Canadian Science Publishing | Environmental Reviews

Application of machine learning methods in forest ecology: recent progress and future challenges

Journal:	*Environmental Reviews*
Manuscript ID	er-2018-0034.R2
Manuscript Type:	Review
Date Submitted by the Author:	28-Jun-2018

Complete List of Authors:	Liu, Zelin; University of Quebec at Montreal, Biological Sciences Peng, Changhui; University of Quebec at Montreal Timothy, Work; UQAM Departement des sciences biologiques Candau, Jean-noel; Natural Resources Canada, Great Lake Forestry Centre, Canadian Forest Service Desrochers, Annie; Université du Québec en Abitibi-Témiscamingue, Institut de Recherche sur les Forêts Daniel, Kneeshaw; UQAM Departement des sciences biologiques

Keyword:	decision trees learning, artificial neural network, support vector machine, species classification, hazard assessment

Application of machine learning methods in forest ecology: recent progress and future challenges

Zelin Liu[1], Changhui Peng[1,*], Timothy Work[1], Jean-Noel Candau[2], Annie DesRochers[3], Daniel Kneeshaw[1]

[1]Department of Biological Sciences, University of Quebec at Montreal, Montreal, QC H3C 3P8, Canada

[2]Great Lake Forestry Centre, Canadian Forest Service, Natural Resources Canada, ON P6A 2E5, Canada

[3]Institut de Recherche sur les Forêts, Université du Québec en Abitibi-Témiscamingue, QC J9T 2L8, Canada

To be submitted to *«Environmental Reviews»*

*Corresponding author:
Professor Changhui Peng
Department of Biological Sciences
University of Quebec at Montreal, QC, Canada

Email: peng.changhui@uqam.ca

Word count (including abstract, without acknowledgements and references): 7211

Abstract

Machine learning, an important branch of artificial intelligence, is increasingly being applied sciences such as forest ecology. Here, we review and discuss three commonly used methods of machine learning including decision tree learning, artificial neural network, and support vector machine, and their applications in five different aspects of forest ecology over the last decade. These applications include: (1) species distribution models (SDMs), (2) carbon cycles, (3) hazard assessment and prediction, and (4) other applications in forest management. While machine learning approaches are useful for classification, modeling, and prediction in forest ecology research, further expansion of machine learning technologies is limited by the lack of suitable data and the relatively "higher threshold" of applications. However, the combined use of multiple algorithms and improved communication and cooperation between ecological researchers and machine learning developers still present major challenges and tasks for the betterment of future ecological research. We suggest that future applications of machine learning in ecology will become an increasingly attractive tool for ecologists in the face of "big data" and that ecologists will gain access to more types of data such as sound and video in the near future possibly opening new avenues of research in forest ecology.

Key words: decision trees learning, artificial neural network, support vector machine, species classification, hazard assessment, forest management

1. Introduction

Forests cover approximately 30% of the world's land area and are the dominant terrestrial ecosystem on Earth (Schmitt et al. 2009). As such, forest ecosystems have historically received much attention from scientists who have been trying to understand the complex interactions between the various ecological processes that drive the dynamics of these systems. The recent increase in the availability of large amounts of data and the development of data analysis methods capable of handling large datasets are providing new opportunities to study these complex systems (Flach 2001; Crisci et al. 2012). Machine learning (ML) is an important branch of artificial intelligence (AI), which provides some significant advantages over traditional statistical methods for analyzing forest ecological data when sufficiently large data sets are available as model training sets. The ML application processes mainly include: (1) the selection of relevant data and its pre-processing; (2) the selection of adequate algorithms; and (3) its quality assessment solutions (Muhamedyev 2015).

Since the 1990s, ML has increasingly been used in environmental sciences (Hsieh 2009). Previous reviews and books (Haupt et al. 2008; Hsieh 2009; Thessen 2016) focused on several fields of research that included oceanography, hydrology, and atmospheric sciences, but they rarely reported on how ML was used to study forest ecosystems although these methods had increasingly become popular over the last decade in forest ecosystem research as reflected by the increasing number of publications (see Fig. 1). In this research field, ML approaches provided powerful and efficient ways to deal with data that was nonlinear, had high dimensionality, contained complex interactions and/or missing values (Bhattacharya 2013; Thessen 2016). For example, using modern remote sensing and mapping techniques, ML methods effectively improved the accuracy of species distribution models (SDMs) (Garzón et al. 2008; Vaca et al. 2011; Pouteau et al. 2012; Faleiro et al. 2013; Périé and Blois 2016), or in combination with traditional processes or empirical models, they were used to predict carbon (C) and energy fluxes (Papale and Valentini 2003; Papale et al. 2015; Shoemaker and Cropper 2008, 2010; Tramontana et al. 2015, 2016). ML methods were also used in hazard assessment and forest management (Rogan et al. 2008; Hlásny et al. 2011; Hlásny and Turčáni 2013; Fassnacht et al. 2014; Bai et al. 2014; Satir et al. 2016; Vahedi 2016; Hengl et al. 2017).

Here we present a concise review of ML approaches applied to forest ecosystem studies along with an elaboration of barriers that prevent wider ML adoption. To do so, we briefly:

1) describe the general framework of ML, and then focus three particular ML algorithms for forest ecosystem research.

2) review and synthesize recently (mostly after 2008) published applications of ML in forest ecosystems.

3) discuss two bottlenecks of ML in forest ecology and some future relevant research.

4) present key conclusions with outlooks on the application of ML methods.

2. Machine learning

2.1 Background

Machine learning technology helps computers find patterns in data and use these patterns to improve predictions. At its core, the concept of ML is relatively simple and mirrors a similar process by which humans use information, experiences, and trials and errors in learning (Fig. 2). People accumulate historical experiences and generalize these experiences to speculate on novel problems where underlying assumptions or processes may not be known with the goal of predicting specific outcomes. The “training” and “predicting” processes in ML can correspond to human “generalization” and “speculation” processes. Just as experience is important in learning, historical datasets play a decisive role in ML. Within the data analytics field, ML is an approach that is used to design complex models that implement themselves for prediction (Carbonell et al. 1983). In environmental or ecological studies, these analytical models allow researchers to “output reliable, repeatable results” and discover “unknown relationships” through learning from historical datasets (Crisci et al. 2012; Domingos 2015).

2.2 Three specific machine learning algorithms used in forest ecology

ML can be divided into two large categories: supervised learning and unsupervised learning. Supervised learning provides a clear expectation of outputs after input samples have been trained through the model, such as classification and regression. Unsupervised learning is relatively unpredictable in what type of output is generated after input samples have been trained on the model. A typical example is clustering, that is, bringing together similar items (Fig. 3). Figure 4 shows the taxonomy of ML and some widely used algorithms. In the following sections, we provide a short description of the three well known and widely used ML algorithms in forest ecosystems: decision tree learning, artificial neural network (ANN), and support vector machine (SVM).

2.2.1 Tree-based learning

Decision tree (DT) learning is a predictive model and a support tool that combines a decision graph (such a bifurcating flow charts, dichotomous keys and even ‘choose your own adventure’ books designed for children) with possible outcomes or results. DT have a simple recursive structure composed of the root node, internal nodes, and leaf nodes and branches that represents the knowledge extracted from data (see Fig. 5A) (Quinlan 1987). Each internal node represents an attribute which is associated with a test or decision rule relevant to data classification. Each leaf node represents a class label and each branch represents the outcome of the test. The path from root to leaf represents the value of the target variable that is conditional to the value of the input variables. DT correspond to a logical expression and thus are often referred to “white box” models (Breiman et al. 1984). A consequence of successive partitioning is that nonlinear relationships between parameters do not affect tree performance; likewise, complex interactions are readily interpretable as successive partitioning often identify and isolate conditional variables in the initial splits of a tree. These are major advantages of tree-based models over methods such as regularized discriminant analysis (RDA) or canonical correspondence analysis (CCA). Tree methods also can be used as a good extension to a large database, while its size is independent of the database size. However, DT has more difficult to deal with missing data.

The classification and regression tree (CART) model is one of the most popular tree-based methods introduced by Breiman et al. (1984). As the name suggests, CART performs classification and regression analysis; however, it also can manage mixed variable types and missing values that DT cannot (Bell 1999). This approach has been expanded to include multivariate datasets (De’ath 2002) and thus is becoming increasingly utilized in biodiversity assessments (Work et al. 2008; Work et al. 2010; Paradis et al. 2011; Graham-Sauvé et al 2013). Another DT approach is random forest (RF) (Breiman 2001). RF methods generate and aggregate results of multiple trees using bootstrap samples of the input data (Svetnik 2003). All DT may suffer from overfitting; whereby trees are overgrown and provide terminal nodes which may not be statistically different. In CART, overfitting is avoided using cross-validation procedures. In RF reliance on multiple trees is used to avoid overfitted trees. It mainly depends on three random processes: the samples that generate DTs are randomly generated; the eigenvalues of building a DT are randomly selected; and the random direction is chosen for tree fission selection during the production process.

