmentation, and seasonality. All these parameters are crucial for developing a wetland habitat condition (Pal and Talukdar 2018b; Saha and Pal 2019). Several other parameters are essential to be incorporated as these factors drive a wetland habitat toward vulnerable condition (Jiang et al. 2017), such as crop and vegetation presence frequency, effect of human settlement, distance from rivers, and distance from roads. WPF is the frequency of an area that appears with water and it denotes the dynamicity of

average water stagnation (Borro et al. 2014). The wetland area belongs to low WPF which may not afford a healthy wetland habitat, and in addition, inconsistency in WPF promotes instability in the wetland ecosystem (Pal and Talukdar 2018b); thus, it is important to be included in the study. The map of WPF was prepared firstly by demarcating the wetland area from Landsat image using the Re-modified Normalized Difference Water Index (RmNDWI) (Debanshi and Pal 2020b) (Eq. (3)) for the last 30 years (1989–2018). Here the entire temporal span of the study was divided into three phases each comprising 10 years, in order to monitor the dynamicity. Then, Eq. (4) was used to calculate the WPF of each water pixel.

$$RmNDWI = \frac{(b_R - b_{MIR})}{(b_R + b_{MIR})} \quad (3)$$

where b_R is the red band brightness value and b_{MIR} is the middle infrared band brightness value

$$WPF_p = \frac{\sum_{n=1}^i X_j}{N} \quad (4)$$

where WPF_p is the calculated water presence frequency for p pixel, X_j is the frequency of j th pixel in image X having water appearance, and N is the number of years taken.

Depth of water in the wetlands is also very crucial for sustaining a healthy habitat and producing ecological services (Duan and Niu 2018). The spatial thickness of water in the wetlands can be extracted from the index scores of the water body extraction index (Rokni et al. 2014; Donchyts et al. 2016) and these data have been used here to prepare the map of average water depth in the wetlands. The appearance of water in the wetlands and its depth show high dynamicity (Cong et al. 2019; Zheng et al. 2019); therefore, there changes with time need to be considered to assess the stability of habitat suitability in wetlands (Adel 2013). The changes in WPF and depth of the wetlands between two successive phases were mapped and used as two data layers of this study. In the floodplains of monsoon climate, seasonality plays and creates a vital role behind the water and the water appearance in the wetlands. It creates a massive change in the wetlands during the lean season (Mukherjee et al. 2018). Keeping this in mind, the data layer of seasonality was prepared by comparing the RmNDWI score of the same wetland during two seasons to trace the seasonal wetlands and record the reduction of water thickness in permanent wetland. The tiny wetlands in the monsoon climate become dried up so rapidly with drought season and lowering of the groundwater table (Tockner et al. 2010). The fragmentation of the wetland here is largely responsible for such phenomena. Apart from that, the larger wetlands are more favorable to compose habitat conditions than the smaller and fragmented wetlands. Therefore, the fragmentation of wetlands was included and mapped in this study with the help of the GIS tool developed by the Centre for Land Use Education and Research (CLEAR) (CLEAR 2002; Parent et al. 2007). In this way, the wetland were identified into six fragments of varying size; those are large core, medium core, small core, perforated, edge, and patch (Vogt et al. 2007). In the floodplains of developing countries, occupancy of wetland area especially the fringe part by crop field and other land use type is quite common phenomena (Pal and Akoma 2009; Jiang et al. 2017; Das and Pal 2017). The adjacency of the wetland area with crop field poses another harm to the wetland habitat by supplying and mixing the fertilizers and pesticides into the water (Ziaul and Pal 2017; Das and Pal 2018). Thus, in this study, this factor was chosen as a driver of vulnerability and the map was prepared following the same method of WPF just replacing the water indices map with the map of vegetation and cropland demarcated from normalized difference vegetation index (NDVI) (Townshend and Justice 1986) (Eq. (5)). Harm to the wetland habitat also poses by human settlement by producing and dumping toxic garbage and sewage to the wetlands, not only the wetland habitat but it causes the loss of physical wetland by expanding its area into wetlands (Mahmud et al. 2011). The dense population having a good number of large municipalities and urban settlement insists on the inclusion of this factor in the study. To develop the data layer of this factor, the

density of the built-up area was extracted from the normalized difference built-up index (NDBI) (Eq. (6)) (Zha et al. 2003). The wetlands in such kind of deltaic floodplain landscape largely depend on the nearby river for the supply of water and nutrition content during inundation (Talbot et al. 2018). The distance of the wetlands from the nearby river can be a factor as the river water may not rich in the wetland. The data layer of this map has been prepared by calculating the Euclidian distance of the wetlands from the nearby river network in the GIS environment. Construction of transport network is another issue in this region responsible for habitat fragmentation (Grzybowski and Glińska-Lewczuk 2019) as this region is well connected by closely spaced road and railway network. In addition, the faunal species of the wetland near the transport network is more prone to be harmed by the toxicity and pollution generated during transportation. This factor has been included in this study by preparing a data layer of the distance map from OpenStreetMap (OSM) road network. All the data layers of both the phases are shown in Figs. 2, and 3. It would have been better to generate the models for three phases—phase I (1989–1998), phase II (1999–2008), and phase III (2008–2018), but due to lack of the data layer of few parameter-like changes in water presence frequency and average depth, the models of wetland habitat vulnerability were generated for phase II and phase III only.

$$\text{NDVI} = \frac{(b_{\text{NIR}} - b_R)}{(b_{\text{NIR}} + b_R)} \quad (5)$$

$$\text{NDBI} = \frac{(b_{\text{MIR}} - b_{\text{NIR}})}{(b_{\text{MIR}} + b_{\text{NIR}})} \quad (6)$$

where b_R is the red band brightness value, b_{MIR} is the middle infrared band brightness value, b_{NIR} is the infrared band brightness value.

Modeling wetland habitat vulnerability

Bivariate statistics Frequency ratio

FR depicts the possibility of the occurrence of given characteristics by obtaining the relationship between existing vulnerability and its distribution in different classes of determining factors (Bonham-Carter 2014; Nohani et al. 2019). It obtains the ratio between wetland vulnerability in the class and the areal extent of the class as a percentage of total vulnerability and total study area (Meten et al. 2015). Greater FR values than 1 indicate a greater correlation of wetland vulnerability with any particular class, whereas lesser FR values than 1 indicate poor correlation (Nohani et al. 2019). The algorithm of FR can be formulated as (Eq. (9))

$$\text{FR}_{ij} = \frac{\text{FV}_{ij}}{\text{FNV}_{ij}} \quad (9)$$

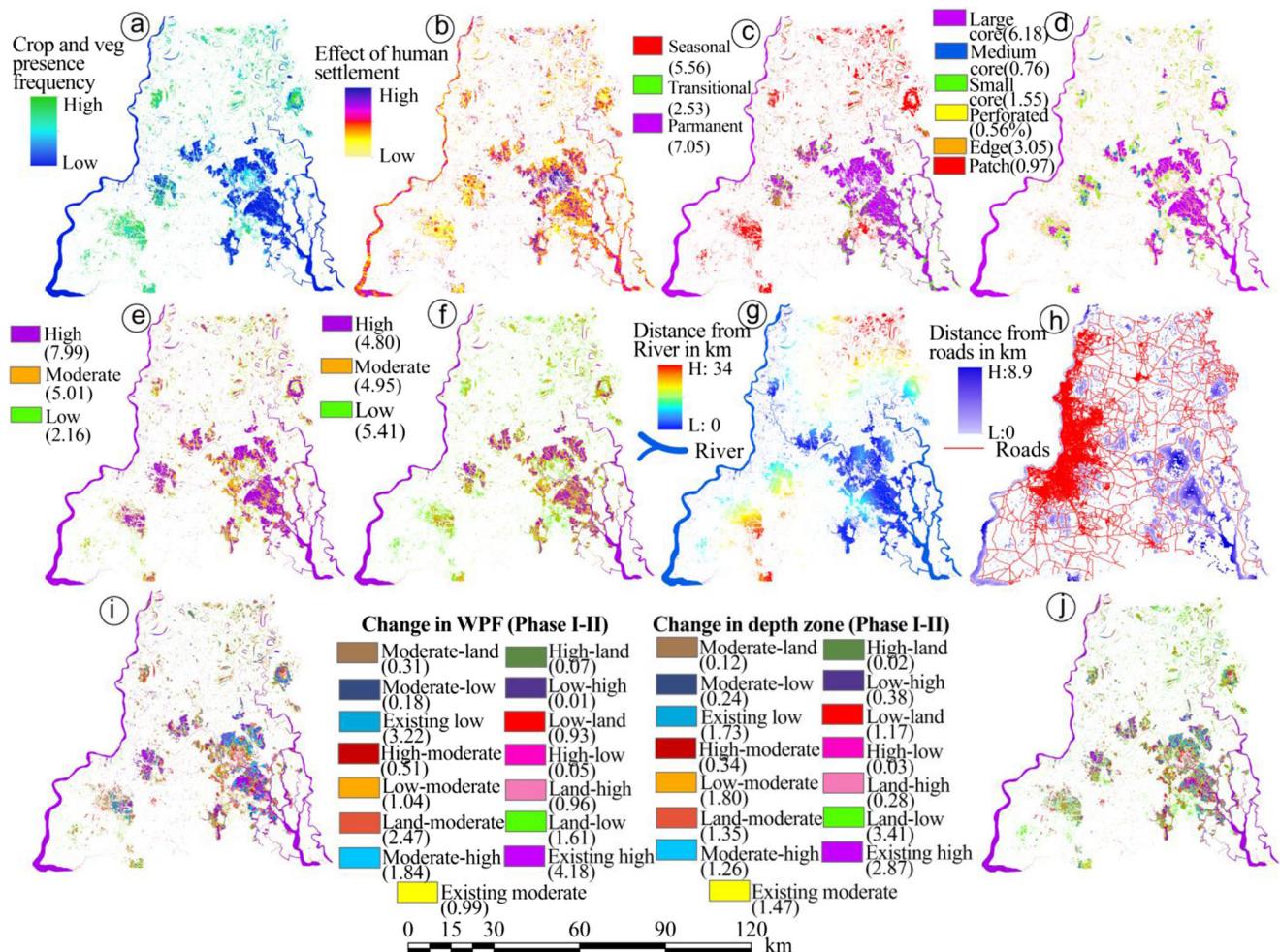


Fig. 2 Data layers of phase II: **a** crop and vegetation presence frequency, **b** effect of human settlement, **c** seasonality, **d** wetland fragmentation, **e** WPF, **f** depth zone, **g** distance from roads, **i**

change in WPF, and **j** change in the depth zone (values within parentheses show the percentage of areal coverage under different)

where FR_{ij} is the frequency ratio of i class of j factor; FV_{ij} is the frequency of the presence of vulnerable wetlands for i class of factor j ; and FNV_{ij} is the frequency of presence of non-vulnerable wetlands for i class of factor j .

To prepare the wetland habitat vulnerability model, FR values of the influencing factor were summed up (Youssef et al. 2016) (Eq. (10))

$$WHV = \sum_{j=1}^n FR_{ij} \quad (10)$$

where FR_{ij} is the weight of i class of j factor derived from frequency ratio and n is the number of influencing factor.

Shannon entropy

Entropy measures the changeability or unstable behavior of any system, in terms of its uncertainty or deviation from the ideal situation (Pourghasemi et al. 2012). It describes the abnormality between the results and the causes behind it (Arabameri et al. 2019a, b). The entropy index tries to estimate

the inter-group difference of average shear within the whole system. SE is the average imbalance in a random variable which is considered as equivalent to its information content (Al-Abadi 2017). The entropy of wetland habitat vulnerability refers to the extent that various controlling factors determine the vulnerability. Hence, the exact entropy value can be used as a weight of the index system (Jaafari et al. 2014). Here the following formulae (Eqs. (11)–(15)) have been used to calculate the information coefficient (W_j), which denotes the weight value of each influencing factor (Al-Abadi 2017). The range of W_j varies from 0 to 1 and the values closer to 1 indicate a higher degree of inconsistency.

$$E_{ji} = \frac{FR}{\sum_{j=1}^n FR} \quad (11)$$

where FR is the frequency ratio and E_{ji} is the probable density of each class.

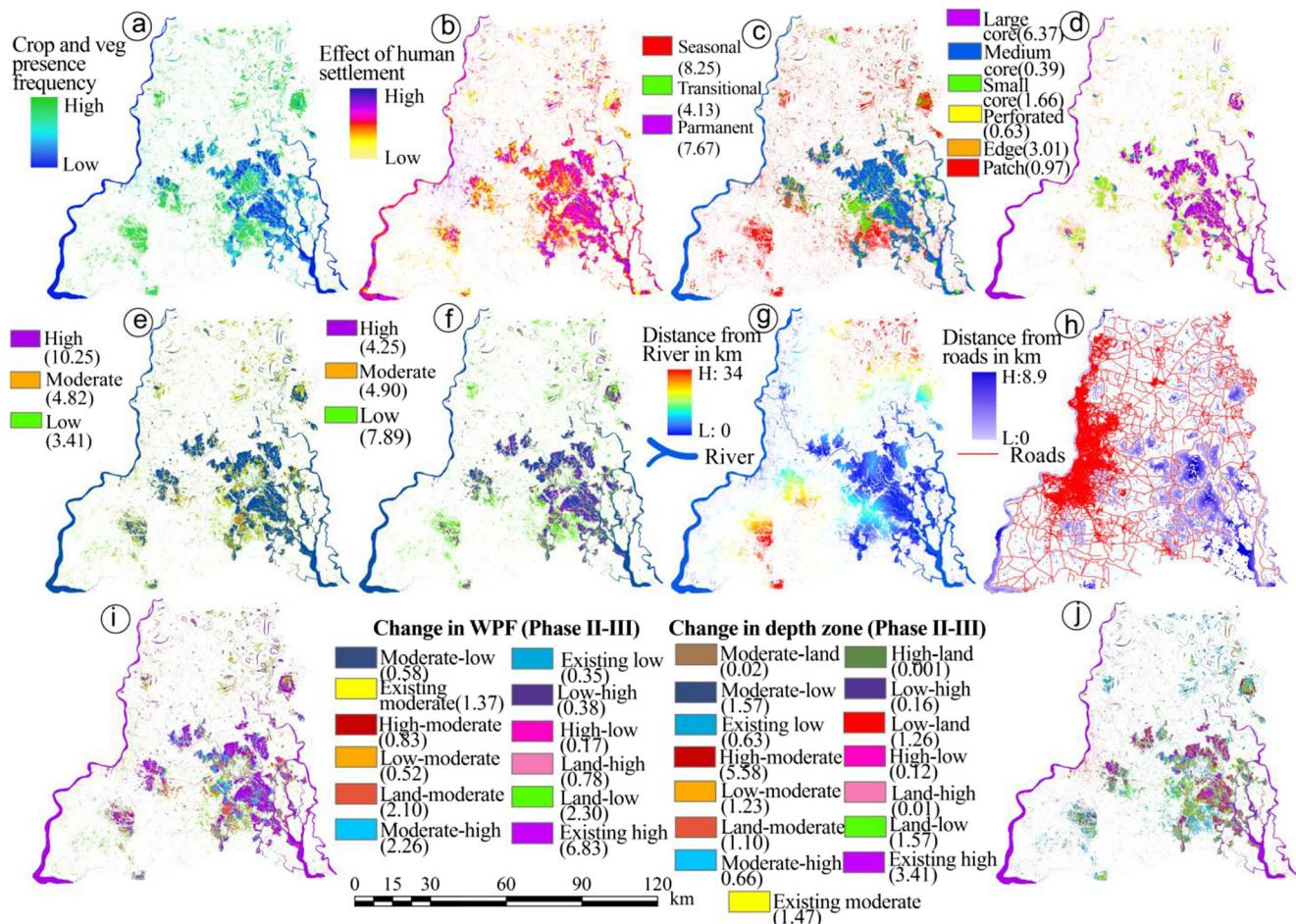


Fig. 3 Data layers of phase III: **a** crop and vegetation presence frequency, **b** effect of human settlement, **c** seasonality, **d** wetland fragmentation, **e** WPF, **f** depth zone, **g** distance from the river, **h** distance from roads, **i**

change in WPF, and **j** change in the depth zone (values within parenthesis show the percentage of areal coverage under different zones)

uses to calculate the positive weight (W^+) and negative weight (W^-) of each factor class then the contrast weight (Cw) of them. There can be four combinations of each factor which is possible to estimate at pixel scale and can be calculated in GIS software (Dube et al. 2014).

