USING ROC CURVES TO COMPARE NEURAL NETWORKS AND LOGISTIC REGRESSION FOR MODELING INDIVIDUAL NONCATASTROPHIC TREE MORTALITY

Susan L. King[1]

ABSTRACT.—The performance of two classifiers, logistic regression and neural networks, are compared for modeling noncatastrophic individual tree mortality for 21 species of trees in West Virginia. The output of the classifier is usually a continuous number between 0 and 1. A threshold is selected between 0 and 1 and all of the trees below the threshold are classified as mortality trees and all of the trees above the threshold are classified as survival trees. Selecting the threshold that has both a high sensitivity and specificity is a major decision. A receiver operating characteristic (ROC) curve graphically describes the performance of the classifier without the requirement of a threshold. Its accuracy is measured by the area under the curve (AUC). A neural network is the superior classifier because it has a higher AUC statistic.

The output from a statistical procedure, such as logistic regression or neural networks (Bishop 1995), is traditionally classified as to group membership, i.e., mortality or survival, based on whether the output is above or below a threshold or cutoff. The threshold is selected based on historical information, such as the survival or mortality rate of the population, or by chance (cutoff = 0.5). The comparison statistic is usually the correct number of matches, or overall accuracy, between the predicted and observed outcome.

The comparison statistic suffers from several disadvantages. First, different procedures perform better at different decision thresholds. For example, logistic regression may be superior over neural networks at low decision thresholds, whereas neural networks may be superior at higher thresholds. Second, in some applications there is a higher cost associated with different types of misclassifications. Third, this statistic is extremely sensitive to the prevalence of the classes in the data, being biased toward the majority class.

In any noncatastrophic tree-mortality model, the overwhelming population is survival trees. A classification technique could have a high overall accuracy and completely misclassify mortality trees because of their low prevalence. Another classification statistic, such as the kappa statistic or receiver operating characteristics (ROC), is required. This paper applies ROC analysis to determine the classification superiority of either neural networks or logistic regression for modeling individual tree mortality. The most common type of neural networks, a feed-forward backward propagation neural network, was selected for this study. A neural network is not restricted to a categorical dependent variable.

ROC CURVES

ROC curve analysis developed in electrical engineering in the 1950s to detect electromagnetic signals from noise. In the past 30 years, ROC curve analysis has been applied to problems in medical diagnosis and psychology. The diagnostic decision-making process is essentially the same across disciplines (Swets and others 2000).

Environmental applications of ROC curves are scarce and a goal of this paper is to introduce this topic to the forest resource community. One application is a land-cover change model (Pontius and Schneider 2001) and another environmental paper evaluated tree mortality following fire damage (Saveland and Neuenschwander 1990). Several medical imagining applications using neural networks are described in Salchenberger and others (1997)

[1] Operations Research Analyst (SLK), Northeastern Research Station, USDA Forest Service, 11 Campus Blvd., Ste. 200, Newtown Square, PA 19073-3294. SLK is corresponding author: to contact, call (610) 557-4048 or e-mail at sking01@fs.fed.us.

Citation for proceedings: Van Sambeek, J.W.; Dawson, J.O.; Ponder, F., Jr.; Loewenstein, E.F.; Fralish, J.S., eds. 2003. Proceedings, 13th Central Hardwood Forest conference; 2002 April 1-3; Urbana, IL. Gen. Tech. Rep. NC-234. St. Paul, MN: U.S. Department of Agriculture, Forest Service, North Central Research Station. 565 p. [Peer-reviewed paper from oral presentation].

and Wu and others (1993). Other useful references and applications are found in Swets and Pickett (1982), Hanley and McNeil (1982), Gohagan and others (1984), and Swets (1996).

In any application, a threshold distinguishes between noise and a signal, survival and mortality, cancerous and non-cancerous cells. Anything below the threshold is classified as "noise" and anything above the threshold is classified as a "signal." If the threshold is too low, not only is it more likely that a "signal" will be detected, but also that "noise" will be mistaken for a "signal." Similarly, if the threshold is too high, it is very likely that a signal will be misclassified as noise.

At any threshold, a confusion matrix is generated by tabulating the true positives, the true negatives, and the two misclassifications (table 1). In this application, true positives are the number of trees that were predicted to die that actually died and true negatives are the trees that were predicted to survive that actually survived. False positives occur when the trees are predicted to die but actually survive. The opposite is a false negative. The overall accuracy is the sum of the true positives (M_{11}) and true negatives (M_{22}) divided by the total number of observations in the matrix (N).

