INTERNATIONAL JOURNAL OF FOREST ENGINEERING
https://doi.org/10.1080/14942119.2024.2380230

Machine learning applications in forest and biomass supply chain management: a review

Jinghan Zhao[a,b], Jingxin Wang[a,b,c], and Nathaniel Anderson[d]

[a]Center for Sustainable Biomaterials & Bioenergy, West Virginia University, Morgantown, WV, USA; [b]Division of Forestry and Natural Resources, West Virginia University, Morgantown, WV, USA; [c]Department of Forest Biomaterials, North Carolina State University, Raleigh, NC, USA; [d]Research and Development, USDA Forest Service Rocky Mountain Research Station, Missoula, MT, USA

ABSTRACT

Forest and biomass crops for bioenergy and bioproducts can promote a sustainable bioeconomy while effectively reducing greenhouse gas (GHG) emissions to mitigate global warming. One of the most concerning issues is selecting and using appropriate modeling and analytical technologies to optimize the benefits of multi-feedstock biomass supply chains, including logistics. Machine learning (ML) has been used to solve increasingly complex supply chain problems, providing powerful tools for sustainable forest management and biomass resource development. Existing research is extensive and spans many different ML techniques, but synthesis is needed to help guide the adoption of these rapidly evolving tools. This review summarizes ML applications in forest and biomass supply chain management in terms of data, algorithms, and process examples, with an emphasis on direct application to supply chain management. ML is a viable technique to support strategic, operational, and tactical planning and decision-making in this field and can enhance the environmental and economic performance of diverse forest and biomass supply chains.

ARTICLE HISTORY
Received 23 July 2023
Accepted 24 June 2024

KEYWORDS
Machine learning; algorithm; modeling; forest resources; biomass; supply chain management

Introduction

Globally, bioenergy accounts for 55% of renewable energy and over 6% of the total energy supplied globally, with that share expanding to meet targets for greenhouse gas (GHG) emissions reductions (IEA 2022). In 2022, bioenergy constituted 38.5% of all renewable energy in the United States, with the majority of energy from biomass produced in the industrial and transportation sectors, mostly in the form of industrial heat and power and liquid transportation fuels (US Energy Information Administration 2023). The bioproducts industry is also expanding, especially in the chemical and materials sectors. Along with crops like corn grown for ethanol production, lignocellulosic feedstocks from forest management, wood waste, and purpose-grown energy crops (e.g. short rotation coppicing willow, hybrid poplar, and switchgrass) which are a critical source of biomass for both bioenergy and bioproducts.

The complexities inherent in forest and biomass supply chains which encompass feedstock cultivation and establishment, harvest, processing, storage, transportation, and conversion (Figure 1), directly impact efficiency and cost due to factors such as feedstock diversity, geographical variability, and seasonal dependence. Managing these intricacies is crucial for sustaining and developing forest and energy crop resources in the long term, enhancing the environmental and economic performance of associated bioproducts. Mathematical programming and heuristic algorithms have traditionally been employed for forest and biomass supply chain management in tactical and practical approaches. For example, Han et al. (2018) combined forest operations research (Han et al. 2018) with a mixed-integer linear programming (MILP) model and a network algorithm to optimize forest biomass feedstock processing and transportation for a state forest in Colorado, USA. While Acuna et al. (2019) and Zhang et al. (2020a) have highlighted data-driven modeling, analytics methods, and optimization techniques, recent advancements in enhanced remote sensing (RS), geographic information system (GIS), and global positioning system (GPS) data services and tools have paved the way for machine learning (ML) techniques to process and analyze big data. ML provides a scientific basis for decision-making that goes beyond traditional approaches in operations research and supply chain management.

ML is a subset of artificial intelligence (AI) that involves crafting algorithms and statistical models that enable computer systems to learn from data and generalize to unseen data without being explicitly programmed. ML encompasses a range of techniques, including supervised learning (SL), semi-supervised learning (SSL), unsupervised learning (USL), and reinforcement learning (RL), finding applications across diverse fields like natural language processing, image and speech recognition, predictive analytics, and manufacturing and quality control. Several reviews have explored the applications of ML in biomass supply chain management. Rodias et al. (2019) focused on modern technological tools and identified AI methods associated with ML for potential applications in woody biomass supply chain management. He and Turner (2021) discussed Industry 4.0 implications, including ML technologies, for forest supply chains, addressing opportunities and

CONTACT Jingxin Wang jingxin_wang@ncsu.edu Department of Forest Biomaterials, North Carolina State University, Raleigh, NC 27695-7103

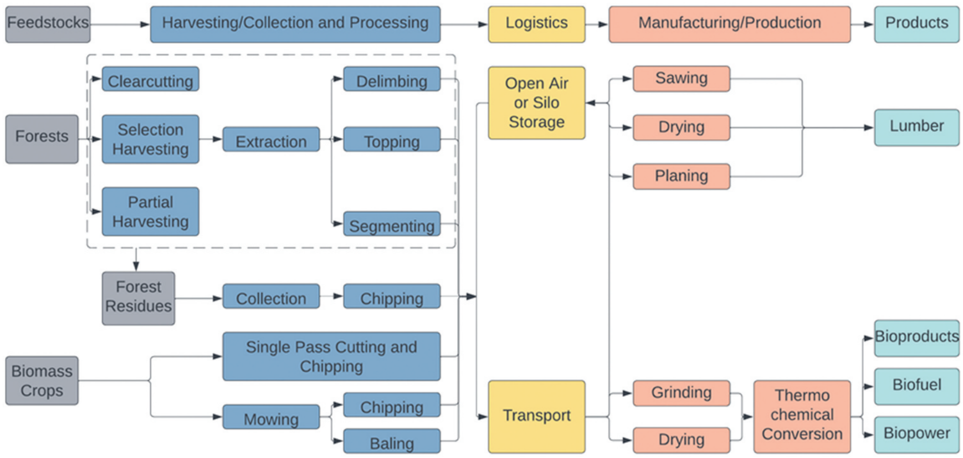

Figure 1. Forest and biomass supply chain and related components.

challenges. Xing et al. (2021) applied ML and AI to optimize biodiesel production processes. Wang and Yao (2023) investigated ML's upstream use in biomass-derived material systems, addressing material, process, and supply chain design challenges. Helal et al. (2023) identified research gaps and opportunities for ML and AI in biomass-to-bioenergy supply chains.