$$N_{\text{fac}} = \text{number of pixel in the factor}$$

$$N_{\text{vul}} = \text{vulnerable pixel}$$

$$N_{\text{class}} = \text{number of pixel in the class}$$

$$N_{\text{vulclass}} = \text{number vulnerable of the pixel in the class}$$

Now the four combinations could be-

$$N_{\text{pix1}} = N_{\text{vulclass}}$$

$$N_{\text{pix2}} = N_{\text{vul}} - N_{\text{vulclass}}$$

$$N_{\text{pix3}} = N_{\text{class}} - N_{\text{vulclass}}$$

$$N_{\text{pix4}} = N_{\text{fac}} - N_{\text{vul}} - N_{\text{class}} + N_{\text{vulclass}}$$

$$W^+ = \ln \left[\left\{ N_{\text{pix1}} / (N_{\text{pix1}} + N_{\text{pix2}}) \right\} / \left\{ N_{\text{pix3}} + N_{\text{pix4}} \right\} \right] \quad (16)$$

Weight of evidence

WoE combines some binary patterns to predict binary patterns like habitat vulnerability or non-vulnerability (Nohani et al. 2019). It is based on the Bayesian probabilistic method and some other benefits of this method make it popular than other traditional binary statistical methods (Arabameri et al. 2019a, b). This method first

$$W^- = \ln [\{N_{pix2}/(N_{pix1} + N_{pix2})\}/\{N_{pix4}/(N_{pix3} + N_{pix4})\}] \quad (17)$$

$$C_w = W^+ - W^- \quad (18)$$

Now the weight contrast shows the relationship between the factors determining the vulnerability. Then, the standard deviation of positive and negative weights $S_{(c)}$ is calculated using the variance of positive weights (SW^+) and negative weights (SW^-)

$$S_{(c)} \sqrt{SW^+ + SW^-} \quad (19)$$

Now finally, the studentized contrast which is the ratio of the contrast to its standard deviation to get the final weight (W_f) of any factor class.

$$W_f = (C/S_{(c)}) \quad (20)$$

Machine learning algorithms *Logistic regression*

LR is a preliminary vulnerable area-based data-driven model frequently used in vulnerability study (Wu et al. 2017; Chen et al. 2017a, b; Patriche et al. 2016). It generally proposes a best-fitted model correlating the dichotomous interest (vulnerability/non-vulnerability) with the exposure factor (Rasyid et al. 2016). It allows compositing of both continuous or discrete data layer and a combination of both (Raja et al. 2017). This model requires a randomly ground truth training dataset from both vulnerable and non-vulnerable wetlands. In the LR model, the relationship between wetland habitat vulnerability and its determining factor can be obtained through the following equation (Eq. (21)).

$$P_v = \frac{1}{1 + \exp^{-z}} \quad (21)$$

where P_v is the probability of vulnerability occurrences and z is the linear regression model.

$$z = b_0 + b_1x_1 + b_2x_2 + \dots + b_nx_n \quad (22)$$

where b_0 is the intercept of the model, n is the number of vulnerable conditioning factors, b_n is the weight of each indicator, and X_n is the vulnerability of causative factors.

Reduced error pruning tree

The algorithms of REP tree works as fast machine learning which combines reduced error pruning with decision tree and uses to split and prune steps (Quinlan 1987). The use of the decision tree in this model is to simplify the modeling process with the help of a training dataset. The REP process is used to reduce the complexity of the tree structure, when the decision tree output is large (Mohamed et al. 2012). In this algorithm, the backwardness of the overfitting problem is removed with the pruning process of the REP tree. This algorithm generates

a series of pruned trees directly using the test data and then identify the most accurate sub-tree based on the post-pruning method (Pham et al. 2019a). The performance of this model is based on reduced error pruning techniques and the information is extracted for entropy or minimizing the variance (Srinivasan and Mekala 2014).

Random forest

The RF was introduced by Breiman (2001) which is quite a user-friendly and efficient computational algorithm (Chen et al. 2018b). It is a novel parametric ML algorithm that can manage a variety of data with an unbalanced structure (Herrera et al. 2019). This algorithm combines the performances of numerous decision tree algorithms for classifying and predicting the variable (Rodriguez-Galiano et al. 2012) and raises the accuracy level (Youssef et al. 2016; Chen et al. 2017b; Behnia and Blais-Stevens 2018). After receiving any input data (x) with differential evidence feature, random forest first analyzes it for a given training sets and then produces number K of regression tree and sums up it for averaging the result (Rodriguez-Galiano et al. 2015) after K the trees are grown as $\{T(x)\}_1^K$ and the regression predictor can be expressed as

$$\tilde{f}_{rf}(x) = \frac{1}{K} \sum_{k=1}^K T(x) \quad (23)$$

The setup of the RF model is based on two parameters, called the base of the model. These are number of tree (“n-tree”) and number of features in each split (“m-try”). For generating the most suitable predictive model, it produces a bunch of classification trees based on binary decisions and uses them with a voting majority (Sirbu et al. 2019). Thus, the accurate classification trees are chosen by the majority vote from the trees of the entire forest. The similarity between different trees in this algorithm is avoided by increasing the multiplicity and producing them through different training subsets created through bagging. If the samples are not chosen for the training of the K-th tree, then the bagging process becomes a part of another subset called out of bag (OOB). These OOB elements are used to evaluate the performance of the K-th tree (Peters et al. 2007). By this process, the RF model becomes enable for estimating the generalization error with zero biasness without using any external data subset (Breiman 2001). It provides the model greater stability and makes it the robust model when it faces the variation in input data and also helps to achieve greater prediction accuracy. This model is developed without any pruning and that is why it is lighter from a computational perspective. The error of generalization converges with an increasing number of trees in the forest, which makes it not prone to overfit with the input data and also able to assess the relative importance of evidential features. This aspect is very crucial for any

multi-source study. As the present study consists of very high dimensionality of data, then it is important to include the influences of each feature so that the model can predict habitat vulnerability of the wetland with a higher degree of accuracy.

Artificial neural network

ANN is capable of mathematical modeling and parallel processing and then enables them to mimic the biological system of the human brain like a neural network using interconnected units neuron (Mosavi et al. 2018; Daniel 2013). It is considered versatile and more efficient than other traditional statistical models to comprehend any complex environmental features like WHV with a higher degree of accuracy (Abbot and Marohasy 2014). ANN algorithm uses to observe the relation between the data inputted (Arekhi and Jafarzadeh 2014; Tayyebi and Pijanowski 2014) and runs the data mining processes with transition rule from the underlying layer (Yang et al. 2016). The ANN algorithm is analogous to the human brain like the human brain consisting of units of neurons and nodes. The units of the neural network are treated as layer functions through the unidirectional flow of information originated from the input layer via hidden layer to the final output layer, where the first input layer generates a signal corresponding to an element and then, the output value is generated by each neuron which is used as input value by the next layer. The learning of ANN often takes place following a pattern through training of a known set of input and output dataset for the training algorithms to iteratively generate the link weight. The present study uses a standard multilayer ANN whose j in $l+1$ layer produces $x_j^{l+1} = f\left(\sum_i w_{ij}^l x_i^l + w_{bj}^l\right)$, where w_{ij}^l is the weights connecting to the neuron i in l layer to neuron j in $l+1$ layer, w_{bj}^l indicates the bias term of neuron in neuron j in l layer, and f is the logistic activation function. The ANN prediction for sample x_i is signified as $f(x_i)$. This algorithm tries to find out a set of weights that ensure each input vector to produce the same or at least nearer vector from the network, to the desired output vector. This kind of multilayer ANN model can derive meaning from determining factors and thus considered as a reliable data-driven tool in this study for integrating exposure factors of WHV.

Support vector machine

An SVM is a widely used supervised learning method that performs dichotomous classification and regression of multidimensional vector feature based on statistical learning theory and follows the principle of structural risk minimization (Vapnik 1963; Vapnik and Chervonenkis 1964). It was introduced by Vapnik (2013) and is able to distinguish the original input space into a feature space differentiated by

high dimension, using the training datasets (Kanevski et al. 2009). A decision surface is used to make a distinction between classes and simultaneously, the margin of the surface is maximized and minimizes the error between the classes. The surface is called hyper-plane and the closest data points to that plane are referred to as support vector. Hence, the hyper of a larger margin can resist more classification error than the hyper-plane having narrower margin and produces better generalization (Kanevski et al. 2009). This property makes SVM a unique and advantageous machine learning algorithm and researchers frequently apply it in their study field (Micheletti et al. 2011). While classifying a target variable, the support vectors play a crucial role to set an optimal hyper-plane (Pradhan 2013). Two basic functions of SVM modeling take part in the discriminant type of statistical problem, discriminating the data pattern using optimal hyper-plane and enabling the kernel function for converting original nonlinear dataset to a linearly separable format with high-dimensional feature space (Lee et al. 2017). The ability of this algorithm largely depends on this kernel function. The frequently used kernel functions of the SVM algorithm are polynomial kernel (PL); sigmoid kernel (SIG); radial basis function (RBF); and linear kernel (LN). The detailed procedure of SVM algorithms as explained by Rodriguez-Galiano et al. (2015) can be scripted.

In a given dataset $\{X_n\}_{n=1}^N$ with N samples, where $X \in R^L$ a vector of L is input features and the features corresponding to it are known as output features $\{y_n\}_{n=1}^N$, with $y_n \in \{-1, 1\}$, then the regression of SVM is defined as

$$f(X) = W^T \phi(X) + b \quad (24)$$

where $\phi: x \rightarrow \phi(x) \in R^H$ is any nonlinear function that converts the input data into the high-dimensional feature space with $H \geq L$. Originally, assuming linearly separable features, this function was trivially defined as $\phi(X) = X$. On the other hand, W is the unknown parameter of the model which is a weight vector and is normal to the hyper-plane and its bias b .

Then, the SVM model for regression is defined to allow the misclassification error to cope with non-separable features. Therefore, the above SVM model presented becomes subject to the following constraints:

$$\begin{aligned} y_n - f(X_n) &\leq \xi_n + \epsilon \\ f(X_n) - y_n &\leq \xi_n^* + \epsilon \\ \epsilon, \xi_n, \xi_n^* &\geq 0, \forall n \end{aligned} \quad (25)$$

where ϵ indicates the (in) sensitivity (maximum misclassification error) and $\{\xi_n, \xi_n^*\}_n^N = 1$ are slack variables which quantify the deviation of output features from the positive and negative classes.

The optimization of the previous model is subject to soft margin constrain that defines a hyper-plane which separates

the training data maximizing the margin. The Lagrange multipliers method can be used to solve the optimization problem which is discussed in detail by Vapnik (2013), and here the next cost function is yield:

$$L(\{a_n, a_n^*\}_{n=1}^N) = -\frac{1}{2} \sum_{i,j=1}^N (a_i - a_j^*) (a_i - a_j^*) K(X_i, X_j) - \varepsilon \sum_{i=1}^n (a_i + a_i^*) \sum_{j=1}^n (a_j - a_j^*) y_j \quad (26)$$

where on the one hand, $\{\hat{a}_n, \hat{a}_n^*\}_{n=1}^N$ are the Lagrange multipliers and on the other hand, $K(X_i, X_j)$ is the Kernel function, which is defined as the inner product of the transformed input feature vectors:

$$K(X_i, X_j) := \langle \phi(X_i) | \phi(X_j) \rangle. \quad (27)$$

Now here the kernel notation is used to simplify the optimization process of the cost function. The designing of any mapping function is replaced, and transforming the data and later computing the inner products, SVM directly uses the kernel for functioning the input feature vector. The following are some of the kernel functions typically used in SVM applications:

$$\begin{aligned} K_{linear}(x, x') &= x, x' \\ K_{polynomial} &= (\gamma x x' + r)^p \\ K_{RBF}(x, x') &= \exp(-\gamma \|x - x'\|^2) \\ K_{sigmoid}(x, x') &= \tanh(\gamma x x' + r). \end{aligned} \quad (28)$$

After estimating the $\{\hat{a}_n, \hat{a}_n^*\}_{n=1}^N$ by maximizing the cost function defined above, the margin can be inferred as:

$$\hat{w} = \sum_{n=1}^N (\hat{a}_n - \hat{a}_n^*) \phi(X_n) \quad (29)$$

Such as $f(X)$ can be directly estimated as;

$$\hat{f}(X) = \sum_{n=1}^N (\hat{a}_n - \hat{a}_n^*) K(X_n, X) + \hat{b} \quad (30)$$

where the pre-processing and centralization of data can conveniently drop out the computation of b by forcing the bias to be zero.

In this study, the best-fitted model of LR has been performed in Statistical Package for Social Science (SPSS) software and compositing of data layers has been done in the GIS environment (Debanshi and Pal 2020c; Pal and Talukdar 2018b). Other ML algorithms mentioned above have been performed in WEKA software in order to produce suitable models (Kadavi et al. 2018). The following table (Table 1) shows the optimization of hyperparameters associated with ML algorithms (Pal et al. 2020; Pham et al. 2020).

