Biol Invasions (2018) 20:2485–2503
https://doi.org/10.1007/s10530-018-1715-2

ORIGINAL PAPER

Evaluation of machine learning methods for predicting eradication of aquatic invasive species

Yanyu Xiao · Russell Greiner · Mark A. Lewis

Received: 4 February 2017 / Accepted: 19 March 2018 / Published online: 27 March 2018

Abstract In the work, we evaluate the performance of machine learning approaches for predicting successful eradication of aquatic invasive species (AIS) and assess the extent to which eradication of an invasive species depends on the certain specified ecological features of the target ecosystem and/or features that characterize the planned intervention. We studied the outcomes of 143 planned attempts for eradicating AIS, where each attempt was described by ecological and eradication-strategy-related features of the target ecosystem. We considered several machine learning approaches to determine whether one could produce a classifier that accurately predicts weather an invasive species will be eradicated. To assess each learner's performance, we examined its tenfold cross-validated prediction accuracy as well as the false positive rate, the F-measure, and the Area Under the ROC Curve. We also used Kaplan–Meier survival analysis to determine which features are relevant to predicting the time required for each eradication program. Across the five typical machine learning approaches, our analysis suggests that learners trained by the decision tree work well, and have the best performance. In particular, by examining the trained decision tree model, we found that if an occupied area was not large and/or containments of AIS dispersal were employed, the eradication of AIS was likely to be successful. We also trained decision tree models over only the ecological features and found that their performances were comparable with that of models trained using all features. As our trained decision tree models are accurate, decision makers can use them to estimate the result of the proposed actions before they commit to which specific strategy should be applied.

Keywords Aquatic species · Machine learning · Survival analysis · Ecological features · Planned intervention

Y. Xiao (✉)
Department of Mathematical Sciences, University of Cincinnati, Cincinnati, OH 45221-0025, USA
e-mail: yanyu.xiao@uc.edu

R. Greiner
Department of Computing Science, University of Alberta, Edmonton, AB T6G 2E8, Canada
e-mail: rgreiner@ualberta.ca

M. A. Lewis
Center for Mathematical Biology, Department of Mathematical and Statistical Sciences, University of Alberta, Edmonton, AB T6G 2G1, Canada
e-mail: mark.lewis@ualberta.ca

M. A. Lewis
Department of Biological Sciences, University of Alberta, Edmonton, AB T6G 2E9, Canada

Introduction

Non-indigenous species are one of the greatest threats to the biodiversity of ecosystems (Cambray 2003;

Houlahan and Findlay 2004). Due to growth of commerce, recreation, and global transportation, the rate of introduction and establishment of aquatic invasive species (AIS) worldwide is likely to increase in coming years. These invasive alien species may radically alter local ecosystems, rendering the habitats inhospitable to native species. The establishments of aquatic invaders, such as Asian carp, zebra mussels and round goby in North America, have had devastating impacts on native ecosystems (Kolar and Lodge 2001, 2002; Ricciardi et al. 1998; Gurevitch and Padilla 2004). Existing and imminent threats from AIS motivate conservationists to create management strategies to eradicate these invaders. Current strategies often use information from managers' experience and/or from experts' advice to create and set policies to prevent, control, and eradicate invasive species (Pullin et al. 2004). Unfortunately, this approach is not always informed by systematic review of the outcomes of previous actions (Drolet et al. 2014). In addition, if one can use only ecological features, such as taxonomy, status of invaders, habitat, etc., to train risk prediction models, it is possible to predict the survival of AIS in an target ecosystem even before an eradication attempt. This can be useful in the management of AIS.

Many scientific approaches have been used to assess the feasibility of eradicating AIS and to determine primary factors leading to successful eradication. From a mechanistic perspective, Raymond et al. (2011) adopted a mathematical model, based on differential equations, to predict outcome of eradication campaigns on three pests—rabbits, rats, and mice—from the Macquarie Island. From a statistical perspective, Drolet et al. (2014) developed a user-friendly tool, 'Model Informing Probability of Eradication of aquatic non-indigenous species' (MIPE), that used logistic regression to predict the probability that a planned intervention of AIS will succeed, and to prioritize ecological and eradication-strategy-related features that should be collected/selected first, to increase the reliability of model estimation. The performance of the model was shown to surpass predictions based on the advice from experts (Drolet et al. 2015). Recently, machine learning approaches have been applied to ecological problems. They have been widely adopted to identify the complex structure of datasets, and to train risk prediction models in ecology (Fielding 1999; Olden et al. 2008). Bayesian belief networks (Boets et al. 2015) and decision trees (Reichard and Hamilton 1997) have been used to classify invaders by the level of invasiveness (for alien macro-invertebrates and plants in North America, respectively). Artificial neural networks, systematically discussed by Olden and Jackson (2002) in the application of ecological modelling, have been applied to monitor and predict the density of invasive species, and also have been used as a tool to suggest eradication strategies (Pu et al. 2008; Lek and Guacgan 1999). Drake et al. (2015) successfully used the classification trees to predict the risk of jeopardizing local environment by anthropegenic release of fish (including some AIS) into the wild.

Another way to access the efficiency of eradication attempts is to observe the time that an invasive species will continue to survive under an eradication intervention. Survival analysis tools, such as Kaplan–Meier (K–M) curves (Cox and Oakes 1984; Lawless 2002; Kleinbaum and Klein 2005) and log-rank tests (Mantel and Haenszel 1959; Kleinbaum and Klein 2005), have been widely used to deal with survival times in medical disciplines. Others have adopted this type of survival analysis to estimate the survival rate of invasive species. For example, K–M curves and log-rank statistics have been used to estimate and compare the survival of groups of invasive Argentine ant population under varying climate conditions (Cooling et al. 2011) and the survival rates of both invasive and native ladybird species at different temperatures (Barahona-Segovia et al. 2015). Also, Nagar and Shenkar (2016) have used the survival curves to reveal the sensitivity of the aquatic invader, *Microcosmus exasperatus*, to three different salinities and temperatures (Nagar and Shenkar 2016).

The goal of this paper is to develop and test an accurate and interpretable model that can predict the outcome of eradication attempts. We use several machine learning algorithms to train classifiers—here, artificial neural networks (ANN), decision trees (DT), logistic regression (LR), naive Bayes (NB) and support vector machines (SVM)—to predict the success of eradication. Each instance is described by six features: four are ecological features, describing the intrinsic characteristics of both target ecosystem and invader; and the two are related to eradication strategies. To assess model performance, we examine the prediction accuracy as well as false positive rate, F-measure, and the Area Under the Curve (AUC) for each of the five algorithms. ("Appendix 1" explains these terms.)

We show that DT models are significantly more accurate than other models. Moreover, DT models can provide a clear and easy-to-read model—which matches our secondary goal of interpretability. This means that the trained model can be used as a decision support tool to help decision-makers estimate the effectiveness of a proposed approach to eradicate AIS, which can be used to decide whether to commit to the approach. We further train a DT model based on only *ecological features*; this simpler system has an accuracy of 80.40%. To evaluate the performance of eradication strategies, we also train a third set of models using only the features that characterize the eradication attempt—here, *Containment* and *Method of eradication*. In addition, We perform survival analysis (K–M) to identify features that help determine the effectiveness of the eradication action. Finally, we use DT models to further predict the outcome of a planned eradication attempt in a fixed time period.

Data and methods

Study data

We adopted the dataset from the Drolet et al. (2014) paper, which was collected from hundreds of relevant articles that described 143 attempts of eradicating AIS from many aquatic target ecosystems. The dataset records observations of six of features for each attempt, including four ecological features of the invasive species itself and two features related to planned interventions. The ecological features of AIS are: (a) the **taxonomy** of the target species (plant/algae, invertebrate, or vertebrate); (b) the **status** of an invading species ('introduced' if no evidence of reproduction, 'established' if reproducing, or 'invasive' if reproducing and causing economic or ecological harm); (c) the **habitat** or the type of ecosystem (marine intertidal, marine subtidal, river/stream, or lake/pond); and (d) the **area** occupied by the invading population and treated in an eradication attempt (unit: m^2; we use $\log_2(area)$ for better visual presentations only); see Fig. 1a–d. Here, features (a) and (b) describe invaders, while (c) and (d) describe environmental factors. The remaining two features describe planned interventions: (e) the **method of eradication attempts** used (mechanical, chemical, biological or a combination of methods) and (f) **containment** ('yes' if some actions were taken to prevent natural or anthropogenic dispersal to or from the target areas, and 'no' if no action was taken); see Fig. 1e, f. In each eradication attempt, we considered two possible outcomes: 'Success' or 'Failure', where 'Success' means that no individuals of AIS were detected during surveys conducted after the eradication intervention. If multiple independent interventions were reported in the same article or report (e.g., arising from different water bodies or timeline), they were recorded as separate instances. Drolet et al. (2014) also records the duration of each attempt, which was used as an input feature. However, since the duration is not known when the attempt is initiated, we do not consider it as an input feature, but rather as an outcome, along with the label that specifies the outcome of attempts, see Fig. 2.

