Received: 19 July 2022 | Revised: 28 October 2022 | Accepted: 16 November 2022

DOI: 10.1111/efp.12783

Forest Pathology WILEY-BLACKWELL WILEY

ORIGINAL ARTICLE

Prediction of pine mistletoe infection using remote sensing imaging: A comparison of the artificial neural network model and logistic regression model

Ayhan Usta | **Murat Yilmaz**

Department of Forest Engineering (former members), Faculty of Forestry, Karadeniz Technical University, Trabzon, Turkey

Correspondence
Ayhan Usta, Department of Forest Engineering (former member), Faculty of Forestry, Karadeniz Technical University, Trabzon 61080, Turkey.
Email: ayhanusta@hotmail.com

Funding information
The Scientific and Technological Research Council of Turkey, Grant/Award Number: TOVAG-112O258

Editor: S. Woodward

Abstract

In this study, the prediction of pine mistletoe distribution in Scots pine ecosystems was explored using remote sensing variables to compare the multilayer perceptron (MLP) artificial neural network (ANN) and logistic regression (LR) model performances. For this purpose, 109 sample plots were distinguished in pure Scots pine forests (natural) in the Eastern Black Sea Region of Turkey. Distinguishing mistletoe-infected stands (69) and uninfected stands (40) was performed with field observations. The variables acquired from Landsat 8 (Level 1) images were used as independent variables for independent-sample *t*-test, MLP ANN and LR models. Remote sensing variables indicated that mistletoe-infected stands were in drier areas with a lower vegetation-leaf area index. Based on the performance results of both models, the sensitivity (SEN), specificity (SPE), positive predictive value (PPV), negative predictive value (NPV) and accuracy of the MLP ANN model were superior to those of the LR model. The prediction percentages (SEN, SPE, PPV and NPV) of mistletoe-infected stands were better than the prediction percentages of uninfected stands. The prediction accuracies of LR and MLP ANN models were 74.3% and 89.6%, respectively. However, all remote sensing variables were included in the prediction equation of the MLP ANN model, while the thermal infrared 1 (TIRS1) variable was included in the LR model. In the MLP ANN model, the TIRS1 variable also had the highest normalized importance (100%). The area under the curve (AUC) value for identifying the mistletoe-infected stands of Scots pine forests used by the MLP ANN model (0.892 ± 0.034) was higher than in the LR model (0.838 ± 0.039), explaining the more accurate predictions obtained from the MLP ANN model. The MLP ANN model showed much better performance than the LR model. The results of this study are expected to make important contributions to the identification of potential mistletoe-infected areas.

KEYWORDS
forest healthy, independent-samples *t*-test, Landsat 8, Scots pine forests, the predicted pseudo-probability

1 | INTRODUCTION

Undoubtedly, pines are among the most important tree species in the world, distributed from coastal to mountainous areas and alpine areas between the subtropics and subarctic latitudes of the northern hemisphere (Richardson et al., 2007). Scots pine (*Pinus sylvestris* L.) has the widest geographical distribution of all species in the genus, occurring in large mixed and pure forests, and its timber is highly valued commercially (Alemdağ, 1967). Scots pine is also a host of European pine mistletoe (*Viscum album* subsp. austriacum (Wiesb.) Vollm.) within the distribution area (Zuber, 2004). Dobbertin and Rigling (2006) reported that the mortality rate of mistletoe-infected pine was more than double that of uninfected trees, and mistletoe-containing trees were among the most stressed individuals.

Pine mistletoe renders affected trees more susceptible to drought stress, increasing needle losses and the risk of drought-related mortality (Rigling et al., 2010). Mistletoe infestation does not rapidly result in tree mortality; the infections in a host amplify the effects of drought and greatly reduce the capacity to absorb carbon in xeric conditions (Zweifel et al., 2012). Moreover, trees severely infected with mistletoe are more susceptible to attacks by other damaging agents, such as insects or decay fungi, which often cause further decline of the mistletoe-infected tree (Hawksworth & Wiens, 1996). The pathological effects of mistletoe vary considerably, depending on their capacity to obtain water, minerals and carbon from their hosts (Ehleringer & Schulze, 1985). As with other pathogenic agents, the effects of mistletoe are highly influenced by the habitat conditions in which the hosts grow and the size, age and density of infected plants (Knutson, 1983).

Globally, forest health is threatened by climate change, increasing pollution, pest and pathogen activities, many of which are associated with varied human activities. Assessment and monitoring of forest health play a critical role in sustainable forest management. Considering that threats to forest healthy are increasing globally (Lausch et al., 2016), the ability to detect ecosystem deterioration due to biotic, abiotic and anthropogenic factors is extremely important.

Remote sensing systems provide rapid, spatially defined and practical paths to monitor and assess ecosystem health in comparison with field studies. In determining and monitoring stress in the ecosystem, spectral (multiple, hyper) sensors are valuable in detecting information outside the visible spectrum, allowing for the prediction of plant biochemical characteristics, such as leaf pigment contents and canopy water content (Stone & Mohammed, 2017). Remote sensing images offer a spatially objective approach and are often more economical and effective than fieldwork processes applied on a large scale. Thus, remote sensing images provide more convenient information for assessing and monitoring environmental conditions related to the biophysical and biochemical characteristics of plant species at a landscape scale (He et al., 2019; Stone & Mohammed, 2017; Zhang et al., 2019).

In predictive studies, logistic regression is a widely used statistical modelling technique that expresses the probability of an outcome occurring based on many potential predictors. However, logistic regression models have some limitations and statistically more standardized training for the development of the model, and they cannot determine all probable relationships between independent attributes. In comparison with the logistic regression techniques, artificial neural networks can overcome some of these restrictions (Shahmoradi et al., 2018).

The aim of the work described here was to establish an artificial neural network (ANN) and logistic regression model to predict the incidence of pine mistletoe infections using remote sensing images. In the structure of the ANN model, various configurations were evaluated and the model's overall performance improved by altering the number of hidden layers and neurons to obtain a suitable model for pine mistletoe infection. During the selection of the remote sensing images evaluated in the work, the dry season where the field studies were carried out was considered.

2 | MATERIAL AND METHODS

2.1 | Study area

The study area was located between 38°53′–39°56′E longitudes and 39°50′–40°45′N latitudes within the boundaries of Gümüşhane province in north-eastern Turkey (Figure 1). The study was performed in pure Scots pine forests on a total area of 5718.8 km^2. The altitude in the study area was between 1003 and 2228 m.

The location of the study area, due to the Turkish Black Sea mountains running parallel to the coast, cannot gain from the humid and rainy climate of the coast itself. The humid and rainy air masses of the Black Sea coast, which are carried inland along the Harşit and Kelkit Streams in the region close to the Torul district of Gümüşhane, only partially affect the study area.

The annual average temperature in the study area is 9.7°C, and total annual precipitation is 462.1 mm (Table 1). In the study area, together with Scots pine (*Pinus sylvestris* L.), Nordmann fir (*Abies nordmandiana*), Oak (*Quercus* spp.) and Juniper (*Juniperus* spp.) are also distributed (Usta & Yilmaz, 2021).

2.2 | Data set

Circular sampling points of 1000 m^2, based on a 2 × 2 km grid system, were established in the study area. Open space areas and Scots pine with a diameter at breast height ≤20 cm were not sampled (Bilgili et al., 2020). Mistletoe-infected and uninfected trees were recorded in the field.