Accuracy assessment

The prediction and classification of any environmental process or phenomenon using multiple modeling approaches often produce variation between different inter-model classification results. In the present study, it is observed as different models produce different zonation of WHV. Here the statistical significance of this variance is measured using the analysis of variance (ANOVA) and Friedman test. Apart from that, the ground truth reality of any predictive study is also necessary in order to assess the accuracy of the prediction and classification (Rasyid et al. 2016). In the present study, therefore, the receiver operating characteristic (ROC) curves have been constructed on each model of habitat vulnerability. The area under curve (AUC) values are considered to be efficient for determining the ground truth reality of the models (Nohani et al. 2019). In order to carry out the validation of this study, 400 random GCPs were selected. The ROC curves have been constructed considering the ground data and data extracted from the WHV models. Besides ROC, a GCP-based confusion matrix containing four sets of ground-truth reality, i.e., true positive (TP), false positive (FP), true negative (TN), and false negative (FN) can also be used to assess the success rate of the models (Chen et al. 2017c) calculating precision, recall, and F1 score of the models (Dang et al. 2020). The F-measure (precision, recall, and F1 score) is recently considered a critical evaluation for measuring the prediction performance of spatial models especially where ML is involved (Liu et al. 2017). The value of F-measure varies from 0 to 1, where a greater value indicates high prediction and classification performance (Tharwat 2020). Therefore, in this study, the F-measure has been carried out using the following equations (Sokolova et al. 2006).

$$Precision = TP / (TP + FP) \quad (31)$$

$$Recall = TP / (TP + FN) \quad (32)$$

$$F1score = 2 \times (Precision \times Recall) / (Precision + Recall) \quad (33)$$

The models have also been cross-verified by calculating the root-mean-square error (RMSE) using the following equation.

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (X_{obsi} - X_{modeli})^2}{n}} \quad (34)$$

where X_{obsi} and X_{modeli} are the observed and predicted values of model i and n is the number of observations.

Modeling with field parameters In order to explicit the ground truth reality of all the data-driven models, here a field data-

| Machine learning algorithms | Model optimization |
|-----------------------------|---|
| REP tree | Batch size: 100, seed: 2, max depth: 2, minimum number: 2.0, minimum variance proportion: 0.001, spread initial count: TRUE |
| RF | Batch size: 100, seed: 2, number of iteration: 150, max depth: 2, calc out of bag: TRUE, compute attribute importance: TRUE |
| ANN | Hidden layer: 10, learning rate: 0.32, momentum: 0.24, seed: 5, training time: 500, validation threshold: 20, normal to binary filter: TRUE |
| SVM | Complexity parameter: 1, epsilon: 1.0E-12, kernel: Puk, random seed: 4, number of fold: 1, tolerance parameter: 0.001 |

based modeling approach has been adapted. To do so, here 400 GCPs were used to extract field data of several wetland- and habitat-related parameters. These parameters are hydro-duration of wetland in wetlands, average depth of the wetlands, positive hydro-duration anomaly (hydro-duration above-average depth), and frequently found fish species. To prepare the data layers of these parameters, the field extracted data were interpolated for the entire study area using the inverse weighted distance (IDW) tool in the GIS environment. Analytic hierarchy process (AHP) has been used to derive the weight of the selected field parameters and then, the field-based wetland habitat vulnerability model has been generated by linear compositing of weighted parameters.

For testing the similarity of the habitat vulnerability models with the field-based model, the correlation coefficient has been computed. Residuals of standard regression are also computed using geographically weighted regression of the wetland vulnerability model parameters and field-based vulnerability model to judge the performance level of representative parameters on model building. The standard error is then interpolated to see the proportion of wetland area registers very low residual showing the representativeness of the predicted model with observed field model.

Ensemble modeling

The multitude of the ML modeling approach has been a strong background (Dietterich 2000); thus, the single-model prediction has shifted toward the ensemble of models (Mosavi et al. 2018). In this present study, the individual bivariate and ML models which have achieved more than 0.8 AUC values have been ensembled. Among the bivariate models, the calculated values of FR and WoE have been used as training data of SE (Chen et al. 2018a, 2019). Similarly, the ensemble of bivariate statistics with ML algorithms is calculated; the result of bivariate statistics, which provides weights for each class of influencing factors such as ER and WoE, has taken into consideration, and training of ML algorithms has been done in the WEKA environment. The ensemble of ML algorithms has a long tradition of vulnerability prediction (Mosavi et al. 2018). In order to increase the accuracy of the prediction system, the ensemble prediction system (EPS) has been proposed (Sajedi-Hosseini et al. 2018) which generally use multiple ML algorithms and tries to achieve higher result accuracy using automated assessment and weighting system (Mosavi et al. 2018). The present study therefore involves the ensemble of ML algorithms with each other tries to achieve greater accuracy and increase the robustness of prediction. The ensemble of machine learning algorithms which have been done is REP Tree-RF, REP Tree-ANN, REP Tree-SVM, RF-ANN, RF-SVM, and ANN-SVM for both phases. Figure 4 shows the methodological proceedings of this study.

Table 1 Parameter optimization ML algorithms for generating WHV models

Results

Wetland habitat vulnerability state

The spatial organization of wetland habitat vulnerability of both phases derived from all the modeling techniques is shown in Fig. 5. Table 2 shows their areal coverage. More or less 150 km² of the wetland area has been demarcated as very high vulnerable wetland habitat in bivariate statistical models. This figure is recorded more than 350 km² in ML models in phase II. Phase-wise monitoring of the wetland vulnerability zones shows an increase of areal coverage under very high vulnerable zone from phase II to phase III. In the case of bivariate statistics, such increase is by 30–35% and SE shows the maximum increase. ML algorithms also show a similar trend by 10–12% where the maximum increase can be observed in the RF model. Areal shrinkage of relatively safe wetland in phase III as compared to phase II is also common among all the bivariate statistical models where an average of 15% decrease can be observed. On the other hand, such shrinkage is not so evident in cases of ML models except LR and RF. Apart from very high vulnerability, high vulnerability also occupies a considerable area in phase II and which increases in phase III. Maps of wetland habitat vulnerability (Fig. 5) show most of the relatively safe wetlands are concentrated in the largest stretch of wetland situated in the middle of

the study area. Larger parts of the wetland situated outside the largest stretch belong to greater vulnerability. Some of the patches of these scatter wetlands and ox-bow lakes are seen in lesser vulnerability situation in bivariate models but most of these patches are counted as high and very high vulnerable wetland habitat in ML models.

The ensemble models do not show much variability in the spatial organization of different vulnerable zones (Figs. 6, 7, and 8). Among the ensembled bivariate statistical models, the ensemble of FR-SE categorizes greater wetland area under higher vulnerability as compared to individual FR and SE model. Similarly, when SE is ensembled with WoE, it reduces the areal coverage of very high vulnerable wetland than those two individual models but shrinkage of safer wetlands over the phases is observed in this ensemble model. The ensembles of ML algorithms and bivariate statistics considerably increase the area coverage of very high and high vulnerable wetlands than the individual bivariate models and at the same time, it reduces the area under safer wetland conditions. For example, the FR model demarcates 270 km² as very high vulnerable wetland and it increases to almost 400 km² when it is an ensemble with ML algorithms, whereas more than 250 km² of relatively safe wetland reduces to about 180 km². The ensembles among the ML model do not show much variation in the areal coverage and their spatial organization in both phases from the individual models of ML algorithms.

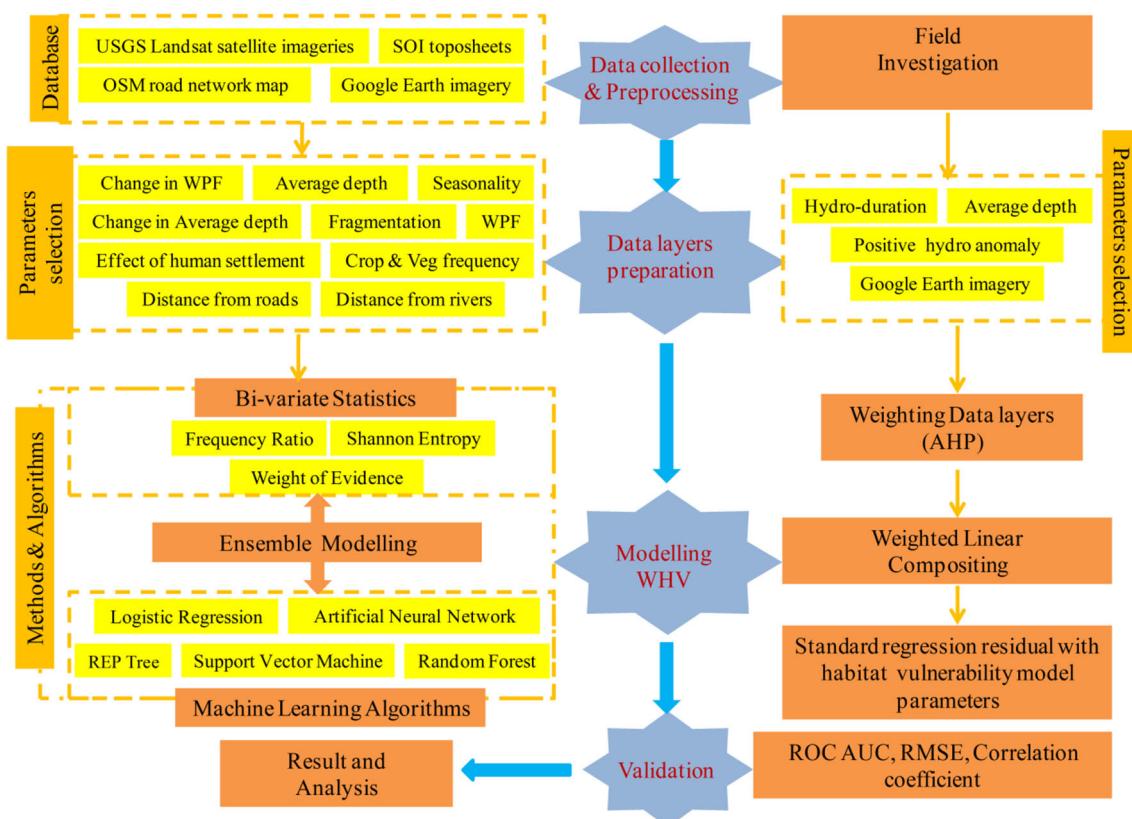


Fig. 4 Methodological proceedings of the study

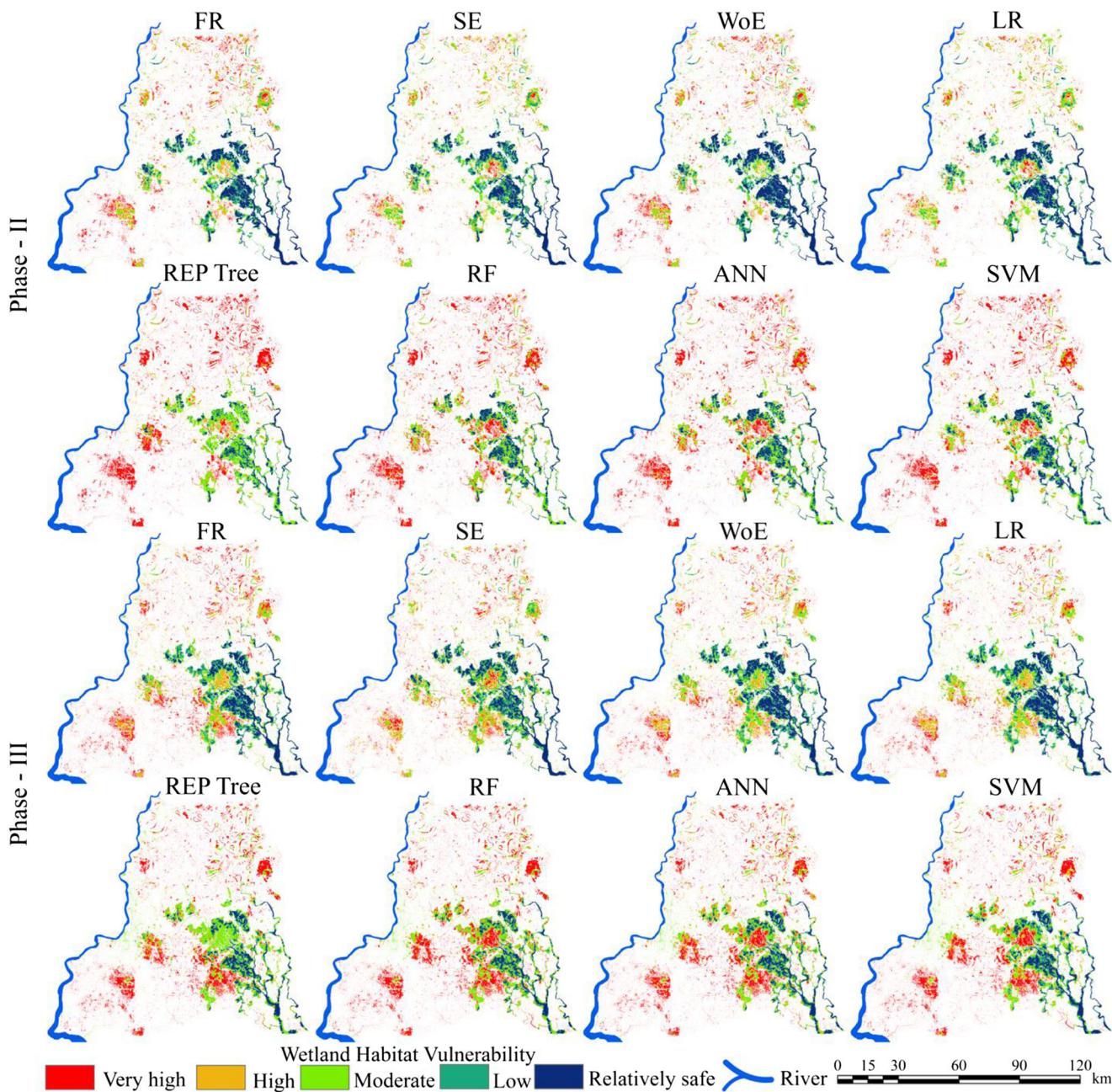


Fig. 5 Wetland habitat vulnerability zones derived from bivariate statistics and ML algorithms

Accuracies of the models

Table 3 shows the result of ANOVA and Friedman tests, and the P values of both the tests indicate statistical insignificance of the variation in vulnerability zonation of the models in both the phases. However, to justify the best suitable among the applied models, the accuracies of the models regarding ground-truth reality are tabulated in Table 4. Among the bivariate models except for FR in phase II, all the models have achieved fair accuracy in both the phases. The overall accuracy of the ML models is higher than in some of the bivariate models. The F -

measures show in most cases more than 80% and in some cases like in ML algorithms and ensemble of ML more than 90% of accuracy which indicates high prediction and classification performances of the models. On the other hand, some of the bivariate statistics have claimed almost the same accuracy of the ML algorithms. This does indicate the adoption of any ML model is not always better than the bivariate model. The ensemble of individual bivariate models or ML algorithms with the bivariate statistics fails to increase the accuracy for predicting vulnerability of wetland habitat, but the accuracy of some ML algorithms have fairly increased after ensemble with another ML algorithm, like the ensemble of REP tree and RF with ANN and