2.2.2 Artificial neural network

An ANN is a mathematical or computational model that mimics the structure and function of biological neural networks (i.e., the central nervous system of animals, especially the brain) (Bishop 1995; Liu et al. 2010). An ANN is composed of a large number of artificial neurons which may be favored or disfavored through a weighting processes as learning proceeds. Neurons in the ANN are organized into different layers which may perform different types of transformations on their inputs. Figure 5B shows a schematic of a typical multilayer feedforward network which includes the input layer, the output layer, and the hidden layer. Signals travel from the first (input) to the last (output) layer, possibly after traversing the layers multiple times. In most cases, ANNs can change their internal structure on the basis of external information; thus, it is an adaptive system (Haykin 2001). In other words, learning process of ANN cannot be observed directly and thus these methods are often referred to as "black box" methods. This can lead to output that is difficult to explain. However, ANNs have the ability to learn, model nonlinear and complex relationships and are also robust and fault tolerant to noisy data.

There are many powerful ANN algorithms that are used for various studies in forest ecology. The backpropagation algorithm is often used to train neural networks that calculate the errors at the output layers and are distributed back through the network layers. The backpropagation method calculates the gradient of the loss function for all weights in a network. This gradient is fed back to the optimization method to update weights by which to minimize the loss function (Rojas 2013). A cascade correlation artificial neural network (CCANN) is a supervised algorithm which was developed by Fahlman and Lebiere (1990), with the main objective of managing several perceived problems deriving from the backpropagation method. It starts with a minimal network and adds new hidden units step by step to the hidden layer. The CCANN algorithm learns very quickly because the network determines its own size and topology (Fahlman and Lebiere 1990). The self-organizing map (SOM) is a type of ANN that uses unsupervised learning to generate a low-dimensional (usually two-dimensional), discretized representation of the input space of the training sample. Unlike other ANN methods, SOM is a topographic organization for which nearby locations in the map represent inputs with similar properties (Shah-Hosseini 2011).

2.2.3 Support vector machine

The SVM algorithm uses non-parametric kernel-based techniques derived from statistical learning theory, which was primarily invented and developed by Vapnik (Vapnik 2013; Vapnik and Chervonenkis 2015). Since the mid-1990s, SVMs have been particularly appealing in addressing nonlinear classification, regression, and density estimation problems. Moreover, SVM often uses kernel functions to project the multidimensional space of data in the form of points, and then finds the best classification of the hyperplane, finally being classified according to this plane (Vapnik 2013). For example, in Fig. 5C, although both “a” and “b” are classified as hyperplanes, neither are optimal. This is because they are too close to the samples, which are highly sensitive to noise and poorly generalized. The essence of the SVM algorithm is to find a hyperplane (as seen in Fig. 5C “c”) that maximizes a value, which is the minimum distance between the hyperplane and all training samples. This minimum distance is called “margin” in SVM terms.

Furthermore, SVM has a core function which is the sequence minimum optimization (SMO) algorithm. The aim of the SMO is to find the optimal parameter α and calculate the hyperplane for classification. The SMO method can decompose a large optimization problem into several small optimization problems that greatly simplifies resolution processes (Were et al. 2015). Another important part of SVM is the kernel function. Its main function is to map data from low- to high-dimensional space and resolve nonlinear data problems without considering mapping processes (see Fig. 5D). In SVM theory, the use of different kernel functions will result in different SVM algorithms (Cristianini et al. 2000). Moreover, SVM can cope well with noisy conditions; this is because it automatically identifies and incorporates support vectors during training processes and prevents the influence of non-support vectors over the model (Cherkassky et al. 2004; Yu et al. 2006). It is also fairly robust against overfitting, especially in high-dimensional space. In addition, SVM can be trained with a few meaningful pixels and is able to fit limited information (e.g. Pouteau et al. (2012)). The main weakness of SVM is that it can be very time consuming to find the suitable kernel function (Sujay et al. 2014).

In this article, we also summarize several advantages and disadvantages of these methods (see Table 1). For more detailed information of other ML algorithms see previous reports (Haupt et al. 2008; Hsieh 2009; Michalski et al. 2013; Muhamedyev 2015; Thessen (2016).

3. Application of machine learning techniques in forest ecology

ML has been widely adopted and put into practice by researchers in light of increasing concerns over forest ecosystems, including (1) species distribution modeling; (2) C cycles; (3) hazard assessment and prediction; and (4) other applications in forest management (Table 2).

In this article, we only focus on relatively recent (mostly after 2008) published applications of ML in forest ecosystems. Previous ML applications, in environmental science and ecology, were reviewed by Haupt et al. (2008) and Hsieh (2009).

3.1 Species distribution models

Many ML algorithms have been used to study the impact of environmental changes in biodiversity. A regional-scale study of the effect of global warming on forest distribution was reported by Garzón et al. (2008). They predicted future tree species distribution using the RF algorithm in the Iberian Peninsula. Additionally, their research simulated the distributions of 20 tree species that could be impacted by climate change under four Intergovernmental Panel on Climate Change (IPCC) scenarios (i.e., A1, A2, B1, and B2) and for the time points 2020, 2050, and 2080. The results indicated that the distribution of temperate broad-leaved species and Mediterranean and sub-Mediterranean species will decrease, while the potential area of mountain conifer species will rapidly decrease suggesting that climate change could have serious potential impacts in the Iberian Peninsula. However, their models did not consider the effects on land-use change and other factors on forest distribution that could limit accurate predictions by the model. Another uncertainty is that they did not simulate species that could expand into the Iberian Peninsula (e.g., from North Africa). In addition, Vaca et al. (2011) also used the RF method to potentially improve the accuracy of coarse resolution vegetation maps by downscaling to finer resolution climatic grids. Finally, Périé and Blois (2016) assessed habitat suitability with climate change for five dominant tree species in Québec (Canada). Their SDMs used eight modeling techniques which were produced using default BIOMOD (Groner et al. 1971; Thuiller 2003; Thuiller et al. 2009) parameters where appropriate. For each species, they used a random subset of data containing 70% of the 20×20-km cells (i.e., 4,493 cells) to build SMDs, and then they evaluate the predictive performance of the models using the remaining 30% (i.e., 1,925 cells). Based on their results, they suggested that traditional whole regional vegetation assemblages could become less adapted to their traditional regions, which would significantly impact the forest economies in these regions.

SDMs are frequently used to illustrate and predict species distributions and environmental preferences. Identifying priority areas for environmental conservation is a main application for SDMs. SDMs are able to simulate the dispersal capacity of each species in order to minimize the distance between their present and future distributions while determining priority sites for conservation (Loyola et al. 2012). Through statistical and ML methods, SDMs are most regularly built by inferring occurrence–environment relationships of species. Faleiro et al. (2013) developed spatial conservation designs using SDMs to predict range migrations affected by climate and landscape changes. They also measured and reduced uncertainties associated with SDMs, which include three distance methods (i.e., BIOCLIM, Euclidian, and Gower distances); three statistical methods (i.e., GLM, GAM, and MARS); and three ML methods (i.e., RF, maximum entropy, and genetic algorithms (GA)). ML approaches outperformed these other models in terms of accuracy (i.e., the highest true skill statistics (TTS) values). The TSS range from -1 to +1, where values that equal +1 represent a perfect prediction and values equal or less than zero (0) represent a prediction no better than random, in prediction species distributions. Pouteau et al. (2012) used the SVM approach to predict rare plant distributions in island forest ecosystems. In their study, SVM performed significantly better than RF, especially when observational records were limited (the number of available training pixels ranged from 13 to 54). They reported that high conservation priority should be given to the rare species found in the low- and mid-elevation forests of Pacific islands since such areas are much more prone to extinction.

Multi-SDMs may have the potential for researchers to understand and predict the structure of forest ecosystem communities. Chapman and Purse (2011) predicted assemblages of 701 native plant distributions across Great Britain at a 10-km^2 grid scale. They compared single- and multi-SDM versions (univariate and multivariate) of two ML distribution models, based on CART and ANN algorithms. They found that multi-SDMs were slightly less accurate than single-SDMs; however, they also claimed that multi-SDM models provide a highly simplified way in which to model spatial patterns, and that fact in itself counteracts their inferior performance. Another reason is that multi-SDMs can generate more sufficiently realistic response curves when modeling shared environmental responses.

3.2 Carbon cycles

Traditional modeling approaches (both empirical and process-based modeling) have a great capacity to quantify and predict C cycles (e.g., Peng et al. 2002; Kurz et al. 2009), which mostly depends on the data used for parameterization and identification of input-output relationships, and can be upscaled from local to regional or global scales. However, the adaptability of these models are typically unsatisfactory, which generally leads to uncertain predictions if spatial and temporal conditions change. Fortunately, the adaptability of ANNs to environmental conditions is strong if training and test data are sufficient. Papale and Valentini (2003) reported on how C flux data obtained from the EUROFLUX project was used to train ANN models and to provide spatial (1 km × 1 km) and temporal (weekly) estimates of C flux (i.e., C uptake was 0.47Gt C yr^{-1}) in forests on a continental scale (Europe as a whole). Later, Papale et al. (2015) again attempted to further develop ANN methods for the prediction of gross primary production (GPP), latent heat flux (LE), and net ecosystem exchange (NEE) of CO_2 while primarily trying to assess the uncertainties in extrapolation due to sample selection. Their results showed that ANN models had a higher accuracy with both GPP and LE than with NEE. A possible reason for this is the marked influence of management practices, disturbances, and site history of NEE. However, they did not include these variables as drivers in their ANN models. Papale et al. (2015) also validated models using data obtained from different continents, and they found that the extrapolation of similar climatic and vegetation conditions was possible. However, accuracy decreased when the extrapolation was applied to regions under differing seasonal cycles.