Existing literature reviews on ML applications in integrated forest and biomass supply chains have valuable insights, but their scope is limited. To date, no comprehensive review has systematically evaluated algorithms and approaches specific to these biomass systems. Addressing this knowledge gap is crucial, given the significance of big data, modern digital technologies, and rapid AI advancements, including ML techniques. This review aims to explore and analyze the mechanisms and characteristics of algorithms, data requirements, and future challenges of ML applications throughout the forest and biomass supply chain. We focused on forest biomass and purpose-grown lignocellulosic energy crops like willow, poplar, and switchgrass for several reasons. First, they share similar supply chains when compared to other biomass sources like algae, food waste, sewage sludge, and biofuel crops primarily grown for sugar and starch content. This review thoroughly covers the four stages of a typical forest and energy crop supply chain, including feedstock production, harvesting/collection and processing, transportation and storage, and manufacturing (conversion). Second, their lignocellulosic characteristics result in relatively similar end uses, often used as feedstock in combustion, gasification, and pyrolysis conversion, and sometimes in biochemical conversion to fuels, chemicals, and sustainable biomaterials. Third, they are often cultivated on marginal or unsuitable lands for agricultural production and are easily substitutable and blendable in various conversion pathways. These multi-feedstock biomass supply chains offer flexibility and resiliency in biomass supply while also posing increased complexity, making them well-suited for emerging ML approaches.

Specifically, this review delves into enhancing ML applications in forest and biomass management through two main aspects: (1) exploring ML types, algorithms, and processes, and (2) examining ML applications in forest and biomass estimation, harvest and processing, and technology selection and optimization for converting biomass to bioproducts. The objective is to provide a road map for utilizing ML tools effectively in supply chain management, guiding the selection of appropriate approaches and tools for various biomass sources to meet diverse supply chain requirements.

Overview of machine learning

Types of machine learning

Depending on the expected outcome of the algorithm, ML types include SL, SSL, USL, and RL. Figure 2 shows the four types of ML workflows. SL deals with labeled data, in which each input is associated with a corresponding output, aiming to learn a mapping between inputs and outputs (Ayodele 2010). USL focuses on finding patterns or structures in unlabeled data. SSL is a combination of both SL and USL, the algorithm is trained on a dataset that contains both labeled and unlabeled examples. RL involves an agent learning by interacting with an environment and receiving feedback in the form of rewards or penalties (Ayodele 2010). Given that the majority of real-world data problems in forest and biomass supply chain management involve regression or classification tasks, SL emerges as more prevalent. However, USL finds applications in scenarios requiring clustering or dimensionality reduction.

Machine learning algorithms

ML relies on algorithms for data problem-solving, without a one-size-fits-all solution. The selection of the algorithm depends on factors such as the problem's nature, the number of variables, and the best-suited model. Commonly used algorithms in ML include linear regression (LR), decision tree (DT), logistic regression (LogR), random forest (RF), support vector machine (SVM), clustering methods like k-means clustering (KMC), and neural network (NN).

LR is a fundamental SL algorithm used for predicting a continuous outcome based on one or more input features. The primary goal of LR is to find the best-fitting linear relationship (line) that minimizes the difference between the predicted values and the actual values in the training dataset

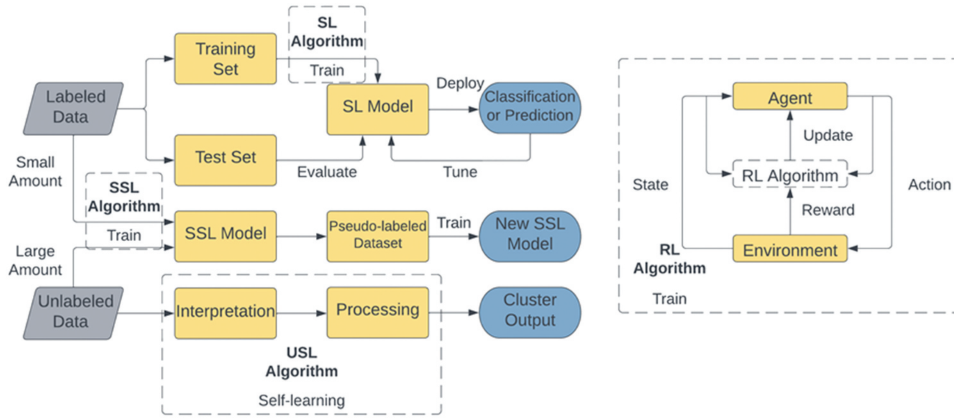

Figure 2. Comparison of four machine learning algorithm types and workflows.

(Maulud and Abdulazeez 2020). This relationship is expressed by a linear equation: $Y = \beta_0 + \beta_1 X_1 + \cdots + \beta_n X_n + \varepsilon$, where Y is the dependent variable, $X_1, \cdots, X_n$ are independent variables, β_0 is the intercept, $\beta_1, \cdots \beta_n$ are the coefficients, and ε is the error. LogR uses the logistic function (sigmoid function) for binary classification, where the outcome variable is categorical and has two classes. The logistic function is defined as: $P(Y = 1) = \frac{1}{1+e^{-(\beta_0+\beta_1 X_1+\cdots+\beta_n X_n)}}$, where $p(Y = 1)$ is the possibility of the dependent variable being in class 1.

SVM is used for both classification and regression tasks, demonstrating effectiveness in smaller datasets with high dimensionality. SVM also aims to find a hyperplane that best separates the data into two classes. SVM can perform nonlinear classification through kernel tricks, enabling the capture of complex relationships in data (Noble 2006). The optimal hyperplane is the one that maximizes the margin, defined as the distance between the hyperplane and the nearest data points from each class, known as support vectors. The prediction accuracy of SVM is highly dependent on the appropriate configuration of hyperparameters C, ε, and kernel parameters (Cherkassky and Ma 2004). The hyperparameter C controls the trade-off between model complexity and the degree to which deviations larger than ε are tolerated in the optimization formulation. The parameter ε, on the other hand, impacts the number of support vectors used to construct the regression function.

