ARTICLE

A machine-learning approach to predict success of a biocontrol for invasive Eurasian watermilfoil reduction

Diana T. White[1] | **Thibaud M. Antoniou[1]** | **Jonathan M. Martin[1]** |
William Kmetz[2] | **Michael R. Twiss[2]**

[1]Department of Mathematics, Clarkson University, Potsdam, New York, USA

[2]Department of Biology, Clarkson University, Potsdam, New York, USA

Correspondence
Diana T. White
Email: dtwhite@clarkson.edu

Handling Editor: Éva Elizabeth Plaganyi

Abstract

Myriophyllum spicatum, more commonly known as Eurasian watermilfoil (EWM), is one of the most invasive aquatic plants in North America, causing negative ecological and economic impacts in ecosystems where it proliferates. Many control strategies have been developed and implemented to mitigate EWM growth and spread, although the results are mixed and there is no consensus on lake-specific strategies. Here, we describe the development of a predictive model using a support vector technique, that predicts the success of biological pest control using *Euhrychiopsis lecontei* (the milfoil weevil), a milfoil specialist, to reduce EWM in lakes. Such a model is informed by lake characteristics (limnological and landscape) and augmentation strategies. To develop our predictive model, we performed a metadata analysis from 133 - published peer-reviewed literature and professional reports of milfoil weevil augmentation field experiments that contained information on lake characteristics. The predictive model's algorithm uses a support vector machine (SMV) to learn patterns among lake characteristics, along with the recorded augmentation strategy and the reported success of each study, where success is a measure of EWM change over a season and is recorded in a variety of ways (e.g., EWM biomass change, EWM percent change, EWM visual change, etc.,). Overall, the model results suggests that shallower lakes, more frequent weevil augmentations, and larger weevil overwintering habitat are the most important predictors for EWM reduction success by weevil augmentation. Although watermilfoil weevil augmentation is a promising mitigation strategy, it may not work for all lakes. However, in terms of suggesting weevil augmentation, our model is a valuable tool for lake stakeholders and resource managers, who can use it to determine whether milfoil weevil augmentation, which can be very costly due to the difficulties in finding and raising milfoil weevils, will be a useful and sustainable approach to control EWM in their lake community.

INTRODUCTION

Eurasian watermilfoil as an aquatic invasive species and its control

One of the most widespread invasive aquatic plant species throughout North America is *Myriophyllum spicatum*, more commonly known as Eurasian watermilfoil (EWM) (Les & Mehrhoff, 1999; Parsons et al., 2011; Smith & Barko, 1990). EWM is one of the many invasive varieties of watermilfoils that are believed to have been introduced to North America in the 1940s, following purposeful introduction from Europe, Asia, and North Africa, with its first reliable record documented in 1942 (Les & Mehrhoff, 1999). EWM negatively impacts lake ecosystems (Les & Mehrhoff, 1999; Smith & Barko, 1990), causing reduction of dissolved oxygen, changes in water temperature (Caspers et al., 2009), and the formation of dense monocultures that push out native aquatic plants (Smith & Barko, 1990). One of the most invasive qualities of this plant is its multiple modes of reproduction (Smith & Barko, 1990). In addition to sexual reproduction and seed dispersal, vegetative reproduction in EWM occurs by the formation of new stems at the parent root mass, and by stem fragmentation (Martin & Valentine, 2014). Stem fragmentation occurs when pieces of the parent plant break off and settle at the lake bottom where they take root and continue growth as a new plant clone. In addition to undesired ecological effects, EWM causes negative social impacts, decreasing the quantity and quality of recreational activities like boating, swimming, and fishing (Hussner et al., 2017; Schultz & Dibble, 2012), as well as economic losses due to increased cost of degraded shoreline land use values (Rockwell, 2003) and lake-front property devaluation (Olden & Tamayo, 2014).

Much experimental and field work has been done to implement control strategies to mitigate invasive watermilfoil growth and spread (Coetzee et al., 2011; Creed & Sheldon, 1995; Gross et al., 2020; Hussner et al., 2017; Laitala et al., 2012; Marko & White, 2018). Control strategies incorporate localized methods, such as hand harvesting and strategic placement of mats (Laitala et al., 2012), as well as global (i.e., lake-wide strategies) that include mechanical harvesting (Gross et al., 2020; Hussner et al., 2017), herbicidal treatment (Gross et al., 2020; Hussner et al., 2017; Marko & White, 2018), and the use of various biocontrols (McKnight & Hepp, 1995; Newman, 2004). Localized methods can be useful for small-scale removal but are costly

(in terms of time) and inefficient for large-scale removal. Global methods target invasive plants but can cause more harm to native plants than local methods, which can target just invasives. In addition, global methods can be very expensive in terms of physical cost, and herbicides may require additional permits depending on state (Wersal et al., 2010). Although some control methods have proved successful in certain lakes, many have not. Here, we address mitigation of EWM by offering insight into a sustainable control strategy using biocontrol with a native insect. This approach is sustainable in terms of being a cost- and labor-efficient method that does not cause harm to the local environment.

The sustainable control strategy we analyze is the augmentation (addition of a native species) of an infested water body with the biocontrol insect, *Euhrychiopsis lecontei*, commonly known as the milfoil weevil (Newman, 2004; Newman & Biesboer, 2000; Reeves et al., 2008; Sheldon & Creed, 1995). These weevils are native to North America and are milfoil specialists, known to preferentially target EWM (as well as other invasive, hybrid, and native species of watermilfoil) without causing damage to native plants (Creed & Sheldon, 1995). Some biocontrols can have adverse effects on native species, however previous field studies have shown that native aquatic plants are rarely used by the milfoil weevil for rearing and feeding and thus the milfoil weevil does not impact their growth (Sheldon & Creed, 1995). Adult milfoil weevils can fly, but generally remain submersed where they lay their eggs in beds of EWM and the larval instars burrow into EWM tissues (Newman, 2004). In the winter, milfoil weevils overwinter in leaf debris until spring (Thorstenson et al., 2013). Hollowed-out EWM tissues reduce the plants fitness and reduces its buoyancy (Creed & Sheldon, 1993), making colonization by fragmentation from weevil-impacted plants less of a concern. Milfoil weevils have been studied in the past as a biocontrol for EWM, where augmentation in the field has shown mixed results (Havel et al., 2017; Jester et al., 2000). In some experiments, augmentation of local weevil populations through control interventions have shown reduction in EWM biomass when augmentation targets existing stands of EWM (Creed & Sheldon, 1993, 1995; Newman, 2004; Reeves et al., 2008; Sheldon & Creed, 1995). However, other experiments have shown little to no change, and sometimes even increases in overall EWM biomass (Jester et al., 2000; Reeves et al., 2008). Most other biocontrols for EWM are either ineffective or detrimental, as in

the case of addition of *Ctenopharyngodon idella* (grass carp) (McKnight & Hepp, 1995), applications of the fungus *Mycoleptodiscus terrestris*, or addition of the native aquatic midge *Cricotopus myriophyll* (McKnight & Hepp, 1995). One limitation of milfoil weevil augmentation, and with augmentation in general, includes the fact that augmentation may need to be repeated (Michaud, 2018). For this reason, we provide details of riparian zone restoration in terms of restoration of milfoil weevil overwintering habitat (Thorstenson et al., 2013), as an additional part of the augmentation strategy to ensure natural stabilization of milfoil weevil populations.

Previous mathematical models of aquatic macrophyte growth have been developed to study both seasonal (Herb & Stefan, 2003; Miller et al., 2011) and yearly plant growth (Best et al., 2001). These predictive models are mechanistic in nature, taking into account how macrophytes are dependent on lake characteristics such as water clarity, available nutrients, water temperature, as well as the metabolic properties of the plant itself. Miller et al.'s model focuses on EWM growth, while also taking into account the effect of milfoil weevils on season biomass (Miller et al., 2011). Other models take geospatial and field data into account (plant location and lake boundary), merging this with other model variables that characterize the growth characteristics of invasive macrophytes, like EWM (Buchan & Padilla, 2000; Olson et al., 2012). In addition to the above predictive models for growth and spread, other models have been developed to provide information on the likelihood of an invasive watermilfoil invasion, based on lake and plant characteristics. For example, in (Thum & Lennon, 2010), the authors perform a Principal Component Analysis (PCA) using a correlation matrix from 17 environmental variables to identify the major environmental gradients in lakes across a given test region to gain qualitative insight into environmental characteristics of lakes occupied by invasive watermilfoil. Here, we complete an analysis of the same light, in the sense that we build a model that provides details on the probability of success of EWM reduction by weevil augmentation.