Machine learning algorithms

Our goal is to produce classifiers that take, as inputs, features of each eradication attempt, and predict whether that AIS in that target ecosystem would be eliminated. We explored ways to 'train' such a classifier based on the entire set of observations from our dataset with labeled outcomes, using standard machine learning techniques. We applied five machine learning algorithms to our training dataset to search for patterns in features (both characteristics of the target ecosystem and eradication strategies) that explain the success of eradication of AIS and lead to classifiers that can accurately predict the chance of success. (The next subsection discusses how to validate such trained classier.)

Below, we give brief descriptions and references for these learning algorithms. For notations, we use $\mathbf{x} = [x_1, x_2, \ldots, x_6]$ to represent the values of the features describing an instance/attempt, and $c \in \{\text{Success}, \text{Failure}\}$ to refer to the class label. For probability models, we used 'Success': prob ≥ 0.5; 'Failure': prob < 0.5.

Naive Bayes (NB) Naive Bayes is a probabilistic linear classifier that is popular in biological applications (Mitchell 1997; Keller et al. 2011). In general, the posterior probability $P(c|\mathbf{x})$ of each class label c, given the values of the features $\mathbf{x}$, is

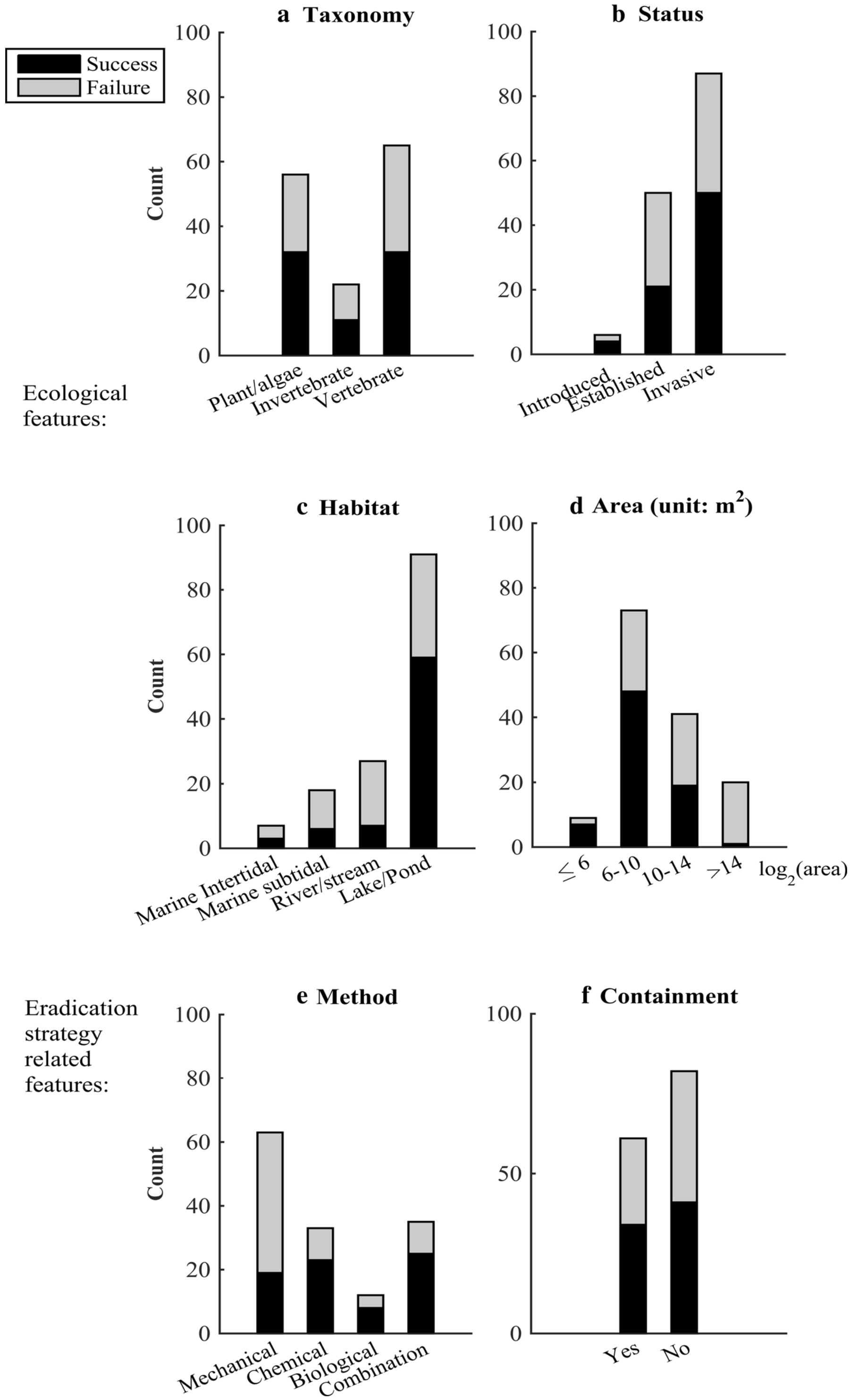
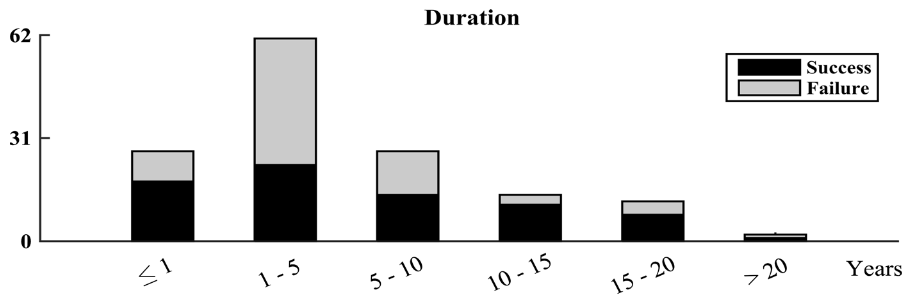


Fig. 1 **a–d** The bar plots for the four features that characterize AIS and target ecosystem; **e**, **f** the two features related to the eradication strategies. The height in each column reflects the number of instances, and the two colors represent the outcome of planned interventions, 'Success' (dark) or 'Failure' (grey)

Fig. 2 Feature: duration of eradication attempts



$$P(c|\mathbf{x}) = \frac{P(\mathbf{x}|c)\ P(c)}{P(\mathbf{x})}. \tag{1}$$

A Naive Bayes classifier assumes that, for each instance, the value of one feature is conditionally independent of that of the other features, given the class label—that is

$$P(\mathbf{x}|c) = \prod_i P(x_i|c). \tag{2}$$

After computing $P(c = \text{Success}|\mathbf{x})$, the classifier then returns 'Success' if $P(c = \text{Success}|\mathbf{x})$ is above 0.5, and 'Failure' otherwise. The learning process involves estimating the values of $P(c)$ and $P(x_i|c)$ for each value of c and x_i (for each $i = 1, 2, \ldots, n$[1]). This classifier is relatively insensitive to irrelevant features that are independent of the class label. When the assumption of independence, Eq. (2), holds, the Naive Bayes classifier often performs better than other models.

Logistic regression (LR) Logistic regression is a generalized linear model that takes the form,

$$\mathrm{LR}(\mathbf{x}; \mathbf{b}) = \frac{1}{1 + e^{-b_0 - \sum_{i=1}^{n} b_i x_i}}, \tag{3}$$

where $\mathbf{b} = [b_0, b_1, \ldots, b_n] \in \Re^{n+1}$ is the vector for the constant term and coefficients of n input variables, $x_1, x_2, \ldots, x_n$. The learning process involves computing the appropriate values of $\mathbf{b}$, which maximize the log likelihood. This model forms the basis of MIPE, the assessment tool developed by Drolet et al. (2014) that was previously applied to this dataset.

Artificial neural networks (ANNs) Artificial neural networks are computational models motivated loosely by biological neural networks and widely used in medical science and control. Neural networks are typically organized by layers of nodes, with the first layer corresponding to the input values (here, the six features defining an instance, $\mathbf{x}$), and the final layer being a single node, corresponding to the class label. In between are one or more 'hidden layers'. The learning algorithm trains the values of weights connecting the output of one node to the input of a succeeding node, which in our case corresponds to a logistic regression function, like (3); see (Lawrence 2005).

Support vector machines (SVMs) This SVM algorithm is popular for learning classifiers in bioinformatics (Keller et al. 2011). An SVM learner seeks 'optimal' hyperplanes for categorizing new instances. It can often perform well even if the training data is not linearly separable in the base feature space, by using some nonlinear kernels. Here, we used a polynomial kernel.