Landsat-8 (Level 1) obtained from the United States Geological Survey (USGS) Earth Explorer (USGS, 2018) was used for the work. The dates of remote sensing images used were selected as 14 June

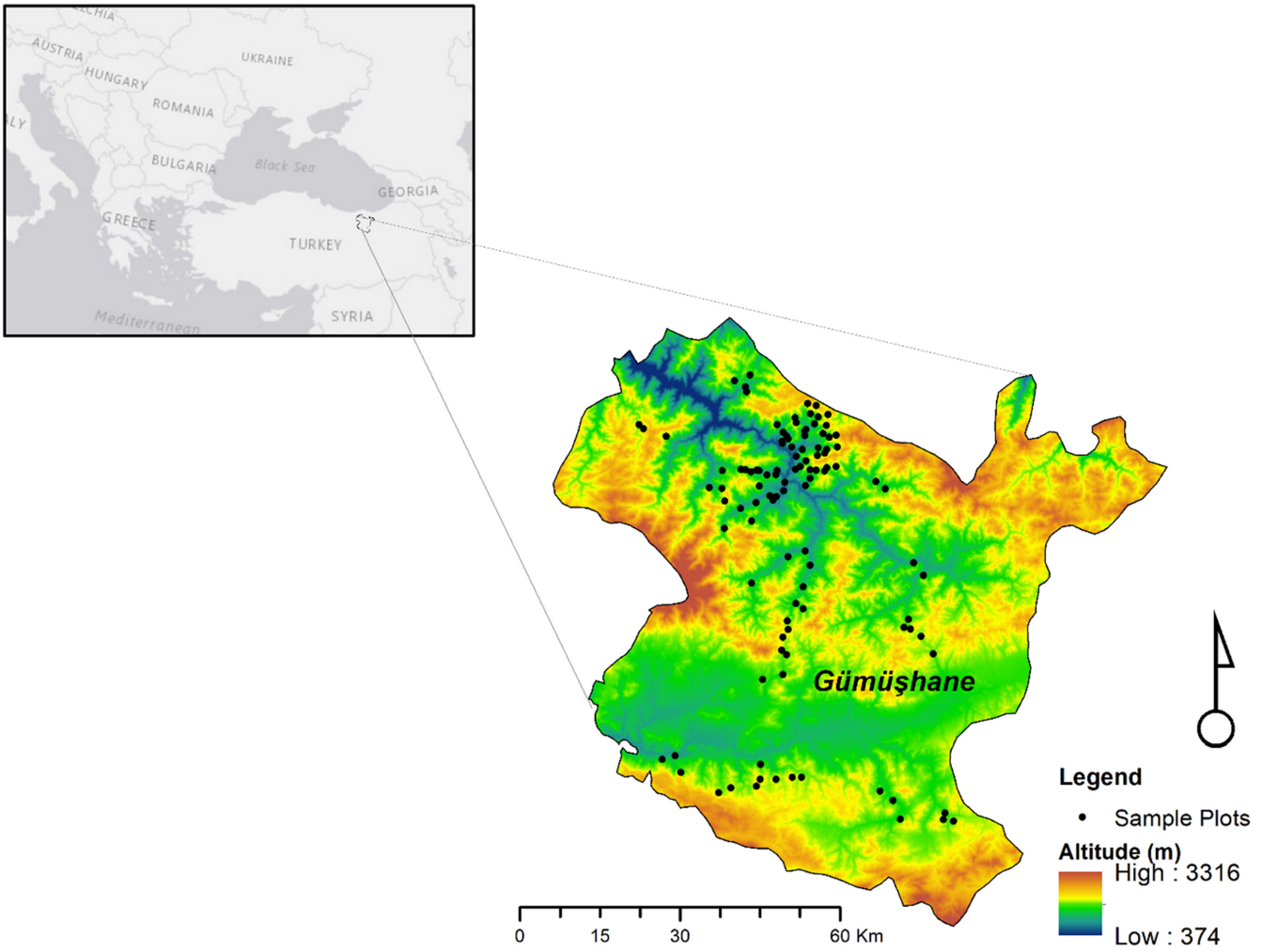

FIGURE 1 Location of the study area

2014, as the fieldwork was carried out between 2012 and 2014. The remote sensing variables derived from Landsat images were: enhanced vegetation index (EVI), soil-adjusted vegetation index (SAVI), transformed soil-adjusted vegetation index (TSAVI), normalized difference moisture index (NDMI), modification of normalized difference water index (MNDWI), moisture stress index (MSI), soil moisture index (SMI), thermal infrared 1 (TIRS1), thermal infrared 2 (TIRS2), normalized difference vegetation index (NDVI), short-wave infrared 1 (SWIR1) and short-wave infrared 2 (SWIR2) (Table 2; Figure 2).

Remote sensing input variables were calculated, according to the USGS (2017), as shown in Table 2. SWIR1, SWIR2, TIRS1 and TIRS2 variables consist of 6, 7, 10 and 11 bands of Landsat-8, without the calculation, respectively. The formulae used were:

$$\mathrm{EVI} = 2.5 \times \frac{(\mathrm{Band\ 5} - \mathrm{Band\ 4})}{(\mathrm{Band\ 5} + 6.0 \times \mathrm{Band\ 4} - 7.5 \times \mathrm{Band\ 2} + 1.0)} \quad (1)$$

$$\mathrm{SAVI} = \frac{\mathrm{Band\ 5} - \mathrm{Band\ 4}}{\mathrm{Band\ 5} - \mathrm{Band\ 4} + 0.5} \times 1.5 \quad (2)$$

$$\mathrm{TSAVI} = \frac{a(\mathrm{Band\ 5} - a\mathrm{Band\ 4} - b)}{\mathrm{Band\ 4} + a(\mathrm{Band\ 5} - a) + 0.08(1 + a^2)} \quad (3)$$

where a: the slope (the soil line), b: y-intercept a (Baret et al., 1989).

$$\mathrm{NDMI} = \frac{\mathrm{Band\ 5} - \mathrm{Band\ 6}}{\mathrm{Band\ 5} + \mathrm{Band\ 6}} \quad (4)$$

$$\mathrm{MNDWI} = \frac{\mathrm{Band\ 3} - \mathrm{Band\ 6}}{\mathrm{Band\ 3} + \mathrm{Band\ 6}} \quad (5)$$

$$\mathrm{MSI} = \frac{\mathrm{Band\ 6}}{\mathrm{Band\ 5}} \quad (6)$$

$$\mathrm{NDVI} = \frac{\mathrm{Band\ 5} - \mathrm{Band\ 4}}{\mathrm{Band\ 5} + \mathrm{Band\ 4}} \quad (7)$$

$$\mathrm{SMI} = \frac{T_{\mathrm{smax}} - T_{\mathrm{s}}}{T_{\mathrm{smax}} - T_{\mathrm{smin}}} \quad (8)$$

TABLE 1 Gümüşhane meteorology station (TSMS, 2020) climate data

Climate parameters	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	Annual
	Measurement period (1961–2019)												
Temperature, °C	−1.7	−0.4	3.8	9.4	13.7	17.2	20.2	20.3	16.7	11.4	5.1	0.5	9.7
Max. temperature, °C	2.8	5.0	9.8	16.1	21.0	24.9	28.1	28.8	25.3	18.6	10.5	4.7	16.3
Min. temperature, °C	−5.6	−5.0	−1.1	3.7	7.6	10.6	13.7	13.8	10.0	5.8	0.7	−3.1	4.3
Sunshine duration (h)	1.3	3.7	5.0	6.0	7.4	9.0	10.0	9.7	7.9	5.4	2.2	0.8	68.4
Number of rainy days	11.0	10.7	12.6	13.8	15.6	10.3	4.0	3.6	5.4	9.5	10.1	11.5	118.1
Monthly total precipitation (mm)	36.2	32.3	43.5	60.4	68.2	46.8	12.1	12.9	21.7	45.1	41.9	41.0	462.1