DT is a tree-structured recursive partition of the dataset that comprises internal nodes for feature judgments, branch nodes for decision outcomes, and leaf nodes for classification results (Singh et al. 2016). Ensembles of DTs, such as RF, are often used to enhance predictive performance. The individual DTs are trained using in-bag samples drawn with replacements from the training set. Meanwhile, the remaining out-of-bag samples are used for cross-validation, with the out-of-bag error recorded for each training datum. RF then combines the predictions of the individual DTs by averaging, and the final classification result is determined by the plurality of the output classes of the DTs. The modeling process of RF requires the specification of two key parameters: the number of DTs (*ntree*) and the number of variables (*mtry*). Previous research highlighted that *mtry* significantly impacts classification accuracy more than *ntree* (Breiman 2001; Belgiu and Drăguț 2016).

KMC is used for partitioning data into clusters based on similarity. The KMC algorithm randomly selects k points from the dataset as the initial cluster centers, then calculates the Euclidean distance from a sample point to each cluster center and assigns the sample point to the subset corresponding to the closest cluster center. The mean value of all samples in each cluster is then calculated and used as the new cluster center. This process is iteratively repeated until either the maximum number of iterations is reached, or the change in the position of the cluster center is less than a predefined threshold (Likas et al. 2003).

NNs represent computational models inspired by the structure and functioning of the human brain, including artificial neural networks (ANN), recurrent neural networks (RNN), and convolutional neural networks (CNN). Composed of interconnected nodes known as neurons, organized into layers, an NN processes input data through weighted connections, incorporating activation functions at each neuron to introduce non-linearity (Choi et al. 2020). During the training process, the NN improves its performance by adjusting weights through backpropagation and optimization algorithms, minimizing the difference between predicted and actual outputs (Choi et al. 2020). NNs with multiple hidden layers are termed deep neural networks, and the broader field of deep learning (DL) involves training these networks to automatically derive hierarchical representations of data, thereby enabling them to grasp intricate patterns and features more effectively.

Basic process of machine learning

The fundamental procedure of ML involves four main steps: data collection and collation, model training, evaluation, and optimization, as illustrated in Figure 3. With the increasing use of big data, which is characterized by enormous volume and dimensionality, preprocessing is necessary to extract relevant information. This preprocessing includes data visualization, cleaning, and feature selection. Data visualization enables the observation of data structures and patterns. Data cleaning is helpful to improve data quality by identifying and eliminating errors and inconsistencies. Feature selection aims to identify the optimal subset of relevant features while eliminating irrelevant or redundant ones. The modeling process should be

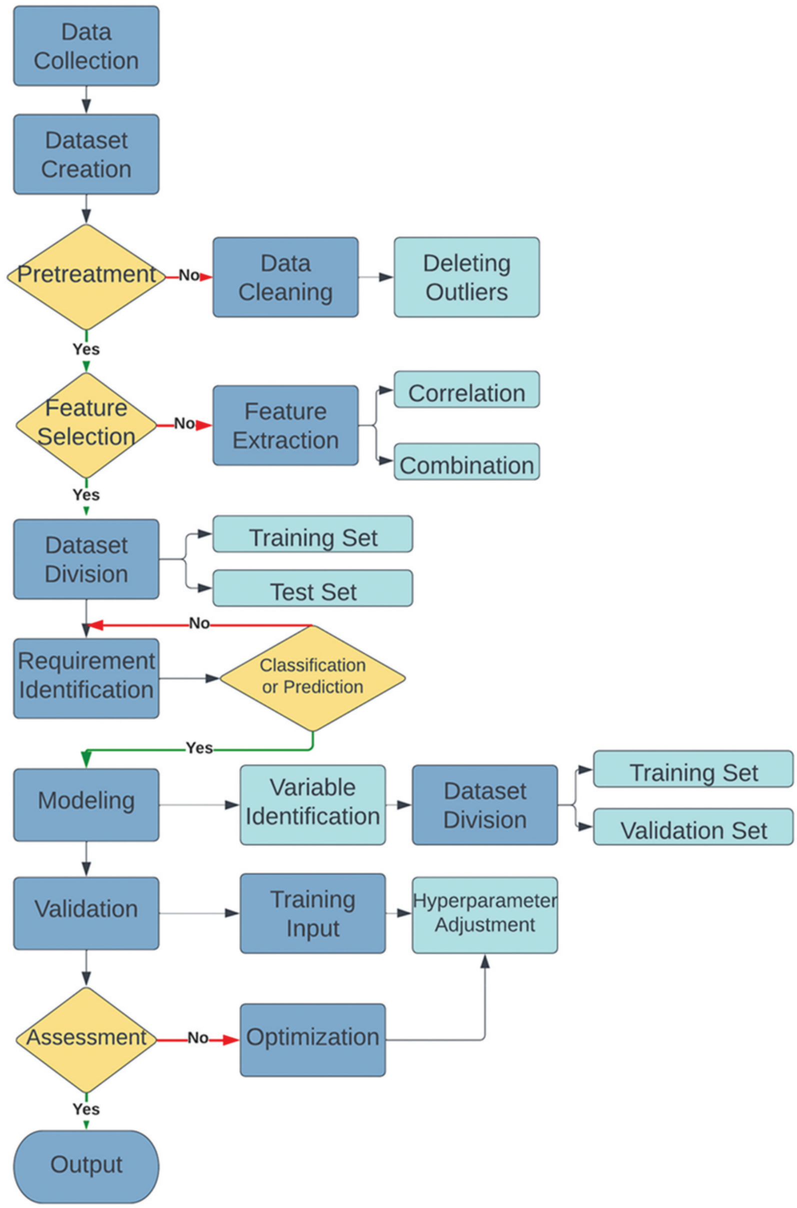

Figure 3. The basic process of machine learning.

goal-oriented regarding the problem to be solved while considering the data characteristics such as distribution and balance. The dataset is initially divided into a training set and a test set, and the training set is further divided into a training subset and a validation set to validate the model's configuration and training. The model's accuracy is evaluated using the test set, which cannot be used for parameter tuning or model selection. ML algorithms have several settings, including model parameters updated through data learning and hyperparameters that regulate the algorithm's performance by balancing bias and variance.