Decision trees (DTs) Decision tree learners (Breiman et al. 1984; Quinlan 1993; Hall et al. 2009) produce a classifier represented by a tree where each internal node corresponds to a feature and each leaf node corresponds to a class label (here, Success or Failure). Each arc descending from an internal node corresponding to feature x_i is labeled with one value of that feature x_i. Jumping ahead to Fig. 4, one can observe that there are four arcs descending from the intermediate node labeled $x_3 \equiv$ '*Habitat*', one for each of its values. An instance with *Habitat* = Marine SubTidal would follow the first of these arcs. In this way, an instance $\mathbf{x}$ will traverse a tree, from the initial node (root) to a leaf node; the decision tree will then assign that instance the label associated with that leaf node. This machine learning approach has been previously applied for risk assessment in invasion biology (Kolar and Lodge 2001; Keller et al. 2011). We used C4.5, a classification tree algorithm developed by Quinlan (1993), to train our DT models. This

[1] Here, $n = 6$ as we are considering 6 features.

algorithm uses information gain ratio as the splitting criteria and error-based pruning during the training process.

With the consideration of the complexity of models and efficiency of training processes, the ANN model we trained has at most nine nodes in one hidden layer, the SVM model was trained by selecting a polynomial kernel with the highest degree up to five, and the DT model was pruned based on error to avoid overfitting.

Model validation and evaluation

We first ran each of our five learners on the entire dataset, to produce five classifiers. Then, to evaluate the performance of each classifier, we ran tenfold cross-validation, for each of the five learning algorithms. Here, we randomly partitioned the dataset into ten roughly equal-sized subsets with balanced class labels in each subset. Of the ten subsets, we used nine subsets to train each of the classifiers, and the remaining subset to test the resulting model. The training process was repeated ten times, such that each subset was used exactly once as the validation data. We evaluated each classifier, denoting by C, on each held-out dataset S, by

$$\text{Accuracy}(C, S) = \frac{TP + TN}{TP + FP + FN + TN}. \quad (4)$$

where TP, FP, FN, TN come from the following confusion matrix.

		Prediction	
		Success	Failure
Truth	Success	TP	FN
	Failure	FP	TN

We also applied paired t test (McDonald 2014) to determine whether accuracy of the five classifiers are significantly different. In addition, we checked other statistics of model performances, such as precision, recall, and F-measure. "Appendix 1" provides detailed definitions and formulae for these criteria.

Survival analysis and censored data

As mentioned earlier, we considered 'the duration of an eradication attempt' in a target ecosystem as an *outcome variable* rather than an input. If the eradication is successful, the duration will be considered as the time required to claim the eradication of the AIS.[2] However, if an eradication attempt failed, then the recorded duration time is an underestimate of the time required for eradication (which may be in finite, if this eradication never happens).

We therefore applied survival analysis, where we consider the duration time to be "right censored" if an attempt is a failure. We want to calculate the probability that an invasive species will survive for at least T years—i.e., $P(\textit{time until Eradication} \geq T)$. We adopted standard Kaplan–Meier survival analysis to display these probabilities for various classes of instances—here, defined based on values of certain features (e.g., is the eradication time longer for vertebrate or for invertebrate, ceteris paribus.). We then ran the log rank test to see whether each value of feature (individually) made a significant difference in AIS eradications (Mantel 1966; Peto and Peto 1972). "Appendix 2" describes how to calculate these Kaplan–Meier survival curves. When K–M curves cross (for example, survival curves shown in Fig. 7b–d), we run the Kolmogorov–Smirnov test (K–S test) (Massey 1951; Miller 1956) to determine if two curves differ significantly (Klein and Moeschberger 1997).

Based on the duration of the attempts/the time required for eradication, we also examined the efficiency of intervention during fixed periods—e.g. whether a species could be eradicated in 1 year? The right-censoring problem makes this challenging: recall that an attempt, whose outcome is 'Failure', is considered censored. Therefore, we created a new dataset (named '1-year survival data') by the following two steps. We first excluded attempts whose outcomes were 'Failure' and durations were less than 1 year, as we do not know whether the AIS were eradicated by 1 year. Next, we labeled attempts whose

[2] Some successful eradication attempts had records of several annual follow-up surveys at the end of the attempts (Rowe and Champion 1994; Akers 2009). (This is because confirmations of some species being eradicated may need several years of continuous observations on target ecosystems and assessments on the trade-offs arising in any decisions.) Here, we defined these recorded durations as the time of that final follow-up survey—i.e., as time required to confirm the eradication of the AIS. For the other successful trials, without records of follow-up surveys, we set the recorded time as the eradication time.

durations were greater than 1 year as ‘Failure’ (fail-1 year, independent of the original labels) as we know that the AIS was not eradicated within 1 year. Finally, we labeled every attempt that was originally labeled ‘Success’ and with durations no more than 1 year as ‘Success’ (success-1 year). We ran machine learners on this modified dataset (same six features, but with labels of “1-year success/failure”), to produce a predictor of 1-year survivorship of AIS. We also modified our dataset to train a 5-year survival (resp., 10-year survival) dataset and to produce 5-year survival (resp., 10-year survival) models.

Data preparation and model comparison

In our analysis, we converted all nominal features to one hot encoding (binary form, e.g., the three categories of feature ‘Taxonomy’ are coded as 001, 010, 100) to train the five machine learning models (NB, LR, ANN, SVM, DT). We undertook several tasks. First, we took all six features as input to train different classifiers and compared their performances. Next, we applied only the four features that characterize the target ecosystem (in Fig. 1a–d) along with the outcome labels, to train DT models. This helped identify the importance of the four characteristics of target ecosystems in successful eradication of AIS. We also trained DT models using only the two features related to strategies. Finally, we used the k-year survival dataset to train DT models that predict the outcome of an eradication attempt during fixed periods (i.e., $k = 1, 5, 10$). Table 1 summarizes the features and outcomes involved in each analysis.

Result

Performance of machine learning models

We trained models that predict the outcome of an eradication attempt, using five different machine learning models. The performances of these models trained are displayed in Table 2; we see that DT has the largest tenfold cross-validated accuracy. Figure 3 shows the median and 1.5 interquartile range of accuracy of the models.

Since the statistics presented in Table 2 shows that the DT model surpasses the other models, we only present the detailed results of DT models in the following situations.

		Prediction	
		Success	Failure
Truth	Success	61	14
	Failure	18	38

Also, as a quick comment on all of the DT models presented in the following figures: each leaf node includes a pair of numbers—e.g. the “(2, 29)” shown in Fig. 4 under its far right node—represents the number of succeeded or failed attempts. As these reflect the results on the *training set* (rather than a held-out set), these values only suggest the performance of this part of the tree; *n.b.*, they are not part of the evaluation itself. The performance statistics for the DT model displayed in Fig. 4 is based on the confusion matrix below.

Table 1 Features and outcomes of eradication attempts used for various tasks

	Features		Outcomes	
	Ecological (Fig. 1a–d)	Strategic (Fig. 1e, f)	Duration (Fig. 2)	Label
Predict from all features (Fig. 4)	✔	✔		✔
Predict from ecological features (Fig. 5)	✔			✔
Predict from strategic features		✔		✔
Survival analysis (Fig. 7)	✔	✔	✔	✔

Table 2 Comparisons of model performances

Method	Accuracy (%)	AUC	F-measure	Precision	Recall
NB	65.03	0.76	0.62	0.69	0.65
LR	71.33	0.81	0.71	0.71	0.71
ANN	69.23	0.78	0.69	0.69	0.69
SVM	70.63	0.71	0.71	0.71	0.71
DT	**77.62**	0.75	0.79	0.77	0.81

Of all 143 attempted eradication trials, 52.45% succeeded. The formulae for Accuracy (Eq. 4), AUC, F-measure, Precision and Recall are listed in "Appendix 1"

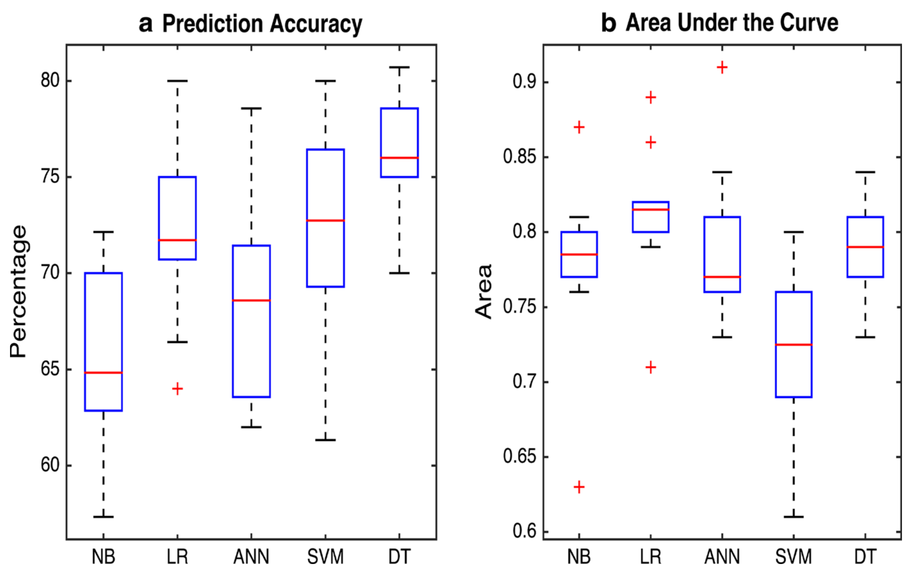


Fig. 3 Comparison of **a** prediction accuracy and **b** area under the curve, for NB, LR, ANN, SVM, DT models. The top and bottom of each box are the 25th and 75th percentiles. Lines extending vertically from the boxes (whiskers) indicate variability outside the 95th and 5th percentiles. The red line inside the box is the median and outliers are marked by '+'s