TABLE 2 Remote sensing variables used for the prediction of mistletoe infection

Input variables (12)	Acronym	Data
Enhanced vegetation index[a]	EVI	Band 2, 4, 5
Soil-adjusted vegetation index[a]	SAVI	Band 4, 5
Transformed soil-adjusted vegetation index[a]	TSAVI	Band 4, 5
Normalized difference moisture index[b]	NDMI	Band 5, 6
Modification of normalized difference water index[b]	MNDWI	Band 3, 6
Moisture stress index[b]	MSI	Band 5, 6
Soil moisture index[b]	SMI	Band 4, 5, 10, 11
Thermal infrared 1[c]	TIRS1	Band 10
Thermal infrared 2[c]	TIRS2	Band 11
Short-wave infrared 1[b]	SWIR1	Band 6
Short-wave infrared 2[b]	SWIR2	Band 7
Normalized difference vegetation index[a]	NDVI	Band 4, 5

[a]These variable limits soil moisture loss.
[b]This variable is linked to surface water.
[c]This variable is linked to the soil moisture evaporation.

where T_{smax} is the maximum surface temperature for normalized difference vegetation index (NDVI), and T_{smin} is the minimum surface temperature for NDVI.

2.3 | Artificial neural network analysis (ANN)

Artificial neural network is an information processing technology inspired by biological nervous systems (Usta et al., 2018). ANN also generates nerve cells related to each other, as in biological neural networks. In a prediction issue that can be solved by ANN and is based on a cause-and-effect relationship, inputs are generally independent variables or variables, and the output is the dependent variable. The stated functional relationship can be written as:

$$y = f(x_1, x_2, x_3, \ldots x_k)$$

In the functional relationship, k is the independent variable, and y is the dependent variable. For this reason, ANN can be considered a non-linear regression model. The type of ANN most used in predictions is the Multilayer Perceptron (MLP). MLP has an input layer, output layer, and a feed-forward structure, which is usually composed of one and sometimes two or more layers between these two layers (Figure 3; Hamzaçebi, 2011).

2.4 | Logistic regression model (LR)

In this study, the LR model was built using the binary response variable (1: Mistletoe-infected; 0: Uninfected stands) and some explanatory variables. LR models are formulated as follows,

$$p = \mathrm{E}(Y) = \frac{\exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_i X_i)}{1 + \exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_i X_i)}$$

where p is the possibility of mistletoe infection; E (Y) is the expected value of the binary dependent variable Y; β_0 is the constant to be estimated, β_i is the estimated coefficient of the independent variable X_i (Schneider & Pontius, 2001). In this work, statistical analyses were carried out using SPSS version 26.0.

3 | RESULTS

3.1 | Independent-samples *t*-test

There were significant differences between mistletoe-infected and uninfected stands for remote sensing variables (Table 2). These variables were normally distributed based on Skewness and Kurtosis values.

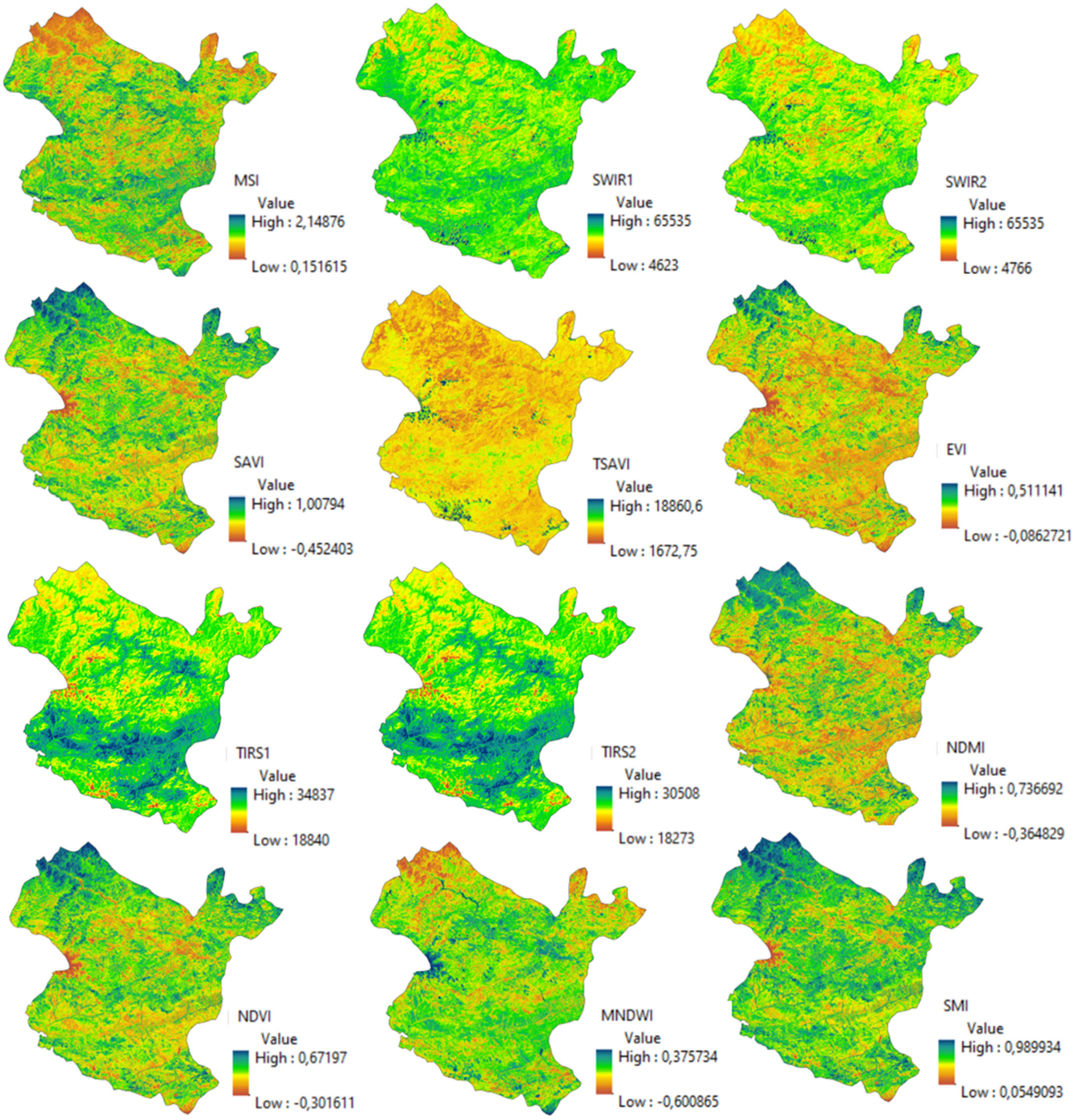

FIGURE 2 Images of remote sensing variables for Gümüşhane province.

With the exceptions of the variables SAVI, EVI, NDVI, MNDWI and SMI, all mean values of the variables MSI, SWIR1, SWIR2, TSAVI, TIRS1, TIRS2 and NDMI were significantly different in mistletoe-infected and uninfected stands ($p < .01$ and $p < .05$) (Table 3).