Table 2 Areal coverage of different wetland habitat vulnerable zones

| Method | Model | Zones | Phase II | | Phase III | |
|--|----------------------|-------------------------|-------------------------|------|-------------------------|------|
| | | | Area in km ² | % | Area in km ² | % |
| Bivariate statistics | FR | Very high vulnerability | 195.11 | 3.07 | 273.44 | 4.3 |
| | | High vulnerability | 182.29 | 2.87 | 241.73 | 3.8 |
| | | Moderate vulnerability | 164.63 | 2.59 | 193.83 | 3.05 |
| | | Low vulnerability | 191.74 | 3.02 | 225.52 | 3.55 |
| | | Relatively safe | 294.77 | 4.64 | 265.49 | 4.18 |
| | SE | Very high vulnerability | 119.59 | 1.88 | 207.42 | 3.26 |
| | | High vulnerability | 197.5 | 3.11 | 281.49 | 4.43 |
| | | Moderate vulnerability | 229.46 | 3.61 | 227.39 | 3.58 |
| | | Low vulnerability | 230.88 | 3.63 | 235.44 | 3.7 |
| | | Relatively safe | 268.75 | 4.23 | 248.27 | 3.9 |
| Machine learning algorithms | WoE | Very high vulnerability | 174.17 | 2.74 | 191.53 | 3.01 |
| | | High vulnerability | 161.04 | 2.53 | 253.83 | 3.99 |
| | | Moderate vulnerability | 157.59 | 2.48 | 207.83 | 3.27 |
| | | Low vulnerability | 183.96 | 2.89 | 235.89 | 3.71 |
| | | Relatively safe | 349.5 | 5.5 | 301.73 | 4.75 |
| | LR | Very high vulnerability | 110.18 | 1.73 | 193.61 | 3.04 |
| | | High vulnerability | 211.63 | 3.33 | 274.07 | 4.31 |
| | | Moderate vulnerability | 206.22 | 3.24 | 207.99 | 3.27 |
| | | Low vulnerability | 226.84 | 3.57 | 242.78 | 3.82 |
| | | Relatively safe | 291.3 | 4.58 | 281.57 | 4.43 |
| Machine learning algorithms | REP tree | Very high vulnerability | 423.19 | 6.66 | 400.17 | 6.29 |
| | | High vulnerability | 97.01 | 1.53 | 156.31 | 2.46 |
| | | Moderate vulnerability | 307.56 | 4.84 | 371.14 | 5.84 |
| | | Low vulnerability | 112.6 | 1.77 | 118.58 | 1.87 |
| | | Relatively safe | 105.81 | 1.66 | 152.94 | 2.41 |
| | RF | Very high vulnerability | 402.38 | 6.33 | 472.59 | 7.43 |
| | | High vulnerability | 139.13 | 2.19 | 175.43 | 2.76 |
| | | Moderate vulnerability | 242.63 | 3.82 | 305.34 | 4.8 |
| | | Low vulnerability | 119.67 | 1.88 | 108.4 | 1.7 |
| | | Relatively safe | 142.36 | 2.24 | 137.4 | 2.16 |
| Machine learning algorithms | ANN | Very high vulnerability | 396.59 | 6.24 | 432.45 | 6.8 |
| | | High vulnerability | 152.02 | 2.39 | 210.89 | 3.32 |
| | | Moderate vulnerability | 217.33 | 3.42 | 267.69 | 4.21 |
| | | Low vulnerability | 132.52 | 2.08 | 133.88 | 2.11 |
| | | Relatively safe | 147.71 | 2.32 | 154.24 | 2.43 |
| | SVM | Very high vulnerability | 366.32 | 5.76 | 425.8 | 6.7 |
| | | High vulnerability | 151.95 | 2.39 | 198.58 | 3.12 |
| | | Moderate vulnerability | 222.94 | 3.51 | 250.82 | 3.94 |
| | | Low vulnerability | 134.88 | 2.12 | 149.75 | 2.36 |
| | | Relatively safe | 170.09 | 2.68 | 174.19 | 2.74 |
| Ensemble of bivariate statistics | Ensemble FR-SE | Very high vulnerability | - | - | 281.56 | 4.43 |
| | | High vulnerability | - | - | 247.48 | 3.89 |
| | | Moderate vulnerability | - | - | 204.86 | 3.22 |
| | | Low vulnerability | - | - | 242.24 | 3.81 |
| | | Relatively safe | - | - | 223.88 | 3.52 |
| | Ensemble WoE-SE | Very high vulnerability | 200.24 | 3.15 | 187.13 | 2.94 |
| | | High vulnerability | 140.33 | 2.21 | 266.33 | 4.19 |
| | | Moderate vulnerability | 150.53 | 2.37 | 257.05 | 4.04 |
| | | Low vulnerability | 173.56 | 2.73 | 296.48 | 4.66 |
| | | Relatively safe | 331.42 | 5.21 | 184.86 | 2.91 |
| Ensemble of bivariate statistics and machine learning algorithms | Ensemble FR-LR | Very high vulnerability | - | - | 228.04 | 3.59 |
| | | High vulnerability | - | - | 297.28 | 4.68 |
| | | Moderate vulnerability | - | - | 182.12 | 2.86 |
| | | Low vulnerability | - | - | 313.84 | 4.94 |
| | | Relatively safe | - | - | 178.72 | 2.81 |
| | Ensemble WoE-LR | Very high vulnerability | - | - | 182.76 | 2.87 |
| | | High vulnerability | - | - | 121.38 | 1.91 |
| | | Moderate vulnerability | - | - | 390.54 | 6.14 |
| | | Low vulnerability | - | - | 239.5 | 3.77 |
| | | Relatively safe | - | - | 265.82 | 4.18 |
| Ensemble of bivariate statistics and machine learning algorithms | Ensemble FR-REP tree | Very high vulnerability | - | - | 428.98 | 6.75 |
| | | High vulnerability | - | - | 189.63 | 2.98 |
| | | Moderate vulnerability | - | - | 262.52 | 4.13 |

Table 2 (continued)

| Method | Model | Zones | Phase II | | Phase III | |
|---|----------------------|-------------------------|-------------------------|------|-------------------------|------|
| | | | Area in km ² | % | Area in km ² | % |
| Ensemble FR-RF | | Low vulnerability | - | - | 144.16 | 2.27 |
| | | Relatively safe | - | - | 173.19 | 2.72 |
| | | Very high vulnerability | - | - | 404 | 6.35 |
| | | High vulnerability | - | - | 237.35 | 3.73 |
| | | Moderate vulnerability | - | - | 244.66 | 3.85 |
| | | Low vulnerability | - | - | 139.21 | 2.19 |
| Ensemble FR-ANN | | Relatively safe | - | - | 173.53 | 2.73 |
| | | Very high vulnerability | - | - | 381.89 | 6.01 |
| | | High vulnerability | - | - | 216.24 | 3.4 |
| | | Moderate vulnerability | - | - | 250.53 | 3.94 |
| | | Low vulnerability | - | - | 159.79 | 2.51 |
| | | Relatively safe | - | - | 184.32 | 2.9 |
| Ensemble FR-SVM | | Very high vulnerability | - | - | 425.58 | 6.69 |
| | | High vulnerability | - | - | 185.8 | 2.92 |
| | | Moderate vulnerability | - | - | 262.51 | 4.13 |
| | | Low vulnerability | - | - | 145.72 | 2.29 |
| | | Relatively safe | - | - | 143.43 | 2.26 |
| | | Very high vulnerability | 281.56 | 4.43 | 434.12 | 6.83 |
| Ensemble WoE-REP tree | | High vulnerability | 247.48 | 3.89 | 221.69 | 3.49 |
| | | Moderate vulnerability | 204.86 | 3.22 | 264.83 | 4.17 |
| | | Low vulnerability | 242.24 | 3.81 | 120.15 | 1.89 |
| | | Relatively safe | 223.88 | 3.52 | 158.79 | 2.5 |
| | | Very high vulnerability | 366.87 | 5.77 | 453.81 | 7.14 |
| | | High vulnerability | 145.39 | 2.29 | 193.62 | 3.05 |
| Ensemble WoE-RF | | Moderate vulnerability | 232.69 | 3.66 | 272.48 | 4.29 |
| | | Low vulnerability | 135.76 | 2.14 | 121.61 | 1.91 |
| | | Relatively safe | 163.81 | 2.58 | 158.26 | 2.49 |
| | | Very high vulnerability | 349.38 | 5.49 | 453.81 | 7.14 |
| | | High vulnerability | 146.4 | 2.3 | 193.62 | 3.05 |
| | | Moderate vulnerability | 176.68 | 2.78 | 272.48 | 4.29 |
| Ensemble WoE-ANN | | Low vulnerability | 138.67 | 2.18 | 121.61 | 1.91 |
| | | Relatively safe | 228.86 | 3.6 | 158.26 | 2.49 |
| | | Very high vulnerability | 362.58 | 5.7 | 453.81 | 7.14 |
| | | High vulnerability | 147.17 | 2.31 | 193.62 | 3.05 |
| | | Moderate vulnerability | 219.27 | 3.45 | 272.48 | 4.29 |
| | | Low vulnerability | 140.65 | 2.21 | 121.61 | 1.91 |
| Ensemble WoE-SVM | | Relatively safe | 176.22 | 2.77 | 158.26 | 2.49 |
| | | Very high vulnerability | 141.2 | 2.22 | 231.46 | 3.64 |
| | | High vulnerability | 97.37 | 1.53 | 153.26 | 2.41 |
| | | Moderate vulnerability | 308.96 | 4.86 | 371.09 | 5.84 |
| | | Low vulnerability | 110.77 | 1.74 | 117.34 | 1.85 |
| | | Relatively safe | 104.23 | 1.64 | 151.43 | 2.38 |
| Ensemble of machine learning algorithms | Ensemble REP tree-RF | Very high vulnerability | 428.51 | 6.74 | 304.89 | 4.8 |
| | | High vulnerability | 112.85 | 1.77 | 179 | 2.82 |
| | | Moderate vulnerability | 279.66 | 4.4 | 359.56 | 5.66 |
| | | Low vulnerability | 121.36 | 1.91 | 144.25 | 2.27 |
| | | Relatively safe | 103.79 | 1.63 | 147.46 | 2.32 |
| | | Very high vulnerability | 400.32 | 6.3 | 426.35 | 6.71 |
| Ensemble REP tree-ANN | | High vulnerability | 151.14 | 2.38 | 204.25 | 3.21 |
| | | Moderate vulnerability | 218.24 | 3.43 | 251.09 | 3.95 |
| | | Low vulnerability | 131.13 | 2.06 | 145.04 | 2.28 |
| | | Relatively safe | 145.33 | 2.29 | 173.28 | 2.73 |
| | | Very high vulnerability | 367.52 | 5.78 | 442.16 | 6.95 |
| | | High vulnerability | 152.23 | 2.39 | 207.3 | 3.26 |
| Ensemble REP tree-SVM | | Moderate vulnerability | 222.74 | 3.5 | 265.3 | 4.17 |
| | | Low vulnerability | 136.38 | 2.14 | 132.75 | 2.09 |
| | | Relatively safe | 167.3 | 2.63 | 152.13 | 2.39 |
| | | Very high vulnerability | 370.42 | 5.83 | 429.88 | 6.76 |
| | | High vulnerability | 150.95 | 2.37 | 201.75 | 3.17 |
| | | Moderate vulnerability | 221.75 | 3.49 | 248.53 | 3.91 |
| Ensemble RF-ANN | | Low vulnerability | 135.98 | 2.14 | 237.18 | 3.73 |
| | | Relatively safe | 167.07 | 2.63 | 172.67 | 2.72 |
| | | Very high vulnerability | 371.14 | 5.84 | 141.31 | 2.22 |
| | | High vulnerability | - | - | - | - |
| | | Moderate vulnerability | - | - | - | - |
| | | Low vulnerability | - | - | - | - |
| Ensemble RF-SVM | | Relatively safe | - | - | - | - |
| | | Very high vulnerability | - | - | - | - |
| | | High vulnerability | - | - | - | - |
| | | Moderate vulnerability | - | - | - | - |
| | | Low vulnerability | - | - | - | - |
| | | Relatively safe | - | - | - | - |
| Ensemble ANN-SVM | | Very high vulnerability | - | - | - | - |

Table 2 (continued)

| Method | Model | Zones | Phase II | | Phase III | |
|--------|-------|------------------------|-------------------------|------|-------------------------|------|
| | | | Area in km ² | % | Area in km ² | % |
| | | High vulnerability | 162.14 | 2.55 | 290.11 | 4.56 |
| | | Moderate vulnerability | 218.36 | 3.43 | 252.65 | 3.97 |
| | | Low vulnerability | 133.62 | 2.1 | 262.99 | 4.14 |
| | | Relatively safe | 160.91 | 2.53 | 252.94 | 3.98 |

SVM. RMSE values also support the accuracy of the models. Figure 9 shows the wetland habitat vulnerability model generated using field-driven data. It shows a considerable similarity of vulnerability zones with predictive models (Fig. 9e). The correlation values between the field-based vulnerability model and predictive models are greater than 0.7 which is significant at a 99% confidence level. However, this relation is stronger in the case of ML and ensemble prediction models. Geographically weighted regression analysis among the predictive parameters and field-based model illustrated that about 70% wetland area is predicted very well lying within very lesser range of residuals of standard regression (± 0.5). It also statistically proves the justification of the selection of parameters for the predictive model.