In medical diagnostics, the rate of true positives (percent of positive events correctly classified) is called sensitivity, whereas the rate of true negatives (percent of nonevents correctly classified) is called specificity. Sensitivity is a measure of accuracy for predicting events, whereas specificity is a measure of accuracy of predicting nonevents. An ROC curve is a plot of the sensitivity versus one minus the specificity (1–specificity), false positives, for a locus of decision thresholds. Both the ordinate and abscissa range from 0 to 1.

The ROC curve for a perfect test would be an inverted "L", going up the ordinate from 0 to 1 and then traveling in a horizontal line from (0,1) to (1,1). A test that makes decisions by chance would have the 45-degree line as the ROC curve. The area under the curve (AUC) is calculated using the trapezoidal rule from calculus or another numerical integration technique and is a measure of the performance of the classifier. Each decision threshold is a point on the curve. A perfect test would have an AUC of 1, whereas a test by chance would have an AUC of 0.5. The AUC provides a measure of performance that is less sensitive to the prevalence of a class in the data and measures the performance of the classifier across the entire range of thresholds.

For logistic regression, the OUTROC option in PROC Logistic in SAS (SAS Institute, Cary, NC) provided both the AUC statistics and the locus of points for the ROC curve. For neural networks, a user written program in SAS provided the AUC curve statistics and the locus of points for the ROC curve.

INDIVIDUAL TREE MORTALITY

An important part of forest management is predicting the growth and yield of individual trees. Growth models have four components: survivor growth, ingrowth, cutting, and mortality.

Table 1.—*Classification results for a threshold*

		Observed		
		Mortality	Survival	Totals
Predicted	Mortality	True Positives (M11)	False Positives (M12)	M11 + M12 = N1
	Survival	False Negatives (M21)	True Negatives (M22)	M21 + M22 = N2
	Totals	M11 + M21	M12 + M22	N = M11 + M12 + M21 + M22

Predicting noncatastrophic individual tree mortality is the most difficult of the four components to model because an overwhelming number of trees survive from the first measurement to the second measurement. King and others (2000) evaluates neural networks, logistic regression, and two support vector methods to predict noncatastrophic individual tree mortality. They determined that neural networks were the superior technique for evaluating individual tree mortality. ROC curve analysis was one of their tools for reaching this conclusion.

DATA AND VARIABLES

The data for this project came from the 1964, 1975, and 1988 forest inventories of West Virginia collected by the USDA's Northeastern Forest Inventory and Analysis (NEFIA) unit. This analysis uses trees that were alive at the second measurement (1975) and were larger than 5 inches in diameter at breast height (d.b.h.). Some of the trees that were measured in the 1975 and 1988 inventories were not measured in the 1964 inventory. Their diameters were estimated using a linear regression procedure for estimating previous diameters developed by King and Arner (1999).

Table 2 shows the distribution of the species into 21 groups and the number of survival and mortality trees in each species group. The species groups were classified according to growth and competitive characteristics. The same set of independent variables was included in each species group. In West Virginia, there were no trees in species group 13, sweetbay.

For modeling, each species group was broken into a model and a validation data set. The modeling procedures require each pattern, or combination of dependent and independent variables, to be represented in the model data set. This is an impossible objective given several continuous independent variables. The solution is to obtain a proportional representation of the mortality by diameter class and hopefully a proportional representation of each pattern, the data were divided into 2-inch diameter classes. All trees 23 inches d.b.h. and larger were grouped into one class. Using a random number generator, approximately 75 percent of the trees in each diameter and mortality class were placed in the model data set and the remaining trees were placed in the validation data set. Mortality is highest in the lower and upper

Table 2.—*Species group assignments*

Species groups	Common names	Number of survival trees	Number of mortality trees
1	Red spruce	107	31
2	Eastern redcedar, shortleaf pine, table mountain pine, pitch pine, Eastern white pine	283	113
3	Virginia pine	495	203
4	Eastern hemlock	304	15
5	Boxelder, sugar maple	908	94
6	Red maple	1298	124
7	Yellow birch, sweet birch	70	193
8	Hickory	124	219
9	American beech	1031	167
10	White ash, green ash	31	62
11	Butternut, black walnut	145	48
12	Yellow-popla	1344	140
13	Sweetbay	0	0
14	Black cherry	44	62
15	White oak, chestnut oak, post oak Scarlet oak, Southern red oak, overcup oak, chinkapin oak, pin oak,	3635	423
16	Willow oak Willow oak	541	129
17	Northern red oak, black oak	2076	320
18	Black locust	335	226
19	American basswood	355	43
20	All other commercial species	761	203
21	All other noncommercial species	432	306

diameter classes. Smaller trees are more likely to die from competition, whereas trees that are larger, and therefore older, suffer from cumulative life stresses.