3.2 | Logistic regression analyses

Binary LR analysis was used to examine whether mistletoe infection could be predicted by remote sensing variables. LR analysis and a single regression model were determined for mistletoe infection. The results of the LR model (Step 1) showed that the probability of mistletoe infection could be estimated using a remote sensing variable, such as TIRS1 (Wald = 23.167, df = 1) (Table 4).

3.3 | MLP artificial neural networks analyses

Two hidden layers were preferred for the prediction of mistletoe infection with MLP ANN using remote sensing variables (MSI, SWIR1,

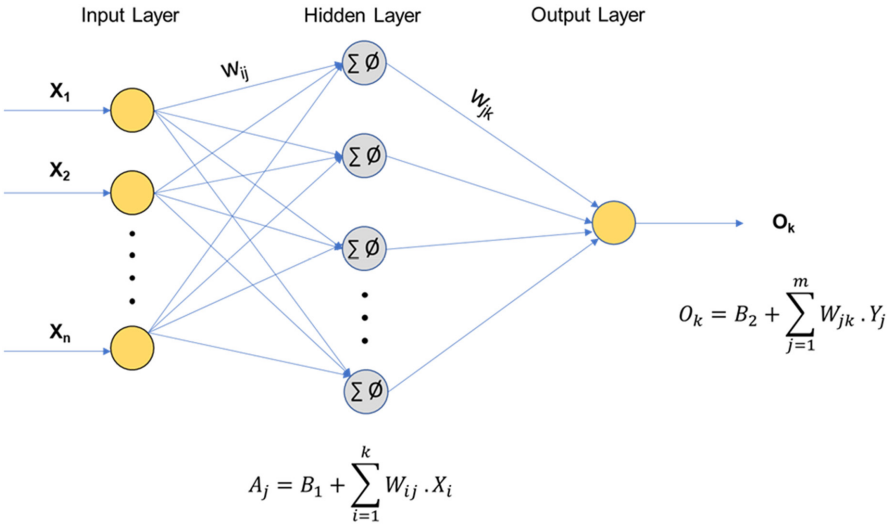

FIGURE 3 Architecture of an MLP ANN

TABLE 3 Results of independent-sample *t*-tests and normality test

Variables	Infection	N	Mean	SE	Sig. (2-tailed)	Mean difference	SE difference	Skewness	Kurtosis
MSI	Absence	40	0.75	0.02	0.028	−0.06	0.03	0.474	0.254
	Presence	69	0.81	0.02					
SWIR1	Absence	40	11,843.05	342.64	0.000	−1570.57	436.09	0.567	0.790
	Presence	69	13,413.62	266.04					
SWIR2	Absence	40	8935.26	253.10	0.001	−1263.03	359.86	0.842	0.669
	Presence	69	10,198.29	231.25					
SAVI	Absence	40	0.51	0.02	0.280	0.03	0.03	0.143	0.007
	Presence	69	0.48	0.02					
TSAVI	Absence	40	3932.94	68.90	0.001	−286.11	87.04	0.457	0.398
	Presence	69	4219.05	52.88					
EVI	Absence	40	0.16	0.01	0.509	0.01	0.01	0.712	0.316
	Presence	69	0.15	0.01					
TIRS1	Absence	40	25,299.78	157.99	0.000	−1651.63	244.09	0.389	−0.545
	Presence	69	26,951.42	161.57					
TIRS2	Absence	40	23,407.10	124.06	0.000	−1257.08	186.62	0.332	−0.611
	Presence	69	24,664.17	122.45					
NDMI	Absence	40	0.15	0.01	0.032	0.04	0.02	−0.064	−0.297
	Presence	69	0.11	0.01					
NDVI	Absence	40	0.34	0.01	0.280	0.02	0.02	0.143	0.007
	Presence	69	0.32	0.01					
MNDWI	Absence	40	−0.30	0.01	0.598	−0.01	0.01	−0.310	0.094
	Presence	69	−0.29	0.01					
SMI	Absence	40	0.81	0.01	0.220	0.01	0.01	−0.229	0.423
	Presence	69	0.80	0.01					

SWIR2, SAVI, TSAVI, EVI, TIRS1, TIRS2, NDMI, NDVI, MNDWI and SMI). Figure 4 shows the structure of the MLP artificial neural network model. In the model, a standardized rescaling method for covariates was used. The number of neurons in the hidden layers was three units in hidden layer 1 and two units in hidden layer 2. Activation functions of both the hidden and output layers were 'Sigmoid'. Case processing and model summaries in MLP ANN modelling are given in Table 5.

Figure 5 shows box plots of predicted pseudo-probabilities for training and testing samples. For the dependent variable 'mistletoe infection', the box plots classify the estimated pseudo-probabilities using the whole dataset (mistletoe-infected and uninfected stands).

TABLE 4 Results of binary logistic regression analysis

	Variables	*B*	SE	Wald	df	Sig.	Exp (*B*)
Step 1	TIRS1	0.001	0.000	23.167	1	0.000	1.001
	Constant	−30.964	6.492	22.750	1	0.000	0.000

