FOREST FIRE SUSCEPTIBILITY MAPPING OF BHUTAN USING LOGISTIC REGRESSION AND FREQUENCY RATIO MODEL

Hem Raj Acharya[1], Jigme Tenzin[2], Mim Prasad Phuyel[3], Karma Tshomo[4]

Department of Civil Engineering and Surveying, Jigme Namgyel Engineering College, Dewathang [1-4]

hemrajacharya337@gmail.com[1], 05190050.jnec@rub.edu.bt[2], 05190054.jnec@rub.edu.bt[3], 05190051.jnec@rub.edu.bt[4]

DOI: 10.54417/jaetm.v3i1.113

Abstract— *Forest fire is not only observed as one of the most significant sources of forest degradation in Bhutan but also a serious danger to national conservation efforts. As a result, forest fire susceptibility analysis is recognised as an important part of Bhutan's forest fire management strategy. The study's major goal is to create a forest fire susceptibility map for Bhutan using logistic regression (LR) and frequency ratio (FR) models. The study gathered the number of fire-influencing factors, evaluated them, and created susceptibility maps. Using the relative operating characteristics technique, the efficiency of each of the two models was analysed and compared to select the best model. The Receiver Operating Characteristics (ROC) curves with the area under the curve (AUC) were used to check the correctness of the maps produced by the modelling procedure. The prediction and success rates of the LR model were 88.8% and 87.5%, while for the FR model, they were 85.4% and 85.1%, respectively. The results showed that both models are good predictors of forest fire with the LR model performing fairly better than the FR model. So, the LR model was chosen as an optimum model for forest fire susceptibility mapping. The susceptibility map obtained from the optimum LR model was classified into five categories such as; very low, low, moderate, high, and very high.. The findings of this study give useful spatial information for implementing forest management techniques.*

Keywords—Forest fire susceptibility mapping, Remote sensing, Geographic Information System, Logistic regression, frequency ratio

1. INTRODUCTION

1.1. General

Bhutan has pristine forest areas covering 80.9 percent of the total land area, with trees covering 70.46 percent [1]. This has contributed to the country's status as the world's first carbon-negative country making a substantial contribution to a world endangered by climate change [2]. These abundant forest resources also contribute to Bhutan's overall development by supporting the hydropower industry, rural livelihoods, and food subsistence. As a result, the people's and country's future economies are dependent on the protection, conservation, and scientific management of forest resources [3].However, among many natural disasters, a forest fire is one of the most serious and constant risks, posing potential damage to the physical, biological, and ecological surroundings. Wildfires are expected to cause the loss of roughly 10,000 acres of forest cover per year [4].

As a result, forest fire susceptibility research is an important part of Bhutan's forest fire management plan. The study's main goal is to generate a forest fire susceptibility map for Bhutan using logistic regression (LR) and frequency ratio (FR) models. The mapping of forest fire susceptibility is a crucial step in preventing forest fire damage. Forest fire susceptibility maps (FFSMs) can help identify regions where forest fires are likely to occur [5].

1.2. Study Area

Bhutan is situated on the eastern Himalayan foothills, landlocked between two great Asian civilizations: Tibet (China) to the north and the Indian states of Assam, Arunachal Pradesh, Sikkim, and West Bengal to the east, west, and south. Bhutan has a land area of 38,394 with a latitude of 27°30'N and a longitude of 90°30' E. The topography rises from the southern foothills at 200 metres above sea level to the northern mountains at 7000 meters.