The models require three time periods of data. The independent variables are derived from the variables collected in the first two measurements. The dependent variable, mortality, is from the third measurement. Many possible independent variables were investigated. Variables such as crown class and crown ratio were not available for all of the trees in the second measurement. Modeling of the missing crown classes and crown ratios was not successful.

In many other individual tree mortality models, one of the independent variables is site quality. Site index, the height of a free-growing dominant or codominant tree at an index age, is used to measure site quality. Most of the site-index equations are based on the assumption that the trees on the plot or in a stand are of the same species and age. These assumptions are violated in most of the Northeastern states. From NEFIA's experience, site is not a useful predictor of growth and slow growth is an indicator of susceptibility to mortality.

Six independent variables were included in the final model: basal area at the second inventory (BA2); basal area increment (BAI); relative basal area increment (RBAI); basal area larger than the subject tree (BAL); the interaction of basal area and basal area increment (BABAI); and soundness. Guan and Gertner (1991) used d.b.h., d.b.h. increment, the interaction between d.b.h. and d.b.h. increment, and soundness as independent variables in their comparison of logistic regression and neural networks for modeling individual tree mortality. In addition to an expanded variable list, three time periods of data were used instead of two.

BAI is the change in the basal area of an individual tree between the first and second inventories divided by the number of years between NEFIA plot measurements. It is a measure of the vigor of the tree and is indicative of its survival. A tree with low vigor is more likely to die.

RBAI is the ratio of each individual tree's BAI with the total BAI of all of the trees on the plot. RBAI is a measure of an individual tree's vigor as compared to the vigor of the other trees on the NEFIA plot.

BAL measures the competitive position of an individual tree within a plot as expressed by the sum of the basal area of all trees larger than the subject tree. The largest tree on the plot has a BAL of zero and the smallest tree has a BAL equal to the sum of the basal areas of the larger trees. Teck and Hilt (1990) report that BAL had a higher correlation with survival than the ratio of d.b.h. to quadratic mean stand diameter and the ratio of tree basal area to plot basal area. Also King and Arner (1999) found that BAL was a better explanatory variable for competition than either relative diameter or relative quadratic diameter.

BABAI captures the relationship between tree size and vigor. Buchmann and others (1983) postulate that the survival rate for low vigor trees depends on tree size and that the survival peaks between the juvenile stage and senescence. The relationship between species and size is species dependent. Long-lived trees with low vigor are more likely to survive if they are past the sapling stage. Short-lived trees with low vigor are more likely to survive when they are small.

Soundness is an estimated variable. It is a measure of rot and decay in the tree. Not all of the information was available for every tree in the second measurement. Merchantability (MRCH2) is a class variable that captures cull and form of the tree bole. Trees of poor form and high cull die more frequently. The mean of the soundness calculations for the third measurement was assigned to a tree based on its MRCH2 classification. The following values were assigned to soundness.

MRCH2	Soundness
1	99
2	98
3	81
5	90
6	45

Growing stock trees have a MRCH2 of 1, 2, or 3 and cull trees have a MRCH2 of 5 or 6. Trees with a MRCH2 of 1 have no cull, whereas trees with a MRCH2 of 2 or 5 have only sound cull and trees with a MRCH2 of 3 or 6 have rotten cull. A frequency analysis between mortality and cull shows that mortality occurs more frequently if the cull is rotten.

RESULTS AND CONCLUSIONS

The ROC curves for the model and validation data sets were created and analyzed for each species group. Each curve has 100 points, each corresponding to a threshold. The population mortality ranged from 4.5 percent for eastern hemlock to 42.2 percent for all other noncommercial species. The model data set is used to establish the cutoff. However, the superiority of a technique is determined by the model's ability to classify new data. This is evaluated by the validation data set. The results show a variety of ROC curves. By definition all of the curves start at the origin and end at (1,1). At the origin, all of the survival trees are perfectly classified and all of the mortality trees are misclassified, whereas at (1,1) all of the mortality trees are perfectly classified.

