


# Deforestation probable area predicted by logistic regression in Pathro river basin: a tributary of Ajay river

Amiya Gayen<sup>1</sup> · Sunil Saha<sup>1</sup> 

Received: 5 September 2017 / Revised: 15 November 2017 / Accepted: 16 November 2017 / Published online: 24 November 2017  
© Korean Spatial Information Society 2017

**Abstract** Deforestation threatens biodiversity in remaining forest in India. Today majority of populated areas are facing huge anthropogenic deforestation and it is one of the greatest problems in our country. For the sustainable management of forest there is a need of prediction about the probability of deforestation, i.e. which areas are most susceptibility to deforestation. This study reveals a methodology for predicting the areas of deforestation based on cultural and natural landscape. Geographical information system and logistic regression have been used to predict the greatest propensity for the deforestation of Pathro river basin. The logistic regression model has proven that the deforestation is an integrated function of altitude, slope, slope aspect, distance from road, settlement, river and forest edge. The independent variables are strongly correlated with deforestation. Finally, the receiver operating characteristic curve has been drawn for the validation of deforestation probability map and the area under the curve (AUC) is commuted for verification and measurement of level of accuracy. The AUC for the logistic regression model has shown 76.6% prediction accuracy. The result reveals that the performance logistic regression is good enough in simulation of deforestation process. This model also predicted the areas with high potential for future deforestation.

**Keywords** Logistic regression · ROC curve · Deforestation · Forest loss · Area under the curve

## 1 Introduction

Deforestation is a quasi-natural processes occurring presently over the earth's surface. Some of the major problems associated with deforestation are climate change, disruption in atmospheric carbon balance and loss of biodiversity. Deforestation and forest degradation is the second largest source of carbon emission [1]. Today most of the county's population are continuing expectations of improvement their standard of living which leads the pressure on natural resources [2]. So, as a result, increasing pressure of human development is reflected on forest degradation [3], fragmentation of ecological niche [4], loss of wildlife corridors [5], enhanced rate of soil erosion [6] and finally it initiates the human animal conflict [7]. About 9 million sq. km of tropical humid forest area are lost in less than 50 years but surpassingly the present rate of deforestation is not only high but its increasing continuously. Today a number of methods are used for assessing the rate of deforestation at different scale (i.e., local, regional and global). The models proposed in the literature are commonly subdivided into two main categories; such as quantitative and qualitative [8]. The qualitative models are used for risk assessments and quantitative models are helpful for estimation of deforestation rate, based on measured data or modelling. Various models based on various methods (i.e., empirical and physical based models) are used to quantify deforestations. Today a large number of studies are completed based on remote sensing and geographical information system (GIS) for effective forest cover management [5, 9]. Recent advancement in technologies, such as remote

---

✉ Sunil Saha  
sunilgeo.88@gmail.com

Amiya Gayen  
mr.amiyagayen@gmail.com

<sup>1</sup> Department of Geography, University of Gour Banga, Malda, West Bengal, India

sensing and GIS, and numerical modelling techniques are not only developed as powerful tools for ecological and environmental assessment but also predict land use/land cover assessment more efficiently than ever-before [10]. Numbers of approaches are developed to model and predict the dynamics of land use/land cover [11–15]. Forest cover analysis based on temporal basis correlated with geospatial modelling is no doubt helpful in forecasting the future forest cover scenarios, but spatial modelling can be an immensely useful activity to understand the future of the forests. As various factors like Deforestation, logging, diversion of forests for non-forestry purposes etc. will be operative in future too, which cause continuous changes in forest cover [16].

The present study aims to analysis of deforestation probability at Pathro river basin based on logistic regression model (LRM). Preparation of deforestation probability map is useful to decision makers for identifying deforestation prone areas and existing forest management. In this context, the objective of this study is to identify the probable deforested areas applying logistic regression model which will give us a comprehensive idea about the deforestation rate.

## 2 Study area

Pathro river basin is a 6th order tributary of Ajay River of Jharkhand (Fig. 1). From the ultimate point of view, the study area is located between  $24^{\circ}9'58''\text{N}$ – $24^{\circ}29'50''\text{N}$  latitudes and  $86^{\circ}15'34''\text{E}$ – $86^{\circ}48'18''\text{E}$  longitudes, covering an area of 709 sq. km. Elevation of this basin ranges from 168 to 461 m. Maximum area of this basin fall under moderate to moderately steep slopes with ranges from 50 to 530. Average annual rainfall receives of this basin is 1247 mm mainly intermediate July and October. This area are facing drought like situation from December to June due to high surface runoff potential, and poor infiltration. Due to continuous changes in climatic conditions as well as land use pattern, this region is now facing deforestation problems through environmental degradation.

Local people of this study area use wood immensely for repairing and construction of houses, manufacturing equipment and as a source of fuel. People of surrounding villages cut down lots of valuable wood producing trees illegally from the adjoining forest. The major threat of this area is illegal grazing in the forest which leads to the disruption of forest regeneration.