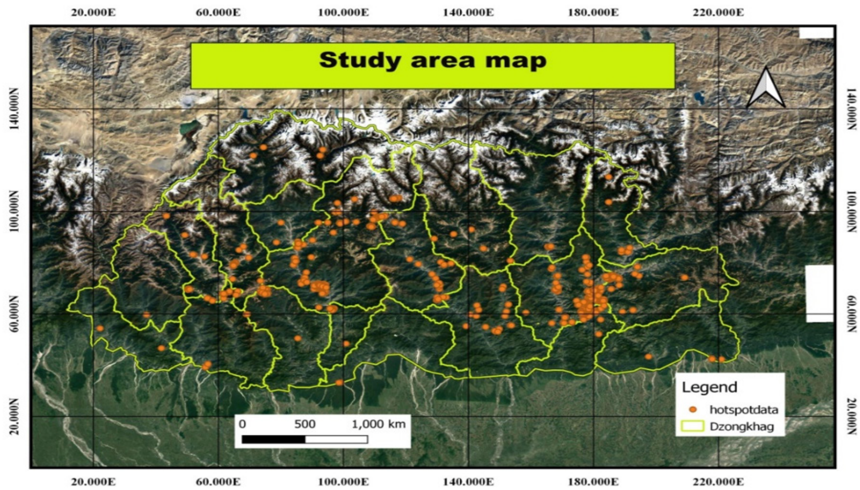

Fig. 1. Study area

2. MATERIALS AND METHODS

2.1. Data and Source

TABLE 1. Basic remote sensing and GIS input data for forest fire susceptibility analysis

Sl.NO	Data	Data format	Resolution	Source
1	NOAA hotspot data	Vector	1 Km	NASA FIRMS (LANCE)
2	SRTM DEM	Raster	30m	USGS Earth Explorer
3	LULC map	Vector	10m	Ministry of Agriculture and Forest (MoAF)
4	Population data	Vector	NA	National Statistical Bureau (PHCB – 2016)
5	Settlement data	Vector	NA	National Land Commission (NLCS), Bhutan
6	River & Road data	Vector	NA	National Land Commission (NLCS), Bhutan
7	Rainfall data	Vector	NA	CRU (Climate Research Unit)

2.1.1. Dependent variable (Hotspot)

The dependent variable in the analysis is the forest fire inventory map, which reflects the spatial location of forest fire points. However, spatial data on forest fire incidents in the research area is either unavailable or non-existent. The spatial locations of forest fire hotspots was received from NASA Fire Information for Resource Management System (FIRMS) for four years (2019 January -2022 March) via E-mail (https://firms.modaps.eosdis.nasa.gov) from NOAA VIIRS satellites. The data from the collected hotspot were re-projected into a standard coordinate system. The hotspots were then converted to a raster format with a 30 m cell size as the dependent variable. The study region has a total of 695 hotspot pixels which was used in the analysis.

2.1.2. Independent variables

Twelve important forest fire factors were retrieved from the input databases in this study, with three major categories: environmental, climatic, and anthropogenic variables. Environmental parameters include topographic features. Topography is one of the main factors applied in any fire hazard rating system because it characterizes the landscape features and it is strongly recommended in forest fire studies [6]. Using surface analysis tool in the ESRI ArcGIS program, topographic variables such as elevation, slope, and aspect were extracted from the 30 m resolution of the STRM digital elevation model (DEM). Similarly, using the hydrological tool in the ESRI ArcGIS software, the topographic wetness index (TWI) that represents moisture content was calculated from the DEM.

The enhanced vegetation index (EVI) was generated from Landsat 8 OLI employing 2, 4, and 5 bands. Land use data from the Bhutan Land Cover Assessment 2016 (LCMP-2016) were categorized into nine categories: coniferous forest, water bodies, shrubs and meadows, agriculture, non-built up, rock outcrops, built up, broad leaf, snow, and glacier.

Climatic conditions are known to affect fuel accumulation and the moisture content (Syphard et al., 2008).They play a major influence in generating fire-prone environments by determining the type of vegetation in a region. The inverse distance weighted (IDW) interpolation technique was used to generate rainfall map from Climate Research Unit (CRU) data. Land surface temperature (LST) data was generated from Landsat OLI imagery using a 100 meter resolution composite of thermal bands (10 and 11).

Human accessibility to regions where fires can occur is represented by the proximity variables. Because of habitation/cultural practices, forest near roads, communities, and agricultural area is more prone to fire. Euclidean distance tools in the ERSI ArcGIS software ws used to produce proximity variables such as distance to roads, rivers and settlement.

Population density data was obtained from The Population and Housing Census of Bhutan (PHCB) 2016, from the National Statistics Bureau. The GRASS tool in the ERSI ArcGIS program was used to generate missing rivers and streams from the SRTM.

