IdentiFlight®

Appendix - Model Evaluation and Selection Metrics

Metric	Formula	Focus
True Positive (TP)		These are the correctly predicted positive values. That is Golden-Eagle as Golden Eagle or Bald-Eagle as Bald-Eagle. Use curtail\no-curtail confusion matrix to determine these numbers.
True Negative (TN)		These are the correctly predicted negative values. That is Turkey-Vulture as Turkey-Vulture, Hawk as Hawk, etc.
False Positive (FP)		When actual class is negative the predicted value is positive. Turkey-Vulture as Bald-Eagle, Raven as Golden-Eagle, etc.
False Negative (FN)		When actual class is positive the predicted class is negative. Golden-Eagle as Turkey Vulture, Bald-Eagle as Hawk, etc
Accuracy (ACC)	$\frac{TP+TN}{TP+TN+FP+FN}$	Measures the ratio of correct predictions over the total number of frames evaluated.
Error rate (ERR)	$\frac{FP+FN}{TP+TN+FP+FN}$	The ratio of incorrect predictions over the total number of frames evaluated.
Sensitivity (SN)	$\frac{TP}{TP+FN}$	The fraction of positive patterns that are correctly classified.
Specificity (SP)	$\frac{TN}{NP+FP}$	The fraction of negative patterns that are correctly classified.
Precision (P)	$\frac{TP}{TP+FP}$	What proportion of positive identifications were correctly classified? Precision is used to measure the positive patterns that are correctly predicted from the total number of predicted patterns in a positive class.
Recall (R)	$\frac{TP}{TP+FP}$	The fraction of positive patterns that are correctly classified.
F-Measure (FM)	$\frac{2 \cdot P \cdot R}{P+R}$	The harmonic mean between recall and precision values.

Metric	Formula	Focus
Geometric-Mean (GM)	$\sqrt{TP \cdot TN}$	Used to maximize the TP rate and TN rate, and simultaneously keeping both rates relatively balanced
MICRO-Average		Aggregate contribution from each class and compute the average.
MACRO-Average	$M_{pr} = \frac{\sum_i^N Pr_i}{N}$	Compute metric for each class independently then take the average (treats every species equally)

Appendix - Model Evaluation Plots

Metric	Focus
Receiver Operating Characteristic curve (ROC)	The ROC curve is plotted with TP-rate against the FP-rate. It shows the trade off between sensitivity and specificity. A curve close to the 45-degree line means the model makes random guesses. The closer the curve follows the left-hand border, or the steeper it is, the more accurate the mode model is.
Area Under the Curve	AUC represents the degree or measure of separability. It tells how much the model is capable of distinguishing between classes Any value above 0.95 is considered very good
Confusion Matrix	If we label Bald-Eagle, Golden-Eagle as positive classes, and the rest as negative classes, we can use the species-based matrix to import_pd the False Positive and False Negative rate of a species. For instance, cell (0,0) is the FP rate of Bald-Eagle which measures how often Bald-Eagle is confused for Golden-Eagle. In general, if viNet misclassifies any positive class for another positive class (GE as BE or vice versa), that is considered FP. If it misclassifies any positive class as a negative class (GE as Raven) that is considered FN. On the other hand, if it misclassifies a negative class as any positive class (Raven as GE), that is considered FP. Finally, if viNet misclassifies a negative class as another negative class, that is considered FN (Raven as Hawk). Correctly classified positive and negative classes are captured by TP and TN rates respectively. This same process can be applied to assess Curtail, No-Curtail call performance where curtail and no-curtail are treated as binary classes representing protected and unprotected species respectively