The lower cutoffs are associated with a low sensitivity, a low 1-specificity, and are biased toward classifying survival trees correctly. The higher cutoffs are associated with higher sensitivity values and are biased toward correctly predicting mortality at the expense of survival trees. The overall accuracy is high at low cutoffs and is low at the higher cutoffs due to the large number of survival trees. Landis and Koch (1977) classified different ranges of kappa. These classifications are poor ($\breve{\kappa} \leq 0.4$), good ($0.4 < \breve{\kappa} \leq 0.75$), and excellent ($\breve{\kappa} > 0.75$).

Since the purpose of this paper is to illustrate the technique, only the results for three species are presented: black cherry, Virginia pine, and yellow-poplar. For species group 14, black cherry, figure 1a shows neural networks trading positions of superiority with logistic regression over the decision space. Overall, neural networks are slightly superior with an AUC of 0.852 to an AUC of 0.809 for logistic regression. Since it is difficult to discern the changes in the graph, Appendix table 1 presents selected thresholds used to create the graphs. From Appendix table 1 and from figure 1a, the two procedures are equivalent until a threshold of 0.06. The overall accuracy is high, over 88 percent, yet the kappa statistic is low, less than 0.01, indicating a lack of agreement between the two classifiers. From a threshold of 0.06 to 0.18, logistic regression is superior as indicated by the larger sensitivity. Both procedures have the same 1–specificity value.

It is difficult to discern the difference graphically, but the logistic regression curve is slightly to the left of the neural network curve in figure 1a. Both procedures are classifying survival the same, but logistic regression correctly classifies more trees that actually died. From a threshold of 0.18 to a threshold of 0.90, neural networks are superior to logistic regression. The latter threshold can be found from the graph and Appendix table 1. The two curves cross at approximately a sensitivity of 0.84 and a 1-specificity of 0.38. This corresponds to a cutoff between 0.90 and 0.92 in Appendix table 1. At a threshold of 0.92, the overall accuracy is 58.5 percent for neural networks and 64.8 percent for logistic regression. The procedures swap positions of superiority twice again before the final cutoff of 1.0. The highest overall number of correct matches for neural networks and logistic regression occur at thresholds 0.64 and 0.44 to 0.46, respectively. The highest kappa for neural networks is 0.523, which occurs at a threshold of 0.68, and the maximum kappa for logistic regression is 0.472, which occurs at a threshold of 0.72.

Whether the optimal threshold is from the maximum value of the kappa statistic, maximum overall accuracy, by chance, or another criterion, this threshold is found from the model data set and applied to the validation data set. The validation data set has an AUC of 0.745 for

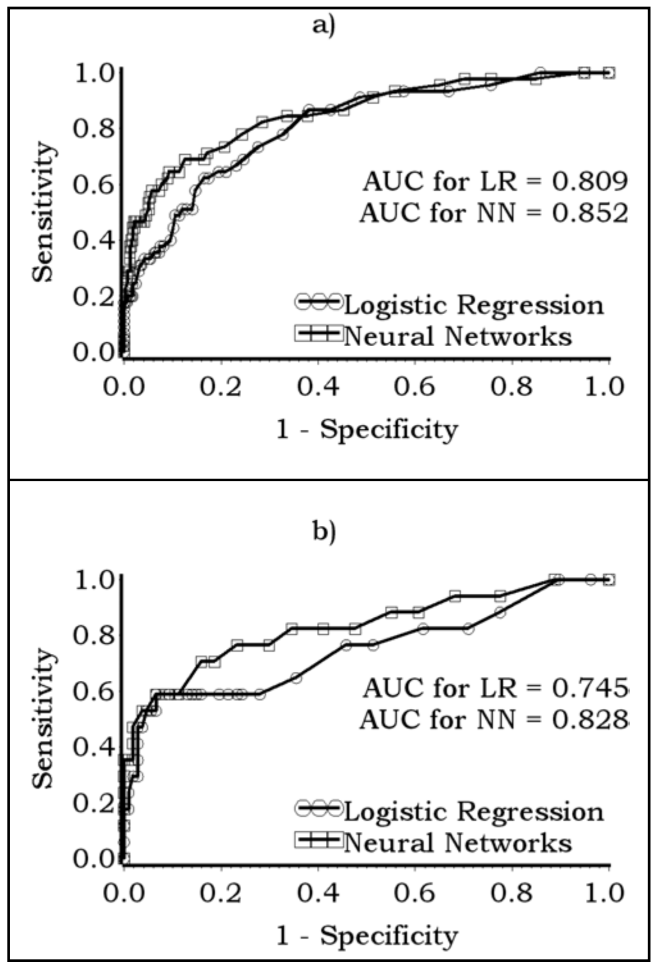

Figure 1.—*Black cherry ROC curves for: a) the model data; and b) the validation data set. Of the 378 trees in the model data set, 45 were mortality trees and of the 124 trees in the validation data set, 17 were mortality trees. Neural networks are the superior classifier due to the higher AUC statistics. Each point on the curve represents a threshold.*