Machine learning in forest and biomass supply chain processes

A substantial virtual dataset for the forest and biomass supply chain is formed by aggregating extensive data generated across processes, categorized into input data (resources, energy, materials, and labor consumption), output data (products and emissions), and socio-economic data (costs and prices) based on the relationship among chain components (Zhang et al. 2020a). ML is essential for extracting valuable insights from these raw data, contributing to improved management efficiency and cost-effectiveness by enhancing knowledge of the system, processes, technologies, and management frameworks. Table 1 is a comprehensive overview of the prevailing ML applications within the specific domains of forest inventory, biomass estimation, harvesting and processing, storage and transportation, and bioproduct conversion. Input data, modeling algorithms, and inherent challenges are listed for each of the application case scenarios.

Forest inventory

Forest inventory encompasses the systematic measurement and assessment of forest resources and forms the foundation

Table 1. Summary of machine learning applications in forest and biomass supply chain management.

Classification/Prediction Tasks	Input Data	Recommended Models	Challenges
Volume/volume increment	Inventory, ALOS and Sentinel data	ANN and SVM	Spatial autocorrelation, hidden layers and transfer functions
Canopy height	Landsat imagery, SAR and LiDAR data	RF and SVM	High data dependency and model transfer
Forest attributes	QuickBird imagery, field inventory and ASTER data	RF and SVR	High-resolution image unavailability
Forest residues	UAV imagery, ALS and TLS data	CNN and RF	Occlusion issue
AGB	LiDAR data, satellite and UAV imagery	ANN, RF, SVR and Tree-based models	Data unavailability
Harvesting area	Inventory data	Clustering and Monte Carlo tree	Multivariate consideration
Machine productivity/ performance	Inventory, environment and operation data	Clustering, Ensemble, GBM, RF and SVM models	Data availability and quality
Stand susceptibility	Inventory data	BRT	Model generalizability
Biomass supply	Historical data	DT and SVM	Methodological decision-making approaches
Facility candidate sites	GIS data	DT and LR	Large-scale network problems
Vehicle routing	GIS data	GA and hybrid models	Cross-cutting sustainability
Uncertainty parameters	Biomass properties and distance matrix	SVM	Cross-cutting sustainability
Biochar yield/carbon content	Proximate, ultimate, structural analysis of biomass and pyrolysis/torrefaction process conditions	ANN, GTB, NN, RF, XGBoost	Limited existing dataset
Bio-oil yield/hydrogen content/characteristics	Proximate, ultimate, structural analysis of biomass and pyrolysis/torrefaction process conditions	NN, RF	Limited existing dataset
HHV	Experimental elemental compositions and HHVs	SVM hybrid models	Limited existing dataset
H_2 concentration	Experimental parameters	DT, KNN and SVR	Limited existing dataset
Heavy metal sorption/ immobilization efficiency	Biochar characteristics, metal sources and environmental conditions	NN, RF, SVM and SVM-ANN ensemble models	Limited existing dataset

Note: ALOS: advanced land observing satellite; SAR: synthetic aperture radar; LiDAR: light detection and ranging; ASTER: advanced spaceborne thermal emission and reflection radiometer; UAV: unmanned aerial vehicle; ALS: airborne laser scanning; TLS: terrestrial laser scanner; GBM: gradient boosted machine; BRT: boosted regression tree; GTB: gradient tree boosting; NN: neural network; XGBoost: extreme gradient boosting; HHV: higher heating value.

of forest planning and management. Forest attributes function as indicators to assess ecological functions and ascertain forest stability and resilience in response to biotic or abiotic stresses (Zhao et al. 2019). Additionally, attributes related to timber volume, growth, and stock can be used to estimate biomass and indirectly quantify carbon sequestration and carbon dioxide (CO_2) cycling. However, traditional forest inventories consume a lot of time, money, and labor and often fail to provide detailed, landscape-scale information over large areas because they tend to be confined to single ownerships or land uses, such as private timberlands. Forest resource data in the United States is available from the U.S. Forest Service (USFS) and the U.S. Geological Survey (USGS). The USFS's Forest Inventory and Analysis (FIA) program provides insights into forest area, tree species, size, and health, as well as overall tree growth, mortality, and harvest removals. It also covers wood production, utilization rates for different products, and forest land ownership. The USGS manages high-resolution satellite imagery, such as the National Land Cover Database (NLCD), useful for monitoring land cover changes, forest health, and biomass estimation. ML techniques find utility in this domain by leveraging data such as RS imagery, satellite data, and aerial surveys to facilitate accurate and efficient estimation of forest characteristics.

Support vector regression (SVR) with multi-sensor satellite data paired with field measurements can predict stand volume more effectively than other tested algorithms, and choosing the appropriate kernel function or creating hybrid models can help improve the model's accuracy (de Souza et al. 2019; Chen et al. 2020). Integration of topographic indices and texture features from different bands and vegetation indices from optical sensors contributes to explaining spatial variations of stand volume (de Souza et al. 2019; Chen et al. 2020). Several theories have also been proposed for modeling forest canopy height to determine long-term stand changes and trends. SVR models using radar data as input can accurately estimate canopy heights at high resolution (Pourshamsi et al. 2021). When combined with satellite imagery, RF and DL models further enhance the predictive capabilities, extending to the region level at resolutions of 30–250 meters (Staben et al. 2018; Li et al. 2020).

ML can also predict forest properties based only on historical inventories and topographic data. Both ANN and SVM enable accurate estimation of incremental volume using field-measured terrain slope and aspect as input variables (Hamidi et al. 2021). In another example, using field plot data, aerial photography, and Landsat imagery, Hogland et al. (2019) combined softmax neural networks with LR models to estimate key forest characteristics across 2.3 million hectares in Georgia, USA. RF has been found to have great advantages in predicting stand information, and it has higher prediction accuracy and lower error for stand age, diameter at breast height, tree height, stand density, and canopy-related metrics (Shataee et al. 2012; Zhao et al. 2019). Various approaches have been reported for quantifying logging residues with the potential for predicting forest biomass availability. The CNN, the refined ground algorithm, and RF have been utilized to detect, segment, and count fine and coarse woody debris from different data sources, including drone images, airborne laser scanning returns, and terrestrial laser scanner data (Windrim et al. 2019; Shokirov et al. 2021).