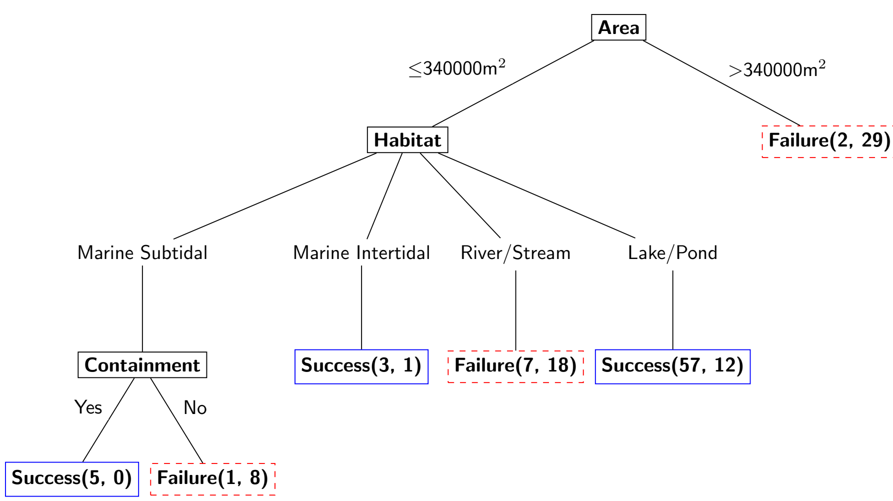


Fig. 4 The decision tree model trained over all six features (in the following context, we name it the "six-feature model"). Each leaf node is labeled '{Success, Failure} (n_s, n_f)', where n_s is the number of successful eradication attempts (in the training set) that reach this node, and n_f is the number of failed attempts

Ecological features

Our analysis suggests that this DT model, displayed in Fig. 4, can effectively evaluate the feasibility of an eradication attempt on AIS. We now explore the other applications of DT models in AIS management. Recall, there are two groups of input features: the first group includes ecological features describing intrinsic characteristics of AIS (Fig. 1a–d) and the second group characterize eradication strategic features (Fig. 1e, f).

We learned a decision tree using only ecological features; this produced the DT model, shown in Fig. 5, which is as good as the model trained by all six features in Fig. 4; see Table 3. The result of a paired *t* test statistically shows that there is virtually no difference between the two models; see also Fig. 6. To further confirm that training a decision tree model with only ecological features is feasible, we ran several statistical correlation tests. Here, we found that the strategy-related features ('Method' and 'Containment') were highly correlated with the ecological feature 'Taxonomy' (correlation coefficients: -0.5638 and 0.4493). This may because managers implicitly used the taxonomy information to determine eradication strategies.

We also trained a decision tree model using only the two features related to strategies, leading to the "two-feature model". The performance of this model does not surpass the six- and four-feature models, so we only present the statistics of its performance in Table 3.

Survival analysis

We ran survival analysis tests on our dataset to explore the roles of these features on the survival of AIS under a planned intervention. Figure 7 presents the Kaplan–Meier survival curves with respect to all six features; see "Appendix 2" for detailed calculations. Regarding the ecological features, Fig. 7a suggests that the eradication of invasive plants (Taxonomy = Plant) is different from that of invasive invertebrates and vertebrates; this is confirmed by the log-rank tests on their survival curves ($\max\{p\} = \max\{p_{plant,vertebrate}, p_{plant,invertebrate}\} \leq 0.0008$. We observed that it took longer to lower the survival probability of invasive plants under an intervention. Figure 7c suggests that AIS living in lakes or ponds have a different survival curve compared to those living in other habitats. We also performed K–S test on the survival curves sorted by habitat and get $\max\{p\} \leq 0.0008$. Figure 7d shows the distinctions between the eradications of invasive species with feature *Area* below versus above 2^{10} m^2—i.e., for $\log_2(\text{area}) \leq 10$ versus > 10, $\max\{p\} \leq 0.025$ by the log-rank test.

Regarding the strategy-related features, Fig. 7e deals with 'Method', showing that the survival curve related to chemical methods is significantly different from that of other three ($\max\{p\} \leq 0.02$ by the log-rank test). The K–M curves show that chemical

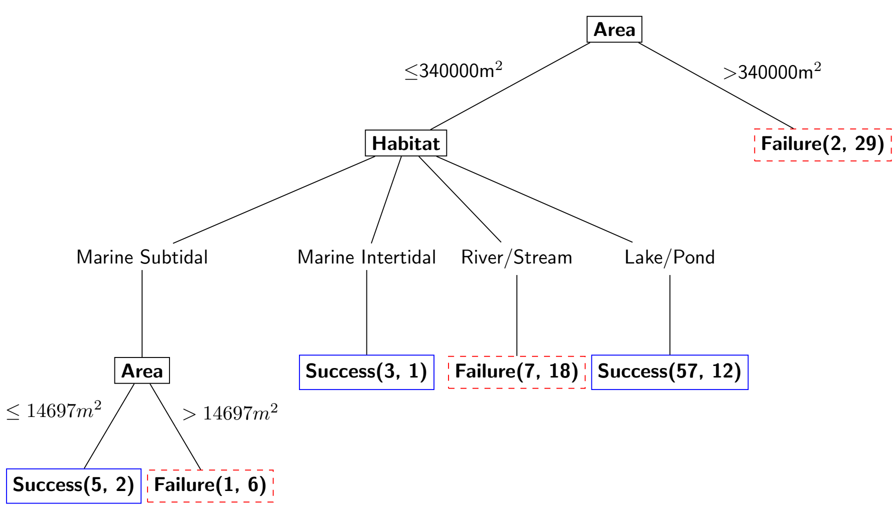


Fig. 5 Decision tree model trained on only the four ecological features (four-feature model)

Table 3 Performance statistics for the six-, four- and two-feature models

Features	Accuracy (%)	AUC	F-measure	TPR	TNR
Six-feature model (Fig. 4)	77.62 + 2.31	0.75	0.78	0.81	0.74
Four -feature model (Fig. 5)	80.40 + 2.63	0.75	0.80	0.83	0.78
Two-feature model	69.93 + 2.16	0.70	0.80	0.75	0.65

For accuracy, it shows mean ± standard deviation, based on the 10 values obtained in the tenfold cross validation. TPR (TNR) abbreviates for the true positive rate of 'Success' ('Failure') class