logistic regression and 0.828 for neural networks. Neural networks and logistic regression are equivalent until a sensitivity of 0.12, which occurs at a threshold of 0.06. Logistic regression is superior until a threshold between 0.18 and 0.20, which corresponds to a sensitivity of around 0.2. The two procedures are close with neural networks having the edge until a sensitivity of 0.6, which corresponds to a threshold of around 0.86. Neural networks are superior to logistic regression over the remaining thresholds. For any of the optimal cutoffs from the model data set, neural networks would be the superior procedure at that threshold.

Virginia pine is an example of an almost perfect classifier (fig. 2). The AUC for neural network and logistic regression model data sets is 0.941 and 0.937, respectively. This indicates that neural networks are a slightly better classifier and that both procedures are almost perfect classifiers. The AUC for the validation data sets is 0.917 for neural networks and 0.911 for logistic regression. The curves have an inverted "L" shape, with high sensitivity and low 1–specificity values.

Conversely, yellow-poplar is an example of a poor classifier (fig. 3). For both the model and validation data sets, the AUC statistics are between 0.7 and 0.8. Although, these statistics are above 0.5, they are still quite low. The graphs in figures 3a and 3b rise slowly after a sensitivity of 0.3.

FURTHER APPLICATIONS

In many applications, especially in medical diagnostics, there is a greater cost associated

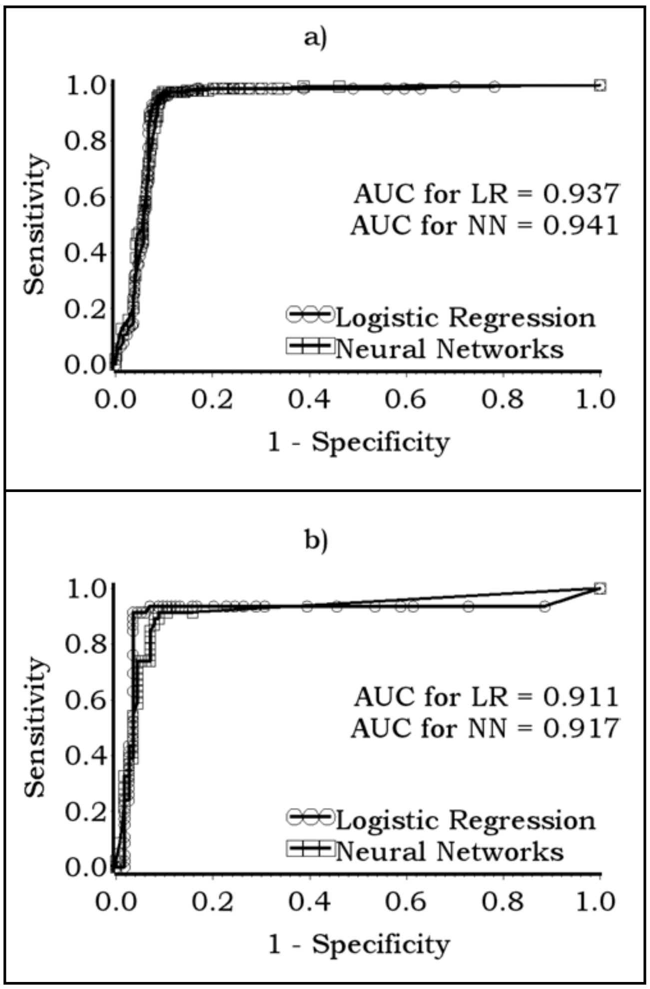

Figure 2.—*Virginia pine ROC curves for: a) the model data set; and b) the validation data set. Of the 538 trees in the model data set, 157 were mortality trees and of the 160 trees in the validation data set 46 were mortality trees. Both classifiers are equivalent and their inverted "L" shape curve indicates that they are close to perfect.*