## 3 Methodology

### 3.1 Database and methods

Deforestation is influenced by complex factors, such as, surface cover and environmental factors. Therefore, to assess the deforestation hazard of the area a range of evaluation criteria, objectives and attributes should be identified with respect to the problem situation. For deforestation probability analysis, seven deforestation conditioning factors are used based on field survey and expert knowledge such as, distance from settlement, distance from forest edge, distance from road, distance from river, slope in degree, slope aspect and altitude. In this study, first the forest cover areas are extracted from both 1991 and 2016 digital maps and finally obtained deforestation map and then the values of 0 and 1 are labelled to non-forested and forested areas, respectively (Fig. 2). With the help of Logistic Regression Model using said deforestation conditioning factors, the deforestation probability map of the Pathro river basin has been prepared.

Finally, the receiver operating characteristic (ROC) curves of the deforestation probability model has been constructed and the area under the curve (AUC) has commuted for verification and accuracy purposes (Fig. 3).

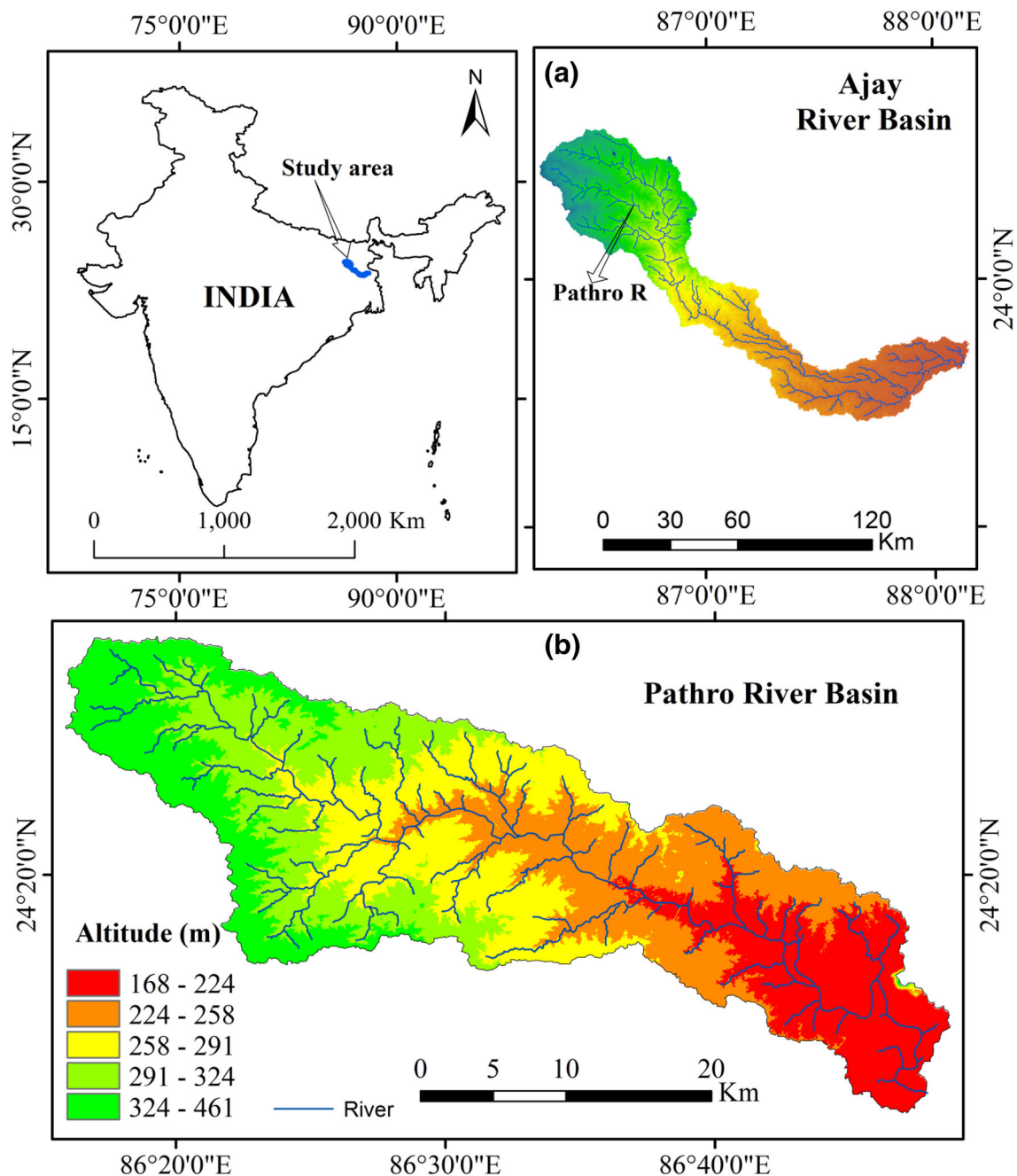
### 3.2 Preparation a data base of effective factors

#### 3.2.1 The dependent variable: forest cover change

Regards this model, the forest cover changed (Since 1991–2016) is considered as dependent variable (Fig. 4). First, with the help of GIS software normalized difference vegetation index (NDVI) has been calculated, and the NDVI value greater than .3 is consider as forest in both cases [17]. Hence, a Boolean image with the categories ‘forest change’ (forest to no forest) and ‘no change’ (forest remained unchanged) were generated for the period 1991–2016 by subtracting the forest cover of 1991 from 2016 (Fig. 5).