Biomass estimation

Aboveground biomass (AGB) is one of the key measurements for carbon budget accounting, carbon flux monitoring, and climate change studies (Dang et al. 2019). The rapid and accurate estimation of AGB is critical for assessing biomass availability and performing management practices. However, the accuracy of empirical models is often unsatisfactory due to the influence of environmental factors and small sample sizes, as in allometric equations used in biomass estimation for specific tree species and locations (Chung et al. 2017). As a result, ML methods have been increasingly adopted by utilizing input data derived from sources such as LiDAR, hyperspectral imagery, and field measurements. The US DOE Oak Ridge National Laboratory (ORNL) and National Renewable Energy Laboratory (NREL) provide biomass data online, including information on the quantity, distribution, and diversity of biomass resources. This data is critical for informed decision-making in biomass energy development, covering agricultural and forestry products, animal wastes, and diverse residues, with potential applications in power, heat, transportation fuels, and chemical production. A large number of existing studies have confirmed that RF, SVR, ANN, and a series of DT-based algorithms in corporations with RS data have a great potential in accurately estimating AGB in various applications.

Biomass predictive modeling typically considers multidimensional variables, including stand or vegetation characteristics, topographic features, climate conditions, and management and disturbance history. The present studies reveal that the RF outperformed the stochastic gradient boosting algorithm in accurately predicting the spatial distribution of AGB (Ghosh and Behera 2018; Dang et al. 2019). Furthermore, the RF possesses the ability to further variable selection, thus enabling the estimation of AGB with a minimal set of variables. SVR combined with airborne LiDAR point clouds has been evaluated for estimating AGB. Although the performance of the SVR model is slightly inferior to that of the ordinary least square model, its estimations are credible (Rex et al. 2020). However, the stability of the model fluctuates depending on the division of the dataset (Santi et al. 2020).

ANNs have been applied in AGB estimation due to their unique information processing and computational capabilities compared to empirical statistical models. ANN can also be used in combination with LR models (Hogland et al. 2019). ANN shows substantial advantages for handling high-dimensional nonlinear data (Rex et al. 2020; Santi et al. 2020). Although ANNs are black-box models with unclear mechanisms and not easy to generalize guiding procedures, they have great potential for large-scale AGB estimation. Other ML algorithms, such as cubist regression, stochastic gradient boosting, extreme gradient boosting, gradient boosting trees, extremely randomized trees, and light gradient boosting machines, have also been applied in forest AGB estimation (Gleason and Im 2012; Li et al. 2019; Rex et al. 2020; Zhang et al. 2020b). ML has also been applied to estimate the biomass of dedicated energy crops, including perennial grasses such as switchgrass and miscanthus and fast-growing shrubs such as hybrid willow. The estimation is generally based on vegetation indices derived from unmanned aerial vehicle (UAV) images or point clouds. RF models have been used to accurately estimate the biomass of shrub willow and miscanthus, though CNNs achieve higher accuracy in estimating shrub willow biomass (Tamiminia et al. 2021; Impollonia et al. 2022).

Harvesting and processing

The dynamic and decentralized nature of forest biomass supply often results in high logistics costs (Shadbahr et al. 2021), which typically include felling, hauling, and processing in the forest and processing, loading, and transportation between the forest and the facility. Factors such as equipment type, job layout, equipment movement, fuel consumption, and operator skill can determine environmental impacts, especially with regard to negative soil impacts, residual stand damage, and GHG emissions. To mitigate these impacts, ML models can be used to optimize logistics and equipment required for strategic and operational planning of biomass harvesting, and to help operators minimize negative environmental outcomes. Strategic planning of feedstock harvesting aims to determine the approximate size and boundaries of the harvest area and the amount to be harvested. Optimized harvesting strategies can help extract feedstocks that maximize benefits to stakeholders and ensure sufficient regeneration and restoration to meet sustainability criteria.

An ensemble coverage model used a spectral clustering algorithm to identify felling area clusters with minimal spatial dispersion and roughly equal timber volume among clusters (Mobtaker et al. 2018). The Monte Carlo simulation was used to eliminate the multi-objective constraint problem of the clearcut area, total habitat area, and core area inside habitats, thus promoting biological conservation and avoiding forest fragmentation caused by small clearcuts (Neto et al. 2020). Besides, ML models can help minimize the damage caused to residual stands during harvesting operations, especially during skidding operations. For example, the boosted regression tree was applied to predict stand susceptibility to damage, outperforming LogR and RF (Shabani et al. 2020, 2021). Generated stand damage susceptibility maps have practical implications for harvest planning and demonstrate the potential of ML methods in forest operations and engineering.

ML techniques have also been used to estimate the productivity of harvesting machines by involving sensor data from logging equipment, video feeds, and operational records. Lopes et al. (2022) applied ML to estimate the productivity of an excavator-based grapple saw based on its harvesting condition and operation data records. They indicated a combination of genetic algorithm (GA)-RF feature selection with RF prediction can yield better accuracy and robustness of the productivity estimation. The gradient boosted machine and SVM showed promise in predicting mechanized cut-to-length productivity while considering variables related to the environment, machinery, and operator (Liski et al. 2020). Munis et al. (2022) tested 24 algorithms to estimate the productivity of harvesters and found that combining ensembles achieved acceptable accuracy but was not a significant improvement over a single algorithm.

Melander et al. (2020) employed the clustering method to fuse open forest data and machine fieldbus data (e.g. performance data from actuators, sensors, and controllers), helping to explain the relationship between machinery performance and forest attributes.

Storage and transportation

The logistical components of forest and biomass supply chains are multi-feedstock flows to storage, preprocessing, and conversion facilities during the post-harvest phase. However, sustainable, robust, and high-quality feedstock supply faces challenges related to cost- and time-sensitive storage and transportation issues. Biomass resources undergo a growth cycle, from planting to cultivation to harvest, and some exhibit seasonal characteristics (Rentizelas et al. 2009). While the supply of feedstocks is seasonal, the demand from facilities remains consistent throughout the year. Selecting the least costly storage method while preventing biomass degradation is always challenging (Yue et al. 2014). Compared with fossil energy, biomass resources are spatially dispersed, and their large volume and low energy density create significant barriers to storage and transportation. Consequently, well-planned logistics require extraordinary efforts to integrate supply chain networks, facility locations, vehicle routing, and the capacity of depots and vehicles.