Discussion

This study successfully explored the methodological possibilities for modeling WHV using multiple modeling approaches. The results derived from both bivariate statistics and ML algorithms were compared in terms of their ground truth reality of prediction. Thereafter, the ensemble of the models was done in order to increase accuracy. It was seen that among the individual bivariate models, SE and WoE; four ML

models namely REP tree, RF, ANN, and SVM; ensemble of WoE and SE; and ensembles among the ML algorithms gave highly accurate prediction over other models which are highlighted in Table 4. Few other individual models used in this study like FR and LR did not give highly accurate result especially in phase II. These models were reportedly given better result in the studies of Pal and Talukdar (2018b); Saha and Pal (2019); and Kim et al. (2019), whereas FR was reported to have poor accuracy by Yalcin et al. (2011). Both the mentioned models were seen to be increasing the accuracy in phase III which indicates their inability to manage historical data with the spatial entity. Some bivariate statistical techniques like SE or WoE have also been proved to be able to provide better results with almost equal accuracy as achieved by relatively advanced ML methods in both phases. This result supports the decision of applying bivariate statistics for modeling habitat vulnerability as some study shows the successful use of these models to vulnerability study in their study (Gayen and Saha 2017; Yang et al. 2018; Nohani et al. 2019; Sarkar and Mondal 2020). Pradan et al. (2012) concluded with the inability of bivariate models to measure the importance of each factors and mentioned these drawbacks may arise in other predictive findings in different cases. In predictive vulnerability study, ML algorithms have proved their superiority over other traditional methods in several types of research in

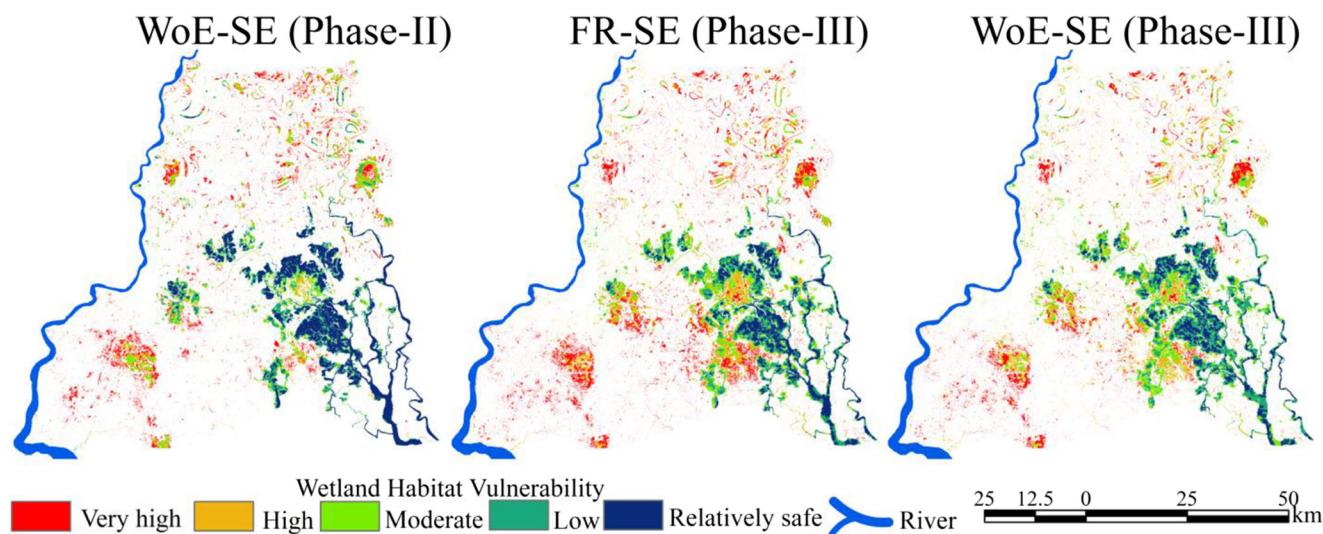


Fig. 6 Wetland habitat vulnerability derived from the ensemble model of bivariate statistics

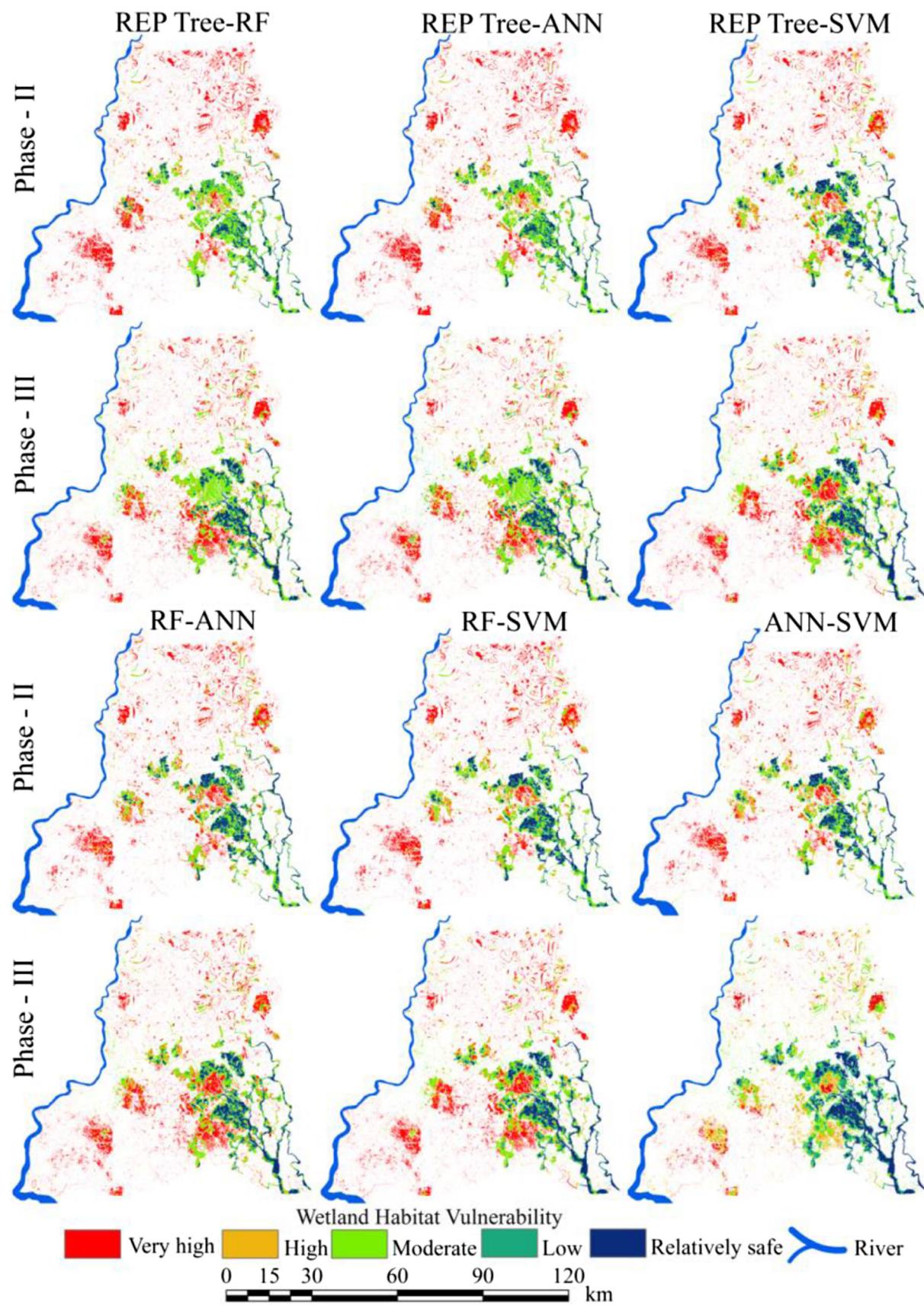


Fig. 7 Wetland habitat vulnerability derived from the ensemble model of ML algorithms

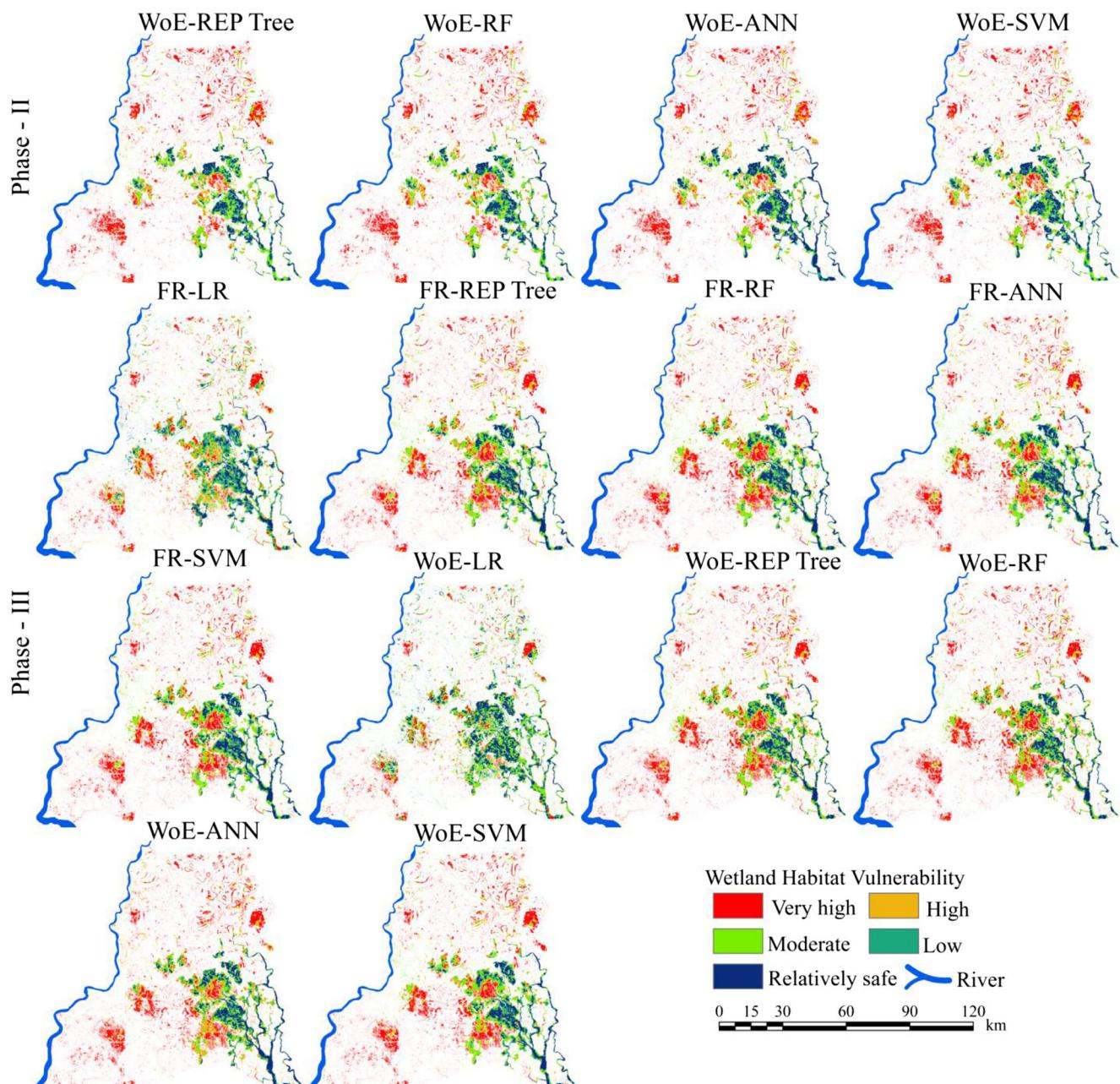


Fig. 8 Wetland habitat vulnerability derived from ensemble model of bivariate statistics and ML algorithms

recent times. This study found the ML technique for measuring WHV to have greater accuracy than bivariate models as the recent studies (Han et al. 2019; Vakhshoori et al. 2019; Fan et al. 2019; Dodangeh et al. 2019) have shown in the field of vulnerability and prediction modeling. In nonlinear modeling, big data handling capability makes this technique superior over other data-driven models. The ensembles of individual models were also done to investigate the capability to increase the prediction accuracy. Several studies recommended better accuracy of ensemble modeling over individual models (Chen et al. 2018a, b, c; Han et al. 2019), but here the individual bivariate models having poor prediction accuracy have failed to achieve

greater accuracy after ensembling with other bivariate or ML models. The accuracy of some individual models like SE and WoE was fairly high but was degraded a lot after ensembling with ML algorithms. In few cases of individual ML algorithms like REP tree and RF, better accuracy was gained after ensembling with ANN and SVM. This indicates the ability of ANN and SVM to generate predictive result with greater accuracy. It should be mentioned here that the areal coverage of different WHV zones varies through their spatial arrangements which are quite similar to each other. The ANOVA or Friedman test shows that variations between the WHV zones of different models are not statistically significant.

Table 3 Measure of variance between the areal coverage of WHV zones derived from various models

| Variance sources | Phase II | | | | Phase III | | | |
|-------------------------------------|----------|-------|------------------------|-------|-----------|-------|------------------------|-------|
| | ANOVA | | Friedman test P values | | ANOVA | | Friedman test P values | |
| | F | P | F crit | | F | P | F crit | |
| Individual bivariate models | 0.006 | 0.994 | 3.885 | 0.549 | 0.005 | 0.995 | 3.885 | 0.549 |
| Individual ML models | 3.3E-10 | 1 | 2.866 | 0.938 | 2.2E-06 | 1 | 2.866 | 0.78 |
| Individual bivariate and ML models | 0.002 | 1 | 2.313 | 0.939 | 0.0002 | 1 | 2.313 | 0.939 |
| Ensemble among bivariate models | - | - | - | - | 0.004 | 0.951 | 5.318 | 0.948 |
| Ensemble of bivariate and ML models | 0.190 | 0.902 | 3.239 | 0.782 | 0.002 | 1 | 2.124 | 0.931 |
| Ensemble among ML models | 0.237 | 0.942 | 2.621 | 0.43 | 0.154 | 0.977 | 2.621 | 0.537 |
| All models | 0.128 | 1 | 1.741 | 0.961 | 0.034 | 1 | 1.612 | 0.998 |

Apart from the methodological exploration, the results articulate serious concern about the habitat vulnerability of the wetland in the present study area. It is seen that more than 60% of the total wetland habitat is suffering from moderate to a critical vulnerability. Milton et al. (2018) articulated the stress on the wetland on a global scale and marked a general conclusion of this phenomenon to be attributed majorly to the climate change and human effect. Growing population pressure and exploitation in the present study area are also seen to be playing major role behind the vulnerability of the wetlands. Researchers like Liu et al. (2015) and Fang et al. (2019) found a similar trend of wetland under stress in their study area and concluded the damning effects of it. The influencing factors behind it can be predicted in the study area as, the lowering of WPF and decreasing depth of the wetlands. These causes are not so unique in the present study area, as Pal and Saha (2018) reported the attenuation of 50% or more area of the wetland due to a reduction in WPF in the lower Atreyee basin of Indo-Bangladesh, environmentally which does not differ a lot from the present study area. The present study show 50% of the total wetland belongs to moderate to low WPF and depth which is in increasing way. The decreasing depth of the wetland with curtails in WPF makes the wetland more vulnerable and weakens the regulatory process to compose a healthy habitat. Paul and Pal (2019) found such kind lowering of wetland water and identified overextraction of wetland water as well as the groundwater responsible for this in very adjacent part of the present study area at the moribund deltaic region. Besides the lowering of water stagnation in the wetlands, over the years, the depth of the wetlands fluctuates with very irregular pattern of rainfall. This factor along with the human-induced reduction of wetland area and depth to pose an immense threat to the aquatic organisms and also significantly limits reproduction possibilities. Apart from that, the climate-induced long-term summer season makes the wetlands to be appeared as non-perennial for several months of a year. The hydrological

effect of this seasonality does not allow the wetlands to hold a rich aquatic biodiversity during the lean season. As the wetland phenomena in this region is largely dominated by the bheri culture (the embankment for fish cultivation is locally called “bheri”) which often fed by tidal saltwater; therefore, the freshwater wetland in this region consists of shallow lakes, ponds, and other temporary water bodies, which more easily dries up due to the drought insist of climate. Hence, the smaller temporary wetlands are converted into paddy field for some economic gain. Phethi and Gumbo (2019) emphasized this subsistence type of agriculture for degrading the wetland in poor economics like South Africa. As a developing region of South Asia, the present area study exhibits quite similar incidents of wetland deterioration. Physical wetland loss is another dimension of wetland vulnerability. The smaller and scatter wetlands face conversion-related issues very often. Davidson (2018) studied the trend of wetland loss and reported the most decaying trend of wetlands in the twentieth century and made an anxiety of continuity of it in the twenty-first century also. In the present study area, the historically long event of urbanization of Kolkata and its surrounding townships is one of the most prominent causes of occupying the wetlands. In addition, the present urbanization status of this region according to the last Census (2011) stands with the further potentiality to harm the wetland habitat. Such kind of wetland loss puts huge pressure on the species dependent on the wetlands and pushes them to the doorsteps of destruction. A study by Grzybowski and Glińska-Lewczuk (2019) also claimed such pressure which is several times higher as compared to the other species dependent on other kinds of habitat. Some other drivers of wetland vulnerability have been traced here like the development of infrastructure, overexploitation of water for meeting the urban water demand and regulating sewage material, eutrophication and pollution, and introduction of exotic fishes for commercial fish cultivation. Galatowitsch (2016) said that these causes are quite common to degrade wetland habitat. Infrastructure