In general, ML methods combined with traditional models (e.g., process-based models) are an efficient way to study C cycles. Shoemaker and Cropper (2008, 2010) developed a generalized southern pine leaf area index (LAI) predictive model (GSP-LAI) based on ANN methods, yielding coefficients of determination (R^2) of 0.77 and root mean square errors (RMSE) less than 0.50 during validation tests. They applied the model to predict LAI values, and then they estimated NEE through a process-based model (SPM-2) in a slash pine forest (North Central Florida). Tramontana et al. (2015) reported on the application of the RF algorithm to predict uncertainties in GPP at three different spatial scales: the site itself, the ecosystem, and Europe as a whole. Given that the use of satellite-measured data can avoid the propagation of uncertainties related to the modeled grid, they were able to confirm the importance of remote sensing data in the spatial upscaling of GPP. Subsequently, in 2016, they conducted a new study using 11 ML algorithms while applying four broad approaches (tree-based methods, regression splines, neural networks, and kernel methods) predict CO_2 and energy flux (i.e., NEE, ecosystem respiration, GPP, LE, sensible heat, and net radiation) across various ecosystem types (Tramontana et al. 2016). In their study, better predictions of flux were achieved for forested and temperate regions compared to areas under extreme climate conditions or with less data (e.g., tropic sites). They found that ML methods were able to predict across-site variability and mean seasonal cycles of the observed flux well (R^2>0.7); however, they obtained uncertainty results with 8-day deviations from the mean seasonal cycle (R^2<0.5).

Additionally, ANN methods also have excellent data mining capabilities that allow relationships to be extracted directly from the data to predict C flux. For example, Moffat et al. (2010) developed a feedforward ANN using a backpropagation algorithm to forecast daytime C flux for a deciduous broadleaf forest in Germany. Their results showed that the first dominant control of daytime response was total photosynthetic photon flux (PPF) density, and the vapor pressure deficit (VPD) was the most important non-radiative control. If climate change had caused changes in ecosystem response to its relevant climatic controls, they believed their ANN model would be able to detect these directly in the historical data better than purely empirical models would alone. Were et al. (2015) used support vector regression (SVR), ANN, and RF models to create prediction maps of soil organic carbon (SOC) stocks in forest ecosystems in Kenya, Africa. Given that the RMSE was 14.9 Mg C ha^{-1} and R^2 was 0.6, the SVR model based on an SMO algorithm was found to be the best approach in predicting SOC stocks. They remarked that data quality was very important for predictions, and that total nitrogen (TN) was the most important factor in explaining the observed variability of SOC stocks in forest ecosystems.

More recently, Li et al. (2017) developed a three-layer backpropagation neural network (BPNN) model to quantify the response of global terrestrial net primary production (NPP) to multifactor global change data from 1961 to 2010. Their results indicated that the ANN method is capable of simulating and predicting global terrestrial ecosystem NPP, yielding a simulation accuracy of 0.72 and a prediction accuracy of 0.60. Song et al. (2014) reported soil respiration (Rs) estimates for seven sub-forest types across China using an ANN model. In addition, based on comprehensive global Rs databases, Zhao et al. (2017) developed ANN models which were capable of spatially estimating global Rs and evaluating the effects of interannual climate variation on 10 major global biomes. The development of reliable global NPP and Rs databases that could be incorporated into a comprehensive benchmarking system for global land and soil C models will aid in our understanding of the mechanisms underlying variations in vegetation and soil C dynamics and in quantifying uncertainty in global C cycles.

3.3 Hazard assessment and prediction

Natural hazards caused by insect outbreaks are among the most widespread disturbances that impact the balance of forest ecosystems in different regions. Fassnacht et al. (2014) implemented a supervised classification technology (i.e., SVM) combined with an improved feature-selection approach (i.e., genetic algorithm; GA) to assess the potential of hyperspectral imagery, and they generated a map of bark beetle-induced tree mortality. Their results showed that the overall accuracy (OA) of mapping dead trees was 84%–96%, and the OA of the separation between healthy and dead trees was 94%–97%. Hlásny and Turčáni (2013) used spatial-dependence analysis, ordinary kriging, and neural network-based regression modeling to investigate the patterns of bark beetle outbreaks and the casual relationships in secondary Norway spruce forest ecosystems. They inferred that two bark beetle outbreaks (1995–1999 and 2001–2004) resulted in unsustainable secondary spruce forests in Central Europe.

The sensitivity rating of trees for diseases and pests can provide information that could be used to evaluate current or future hazardous situations in forests of concern (Mason et al. 1985). In order to enhance the predictive accuracy of forest pest occurrences, Bai et al. (2014) promoted a type of forecasting approach based on a combination of three technologies: rough set theory, particle swarm optimization (PSO), and BPNN. In view of their findings, rough set theory could effectively discard the characteristic dimension and the new POS-BP model could decrease iteration times, with an average accuracy of 94.8 %. Though Bai et al. (2014) yielded results that supported their methodology for classifying dead trees, their attempt to accurately map different mortality stages was defective. In addition, three remote sensing indicators (temperature/Vegetation Dryness Index, LAI, and canopy water content) were defined by Wang et al. (2010) and combined with ANNs to predict pest hazards in a larch forest.

Self-organizing maps (SOMs) are another type of ANN that has been well established in dealing with high-dimensional data (e.g., gridded meteorological data). Thus, SOMs are typically applied in hazard assessment models primarily driven by meteorological factors. Park et al. (2013) applied SOM and RF to identify the hazard ratings of trees (on an individual scale) and forests (on a stand scale) infested by the pine wood nematode (PWN) and consequently impacted by the pine wood disease (PWD). In their study, they combined SOM with RF to predict both the number and rate of infested trees, and they even applied this approach to evaluate the relative significance of each environmental factor in determining PWD infestation. They found that large trees which were taller and had wider crown volumes were at higher risk for PWD, and a drop in tree vigor may be caused by a susceptibility to PWD.

Fire is one of the most important disturbances for ecosystem hazard assessments as well as being a primary cause of forest destruction. Lagerquist et al. (2017) developed a new fire-weather prediction model that could be deployed in real time in northern Alberta (Canada). They implemented SOMs to predict fire spread days using six key predictors (e.g. sea-level pressure, 500 hPa geopotential height, etc.). Spread day mean extreme threshold values of three Canadian Fire Weather Index System (CFWIS) variables include the fine fuel moisture code, the initial spread index, and the fire weather index. The BPNN method is suitable for dealing with nonlinear problems due to it being capable of nonlinear mapping. Satir et al. (2016) successfully used the BPNN method to map forest fire probability in the Upper Seyhan Basin River (Turkey). Results were validated by relative operating characteristic analysis, which indicated that BPNN yielded a good coefficient of accuracy (R=0.83). Safi and Bouroumi (2013) also tested their BPNN model using a real fire dataset from the Montesinho Natural Park in Portugal and obtained a satisfactory prediction of forest fire occurrences. Sakr et al. (2010) favorably utilized a SVM algorithm to predict the fire hazard level of a day based only on previous (from 2000 to 2008) meteorological data in Lebanon.

Snow hazard is an important natural disturbance of forest growth, regeneration, and distribution (Valinger and Fridman 1999). Hlásny et al. (2011) studied a neural network-based regression model to assess snow damage in Norway spruce forests. They analyzed the relationship between environmental parameters and various types of snow damage (i.e., tree top breakage, crown breakage, stem breakage, and uprooting). Their results showed that snow hazards were largely associated with the developmental stage of the forests (i.e., forest age, height, and diameter) and not closely related to forest density or tree taper. Thus, they proposed that more ways in preserving forest health and productivity should be applied to spruce forest management to deal with snow disturbances.

3.4 Other applications in forest management

Forest mapping is a key measure in forest management. Especially in semiarid environments, soil moisture can limit leaf structure and orientation; thus, the classification of tree species becomes more difficult. Increasingly, ML methods have been used in land-cover mapping and monitoring based on remote sensing data (Lees and Ritman 1991; Gopal et al. 1999; Lawrence and Wright 2001; Pal and Mather 2003). Rogan et al. (2008) compared a Fuzzy ARTMAP neural network algorithm with two classification tree algorithms (i.e., C4.5 and S-Plus) in mapping land-cover modifications over large areas (California, USA). In their study, the ARTMAP neural network algorithm led to higher accuracy (approximately 84%) compared to the two classification tree algorithms in land-cover mapping. Moreover, the ARTMAP was also less impacted by noise and produced more stable results in large area mapping. In addition, Adelabu et al. (2013) conducted an experiment to separate *Colophospermum mopane* from coexisting species around Botswana's Central District by means of high-resolution (5 m) satellite images (i.e., RapidEye) and ML classification algorithms (i.e., RF and SVM). They proposed that SVM could be used to map plant species based on a small pixel sample, and this approach had a higher accuracy compared to RF. However, no significant difference was detected between SVM and RF when sufficient training data was available.