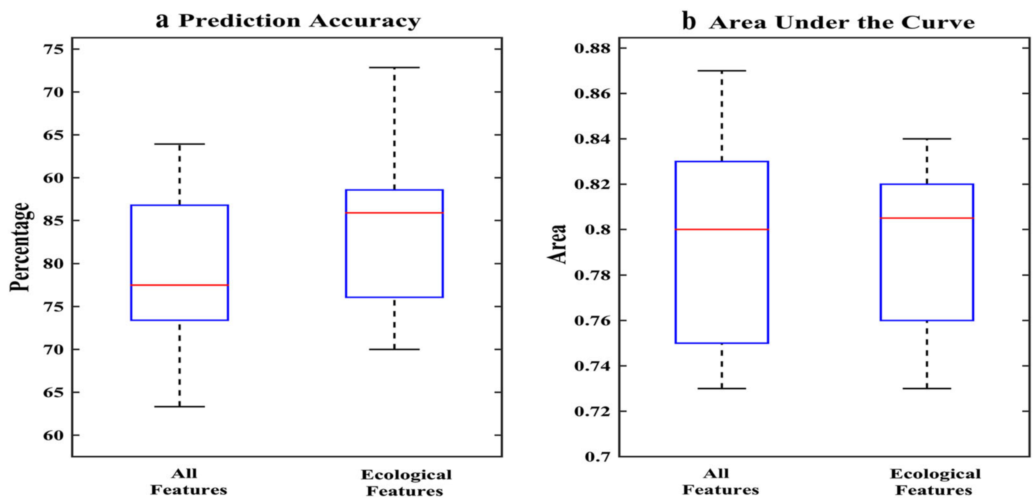


Fig. 6 Comparisons of the performance of six- and four-feature models

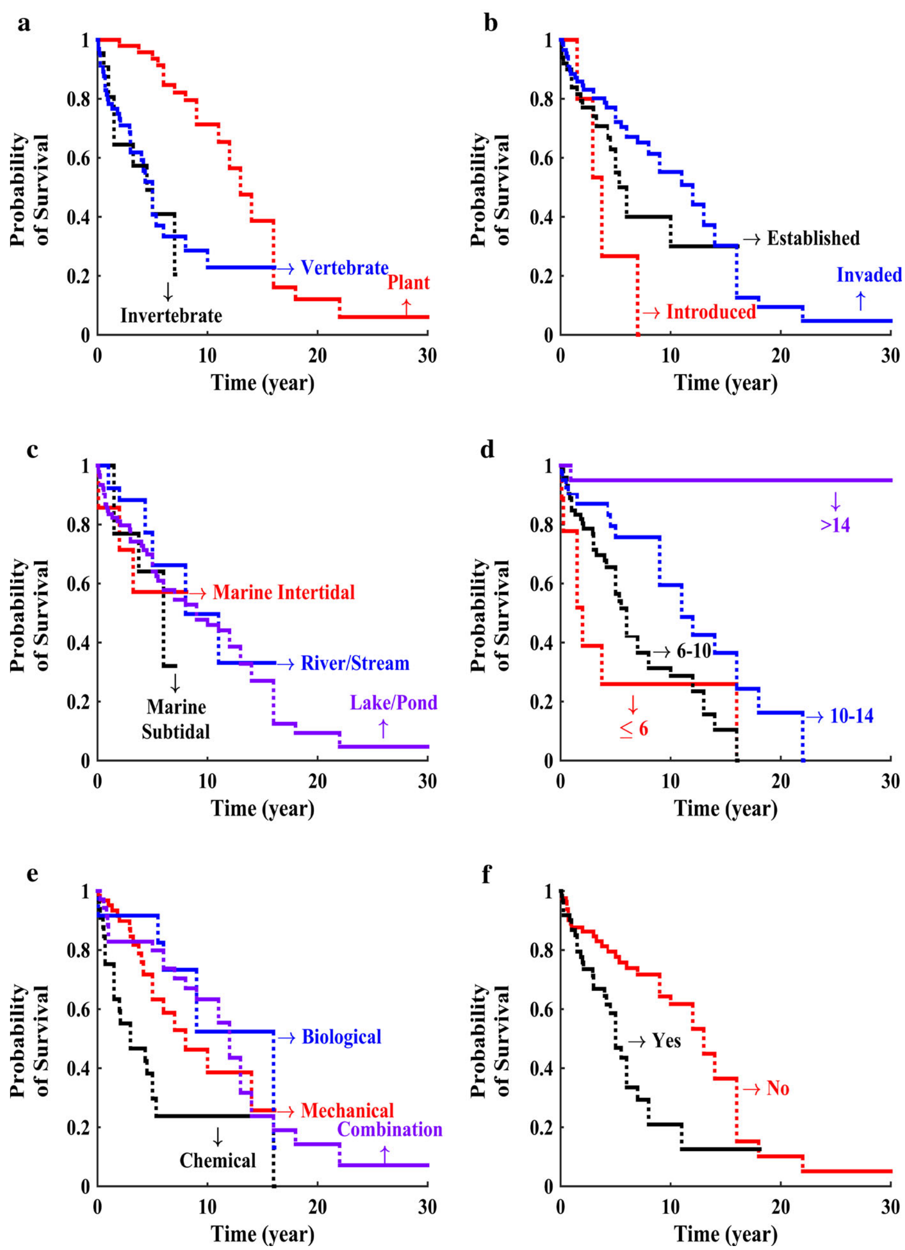
methods eradicated AIS faster and more efficiently than other methods. The survival curve related to combined methods (Method = Combination) also differs from that of the other three ($\max\{p\} \leq 0.008$ by the K–S test). As suggested by Fig. 7f, containments (Containment = Yes) played an important role in the successful eradication of AIS ($p = 0.00097$ by the log-rank test).

Finally, we evaluated the outcome of attempts after fixed durations of interventions—here of 1, 5 and 10 years. The probabilities of attempts being successful were 12.86, 41, and 65.48%, correspondingly. We then trained DT models, on the three modified datasets, to predict the outcome of attempts after these fixed durations leading to 1-, 5- and 10-year models; see Figs. 8, 9 and 10. The tenfold cross-validated prediction accuracy (AUC) for these trees were 91.43% (0.89), 81.00% (0.81) and 82.14% (0.67). We further trained a DT model based on the two strategy-related features and modified labels for 5-year survival data (see Fig. 11). This model predicted (with 76% accuracy) whether AIS in target ecosystems is successfully eradicated after a 5-year intervention based on a selected eradication strategy.

Discussion

We applied the six features, including four that were ecological (i.e., taxonomy, habitat, status and area) and two that were strategy-related (i.e., method and containment), to the 143 recorded eradication attempts from an existing data set from Drolet et al. (2014) as so to train classifier and to assess the feasibility of producing a model for eradications of AIS. We first used all six features to train five machine learning classifiers. The statistics presented in Table 2 and Fig. 3 show that the six-feature DT model is more accurate than the other

Fig. 7 Survival curves for invasive species under eradication attempt by **a** taxonomy; **b** status; **c** habitat; **d** area ($\log_2$); **e** method; **f** containment



four (NB, LR, ANN and SVM). We next trained DT models on first just ecological features, and then just strategy-related features, and found that the accuracy of our four-feature model (80.40%) surpasses the six-feature model (77.62%) and the two-feature model (69.93%). This suggests that the four-feature model can be invoked by managers as a pre-evaluation tool before they commit to a specific strategy; see statistics of these models in Table 3. We also performed Kaplan–Meier survival analysis to examine the efficiency of ecological and strategy-related features on eliminating AIS. This shows the key roles of the size of occupied area and the settings of containments in eradicating AIS. Finally, we adopted DT models to predict the possibility of successful eradication of AIS after a fixed period of intervention, for varying durations.

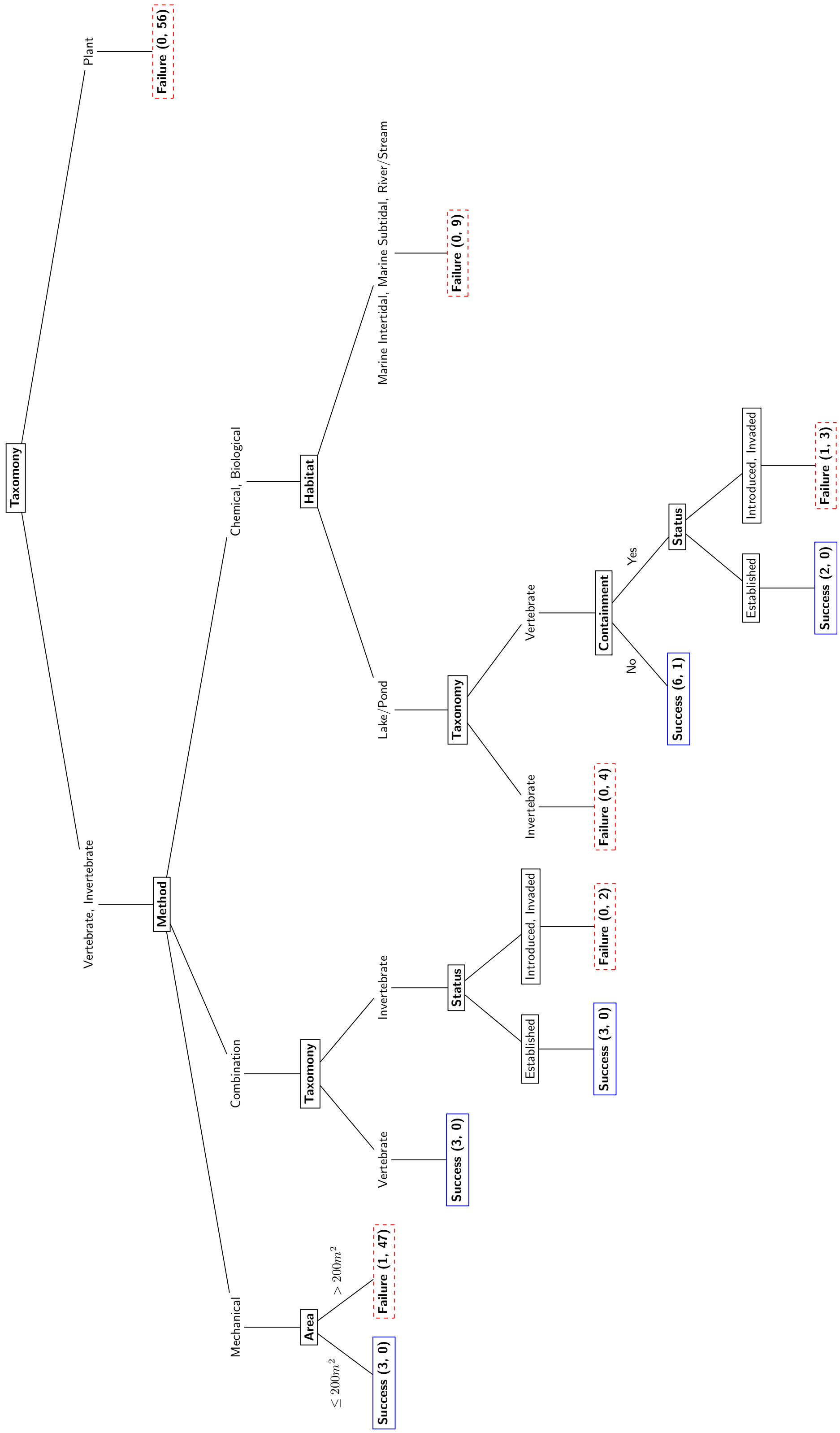


Fig. 8 The 1-year model that predicts the outcome of a 1-year eradication attempt, trained on 140 trials (18 cases of 'Success' and 122 cases of 'Failure')