2.2. Methodology

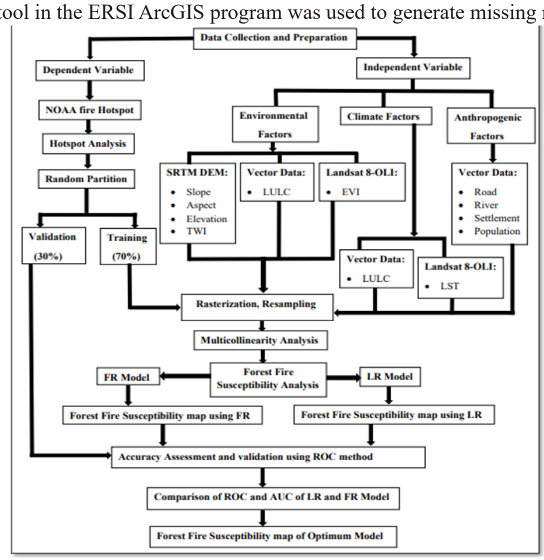

Fig 2: Methodology flow chart

3. METHOD AND METHODOLOGY

3.1. Frequency Ratio model

The association between hotspot locations and the factors in the study region is revealed using frequency ratio techniques, which are based on observed relationships between hotspot distribution and each hotspot-related factor [7]. The spatial correlations between hotspot-occurrence location and each component contributing to hotspot occurrence were obtained using the frequency ratio model. The frequency is determined by looking at the relationship between the hotspot and the influential factors[7]. As a result of their link with hotspot events, the frequency ratios of each factor's kind or range were calculated.

Thematic maps of all 12 forest fire influencing factors were classified for the FR analysis based on the purpose, data accuracy and scale, and literature reviews. The frequency ratio (FR) of each element is calculated in the first step of the FR analysis using the equation below [8]:

$$FR = \frac{\%\text{ of hotspot pixel }(A)}{\%\text{ of pixel in each class }(B)} \qquad (1)$$

The next stage in the FR analysis is to use the reclassification feature of the spatial analyst tool to assign those computed FR values to each class of factors. Finally, using the equation below, all of the factor maps with assigned FR values were combined to produce a forest fire susceptibility index (FSI) map.

$$FSI = FR_1 + FR_2 + FR_3 + \ldots\ldots\ldots\ldots\ldots FR_n \qquad (2)$$

Where FSI represents the forest fire susceptibility index; it indicates the relative susceptibility to forest fire occurrence, where higher values are associated with high susceptibility and lower values represent low susceptibility [7]. FSI represents the weighted factor maps of forest fire influential factors.

3.2. Logistic Regression

The independent variables were used to build a formula that was used to calculate the chance of forest fire in a given area. This approach requires dependent data consisting of values of 0 and 1, which indicate the absence and presence of disasters, respectively. The logistic regression analysis was carried out using SPSS software.

The coefficients were measured and are listed in Table 5. The higher the logistic coefficient, the greater the expected impact on forest fire occurrence. The probability (p) of fire occurrence was computed using the derived logistic coefficients as follows:

$$p = 1/\left(1 + e^{-z}\right) \qquad (3)$$

Where, p is the probability of fire. Z is the linear combination and it follows that logistic regression involves fitting an equation of the following form to the data:

$$Z = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_nx_n \qquad (4)$$

Where b0 is the intercept of the model, b^i (i = 0, 1, 2… n) represents the coefficients of the logistic regression model, and x_i (i = 0, 1, 2… n) denotes the independent variables [8].

3.2.1. Multicollinearity Analysis

Logistic regression is sensitive to collinearity among the factors and it is important to check the collinearity among the factors [9]. It is important to examine the collinearity among the independent factors. Multicollinearity is caused by the high correlation between the independent factors [10]. The factors were considered non-collinear if TOL is more 0.1 and VIF is less than 5 [9] The tolerance (TOL) and variance inflation factor (VIF) were commonly used to check multi-collinearity. The TOL and VIF are calculated using Equations 5 and 6.

$$TOL = 1 - r^2 \qquad (5)$$

$$VIF = \frac{1}{1 - r^2} \qquad (6)$$

Where, TOL is tolerance and VIF is variance inflation factor.