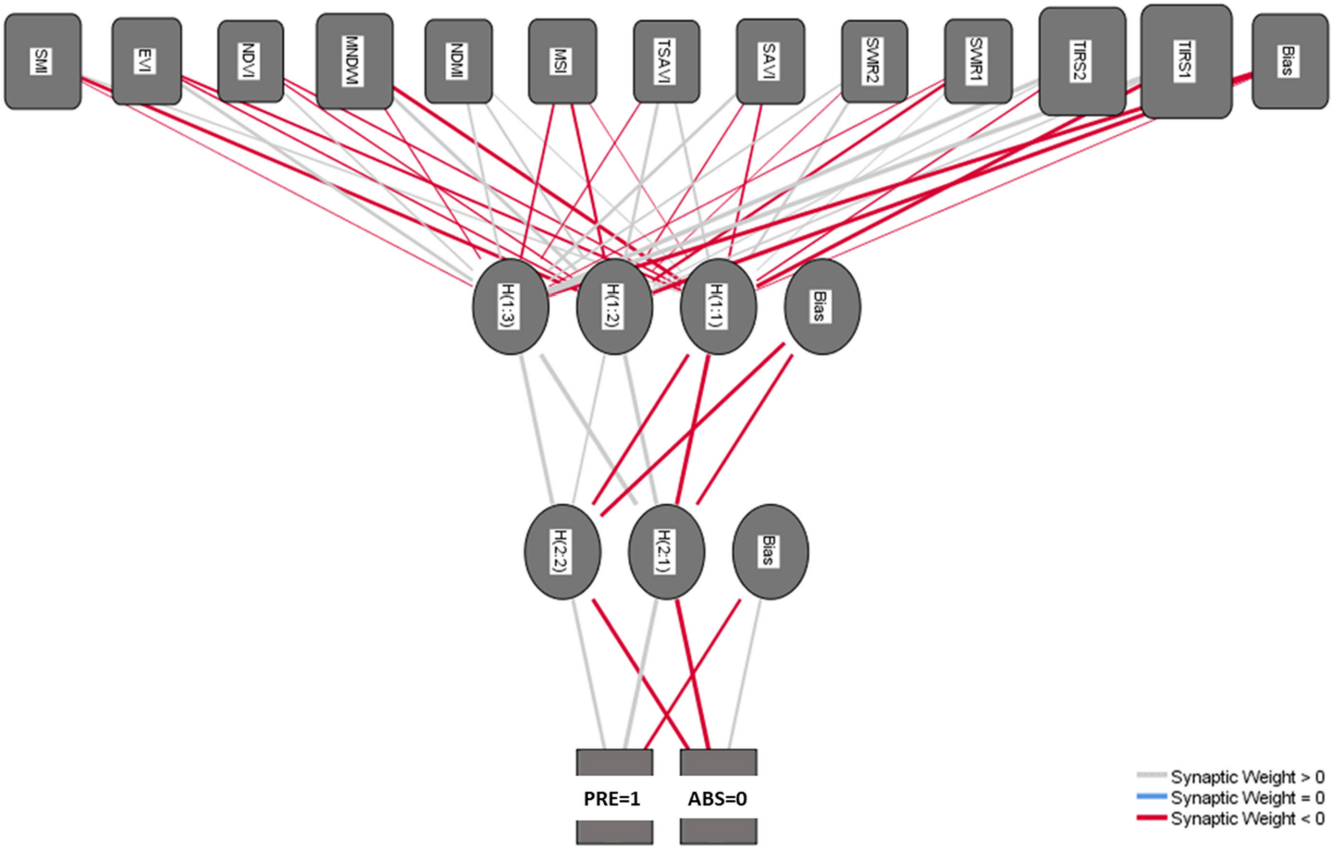

FIGURE 4 Structure of the MLP ANN model for prediction of mistletoe infection.

TABLE 5 Summary statistics for MLP ANN model

	Case processing summary		Model summary	
	N	Percent	Sum of squares error	Relative error
Training	61	56.0	7.263	0.148
Testing	48	44.0	5.335	0.104

The leftmost and next boxplots in Figure 5 display cases that have observed stands without mistletoe infections, the predicted pseudo-probability of uninfected stands and mistletoe-infected stands, respectively. The third and fourth boxplots display cases from mistletoe-infected stands, the predicted pseudo-probability of uninfected stands and mistletoe-infected stands, respectively. For each box plot, the number above 0.5 displays correct predictions, and the number below the 0.5 mark displays incorrect predictions.

The predicted pseudo-probability of mistletoe-infected stands gave better results than mistletoe-uninfected stands (Figure 5). In the observed mistletoe-infected stands, the predicted pseudo-probability of mistletoe-infected stands was close to 1.0 (0.9907). However, for the observed uninfected stands, the predicted pseudo-probability was 0.7124.

During training for the MLP ANN model, the weight proportion of TIRS1, TIRS2, MNDWI and SMI had much larger impacts than other variables in the input layer and were significant among all independent variables for predicting mistletoe infection; their normalized importance levels were 100%, 85.6%, 52.0% and 51.3%, respectively (Figure 6).

3.4 | Comparisons between MLP ANN and LR models

The evaluation criteria of the MLP ANN and LR models are shown in Table 6. There were obvious differences between the two models relating to SPE, SEN, NPV, PPV, accuracy and AUC. The prediction accuracies of the LR and MLP ANN models were 74.3% and 89.6%, respectively (Table 6). However, all remote sensing variables were included in the prediction equation of the MLP ANN model, while

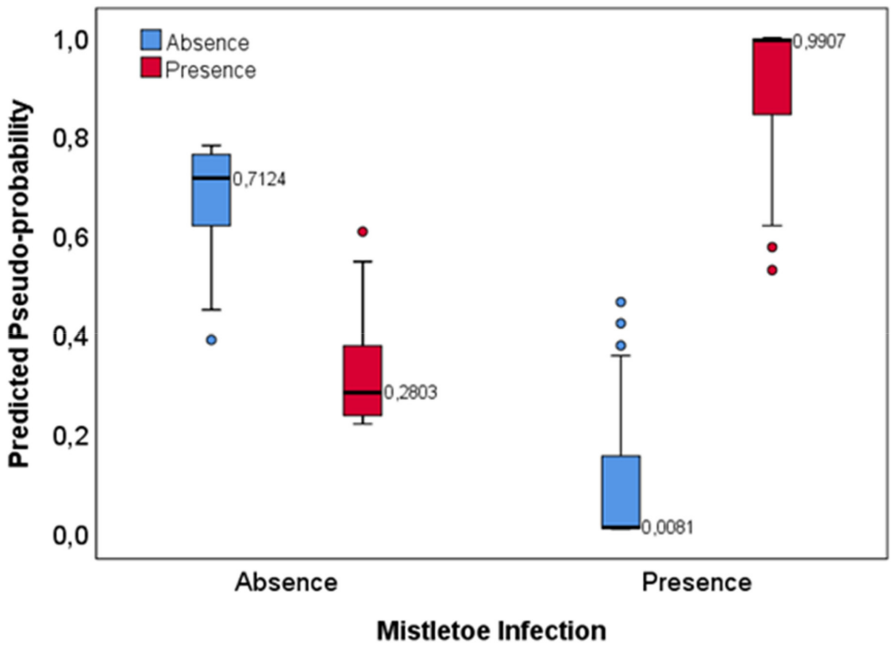

FIGURE 5 Predicted-by-observed chart for absence–presence of mistletoe infection.

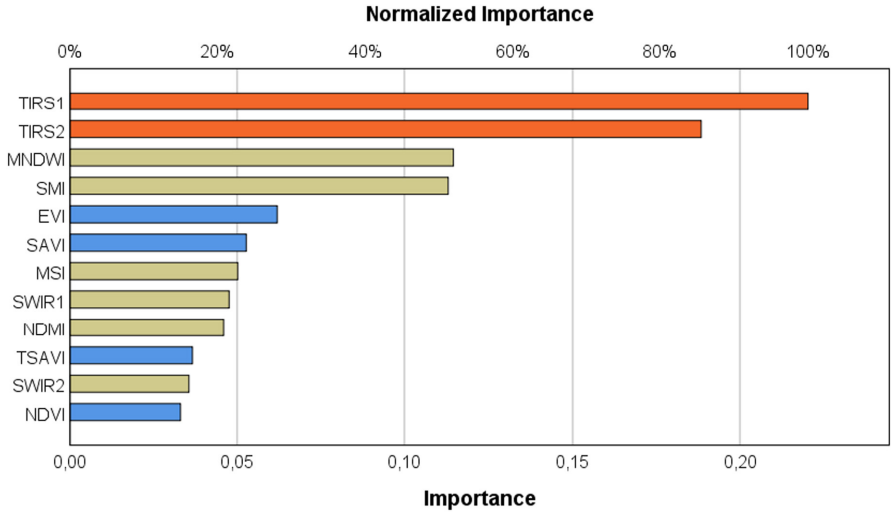

FIGURE 6 Importance of variables for predicting mistletoe infection in an MLP ANN model.

TABLE 6 Comparison of performances results for the ANN and LR models in the prediction of mistletoe infection

Variable	SEN (%)	SPE (%)	PPV (%)	NPV (%)	Accuracy (%)	AUC
ANNs model	90.63	87.50	93.55	82.35	89.6	0.892 ±0.034
LR model	78.87	65.79	81.20	62.50	74.3	0.838 ±0.039

only the TIRS1 variable was included in the prediction equation of the LR model (Figure 6; Table 4). In the MLP ANN model, the TIRS1 variable was also the most important variable (Figure 6).

Figure 7 displays the ROC curve acquired from the LR and MLP ANN models generated by the test data set. The area under the ROC curve (AUC) value for identifying mistletoe infection was 0.892±0.034 when the MLP ANN model was used, demonstrating that the MLP ANN model was superior to the LR model (0.838±0.039) in the overall performance of the prediction.