Table 4 Ground truth accuracies of the models

| Method | Model | Phase II | | | Phase III | | |
|--|-----------------------|----------|------|-------------------------|-----------|--------|----------|
| | | ROC AUC | RMSE | Correlation coefficient | Precision | Recall | F1 score |
| Bivariate statistics | FR | 0.69 | 0.42 | 0.61 | 0.707 | 0.714 | 0.80 |
| | SE | 0.86 | 0.27 | 0.76 | 0.886 | 0.894 | 0.88 |
| Machine learning algorithms | WoE | 0.89 | 0.25 | 0.79 | 0.916 | 0.932 | 0.89 |
| | LR | 0.80 | 0.33 | 0.71 | 0.822 | 0.837 | 0.829 |
| | RFP Tree | 0.86 | 0.27 | 0.76 | 0.887 | 0.903 | 0.895 |
| | RF | 0.85 | 0.28 | 0.75 | 0.875 | 0.890 | 0.882 |
| | ANN | 0.85 | 0.28 | 0.75 | 0.877 | 0.892 | 0.884 |
| | SVM | 0.86 | 0.28 | 0.76 | 0.883 | 0.899 | 0.891 |
| Ensembling of bivariate statistics | Ensemble FR-SE | - | - | - | - | - | - |
| | Ensemble WoE-SE | 0.88 | 0.25 | 0.78 | 0.907 | 0.923 | 0.915 |
| Ensembling of bivariate statistics and machine learning algorithms | Ensemble FR-LR | - | - | - | - | - | - |
| | Ensemble WoE-LR | - | - | - | - | - | - |
| | Ensemble FR-RP tree | - | - | - | - | - | - |
| | Ensemble FR-RF | - | - | - | - | - | - |
| | Ensemble FR-ANN | - | - | - | - | - | - |
| | Ensemble FR-SVM | - | - | - | - | - | - |
| Ensembling of machine learning algorithms | Ensemble WoE-RP tree | 0.66 | 0.34 | 0.58 | 0.676 | 0.688 | 0.682 |
| | Ensemble WoE-ERF | 0.66 | 0.34 | 0.58 | 0.676 | 0.688 | 0.682 |
| | Ensemble WoE-ANN | 0.65 | 0.34 | 0.58 | 0.672 | 0.684 | 0.678 |
| | Ensemble WoE-SVM | 0.67 | 0.34 | 0.59 | 0.687 | 0.699 | 0.693 |
| | Ensemble REP tree-RF | 0.86 | 0.26 | 0.76 | 0.887 | 0.903 | 0.895 |
| | Ensemble REP tree-ANN | 0.86 | 0.26 | 0.76 | 0.891 | 0.907 | 0.899 |
| | Ensemble REP tree-SVM | 0.85 | 0.26 | 0.75 | 0.877 | 0.892 | 0.884 |
| | Ensemble RF-ANN | 0.86 | 0.26 | 0.76 | 0.885 | 0.901 | 0.893 |
| | Ensemble RF-SVM | 0.86 | 0.26 | 0.76 | 0.883 | 0.899 | 0.891 |
| | Ensemble ANN-SVM | 0.85 | 0.26 | 0.76 | 0.882 | 0.898 | 0.890 |

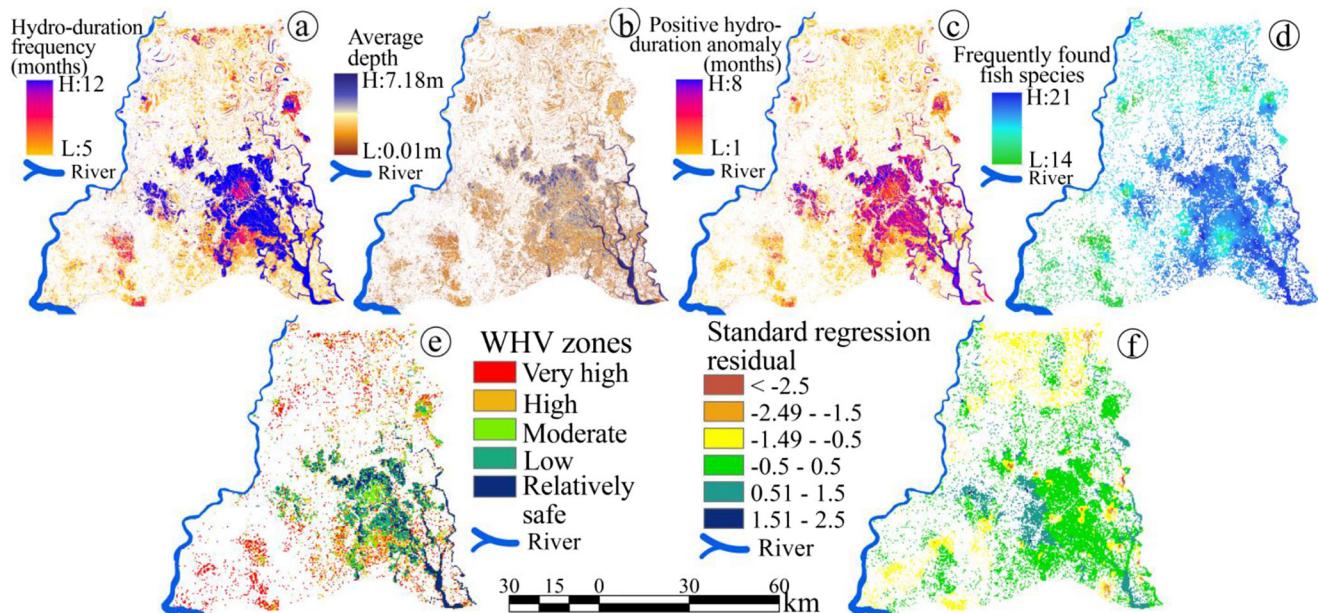


Fig. 9 Field-based data layers WHV model and its residuals of standard regression with the representative data layers of WHV models: **a** hydro-duration (month), **b** average depth, **c** positive hydro-duration anomaly (number of months having water stagnation above-average depth), **d**

number of frequently found fish species, **e** field-based wetland habitat vulnerable zone, **f** residual of standard regression between predicted parameters and field-based wetland vulnerable model

development like construction of roads and railways and concrete structure beside wetland is a driver of wetland squeeze and fragmentation of larger wetland into smaller patches and makes them more easily dryness prone during the summer season. This fragmentation of continuous wetland stretch also affects the breeding of aquatic species. Due to the fragmentation, not only the larger wetland areas are converted into patches but construction projects also cut or infill the tie channels of the ox-bow lake like smaller wetlands with the main fiddling channels which is considered to be essential for composing a sound and salubrious wetland. Mixing of urban sewage and dumping of toxic and polluted garbage into the wetland from urban settlements stop ecological functions and push the wetlands into an unproductive state. For being a leading urban agglomeration of the country since the colonial period, this problem has been very prominent for degrading wetland like other populated parts of the world as reported abroad researchers like Xu et al. (2019) and Xu and Chen (2019). A large portion of the present wetland phenomenon is occupied by the embankment for fish cultivation locally called “bheri” where the introduction of exotic fish species for commercial purposes creates a stressful situation for the indigenous species. These fish species are of main characteristics for the outnumbering of reproduction and hybrid growth. Research like Alam (2014) and Bansal et al. (2019) found a significant role of such an introduction of exotic species on putting stress on the growth of indigenous species. The modeling of WHV shows very little wetland area is having a well condition of wetland habitat, most of which is confined within the largest stretch of wetland in the middle part of the

study area, while most of the smaller wetlands and ox-bow lakes situated outside this largest stretch are suffering from severe habitat vulnerability. The contributions of some healthy habitat developing factors like lesser disturbance due to countryside location, lesser density of transport network, and supply of the freshwater and nutrients from some adjacent perennial river like Ichhamoti, Kalindi, Dansa, and Matla are also seen in some parts of the region. Although the local pisciculture fragments the largest stretch of wetland but also provide well maintenance for sustaining as good habitat in terms of water supply and keeping free from pollutants. In the last phase, a better habitat condition can be seen in some parts of the wetland on the outskirts of the largest stretch. Xue et al. (2018) marked such incident as unusual in wetland vulnerability trend but it is also attributable to the local bheri culture, as it expands outward in search of new ponds and artificial excavation takes place and the newly constructed bheries are taken well care of.

Field-based modeling of WHV using the data directly extracted from the field shows similarity with WHV models. The Pearson correlation coefficient between the field-based model and predictive WHV shows a statistically significant relation stating the validity of the predictive models. About 70% of wetland habitats represent very lesser range (± 0.5) of residuals of standard regression between representative data layers and field-based WHV models. This fact indirectly supports the rationale for parameter selection. This study is well validated with statistical and field data hence could be treated as supporting literature for encouraging predictive parameter-based modeling where ground data is difficult to acquire or time-consuming, but

the result is required very fast. This study would have been more accurate if the information of several factors were spatially available, like river discharge, rainfall map of evenly distributed weather stations, water quality of the wetlands, and most importantly, the changing species composition of the flora and fauna community of the wetlands.

Conclusion

Wetland habitat vulnerability is successfully measured in this study using multiple modeling approaches. The study reveals the status of vulnerability is in a critical state in some patches of wetland. From a management perspective, this would be crucial to provide safeguard to the vulnerable wetlands as the study area holds a dense population and other economic activities. Small wetlands and ox-bow lakes mostly belong to such vulnerability status. Besides the larger wetland stretch, the contribution of these scatter and smaller wetlands to ecological sustainability is not of lesser importance. Area shrinkage, curtail in hydro-duration, and depth reduction are some serious concern provoking issues which are seen frequently in the deltaic landscape. Outcomes of this study can be used for devising a viable preventive and management strategy. Furthermore, this study tries to build up a methodological knowledge addition which will be helpful for future vulnerability study. Some strategies have already been adopted like sewage treatment of some municipalities and the production of fishes from the sewage pond. Such measures may give economic turnover but not sufficient for managing vulnerability from the toxicity of garbage and sewage. A major part of the largest wetland stretch shows the relatively better condition in terms of vulnerability, which is due to local fishing activity. Such an incident proves the viability of wetland management which can open up livelihood options to the local people. Linking wetland usages with an ecologically balanced economy are the proposed option to minimize wetland reclamation for non-promising other activities and help people to rethink wetland is not the wasteland. Immense population pressure and wetland conversion cannot be restricted only stating the priceless serviceability of this unique environmental component.

Acknowledgments For this study, the authors would like to convey their gratitude to USGS for providing Landsat imageries. They are also thankful to Swapan Talukdar, Susanta Mahato, Sk Ziaul, Pankaj Singha, and Rajesh Sarda for their assistance during field survey and software handling.

Authors' contributions Both the authors contributed to the study conception and design. Conceptualization; methodology designing; supervision; editing; and reviewing were performed by Swades Pal. Data curation; investigation; formal analysis; validation; and writing of original draft were performed by Sandipta Debanshi. Both the authors read and approved the final manuscript.

Data Availability The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval Not applicable

Consent to participate Not applicable

Consent to publish Not applicable

References

- Abbot J, Marohasy J (2014) Input selection and optimisation for monthly rainfall forecasting in Queensland, Australia, using artificial neural networks. *Atmos Res* 138:166–178
- Adel MM (2013) Upstream water piracy, the strongest weapon of cornering a downstream nation. *Environ Ecol Res* 1(3):85–128
- Al-Abadi AM (2017) Modeling of groundwater productivity in northeastern Wasit Governorate, Iraq using frequency ratio and Shannon's entropy models. *Appl Water Sci* 7(2):699–716
- Alam MZ (2014) Status of biodiversity at wetland ecosystem of Mohangonj Upazila in Netrakona District. *Adv Ecol* 2014:1–8
- Almuktar SA, Abed SN, Scholz M (2018) Wetlands for wastewater treatment and subsequent recycling of treated effluent: a review. *Environ Sci Pollut Res* 25(24):23595–23623
- Arabameri A, Cerdá A, Tiefenbacher JP (2019a) Spatial pattern analysis and prediction of gully erosion using novel hybrid model of entropy-weight of evidence. *Water* 11(6):1129
- Arabameri A, Pradhan B, Rezaei K, Lee CW (2019b) Assessment of landslide susceptibility using statistical-and artificial intelligence-based FR–RF integrated model and multiresolution DEMs. *Remote Sens* 11(9):999
- Arekhi S, Jafarzadeh AA (2014) Forecasting areas vulnerable to forest conversion using artificial neural network and GIS (case study: northern Ilam forests, Ilam province, Iran). *Arab J Geosci* 7(3): 1073–1085
- Asomani-Boateng R (2019) Urban wetland planning and management in Ghana: a disappointing implementation. *Wetlands* 39(2):251–261
- Bagchi K (1944) The ganges delta. Calcutta University Press, Calcutta, p 157
- Bagchi K, Mukerjee KN (1983) Diagnostic survey of West Bengal (s). Dept Geogr Calcutta Univ Pantg Delta Rarh Bengal 42(58):17–19
- Balvanera P, Quijas S, Karp DS, Ash N, Bennett EM, Boumans R et al (2017) Ecosystem services. In: The GEO handbook on biodiversity observation networks. Springer, Cham, pp 39–78
- Bansal S, Lishawa SC, Newman S, Tangen BA, Wilcox D, Albert D et al (2019) Typha (Cattail) Invasion in North American Wetlands: biology, regional problems, impacts, ecosystem services, and management. *Wetlands* 39(4):645–684
- Bassi N, Kumar MD, Sharma A, Pardha-Saradhi P (2014) Status of wetlands in India: a review of extent, ecosystem benefits, threats and management strategies. *J Hydrol Reg Stud* 2:1–19
- Behnia P, Blais-Stevens A (2018) Landslide susceptibility modelling using the quantitative random forest method along the northern portion of the Yukon Alaska Highway Corridor, Canada. *Nat Hazards* 90(3):1407–1426