It is often necessary for forest managers to predict aboveground biomass (AGB). Vahedi (2016) conducted a study that compared ANN to allometric equations for forecasting AGB in mixed-beech forests in Hyrcania, Iran. The diameter at breast height (DBH), tree height, and wood density were used to train and test ANN and the allometric equation models. They reported that some statistical issues (e.g., reliability of parameters and collinearity among the parameters) influenced the development of allometric equations; however, they found that ANN did not experience these problems. Their ANN model was designed by two hidden layers and 20 neurons per layer. Their results showed that ANN resulted in the better prediction of AGB (RMSE% = 7.3) compared to allometric equations in natural forest ecosystems. In order to better manage cork-oak plantations and ensure the sustainability of the Maâmora forest, Lahssini et al. (2015) proposed an assessment to evaluate the suitability of cork oak, based on a random forest algorithm. In their evaluation model, the most important indicator was the success rate of cork-oak plantations. The model produced a map which enabled forest managers to choose the most suitable area for planting and regeneration.

Accurate measurements associated with wood volume, tree height, and stem taper are critical in forest management. First, Diamantopoulou and Milios (2010) used a CCANN model to predict the total volume of dominant pine trees in north-eastern Greece. They used the Kalman filter method (Brown and Hwang 1992) to obtain the best weight estimates in their study. They also compared results between multiple linear regressions (MLR), nonlinear regressions (NLR), and CCANN. The CCANN model performed best and proved to be a useful tool for predicting the total volume of dominant pine trees. Second, Özçelik et al. (2013) conducted a comparative analysis between three methods (i.e., mixed-effects models, generalized models, and BPNN) to obtain tree height predictions in the south and southwestern region of Turkey. This study showed that both nonlinear mixed-effects regression and BPNN could both successfully predict tree bole height with high accuracy when the variability of the height-diameter relationship from stand to stand was incorporated into the model. Finally, Nunes et al. (2016) evaluated ANN and RF methods in modeling stem taper, while comparing their results to traditional techniques (i.e., taper-based equations) across three different forest types, a tropical savanna, a seasonal semi-deciduous forest, and a rainforest in Brazil. They found that RF was not good at predicting diameter and wood volume; it tended to over predict low diameter- and under predict high-diameter values. However, the ANN model performed well in taper predictions and was determined to be better than taper-based equations.

Forest property data are typically collected by point sampling. However, researchers and forest managers often require spatially continuous data over a region of interest to make informed decisions. Although geographic information systems (GIS) and modeling techniques have traditionally been powerful tools in forest ecosystem management and conservation, spatially continuous data of environmental variables is increasingly required. Machine learning methods, such as RF and SVM, have proven their predictive accuracy in data mining fields and superior performance in various disciplines (Drake et al. 2006; Shan et al. 2006; Cutler et al. 2007; Marmion et al. 2009). Hengl et al. (2017) primarily used tree-based models, such as RF and gradient tree boosting, to account for local relationships between soil variables and covariates. For example, they used 150 000 soil profiles for training and a stack of 158 remote sensing-based soil covariates (primarily derived from MODIS land products, Shuttle Radar Topography Mission (SRTM) digital elevation model (DEM) derivatives, climatic images, and global landform and lithology maps), which they used to fit an ensemble of ML methods using the R package. The results from greater than tenfold cross validations showed that the ensemble models explained between 56% (coarse fragments) and 83% (pH) of variation with an overall average of 61%. However, this approach suffered from two limitations: (1) It is difficult to derive spatially explicit measurements of prediction accuracy using ML approaches. Although they calculated accuracy measurements using greater than tenfold cross validations, these were only global measurements. (2) Machine learning approaches are highly opaque due to the “black box” effect, and it is difficult to incorporate knowledge of soil formation processes and soil properties in the prediction algorithm. Recently, Li et al. (2011)

successfully applied ML methods, such as SVM, RF, and combined methods (e.g., a hybrid method where RF is combined with ordinary kriging; RFOK) to spatial interpolation, and its prediction error (relative mean absolute error; RMAE) was less than 19% of the control (using the inverse distance squared (IDS)). Prediction errors of this method were also less than 30% compared to the best methods published in literature. This study demonstrated that combined ML method approaches and other existing spatial interpolation techniques have a great potential and shed a new light on a potential direction for future studies in order to select statistical methods for spatial interpolation.

4. Discussion

4.1 Bottlenecks of machine learning in forest ecology

As described above, ML is a powerful classification, modeling, and prediction tool in forest ecology research. Specifically, ML models have a higher accuracy and faster capacity in resolving complex issues, analyzing interactions, and predicting nonlinear system behavior. However, there are two bottlenecks that limit further expansion of ML technologies.

On the one hand, the lack of suitable data (in both quantity and quality) is a major bottleneck that prevents the widespread application of ML methods in forest ecology. None the less, compared to traditional empirical and process-based models that require frequent parameter initialization and adjustment under different conditions (e.g., climate, region, or disturbance), ML technologies have stronger environmental adaptability. This strength of ML methods, however, requires more rigorous training and test data. However, long-term and highly-accurate monitoring is expensive monitoring, data collection, storage and up-dating can be disrupted by: reduced funding, instrument failure, limitations of historical technologies, interference by human activities, and so on. For example, the loss of historical data required

Wang et al. (2010) to perform additional analysis, which made validation difficult. Additionally, Papale et al. (2015) suggested that more study sites would be needed to provide the necessary data as well as to reliably forecast GPP, LE, and NEE fluxes that are crucial for predicting global C cycles. Tramontana et al. (2016) also proposed that the number of eddy-covariance sites be increased, especially in poorly represented regions (e.g., the tropics, which account for a disproportionate share of global terrestrial water and C flux), to improve the predictive capacity of ML methods. In the coming years, the development of big data research and data sharing may be an effective way to resolve the problem of insufficient data.

On the other hand, the relatively "higher threshold" of application is another key constraint for the widespread use of ML. For example, the different algorithms used in ANN determine the number of processing elements in the hidden layer(s) as well as the number of hidden layers. The development of black box package algorithm has greatly simplified the application of ANN since many ecologists only need to know what the characteristics of the different algorithms are. They will thus now be able to apply the ANN method in their own research. That being said, there are no new algorithms specifically designed for forest ecosystem research. Currently, not only ANN but also most ML algorithms are very complex. These algorithms typically require strong mathematical skills and major investments in time to understand them in detail (Thessen 2016) as well as to avoid "black box" and overfitting problems. Although the unfamiliarity with "black box" effects does not necessarily hamper the use of ML algorithms, it may influence which algorithms are selected by users. More specifically, it also influences the adaptation of the algorithm itself to different environments by developers. Overfitting is usually a product of nonparametric and nonlinear ML models that have more flexibility when learning a target function. To avoid overfitting in ML models, users must understand in greater detail the principles of different algorithms. For instance, Diamantopoulou and Milios (2010) randomly divided their data into training (80% of the total data) and test (the remaining 20%) datasets. The test data were only used to examine model performance with the new dataset. In addition, having a certain level of programming skill is also necessary for applying ML methods. Although convenient tools (e.g., MATLAB and the R programming language) have already provided powerful and friendly user interfaces (UI) which aim to reduce many user barriers, ML users still need to master some necessary skills that, many ecologists lack unfortunately, to debug parameters. Finally, if ML is to be more frequently used in forest ecology, ecologists need better mathematical proficiency and more training skills in programming (e.g., through taking workshops or applying for summer courses) to ensure that they understand algorithms and potential problems such as overfitting.

4.2 Challenges and future directions

To help better understand the various ecological mechanisms in forests and to find new ways to address problems, we suggest applying a combination of different ML methods as well as a combination of ML methods with traditional statistical methods. With this in mind, forest ecologists must understand that there is no universal best ML method. The choice of a specific method or a combination of methods depend on specific users and the questions they're asking (Flach 2001; Crisci et al. 2012; Bhattacharya 2013). Park et al. (2013) confirmed that a combination of SOM and RF was effective in extracting ecological information from a dataset. Diamantopoulou and Milios (2010) also built a novel model that used multivariate analysis to decrease and select a minimum set of tree measurements. They then introduced this set to CCANN models in which the Kalman learning algorithm was embedded for training. Bai et al. (2014) applied rough set theory to eliminate redundancy attributes, for which input factors could be reduced from 16 to eight. Following this, Bai et al. (2014) used a PSO algorithm to optimize weights and thresholds in BPNN.