#### 3.2.2 The independent variables of forest cover change

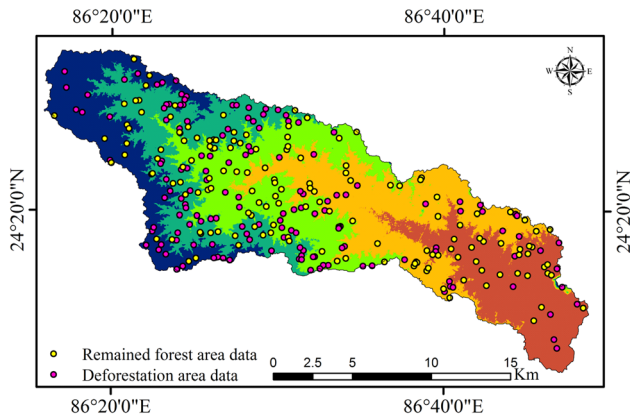
Distance from settlement, forest edge, roads, river, slope, slope aspect and altitude are considered as potential independent variables for forest cover change (Fig. 2). Distance form forest edges are considered as one of the important explanatory variable because deforestation tends to start from the edge of existing forest [18] so, forest edge considered as of high probability of deforestation [19]. Deforestation is also highly correlated with roads and settlements [19] and hence these are considered as



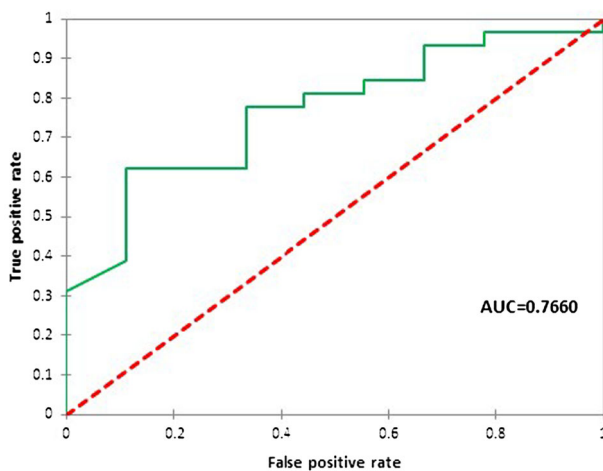
**Fig. 1** Location of the study area **a** The Ajay river basin in south-western part of West Bengal, **b** Pathro river basin with altitude and drainage network

independent variables. Due to Construction of roads, railways and bridges open up the land for further developmental projects and that ultimately attract large number of population to the forest frontiers [20, 21]. These population are usually colonized the forest by using logging trails or construct the new roads to access the forest for subsistence land [20]. Topographical factors such as altitude, slope and slope aspect are strongly correlated with forest cover change [16]. Altitude considered as independent variable

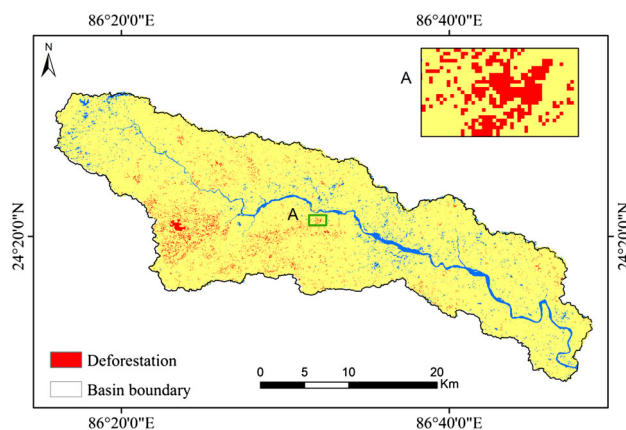
due to its significant direct relationship with total vegetation cover and a significant inverse relationship with annual grass cover [22]. Slope steepness was also showed an inverse relationship with vegetation cover [22] but in respect to slope aspect it's found that north-facing forests had more tree species and higher tree density than the south-facing forests [23]. Beside this expansion of agricultural land is one of the main leading causes of deforestation [24]. Usually the agricultural land is mostly



**Fig. 2** Location of remained forest area and deforestation area of Pathro river basin



**Fig. 3** Validation of LRM prediction (AUC/ROC)



**Fig. 4** Forest cover change, during 1991–2016 (Dependent variable), used for calibration and prediction of LRM

growing near the river side. Therefore distance from river is considered as an important factor for deforestation analysis (Fig. 6c).