4 | DISCUSSION

4.1 | Independent-samples *t*-test

Independent-samples *t*-test results, except for SAVI, EVI, NDVI, MNDWI and SMI variables, showed significant differences between mistletoe-infected and uninfected stands for the MSI ($p < .05$), SWIR1 ($p < .01$), SWIR2 ($p < .01$), TSAVI ($p < .01$), TIRS1 ($p < .01$), TIRS2 ($p < .01$) and NDMI ($p < .05$) variables (Table 2). Means of

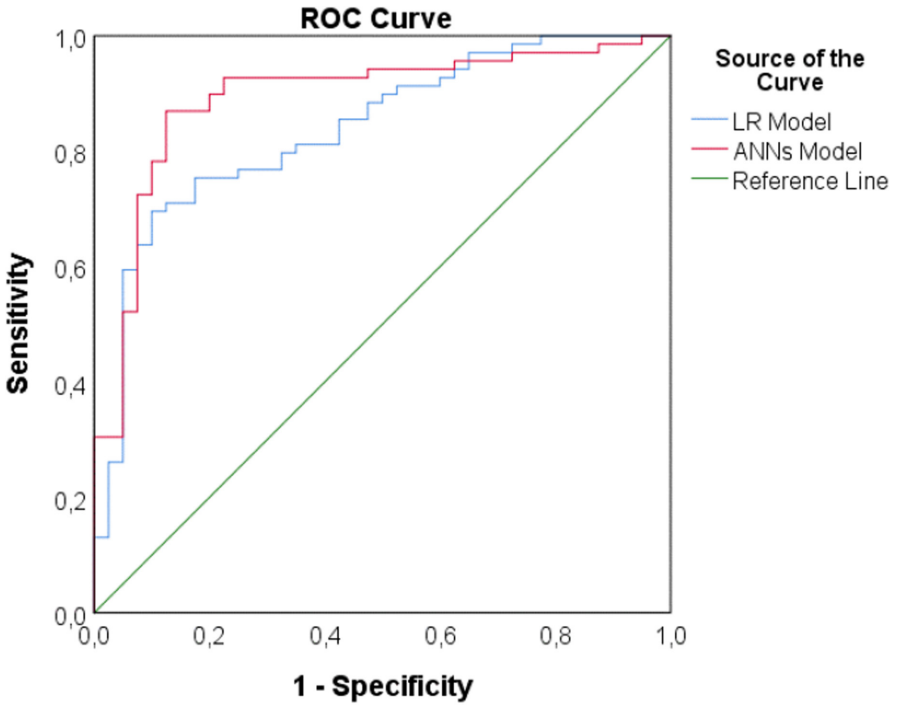

FIGURE 7 Receiver operating characteristic (ROC) curve for predicting the mistletoe infection using LR and ANNs models.

mistletoe-infected (Presence) stands for remote sensing variables were higher than uninfected (Absence) stands, except for NDMI.

The MSI index is used to assess leaf water content. Higher values indicate dry ecosystems (more water stress) and lower values indicate moist and healthy ecosystems (Markov, 2018). In this study, mistletoe-infected stands had higher MSI values than uninfected stands, suggesting that the mistletoe-infected stands were in drier areas than the uninfected stands. When evaluated according to leaf water content, this effect can be explained by water deficit because of the mistletoes in infected stands. Although mistletoe infestation does not directly cause tree mortality, its long-term presence increases the impact of drought (Zweifel et al., 2012). Similarly, the SWIR1, SWIR2 and TSAVI variables support this conclusion. Thus, there is a negative correlation between SWIR and leaf water content. An increase in SWIR values means a decrease in leaf water content. These results are compatible with previous studies that indicated strong absorption of water in SWIR (Fensholt et al., 2010; Galvao et al., 2012). In addition, SAVI, and especially TSAVI compared to NDVI were more accurate in differentiating vegetation cover and bare soil due to the independence of the sensor used (Bannari et al., 1995). However, NDVI and SAVI gave similar results: there was no significant difference (independent *t*-test). These variables were represented by lower values in mistletoe-infected stands. TSAVI (Baret et al., 1989), a soil-adjusted vegetation index, would also be expected to show similar results to SAVI and NDVI (Table 3). Zhang et al. (2005) reported a negative correlation between LAI (Leaf area index) and TSAVI, as well as a positive correlation between TSAVI and proportion of bare ground. In this study, it was interesting that the average TSAVI in mistletoe-infected stands was higher than in uninfected stands.

Thermal infrared sensors (TIRS1 and TIRS2) provide important information on surface temperature. These sensors can also be used for drought research, providing predictions of soil moisture (Ceccato et al., 2001). Average TIRS1 and TIRS2 in mistletoe-infected stands were higher than in uninfected stands (Table 3). Numerous scientific studies have demonstrated the important potential of TIRS data to explain plant biophysical and biochemical characteristics (Buitrago et al., 2016). For instance, the primary absorption feature linked with leaf water content can be shown in the thermal infrared and short-wave regions of the light spectrum (Fabre et al., 2011). Moreover, multiple studies have reported that TIRS data have important potential to define plant disease before the plants show visual stress symptoms (Moller et al., 2007; Ni et al., 2015). Plant functions, such as evapotranspiration, are controlled by stomatal conductivity and are related to surface temperature (Kim et al., 2016). Therefore, changes in these processes will lead to alterations in the air and the surface temperature of leaves (Gersony et al., 2016; Vanderhoof et al., 2013). In addition, plant physiological studies indicated that a decrease in leaf water content breaks stomatal conductance and causes an increase in leaf surface temperature (Pierce et al., 1990; Schulze & Hall, 1982). The results obtained here from NDMI also supported the TIRS results. NDMI has been assessed to predict the moisture contents of sites such as rocks, soils and vegetation, and shown to be valuable in the prediction of ecological and environmental exposures (Nguyen et al., 2016). A high moisture ratio shown by NDMI values >0.1 and values near −1 indicates a low moisture level. The average NDMI in mistletoe-infected stands was significantly lower than in uninfected stands, indicating that mistletoe-infected stands are drier.

Based on the other vegetation indices (SAVI, EVI, NDVI), the vegetation ratio of the mistletoe-infected stands was not significantly different between infected and uninfected stands. Mistletoe-infected stands and uninfected stands were similar in terms of the variable MNDWI and SMI. With the MNDWI, water bodies (such as rivers or lakes) have positive values, and vegetation and soil are generally linked to zero or negative values (McFeeters, 1996). The SMI variable ratio is between 0 and 1 and indicates the relative soil moisture within the site with 0 being the lowest soil moisture and 1 being the highest on a specific day (Tajudin et al., 2021).

Overall, the remote sensing variables showed that mistletoe-infected stands were drier areas with a lower leaf area index.

4.2 | Comparisons between MLP ANN and LR models

In this study, all remote sensing variables were used for the prediction of both models. In the prediction equations of models, all variables (MSI, SWIR1, SWIR2, SAVI, TSAVI, EVI, TIRS1, TIRS2, NDMI, NDVI, MNDWI and SMI) were included in the MLP ANN model, while only TIRS1, which shows surface temperature was included in the LR model. In the LR model, TIRS1 was significantly positively correlated with the prediction of mistletoe-infected stands. Similarly, according to the importance of variables for predicting mistletoe infection in an MLP ANN model, TIRS1 (100%) was of the highest importance, followed by TIRS2, MNDWI, SMI, EVI, SAVI, MSI, SWIR1, NDMI, TSAVI, SWIR2 and NDVI, respectively (Figure 5).