With the rapid development of computing power, more complex ML algorithms can be implemented more rapidly when trained by larger datasets. This trend will promote the more extensive application of ML. For instance, most SDMs based on ML methods incorporate remote sensing as extremely large data sets are being generated (e.g., Garzón et al. 2008; Chapman and Purse 2011; Vaca et al. 2011; Pouteau et al. 2012; Faleiro et al. 2013). Deep learning is a machine learning method based on feature learning. From 2006 to 2014, deep learning, especially deep neural network, had achieved rapid development (Schmidhuber 2015). In 2006, Hinton et al. (2006) introduced the Deep Belief Network (DBN) that took ML into a new phase of deep learning. Feature learning aims to find better representations and to create better models for learning from large unlabeled data (LeCun et al. 2015). Today, deep learning has swept across many fields, such as computer vision, speech recognition, natural language processing, audio recognition, and bioinformatics fields (Turovsky 2016; Ghasemi et al. 2017). However, the higher data requirements and more complex model architectures have slowed the application of deep learning to forest ecosystems and sustainable development. However, a successful example is that of Jean et al. (2016) who combined satellite imagery and a novel convolutional neural network (CNN) model (a type of deep learning method) to quantify and predict poverty in developing countries in Africa. This demonstrates that ML techniques can still be powerful when applied to a setting with limited training data. Ecologists need to be more effectively engaged with ML research in the future, especially deep learning developers so that more techniques can be specifically developed for the field. For instance, SDM studies will potentially achieve higher accuracy and wider application by combining remote sensing and deep learning methods if developers can create more suitable models with the help of ecologists. In addition, since ML promotes the development and application of smart devices more specialized and more intelligent observation equipment or technologies will help ecologist to increase the quantity of empirical data.

Uncertainty plays a fundamental role in all ML applications. Many aspects of ML crucially depend on a careful probabilistic representation of uncertainty. One way to deal with uncertainties effectively is to develop a probabilistic machine learning (PML) approach which provides a framework for representing and manipulating uncertainty related to data, models, and predictions (Ghahramani 2015). The PML approach to ML and artificial intelligence is a very active area of research with wide-ranging impacts beyond conventional pattern-recognition problems. It will thus continue to play a central role in the development of ever more powerful ML systems for future application in forest ecosystems.

5. Summary

This study primarily dealt with literature related to ML applications focused on forest ecology. However, despite our encouragement for greater use of ML, ML approaches are not meant to and will never be able to answer all issues related to forest ecology. ML methods, however, provides many useful tools that should be more critically considered to deal with some relevant ecological problems.

In this study, we propose two outlooks for future research that is related to this topic:

(1) Data mining, especially deep mining, has consistently been the strength of ML approaches. As greater and greater data sharing becomes a reality, ML approaches will be the best choice

for ecologists in the face of an influx of a massive amount of research data, especially at a global scale.

(2) With the rapid development of deep learning techniques, image and voice recognition technologies have progressively improved although not yet perfected. We thus boldly predict that researchers will be able to apply not only the statistics and analysis of numerical and remote sensing data, but also the application of other types of data (such as sound and video) to study forest ecology in the near future.

Acknowledgements

This study was a part of a research project recently funded by the Fonds de recherche du Québec (FQRNT) program and the Natural Sciences and Engineering Research Council of Canada (NSERC) Discover Grant.

References

Adelabu, S., Mutanga, O., Adam, E., & Cho, M. A. 2013. Exploiting machine learning algorithms for tree species classification in a semiarid woodland using RapidEye image. *Journal of Applied Remote Sensing,* 7(1), 073480–073480.

Bai, T., Meng, H., & Yao, J. 2014. A forecasting method of forest pests based on the rough set and PSO-BP neural network. *Neural Computing and Applications,* 25(7-8), 1699–1707.

Bell, J. F. 1999. Tree-based methods. *Machine learning methods for ecological applications,* 89–105.

Bhattacharya, M. 2013. Machine Learning for Bioclimatic Modelling. *International Journal of Advanced Computer Science and Applications.* (4), 1.

Bishop, C. M. 1995. Neural networks for pattern recognition. Oxford university press. Inc. New York, NY, USA. P.482.

Breiman, L., Friedman, J., Stone, C. J., & Olshen, R. A. 1984. Classification and regression trees. CRC press.

Breiman, L., 2001. Random forests. *Mach. Learning,* 45, 5–32.

Brown, R. G., & Hwang, P. Y. 1992. Introduction to random signals and applied Kalman filtering. Willey, New York.

Carbonell, J. G., Michalski, R. S., & Mitchell, T. M. 1983. Machine learning: A historical and methodological analysis. *AI Magazine,* 4(3), 69.

Chapman, D. S., & Purse, B. V. 2011. Community versus single-species distribution models for British plants. *Journal of biogeography,* 38(8), 1524–1535.

Cherkassky, V., & Ma, Y. 2004. Practical selection of SVM parameters and noise estimation for SVM regression. *Neural networks,* 17(1), 113–126.

Crisci, C., Ghattas, B., & Perera, G. 2012. A review of supervised machine learning algorithms and their applications to ecological data. *Ecological Modelling,* 240, 113–122.

Cristianini, N., & Shawe-Taylor, J. 2000. An introduction to support vector machines and other kernel-based learning methods. Cambridge university press.

Cutler, D. R., Edwards, T. C., Beard, K. H., Cutler, A., Hess, K. T., Gibson, J., & Lawler, J. J. 2007. Random forests for classification in ecology. *Ecology,* 88(11), 2783–2792.

De'ath, G. 2002. Multivariate regression trees: a new technique for modeling species–environment relationships. Ecology, 83(4), 1105-1117.

Diamantopoulou, M. J., & Milios, E. 2010. Modelling total volume of dominant pine trees in reforestations via multivariate analysis and artificial neural network models. *Biosystems engineering,* 105(3), 306–315.

Domingos, P. 2015. The master algorithm: How the quest for the ultimate learning machine will remake our world. *Basic Books.*

Drake, J. M., Randin, C., & Guisan, A. 2006. Modelling ecological niches with support vector machines. *Journal of Applied Ecology,* 43(3), 424–432.

Fahlman, S. E., & Lebiere, C. 1990. The cascade-correlation learning architecture. *In Advances in neural information processing systems* (pp. 524–532).

Fassnacht, F. E., Latifi, H., Ghosh, A., Joshi, P. K., & Koch, B. 2014. Assessing the potential of hyperspectral imagery to map bark beetle-induced tree mortality. Remote sensing of environment, 140, 533–548.

Faleiro, F. V., Machado, R. B., & Loyola, R. D. 2013. Defining spatial conservation priorities in the face of land-use and climate change. *Biological Conservation,* 158, 248–257.

Flach, P. A. 2001. On the state of the art in machine learning: A personal review. *Artificial Intelligence,* 131(1-2), 199–222.

Garzón, M. B., de Dios, R. S., & Ollero, H. S. 2008. Effects of climate change on the distribution of Iberian tree species. *Applied Vegetation Science,* 11(2), 169–178.

Ghahramani, Z. 2015. Probabilistic machine learning and artificial intelligence. *Nature,* 521(7553), 452–459.

Ghasemi, F., Mehridehnavi, A. R., Fassihi, A., & Pérez-Sánchez, H. 2017. Deep Neural Network in Biological Activity Prediction using Deep Belief Network. *Applied Soft Computing.*

Gopal, S., Woodcock, C. E., & Strahler, A. H. 1999. Fuzzy neural network classification of global land cover from a 1 AVHRR data set. *Remote Sensing of Environment,* 67(2), 230–243.

Graham-Sauvé, L., Work, T. T., Kneeshaw, D., & Messier, C. 2013. Shelterwood and multicohort management have similar initial effects on ground beetle assemblages in boreal forests. Forest ecology and management, 306, 266-274.

Groner, G. F., Clark, R. L., Berman, R. A., & DeLand, E. C. 1971, November. BIOMOD: an interactive computer graphics system for modeling. In Proceedings of the November 16-18, 1971, *fall joint computer conference* (pp. 369–378). ACM.

Gunn, S. R. 1998. Support vector machines for classification and regression. *ISIS technical report,* 14, 85–86.

Haupt, S. E., Pasini, A., & Marzban, C. (Eds.). 2008. Artificial intelligence methods in the environmental sciences. Springer Science & Business Media.

Haykin, S. S. 2001. Neural networks: a comprehensive foundation. Tsinghua University Press.

Hengl, T., de Jesus, J. M., Heuvelink, G. B., Gonzalez, M. R., Kilibarda, M., Blagotić, A., & Guevara, M. A. 2017. SoilGrids250m: Global gridded soil information based on machine learning. *PloS one,* 12(2), e0169748.

Hinton, G. E., Osindero, S., & Teh, Y. W. 2006. A fast learning algorithm for deep belief nets. *Neural computation,* 18(7), 1527–1554.

Hlásny, T., Křístek, Š., Holuša, J., Trombik, J., & Urbaňcová, N. 2011. Snow disturbances in secondary Norway spruce forests in Central Europe: regression modeling and its implications for forest management. *Forest Ecology and Management,* 262(12), 2151–2161.

Hlásny, T., & Turčáni, M. 2013. Persisting bark beetle outbreak indicates the unsustainability of secondary Norway spruce forests: case study from Central Europe. *Annals of forest science,* 70(5), 481–491.

Hsieh, W. W. 2009. Machine learning methods in the environmental sciences: Neural networks and kernels. Cambridge university press.