To build our predictive model of EWM reduction by weevils, we first complete a metadata analysis of all milfoil weevil augmentation studies to date, to collect information on (1) lake characteristics important for EWM growth and weevil survival and (2) weevil augmentation strategies. We then use this metadata to run a machine-learning algorithm that finds connections and patterns among lake sites that were successful at EWM reduction and those that were not. First, we discuss the methods used in data collection, which incorporates the selection of model predictors (which we call features). Model features include data we believe important for determining

weevil success at EWM reduction and weevil survival, in addition to those related to EWM growth and spread. Model features also include the augmentation strategy employed, such as how many weevils are added and how often they are added. To build our machine-learning algorithm, a support vector machine (SVM), we train the model to data based on known outcomes of success (EWM reduction or increase) called our model target, where the target is recorded in a myriad of ways (e.g., measurements of biomass, records of percent change, or visual inspection). Second, we discuss the SVM construction, which is used to find the connections between our defined features and targets. We then test our model using data that is not incorporated in the training data set. In particular, we split the data into a training set (three-fourths of the data), and test set (one-fourth of the data) randomly. This random splitting process is repeated 500 times until the model gives a relatively consistent/constant output for model accuracy (the final results of each run are averaged). Finally, we describe the results of our study, which indicates that our model can be 87% effective at predicting weevil success or failure in a given lake. Thus, any interested stakeholder who would want to consider weevils as a sustainable control strategy for their lake can implement our model and base their decision to augment a lake on the model prediction. In particular, decision making can be informed by determining the probability of success of a given augmentation strategy for a given lake.

METHODS

Metadata assembly: Literature search of peer-reviewed articles and contractor reports

The overall project goal was to predict whether an augmentation of the milfoil weevil would prove successful in controlling EWM in a given lake, given known characteristics of the lake and an augmentation strategy. As stated previously, the definition of success relies not only on how stakeholders define success, but also on lake-specific data that is publicly available. As we will show later in the results section, this second point more heavily guided us towards our definition of success. To run our machine-learning algorithm, we collected data for the set of features related to a lake's limnological and landscape characteristics, as well as the weevil augmentation strategy. The collection and sorting of this data are what we refer to as our metadata analysis. Such an analysis can be quite laborious since it requires the collection of data from a variety of sources,

including peer-reviewed articles, technical reports from industry, along with reports from local lake associations. We included all papers and reports that indicated, at minimum, EWM changes and augmentation strategies, and we removed a handful of papers and reports where a large portion of data were missing. After obtaining an initial set of data, we found that the data collected were often reported differently between studies (e.g., certain data might have different units, or be collected at different times in the year, etc.,), and organizing this data in such a way that the features are consistent such that they can be incorporated into a predictive model can be tricky. In this section, we describe the collection of such data, citing sources as well as other tools and methods used to interpret and record information. We break the data up into two categories: augmentation features and lake characteristics, where lake characteristics correspond to limnological and landscape features. Limnological characteristics include information such as lake depth, chemical composition, and Secchi depth, and landscape features provide information on the shoreline (habitable shoreline for overwintering weevils), as well as latitude, and lake area.

The metadata analysis resulted in 133 cases of milfoil weevil augmentation (these data are summarized in part in Table 1; the complete table can be found in White, 2022). Within the 133 cases, 34 lakes throughout the United States and Canada were represented.

Augmentation features included the number of weevils added, as well as the augmentation treatment frequency (corresponding to the number of times weevils are stocked). The number of weevils added was reported in one of two ways: studies either recorded the absolute number of weevils added to a milfoil bed, or they recorded the ratio of weevils added to each EWM stem. The majority of studies (114 of the 133 cases) recorded absolute number of weevils added, so this measurement was used in the model. For the remaining 19 cases, we took the average value of weevils added from all other studies, which was 5036 weevils per lake, and used this value for the number of weevils added. As stated later in the results section, we test our model using (1) all data, which includes the averaged weevil data for the 19 missing cases, and (2) only data for which the absolute value of weevils was recorded. A second model feature describing the weevil augmentation process is the treatment

TABLE 1 Table of model features and targets for 13/133 studies (full table provided in White, 2022)

Study site	Latitude	Area (ha)	Maximum depth (m)	Buffer zone (km)	Phosphorus (µg/L)	Secchi depth (m)	Treatment frequency	Augmentation (average no. weevils)	Biological success (Y = 1, N = 0)
Big Sand Lake (Jester et al., 2000)	46.06	563.2	19.8	11.3	19.0[a]	2.9[a]	1	N/A	1
Eagle Lake (Jester et al., 2000)	42.70	208.0	3.6	9.3	19.0[a]	2.4	1	N/A	1
Lower Spring Lake (Jester et al., 2000)	42.88	41.6	3.3	2.2	44.3[a]	1.3[a]	1	N/A	1
Whitewater Lake (Jester et al., 2000)	42.76	256.0	11.6	5.8	15.0[a]	1.3[a]	1	N/A	0
Nancy Lake (Jester et al., 2000)	46.09	309.8	11.9	14.2	14.6[a]	4.2[a]	1	N/A	0
Pearl Lake (Jester et al., 2000)	44.09	36.8	15.2	1.3	12.3[a]	5.8[a]	1	N/A	0
Beaver Dam Lake (Jester et al., 2000)	45.55	444.8	32.3	13.9	9.9[a]	4.0[a]	1	N/A	0
Cedar Lake (Ward & Newman, 2006)	44.96	68.0	16.0	2.7	25.0[a]	2.8, 2.8, 2.5, 2.5	1, 1, 2, 2	N/A,N/A, NA,N/A	1, 0, 0, 0
Little Bearskin (Havel et al., 2017)	45.71	74.0	8.1	3.7	20.8, 66.4, 33.1, 20.8, 66.4,33.1	1.8	1	2157,0,0, 2719,0,0	1, 0, 0, 1, 1
Manson Lake (Havel et al., 2017)	45.56	96.0	16.2	4.5	11.0, 12.6, 14.7, 11.0, 12.6, 14.7	4.8[a]	1	3013, 0, 0, 2912, 0, 0	0, 0, 0, 0, 0, 1

Note: Metadata analysis for 13/133 studies. Full table of metadata analysis shared here. Information for model target (biological success 0 = failure, 1 = success) and the eight most highly ranked model features, Lake location (name and latitude), area, maximum depth, buffer zone (measure of shoreline suitable for weevil overwintering), phosphorus (P), Secchi depth, augmentation strategy (number of weevils added), and treatment frequency.

[a]Corresponds to estimate values. Where N/A reported, average weevil number 5036 used.

frequency, defined as the number of times an EWM patch had been augmented. The number of weevils added and the treatment frequency for each of the 133 cases examined are summarized in Table 1.

Limnological features include the lake's maximum depth, as well as the features phosphorus and Secchi depth. Lake depth is an important feature, since deeper lakes inhibit plant growth as light availability decreases exponentially with depth. The maximum depth was recorded for each body of water when available. In two bays, the maximum depth was not recorded, and was estimated by using contour maps of lake depth. Specifically, the depth was calculated by averaging the greatest contour depth with the contour line adjacent to it.

The features phosphorus and Secchi depth were also included in our study and provide measures as to how habitable a lake is for EWM growth. Phosphorus (concentrations are listed as μg/L) is generally the primary limiting nutrient in most freshwaters (Cao et al., 2012). In general, phosphorus concentrations fluctuate greatly among bodies of water, and result in ranges of aquatic macrophyte and phytoplankton proliferation. Secchi depth is a simple and reliable method to measure the amount of light penetrating to a certain depth in a lake due to both water color and turbidity and is measured by placing a Secchi disk in the water and recording the depth (in m) that the disk is no longer visible. In most reports examined, Secchi depths were recorded in varying years, at varying times throughout the growing season, and sometimes not the same year as the augmentation study. In addition, these measurements were often recorded in multiple locations throughout the body of water. Whenever data was available for phosphorus and Secchi depth at multiple locations, averages were taken. If data for phosphorus concentrations or Secchi depth were missing in a particular study, then all available data (data for every study) for phosphorus and Secchi depth were averaged, and our model was run using (1) all data (including averages used for missing data), and (2) only the data where the values for phosphorus and Secchi depth were known. When all data sets with missing phosphorus, Secchi depth, and weevil numbers were removed from the model, the resulting model included a total of $n = 54$ studies. Thus, we ran our model with all 133 data sets, and with 54 data sets (the first data set we refer to as "All augments" and the second we refer to as "All augments[a]"). In future work, if more augmentation studies are reported, we will develop a model that works to fill in missing data using more comprehensive missing data schemes (Little & Rubin, 2002).

Landscape features that describe the terrain surrounding a lake were calculated using the distance drawing tool on satellite images in Google maps. Landscape features correspond to total shore length, buffer zone length, and lake surface area, where the buffer zone is defined as the perimeter of habitable land for milfoil weevils to overwinter. A buffer zone is terrestrial terrain that contains duff (decaying leaf litter and debris), typically found below shrubs and forest that affords shelter for over-wintering weevils and is 3–5 m in breadth (Thorstenson et al., 2013). Finally, lake surface area was collected from lake association databases, and when surface area could not be found from such databases, it was determined using the draw tool in Google Maps.