Various mathematical programming models have been developed for forest and biomass storage and transportation optimization. Linear programming improves woody biomass logistics (Lin et al. 2016), while MILP optimizes the bio-based supply chain and forest biomass supply network (Mirkouei and Haapala 2015; Guo et al. 2020). Simulation methods like discrete event simulation analyze inventory management impacts on demand fulfillment, costs, and emissions in forest biomass supply chains (Akhtari et al. 2019). GA effectively tackles vehicle routing problems in biomass transportation (Gracia et al. 2014) and combining GA and particle swarm optimization in discrete event simulation enhances vehicle scheduling efficiency (Pinho et al. 2018).

However, mathematical methods often struggle to find optimal solutions for non-convex or large-scale problems (Castillo-Villar 2014), such as supply chains that involve multiple feedstocks flowing through a large network of suppliers and intermediate storage facilities (Helal et al. 2023). Since ML focuses on classification and regression learning, its primary function in solving optimization problems is reflected in variable classification and parameter prediction to simplify calculations. For instance, ML methods have been applied to select biomass feedstock suppliers using DT to define feedstock types and SVM to evaluate suppliers (Mirkouei and Haapala 2014). Similarly, potential biomass harvesting sites were classified using DT analysis, while SVMs were used to determine patterns of uncertainties (Mirkouei et al. 2016, 2017). ML methods were also employed to select potential depot locations for the supply chain of biomass co-firing. Goettsch et al. (2020) incorporated LogR, DT, RF, and multilayer perceptron neural networks into a MILP model to reduce costs, with both LogR and DT achieving better results.

Bioproduct conversion

Thermal conversion processes, including combustion, pyrolysis, torrefaction, hydrothermal liquefaction, and gasification, are widely used to produce value-added solid, liquid, and gaseous bioproducts (Naqvi et al. 2023). However, the reaction process's complexity and nonlinear parameters necessitate the modeling of thermal conversion processes to determine optimal operating conditions and assess the effects of parameters on the output. To this end, mathematical and kinetic models have been constructed and used to analyze process behavior and evaluate the effects of the various parameters. These models mainly rely on transport phenomena and kinetic equations, which require extensive experimental efforts, consuming time and resources. ML has been increasingly utilized to control and optimize thermal conversion processes and predict bioproduct properties.

Co-pyrolysis of biomass with other feedstocks has demonstrated the potential to enhance the yield of liquid fuels or upgrade bio-oil. However, conventional pyrolysis kinetic models do not account for the synergistic effects of multiple feedstocks. Recent studies have utilized RF modeling to accurately predict coal and biomass co-pyrolysis (Wei et al. 2022). Extreme gradient boosting and neural network models have also been employed to predict biochar and bio-oil yield in the co-pyrolysis of plastics and biomass (Alabdrabalnabi et al. 2022). As torrefaction can significantly alter the calorific value of biomass, hybrid ML models were proposed to detect higher heating values, as highlighted by García Nieto et al. (2019) and Nieto et al. (2019). In hydrothermal liquefaction, water is typically a solvent, and process parameters (temperature, pressure, and time) are set based on the desired products (Marzbali et al. 2021). ML techniques have been combined with experiments and molecular dynamics simulations to estimate the impact of solvents on the reaction rate of liquid-phase acid-catalyzed biomass conversion (Walker et al. 2020). Furthermore, a decision support system that leverages ML algorithms has been developed to predict optimal hydrothermal liquefaction process parameters (Gopirajan et al. 2021).

On the other hand, the thermal conversion of biomass can result in a wide range of energy, fuel, chemical, and material co-products. This highlights the need for further exploration of ML methods applied to specific supply chains and products. For example, ML can predict biochar yields based on biomass composition/characteristics and pyrolysis conditions. RF, gradient tree boosting, and multilayer perceptron neural network models have high accuracy and provide insights for system optimization and environmental footprint assessment of biochar production (Zhu et al. 2019; Onsree and Tippayawong 2021; Li et al. 2022). The ANNs incorporating metaheuristic algorithms also simplified the prediction of biochar yield (Khan et al. 2022). Additionally, ML can predict metal sorption efficiency based on biochar properties and environmental conditions, opening up a direction for soil and water pollution treatment (Zhu et al. 2019; Ke et al. 2021; Palansooriya et al. 2022). The feasibility of RF, ML regression, and a fuzzy-based algorithm was demonstrated for bio-oil yield prediction (Tang et al. 2020; Ge et al. 2021; Sujith et al. 2022; Zhang et al. 2022). RF was also successfully applied to the analyses of hydrogen content, viscosity, calorific value, oxygen-carbon ratio, and hydrogen-carbon ratio of bio-oil (Tang et al. 2020; Zhang et al. 2022). Regression models based on KNN, SVM, and DT were studied to predict H_2 concentration during gasification using process parameters as inputs (Ozbas et al. 2019).

Conclusions

A rich body of research shows that ML has emerged as a promising technology in supporting forest and biomass supply chain decision-making at all levels, from strategic to operational and tactical. It can be used to improve supply chain efficiency and environmental and economic sustainability. However, data availability and quality remain pressing challenges to implementing an ML model and application. ML has the potential to revolutionize forest or stand attribute estimation by rapidly training big data derived from RS and resource inventories and applying scalable predictive models to other similar ecosystems or regions. Accurate predictions of forest and resource attributes can facilitate the development of forest or biomass management regimes and policies while reducing environmental impacts. Nonetheless, the effectiveness of ML in generating accurate forest inventories depends on the availability of large amounts of high-quality data, which may be limited in some cases. This is especially true of field-based data, which is often needed for model training and validation. Moreover, combining ML and RS technologies poses technical challenges, although it can enable large-scale biomass estimation in areas where traditional methods are not feasible. Real-time biomass monitoring AI systems can be developed to track spatial and temporal changes in biomass. ML has undoubtedly improved the scale and efficacy of biomass estimation, but acquiring and processing the necessary datasets, including satellite imagery, field measurements, and environmental data, can be both time-consuming and costly.