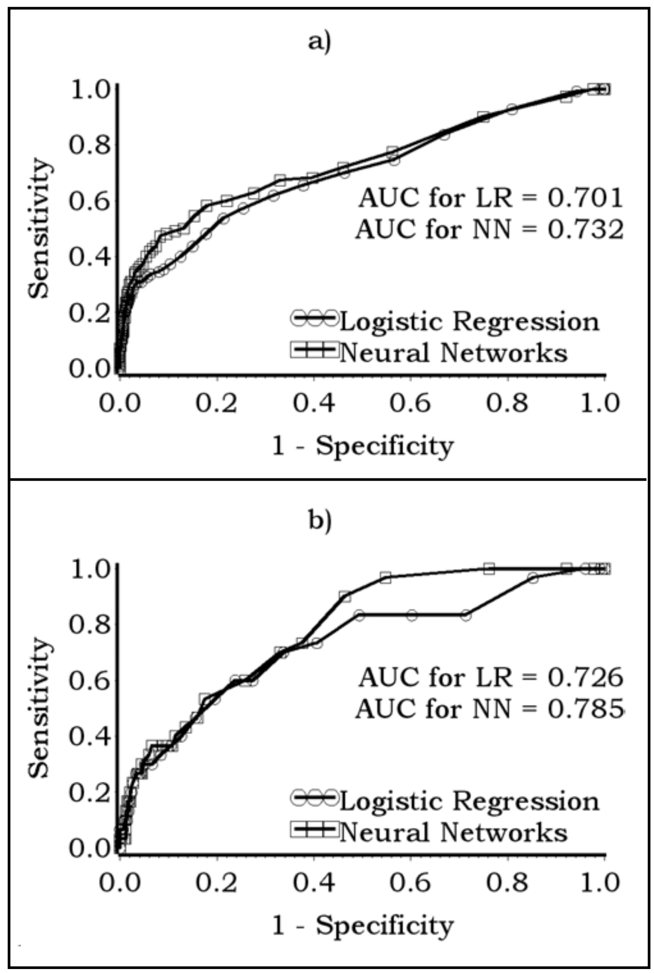

Figure 3.—*Yellow-poplar ROC curves for: a) the model data set; and b) the validation data set. Of the 1,122 trees in the model data set, 110 were mortality trees and of the 362 trees in the validation data set, 30 were mortality trees. Although neural networks are superior, both procedures are poor classifiers as indicated by their closeness to a 45-degree line.*

with false negatives than false positives. A cost or a loss matrix for each of the classes is defined in a confusion matrix. A risk function at each cutoff is defined similarly to that proposed by Bishop (1995) by combining the confusion matrix, cost or loss matrix, and the prior classification probabilities. This measurement is still sensitive to the prior probabilities of class occurrence and the decision threshold. A series of thresholds between 0 and 1 is selected and the risk is calculated at each threshold. The optimal cutoff corresponds to the minimum risk. Mathematically risk is defined as:

$$R=\sum_{j=1}^{c} R_j P(C_j) \qquad (1)$$

where

$$R_j=\frac{1}{N_j}\sum_{k=1}^{c} M_{jk} L_{jk}$$

M_{jk} =number of trees in the confusion matrix at row j and column k

L_{jk} = loss associated with misclassifying a tree

$P(C_j)$ = prior probability that a tree comes from class j

N_j = is the number of trees in class j

c = the number of classes which is equal to j or k

Unlike medical diagnostics, it is difficult in forestry to assign an economic value for loss or even determine which is a more severe misclassification, a false positive or a false negative. In medical diagnostics, it is obviously more costly to make a false negative misclassification error than a false positive misclassification error. With the latter, further testing should reveal the error, whereas no further testing is usually performed in the former case.

In forestry, if the purpose of classification were to estimate volume for an entire state, then the number and size of the misclassified trees would either inflate or deflate the estimates. It is not clear which type of error would be worse. A timber harvester would incur cost of locating the tree on the ground and lost revenue for trees that were predicted to survive but died, a false negative. For a false positive, there would be a missed opportunity cost of harvesting the tree. There would be no loss incurred for correctly classifying the tree.