The final data collected from each study is the change in EWM, which we use to define the model's target of success. As stated previously, due to recorded data coming from different sources, not all EWM changes were recorded in a similar way. In particular, some reports included information on percent decreases and increases, while others gave absolute changes. In addition, most studies simply indicated if there is an increase or decrease in EWM. Therefore, when trying to determine connections between our model features and target, we tested our model using several different definitions of success, where each definition is based on how the change in EWM was recorded. These definitions resulted in five different variables for target success, where the first target included all augmentation studies, and included the qualitative statement of either EWM increase (failure) or decrease (success) ($n = 133$, between 34 lakes that included averaged data, and $n = 54$ when average data were removed). The final four targets used quantitative definitions of success and included EWM stem density measured as stems/m^2 ($n = 47$, among 11 lakes), EWM biomass change measured in g/m^2 ($n = 44$, among nine lakes), EWM percent change in mass from the beginning to the end of a given growing season ($n = 95$, among 30 lakes), and relative abundance of EWM change ($n = 50$, between 11 lakes). Relative abundance records the change in the percent difference of EWM as compared to native plants, within a 1-m^2 region. Here, both EWM and native plant density were measured in g/m^2. Since native plant density was typically only recorded at the beginning of a study, we made a simplifying assumption that native plant density would not change by the end of the study (i.e., change was based only on EWM change).

Building a machine-learning algorithm of milfoil weevil success based on weevil augmentation studies

Here we describe the methods used to predict the success of milfoil weevil augmentation at reducing EWM within a given lake. The overall goal of this modeling approach

is to provide a tool for stakeholders to predict whether weevils can be a successful control method to reduce EWM within a given water body, and if so, to describe an appropriate weevil augmentation strategy. First, we describe the machine-learning algorithm by outlining the details for how we train and validate the model. In addition, we describe the algorithm used to run our model, which uses a SVM to classify weevil augmentation studies as either success or failure.

Our model uses a supervised learning approach to predict our target, based on a set of input features. We implement a classification model, where our target takes on one of two values, success or failure, and the model features include a list of each lake's limnological and landscape characteristics, as well as the weevil augmentation strategy (i.e., the number of weevils added and the frequency of additions in a single season). The model learns (from the training sets) which combinations of lake characteristics and augmentation strategies are likely to correspond to a successful biocontrol program. The model can then use these patterns, by inputting characteristics of new lakes, to predict the success or failure of a biocontrol program for a given augmentation strategy. See Figure 1 for a description of the models utility in developing such future biocontrol programs.
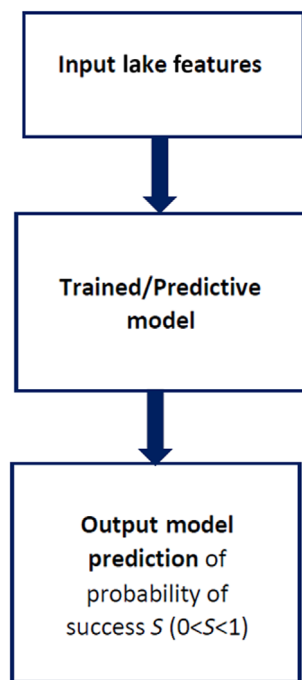
The SVM algorithm

We use an SVM to classify milfoil weevil augmentation experiments as either success or failure (Géron, 2017). SVMs are predictive algorithms that are based on statistical learning frameworks, and work to assign a classification (here our classes are the target's success or failure) to a given input feature set (a group of features with their values). To make predictions, the SVM is first trained with example feature sets with known classifications, by mapping training examples to points in feature space that maximize the distance between groups of points with different classifications (our two target sets of success or failure). Typically, a hyperplane defines a separating plane between these two target sets, where an additional margin (a gap around the hyperplane), called support vectors, exists between the two target sets. The margin can be manipulated and made smaller or larger, so as to best separate these two target sets. To determine the type of hyperplane used to separate the data, a kernel is selected, where each kernel defines (roughly) the shape of the plane that will separate the target sets. Different kernels include linear kernels, polynomial kernels, and radial basis functions RBF (Géron, 2017). For example, the degree/order of a polynomial is selected when using polynomials to define the hyperplane for the SVM. As an example, a linear SVM, which is an order 1 polynomial, uses linear functions to separate target sets. If we only consider two features, for example the number of weevil augmentations and latitude, the separating hyperplane can be a line. Figure 2 describes two variations of this example, where the hyperplane/line is illustrated by the thick black line, and the margin (the support vectors) are described by the dashed black lines. In many cases, a low degree polynomial, such as a line,
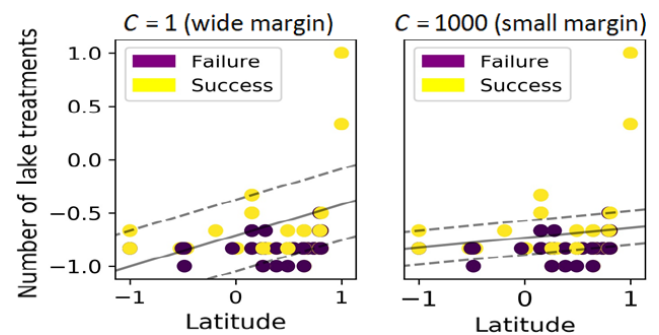

FIGURE 1 A schematic of the models use in future biocontrol programs. Once we have trained, tested, and validated our model, we can input feature sets, including a variety of augmentation strategies, to determine weevil efficacy of Eurasian watermilfoil reduction (with some probability of success)

FIGURE 2 An example of a two-feature linear support vector machine (SVM) describing hyperplane (line) that best separates success and failure. Left image $C = 1$; right image $C = 1000$. The dashed lines are the support vectors that the model attempts to separate the data with. The SVM with $C = 1$ has a larger margin of separation compared to $C = 1000$. Data is scaled between -1 and 1

may be too general and under fit the data. Conversely, if the polynomial degree is too large, the model will over fit.

SVM hyperparameters

Before running the SVM algorithm, the data for features and targets are organized in a matrix, where each study's features are stored in a single row, such that each column represents a single feature across all studies. The final column of the matrix defines the target of success or failure, where the data is formatted such that units for each feature are converted to match. Next, the kernel is selected and the SVM is implemented by setting the SVMs hyperparameters Gamma and C. Such hyperparameters are controllable inputs for the model that change the model behavior. Intuitively, the Gamma parameter defines how far the influence of a single training example reaches, with small values meaning "far" and large values meaning "close." In other words, a large value for Gamma gives target points more weight on the entire separation plane (which can cause overfitting), while a small value for Gamma describes the inverse, that is, a single target point is less likely to change (have an effect on) the location of the hyperplane. The C parameter trades off correct classification of training examples against maximization of the margin (the separation of the support vectors). A large C prioritizes as few miscalculations as possible, whereas a smaller margin is accepted if the decision function is better at classifying all training points correctly. Conversely, a small C prioritizes finding the hyperplane with the widest margin, even if some points are misclassified. In other words, C behaves as a regularization parameter in the SVM.

We choose to use linear and polynomial kernels. Each of these kernels behaves differently, where there are slight variations in the hyperparameters. For the linear model, an array of possible Cs and weights are chosen. For example, in the linear model, the choices for C are 1 and 10 and the possible weights x are between 0.10 and 0.90 (iterated by 0.10), such that there are two C values and nine weightings. The targets defined by success have their C value multiplied by x and the targets defined by failure have their C value multiplied by $1 - x$. Therefore, the model is trained with the 18 possible permutations of hyperparameter sets. For polynomial models, an array of possible Cs, weights, Gamma, and the polynomial degree are chosen. An SVM was made with each possible permutation of C and weightings, and the SVM with the best f1 score (described in *Model validation*) was selected as the final model.

Model training

The SVM is trained by splitting the data into a training set and a test set, where the training set is typically bigger than the test set. The model will know both the feature sets and target values (success or failure) of the training set, and it will use the SVM algorithm, with chosen hyperparameters values as previously described, to find and define the pattern between the features and targets for the training set. The test set is then used to validate the SVM, and is not used in defining/training the model. Thus, the SVM has no knowledge of a study from the test set (i.e., it does not know if a study is classified as a success or failure). Instead the test set of features will later be used in model validation, where using the relationships between features and target already determined during training, the model predicts a target of either success or failure. The SVMs were trained using the technique of cross-validation called Kfold (Géron, 2017). This method is used to determine how best to split the training and test data. We split all the rows of data evenly into K groups (called folds), where we set K equal to 4. The first fold is used as the test set and the remaining $K - 1$ folds are the training sets. Cross validation consists of repeating the above process such that each group/fold is used as a possible test set. We repeated the process 500 times and averaged the results, where 500 times was selected because the f1 score varied only by <0.01 after this point, which is the value we select as a lower threshold.