The predicted pseudo-probability of a stand being mistletoe-infected gave better results than mistletoe-uninfected stands in Scots pine forests (Figure 4). For stands observed to have mistletoe-infected stands, the predicted pseudo-probability of mistletoe-infected stands was 0.9907. Moreover, for the stands observed to be uninfected, the predicted pseudo-probability of uninfected stands was 0.7124. This difference indicates that the prediction of mistletoe-infected stands in the MLP ANN model more reliable than prediction in uninfected stands. It can also be an indication that some of the uninfected stands may still potentially be infected by mistletoe.

According to the performance results for both models, the SEN, SPE, PPV, NPV and accuracy of the MLP ANN model were superior to the LR model. Sensitivity (SEN) and specificity (SPE) are the probabilities that the model will correctly classify a presence (mistletoe-infected stands) and an absence (uninfected stands), respectively (Mouton et al., 2010). PPV and NPV are the percentage of mistletoe-infected stands and uninfected stands, respectively (Zhang et al., 2019). In this work, when the SEN, SPE, PPV and NPV performance analyses were evaluated, the MLP ANN model was superior to the LR model. At the same time, the prediction percentages of mistletoe-infected stands were better than the prediction percentages of uninfected stands. The predicted pseudo-probability of the mistletoe-infected-uninfected stands graph (Figure 4) supports this conclusion.

The AUC value for identifying mistletoe-infected Scots pine forests by the MLP ANN model (0.892 ± 0.034) was higher than produced by the LR model (0.838 ± 0.039), explaining the more accurate prediction of the MLP ANNs model (Table 6, Figure 6). However, all variables were included in the MLP ANN model, while only one variable was entered into the LR model. The accuracy of the LR model including only one variable was 74.3%. The accuracy of the MLP ANN model was 89.6%. The fact that the estimation generated by the LR was poorer than that of the MLP ANN can be explained by the fact that the LR model requires a larger amount of data (Besalatpour et al., 2013). Also, the LR analysis should have little or no multicollinearity among the independent variables, with no high correlation among the independent variables used in the LR analysis (Park, 2013). Therefore, certain variables, such as TIRS1, TIRS2, SWIR1 and SWIR2, used as independent variables in the LR model, may show strong relationships that can result in multicollinearity.

This study developed methods for the prediction of the pine mistletoe distribution in Scots pine ecosystems using remote sensing variables to compare the performances of MLP ANN- and LR-derived models. In the study, variables acquired from Landsat 8 (Level 1) image were the independent variables for the independent-samples *t*-test, MLP ANN and LR models. Remote sensing variables indicated the mistletoe-infected stands were drier areas with a lower vegetation-leaf area index. Performance results (SEN, SPE, PPV, NPV and accuracy) of the MLP ANN model were superior to the LR model. The prediction percentages (SEN, SPE, PPV and NPV) of mistletoe-infected stands were better than in uninfected stands. The prediction accuracies of LR and MLP ANN models were 74.3% and 89.6%, respectively. The MLP ANN model included all remote sensing variables, while the LR model used only TIRS1. The AUC value for identifying the mistletoe-infected stands in Scots pine forests used by the MLP ANN model (0.892 ± 0.034) was higher than in the LR model (0.838 ± 0.039). Generally, it can be said that the MLP ANN model performed better than the LR model. However, the 74.3% accuracy provided by TIRS1 in the LR model is also important in assessing this work. The results of this study provide important contributions to the determination of the potential susceptibility of pine forests to mistletoe-infected forests.

ACKNOWLEDGEMENT

This study was supported by The Scientific and Technological Research Council of Turkey with TOVAG-112O258 (Project No).

DATA AVAILABILITY STATEMENT

Research data are not shared.

ORCID

Ayhan Usta https://orcid.org/0000-0002-9647-2576

REFERENCES

Alemdağ, Ş. (1967). *Structure and yield potential of Scotch pine (Pinus sylvestris L.) forests in Turkey and the principles to be followed in managing these forests* (p. 160). Forestry Research Institute.

Bannari, A., Morin, D., Bonn, F., & Huete, A. R. (1995). A review of vegetation indices. *Remote Sensing Reviews*, *13*, 95–120.

Baret, F., Guyot, G., & Germany Major, D. (1989). TSAVI: A vegetation index which minimizes soil brightness effects on LAI and APAR estimation. In *Proceedings 12th Canadian Symposium on Remote Sensing and IGARSS'90*, 10–14 July 1989, Vancouver, Canada (Piscataway, NJ: IEEE), pp. 1355–1358.

Besalatpour, A. A., Ayoubi, S., Hajabbasi, M. A., Mosaddeghi, M. R., & Schulin, R. (2013). Estimating wet soil aggregate stability from easily available properties in a highly mountainous watershed. *Catena*, *111*, 72–79.

Bilgili, E., Coskuner, K. A., Baysal, I., Ozturk, M., Usta, Y., Eroglu, M., & Norton, D. (2020). The distribution of pine mistletoe (*Viscum album* ssp. austriacum) in Scots pine (*Pinus sylvestris*) forests: From stand to tree level. *Scandinavian Journal of Forest Research*, *35*(1–2), 20–28.

Buitrago, M. F., Groen, T. A., Hecker, C. A., & Skidmore, A. K. (2016). Changes in thermal infrared spectra of plants caused by temperature and water stress. *ISPRS Journal of Photogrammetry and Remote Sensing*, *111*, 22–31.

Ceccato, P., Flasse, S., Tarantola, S., Jacquemoud, S., & Grégoire, J. M. (2001). Detecting vegetation leaf water content using reflectance in the optical domain. *Remote Sensing of Environment*, *77*(1), 22–33.

Dobbertin, M., & Rigling, A. (2006). Pine mistletoe (*Viscum album* ssp. austriacum) contributes to Scots pine (*Pinus sylvestris*) mortality in the Rhone valley of Switzerland. *Forest Pathology*, *36*(5), 309–322.

Ehleringer, J. R., & Schulze, E. D. (1985). Mineral concentrations in an autoparasitic *Phoradendron californicum* growing on a parasitic *P. californicum* and its hosts, *Cercidium floridum*. *American Journal of Botany*, *72*, 568–571.

Fabre, S., Lesaignoux, A., Olioso, A., & Briottet, X. (2011). Influence of water content on spectral reflectance of leaves in the 3–15μm domain. *IEEE Geoscience and Remote Sensing Letters*, *8*, 143–147.

Fensholt, R., Huber, S., Proud, S. R., & Mbow, C. (2010). Detecting canopy water status using shortwave infrared reflectance data from polar orbiting and geostationary platforms. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *3*, 271–285.

Galvao, L. S., Neves Epiphanio, J. C., Breuning, F. M., & Formaggio, A. R. (2012). Crop type discrimination using hyperspectral data. In P. S. Thenkabail, J. G. Lyon, & A. Huete (Eds.), *Hyperspectral remote sensing of vegetation* (pp. 397–422). CRC Press.

Gersony, J. T., Prager, C. M., Boelman, N. T., Eitel, J. U. H., Gough, L., Greaves, H. E., Griffin, K. L., Magney, T. S., Sweet, S. K., Vierling, L. A., & Naeem, S. (2016). Scaling thermal properties from the leaf to the canopy in the alaskan arctic tundra. *Arctic, Antarctic, and Alpine Research*, *48*, 739–754.