- Belle JA, Collins N, Jordaan A (2018) Managing wetlands for disaster risk reduction: a case study of the eastern Free State, South Africa. *Jàmbá J Dis Risk Stud* 10(1):1–10
- Bonham-Carter GF (2014) Geographic information systems for geoscientists: modelling with GIS. vol 13. Elsevier Science, Amsterdam
- Borro M, Morandeira N, Salvia M, Minotti P, Perna P, Kandus P (2014) Mapping shallow lakes in a large South American floodplain: a frequency approach on multitemporal Landsat TM/ETM data. *J Hydrol* 512:39–52
- Bourgoin C, Oszwald J, Bourgoin J, Gond V, Blanc L, Dessard H, Phan TV, Sist P, Läderach P, Reymondin L (2020) Assessing the ecological vulnerability of forest landscape to agricultural frontier expansion in the Central Highlands of Vietnam. *Int J Appl Earth Obs Geoinf* 84:101958
- Breiman L (2001) Random forests. *Mach Learn* 45(1):5–32
- Cao S, Zhang J, Su W (2018) Net value of wetland ecosystem services in China. *Earth's Future* 6(10):1433–1441
- Chakraborty R, Talukdar S, Basu T, Pal S (2018) Habitat identity crisis caused by the riparian wetland squeeze in Tangon River Basin, Barind Region, India. *Spat Inf Res* 26(5):507–516
- Chen W, Xie X, Peng J, Wang J, Duan Z, Hong H (2017a) GIS-based landslide susceptibility modelling: a comparative assessment of kernel logistic regression, Naïve-Bayes tree, and alternating decision tree models. *Geomat Nat Haz Risk* 8(2):950–973
- Chen W, Xie X, Wang J, Pradhan B, Hong H, Bui DT, Ma J (2017b) A comparative study of logistic model tree, random forest, and classification and regression tree models for spatial prediction of landslide susceptibility. *Catena* 151:147–160
- Chen W, Shirzadi A, Shahabi H, Ahmad BB, Zhang S, Hong H, Zhang N (2017c) A novel hybrid artificial intelligence approach based on the rotation forest ensemble and naïve Bayes tree classifiers for a landslide susceptibility assessment in Langao County, China. *Geomat Nat Haz Risk* 8(2):1955–1977
- Chen W, Li H, Hou E, Wang S, Wang G, Panahi M et al (2018a) GIS-based groundwater potential analysis using novel ensemble weights-of-evidence with logistic regression and functional tree models. *Sci Total Environ* 634:853–867
- Chen W, Shahabi H, Zhang S, Khosravi K, Shirzadi A, Chapi K et al (2018b) Landslide susceptibility modeling based on GIS and novel bagging-based kernel logistic regression. *Appl Sci* 8(12):2540
- Chen F, Yu B, Li B (2018c) A practical trial of landslide detection from single-temporal Landsat8 images using contour-based proposals and random forest: a case study of national Nepal. *Landslides* 15(3): 453–464
- Chen W, Sun Z, Han J (2019) Landslide susceptibility modeling using integrated ensemble weights of evidence with logistic regression and random forest models. *Appl Sci* 9(1):171
- CLEAR (2002) Forest Fragmentation in Connecticut: 1985–2006. Center for Land Use Education and Research. University of Connecticut, Middlesex County Extension Centre, USA. <http://clear.uconn.edu/projects/landscape/forestfrag>. Accessed September 7, 2019
- Cong P, Chen K, Qu L, Han J (2019) Dynamic changes in the wetland landscape pattern of the Yellow River Delta from 1976 to 2016 based on satellite data. *Chin Geogr Sci* 29(3):372–381
- Dang VH, Hoang ND, Nguyen LMD, Bui DT, Samui P (2020) A novel GIS-based random forest machine algorithm for the spatial prediction of shallow landslide susceptibility. *Forests* 11(1):118
- Daniel GG (2013) Artificial neural network. In: Runehov A, Oviedo L (eds) Encyclopedia of sciences and religions. Springer, Netherlands, pp 143–143
- 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
- Das RT, Pal S (2018) Investigation of the principal vectors of wetland loss in Barind tract of West Bengal. *GeoJournal* 83(5):1115–1131
- Davidson NC (2018) Wetland losses and the status of wetland-dependent species. In: Finlayson CM, Milton GR, Prentice RC, Davidson NC (eds) The wetland book: II: distribution. Description and Conservation, Springer, Dordrecht, pp 369–381. https://doi.org/10.1007/978-94-007-4001-3_197
- De Groot D, Brander L, Finlayson CM (2018) Wetland ecosystem services. In: Finlayson CM, Everard M, Irvine K, McInnes RJ, Middleton BA, AAV D, Davidson NC (eds) The wetland book: I: structure and function, management, and methods. Springer, Dordrecht, pp 323–333
- Debanshi S, Pal S (2020a) Effects of water richness and seasonality on atmospheric methane emission from the wetlands of deltaic environment. *Ecol Indic* 118:106767
- Debanshi S, Pal S (2020b) Wetland delineation simulation and prediction in deltaic landscape. *Ecol Indic* 108:105757
- Debanshi S, Pal S (2020c) Assessing gully erosion susceptibility in Mayurakshi river basin of eastern India. *Environ Dev Sustain* 22(2):883–914
- Dietterich TG (2000) Ensemble methods in machine learning. In: International workshop on multiple classifier systems. Springer, Berlin, pp 1–15
- Dodangeh E, Choubin B, Eigdir AN, Nabipour N, Panahi M, Shamshirband S, Mosavi A (2019) Integrated machine learning methods with resampling algorithms for flood susceptibility prediction. *Sci Total Environ*:135983
- Donchyts G, Schellekens J, Winsemius H, Eisemann E, van de Giesen N (2016) A 30 m resolution surface water mask including estimation of positional and thematic differences using Landsat 8, SRTM and OpenStreetMap: a case study in the Murray-Darling Basin, Australia. *Remote Sens* 8(5):386
- Dou J, Yunus AP, Tien Bui D, Sahana M, Chen CW, Zhu Z, Wang W, Pham BT (2019) Evaluating GIS-based multiple statistical models and data mining for earthquake and rainfall-induced landslide susceptibility using the LiDAR DEM. *Remote Sens* 11(6):638
- Duan G, Niu R (2018) Lake area analysis using exponential smoothing model and long time-series Landsat images in Wuhan, China. *Sustainability* 10(1):149
- Dube F, Nhapi I, Murwira A, Gumindoga W, Goldin J, Mashauri DA (2014) Potential of weight of evidence modelling for gully erosion hazard assessment in Mbire District-Zimbabwe. *Phys Chem Earth A/B/C* 67:145–152
- El-Askary HM, Lee S, Heggy E, Pradhan B (eds) (2019) Advances in Remote Sensing and Geo Informatics Applications: Proceedings of the 1st Springer Conference of the Arabian Journal of Geosciences (CAJG-1), Tunisia 2018. Springer Nature, Cham
- Erwin KL (2009) Wetlands and global climate change: the role of wetland restoration in a changing world. *Wetl Ecol Manag* 17(1):71–84
- Everard M, Kangabam R, Tiwari MK, McInnes R, Kumar R, Talukdar GH, Dixon H, Joshi P, Allan R, Joshi D, Das L (2019) Ecosystem service assessment of selected wetlands of Kolkata and the Indian Gangetic Delta: multi-beneficial systems under differentiated management stress. *Wetl Ecol Manag* 27(2-3):405–426
- Fan J, Wu L, Zhang F, Cai H, Zeng W, Wang X, Zou H (2019) Empirical and machine learning models for predicting daily global solar radiation from sunshine duration: A review and case study in China. *Renew Sust Energ Rev* 100:186–212
- Fang C, Wen Z, Li L, Du J, Liu G, Wang X, Song K (2019) Agricultural development and implication for wetlands sustainability: a case from Baoqing County, Northeast China. *Chin Geogr Sci* 29(2): 231–244
- Finlayson CM, Milton GR, Prentice RC (2016) Wetland types and distribution. In: Finlayson CM, Milton GR, Prentice RC, Davidson NC (eds) The wetland book: II: distribution. Description and Conservation, Springer, Dordrecht, pp 1–17. https://doi.org/10.1007/978-94-007-6173-5_186-1

- Galatowitsch SM (2016) Natural and anthropogenic drivers of wetland change. In: Finlayson CM, Everard M, Irvine K, McInnes RJ, Middleton BA, Dam AAV, Davidson NC (eds) The wetland book: I: structure and function, management, and methods. Springer, Dordrecht, pp 1–10. https://doi.org/10.1007/978-94-007-6173-5_217-1
- Gayen A, Saha S (2017) Application of weights-of-evidence (WoE) and evidential belief function (EBF) models for the delineation of soil erosion vulnerable zones: a study on Pathro river basin, Jharkhand, India. *Model Earth Syst Environ* 3(3):1123–1139
- Gerbeaux P, Finlayson CM, van Dam AA (2016) Wetland classification: overview. In: Finlayson, CM, Everard M, Irvine K, McInnes RJ, Middleton BA, Dam AAV, Davidson NC (ed) The Wetland Book: I: Structure and Function, Management, and Methods, Springer, Dordrecht, pp 1–8. https://doi.org/10.1007/978-94-007-6172-8_329-1
- Grzybowski M, Glińska-Lewczuk K (2019) Principal threats to the conservation of freshwater habitats in the continental biogeographical region of Central Europe. *Biodivers Conserv* 28(14):4065–4097
- Han J, Park S, Kim S, Son S, Lee S, Kim J (2019) Performance of logistic regression and support vector machines for seismic vulnerability assessment and mapping: a case study of the 12 September 2016 ML5. 8 Gyeongju Earthquake, South Korea. *Sustainability* 11(24):7038
- Herrera VM, Khoshgoftaar TM, Villanustre F, Furht B (2019) Random forest implementation and optimization for Big Data analytics on LexisNexis's high performance computing cluster platform. *J Big Data* 6(1):68
- Hong H, Liu J, Bui DT, Pradhan B, Acharya TD, Pham BT, Zhu AX, Chen W, Ahmad BB (2018) Landslide susceptibility mapping using J48 Decision Tree with AdaBoost, Bagging and Rotation Forest ensembles in the Guangchang area (China). *Catena* 163:399–413
- Islam SN (2016) Deltaic floodplains development and wetland ecosystems management in the Ganges–Brahmaputra–Meghna Rivers Delta in Bangladesh. *Sustain Water Resour Manag* 2:237–256. <https://doi.org/10.1007/s40899-016-0047-6>
- Jaafari A, Najafi A, Pourghasemi HR, Rezaeian J, Sattarian A (2014) GIS-based frequency ratio and index of entropy models for landslide susceptibility assessment in the Caspian forest, northern Iran. *Int J Environ Sci Technol* 11(4):909–926
- Jamali A (2019) Evaluation and comparison of eight machine learning models in land use/land cover mapping using Landsat 8 OLI: A case study of the northern region of Iran. *SN Appl Sci* 1(11):1448
- 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
- Jung HS, Lee S (2019) Special issue on machine learning techniques applied to geoscience information system and remote sensing. *Appl. Sci.* 9(12):2446
- Kadavi PR, Lee CW, Lee S (2018) Application of ensemble-based machine learning models to landslide susceptibility mapping. *Remote Sens* 10(8):1252
- Kanevski M, Timonin V, Pozdnukhov A (2009) Machine learning for spatial environmental data: Theory, Applications, and Software. EPFL Press
- Kim JC, Jung HS, Lee S (2019) Spatial mapping of the groundwater potential of the Geum River Basin using ensemble models based on remote sensing images. *Remote Sens* 11(19):2285
- Krig S (2016) Image pre-processing. In: Computer Vision Metrics. Springer, Cham, pp 35–74
- Lamsal P, Atreya K, Pant KP, Kumar L (2017) People's dependency on wetlands: South Asia perspective with emphasis on Nepal. In: Wetland Science. Springer, New Delhi, pp 407–419
- Landsat project science office (2002) Landsat 7 science data user's handbook. URL: https://landsat.gsfc.nasa.gov/wp-content/uploads/2016/08/Landsat7_Handbook.pdf. Accessed August 2019
- Lary DJ, Zewdie GK, Liu X, Wu D, Levetin E, Allee RJ et al (2018) Machine learning applications for earth observation. In: Mathieu PP, Aubrecht C (eds) Earth observation open science and innovation. Springer Nature, Cham, p 165–218
- Lee S, Hong SM, Jung HS (2017) A support vector machine for landslide susceptibility mapping in Gangwon Province, Korea. *Sustainability* 9(1):48
- Leon A, Tang Y, Chen D, Yolcu A, Glennie C, Pennings S (2018) Dynamic management of water storage for flood control in a wetland system: a case study in Texas. *Water* 10(3):325
- Liu Y, Sheng L, Liu J (2015) Impact of wetland change on local climate in semi-arid zone of Northeast China. *Chin Geogr Sci* 25(3):309–320
- Liu L, Silva EA, Wu C, Wang H (2017) A machine learning-based method for the large-scale evaluation of the qualities of the urban environment. *Comput Environ Urban Syst* 65:113–125
- Liu Y, Yang Y, Jing W, Yue X (2018) Comparison of different machine learning approaches for monthly satellite-based soil moisture downscaling over Northeast China. *Remote Sens* 10(1):31
- Mahmud MS, Masrur A, Ishtiaque A, Haider F, Habiba U (2011) Remote sensing & GIS based spatio-temporal change analysis of Wetland in Dhaka City, Bangladesh. *J Water Resour Prot* 3(11):781
- Malekmohammadi B, Jahanishakib F (2017) Vulnerability assessment of wetland landscape ecosystem services using driver-pressure-state-impact-response (DPSIR) model. *Ecol Indic* 82:293–303
- Malik M, Pai SC (2019) Drivers of land/cover change and its impact on pong dam wetland. *Environ Monit Assess* 191–200:203. <https://doi.org/10.1007/s10661-019-7347-x>
- McInnes RJ (2016) Climate regulation and wetlands: overview. In: Finlayson CM, Everard M, Irvine K, McInnes RJ, Middleton BA, Dam AAV, Davidson NC (eds) The wetland book: I: structure and function, management, and methods. Springer, Dordrecht, pp 1–7. https://doi.org/10.1007/978-94-007-6172-8_231-1
- Meng L, Dong J (2019) LUCC and ecosystem service value assessment for wetlands: a case study in Nansi Lake, China. *Water* 11(8):1597
- Meten M, PrakashBhandary N, Yatabe R (2015) Effect of landslide factor combinations on the prediction accuracy of landslide susceptibility maps in the Blue Nile Gorge of Central Ethiopia. *Geoenviron Disasters* 2(1):9
- Metzger MJ, Schröter D (2006) Towards a spatially explicit and quantitative vulnerability assessment of environmental change in Europe. *Reg Environ Chang* 6(4):201–216
- Micheletti N, Foresti L, Kanevski M, Pedrazzini A, Jaboyedoff M (2011) Landslide susceptibility mapping using adaptive support vector machines and feature selection. *Geophysical Research Abstracts, EGU*, 13
- Miller F, Osbahr H, Boyd E, Thomalla F, Bharwani S, Zier vogel G et al (2010) Resilience and vulnerability: complementary or conflicting concepts? *Ecol Soc* 15(3):11. <http://www.ecologyandsociety.org/vol15/iss3/art11/>
- Milton GR, Prentice RC, Finlayson CM (2018) Wetlands of the world. In: Finlayson CM, Milton GR, Prentice RC, Davidson NC (eds) The wetland book:II: Distribution, description and conservation, Springer, Dordrecht, pp 3–16. http://doi.org/443.webvpn.fjmu.edu.cn/10.1007/978-94-007-4001-3_182
- Mohamed WNHW, Salleh MNM, Omar AH (2012) A comparative study of reduced error pruning method in decision tree algorithms. In: 2012 IEEE International conference on control system, computing and engineering. IEEE, Penang, pp 392–397). IEEE
- Mandal DK, Kaviraj A (2009) Distribution of fish assemblages in two floodplain lakes of North 24-Parganas in West Bengal, India. *J Fish Aquat Sci* 4(1):12–21
- Mondal B, Dolui G, Pramanik M, Maity S, Biswas SS, Pal R (2017) Urban expansion and wetland shrinkage estimation using a GIS-based model in the East Kolkata Wetland, India. *Ecol Indic* 83: 62–73