Model validation

The model's strength is evaluated by comparing the predicted targets to the actual targets. A good model will consistently predict the correct target and a bad model will often misclassify the targets. One way to test the strength of a model is to calculate the number of true positive (TP), false positive (FP), true negative (TN), and false negative (FN) results. TP corresponds to the number of positive examples correctly predicted by the classification model, whereas FP corresponds to the number of negative examples wrongly predicted as positive by the classification model. Similarly, TN corresponds to the number of negative examples correctly predicted by the classification model, and FN corresponds to the number of positive examples wrongly predicted as negative by the classification model. In addition to calculating TP, FP, TN, and FN results, we can calculate the recall R and precision P, for the same set of simulation results. Recall and precision are two widely used metrics employed in applications where successful detection of one of the classes is considered more significant than the detection of the other classes (Tan et al., 2014). Here, it is more important to correctly predict success compared to failure, and so

we use these metrics in our analysis. Definitions for R and P are given by Equations (1) and (2), respectively (Tan et al., 2014). R, defined by Equation (1), is the fraction of the total successes that are labeled correctly. Classifiers with large recall have few positive examples misclassified as a negative class. The recall parameter describes how many lakes with reduced EWM are accounted for in the model's predictions of success, which is the percentage of EWM reductions that the model is able to predict. The denominator represents the total successes (i.e., EWM reduction):

$$R = \frac{\text{True positive}}{\text{True positive} + \text{False negative}}. \tag{1}$$

P, defined by Equation (2), is the fraction of experiments the model correctly labels as a success. The higher the precision is, the lower the number of false positive errors calculated by the classifier. In this study, given a model prediction for success, the probability that a decrease in EWM actually occurs is defined by precision. Stated another way, precision defines how often the model will suggest that weevil augmentation in a lake will result in reduced EWM

$$P = \frac{\text{True positive}}{\text{True positive} + \text{False positive}}. \tag{2}$$

The metric f1 score, given by Equation (3), describes the harmonic mean between P and R. This value is typically closest to the smallest of the values P and R. Thus, if both P and R are reasonably large, then the f1 score is large, too

$$f1 = 2\frac{P \times R}{P + R}. \tag{3}$$

The best hyperparameters were chosen by training the data (both features and targets) over an array of hyperparameter values and choosing the model that results in the best f1 score, defined by Equation (3) as the weighted average of precision P and recall R ($0 < $ f1 score < 1). We choose this metric for selecting hyperparameters since it is typically closest to the smallest of the values P and R, and we get a good indication of accuracy in terms of the "worst" of P and R. This is, if both P and R are large, then the f1 score will be large too, close to 1).

Defining model targets

One of the most important and challenging tasks in this modeling effort is defining the target values as success or failure. As stated in the previous section, some reports record quantitative measurements for EWM biomass at the beginning and end of an augmentation experiment, while others simply indicate whether biomass is reduced or not. Thus, we describe multiple definitions of weevil success, where each definition is based on publicly available data for a given lake. We tested the SVM against a target that uses two classes representing success or failure, marked as 1 and 0, respectively. The target is defined and tested using two methods, increase/decrease (inc/dec—a qualitative measurement) and topx (a quantitative measurement). The inc/dec method, using a qualitative measure, is defined by a statement of EWM decrease or increase by the end of the study ($n = 133$, between 34 lakes). The topx method can only be used for data sets that have quantitative measures for change in EWM. The quantitative sets were additionally tested with splits based on the topx% of EWM decreases. For example, setting topx to 20%, 20% of EWM targets with lowest EWM were marked as success and the remaining rows below the split were set to failure. The overall goal of testing the top percent splits is to test how the model responds to predicting extreme changes in EWM. Since there were few cases with extreme changes of EWM decrease, the 10%–40% split results were similar for all quantitative measures (results not shown), and so the split topx was fixed between 30% and 40% for all quantitative target studies. The two methods (qualitative and quantitative) resulted in six model targets, which each depend on the data for EWM success or failure. The first target, which we call "All augments," uses the qualitative measure of success and failure, and encompasses all of the data we collected and uses the inc/dec method. The second target is a subset of "All augments," called "All augments[a]," and removes all averaged data, such that only data recorded in reports is used. The other four targets, tested with both the inc/dec method and the topx methods, consist of those using quantitative measures of success and failure, including change in EWM stem density, dried EWM biomass, EWM percent, and relative abundance of EWM (amount of EWM compared to other natives). Figure 3 provides a descriptive schematic of the full modeling process, from data collection and filtering to model building, including model training, testing, and validation.

Model reduction and ranking of features

In the first stages of running our SVM, we had selected a large list of potential model features that we thought would be important in defining weevil success at reducing EWM, where a subset of these features are summarized in Table 1. A complete data set that summarizes this initial set of features is given in the linked Open Source Spreadsheet. To simplify the model, we performed feature reduction, a method conducted using a ranking
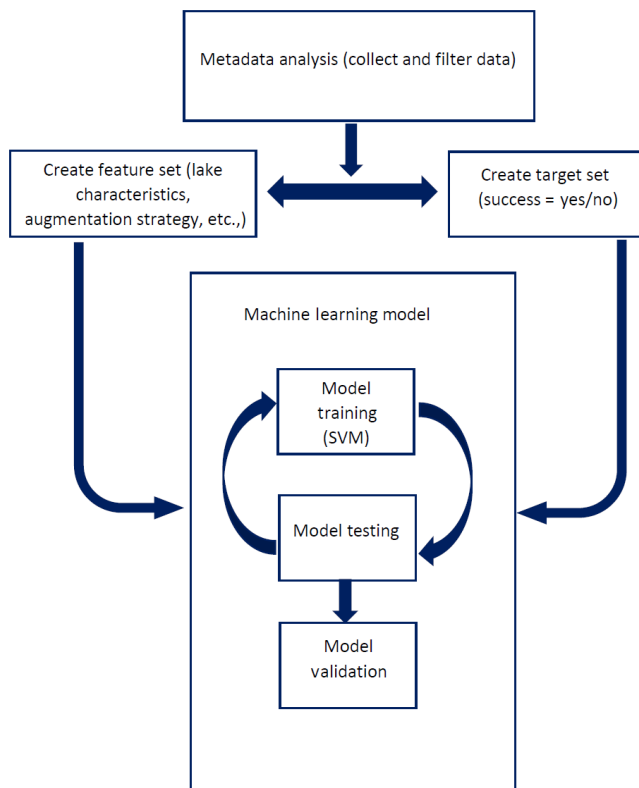
FIGURE 3 A schematic of the data collection process and filtering (the metadata analysis) in connection with the machine learning algorithm. Data is added to a spreadsheet and filtered (lakes with data missing are removed). Once filtered data is recorded as either an input (a model feature such as lake characteristics or augmentation information) or output (model targets of success = 1 or failure = 0). Then, model features and targets are placed into the machine learning algorithm (which uses a support vector machine (SVM) classification algorithm) where data is randomly split into training sets (three-quarters of the data) and training sets (one-quarter of the data) many times. The model is then validated on a set of known targets for which it was not trained

technique to simplify the model such that only the most important features for predicting weevil success at reducing EWM were used. By removing some features from the model, not only was the model easier to build (in terms of computational time), but it's also more practical in the sense that less information needs to be known about a lake and a given augmentation strategy to predict success or failure. Ranking was completed using a recursive feature elimination (RFE) algorithm that removed the least important (lowest ranked) feature first (Géron, 2017). RFE is a wrapper function we applied to the SVM, in an iterative fashion such that one feature was removed at each iteration. After each iteration, the model was retrained with the remaining list of features. The process was repeated until a minimum feature count was reached (i.e., here we set that count to eight

features). Similar to model selection, the minimum feature count was evaluated based on the resulting weighted f1 score.

RESULTS

In this section, we show results for the model's f1 score, precision P, recall R, and accuracy, after training while performing model reduction and ranking, and then testing our SVM for each of the six model targets that describe success and failure. Note that each model target is classified as success or failure based on data recorded as either (1) an increase or decrease in EWM, (2) EWM stem density change, (3) EWM biomass change, (4) percent change in EWM, or (5) the change in relative abundance of EWM, as compared to native plants. In addition to these five targets, we further separated our largest target set All augments into two subsets where (1a) corresponds to All augments and incorporates all $n = 133$ data sets (including averaged missing data), and (1b) corresponds to All augments[a], and includes only data for which values are reported, $n = 54$.

Table 2 summarizes model reduction results, providing the f1 scores for our largest target set, All augments. In particular, the first column illustrates an increasing f1 score up to seven or eight features, stabilizing just above 65% for all additional features (results not shown). Thus, we include eight features in our results for All augments. For all other target sets studied, there were approximately four features that were considered "important" (features 1–4, which correspond to latitude, surface area, lake depth, and buffer zone). In particular, in Table 2, we also illustrate f1 scores for features 3–8 and you can see that for all targets except All augments, the f1 scores do not change after feature 4 (i.e., features 5–8), and in some cases do not change by much. However, for comparison across each model target, we select the same eight features obtained for our largest target set All augments.