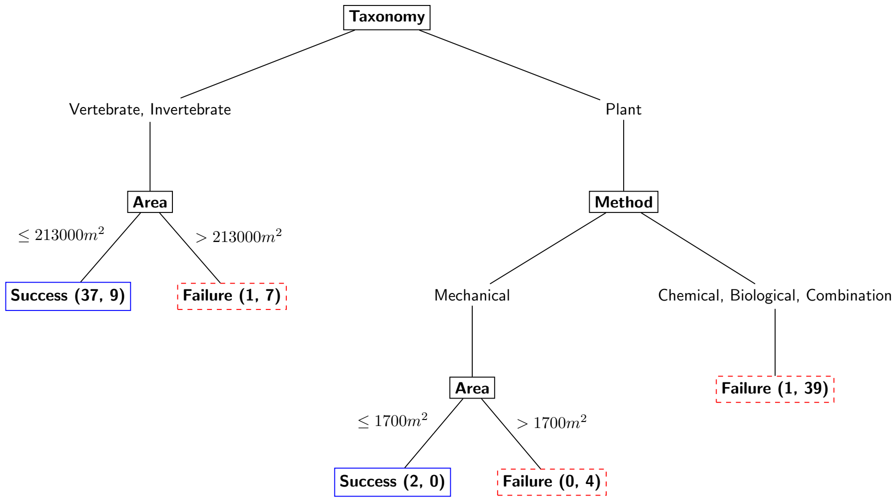


Fig. 9 The 5-year model that predicts the outcome of a 5-year eradication attempt, trained on 100 trials (41 cases of 'Success' and 59 cases of 'Failure')

Our analysis suggests that DT models work well in predicting outcomes of AIS management tasks. We did consider choosing more complicated models (e.g., more complex kernels for SVMs, or more hidden nodes for the ANN), but we chose not to do so as both would lead to models that are harder to understand, which is contrary to our secondary goal of interpretability. It is natural to compare our models, especially the LR model, with the LR model used in Drolet et al. (2014) as both were trained on the same dataset. However, as described in "Data and methods" section, we did not consider the duration of eradication programs to be a feature, as it is unknown when attempts are initiated, which means it cannot help managers make decisions at the beginning of interventions. So our trained models do not use 'duration of attempts' as an input, while Drolet et al. (2014) did. This additional information allows Drolet et al. (2014)'s LR model to have a slightly higher accuracy than our LR model; interestingly, its accuracy is still lower than our DT model.

In general, DT models have several advantages over other models. First, DT models are intuitive and can be easily interpreted. Users with little skills in mathematics/statistics/computer science can apply models to predict the outcome of new attempts by simply answering true/false and/or multiple choice questions. There are no functions, probabilities, or complicated computer codes presented in the final graphical models. Second, this learning algorithm can produce DT models that select key features that impact eradication attempts and will include only the combination of features that optimize classifications. Note that these combinations can be non-linear. Third, users do not need to pre-organize/digitize/normalize features in a dataset to avoid fitting problems that arise from scale differences between featured parameters. DT models consider various values of features as different categories and require less preprocessing of the input data. Some learners require that input features be discretized, or that they be converted to real values. For instance, we need to discretize the values of feature *Area* to several categories for effective training; SVMs needs real input values. These re-codings of feature values may impact model outcomes. Fourth, in our specific situation, the tree structures trained are reasonable in that they align with our understanding of invasion theory. The six-feature model displayed in Fig. 4 shows that the area occupied by invaders (feature: Area) and the availability of spatial dispersal (feature: Containment) play important roles in successful eradication of AIS. If an occupied area is not too large (e.g., Area $\leq$ 340,000 m^2) and/or with containments of AIS dispersal, AIS are more likely to be eliminated. Both the six-feature and the four-feature models in Figs. 4 and 5 indicate that the eradications of AIS fail when the habitat is river/stream—which is not surprising, as flowing waters may provide potential pathways that can be used by invasive species to

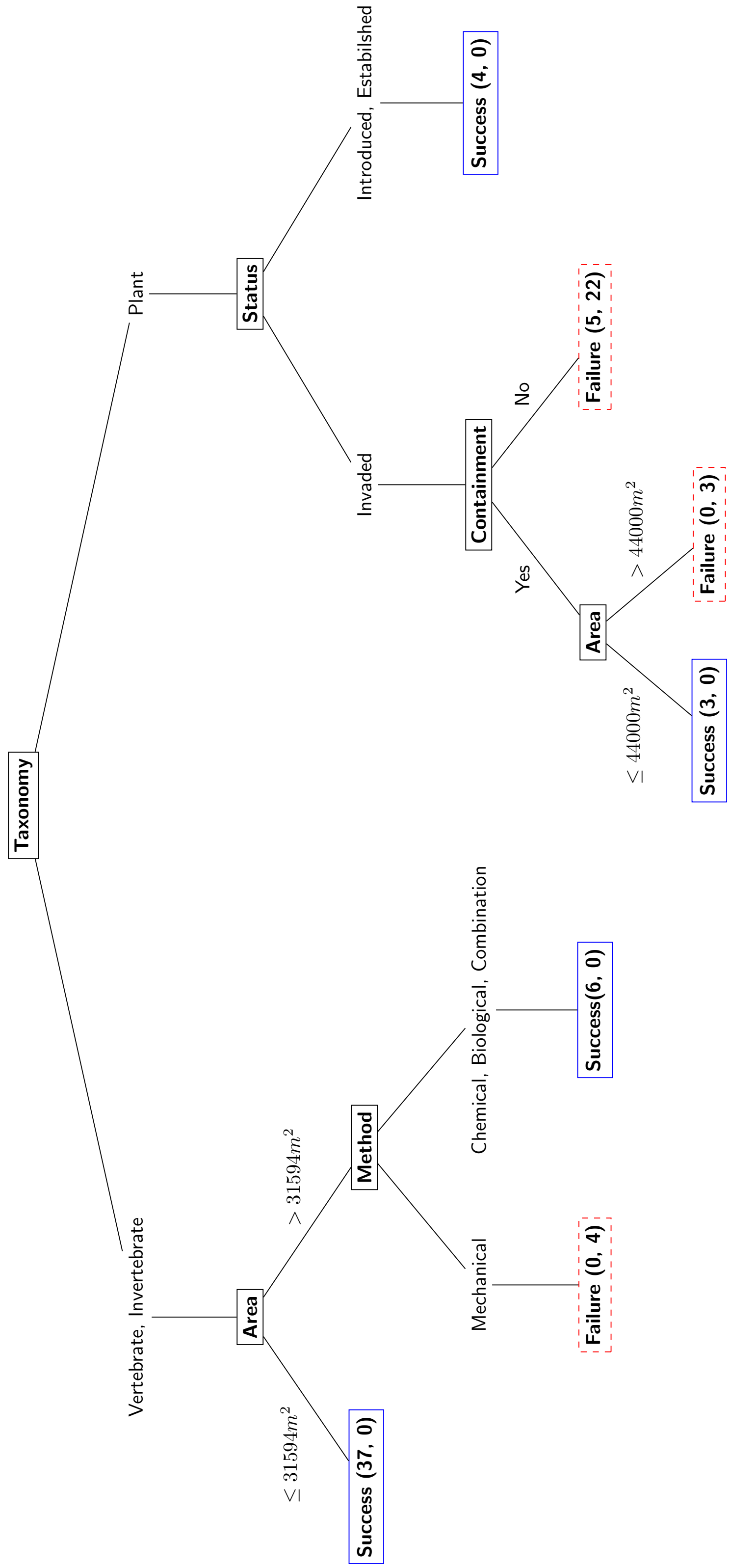


Fig. 10 The 10-year model that predicts the outcome of a 10-year eradication attempt, trained on 84 trials (55 cases of 'Success' and 29 cases of 'Failure')

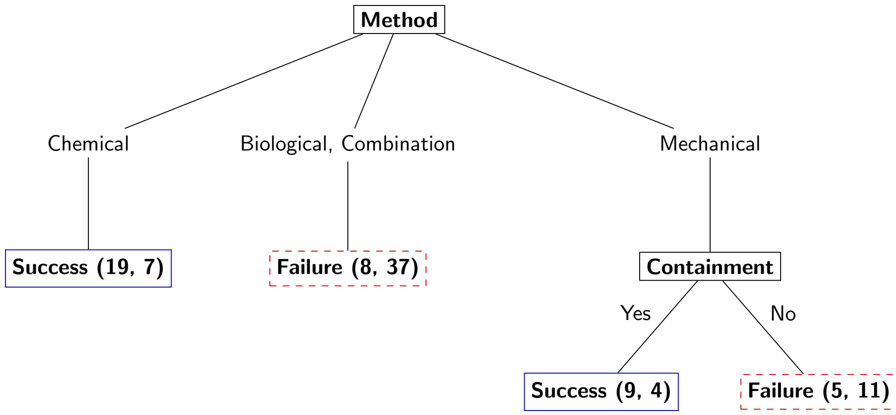


Fig. 11 The decision tree model to predict the outcome of a fixed 5-year eradication attempt with only the two strategy-related features, trained on all 100 relevant trials. The accuracy and AUC of this tree model are 76% and 0.67

expand their territory or enter new ecosystems. When there are no natural barriers and man-made barriers are difficult to set, the eradication effort would be less effective. Large areas and containments correspond to the propagule pressure and ability of dispersal, which have been previously recognized to have great impact on the eradication of AIS in invasive biology (Lockwood et al. 2005).