- Morganti M, Manica M, Bogliani G, Gustin M, Luoni F, Trott P, Perin V, Brambilla M (2019) Multi-species habitat models highlight the key importance of flooded reedbeds for inland wetland birds: implications for management and conservation. *Avian Res* 10(1):15
- Mosavi A, Ozturk P, Chau KW (2018) Flood prediction using machine learning models: literature review. *Water* 10(11):1536
- Mukherjee K, Pal S, Mukhopadhyay M (2018) Impact of flood and seasonality on wetland changing trends in the Diara region of West Bengal, India. *Spat Inf Res* 26(4):357–367
- Nohani E, Moharrami M, Sharifi S, Khosravi K, Pradhan B, Pham BT et al (2019) Landslide susceptibility mapping using different GIS-based bivariate models. *Water* 11(7):1402
- Pal S (2011) Conservation or conversion of wetland in the Riverine Bengal basin: a question of hydro-ecological profit loss. *Pract Geogr Kolkata* 15(1):09–24
- Pal S (2016) Impact of Massanjore dam on hydro-geomorphological modification of Mayurakshi River, Eastern India. *Environ Dev Sustain* 18(3):921–944
- Pal S, Akoma OC (2009) Water scarcity in wetland area within Kandi Block of West Bengal: a hydro-ecological assessment. *Ethiop J Environ Stud Manag*. <https://doi.org/10.4314/EJESM.V2I3.48260>
- Pal S, Saha TK (2018) Identifying dam-induced wetland changes using an inundation frequency approach: the case of the Atreyee River basin of Indo-Bangladesh. *Ecohydrobiol* 18(1):66–81
- Pal S, Talukdar S (2018a) Drivers of vulnerability to wetlands in Punarbhaba river basin of India-Bangladesh. *Ecol Indic* 93:612–626
- Pal S, Talukdar S (2018b) Application of frequency ratio and logistic regression models for assessing physical wetland vulnerability in Punarbhaba river basin of Indo-Bangladesh. *Hum Ecol Risk Assess* 24(5):1291–1311
- Pal S, Talukdar S, Ghosh R (2020) Damming effect on habitat quality of riparian corridor. *Ecol Indic* 114:106300
- Parent J, Civco D, Hurd J (2007) Simulating future forest fragmentation in a Connecticut region undergoing suburbanization. In: ASPRS 2007 Annual Conference Tampa, Florida
- Patriche CV, Pirnau R, Grozavu A, Rosca B (2016) A comparative analysis of binary logistic regression and analytical hierarchy process for landslide susceptibility assessment in the Dobrov River Basin, Romania. *Pedosphere* 26(3):335–350
- Paul S, Pal S (2019) Exploring wetland transformations in moribund deltaic parts of India. *Geocarto Int*. <https://doi.org/10.1080/10106049.2019.1581270>
- Peters J, De Baets B, Verhoest NE, Samson R, Degroeve S, De Becker P, Huybrechts W (2007) Random forests as a tool for ecohydrological distribution modelling. *Ecol Model* 207(2–4):304–318
- Pham BT, Prakash I, Singh SK, Shirzadi A, Shahabi H, Bui DT (2019a) Landslide susceptibility modeling using reduced error pruning trees and different ensemble techniques: Hybrid machine learning approaches. *Catena* 175:203–218
- Pham BT, Shirzadi A, Shahabi H, Omidvar E, Singh SK, Sahana M et al (2019b) Landslide susceptibility assessment by novel hybrid machine learning algorithms. *Sustainability* 11(16):4386
- Pham BT, Avand M, Janizadeh S, Phong TV, Al-Ansari N, Ho LS et al (2020) GIS based hybrid computational approaches for flash flood susceptibility assessment. *Water* 12(3):683
- Phethi MD, Gumbo JR (2019) Assessment of impact of land use change on the wetland in Makhitha village, Limpopo province, South Africa. *Jambah J Disaster Risk Stud* 11(2):693. <https://doi.org/10.4102/jamba.v11i2.693>
- Pourghasemi HR, Mohammady M, Pradhan B (2012) Landslide susceptibility mapping using index of entropy and conditional probability models in GIS: Safarood Basin, Iran. *Catena* 97:71–84
- Pradhan B (2013) A comparative study on the predictive ability of the decision tree, support vector machine and neuro-fuzzy models in landslide susceptibility mapping using GIS. *Comput Geosci* 51: 350–365
- Pradhan AMS, Dawadi A, Kim YT (2012) Use of different bivariate statistical landslide susceptibility methods: a case study of Khulekhani watershed, Nepal. *J Nepal Geol Soc* 44:1–12
- Prasher K (2018) The state of India's disappearing wetlands. *The Weather Channel India*. <https://weather.com/en-IN/india/news/news/2018-11-08-the-case-of-indias-disappearing-wetlands>. Accessed July 2019
- Quinlan JR (1987) Simplifying decision trees. *Int J Man Mach Stud* 27(3):221–234
- Raja NB, Çiçek I, Türkoğlu N, Aydin O, Kawasaki A (2017) Landslide susceptibility mapping of the Sera River Basin using logistic regression model. *Nat Hazards* 85(3):1323–1346
- Ramsar Convention on Wetlands (2018) Global Wetland Outlook: state of the world's wetlands and their Services to people. Ramsar Convention Secretariat, Gland
- 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. *Geoenviron Disasters* 3(1):19
- Rodriguez-Galiano VF, Ghimire B, Rogan J, Chica-Olmo M, Rigol-Sánchez JP (2012) An assessment of the effectiveness of a random forest classifier for land-cover classification. *ISPRS J Photogramm Remote Sens* 67:93–104
- Rodriguez-Galiano V, Sanchez-Castillo M, Chica-Olmo M, Chica-Rivas MJGR (2015) Machine learning predictive models for mineral prospectivity: an evaluation of neural networks, random forest, regression trees and support vector machines. *Ore Geol Rev* 71:804–818
- Rojas C, Munizaga J, Rojas O, Martinez C, Pino J (2019) Urban development versus wetland loss in a coastal Latin American city: lessons for sustainable land use planning. *Land Use Policy* 80:47–56. <https://doi.org/10.1016/j.landusepol.2018.09.036>
- Rokni K, Ahmad A, Selamat A, Hazini S (2014) Water feature extraction and change detection using multitemporal Landsat imagery. *Remote Sens* 6(5):4173–4189
- Saha TK, Pal S (2019) Exploring physical wetland vulnerability of Atreyee river basin in India and Bangladesh using logistic regression and fuzzy logic approaches. *Ecol Indic* 98:251–265
- Sajedi-Hosseini F, Malekian A, Choubin B, Rahmati O, Cipullo S, Coulon F, Pradhan B (2018) A novel machine learning-based approach for the risk assessment of nitrate groundwater contamination. *Sci Total Environ* 644:954–962
- Sarkar D, Mondal P (2020) Flood vulnerability mapping using frequency ratio (FR) model: a case study on Kulik river basin, Indo-Bangladesh Barind region. *Appl Water Sci* 10(1):17
- Sîrbu F, Drăguț L, Oguchi T, Hayakawa Y, Micu M (2019) Scaling landscape variables for landslide detection. *Progress Earth Planet Sci* 6(1):44
- Sokolova M, Japkowicz N, Szpakowicz S (2006) Beyond accuracy, F-score and ROC: a family of discriminant measures for performance evaluation. In: Australasian joint conference on artificial intelligence. Springer, Berlin, pp 1015–1021
- Srinivasan DB, Mekala P (2014) Mining social networking data for classification using reptree. *Int J Adv Res Comput Sci Manag Stud* 2(10):155–160
- 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*. <https://doi.org/10.1007/s13157-018-1039-0>
- Talbot CJ, Bennett EM, Cassell K, Hanes DM, Minor EC, Paerl H, Raymond PA, Vargas R, Vidon PG, Wollheim W, Xenopoulos MA (2018) The impact of flooding on aquatic ecosystem services. *Biogeochemistry* 141(3):439–461
- Tayyebi A, Pijanowski BC (2014) Modeling multiple land use changes using ANN, CART and MARS: comparing tradeoffs in goodness of

- fit and explanatory power of data mining tools. *Int J Appl Earth Obs Geoinf* 28:102–116
- Tharwat A (2020) Classification assessment methods. *Appl Comput Inf*. <https://doi.org/10.1016/j.aci.2018.08.003>
- Tockner K, Pusch M, Borchardt D, Lorang MS (2010) Multiple stressors in coupled river–floodplain ecosystems. *Freshw Biol* 55:135–151
- Townshend JR, Justice CO (1986) Analysis of the dynamics of African vegetation using the normalized difference vegetation index. *Int J Remote Sens* 7(11):1435–1445
- Trevisan DP, da Conceição Bispo P, Almeida D, Imani M, Balzter H, Moschini LE (2020) Environmental vulnerability index: an evaluation of the water and the vegetation quality in a Brazilian Savanna and Seasonal Forest biome. *Ecol Indic* 112:106163
- UNEP, U (2011) Towards a green economy: pathways to sustainable development and poverty eradication. UNEP, Nairobi
- Vakhshoori V, Pourghasemi HR, Zare M, Blaschke T (2019) Landslide susceptibility mapping using GIS-based data mining algorithms. *Water* 11(11):2292
- Vapnik V (1963) Pattern recognition using generalized portrait method. *Autom Remote Control* 24:774–780
- Vapnik, V. (2013). The nature of statistical learning theory. Springer.
- Vapnik V, Chervonenkis A (1964) A note on one class of perceptions. *Autom Remote Control* 25:821–837
- Vogt P, Riitters KH, Estreguil C, Kozak J, Wade TG, Wickham JD (2007) Mapping spatial patterns with morphological image processing. *Landsc Ecol* 22(2):171–177
- VoPham T, Hart JE, Laden F, Chiang YY (2018) Emerging trends in geospatial artificial intelligence (geoAI): potential applications for environmental epidemiology. *Environ Health* 17(1):40
- Warns new report (2018) Wetlands – world's most valuable ecosystem – disappearing three times faster than forests. URL: <https://www.ramsar.org/news/wetlands-worlds-most-valuable-ecosystem-disappearing-three-times-faster-than-forests-warns-new>. Accessed July 2019
- White MP, Weeks A, Hooper T, Bleakley L, Cracknell D, Lovell R, Jefferson RL (2017) Marine wildlife as an important component of coastal visits: the role of perceived biodiversity and species behaviour. *Mar Policy* 78:80–89
- Wondie A (2018) Ecological conditions and ecosystem services of wetlands in the Lake Tana Area, Ethiopia. *Ecohydrol Hydrobiol* 18(2): 231–244
- Wu Z, Wu Y, Yang Y, Chen F, Zhang N, Ke Y, Li W (2017) A comparative study on the landslide susceptibility mapping using logistic regression and statistical index models. *Arab J Geosci* 10(8):187
- Wu C, Chen W, Cao C, Tian R, Liu D, Bao D (2018) Diagnosis of wetland ecosystem health in the Zoige Wetland, Sichuan of China. *Wetlands* 38(3):469–484
- Xiao S, Xiao H, Peng X, Song X (2015) Hydroclimate-driven changes in the landscape structure of the terminal lakes and wetlands of the China's Heihe River Basin. *Environ Monit Assess* 187(1):4091
- Xie J, Song Z, Li Y, Zhang Y, Yu H, Zhan J, Ma Z, Qiao Y, Zhang J, Guo J (2018) A survey on machine learning-based mobile big data analysis: challenges and applications. *Wirel Commun Mob Comput*. <https://doi.org/10.1155/2018/8738613>
- Xu E, Chen Y (2019) Modeling intersecting processes of wetland shrinkage and urban expansion by a time-varying methodology. *Sustainability* 11(18):4953
- Xu T, Weng B, Yan D, Wang K, Li X, Bi W, Li M, Cheng X, Liu Y (2019) Wetlands of international importance: status, threats, and future protection. *Int J Environ Res Public Health* 16(10):1818
- Xue Z, Lyu X, Chen Z, Zhang Z, Jiang M, Zhang K, Lyu Y (2018) Spatial and temporal changes of wetlands on the Qinghai-Tibetan Plateau from the 1970s to 2010s. *Chin Geogr Sci* 28(6):935–945
- Yalcin A, Reis S, Aydinoglu AC, Yomralioglu T (2011) A GIS-based comparative study of frequency ratio, analytical hierarchy process, bivariate statistics and logistics regression methods for landslide susceptibility mapping in Trabzon, NE Turkey. *Catena* 85(3):274–287
- Yan B, Wang J, Li S, Cui L, Ge Z, Zhang L (2016) Assessment of socio-economic vulnerability under sea level rise coupled with storm surge in the Chongming County, Shanghai. *Acta Ecol Sin* 36(2):91–98
- Yang X, Chen R, Zheng XQ (2016) Simulating land use change by integrating ANN-CA model and landscape pattern indices. *Geomat Nat Haz Risk* 7(3):918–932
- Yang W, Xu K, Lian J, Ma C, Bin L (2018) Integrated flood vulnerability assessment approach based on TOPSIS and Shannon entropy methods. *Ecol Indic* 89:269–280
- Yikii F, Turyahabwe N, Bashaasha B (2017) Prevalence of household food insecurity in wetland adjacent areas of Uganda. *Agric Food Secur* 6(1):63
- 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 Wadi Tayyah Basin, Asir Region, Saudi Arabia. *Landslides* 13(5):839–856
- Zha Y, Gao J, Ni S (2003) Use of normalized difference built-up index in automatically mapping urban areas from TM imagery. *Int J Remote Sens* 24(3):583–594
- Zheng Y, Liu H, Zhuo Y, Li Z, Liang C, Wang L (2019) Dynamic changes and driving factors of wetlands in Inner Mongolia Plateau, China. *PLoS One* 14(8):e0221177
- Ziaul S, Pal S (2017) Estimating wetland insecurity index for Chatra wetland adjacent English Bazar Municipality of West Bengal. *Spat Inf Res* 25(6):813–823

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