In contrast to other stages of the supply chain, there have been relatively few ML applications in harvesting and processing due to the dynamic, highly variable, and complex nature of forest biomass feedstocks, which are influenced by various environmental factors. However, ML can still provide benefits in supply chain efficiency, sustainability, and quality control. For example, it can optimize transportation routes and storage locations by considering supply, demand, and procurement distance. It can also be used to predict and monitor changes in feedstock properties over time, including mass loss in storage. But developing ML models that accurately capture multi-objective value chains, including reducing costs and environmental impacts, meeting demand, maintaining quality control, and stimulating regional biomass economies can be challenging. Across diverse conversion pathways, ML has been applied to predict and optimize process parameters in biomass conversion to a wide range of bioproducts. Still, limited data linking feedstock composition, processes, and products make it challenging to train ML models. The opportunities of ML for sustainable development and management of forest and energy crop biomass are significant. Solving big data challenges will be crucial to ensuring the accuracy and reliability of ML models and applications, and fully realizing ML's exciting potential.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

The work was supported by the National Institute of Food and Agriculture [20206801231881].

ORCID

Jinghan Zhao http://orcid.org/0009-0006-6838-8808

References

Acuna M, Sessions J, Zamora R, Boston K, Brown M, Ghaffariyan MR. 2019. Methods to manage and optimize forest biomass supply chains: a review. Curr For Rep. 5(3):124–141. doi: 10.1007/s40725-019-00093-4 .

Akhtari S, Sowlati T, Siller-Benitez DG, Roeser D. 2019. Impact of inventory management on demand fulfilment, cost and emission of forest-based biomass supply chains using simulation modelling. Biosyst Eng. 178:184–199. doi: 10.1016/j.biosystemseng.2018.11.015.

Alabdrabalnabi A, Gautam R, Sarathy SM. 2022. Machine learning to predict biochar and bio-oil yields from co-pyrolysis of biomass and plastics. Fuel. 328:125303. doi: 10.1016/j.fuel.2022.125303.

Ayodele TO. 2010. Types of machine learning algorithms. New Adv Mach Learn. 3:19–48.

Belgiu M, Drăguţ L. 2016. Random forest in remote sensing: a review of applications and future directions. ISPRS J Photogrammetry Remote Sens. 114:24–31. doi: 10.1016/j.isprsjprs.2016.01.011.

Breiman L. 2001. Random forests. Mach Learn. 45(1):5–32. doi: 10.1023/A:1010933404324.

Castillo-Villar KK. 2014. Metaheuristic algorithms applied to bioenergy supply chain problems: theory, review, challenges, and future. Energies. 7(11):7640–7672. doi: 10.3390/en7117640.

Chen L, Ren C, Zhang B, Wang Z. 2020. Multi-sensor prediction of stand volume by a hybrid model of support vector machine for regression kriging. Forests. 11(3):296. doi: 10.3390/f11030296.

Cherkassky V, Ma Y. 2004. Practical selection of SVM parameters and noise estimation for SVM regression. Neural Networks. 17(1):113–126. doi: 10.1016/S0893-6080(03)00169-2.

Choi RY, Coyner AS, Kalpathy-Cramer J, Chiang MF, Campbell JP. 2020. Introduction to machine learning, neural networks, and deep learning. Transl Vision Sci Technol. 9(2):14–14.

Chung W, Evangelista P, Anderson N, Vorster A, Han H, Poudel K, Sturtevant R. 2017. Estimating aboveground tree biomass for beetle-killed lodgepole pine in the rocky mountains of northern Colorado. For Sci. 63(4):413–419. doi: 10.5849/FS.2016-065.

Dang ATN, Nandy S, Srinet R, Luong NV, Ghosh S, Kumar AS. 2019. Forest aboveground biomass estimation using machine learning regression algorithm in Yok Don National Park, Vietnam. Ecol Inf. 50:24–32. doi: 10.1016/j.ecoinf.2018.12.010.

de Souza GSA, Soares VP, Leite HG, Gleriani JM, Do Amaral CH, Ferraz AS, de Freitas Silveira MV, dos Santos JFC, Velloso SGS, Domingues GF, et al. 2019. Multi-sensor prediction of eucalyptus stand volume: a support vector approach. ISPRS J Photogrammetry Remote Sens. 156:135–146. doi: 10.1016/j.isprsjprs.2019.08.002.

García Nieto PJ, Garcia-Gonzalo E, Paredes-Sánchez JP, Bernardo Sánchez A, Menéndez Fernández M. 2019. Predictive modelling of the higher heating value in biomass torrefaction for the energy treatment process using machine-learning techniques. Neural Comput Applic. 31(12):8823–8836. doi: 10.1007/s00521-018-3870-x.

Ge S, Shi Y, Xia C, Huang Z, Manzo M, Cai L, Ma H, Zhang S, Jiang J, Sonne C, et al. 2021. Progress in pyrolysis conversion of waste into value-added liquid pyro-oil, with focus on heating source and machine learning analysis. Energy Convers Manag. 245:114638. doi: 10.1016/j.enconman.2021.114638.

Ghosh SM, Behera MD. 2018. Aboveground biomass estimation using multi-sensor data synergy and machine learning algorithms in a dense tropical forest. Appl Geogr. 96:29–40. doi: 10.1016/j.apgeog.2018.05.011.

Gleason CJ, Im J. 2012. Forest biomass estimation from airborne LiDAR data using machine learning approaches. Remote Sens Environ. 125:80–91. doi: 10.1016/j.rse.2012.07.006.

Goettsch D, Castillo-Villar KK, Aranguren M. 2020. Machine-learning methods to select potential depot locations for the supply chain of biomass co-firing. Energies. 13(24):6554. doi: 10.3390/en13246554.

Gopirajan PV, Gopinath KP, Sivaranjani G, Arun J. 2021. Optimization of hydrothermal liquefaction process through machine learning approach: process conditions and oil yield. Biomass Convers Biorefin. 306:1–10. doi: 10.1016/j.jclepro.2021.127302.

Gracia C, Velázquez-Martí B, Estornell J. 2014. An application of the vehicle routing problem to biomass transportation. Biosyst Eng. 124:40–52. doi: 10.1016/j.biosystemseng.2014.06.009.

Guo X, Voogt J, Annevelink B, Snels J, Kanellopoulos A. 2020. Optimizing resource utilization in biomass supply chains by creating integrated biomass logistics centers. Energies. 13(22):6153. doi: 10.3390/en13226153.

Hamidi SK, Zenner EK, Bayat M, Fallah A. 2021. Analysis of plot-level volume increment models developed from machine learning methods applied to an uneven-aged mixed forest. Ann For Sci. 78(1):1–16. doi: 10.1007/s13595-020-01011-6.