4. RESULT AND VALIDATION

4.1. Frequency Ratio model

TABLE 2. Basic Remote Sensing and GIS input data for forest fire susceptibility analysis

Factor	Class	No of pixel in each class	% of pixel in each class(B)	No of hotstop pixe	% of hotspot pixel(A)	FR =(A/B)
	0 - 16 Degree	7261155	15.08990752	52	1.236623068	0.08195034
	16 - 32 Degree	21265001	44.19226671	193	4.589774078	0.10385921
Slope	32 - 49 Degree	13023006	27.06400788	216	5.136741974	0.18979975
	49 - 65 Degree	1490572	3.097660582	24	0.570749108	0.18425166
	65 - 81 Degree	39089	0.08123355	1	0.023781213	0.29275112
Total		43078823	89.52507624	486	11.55766944	0.85261209
Aspect	(-1) - 71 Degree	7676679	15.96292031	61	12.55144033	0.78628723
	71 - 143 Degree	8879960	18.46502815	116	23.86831276	1.2926226
	143 - 216 Degree	9586822	19.93488012	139	28.60082305	1.43471257
	216 - 288 Degree	8811846	18.3233916	106	21.81069959	1.19032001
	288 - 360 Degree	8123516	16.89207515	64	13.16872428	0.77958002
Total		43078823	89.57829533	486	100	5.48352242

Elevation	76 - 1557 m	9991986	20.74340938	288	6.835983859	0.32954968
	1557 - 3039 m	15085219	31.31698476	129	3.061951104	0.09777286
	3039 - 4520 m	12717084	26.40072549	69	1.6377878	0.06203571
	4520 - 6002 m	5194874	10.78458257	0	0	0
	6002 - 7483 m	89660	0.186134577	0	0	0
Total		43078823	89.43183676	486	11.53572276	0.48935825
Distance from settlement	0 - 1000 m	12659427	99.31793888	283	11.75249169	0.11833201
	1000 - 2000	6648058	52.15650109	119	4.941860465	0.09475061
	2000 - 3000	4609689	36.16473402	19	0.789036545	0.02181784
	3000 - 4000 m	3469433	27.2189993	17	0.705980066	0.02593703
	4000 m<	15692287	123.1118597	48	1.993355482	0.01619142
Total		43078894	337.970033	486	20.18272425	0.27702892
Distance from river	4000 m<	13243945	150.5957734	43	2.584134615	0.01715941
	3000 - 4000 m	5978660	67.98283492	35	2.103365385	0.03093965
	2000 - 3000 m	7144962	81.24475588	59	3.545673077	0.04364187
	1000 - 2000 m	7916976	90.02326148	139	8.353365385	0.09279119
	0 - 1000 m	8794351	99.99981807	210	12.62019231	0.12620215
Total		43078894	489.8464438	486	29.20673077	0.31073427
Distance from road	0 - 1000 m	9358656	99.04080654	209	10.80103359	0.1090564
	1000 - 2000	4971451	52.61188324	85	4.392764858	0.08349378
	2000 - 3000	3836533	40.60127038	55	2.842377261	0.0700071
	3000 - 4000 m	3178476	33.63718323	43	2.222222222	0.06606446
	4000 m<	21733778	230.0042765	94	4.857881137	0.02112083
Total		43078894	455.8954199	486	25.11627907	0.34974256