### 3.3 Statistical test for association between dependent and independent variables: Cramer's V test

$\chi^2$  (for a contingency table larger than two rows by two columns) test is transformed by Cramer's V statistic to a range of 0–1, where unit value represents complete equality between the two nominal variables [25]. In this research work Cramer's V has used to represent the intensity of association between dependent and independent variables. For the Cramer's V test the deforestation conditioning factors have been considered as independent variables and forest change between 1991 and 2016 is considered as dependent variable. The result of the explanatory test procedure for each variable is Cramer's V values and associated p values. The p values signify the probability that the Cramer's V is not significantly different from 0 [18]. Cramer's V represents the relationship between an individual independent variable and forest cover change. Logistic regression model (LRM) is used to provide a perfect sagacity into this (Table 1).

### 3.4 Logistic regression model (LRM)

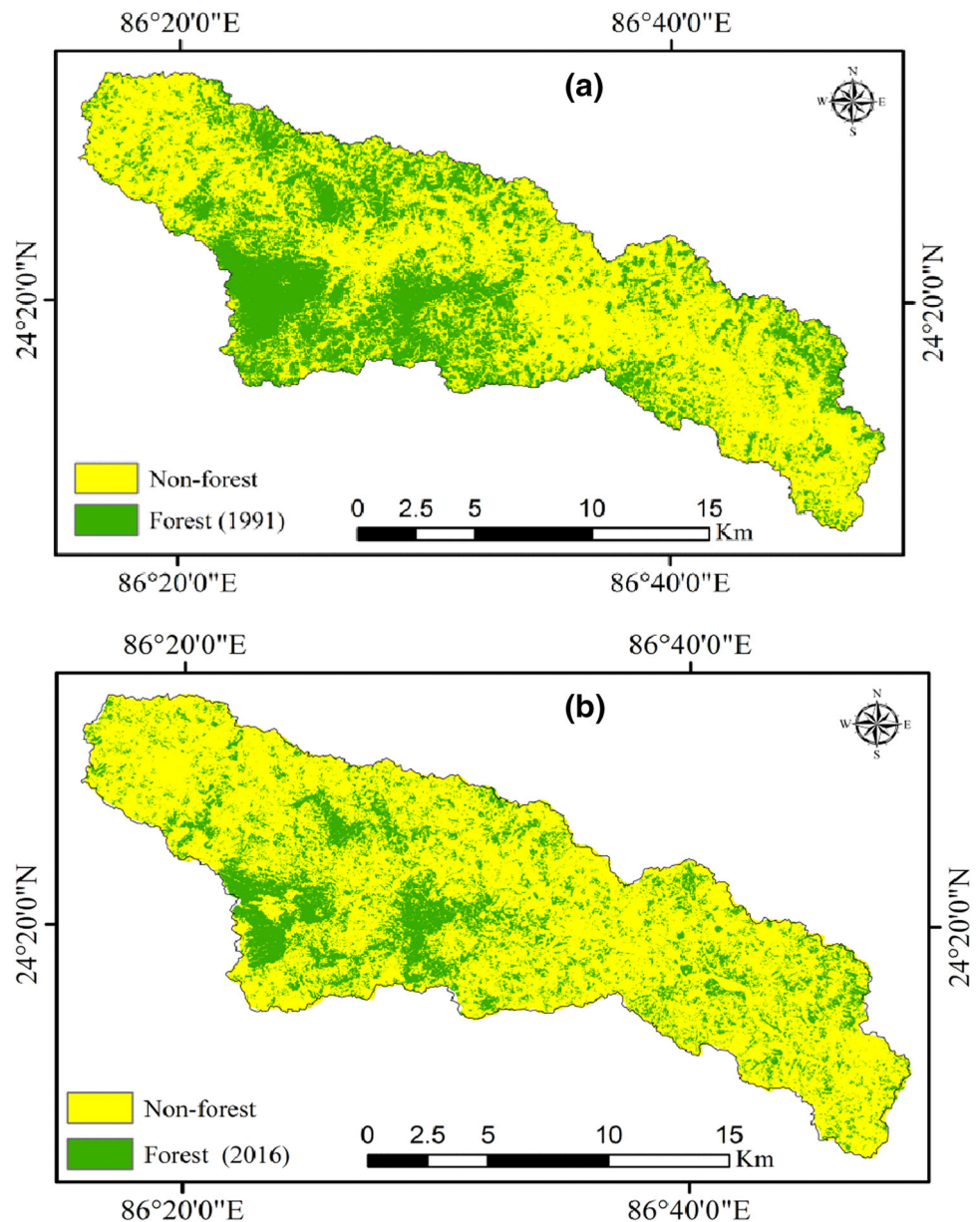
Logistic regression model has been used for the analysis of deforestation probability of Pathro river basin. This LRM is developed based on the binary response variables i.e. 1 for 'forest change' and 0 for 'no change' (Fig. 2) [26] and the explanatory variables (elevation, slope, slope aspect, distance from river, settlements, roads and forest edges). The natural log transformation was done for the continuous variables (distances). The natural log transformation was applied for the continuous variables such as, elevation, slope, distance from river, forest edge, roads and settlements. For the categorical explanatory variable (slope aspect class), the evidence of likelihood transformation was applied. The logistic regression model was calibrated before prediction by including the explanatory variables for 2016 in the Logistic Regression Module as independent variables and the forest change during 1991–2016 as dependent variable.