After performing model reduction, we also looked at ranking the features in terms of importance (i.e., ranking in terms of those that achieve higher f1 scores), where 1 corresponds to the most important feature. Table 3 illustrates the ranking for the top eight features used in predicting weevil success at reduction of EWM from the "all augments" study. These features include lake latitude, surface area, maximum depth, buffer zone, phosphorus, Secchi depth, and weevil augmentation strategy. It is interesting that the top two features for predicting weevil success, using the largest target set All augments (shown in column 1 of Table 3), are weevil treatment frequency and maximum lake depth. For the subset All augments[a], which removes missing data, weevil

TABLE 2 The f1 scores for eight highest ranked features for all five target sets

Feature no.	All augments	All augments[a] (averaged/missing data removed)	EWM stem density change	EWM biomass change	EWM percent change	Relative abundance of EWM
3	60.09 ± 0.19	65.64 ± 0.26	44.32 ± 0.36	65.47 ± 0.28	46.39 ± 0.25	54.86 ± 0.28
4	63.22 ± 0.17	65.39 ± 0.27	48.03 ± 0.40	68.22 ± 0.30	49.91 ± 0.28	56.06 ± 0.30
5	63.88 ± 0.18	65.15 ± 0.27	49.39 ± 0.40	69.11 ± 0.31	51.23 ± 0.28	55.65 ± 0.30
6	64.74 ± 0.17	65.07 ± 0.27	49.51 ± 0.40	69.19 ± 0.31	52.54 ± 0.28	56.33 ± 0.30
7	65.18 ± 0.17	65.05 ± 0.27	49.48 ± 0.40	69.12 ± 0.30	52.61 ± 0.27	56.89 ± 0.32
8	65.15 ± 0.17	64.99 ± 0.27	49.41 ± 0.40	69.07 ± 0.30	52.69 ± 0.26	57.16 ± 0.31

Note: f1 scores for eight highest ranked features for all five target sets. The top eight features are defined by our largest target set, which includes qualitative information on Eurasian watermilfoil (EWM) increase/decrease (any increase is classified as a failure, and any decrease is classified as a success). Overall, each target set showed a large drop in f1 score between three and four features. In addition, the f1 score only started to increase after the first few features were removed (results not shown).

[a]Data removed from adjusted "All augments" is averaged phosphorus, Secchi depth, and number of weevils, which gives us a total of $n = 54$ feature sets.

TABLE 3 Ranking for eight highest ranked features for all six target sets

Features	Targets of success					
	All augments	All augments[a] (averaged/missing data removed)	EWM stem density change	EWM biomass change	EWM percent change	Relative abundance of EWM
Latitude	4.1	5.1	6.3	**1.3**	1.5	**1.1**
Area (ha)	1.7	6.1	**1.3**	3.9	1.2	1.3
Maximum depth (m)	**1.2**	2.4	2.3	4.9	**1.0**	**1.0**
Buffer (km)	2.0	**1.7**	**1**	1.5	2.7	3.5
Phosphorus (µg/L)	2.5	7.1	1.6	2.1	2.0	2.1
Secchi depth (m)	1.4	3.2	3.3	2.9	1.0	2.7
Treatment frequency	**1**	**1**	4.3	**1**	**1**	4.4
Average no. weevils	3.2	4.1	5.3	5.9	3.6	1.7

Note: Ranking of model features from most important (ranked as 1) to least important. The top two highest ranked model features are shown in boldface type in each of the columns (describing the highest ranked features in each target set). It is shown that four out of six of the targets tested show that augmentation treatment frequency (how many times weevils are added) is the most important predictor of weevil success. In addition, three out of six of the targets tested showed that maximum lake depth was an important predictor of weevil success. In addition, two out of six target sets showed buffer zone as one of the most important predictors of weevil success. The model we select as our primary target is the "All augments[a]" target shown in the first column, which illustrates that treatment frequency and buffer zone are the top two most important features needed to predict weevil success.

[a]Data removed from adjusted "All augments" is averaged phosphorus, Secchi depth, and weevil number, which gives us a total of $n = 54$ feature sets.

treatment frequency and buffer zone are the top two features (shown in column 2 of Table 3). In both these studies, the absolute number of weevils added had less of an impact on predicting weevil success as compared to the number of augmentations, which ranked seven out of eight features for All augments, and five out of eight features for All augments[a]. This result suggests that it is more important to consider the number of times one augments a lake, rather than the absolute number of weevils added at any given time. In particular, the more times weevils are stocked in a single season, the better the

outcome of success at EWM reduction. In addition, for the smaller target set All augments[a], the top two features include augmentation frequency and buffer zone, such that lakes with lakes with larger weevil habitat will have a better outcome of success. As weevils need to overwinter to be able to naturally sustain their populations, this result makes sense. In addition to looking at the average ranking of parameters, it is useful to look at their averages and variance. Figure 4 describes a box and whisker plot for the ranking of the eight model features across all six target sets (Figure 4a) as well as across five targets,
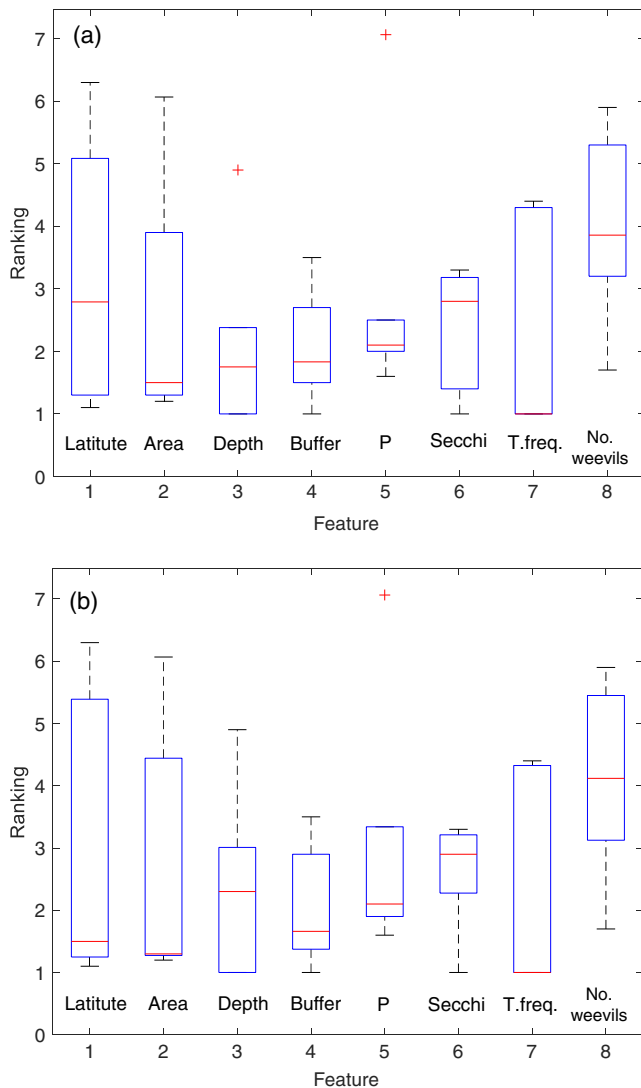
FIGURE 4 Box plot of feature ranking. Here, ranking of features 1 through 8 across 2a: All target sets, and 2b: Target sets excluding "All augments" where missing data is averaged. In each plot, the red line corresponds to the mean, the blue box to the total variation in the data, the black lines to the standard deviation, and red crosses to outliers. Here, we see that in both cases, treatment frequency, lake area, and buffer zone are ranked in the top four of eight features (features are lake location (latitude), area, maximum depth, buffer zone (measure of shoreline suitable for weevil overwintering), phosphorus (P), Secchi depth, augmentation strategy (number of weevils added), and treatment frequency. However, in 2a, lake depth in also ranked in top four, whereas in 2b latitude is ranked in the top four

where we remove the largest target set that used averaged data (Figure 4b).

To test the strength of our model, we tested and compared the SVM prediction results to known results of success or failure. Tables 4 and 5 show the SVM results for each testing strategy 1–6 for the linear and polynomial SVM, respectively, where each table summarizes the models f1 score, precision P, recall R, and accuracy. As milfoil weevil augmentation can be expensive, we treated precision as the most important parameter to indicate model success, in the sense that models that have higher percentages of false positive results (resulting in a lower P) would be a less desirable model result. In a similar light, recommending weevil augmentation that could be beneficial is less important in the sense that we are not wasting resources (i.e., a false positive is less desirable than a false negative).

From Table 4, we conclude that the second target strategy All augments[a] (the subset of All augments) is the best for predicting weevil success in the sense that (1) it has a small number of false positives compared to false negatives; (2) it has the highest recall (aside from stem density, which has the highest false positive count, so we exclude it); (3) it has the highest precision; and (4) it has the second highest f1 score (the target "EWM biomass change" has the highest f1 score, but has a very low precision, so we exclude that target, also). In Table 5, we show the results of our polynomial SVM. As the percentage of false positive results are high in each target (except for the target "Change in the relative abundance of EWM", which has extremely low precision and recall), shown by the values bolded in Table 5, we decide that our linear SVM does a better job at predicting weevil success, as compared to the higher order polynomial model.