Factor	Class	No of pixel in each class	% of pixel in each class (B)	No of hotstop pixe	% of hotspot pixel(A)	FR =(A/B)
Rainfall	1084 - 1355 mm	2339555	5.426783967	0	0	0
	1355 - 1625 mm	10279956	23.84517586	8	0.189888441	0.00796339
	1625 - 1895 mm	12141427	28.16300595	136	3.228103489	0.11462212
	1895 - 2165 mm	14715563	34.13391921	337	7.999050558	0.23434316
	2165 - 2435 mm	3602360	8.355960639	5	0.118680275	0.01420307
Total		43078861	99.92484562	486	11.53572276	0.37113174
LULC	Coniferous Forest	16352258	37.95886922	343	70.57613169	1.85927909
	Non Built up	282369	0.655469596	4	0.823045267	1.25565743
	Shrubs and Meadows	1080160	2.507400029	10	2.057613169	0.82061623
	Agriculture	1186041	2.753184008	28	5.761316872	2.09260146
	Water Bodies	6666	0.015473938	0	0	0
	Rock outcrops	1830957	4.250242218	4	0.823045267	0.19364667
	Built up	83575	0.194004552	2	0.411522634	2.12120091
	Broadleaf	19791117	45.94157101	95	19.5473251	0.4254823
	Snow and Glacier	2465743	5.723785429	0	0	0
Total		43078886	100	486	100	8.76848409
Polpulation Density	3116 - 22228	23096642	53.61491213	119	24.48559671	0.45669378
	22228 - 41340	10160120	23.58498439	273	56.17283951	2.38172044
	41340 - 60452	5737960	13.31969475	64	13.16872428	0.98866562
	60452 - 79564	2088625	4.848386438	12	2.469135802	0.5092696
	79564 - 98676	1995418	4.632022297	18	3.703703704	0.79958676
Total		43078765	100	486	100	5.1359362
TWI	(1) - 5	18166399	42.17013775	241	49.58847737	1.17591452
	(5) - 9	21800252	50.60549588	211	43.41563786	0.85792338
	(9) - 12	2857315	6.632760138	31	6.378600823	0.96168121
	(12) - 16	252996	0.58728624	3	0.617283951	1.05107852
	(16) - 19	1861	0.004319988	0	0	0
Total		43078823	100	486	100	4.04659763
LST	(-24) - -12 Degree Celsius	77264	0.179354947	0	0	
	(-12) -1 Degree Celsius	2732888	6.343924485	1	0.205761317	0.032434
	1 - 13 Degree Celsius	29232236	67.85755498	190	39.09465021	0.576128
	13 - 26 Degree Celsius	10962755	25.44813028	290	59.67078189	2.344800
	26 - 39 Degree Celsius	73712	0.171109596	5	1.028806584	6.012559
Total		43078855	100.0000743	486	100	8.965922
EVI	(-1) - -0.6	4535503	10.52838189	21	4.320987654	0.410413
	(-0.6) - -0.2	23554515	54.67771253	378	77.77777778	1.42247
	(-0.2) - 0.2	14571181	33.82446405	85	17.48971193	0.517072
	0.2 - 0.6	197303	0.458004621	1	0.205761317	0.449255
	0.6 - 1	39686	0.092124151	0	0	
Total		42898188	99.58068724	485	99.79423868	2.799218

TABLE 3. Prediction rates of wildfire influential factors using FR model

Sl.No	Influential factors	Prediction Rate
1	Slope	2.069
2	Aspect	1
3	Elevation	5.637
4	Distance from settlement	3.086
5	Distance from river	2.937
6	Distance from road	2.104
7	Rainfall	5.285
8	LULC	2.025
9	Population density	3.137
10	TWI	2.432
11	LST	5.613
12	EVI	4.253

4.2. Logistic Regression model

4.2.1. Multicollinearity diagonisis

TABLE 4.The value of tolerance (TOL) and Variance Inflation Factor (VIF)

Factors		Collinearity Statistics	
Sl.No	Factors	Tolerance (TOL)	(Variance Inflation Factor) VIF
1	EVI	0.955	1.047
2	Population	0.892	1.122
3	Road	0.455	2.198
4	Rainfall	0.292	3.423
5	LULC	0.883	1.132
6	LST	0.622	1.607
7	Elevation	0.218	4.577
8	Aspect	0.908	1.101
9	Settlement	0.39	2.564
10	Slope	0.845	1.184
11	TWI	0.874	1.144
12	River	0.618	1.618

TABLE 5: Coefficients of LR model and statistics

Factors	B	Exp(B)
EVI	0.001	1
Population	-0.15	0.861
Distance from Road	-0.083	0.92
Distance from River	0.167	1.182
Rainfall	0.311	1.364
LULC	-0.412	0.662
LST	0.743	2.103
Elevation	-0.33	0.719
Aspect	0.032	1.032
Distance from Settlement	-0.435	0.647
Slope	0.429	1.536
TWI	-0.025	0.975
Constant	-1.222	0.295

All of the independent variables were found to be free of charge which means that independent factors are not collinear and all the factors were used for study.

Where B = logistic coefficient Exp (B) = Exponentiated coefficient.

The logistic coefficient coefficients from Table 5 were used to create a linear combination, which was then used to determine the forest fire probability P (Equation 3) and create a forest fire susceptibility map.