For demonstration purposes, assume that the loss matrix is given in table 3. There is no loss in correctly classifying mortality or survival. A false negative has a slightly higher loss than a false positive. Using equation 1, risk is calculated for every threshold value for both logistic regression

Table 3.—*Loss matrix for risk evaluation*

		Observed	
		Mortality	Survival
Predicted	Mortality	$0.00	$600.00
	Survival	$700.00	$0.00

and neural networks. The prior probability for black cherry mortality, 12.35 percent, is taken to be the total number of mortality trees in the sample population, the combined model and validation data sets. Similarly, the prior probability for survival was found and was 87.65 percent. The minimum risk for neural networks was $57.50 and occurred at a threshold of 0.64. The minimum risk for logistic regression was found to be $71.09, occurring at a cutoff of 0.44. These cutoffs would be applied to the validation data set.

CONCLUSIONS

ROC curve analysis allows the user to compare two or more classifiers over the range of thresholds and to visually compare the performance of the classifiers. In most cases, neural networks were superior to logistic regression. The two procedures were generally equivalent at the lower threshold values and neural networks gained its superiority at the higher threshold values. Risk analysis is a useful tool for finding the optimal threshold. If there were a sufficient number of trees, breaking the models into three categories of young, middle-, and older-age trees might improve the results.

ROC curve analysis has many environmental applications and has been underutilized. It is a useful tool for comparing maps produced by different classifiers. The ROC methodology presented above is limited to “yes” or “no” class membership. Extension of the dichotomous ROC curve analysis to multiple cases is a current area of research (Dreiseitl and others 2000, Pontius and Schneider 2001).

LITERATURE CITED

Bishop, C.M. 1995. Neural networks for pattern recognition. New York, NY: Oxford University Press, Inc. 482 p.

Buchmann, R.G.; Pederson, S.P.; Walters, N.R. 1983. A tree survival model with application to species of the Great Lakes region.

Canadian Journal of Forest Research. 13: 601-608.

Dreiseitl, S.; Ohno-Machado, L.; Binder M. 2000. Comparing three-class diagnostic tests by three-way ROC analysis. Medical Decision Making. 20(3): 323-331.

Gohagan, J.K.; Spitznagel, E.L; McCrate, M.; Frank, T.A. 1984. ROC analysis of mammography and palpation for breast screening. Investigative Radiology. 19(6): 587-592.

Guan, B.T.; Gertner, G. 1991. Using a parallel distributed processing system to model individual tree mortality. Forest Science. 37(3): 871-885.

Hanley, J.A.; McNeil, B.J. 1982. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology. 143: 29-36.

King, S.L.; Arner, S.L. 1999. Estimating previous diameter for ingrowth trees on remeasured horizontal point samples. In: Stringer, J.W.; Loftis, D.L., eds. Proceedings, 12th Central Hardwood Forest conference. Gen. Tech. Rep. SRS-24. Asheville, NC: U.S. Department of Agriculture, Forest Service, Southern Research Station: 152-158.

King, S.L.; Bennett, K.P.; List, S. 2000. Modeling noncatastrophic individual tree mortality using logistic regression, neural networks, and support vector methods. Computers and Electronics in Agriculture. 27(1-3): 401-406.

Landis, J.R.; Koch, G.G. 1977. The measure of observer agreement for categorical data. Biometrics. 33: 159-174.

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: 239-248.

Saveland, J.; Neuenschwander, L.F. 1990. A signal detection framework to evaluate models of tree mortality following fire damage. Forest Science. 26(1): 66-76.

Salchenberger, L.; Venta, E.R.; Venta, L.A. 1997. Using neural networks to aid the diagnosis of breast implant rupture. Computers and Operations Research. 24(5): 435-444.

Swets, J.A. 1996. Signal detection theory and ROC analysis in psychology and diagnostics: collected papers. Mahwah, NJ: Lawrence Erlbaum Associates, Inc. 328 p.

Swets, J.A.; Pickett, J.M. 1982. Evaluation of diagnostic systems. New York, NY: Academic Press. 253 p.

Swets, J.A.; Dawes, R.M.; Monahan, J. 2000. Better decisions through science. Scientific American. October: 82-87.

Teck, R.M.; Hilt, D.E. 1990. Individual-tree probability of survival model for the Northeastern United States. Res. Pap. NE-642. Radnor, PA: U.S. Department of Agriculture, Forest Service, Northeastern Forest Experiment Station. 10 p.

Wu, Y.; Giger, M.L.; Doi, K.; Vyborny, C.J.; Schmidt, R.A.; Metz, C.E. 1993. Artificial neural networks in mammography: application to decision making in the diagnosis of breast cancer. Radiology. 187: 81-87.