DISCUSSION

Control and management of aquatic invasive species is one of the most important issues facing lake communities, due to the fact that such invasive species have negative impacts on both the lake's natural ecosystem (Les & Mehrhoff, 1999; Smith & Barko, 1990) and the local economy (Olden & Tamayo, 2014; Rockwell, 2003). With variable success at EWM reduction using current management strategies, in addition to a lack of understanding of the long-term ecological and economic impacts of such invasions, it is becoming increasingly important to determine sustainable mitigation strategies that are cost effective, easily implemented, and cause little to no harm to the natural environment. In this study, we focus on understanding the sustainable approach of augmentation using the milfoil specialist, the milfoil weevil, as a biocontrol for EWM. In particular, we developed a predictive model to determine if EWM reduction via weevil augmentation will be successful in a given lake. Such a predictive model can be used by lake community stakeholders to determine if weevil augmentation is an appropriate EWM mitigation strategy for a given lake community.

TABLE 4 Model outputs for linear SVM for all model targets

Targets of success (across) versus model validation parameters (down)	All augments	All augments[a] (averaged/missing data removed)	EWM stem density change	EWM biomass change	EWM percent change	Relative abundance of EWM
Accuracy	66.48 ± 0.10	67.41 ± 0.14	57.74 ± 0.15	76.11 ± 0.12	58.21 ± 0.14	64.31 ± 0.18
False negative	24.72 ± 0.10	28.77 ± 0.09	**4.42 ± 0.05**	21.58 ± 0.09	34.37 ± 0.12	27.99 ± 0.10
False positive	**8.80 ± 0.08**	**3.81 ± 0.13**	37.84 ± 0.14	**2.31 ± 0.08**	**7.41 ± 0.10**	**7.70 ± 0.16**
Precision	72.62 ± 0.20	86.90 ± 0.46	57.84 ± 0.10	33.97 ± 0.91	60.27 ± 0.59	26.00 ± 0.78
Recall	47.42 ± 0.18	42.92 ± 0.21	92.19 ± 0.09	14.15 ± 0.43	24.76 ± 0.27	13.2 ± 0.39
f1 score	65.04 ± 0.10	64.90 ± 0.15	49.16 ± 0.23	69.61 ± 0.17	52.82 ± 0.18	57.69 ± 0.19

Note: Results for linear SVM model. For each of the targets tested, we can see that false positives are kept at a minimum (are low) for every case except that of stem density (the lowest value between false negative and false positive is shown in boldface type for each target). In addition, we can see that the "All augments" and adjusted "All augments[a]", which included $n = 54$ data sets with missing data for phosphorus, Secchi depth, and weevil numbers removed. Results show similar trends in model accuracy and f1 score and recall. However, there is a significant increase in model precision, and a significant decrease in FP. This results in a better/more accurate model overall.

[a]Data removed from adjusted "All augments" is averaged phosphorus, Secchi depth, and number of weevils, which gives us a total of $n = 54$ feature sets.

TABLE 5 Model outputs for polynomial SVM for all model targets

Targets of success (across) versus model validation parameters (down)	All augments	All augments[a] (averaged/missing data removed)	EWM stem density change	EWM biomass change	EWM percent change	Relative abundance of EWM
Accuracy	64.95 ± 0.14	65.60 ± 0.20	57.20 ± 0.14	69.69 ± 0.24	63.36 ± 0.19	67.40 ± 0.13
False negative	13.54 ± 0.10	18.24 ± 0.14	4.26 ± 0.03	12.64 ± 0.15	17.74 ± 0.13	28.48 ± 0.09
False positive	**21.52 ± 0.11**	**16.18 ± 0.16**	**38.54 ± 0.14**	**17.67 ± 0.20**	**18.90 ± 0.13**	4.12 ± 0.10
Precision	61.04 ± 0.14	67.60 ± 0.26	57.43 ± 0.09	42.78 ± 0.53	59.87 ± 0.22	32.39 ± 0.89
Recall	71.17 ± 0.21	64.30 ± 0.29	92.41 ± 0.06	50.90 ± 0.66	61.12 ± 0.30	11.81 ± 0.34
f1 score	64.79 ± 0.15	65.25 ± 0.21	48.20 ± 0.22	70.22 ± 0.24	63.25 ± 0.19	59.58 ± 0.17

Note: Results for polynomial SVM model (polynomial of order 3). In this particular model, the percentage of false positives is higher than for the linear model, so we stick with our linear SVM, as one of the goals of our decision-making process is to minimize false positives (i.e., it's more costly (in terms of time and money) to suggest weevil augmentation in a lake where it is likely not to work then it is to not prescribe weevil augmentation in a lake where it could work.

[a]Data removed from adjusted "All augments" is averaged phosphorus, Secchi depth, and number of weevils, which gives us a total of $n = 54$ feature sets.

Field reports that studied EWM reduction via weevil augmentation were mixed in the sense that there were no clear indicators as to why some augmentations were successful while others were not (Havel et al., 2017; Jester et al., 2000; Reeves et al., 2008). In particular, there was no obvious connection or pattern between lakes that showed weevil success at reducing EWM. Here, we developed a model, using machine learning techniques, which helped us to determine those connections and patterns. In particular, we simulated a classification SVM that predicts a target of success or failure based on a set of input features, which include lake characteristics such as lake latitude, surface area, maximum depth, buffer zone, phosphorus, Secchi depth, and the weevil augmentation strategy. We used six targets of success, all of which included some measure of EWM decrease or increase, representing success or failure, respectively.

In order to define each target set, and to determine the features listed in Table 1 (those features used in our predictive model), we completed a metadata analysis that consisted of a total of 133 cases collected from 13 studies and 34 lakes. Many of our initial feature ideas did not end up in our final predictive model, due to the fact that (1) data for a wanted feature was missing or incomplete or, (2) after feature ranking, features ranked much lower than the eight recorded. Some features that we collected that did not go into the final analysis were shore length, wetland, open land, average depth, chlorophyll concentration, pH, panfish populations (known to eat milfoil weevils (Ward & Newman, 2006), as well as the start and end month of the study. Features that ranked low included shore length, wetland, open land, and average lake depth. Riparian wetlands include all shorelines adjacent to any wetlands of the body of water and open land

includes any shorelines with <5 m of forested land, including docks and shorelines containing mostly rocks and beaches. Both types of land are not suitable for weevil overwintering, so it makes sense that these features ranked lower at predicting weevil success. Features that had a lot of missing data included fish population information, chlorophyll concentration, water pH, and month of augmentation study. Of those features we removed from the study, we feel that the panfish population size and the month at which the augmentation study started are most important for future consideration, in the sense that no other features used in the current version of our model are linked to these two features. We suggest that average depth ([lake volume]/[lake area]), and chlorophyll concentration correlate with physical and chemical features recorded and used in our model, such as maximum lake depth (Neumann, 1959) and phosphorus concentration (Quinlan et al., 2021), respectively. We should also point out that we initially described buffer zone length in two ways: by measuring the absolute buffer length zone (in m), and the percent buffer zone, measured as a percentage of the total shore length that is adequate for weevil overwintering. We had initially thought that the second measure (percent buffer zone) might be ranked higher than absolute buffer zone length, but this was not the case. Percent buffer zone results were almost identical to absolute buffer length (results not shown), thus we kept absolute buffer length as the primary buffer zone feature.

Performing feature reduction helped to optimize the number of input features required to predict EWM reduction success or failure, in addition to ranking features in terms of importance (1 being the best predictor of success or failure). The results of feature reduction and ranking indicated that eight features were required for the large qualitative study (the All augments target), while four or less were required for all other studies (the five other targets). This result helped to simplify the original model, which consisted of over 20 model features, in terms of computational speed as well as ease in everyday practice (less data needs to be known about a lake than originally assumed). In order to compare f1 scores, precision, and recall across all target sets, we kept all eight features for each model, including lake latitude, surface area, maximum depth, buffer zone length, phosphorus concentration, Secchi depth, number of weevils added, and the number of weevil applications. Lake latitude is related to temperature, such that lakes at higher latitudes tend to be colder. EWM growth rates are directly correlated with lake temperature, such that EWM grows better and faster in warmer environments. Although we might expect more EWM in warmer lakes, we know that weevils also grow better in warmer water (Mazzei et al., 1999). In that case, warmer lakes might also provide a larger EWM habitat for weevils to damage. Overall, this dynamic interplay between EWM growth and weevil success at damaging EWM is not straightforward to quantify highlighting the importance of our modeling approach. Buffer zone is also shown to be important at predicting weevil success. As milfoil weevils require a place to overwinter, it makes sense that lakes with larger buffer zones encourage weevil overwintering, and hence are likely to lead to success at reducing EWM growth (Thorstenson et al., 2013). The water quality parameters such as phosphorus and Secchi depth also rank highly for predicting weevil success. Like lake temperature, it's not entirely straightforward to predict whether low versus high Secchi depth would be a better indicator of weevil success at EWM reduction. In particular, turbid lakes with low Secchi depth often inhibit the growth of EWM since murky water inhibits light penetration at deeper levels (Jones et al., 2012; Smith & Barko, 1990). In Table 1, a low Secchi depth of 1.8 m is recorded for Bearskin Lake, indicating turbid or highly colored water. However, for the five sites tested in that lake, three showed success and two resulted in failure. Finally, it is shown that the weevil augmentation strategy is an important predictor of weevil success (both the number of weevils added and the frequency of augmentations). In particular, stocking more weevils and performing multiple stockings in a single season results in a better outcome of EWM reduction success. This result is an important one, since our work looks to define optimal EWM control strategies in terms of developing weevil augmentation programs.