4.3. Validation

The accuracy assessment is an important step to check the reliability and efficiency of the map [10]. The accuracy assessment of the generated forest fire susceptibility map was done using the Area under the Curve (AUC) of the Receiver Operating Characteristics (ROC) curve. The AUC is calculated using Equation 9. The ROC curve is constructed False Positive Rate (FPR) on the x-axis and True Positive Rate (TPR) on the y-axis [11]. The AUC is interpreted as excellent (0.9-1.0), very good (0.8-0.9), good (0.7-0.8), moderate (0.6-0.7), and poor (0.5-0.6) [9]. The sensitivity and specificity are calculated using the equation 7 and 8.

$$TPR = \frac{TP}{TP + FN} \qquad (7)$$

$$FPR = \frac{FP}{FP + TN} \qquad (8)$$

$$AUC = \frac{\sum TP + \sum TN}{TP + TN + FP + FN} \qquad (9)$$

TP (true positive) and TN (true negative) are the number of samples that are correctly classified as positive (fire class) and negative (non- fire class) observations respectively. FP (false positive) and FN (false negative) are the number of samples that are misclassified. Sensitivity is the percentage of positive (fire class) observations that are correctly classified whereas specificity is the percentage of negative (non- fire class) observations that are correctly identified.

The ROC technique was used to test the accuracy of the forest fire probability maps derived from the LR and FR models based on an independent validation dataset (30 percent). For the final forest fire susceptibility mapping, the model with the best ROC value was chosen. The ROC checks the model's prediction performance to determine whether it is fit or not. The better the model, the greater the ROC value. The ROC's value ranges from 0.5 to 1. A perfect fit is shown by a ROC value of 1, while a random fit is indicated by a ROC value of 0.5.

The forest fire susceptibility map using Logistic Ratio has a success rate of 0.875, which is in the very good category, while the forest fire prediction rate is 0.888, which also falls under very good category. Upon considering the AUC value the logistic regression model outperformed the FR model. Therefore, LR model was used to create final forest fire susceptibility map of Bhutan.

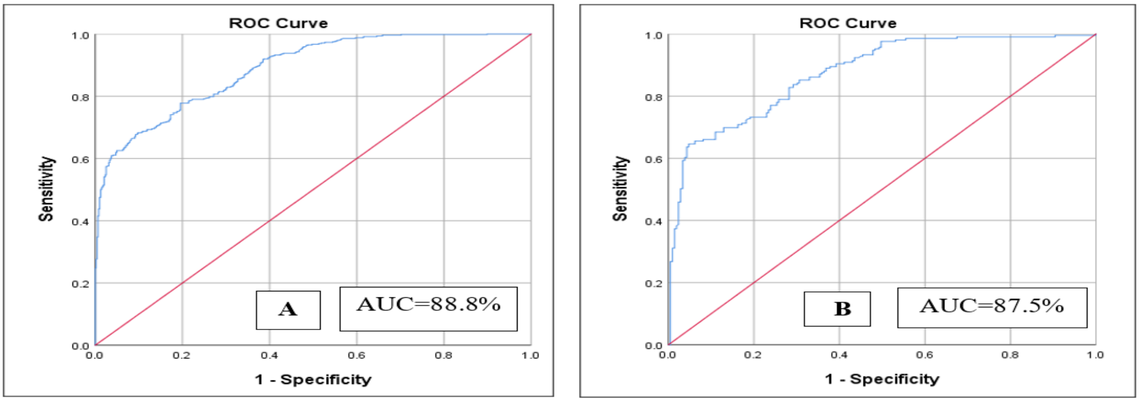

Fig. 5. Forest fire susceptibility mapping of Bhutan (Using optimum model)

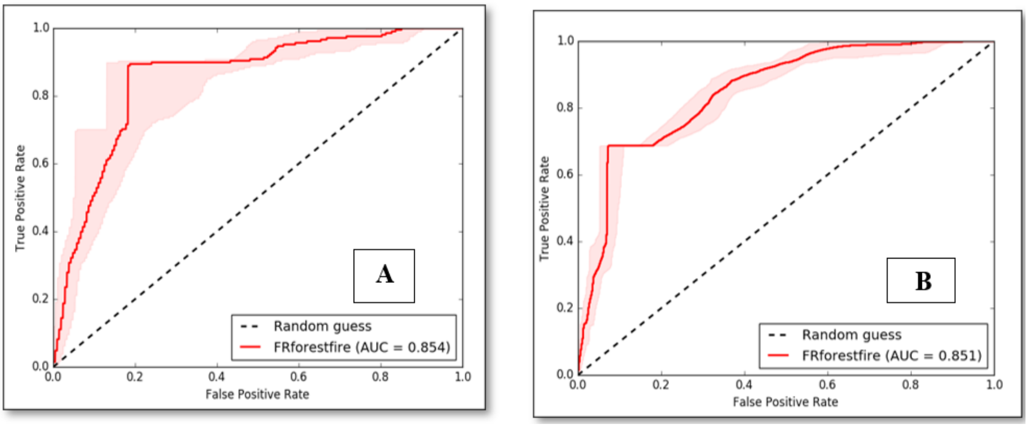

Fig. 4. ROC Curve for Frequency Ratio (A) Success rate curve, (B) Prediction rate curve