The logistic regression provides the probability of forest loss a function of the explanatory variables. The logistic function (Eq. 1) based on Pontius and Schneider [14] results bounded between 0 and 1 as follows:

$$P = E(Y) = \frac{\exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 \cdots + \beta_i x_i)}{1 + \exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 \cdots + \beta_i x_i)} \quad (1)$$

where  $P$  is the probability of forest change,  $E(Y)$  is the expected value of the dependent variable  $Y$ ,  $\beta_0$  is a constant to be estimated,  $\beta_i$  is the coefficient to be estimated for each explanatory variable  $x_i$ . This logistic function

**Fig. 5** Forest covers during **a** 1991 and **b** 2016



(Eq. 1) is transformed (Eq. 2) into a linier function (Eq. 3) which is calculated logistic transformation:

$$\text{Logit}(p) = \log e \left( \frac{p}{1-p} \right) \tag{2}$$

$$\text{Logit}(p) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 \dots + \beta_i x_i \tag{3}$$

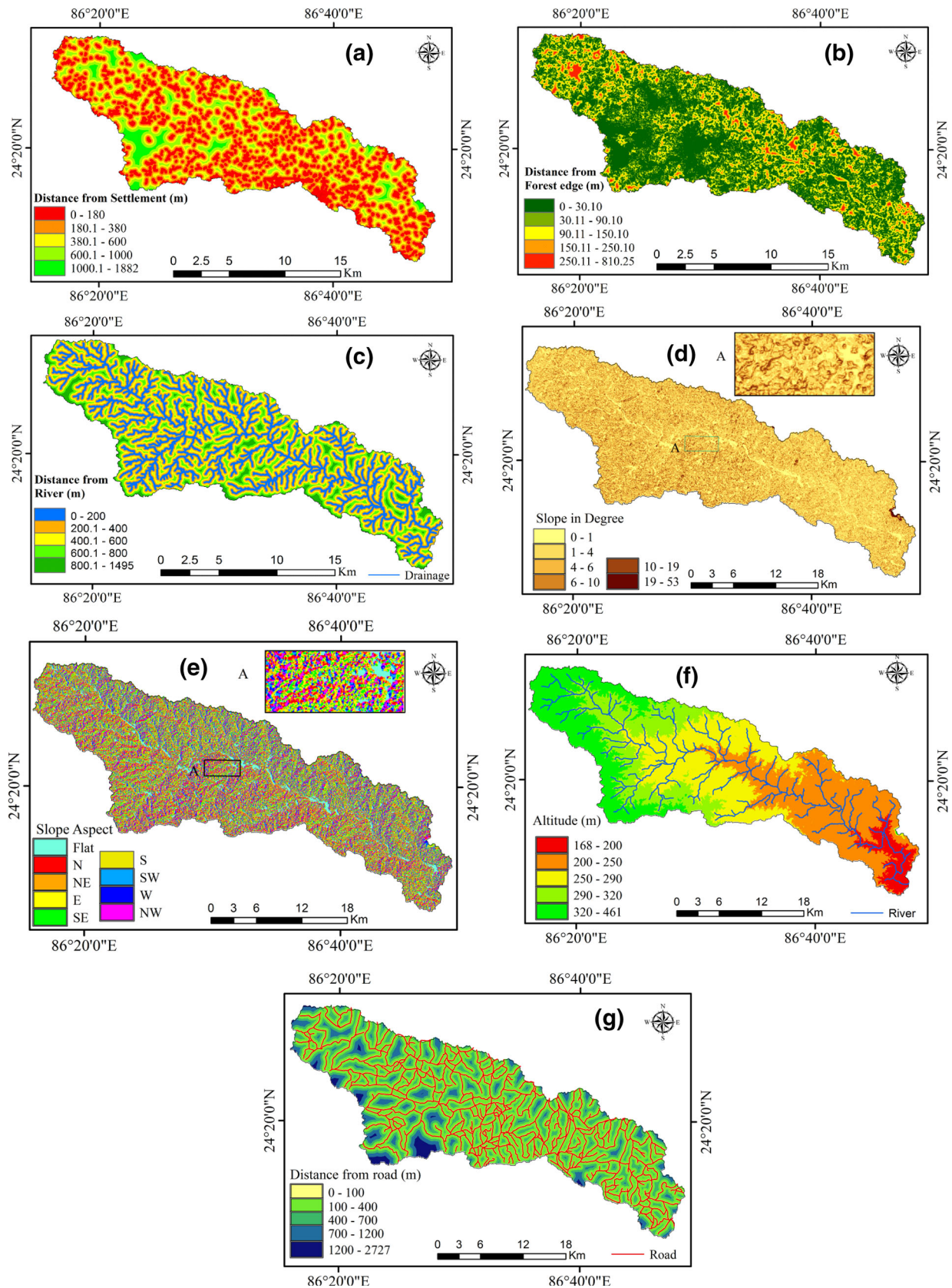
The final result is a probability score (p) for each cell [14].