Appendix Table 1.—*Selected thresholds for black cherry for logistic regression (LR) and neural networks (NN)*

	Model				Validation			
Threshold	LR	% correct kappa sensitivity 1-specific.	NN	% correct kappa sensitivity 1-specific.	LR	% correct kappa sensitivity 1-specific.	NN	% correct kappa sensitivity 1-specific.
0.00	0 45 0 333	88.1 0.00 0.00 0.00	0 45 0 333	88.1 0.00 0.00 0.00	0 17 0 107	86.3 0.00 0.00 0.00	0 17 0 107	86.3 0.00 0.00 0.00
0.02	1 44 0 333	88.4 0.039 0.02 0.00	1 44 0 333	88.4 0.039 0.02 0.00	0 17 0 107	86.3 0.00 0.00 0.00	0 17 0 107	86.3 0.00 0.00 0.00
0.04	1 44 0 333	88.4 0.039 0.02 0.00	1 44 0 333	88.4 0.039 0.02 0.00	0 17 0 107	86.3 0.00 0.00 0.00	0 17 0 107	86.3 0.00 0.00 0.00
0.06	2 43 0 333	88.6 0.076 0.04 0.00	1 44 0 333	88.4 0.039 0.02 0.00	2 15 0 107	87.9 0.187 0.12 0.00	0 17 0 107	86.3 0.00 0.00 0.00
	⋮		⋮		⋮		⋮	
0.18	6 39 0 333	89.7 0.213 0.13 0.00	10 35 1 332	90.5 0.326 0.22 0.00	3 14 0 107	88.7 0.27 0.18 0.00	3 14 0 107	88.7 0.27 0.18 0.00
	⋮		⋮		⋮		⋮	
0.50	8 37 1 332	89.9 0.267 0.18 0.00	17 28 5 328	91.3 0.466 0.38 0.02	3 14 1 106	87.9 0.25 0.18 0.01	6 11 0 107	91.1 0.485 0.35 0.00
	⋮		⋮		⋮		⋮	
0.68	15 30 17 316	87.6 0.323 0.33 0.05	21 24 8 325	91.5 0.523 0.47 0.02	5 12 3 104	87.9 0.342 0.29 0.03	6 11 1 106	90.3 0.457 0.35 0.01
	.		.		.		.	

Thres-hold	Model LR	% correct kappa sensitivity 1-specific.	Model NN	% correct kappa sensitivity 1-specific.	Validation LR	% correct kappa sensitivity 1-specific.	Validation NN	% correct kappa sensitivity 1-specific.
0.72	16 29 24 309	86.0 0.472 0.36 0.07	21 24 12 321	90.5 0.487 0.47 0.04	8 9 3 104	90.3 0.52 0.47 0.03	7 10 2 105	90.3 0.490 0.41 0.02
	⋮		⋮		⋮		⋮	
0.88	30 15 77 256	75.7 0.273 0.67 0.23	35 10 81 252	75.9 0.318 0.78 0.24	10 7 23 84	75.8 0.267 0.59 0.21	12 5 17 90	82.3 0.422 0.71 0.16
0.90	33 12 92 241	72.5 0.258 0.73 0.28	38 7 112 221	68.5 0.253 0.84 0.34	10 7 26 81	73.4 0.235 0.59 0.24	13 4 25 82	76.6 0.349 0.76 0.23
0.92	39 6 127 206	64.8 0.224 0.87 0.38	39 6 151 182	58.5 0.173 0.87 0.45	11 6 38 69	64.5 0.163 0.65 0.36	14 3 37 70	67.7 0.259 0.82 0.35
0.94	41 4 162 171	56.1 0.169 0.91 0.49	42 3 186 147	50.00 0.136 0.93 0.56	13 4 55 52	52.4 0.111 0.76 0.51	14 3 51 56	56.5 0.159 0.82 0.48
0.96	42 3 223 110	40.2 0.085 0.93 0.67	44 1 234 99	37.8 0.085 0.98 0.70	14 3 76 31	36.3 0.04 0.82 0.71	15 2 65 42	46.0 0.107 0.88 0.61
0.98	45 0 286 47	24.34 0.038 1.00 0.86	44 1 283 50	24.9 0.034 0.98 0.85	17 0 96 11	22.6 0.03 1.00 0.90	16 1 83 24	32.3 0.055 0.94 0.78
1.00	45 0 333 0	11.9 0.00 1.00 1.00	45 0 333 0	11.9 0.00 1.00 1.00	17 0 107 0	13.7 0.00 1.00 1.00	17 0 107 0	13.7 0.00 1.00 1.00