Among all DT models we trained, we would recommend the four-feature model (Fig. 5) to the decision-makers for predicting outcomes of planned interventions. This model can be applied to quickly access the feasibility of AIS eradications in target ecosystems before initiating interventions. For example, if the reported area of a target ecosystem is greater than 340,000 m^2, our four-feature model predicts a low chance of a successful eradication (with a training set probability of only $\frac{2}{2+29} \approx 0.065$. Note probability is in [0, 1]); otherwise, if the habitat is marine subtidal and the reported area is less than 14,697 m^2, the chance that an attempt of AIS will succeed is $\frac{5}{7} \approx 0.71\%$.

Although the model we trained based only on ecological features does predict the outcome of eradication attempts well, the features related to eradication strategies are also useful in predicting outcomes of eradications. The six-feature model presented in Fig. 4 shows that, in an attempt with Area $\leq$ 340,000 m^2 and marine subtidal habitat, the setting of containments plays an key role in successfully eradicating AIS.

Figure 7 presents the survival curves for eradication success, with respect for various features. Note that Fig. 7a–d are purely observational, describing the effects of some characteristics of the AIS or target ecosystems. This is fundamentally different from Fig. 7e, f, which describe the effects of some actions by a person, as that action might be based on observations—perhaps about ecological features. For example, while Fig. 7e does show that chemical methods lead to faster eradications than other methods, this might be simply because that managers selectively apply chemical methods to certain specific scenarios—perhaps only the situations where eradications were known to be easy. This does not necessarily mean that chemical methods would have done better in general. In addition, Fig. 7f suggests that the two survival curves are significantly different with versus without containments ($p = 0.00097$). This significance of containments may be because containments were only performed for "easy" ecosystems, where the existing natural barriers or man-made barriers are easy to build. However, consider two attempts, by Caudron and Champigneulle (2011) and Kulp and Moore (2000), where invasive trouts lived in the same habitat (river/stream), with similar reported areas (1463 and 2000 m^2), at the same status (established), and applied with same method (mechanical, more specifically, electro-fishing). The attempt with containments succeeded after 1 year of eradication effort, while the other one without containments failed after 4 years. This suggests that under certain circumstances (e.g., small areas, certain habitats), containments can increase the chance of eliminating AIS.

The correlations between ecological features and strategy-related features indicate that our dataset of actual trials is a biased subsample of all feasible combinations—for instance, the chemical methods are applied to vertebrate pests in most eradication programs (Kaukeinen 1983). Their correlations also suggest that, if attempts are predicted to be 'Failure' by our 1-, 5-, and 10-year models, decision-makers may consider avoiding methods that listed in recorded failed attempts among our dataset with similar ecological features. On the other hand, we should continually collect data after future attempts, to produce a growing dataset that covers larger range of eradication conditions. We anticipate that training on such larger, more comprehensive dataset will increase the accuracy of the resulting models (assuming users continue to follow the standard policy), especially the four-feature model.

To evaluate the impact of the duration on the outcome of attempts, we suggest 1- (resp., 5-, 10-) year model displayed in Fig. 8 (resp., Figs. 9, 10). For the 1-year eradication program, our model predicts that aquatic invasive plants are hard to eradicate: all fifty-six attempts on invasive plants were claimed to fail within 1 year. We noticed that when the duration of attempts extend from 1 year to 5 or 10 years, invasive plants are more likely to eliminate if the feature *Area* is not too large. For example, *Hydrilla verticillata* detected in a lake in California (Akers 2009) with Area $\approx 125{,}500\,\mathrm{m}^2$ (resp., reported in Deer Point Lake in Florida with Area $\approx 170{,}000\,\mathrm{m}^2$ (Van-Dyke et al. 1984) were successfully eradicated after 5- (resp., 9-) year of eradication treatments. When planning eradication programs on invasive vertebrates or invertebrates with interventions up to 10 years, our model indicates that chemical, biological and combination methods may be more effective than mechanical treatments. Our model in Fig. 11 provides some evaluations on the efficiency of eradication methods and the setting of containments under a 5-year intervention. We found that mechanical methods with containments or chemical methods have approximately a chance of 69.23 or 76.08% to eradicate AIS within 5 years, while the total chance of the other two methods is only around 17.78%.

Over our dataset, the six features including ecological and strategic features are important in AIS management, but there are many other aspects that could be part of more accurate classifiers, such as seasons at which attempts should start, costs of different eradication strategies and expertise of practitioners. For example, the four-feature model in Fig. 5 (if applied to the training data) falsely predicts the outcome of two attempts among all 31 attempts at large spatial scales to be 'Failure', even though both attempts were successful. After careful examinations on these two attempts, we suspect that the success of the eradication attempt of an invasive cyprinid (*Gila bicolor*) in Diamond Lake, Oregon (Eilers et al. 2011) is highly related to the timing when the attempt started. The eradication attempt started in winter, which seems to be a perfect time to eliminate the AIS since the reproduction rate is low and survival conditions are poor. The other successful but mistakenly predicted attempt, the eradication of *hydrilla* recorded in Akers (2009), may exaggerate the 'Area' as there was only 100 acres of waters where *hydrilla* were reported. We believe that our classifiers would use such information for better predictions, if they had access to it—i.e., our dataset includes these features.

We plan to explore the application of these useful machine learning approaches for other issues arising in invasive biology, such as planning the initiation time of attempts and evaluating the role of experts in risk assessment of AIS managements. We encourage others to consider this technology in other fields, such as ecology and epidemiology.

Acknowledgements MAL acknowledges support from a Canadian Research Chair, an NSERC Discovery Grant and a Killiam Research Fellowship. RG acknowledges support from NSERC and AMII. YX acknowledges support from the Simon foundations. We thank Boris Beric, David Drolet and Huge MacIsaac for their contribution on data collection and useful comments. This work was partially supported by the Alberta Innovates Centre for Machine Learning, the Canadian Aquatic Invasive Species Network, the Natural Sciences and Engineering Research Council of Canada.

Appendix 1: Definitions

In the main text, we used accuracy, AUC, F-measure, precision and recall to compare the performance of different machine learning algorithms (Powers 2011). Here, we will give a precise description and formula computed based on the following confusion matrix:

		Prediction	
		Success	Failure
Truth	Success	TP	FN
	Failure	FP	TN

Accuracy: The ratio of number of correctly predicted trials and the total number of trials, $\frac{TP+TN}{TP+FP+FN+TN}$.

Precision: the fraction of predicted 'Success' trials that are true: $\frac{TP}{TP+FP}$.

Recall: The fraction of successful trials that are correctly classified, $\frac{TP}{TP+FN}$.

F-measure: Harmonic mean of Precision and Recall: $\frac{2\times\text{Precision}\times\text{Recall}}{\text{Precision}+\text{Recall}}$.

AUC: The area under the receiving operating curve (ROC) for a model; here, we followed the method presented in Ferri et al. (2002) for our decision tree model and methods in Fawcett (2006) for other models.

Appendix 2: Kaplan–Meier analysis

We viewed (various subsets of our) database as 'survival data', where we set 'eradication time' to be the duration of eradication attempts and the 'censor' bit to uncensored if eradications succeeded, and to censored if the eradications failed. We then use this idea to compute a Kaplan–Meier survival curve, which produces $P(\text{Time to eradication} \geq T)$ as a function of time T (Cox and Oakes 1984; Lawless 2002; Kleinbaum and Klein 2005).