Jean, N., Burke, M., Xie, M., Davis, W. M., Lobell, D. B., & Ermon, S. 2016. Combining satellite imagery and machine learning to predict poverty. *Science,* 353(6301), 790–794.

Kurz, W. A., Dymond, C. C., White, T. M., Stinson, G., Shaw, C. H., Rampley, G. J., ... & Metsaranta, J. 2009. CBM-CFS3: a model of carbon-dynamics in forestry and land-use change implementing IPCC standards. *Ecological modelling,* 220(4), 480–504.

Lagerquist, R., Flannigan, M. D., Wang, X., & Marshall, G. A. 2017. Automated prediction of extreme fire weather from synoptic patterns in northern Alberta, Canada. *Canadian Journal of Forest Research,* 47(9), 1175–1183.

Lahssini, S., Lahlaoi, H., Mharzi Alaoui, H., Hlal, E. A., Bagaram, M., & Ponette, Q. 2015. Predicting Cork Oak Suitability in Maâmora Forest Using Random Forest Algorithm. *Journal of Geographic Information System,* 7(02), 202.

Lawrence, R. L., & Wright, A. 2001. Rule-based classification systems using classification and regression tree (CART) analysis. *Photogrammetric engineering and remote sensing,* 67(10), 1137–1142.

LeCun, Y., Bengio, Y., & Hinton, G. 2015. Deep learning. Nature, 521(7553), 436–444.

Lees, B. G., & Ritman, K. 1991. Decision-tree and rule-induction approach to integration of remotely sensed and GIS data in mapping vegetation in disturbed or hilly environments. *Environmental Management,* 15(6), 823–831.

Li, J., Heap, A. D., Potter, A., & Daniell, J. J. 2011. Application of machine learning methods to spatial interpolation of environmental variables. *Environmental Modelling & Software,* 26(12), 1647–1659.

Li, P., Peng, C., Wang, M., Li, W., Zhao, P., Wang, K., Yang, Y., & Zhu, Q. 2017. Quantification of the response of global terrestrial net primary production to multifactor global change. *Ecological Indicators,* 76, 245–255.

Liu, Z., Peng, C., Xiang, W., Tian, D., Deng, X., & Zhao, M. 2010. Application of artificial neural networks in global climate change and ecological research: An overview. *Chinese science bulletin,* 55(34), 3853–3863.

Loyola, R. D., Lemes, P., Faleiro, F. V., Trindade-Filho, J., & Machado, R. B. 2012. Severe loss of suitable climatic conditions for marsupial species in Brazil: challenges and opportunities for conservation. PloS one, 7(9), e46257.

Marmion, M., Parviainen, M., Luoto, M., Heikkinen, R. K., & Thuiller, W. 2009. Evaluation of consensus methods in predictive species distribution modelling. *Diversity and distributions,* 15(1), 59–69.

Mason, G. N., Lorio Jr, P. L., Belanger, R. P., & Nettleton, W. A. 1985. Rating the susceptibility of stands to southern pine beetle attack. *Agriculture handbook-United States Department of Agriculture (USA).*

Michalski, R. S., Carbonell, J. G., & Mitchell, T. M. (Eds.). 2013. Machine learning: An artificial intelligence approach. *Springer Science & Business Media.*

Moffat, A. M., Beckstein, C., Churkina, G., Mund, M., & Heimann, M. 2010. Characterization of ecosystem responses to climatic controls using artificial neural networks. *Global change biology,* 16(10), 2737–2749.

Muhamedyev, R. 2015. Machine learning methods: An overview. *CMNT.*-19 (6), 14-29.

Nunes, M. H., & Görgens, E. B. 2016. Artificial Intelligence Procedures for Tree Taper Estimation within a Complex Vegetation Mosaic in Brazil. *PloS one,* 11(5), e0154738.

Özçelik, R., Diamantopoulou, M. J., Crecente-Campo, F., & Eler, U. 2013. Estimating Crimean juniper tree height using nonlinear regression and artificial neural network models. *Forest ecology and management,* 306, 52–60.

Pal, M., & Mather, P. M. 2003. An assessment of the effectiveness of decision tree methods for land cover classification. *Remote sensing of environment,* 86(4), 554–565.

Papale, D., & Valentini, R. 2003. A new assessment of European forests carbon exchanges by eddy fluxes and artificial neural network spatialization. *Global Change Biology,* 9(4), 525–535.

Papale, D., Black, T. A., Carvalhais, N., Cescatti, A., Chen, J., Jung, M., & Merbold, L. 2015. Effect of spatial sampling from European flux towers for estimating carbon and water fluxes with artificial neural networks. *Journal of Geophysical Research: Biogeosciences,* 120(10), 1941–1957.

Paradis, S., & Work, T. T. 2011. Partial cutting does not maintain spider assemblages within the observed range of natural variability in Eastern Canadian black spruce forests. Forest Ecology and Management, 262(11), 2079-2093.

Park, Y. S., Chung, Y. J., & Moon, Y. S. 2013. Hazard ratings of pine forests to a pine wilt disease at two spatial scales (individual trees and stands) using self-organizing map and random forest. *Ecological informatics,* 13, 40–46.

Peng, C., Liu, J., Dang, Q., Apps, M. J., & Jiang, H. 2002. TRIPLEX: a generic hybrid model for predicting forest growth and carbon and nitrogen dynamics. *Ecological Modelling,* 153(1), 109–130.

Périé, C., & de Blois, S. 2016. Dominant forest tree species are potentially vulnerable to climate change over large portions of their range even at high latitudes. *PeerJ,* 4, e2218.

Pouteau, R., Meyer, J. Y., Taputuarai, R., & Stoll, B. 2012. Support vector machines to map rare and endangered native plants in Pacific islands forests. *Ecological Informatics,* 9, 37-46.

Quinlan, J. R. 1987, August. Generating production rules from decision trees. *In ijcai* (Vol. 87, pp. 304–307).

Rogan, J., Franklin, J., Stow, D., Miller, J., Woodcock, C., & Roberts, D. 2008. Mapping land-cover modifications over large areas: A comparison of machine learning algorithms. *Remote Sensing of Environment,* 112(5), 2272–2283.

Rojas, R. 2013. Neural networks: a systematic introduction. *Springer Science & Business Media* (pp. 151–183)

Safi, Y., & Bouroumi, A. 2013. Prediction of forest fires using artificial neural networks. *Applied Mathematical Sciences,* 7(6), 271–286.

Sakr, G. E., Elhajj, I. H., Mitri, G., & Wejinya, U. C. 2010, July. Artificial intelligence for forest fire prediction. *In Advanced Intelligent Mechatronics (AIM), 2010 IEEE/ASME International Conference on* (pp. 1311–1316). *IEEE.*

Satir, O., Berberoglu, S., & Donmez, C. 2016. Mapping regional forest fire probability using artificial neural network model in a Mediterranean forest ecosystem. *Geomatics, Natural Hazards and Risk,* 7(5), 1645–1658.

Schmidhuber, J. 2015. Deep learning in neural networks: An overview. Neural networks, 61, 85-117.

Schmitt, C. B., Burgess, N. D., Coad, L., Belokurov, A., Besançon, C., Boisrobert, L., ... & Kapos, V. 2009. Global analysis of the protection status of the world's forests. *Biological Conservation,* 142(10), 2122–2130.

Shah-Hosseini, H. 2011. Binary tree time adaptive self-organizing map. *Neurocomputing,* 74(11), 1823–1839.

Shan, Y., Paull, D., & McKay, R. I. 2006. Machine learning of poorly predictable ecological data. *Ecological Modelling,* 195(1), 129–138.

Shoemaker, D. A., & Cropper Jr, W. P. 2008. Prediction of leaf area index for southern pine plantations from satellite imagery using regression and artificial neural networks. *In Proceedings of the 6th Southern Forestry and Natural Resources GIS Conference 2008*, 139–160.

Shoemaker, D. A., & Cropper, W. P. 2010. Application of remote sensing, an artificial neural network leaf area model, and a process-based simulation model to estimate carbon storage in Florida slash pine plantations. *Journal of Forestry Research,* 21(2), 171–176.

Song, X., Peng, C., Zhao, Z., Zhang, Z., Guo, B., Wang, W., ... & Zhu, Q. 2014. Quantification of soil respiration in forest ecosystems across China. *Atmospheric environment,* 94, 546–551.

Sujay Raghavendra, N., Deka, P.C., 2014. Support vector machine applications in the field of hydrology: a review. *Appl. Soft Comput.* 19, 372–386.

Svetnik, V., Liaw, A., Tong, C., Culberson, J. C., Sheridan, R. P., & Feuston, B. P. 2003. Random forest: a classification and regression tool for compound classification and QSAR modeling. *Journal of chemical information and computer sciences,* 43(6), 1947–1958.

Thessen, A. 2016. Adoption of machine learning techniques in ecology and earth science. *One Ecosystem,* 1, e8621.

Thuiller, W. 2003. BIOMOD–optimizing predictions of species distributions and projecting potential future shifts under global change. *Global change biology,* 9(10), 1353–1362.