Here it represents that the logit conversion of dichotomous data confirm that the dependent variable of the regression is continuous, and the new dependent variable (logit transformation of the probability) is infinite. Finally, it confirms that the predicted probability will be continuous within range from 0 to 1 (Fig. 7). The regression equation

of the best-fitted predictor set and the probability of forest change were generated. For evaluate the significance of Logistic regression model, the goodness of fit is an alternative to model  $\chi^2$ . It is calculated based on the differences between the predicted and the observed values of the dependent variable.

### 3.5 Validation of prediction

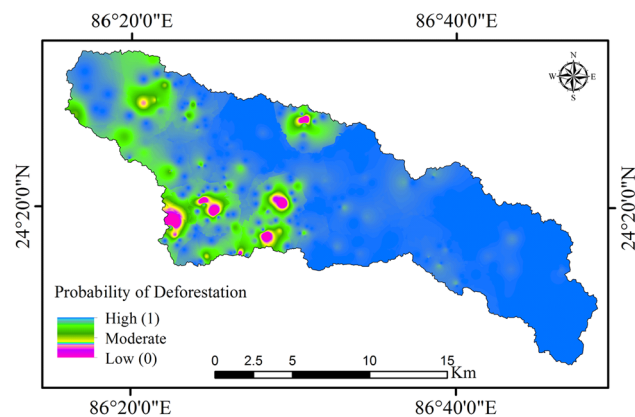
Regard quantitative validation, Receiver Operating Characteristics (ROC) curve has been used by comparing the existing deforestation location in the validation datasets with the deforestation probability map obtained by LRM model. The accuracy of model was evaluated based on



**Fig. 6** Explanatory variables used for calibration of LRM **a** distance from settlements, **b** distance from forest edge, **c** distance from rivers, **d** slope in degree, **e** slope aspect and **f** altitude, **g** distance from road

**Table 1** Association between dependent variable (forest cover change) and explanatory variable using Cramer’s V

Explanatory variables	Cramer’s (V)	P value
Altitude	0.74	0.00
Slope in degree	0.761	0.00
Slope aspect	0.787	0.00
Distance from river	0.58	0.00
Distance from road	0.725	0.00
Distance from settlement	0.915	0.00
Distance from forest edge	0.11	0.00



**Fig. 7** Deforestation Probability map prepared based on logistic regression (LRM)

Google earth, satellite image and field verification. The ROC curve has been constructed based on true positive rate (sensitivity) corresponding false positive rate (1-specificity) with the various cut-off thresholds. Area under the curve (AUC) has used for qualitative analysis of LR model. An AUC value of 1 indicates a perfect model and when AUC equals 0 is indicates a non-informative model. However, the success rate method is useful to define how well the resulting deforestation probability maps are classified the areas of the existing deforestation [27]. Finally, it is stated that GIS-based logistic regression model as an expert knowledge-based approach is very useful for solving complex problems.

**Table 2** Error matrix derived for deforestation probability in Pathro river basin

Probability class	High probability	Moderate probability	Low probability	Total	User’s accuracy (%)
High probability	112	18	13	143	78.32
Moderate probability	11	62	16	89	69.66
Low probability	01	07	60	68	88.23
Total	124	87	89	300	–
Overall accuracy (%)	78%				

Figure 3 show the ROC curve of the deforestation map obtained using LRM models. These curves indicate that the AUC is 76.6% which corresponds overall accuracy of 78% (Table 2), therefore, it can be said that the model applied in this study is showing reasonably good accuracy in spatial prediction of deforestation.

### 4 Result and discussion

Logistic regression model is used to determine the magnitude of correlation between deforestation locations and effective factors. The LRM value was range from 0 to 1 where, value 1 represents highly probability of deforestation. The forest maps of Pathro river basin for 1991 and 2016 are depicted in Fig. 4. Approximately 13.674% (96.948 sq km) of the total area was forest in 1991 but in the year 2016 the forest area is reduced to 8.022% (56.875 sq km). So, within 25 years 40.073 sq km area became deforested.

Based on this analysis it is revealed that the spatial pattern of deforestation is depended on a number of physiographic and anthropogenic factors and the logistic regression model (LRM) is successful to accurately predict future deforestation trend. Among these factors, slope is an important one, as the areas which have steeper slope represent more rugged terrain and less adequate for human activities. It is partly explaining that low slope forests are the most threatened forest type [28]. Finally, its represent that lower slope is more suitable for agricultural practice, another factor that leads deforestation. Roads and built-up areas are also important in determining deforestation patterns [29].