In terms of ranking, there is no single most highly ranked feature across all model targets. However, from the average rankings given in Table 3, we see that both maximum lake depth and treatment frequency (as opposed to the absolute number of weevils added) are the best two predictors for each model (better than any other feature pair). This result suggests that light attenuation, when shortwave radiation from the sun is attenuated by the water (Herb & Stefan, 2003), is one of the most important predictors of success, such that shallower lakes might be more conducive to weevil augmentation than deeper lakes, in addition to more frequent weevil augmentations (stocking per season). Previous work looking a weevil survival fitness has shown that weevils do better in shallow water, which is consistent with the results we have found here (Newman, 2004; Parsons et al., 2011). These authors attribute wave shelter, higher temperature, predation refuge (but not light), are likely contributing factors to weevil success in smaller lakes. In addition, the modeling work by Miller et al. (Miller et al., 2011) has shown diminishing returns between single-season stockings of weevils and EWM control

efficacy. That is, past 50 weevils added per season, the reduction in biomass as described by the models simulated biomass peak at the end of season is negligible. This result is consistent with our results in the sense of defining a successful weevil augmentation program. That is, one should consider the treatment frequency as a better indicator of success than the number of weevils added.

In addition to average model ranking, we illustrate results using a box and whisker plot in Figure 4, to highlight averages and variances in the feature ranking between all target sets. For the All augments target set, in addition to lake depth and treatment frequency, lake surface area and buffer zone are ranked among the top four features when comparing across all target sets. From this perspective buffer zone and treatment frequency, which correlate to weevil survival, in addition to lake surface area, which might correlate to the size of an EWM patch and/or the distance to weevil overwintering sites (lakes with smaller areas are more successful) might be the best long-term predictors of weevil success at reducing EWM.

The results from the SVM simulation showed that a linear kernel in the SVM outperformed that of the polynomial kernel. In general the two models had similar f1 scores (described by Tables 4 and 5, respectively), however the linear model had a lower false positive rate, as compared to the polynomial model. The low false positive rate is important because false positives represent the model suggesting failure of weevil augmentation in a lake. A high false positive rate would be a worst-case outcome for a predictive model like ours, in the sense that we would suggest weevil augmentation in a situation that would waste time and money. Comparatively, the linear model showed higher false negatives, but this simply represents a missed opportunity to reduce EWM and is a more acceptable mistake in the sense that one does not waste time and resources. However, the focus on minimizing false positives by choosing the linear model means that the model is less useful for all lakes; this is seen by the low recall of 43% for All augments[a], meaning that only half of the lakes that could benefit from weevil augmentation are recognized by the model.

A clear model limitation is the amount of publicly available data. Here, we have worked to compile the most complete data set of weevil augmentation reports, which includes 133 cases among 34 lakes. As many of the reports were completed independently from each other, data was often measured or recorded in different ways. In addition, many cases were missing data that we initially thought important for predicting weevil success. The additional data that we feel would improve our study, in the sense that these features are not correlated with any other model features used in the current model, are the month of weevil augmentation, as well as panfish

population density. Panfish are expected to be found in most, if not all, lakes in our study. Fishless lakes are relatively rare in this region and, when they exist, are either ephemeral (e.g., vernal pools), so shallow (maximum depth <4 m; Fang & Stefan, 2000) that winter fish kills due to hypoxia occur during ice cover, they are acidic, or are high elevation lakes where steep slopes prevent colonization by fish (Schilling et al., 2008). Since panfish are the primary predators for weevils (Ward & Newman, 2006), knowledge of their abundance in the lakes of the database of 133 cases weevil augmentations would be useful.

A second limitation also ties into data availability. In particular, we believe that some of our quantitative definitions of success (like overall biomass change, or change in species abundance), would give a more accurate description of weevil success at EWM reduction. However, each of these target sets had only $n = 44$ and $n = 50$ cases for each, respectively, as compared to the qualitative target of All augments, for which $n = 133$ cases. In future augmentation studies, it is important for individuals to measure EWM increase or decrease in a quantitative manor, such as dried mass or stem density, so that (1) data can be more accurately compared across lakes and (2) more data is available to recalibrate (validate) and test our model.

In addition to the above-mentioned limitations of our model, it would be useful to do a more in-depth study of the correlation between a lake's buffer zone to post-augmentation success. In many of the studies looked at, multi-year surveying was not completed to test for success in the years that follow an augmentation. Since most of our studies record data over one growing season, it could be the case that buffer zone might rank more highly if we consider post-augmentation success in our overall definition of success. We would recommend that any groups completing augmentation studies complete surveys in the years that follow and make this data available for future model extensions. In addition to this, we would recommend that in tandem to recording absolute numbers of weevils added, that individuals record weevils per stem (as some studies have done). This measurement is likely to prove more useful than absolute weevil numbers as it gives us information about the number of weevils added relative to the size of a milfoil patch.

Although watermilfoil weevil augmentation is a promising mitigation strategy, it may not work for all lakes. However, results here show that, regardless of the target used in the model development, weevil frequency, buffer zone, lake depth, and lake surface area, are the most important indicators in weevil success. In particular, augmentation programs on lakes that use multiple augmentations and have 1arge buffer zone for weevils to overwinter will be more successful. It should be noted

that, with proper training and funding, lakes with little to no buffer zones can be created/restored by reconstructing shoreline around a lake. Our results also indicate shallower lakes and those with smaller surface areas, are likely to be more successful for weevil augmentation programs. Shallow lakes have already been described as being more successful habitats for weevils (Newman, 2004; Parsons et al., 2011), and we suggest small surface areas could correspond to smaller traveling pathways to weevil overwintering sites.

In terms of suggesting weevil augmentation, our linear model's precision score is around 87% (using All augments[a] as the model target). As false positives are undesirable in the sense that we do not want to suggest weevil augmentation unless we have some level of certainty of success (because it can be expensive in terms of cost and time; Jester et al., 2000), we suggest that this model, because of its low false positive outcome relative to false negatives, is a reasonable model for testing weevil augmentation outcomes. Other sustainable approaches such as small-scale hand harvesting and strategic placement of benthic mats in conjunction with weevil augmentation, may prove the most sustainable approach to biocontrol of EWM (Hussner et al., 2017; Laitala et al., 2012; Marko & White, 2018; Newman, 2004).

ACKNOWLEDGMENTS

CONFLICT OF INTEREST

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

Data (White, 2022) are available in Zenodo at https://doi.org/10.5281/zenodo.6129628.

REFERENCES

Best, E. P. H., C. P. Buzzelli, S. M. Bartell, R. L. Wetzel, W. A. Boyd, R. D. Doyle, and K. R. Campbell. 2001. "Modeling Submersed Macrophyte Growth in Relation to Underwater Light Climate: Modeling Approaches and Application Potential." *Hydrobiologia* 444: 43–70.

Buchan, L. A. J., and D. K. Padilla. 2000. "Predicting the Likelihood of Eurasian Watermilfoil Presence in Lakes, A Macrophyte Monitoring Tool." *Ecological Applications* 10(5): 1442–55. https://doi.org/10.1890/1051-0761(2000)010[1442:PTLOEW]2.0.CO;2.

Cao, J., Y. Wang, and Z. Zhu. 2012. "Growth Response of the Submerged Macrophyte Myriophyllum Spicatum to Sediment Nutrient Levels and Water-Level Fluctuations." *Aquatic Biology* 17(3): 295–303. https://doi.org/10.3354/ab00484.

Caspers, R., R. Selsky, G. Bugbee, and J. White. 2009. "Species Richness of both Native and Invasive Aquatic Plants Influenced by Environmental Conditions and Human Activity." *Botany* 87: 306–14.

Coetzee, J. A., A. Bownes, and G. D. Martin. 2011. "Prospects for the Biological Control of Submerged Macrophytes in South Africa." *African Entomology* 19(2): 469–87. https://doi.org/10.4001/003.019.0203.

Creed, R. P., and S. P. Sheldon. 1993. "The Effect of Feeding by a North American Weevil, *Euhrychiopsis lecontei*, on Eurasian Watermilfoil (*Myriophyllum spicatum*)." *Aquatic Botany* 45(2–3): 245–56. https://doi.org/10.1016/0304-3770(93)90024-Q.

Creed, R. P., and S. P. Sheldon. 1995. "Weevils and Watermilfoil: Did a North American Herbivore Cause the Decline of an Exotic Plant?" *Ecological Applications* 5(4): 1113–21. https://doi.org/10.2307/2269358.

Fang, X., and H. G. Stefan. 2000. "Projected Climate Change Effects on Winterkill in Shallow Lakes in the Northern United States." *Environmental Management* 25(3): 291–304.