5. CONCLUSION

Modeling a forest fire at the regional scale is difficult due to its nonlinear and complex nature. Two easy and strong models for predicting forest fire-prone areas in Bhutan were compared in this study. AUC values showed that the LR model produced a more accurate fire susceptibility map than the FR model. As a result, the LR model is suggested. The usage of LR and FR models can efficiently determine the most important influencing elements of forest fire occurrences and, ultimately, construct the forest fire susceptibility map. The findings may provide useful information to guide and assist Bhutan's successful forest fire management strategy. Furthermore, the methods used in this study may have the potential to be applied in other parts of Bhutan with similar environmental, climatic, and anthropogenic influences. Despite the fact that the fraction of extremely high and high susceptibility zones is lower than other zones, the generated map is consistent with the real fire scenario in the research area. The maps created with the FR and LR maps are extremely relevant and topical for Bhutan's current situation, as the country currently lacks an officially recognized forest fire-prone area map. In Bhutan's center areas, fire-prone classes such as 'very high' and 'high' are expected. Wangdue phodrang, Mongar, Thimphu and Paro have the highest percentage of fire-prone areas, followed by Bumthang and Trashigang. These are locations with dense human populations and a well-developed road system. Therefore, any future fire prevention and control programs in Bhutan must focus these regions.

REFERENCE

[1] NSB, Agriculture survey report 2021. 2021. [Online]. Available: http://www.nsb.gov.bt

[2] “Energy & Climate Intelligence Unit | Insights.” https://eciu.net/insights/p20 (accessed Nov. 19, 2022).

[3] P. Bergen, “Keeping promises,” New Repub., vol. 241, no. 13, pp. 6–9, 2010, doi: 10.1017/9781108386104.009.

[4] D. Gyelmo, “Forest fires burning up Bhutan’s wilderness | The Third Pole.” https://www.thethirdpole.net/en/climate/forest-fires-burning-up-bhutans-wilderness/ (accessed Nov. 19, 2022).

[5] R. K. Jaiswal, S. Mukherjee, K. D. Raju, and R. Saxena, “Forest fire risk zone mapping from satellite imagery and GIS,” Int. J. Appl. Earth Obs. Geoinf., vol. 4, no. 1, pp. 1–10, 2002, doi: 10.1016/S0303-2434(02)00006-5.

[6] H. K. Preisler, D. R. Brillinger, R. E. Burgan, and J. W. Benoit, “Probability based models for estimation of wildfire risk,” Int. J.

Wildl. Fire, vol. 13, no. 2, pp. 133–142, 2004, doi: 10.1071/WF02061.

[7] S. Dorji and S. Ongsomwang, "Wildfire Susceptibility Mapping in Bhutan Using Geoinformatics Technology," Suranaree J. Sci. Technol., vol. 24, no. 2, pp. 213–237, 2017.

[8] S. Lee and B. Pradhan, "Landslide hazard mapping at Selangor, Malaysia using frequency ratio and logistic regression models," Landslides, vol. 4, no. 1, pp. 33–41, 2007, doi: 10.1007/s10346-006-0047-y.

[9] C. Vansarochana and V. Sharma, "Flood Risk Area Mapping with Logistic Regression: A Case Study of Phuntsholing City in Bhutan," Conf. Pap., no. January, 2022, [Online]. Available: https://www.researchgate.net/publication/358233652

[10] M. S. Tehrany, S. Jones, and F. Shabani, "Identifying the essential flood conditioning factors for flood prone area mapping using machine learning techniques," Catena, vol. 175, no. December, pp. 174–192, 2019, doi: 10.1016/j.catena.2018.12.011.

[11] T. Thongley and C. Vansarochana, "Spatial zonation of landslide prone area using information value in the geologically fragile region of samdrup jongkhar-tashigang national highway in bhutan," Environ. Nat. Resour. J., vol. 19, no. 2, pp. 122–131, 2021, doi: 10.32526/ennrj/19/2020171.