According to the logistic regression result at the selected time period the rate of deforestation is noticeably and negatively determined by slope (Wald = 6.908, Exp(B) = 1.863, df = 1), distance from roads (Wald = 5.491, Exp(B) = 1.401 df = 1), and distance from residential areas (Wald = 7.863, Exp(B) = 1.997 df = 1) and Distance from forest edge (Wald = 6.004, Exp(B) = 3.290 df = 1). So, the deforestation is not the result of single factor but it is result of a

**Table 3** Logistic Regression result

Variables	B	S.E.	Wald	Df	Sig.	Exp. (B)
Altitude	0.045	0.019	5.458	1	0.019	1.046
Slope	− 0.147	0.0001	6.908	1	0.001	1.863
Aspect	0.011	0.006	3.014	1	0.043	1.011
Distance from road	− 0.001	0.001	5.491	1	0.014	1.401
Settlement	− 0.003	0.003	7.863	1	0.013	1.997
Distance from river	− 0.006	0.004	2.083	1	0.149	0.994
Distance from forest edge	− 0.671	0.006	6.004	1	0.009	3.290
Constant	− 16.343	6.153	7.055	1	0.008	0.000

group of factors like slope, altitude, distance from the roads and distance from the settlement areas (Table 2).

Though all the factors like distance from roads, distance from the forest edge, altitude and distance from the river have control on the deforestation rate but the determining power of slope and settlements are more than these factors. Odd ratio is greater in case of slope and distance from the settlements than of other factors (Table 3).

The area under the ROC curve of the model is .766. The logistic regression model has considerable amount of accuracy and that can be used for further work [28]. Honesty of the logistic regression is always measured by the Nagelkerke  $R^2$  statistic and  $\chi^2$  value. The probability of  $\chi^2$  value = 378.293; Nagelkerke  $R^2$  = .959; ROC = .766, SE = .016, Sig = .000; Hosmer and Lemeshow test  $\chi^2$  = 7.023, Sig = .219 which are designating the good worthiness of the model in amplification the relationship between independent and dependent variables. This perfectly fit model of deforestation has used to explore the future prosperity of deforestation of the remaining areas of this watershed.

The driving factors of forest cover change may vary from one place to another. In the present study, the selected illustrative variables enclose a substantial share of the factors driving forest cover changes. Especially, the accessibility variables such as distance from settlement seem to be more important than the topographical variables. The present study reveals that the distance from settlement, distance from road, slope aspect and elevation were to be the main drivers of forest cover change in Pathro river basin.

## 5 Conclusion

Present day deforestation is burning issue not only in India but also in the rest of the world. In the present paper using some suitable parameters that are strongly related with the deforestation, probability map has been prepared with the help of logistic regression model of Pathro river basin in

GIS environment. Deforestation data has been collected through analysis of two NDVI maps of 1991 and 2016 respectively. De-forestation is in fact as interplay between several factors. Accessibility has been found to be an important variable for explaining the patterns of deforestation observed in the study area. The results indicated that distance from forest edge, settlement and slope areas have a strongly significant correlation with deforestation. This deforestation probability map is showing satisfied accuracy that has been proved by the ROC. It is the government responsibility to take care of forested areas and keeps these forested areas unaffected from the greedy economic people. This deforestation potentiality map can be used by environmental planners and managers to build up policies intended at controlling the adverse ecological and social effects of deforestation.

**Acknowledgements** The authors would like to convey cordial thanks to our respected teachers of Department of Geography, University of Gour Banga, who have always been mentally, supported us. At last, authors would like to acknowledge all of the agencies and individuals specially, Survey of India, Geological Survey of India and USGS for obtaining the maps and data required for the study.

## References

1. van der Werf, G. R., Morton, D. C., DeFries, R. S., Oliver, J. G. J., Kasibhatla, P. S., Jackson, R. B., et al. (2009). CO<sub>2</sub> emissions from forest loss. *Nature Geoscience*, 2, 737–738.
2. Eastman, J. R. (2001). *IDRISI 32 Andes guide to GIS and image processing*. Worcester: Clark University.
3. Kushwaha, S. P. S., Nandy, S., Ahmad, M., & Agarwal, R. (2011). Forest ecosystem dynamics assessment and predictive modelling in eastern Himalaya. ISPRS Archives XXXVIII-8/W20. In Workshop proceedings: Earth observation for terrestrial ecosystems, 8 November, Bhopal, India.
4. Sun, J., & Southworth, J. (2013). Remote sensing-based fractal analysis and scale dependence associated with forest fragmentation in an Amazon tri-national frontier. *Remote Sensing*, 5, 454–472.
5. Nandy, S., Kushwaha, S. P. S., & Mukhopadhyay, S. (2007). Monitoring Chilla-Motichur corridor using geospatial tools. *Journal for Nature Conservation*, 15(4), 237–244.
6. Gayen, A., & Saha, S. (2017). Application of Weights-of-evidence (WoE) and evidential belief function (EBF) models of soil