Géron, A. 2017. *Hands-on Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*, First ed. Newton MA: O'Reilly.

Gross, E. M., H. Groffier, C. Pestelard, and A. Hussner. 2020. "Ecology and Environmental Impact of Myriophyllum Heterophyllum, an Aggressive Invader in European Waterways." *Diversity* 12(4): 127. https://doi.org/10.3390/d12040127.

Havel, J. E., S. E. Knight, and K. A. Maxson. 2017. "A Field Test on the Effectiveness of Milfoil Weevil for Controlling Eurasian Watermilfoil in Wisconsin Lakes." *Hydrobiologia* 800(1): 81–97. https://doi.org/10.1007/s10750-017-3142-2.

Herb, W. R., and H. G. Stefan. 2003. "Integral Growth of Submersed Macrophytes in Varying Light Regimes." *Ecological Modelling* 168(1–2): 77–100. https://doi.org/10.1016/S0304-3800(03)00206-0.

Hussner, A., I. Stiers, M. J. J. M. Verhofstad, E. S. Bakker, B. M. C. Grutters, J. Haury, J. L. C. H. van Valkenburg, et al. 2017. "Management and Control Methods of Invasive Alien Freshwater Aquatic Plants: A Review." *Aquatic Botany* 136: 112–37. https://doi.org/10.1016/j.aquabot.2016.08.002.

Jester, L. L., M. A. Bozek, D. R. Helsel, and S. P. Sheldon. 2000. "Euhrychiopsis Lecontei Distribution, Abundance, and Experimental Augmentations for Eurasian Watermilfoil Control in Wisconsin Lakes." *Journal of Aquatic Plant Management* 38: 88–97.

Jones, J. I., A. L. Collins, P. S. Naden, and D. A. Sear. 2012. "The Relationship between Fine Sediment and Macrophytes in Rivers: Fine Sediment and Macrophytes." *River Research and Applications* 28(7): 1006–18. https://doi.org/10.1002/rra.1486.

Laitala, K. L., T. S. Prather, D. Thill, B. Kennedy, and C. Caudill. 2012. "Efficacy of Benthic Barriers as a Control Measure for Eurasian Watermilfoil (*Myriophyllum spicatum*)." *Invasive Plant Science and Management* 5(2): 170–7. https://doi.org/10.1614/IPSM-D-09-00006.1.

Les, D., and L. Mehrhoff. 1999. "Introduction of Nonindigenous Aquatic Vascular Plants in Southern New England: A Historical Perspective." *Biological Invasions* 1: 281–300.

Little, R., and D. Rubin. 2002. *Statistical Analysis with Missing Data*, Second ed. New York, NY: Wiley & Sons.

Marko, M. D., and J. C. White. 2018. "Direct Comparison of Herbicidal or Biological Treatment on Myriophyllum Spicatum Control and Biochemistry." *Frontiers in Plant Science* 9: 1–10. https://doi.org/10.3389/fpls.2018.01814.

Martin, C. W., and J. F. Valentine. 2014. "Sexual and Asexual Reproductive Strategies of Invasive Eurasian Milfoil (*Myriophyllum spicatum*) in Estuarine Environments." *Hydrobiologia* 727(1): 177–84. https://doi.org/10.1007/s10750-013-1798-9.

Mazzei, K. C., R. M. Newman, A. Loos, and D. W. Ragsdale. 1999. "Developmental Rates of the Native Milfoil Weevil, Euhrychiopsis Lecontei, and Damage to Eurasian Watermilfoil at Constant Temperatures." *Biological Control* 16: 139–43.

McKnight, S., and G. Hepp. 1995. "Potential Effect of Grass Carp Herbivory on Waterfowl Foods." *Journal of Wildlife Management* 59(4): 720–7.

Michaud, J. 2018. "Problems Inherent to Augmentation of Natural Enemies in Open Agriculture." *Neotropical Entomology* 47: 161–70. https://doi.org/10.1007/s13744-018-0589-4.

Miller, J. K., L. Roketenetz, and H. Garris. 2011. "Modeling the Interaction between the Exotic Invasive Aquatic Macrophyte *Myriophyllum spicatum* and the Native Biocontrol Agent *Euhrychiopsis lecontei* to Improve Augmented Management Programs." *BioControl* 56(6): 935–45. https://doi.org/10.1007/s10526-011-9371-9.

Neumann, J. 1959. "Maximum Depth and Average Depth of Lakes." *Journal of the Fisheries Board of Canada* 16(6): 923–7.

Newman, R. M. 2004. "Invited Review: Biological Control of Eurasian Watermilfoil by Aquatic Insects: Basic Insights from an Applied Problem." *Archiv für Hydrobiologie* 159(2): 145–84. https://doi.org/10.1127/0003-9136/2004/0159-0145.

Newman, R. M., and D. D. Biesboer. 2000. "A Decline of Eurasian Watermilfoil in Minnesota Associated with the Milfoil Weevil." *Journal of Aquatic Plant Management* 38: 105–11.

Olden, J., and M. Tamayo. 2014. "Incentivizing the Public to Support Invasive Species Management: Eurasian Milfoil Reduces Lakefront Property Values." *PLoS One* 9(10): e110458.

Olson, E., S. Ventura, and J. Zedler. 2012. "Merging Geospatial and Field Data to Predict the Distribution and Abundance of an Exotic Macrophyte in a Large Wisconsin Reservoir." *Aquatic Botany* 96: 31–41.

Parsons, J., G. Marx, and M. Divens. 2011. "A Study of Eurasian Watermilfoil, Macroinvertebrates and Fish in a Washington Lake." *Journal of Aquatic Plant Management* 49: 71–82.

Quinlan, R., A. Filazzola, O. Mahdiyan, A. Shuvo, K. Blagrave, C. Ewins, L. Moslenko, D. K. Gray, C. M. O'Reilly, and S. Sharma. 2021. "Relationships of Total Phosphorus and Chlorophyll in Lakes Worldwide." *Limnology and Oceanography* 66(2): 392–404.

Reeves, J. L., P. D. Lorch, M. W. Kershner, and M. A. Hilovsky. 2008. "Biological Control of Eurasian Watermilfoil by *Euhrychiopsis lecontei*: Assessing Efficacy and Timing of Sampling." *Journal of Aquatic Plant Management* 46: 144–9.

Rockwell, H. W. (2003). "Summary of a Survey of the Literature on the Economic Impact of Aquatic Weeds." *The Economic Impact of Aquatic Weeds* 1–18.

Schultz, R., and E. Dibble. 2012. "Effects of Invasive Macrophytes on Freshwater Fish and Macroinvertebrate Communities: The Role of Invasive Plant Traits." *Hydrobiologia* 684(1): 1–14. https://doi.org/10.1007/s10750-011-0978-8.

Schilling, E. G., C. S. Loftin, K. E. Degoosh, A. D. Huryn, and K. E. Webster. 2008. "Predicting the Locations of Naturally Fishless Lakes." *Freshwater Biology* 53(5): 1021–35.

Sheldon, S. P., and R. P. Creed. 1995. "Use of a Native Insect as a Biological Control for an Introduced Weed." *Ecological Applications* 5(4): 1122–32. https://doi.org/10.2307/2269359.

Smith, C., and J. Barko. 1990. "Ecology of Eurasian Watermilfoil." *Journal of Aquatic Plant Management* 28: 55–64.

Tan, P.-N., M. Steinbach, A. Karpatne, and V. Kumar. 2019. *Introduction to Data Mining*, Second Ed. New York, NY: Pearson.

Thorstenson, A. L., R. L. Crunkilton, M. A. Bozek, and N. B. Turyk. 2013. "Overwintering Habitat Requirements of the Milfoil Weevil, *Euhrychiopsis lecontei*, in Two Central Wisconsin Lakes." *Journal of Aquatic Plant Management* 51: 88–93.

Thum, R. A., and J. T. Lennon. 2010. "Comparative Ecological Niche Models Predict the Invasive Spread of Variable-Leaf Milfoil (*Myriophyllum heterophyllum*) and its Potential Impact on Closely Related Native Species." *Biological Invasions* 12(1): 133–43. https://doi.org/10.1007/s10530-009-9437-0.

Ward, D. M., and R. M. Newman. 2006. "Fish Predation on Eurasian Watermilfoil (*Myriophyllum spicatum*) Herbivores and Indirect Effects on Macrophytes." *Canadian Journal of Fisheries and Aquatic Sciences* 63(5): 1049–57. https://doi.org/10.1139/f06-010.

Wersal, R., J. Madsen, T. E. Woolf, and N. Eckberg. 2010. "Assessment of Herbicide Efficacy on Eurasian Watermilfoil and Impacts to the Native Submersed Plant Community in Hayden Lake, Idaho, USA." *Journal of Aquatic Plant Management* 48: 5–1.

White, D. 2022. "DianaTWhite/Weevil-Meta: v1.0 (v1.0)." Zenodo. https://doi.org/10.5281/zenodo.6129628.