Thuiller, W., Lafourcade, B., Engler, R., & Araújo, M. B. 2009. BIOMOD–a platform for ensemble forecasting of species distributions. *Ecography,* 32(3), 369–373.

Tramontana, G., Ichii, K., Camps-Valls, G., Tomelleri, E., & Papale, D. 2015. Uncertainty analysis of gross primary production upscaling using Random Forests, remote sensing and eddy covariance data. *Remote Sensing of Environment,* 168, 360–373.

Tramontana, G., Jung, M., Schwalm, C. R., Ichii, K., Camps-Valls, G., Ráduly, B., & Merbold, L. 2016. Predicting carbon dioxide and energy fluxes across global FLUXNET sites with regression algorithms.

Tripathi, S., Srinivas, V. V., & Nanjundiah, R. S. 2006. Downscaling of precipitation for climate change scenarios: a support vector machine approach. *Journal of hydrology,* 330(3), 621–640.

Turovsky, B. 2016. Found in translation: More accurate, fluent sentences in Google Translate. *Blog. Google.* November, 15.

Vaca, R. A., Golicher, D. J., & Cayuela, L. 2011. Using climatically based random forests to downscale coarse□grained potential natural vegetation maps in tropical Mexico. *Applied Vegetation Science,* 14(3), 388–401.

Vahedi, A. A. 2016. Artificial neural network application in comparison with modeling allometric equations for predicting above-ground biomass in the Hyrcanian mixed-beech forests of Iran. *Biomass and Bioenergy,* 88, 66–76.

Valinger, E., & Fridman, J. 1999. Models to assess the risk of snow and wind damage in pine, spruce, and birch forests in Sweden. *Environmental Management,* 24(2), 209–217.

Vapnik, V. N., & Chervonenkis, A. Y. 2015. On the uniform convergence of relative frequencies of events to their probabilities. *In Measures of complexity* (pp. 11–30). *Springer International Publishing.*

Vapnik, V. 2013. The nature of statistical learning theory. Springer science & business media.

Wang, L., Huang, H., & Luo, Y. 2010, July. Remote sensing of insect pests in larch forest based on physical model. In Geoscience and Remote Sensing Symposium (IGARSS), *2010 IEEE International* (pp. 3299–3302). *IEEE.*

Were, K., Bui, D. T., Dick, Ø. B., & Singh, B. R. 2015. A comparative assessment of support vector regression, artificial neural networks, and random forests for predicting and mapping soil organic carbon stocks across an Afromontane landscape. *Ecological Indicators,* 52, 394–403.

Work, T. T., Koivula, M., Klimaszewski, J., Langor, D., Spence, J., Sweeney, J., & Hébert, C. 2008. Evaluation of carabid beetles as indicators of forest change in Canada. The Canadian Entomologist, 140(4), 393-414.

Work, T. T., Jacobs, J. M., Spence, J. R., & Volney, W. J. 2010. High levels of green‐tree retention are required to preserve ground beetle biodiversity in boreal mixedwood forests. Ecological Applications, 20(3), 741-751.

Yu, P. S., Chen, S. T., & Chang, I. F. 2006. Support vector regression for real-time flood stage forecasting. *Journal of Hydrology,* 328(3), 704–716.

Zhao, Z., Peng, C., Yang, Q., Meng, F. R., Song, X., Chen, S., ... & Zhu, Q. 2017. Model prediction of biome- specific global soil respiration from 1960 to 2012. *Earth's Future* (in press).

Figure captions

Figure 1 The number and proportion of publications searched by the topics “three different ML methods” and “forest ecosystem” on the ISI Web of Knowledge from 2008 to 2017. Three different ML methods are included: ANN represents artificial neural network; SVM represents support vector machine; DT represents decision tree.

Figure 2 Analogy of machine learning and human thinking.

Figure 3 Schematic of clusters.

Figure 4 Taxonomy of machine learning algorithms.

Figure 5 (A) Decision tree schematic; (B) The schematic for a common multilayer feedforward network; (C) Support vector machine (SVM) schematic; and (D) SVM project data from low dimensional to high dimensional space and the determination of the hyperplane for classification.

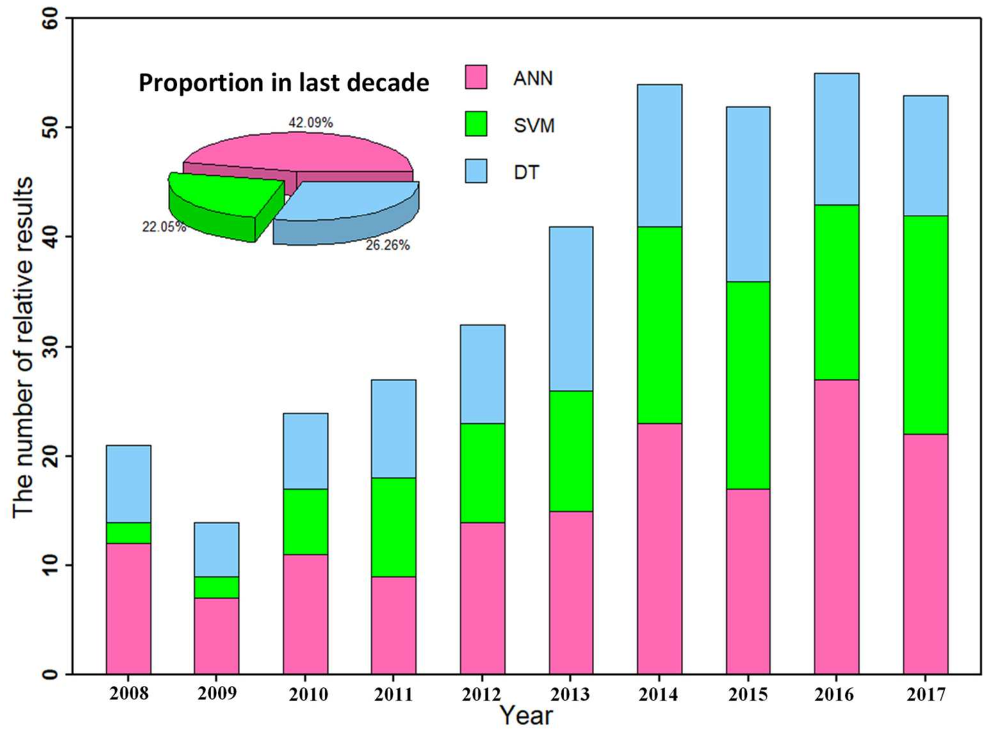

Figure 1 The number and proportion of publications searched by the topics "three different ML methods" and "forest ecosystem" on the ISI Web of Knowledge from 2008 to 2017. Three different ML methods are included: ANN represents artificial neural network; SVM represents support vector machine; DT represents decision tree.

129x95mm (300 x 300 DPI)

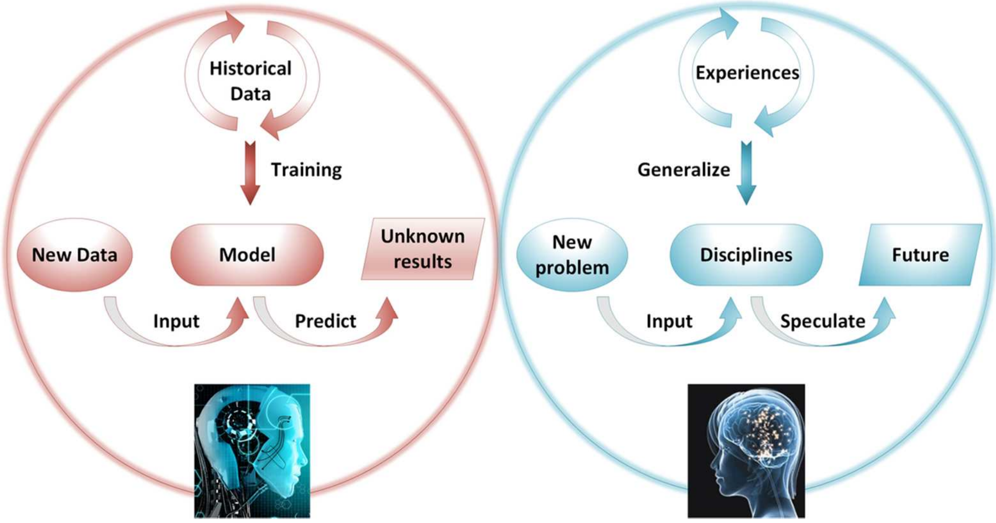

Figure 2 Analogy of machine learning and human thinking.