Hamzaçebi, C. (2011). *Artificial neural networks: Use for estimation, applied matlab and neurosolutions*. Ekin Publications (In Turkish).

Hawksworth, F. G., & Wiens, D. (1996). *Dwarf mistletoes: Biology, pathology, and systematics*. *Agricultural handbook 709* (p. 410). U.S. Department of Agriculture, Forest Service.

He, Y., Chen, G., Potter, C., & Meentemeyer, R. K. (2019). Integrating multi-sensor remote sensing and species distribution modeling to map the spread of emerging forest disease and tree mortality. *Remote Sensing of Environment*, *231*, 111238.

Kim, Y., Still, C. J., Hanson, C. V., Kwon, H., Greer, B. T., & Law, B. E. (2016). Canopy skin temperature variations in relation to climate, soil temperature, and carbon flux at a ponderosa pine forest in central Oregon. *Agricultural and Forest Meteorology*, *226*, 161–173.

Knutson, D. M. (1983). Physiology of mistletoe parasitism and disease responses in the host. In M. Calder & P. Bernhardt (Eds.), *The biology of mistletoes* (pp. 295–316). Academic Press.

Lausch, A., Erasmi, S., King, D. J., Magdon, P., & Heurich, M. (2016). Understanding forest health with remote sensing-part I—A review of spectral traits, processes and remote-sensing characteristics. *Remote Sensing*, *8*(12), 1029.

Markov, B. (2018). Comparison of remote sensing-based indexes for monitoring drought impact on forest ecosystems. In *Annual of Sofia University "St. Kliment Ohridski" Faculty of Geology and Geography Book 2* (pp. 237–246). Universitetsko izdatatelstvo Sv. Kliment Ochridski.

McFeeters, S. K. (1996). The use of the normalized difference water index (NDWI) in the delineation of open water features. *International Journal of Remote Sensing*, *17*(7), 1425–1432.

Moller, M., Alchanatis, V., Cohen, Y., Meron, M., Tsipris, J., Naor, A., Ostrovsky, V., Sprintsin, M., & Cohen, S. (2007). Use of thermal and visible imagery for estimating crop water status of irrigated grapevine. *Journal of Experimental Botany*, *58*(4), 827–838.

Mouton, A. M., De Baets, B., & Goethals, P. L. (2010). Ecological relevance of performance criteria for species distribution models. *Ecological Modelling*, *221*(16), 1995–2002.

Nguyen, A. K., Liou, Y. A., Li, M. H., & Tran, T. A. (2016). Zoning eco-environmental vulnerability for environmental management and protection. *Ecological Indicators*, *69*, 100–117.

Ni, Z., Liu, Z., Huo, H., Li, Z. L., Nerry, F., Wang, Q., & Li, X. (2015). Early water stress detection using leaf-level measurements of chlorophyll fluorescence and temperature data. *Remote Sensing*, *7*, 3232–3249.

Park, H. A. (2013). An introduction to logistic regression: From basic concepts to interpretation with particular attention to nursing domain. *Journal of Korean Academy of Nursing*, *43*(2), 154–164.

Pierce, L. L., Running, S. W., & Riggs, G. A. (1990). Remote detection of canopy water stress in coniferous forests using the NS 001 thematic mapper simulator and the thermal infrared multispectral scanner. *Photogrammetric Engineering and Remote Sensing*, *56*(5), 579–586.

Richardson, D. M., Rundel, P. W., Jackson, S. T., Teskey, R. O., Aronson, J., Bytnerowicz, A., Wingfield, M. J., & Procheş, Ş. (2007). Human impacts in pine forests: Past, present, and future. *Annual Review of Ecology, Evolution, and Systematics*, *38*, 275–297.

Rigling, A., Eilmann, B., Koechli, R., & Dobbertin, M. (2010). Mistletoe-induced crown degradation in Scots pine in a xeric environment. *Tree Physiology*, *30*(7), 845–852.

Schneider, L. C., & Pontius, R. G., Jr. (2001). Modeling land-use change in the Ipswich watershed, Massachusetts, USA. *Agriculture, Ecosystems & Environment*, *85*(1–3), 83–94.

Schulze, E. D., & Hall, A. E. (1982). Stomatal responses, water loss and CO_2 assimilation rates of plants in contrasting environments. In O. L. Lange, P. S. Nobel, C. B. Osmond, & H. Ziegler (Eds.), *Physiological plant ecology II* (pp. 181–230). Springer.

Shahmoradi, L., Safdari, R., Mirhosseini, M. M., Arji, G., Jannat, B., & Abdar, M. (2018). Predicting risk of acute appendicitis: A comparison of artificial neural network and logistic regression models. *Acta Medica Iranica*, *56*, 784–795.

Stone, C., & Mohammed, C. (2017). Application of remote sensing technologies for assessing planted forests damaged by insect pests and fungal pathogens: A review. *Current Forestry Reports*, *3*, 75–92.

Tajudin, N., Ya'acob, N., Ali, D. M., & Adnan, N. A. (2021). Soil moisture index estimation from Landsat 8 images for prediction and monitoring landslide occurrences in Ulu Kelang, Selangor, Malaysia. *International Journal of Electrical and Computer Engineering*, *11*(3), 2101.

TSMS. (2020). *Climate data of Gümüşhane (1961–2019)*. Turkish State Meteorological Service, Ankara.

USGS. (2017). *Product guide: Landsat surface reflectance-derived spectral indices; 3.6 version*. Department of the Interior U.S. Geological Survey (USGS).

USGS. (2018). *Earth explorer*. https://earthexplorer.usgs.gov

Usta, A., & Yılmaz, M. (2021). Effects of land use and topographic variables on distribution of pine mistletoe (*Viscum álbum* subsp. austriacum (Wiesb.) Vollm.) in northeastern Turkey. *Cerne*, *27*, 1–11.

Usta, A., Yilmaz, M., & Kocamanoglu, Y. O. (2018). Estimation of wet soil aggregate stability by some soil properties in a semi-arid ecosystem. *Fresenius Environmental Bulletin*, *27*(12), 9026–9032.

Vanderhoof, M., Williams, C. A., Ghimire, B., & Rogan, J. (2013). Impact of mountain pine beetle outbreaks on forest albedo and radiative forcing, as derived from moderate resolution imaging spectroradiometer, rocky mountains, USA. *Journal of Geophysical Research – Biogeosciences*, *118*, 1461–1471.

Zhang, C., Guo, X., Wilmshurst, J., & Sissons, R. (2005). The evaluation of broadband vegetation indices on monitoring northern mixed grassland. *Prairie Perspectives*, *8*, 23–36.

Zhang, G., Wang, M., & Liu, K. (2019). Forest fire susceptibility modeling using a convolutional neural network for Yunnan Province of China. *International Journal of Disaster Risk Science*, *10*(3), 386–403.

Zuber, D. (2004). Biological flora of central Europe: *Viscum album* L. *Flora - Morphology, Distribution, Functional Ecology of Plants*, *199*(3), 181–203.

Zweifel, R., Bangerter, S., Rigling, A., & Sterck, F. J. (2012). Pine and mistletoes: How to live with a leak in the water flow and storage system? *Journal of Experimental Botany*, *63*(7), 2565–2578.

How to cite this article: Usta, A., & Yilmaz, M. (2023). Prediction of pine mistletoe infection using remote sensing imaging: A comparison of the artificial neural network model and logistic regression model. *Forest Pathology*, *53*, e12783. https://doi.org/10.1111/efp.12783