- erosion vulnerable zones: a study on Pathro river basin, Jharkhand India. *Model Earth System*, 3(3), 1123–1139.
7. Gubbi, S. (2012). Patterns and correlates of human-elephant conflict around a south India reserve. *Biological Conservation*, 148(1), 88–95.
  8. Terranova, O., Antronico, L., Coscarelli, R., & Iaquina, P. (2009). Soil erosion risk scenarios in Mediterranean environment using RUSLE and GIS: an application model for Calabria (southern Italy). *Geomorphology*, 112, 228–245.
  9. Srivastava, S., Singh, T. P., Singh, H., Kushwaha, S. P. S., & Roy, P. S. (2002). Mapping of large-scale deforestation in Sonitpur district. *Assam. CurrSc*, 82(12), 1479–1484.
  10. Rahman, M. R., & Saha, S. K. (2009). Spatial dynamics of cropland and cropping pattern change analysis using Landsat TM and IRS P6 LISS III satellite images with GIS. *Geo-Spatial Information Science*, 12(2), 123–134.
  11. Arekhi, S. (2011). Modeling spatial pattern of deforestation using GIS and logistic regression: a case study of northern Ilam forests, Ilam province Iran. *African Journal of Biotechnology*, 10(72), 16236–16249.
  12. Houet, T., & Hubert-Moy, L. (2006). Modelling and projecting land use and land cover changes with a cellular automaton in considering landscape trajectories: an improvement for simulation of plausible future states. *EARSeL eProceedings*, 5(1), 63–76.
  13. Jenerette, G. D., & Wu, J. (2001). Analysis and simulation of land use change in the central Arizona- Phoenix region USA. *Landscape Ecology*, 16(7), 611–626.
  14. Pontius, R. G., Jr., & Schneider, L. C. (2001). Land-cover change model validation by an ROC method for the Ipswich watershed, Massachusetts, USA. *Agriculture, Ecosystems and Environment*, 85(1–3), 239–248.
  15. Pontius, R. G., Jr., & Malanson, J. (2005). Comparison of the structure and accuracy of two land change models. *International Journal of Geographical Information Science*, 19(2), 243–265.
  16. Kumar, R., Nandy, S., Agarwal, R., & Kushwaha, S. P. S. (2014). Forest cover dynamics analysis and prediction modelling using logistic regression model. *Ecological Indicators*, 45, 444–455.
  17. Weier, J., & Herring, D. (2000). Measuring vegetation (NDVI & EVI). Earth observatory, August 30, NASA, <http://earthobservatory.nasa.gov/Features/MeasuringVegetation>.
  18. Eastman, J. R. (2006). *IDRISI 15 Andes guide to GIS and image processing*. Worcester: Clark University.
  19. Ludeke, A. K., Maggio, R. C., & Reid, L. M. (1990). An analysis of anthropogenic deforestation using logistic regression and GIS. *Journal of Environmental Management*, 31, 247–259.
  20. Amor, D., & Pfaff, A. (2008). Early history of the impact of road investments on deforestation in the Mayan forest. In Working paper, Nicholas School of the Environment and Sanford School of Public Policy, Duke University, Durham, NC, USA.
  21. Wilkie, D., Shaw, E., Rotberg, F., Morelli, G., & Auzels, P. (2000). Roads, development and conservation in the Congo Basin. *Conservation Biology*, 14, 1614–1622.
  22. Bayat, M. F. (2000). Surveying of the relationship between vegetation cover and some environmental variables (altitude, aspect and slope). *Pajouhesh-va-Sazandegi*, 4, 24–27.
  23. Måren, I. E., Karki, S., Prajapati, C., et al. (2015). Facing north or south: Does slope aspect impact forest stand characteristics and soil properties in a semiarid trans-Himalayan valley? *Journal of Arid Environments*, 121, 112–123.
  24. Chakravaty, S., Ghosh, S. K., Suresh, C. P., Dey, A. N., & Shukla, G. (2012). Deforestation: Causes, effects and control strategies, global perspectives on sustainable forest management, Dr. Dr. Clement A. Okia (Ed.), ISBN: 978-953-51-0569-5, In Tech, April 25, <https://doi.org/10.5772/33342>, Available from: <http://www.intechopen.com/books/globalperspectives-on-sustainable-forestmanagement/deforestation-causes-effects-and-control-strategies>.
  25. Liebetrau, A. M. (1983). Comparison of a cellular automata network and an individual-based model for the simulation of forest dynamics. *Ecological Modelling*, 121, 277–293.
  26. Rivera, S., de Martinez, A. P., Ramsey, R. D., & Cowl, T. A. (2012). Spatial modelling of tropical deforestation using socioeconomic and biophysical data. *Small-scale Forestry*, 12, 321–334.
  27. Tien, B. D., Pradhan, B., Lofman, O., Revhaug, I., & Dick, O. B. (2012). Spatial prediction of soil erosion hazards in HoaBinh province (Vietnam): a comparative assessment of the efficacy of evidential belief functions and fuzzy logic models. *CATENA*, 96, 28–40.
  28. Linkie, M., Smitha, R. J., & Leader-Williams, N. (2004). Mapping and predicting deforestation patterns in the lowlands of Sumatra. *Biodiversity Conservation*, 13, 1809–1818.
  29. Amini, M. R., Shataee, S. H., Moaieri, M. H., & Ghazanfari, H. (2009). Deforestation modeling and investigation on related physiographic and human factors using satellite images and GIS (Case study: Armardeh forests of Baneh). *Iranian Journal of Forest and Poplar Research*, 17, 431–443.