To explain this process, consider the subset of 61 instances with 'containment = *yes*'. We first sorted the durations of these instances from the shortest to the longest (total of 25 durations without repetitions); call these times: $[t_1, t_2, \ldots, t_{25}]$. At each time t_i, we defined the 'eradicated trials' for the instances whose durations were t_i and whose outcome was 'Success', and 'censored trials' for attempts with same duration but whose outcomes was 'Failure'. We also defined the number of trials at risk at time t_i to be the number of trials whose durations were no less than t_i. We used these quantities to compute the survival probability corresponding to these t_i's, which are the 25 P_i's; the curve then contains these 25 $[t_i, P_i]$ pairs; see Fig. 7a. The probability can be calculated by the following formula

$$P_i = \Pi_{i:t_i \leq t}\left(1 - \frac{d_i}{n_i}\right),$$

with d_i be the number of events and n_i be the total individuals at risk at time i. The survival probability at each time point are listed in the following table.

Time (year)	Number of eradicated trials	Number of censoring (failed trials)	Number of trials at risk	Survival probability
$t_0 = 0.00$				$P_0 = 1$
$t_1 = 0.08$	1	0	61	$P_1 = 1 - \frac{1}{61}$
$t_2 = 0.17$	1	0	60	$P_2 = P_1 \cdot (1 - \frac{1}{60})$
$t_3 = 0.25$	3	0	59	$P_3 = P_2 \cdot (1 - \frac{3}{59})$
$t_4 = 0.33$	0	1	56	$P_4 = P_3 \cdot (1 - \frac{0}{56})$
$t_5 = 0.83$	1	0	55	$P_5 = P_4 \cdot (1 - \frac{1}{55})$
$t_6 = 1.00$	2	3	54	$P_6 = P_5 \cdot (1 - \frac{2}{54})$
$t_7 = 1.33$	1	2	49	$P_7 = P_6 \cdot (1 - \frac{1}{49})$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$t_{25} = 18.00$	0	1	1	$P_{25} = P_{24} \cdot (1 - \frac{0}{1})$

References

Akers P (2009) Hydrilla eradication program progress report 2009. Technical report, California Department of Food and Agriculture

Barahona-Segovia R, Grez A, Bozinovic F (2015) Testing the hypothesis of greater eurythermality in invasive than in native ladybird species: from physiological performance to life-history strategies. Ecol Entomol 41(2):182–191

Boets P, Landuyt D, Everaert G, Broekx S, Goethals P (2015) Evaluation and comparison of data-driven and knowledge-supported Bayesian belief networks to assess the habitat suitability for alien macroinvertebrates. Environ Model Softw 74:92–103

Breiman L, Friedman J, Stone C, Olshen R (1984) Classification and regression trees. Taylor & Francis, London

Cambray J (2003) Impact on indigenous species biodiversity caused by the globalisation of alien recreational freshwater fisheries. In: Martens K (ed) Aquatic biodiversity: a celebratory volume in honour of Henri J. Dumont. Springer, Dordrecht, pp 217–230

Caudron A, Champigneulle A (2011) Multiple electrofishing as a mitigate tool for removing nonnative atlantic brown trout (*Salmo trutte* l.) threatening a native mediterranean brown trout population. Eur J Widlife Res 5(3):575–583

Cooling M, Hartley S, Sim D, Lester P (2011) The widespread collapse of an invasive species: Argentine ants (*Linepithema humile*) in New Zealand. Biol Lett 8:430–433

Cox D, Oakes D (1984) Analysis of survival data. Chapman & Hall/CRC, London

Drake D, Mercader R, Dobson T, Mandrak E (2015) Can we predict risky human behaviour involving invasive species? A case study of the release of fishes to the wild. Biol Invasions 17:309–326

Drolet D, Locke A, Lewis MA, Davidson J (2014) User-friendly and evidence-based tool to evaluate probability of eradication of aquatic non-indigenous species. J Appl Ecol 51(4):1050–1056

Drolet D, Locke A, Lewis MA, Davidson J (2015) Evidence-based tool surpasses expert opinion in predicting probability of eradication of aquatic nonindigenous species. Ecol Appl 25(2):441–450

Eilers J, Truemper H, Jackson L, Eilers B, Loomis D (2011) Eradication of an invasive cyprinid (*Gila bicolor*) to achieve water quality goals in Diamond Lake, Oregon (USA). Ecol Appl 27:194–204

Fawcett T (2006) An introduction to ROC analysis. Pattern Recognit Lett 27:861–874

Ferri C, Flach P, Hernandez-Orallo J (2002) Learning decision trees using the area under the ROC curve. In: Proceeding ICML '02 proceedings of the nineteenth international conference on machine learning, pp 139–146

Fielding A (1999) Machine learning methods for ecological applications. Springer, New York

Gurevitch J, Padilla DK (2004) Are invasive species a major cause of extinctions? Trends Ecol Evol 19(9):470–474

Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten I (2009) The weka data mining software: an update. SIGKDD Explor 11(1):10–18

Houlahan J, Findlay CS (2004) Effect of invasive plant species on temperate wetland plant diversity. Conserv Biol 18(4):1132–1138

Kaukeinen D (1983) Vertebrate pest control and management materials: fourth symposium. ASTM International, Philadelphia

Keller R, Kocev D, Dzeroski S (2011) Trait-based risk assessment for invasive species: high performance across diverse taxonomic groups, geographic ranges and machine learning/statistical tools. Divers Distrib 17(3):451–461

Klein J, Moeschberger M (1997) Survival analysis—techniques for censored and truncated data—statistics for biology and health. Springer, New York

Kleinbaum D, Klein M (2005) Survival analysis: statistics for biology and health, 2nd edn. Springer, New York

Kolar C, Lodge D (2001) Progress in invasion biology: predicting invaders. Trends Ecol Evol 16:199–204

Kolar C, Lodge DM (2002) Ecological predictions and risk assessment for alien fishes in North America. Science 298(5596):1233–1236

Kulp M, Moore S (2000) Multiple electrofishing removals for eliminating rainbow trout in a small southern appalachian stream. N Am J Fish Manag 20(1):259–266

Lawless J (2002) Statistical models and methods for lifetime data. Wiley-Interscience, Hoboken

Lawrence J (2005) Introduction to neural networks, 2nd edn. California Scientific Software Press, California

Lek S, Guacgan J (1999) Artificial neural networks as a tool in ecological modelling, an introduction. Ecol Model 120:65–73

Lockwood J, Cassey P, Blackburn T (2005) The role of propagule pressure in explaining species invasions. Trends Ecol Evol 20:223–228

Mantel N (1966) Evaluation of survival data and two new rank order statistics arising in its consideration. Cancer Chemother Rep 50(3):163–170

Mantel N, Haenszel W (1959) Statistical aspects of the analysis of data from retrospective studies of disease. J Natl Cancer Inst 22(4):719–748

Massey F (1951) The Kolmogorov–Smirnov test for goodness of fit. J Am Stat Assoc 46(253):68–78

McDonald J (2014) Handbook of biological statistics, 3rd edn. Sparky House Publishing, Baltimore

Miller L (1956) Table of percentage points of Kolmogorov statistics. J Am Stat Assoc 51(273):111–121

Mitchell T (1997) Machine learning. Mc-Graw-Hill Companies Inc, New York

Nagar L, Shenkar N (2016) Temperature and salinity sensitivity of the invasive ascidian Microcosmus exasperatus Heller, 1878. Aquat Invasions 11(1):33–43

Olden J, Jackson D (2002) Illuminating the black box; understanding variable contributions in artificial neural networks. Ecol Model 154:135–150

Olden J, Lawler J, Poff N (2008) Machine learning methods without tears: a primer for ecologists. Q Rev Biol 83(2):171–193

Peto R, Peto J (1972) Asymptotically efficient rank invariant test procedures. J R Stat Soc Ser A 135(2):185–207

Powers D (2011) Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. J Mach Learn Technol 2(1):37–63

Pu R, Gong P, Tian Y, Miao X, Carruthers R, Anderson G (2008) Invasive species change detection using artificial neural networks and CASI hyperspectral imagery. Environ Monit Assess 140(1–3):15–32

Pullin A, Knight T, Stone D, Charman K (2004) Do conservation managers use scientific evidence to support their decision-making? Biol Conserv 119:245–252

Quinlan JR (1993) C4.5: programs for machine learning. Morgan Kaufmann Publishers Inc, San Francisco

Raymond B, McInnes J, Dambacher J, Way S, Bergstrom D (2011) Qualitative modelling of invasive species eradication on subantarctic macquarie island. J Appl Ecol 48(1):181–191

Reichard S, Hamilton C (1997) Predicting invasions of woody plants introduced into North America. Conserv Biol 11(1):193–203

Ricciardi A, Neves RJ, Richard J, Rasmussen J (1998) Impending extinctions of North American freshwater mussels (Unionoida) following the zebra mussel (*Dreissena polymorpha*) invasion. J Anim Ecol 67(4):613–619

Rowe D, Champion P (1994) Biomanipulation of plants and fish to restore Lake Parkinson: a case study and its implications. In: Collier K (ed) Restoration of aquatic habitat. Selected papers from the second day of the New Zealand Limnological Society 1993 annual conference, pp 53–65

Van-Dyke J, Leslie JA, Nall L (1984) The effects of the grass carp on the aquatic macrophytes of four Florida lakes. J Aquat Plant Manag 22:87–95