104x54mm (300 x 300 DPI)

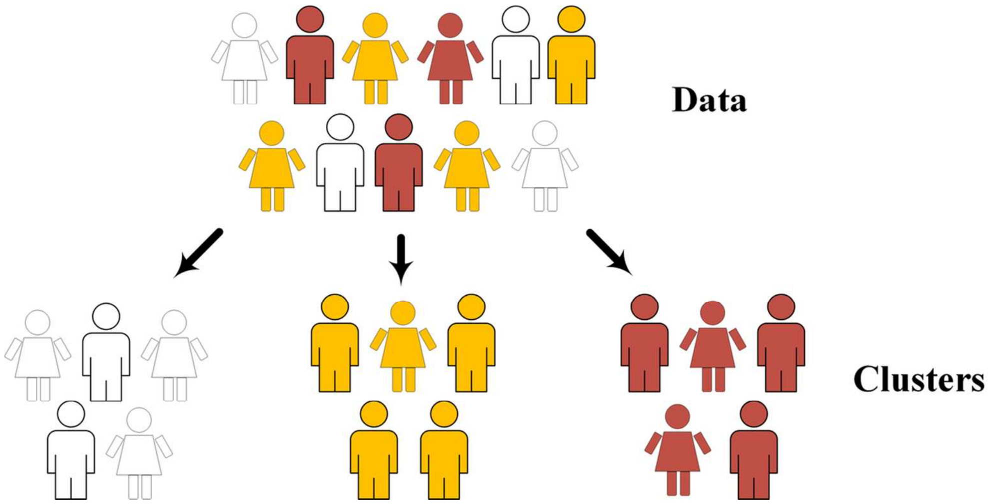

Figure 3 Schematic of clusters.

112x55mm (300 x 300 DPI)

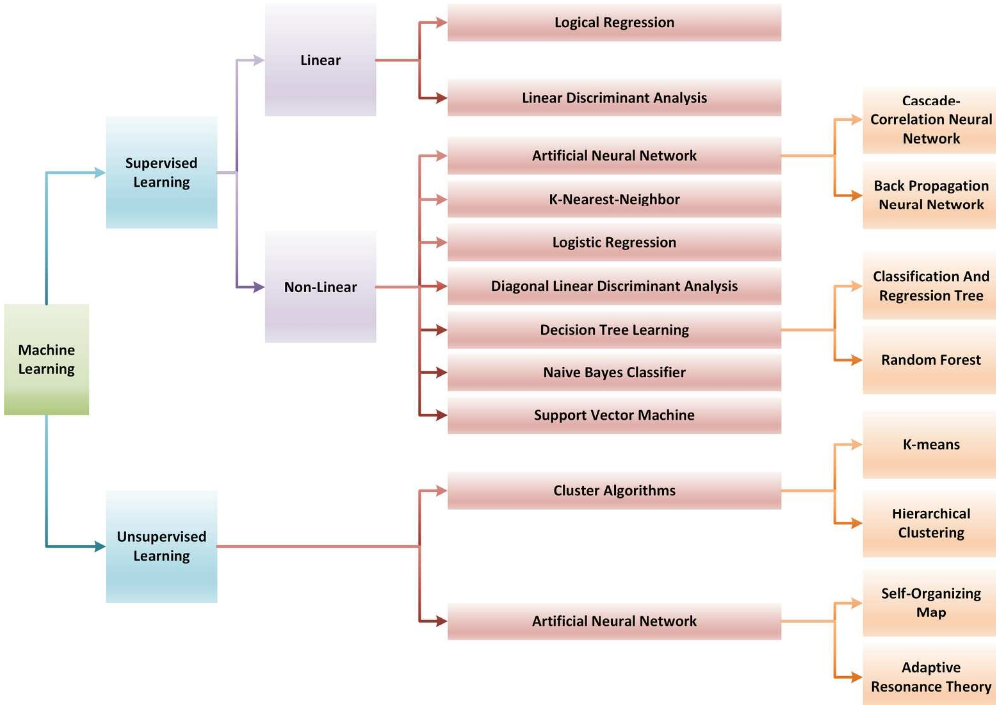

Figure 4 Taxonomy of machine learning algorithms.

188x133mm (300 x 300 DPI)

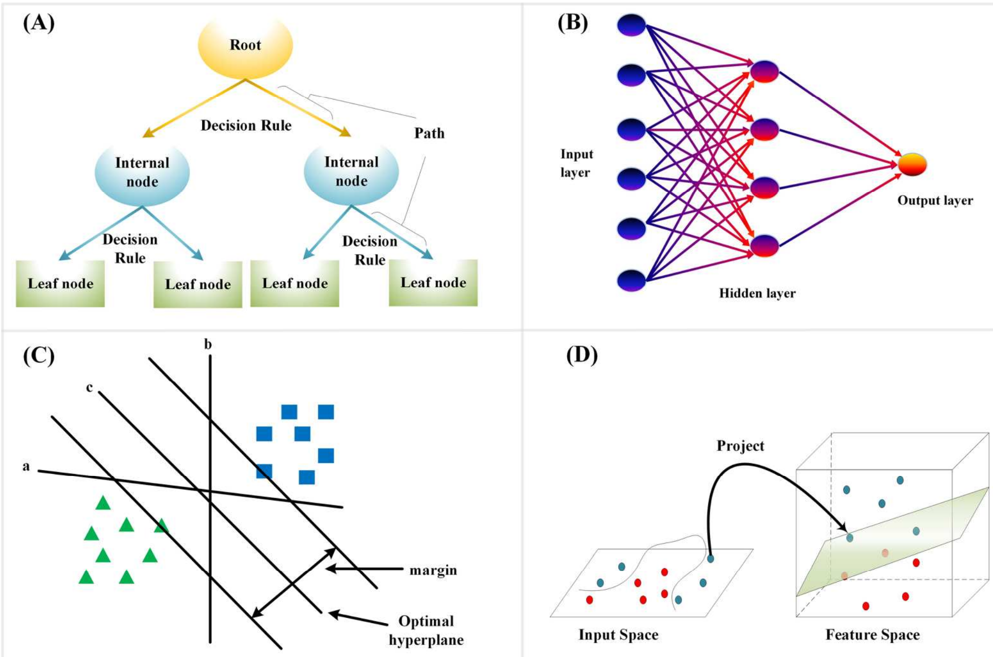

Figure 5 (A) Decision tree schematic; (B) The schematic for a common multilayer feedforward network; (C) Support vector machine (SVM) schematic; and (D) SVM project data from low dimensional to high dimensional space and the determination of the hyperplane for classification.

180x119mm (300 x 300 DPI)

Table 1 Strengths and weaknesses of decision tree learning, artificial neural networks, and support vector machines

	Strengths	Weaknesses	Reference
Decision Tree Learning	Nonlinear relationships between parameters do not affect tree performance; thus, tree methods require relatively little effort for users in data preparation. Ease of interpretation and understanding are the best features of using trees for analytics. Tree methods also can be used as a good extension to a large database, while its size is independent of the database size. CART is robust to the effects of outliers in the output. RF can effectively reduce the risk of overfitting.	Missing data will effect decision trees, and overfitting may result. It has more difficult to deal with missing data.	Breiman et al. 1984; Tramontana et al. 2015
Artificial Neural Networks (ANNs)	ANNs have the ability to learn as well as model nonlinear and complex relationships. They are also robust and fault tolerant to noisy data. ANNs have a strong capacity for parallel processing.	Learning process cannot be observed in a black box, which leads to output that is difficult to explain. ANNs are unable to identify the relative importance and effects of individual environmental variables.	Thuiller 2003; Liu et al. 2010
Support Vector Machine (SVM)	SVMs can model nonlinear decision boundaries, and there are many kernels to choose from. It is also fairly robust against overfitting, especially in high-dimensional space. SVMs can be trained with a few meaningful pixels and is able to fit limited information.	SVM is memory intensive, trickier to tune owing to the importance of picking the correct kernel, and it does not scale well to larger datasets. Poor model extrapolation will result if prior data is inconsistent as the model completely depends on the past records as support vectors.	Gunn 1998; Tripathi et al. 2006; Vapnik 2013; Adelabu et al. 2013

1 Table 2 The application and highlights of machine learning in forest ecology

Applications	Methodology	Highlights	References
Predicted species distribution with climate change	Random Forest	RF can improve the accuracy of predictions	Garzón et al. (2008); Vaca et al. (2011); Faleiro et al. (2013); Périé and Blois (2016)
	Support Vector Machine	SVM performs well when observed records are limited	Pouteau et al. (2012)
Prediction of carbon and energy flux	Artificial Neural Network	ANNs combined with traditional models are an effective way to reduce uncertainty predictions	Papale and Valentini (2003); Papale et al. (2015); Shoemaker and Cropper (2008, 2010); Tramontana et al. (2015, 2016)
		ANNs have the excellent data mining capacity	Moffat et al. (2010); Li et al. (2017); Zhao et al. (2017)
Hazard assessment and prediction	Artificial Neural Network	ANNs can be well applied and deal with high-dimensional data	Wang et al. (2010); Park et al. (2013); Bai et al. (2014)
		ANNs are capable of nonlinear mapping	Satir et al. (2016); Safi and Bouroumi (2013)
	Support Vector Machine	SVM is a powerful tool for resolving classification issues	Sakr et al. (2010); Fassnacht et al. (2014)
Forest management	Artificial Neural Network	ANNs are good at predicting aboveground biomass, wood volume, tree height, and stem taper	Diamantopoulou and Milios (2010); Özçelik et al. (2013); Vahedi (2016); Nunes et al. (2016)
	Random Forest	Combining RF and other spatial interpolation approaches has great potential	Hengl et al. (2017); Li et al. (2011)

2
3