Han H, Chung W, She J, Anderson N, Wells L. 2018. Productivity and costs of two beetle-kill salvage harvesting methods in Northern Colorado. Forests. 9(9):572. doi: 10.3390/f9090572.

Han H, Chung W, Wells L, Anderson N. 2018. Optimizing biomass feedstock logistics for forest residue processing and transportation on a tree-shaped road network. Forests. 9(3):121. doi: 10.3390/f9030121.

He Z, Turner P. 2021. A systematic review on technologies and industry 4.0 in the forest supply chain: a framework identifying challenges and opportunities. Logistics. 5(4):88. doi: 10.3390/logistics5040088.

Helal MA, Anderson N, Wei Y, Thompson M. 2023. A review of biomass-to-bioenergy supply chain research using bibliometric analysis and visualization. Energies. 16(3):1187. doi: 10.3390/en16031187.

Hogland J, Anderson N, Affleck DL, St. Peter J. 2019. Using forest inventory data with Landsat 8 imagery to map longleaf pine forest characteristics in Georgia, USA. Remote Sens. 11(15):1803. doi: 10.3390/rs11151803.

IEA. 2022. Bioenergy. Paris, IEA. https://www.iea.org/reports/bioenergy. License: CC BY 4.0.

Impollonia G, Croci M, Ferrarini A, Brook J, Martani E, Blandinières H, Marcone A, Awty-Carroll D, Ashman C, Kam J, et al. 2022. UAV remote sensing for high-throughput phenotyping and for yield prediction of miscanthus by machine learning techniques. Remote Sens. 14(12):2927. doi: 10.3390/rs14122927.

Ke B, Nguyen H, Bui XN, Bui HB, Choi Y, Zhou J, Moayedi H, Costache R, Nguyen-Trang T. 2021. Predicting the sorption efficiency of heavy metal based on the biochar characteristics, metal sources, and environmental conditions using various novel hybrid machine learning models. Chemosphere. 276:130204. doi: 10.1016/j.chemosphere.2021.130204.

Khan M, Ullah Z, Mašek O, Naqvi SR, Khan MNA. 2022. Artificial neural networks for the prediction of biochar yield: a comparative study of metaheuristic algorithms. Bioresource Technol. 355:127215. doi: 10.1016/j.biortech.2022.127215.

Li W, Niu Z, Shang R, Qin Y, Wang L, Chen H. 2020. High-resolution mapping of forest canopy height using machine learning by coupling ICESat-2 LiDAR with sentinel-1, sentinel-2 and Landsat-8 data. Int J Appl Earth Observation Geoinf. 92:102163. doi: 10.1016/j.jag.2020.102163.

Li Y, Gupta R, You S. 2022. Machine learning assisted prediction of biochar yield and composition via pyrolysis of biomass. Bioresource Technol. 359:127511. doi: 10.1016/j.biortech.2022.127511.

Li Y, Li C, Li M, Liu Z. 2019. Influence of variable selection and forest type on forest aboveground biomass estimation using machine learning algorithms. Forests. 10(12):1073. doi: 10.3390/f10121073.

Likas A, Vlassis N, Verbeek JJ. 2003. The global k-means clustering algorithm. Pattern Recognit. 36(2):451–461. doi: 10.1016/S0031-3203(02)00060-2.

Lin Y, Pan F, Srivastava A. 2016. A linear programming optimization model of woody biomass logistics integrating infield drying as a cost-saving preprocess in Michigan. For Products J. 66(7–8):391–400. doi: 10.13073/FPJ-D-15-00077.

Liski E, Jounela P, Korpunen H, Sosa A, Lindroos O, Jylhä P. 2020. Modeling the productivity of mechanized CTL harvesting with statistical machine learning methods. Int J For Eng. 31(3):253–262. doi: 10.1080/14942119.2020.1820750.

Lopes ILE, Araújo LA, Miranda EN, Bastos TA, Gomide LR, Castro GP. 2022. A comparative approach of methods to estimate machine productivity in wood cutting. Int J For Eng. 33(1):43–55. doi: 10.1080/14942119.2021.1952520.

Marzbali MH, Kundu S, Halder P, Patel S, Hakeem IG, Paz-Ferreiro J, Madapusi S, Surapaneni A, Shah K. 2021. Wet organic waste treatment via hydrothermal processing: a critical review. Chemosphere. 279:130557. doi: 10.1016/j.chemosphere.2021.130557.

Maulud D, Abdulazeez AM. 2020. A review on linear regression comprehensive in machine learning. J Appl Sci Technol Trends. 1(4):140–147. doi: 10.38094/jastt1457.

Melander L, Einola K, Ritala R. 2020. Fusion of open forest data and machine fieldbus data for performance analysis of forest machines. Eur J For Res. 139(2):213–227. doi: 10.1007/s10342-019-01237-8.

Mirkouei A, Haapala KR. 2014. Integration of machine learning and mathematical programming methods into the biomass feedstock supplier selection process.

Mirkouei A, Haapala KR. 2015 June. A network model to optimize upstream and midstream biomass-to-bioenergy supply chain costs. International Manufacturing Science and Engineering Conference, Charlotte, North Carolina, USA, American Society of Mechanical Engineers (Vol. 56833, p. V002T05A011) American Society of Mechanical Engineers.

Mirkouei A, Haapala KR, Sessions J, Murthy GS. 2016. Evolutionary optimization of bioenergy supply chain cost with uncertain forest biomass quality and availability. Proceedings of the IIE-ISERC, Anaheim, California, USA. p. 21–24.

Mirkouei A, Haapala KR, Sessions J, Murthy GS. 2017. A mixed biomass-based energy supply chain for enhancing economic and environmental sustainability benefits: a multi-criteria decision making framework. Appl Energy. 206:1088–1101. doi: 10.1016/j.apenergy.2017.09.001.

Mobtaker A, Montecinos J, Ouhimmou M, Rönnqvist M, Paquet M. 2018. Minimizing spatial dispersion of forest harvest areas using spectral clustering and set covering modelling. Can J For Res. 48(12):1563–1576. doi: 10.1139/cjfr-2018-0210.

Munis RA, Almeida RO, Camargo DA, da Silva RBG, Wojciechowski J, Simões D. 2022. Machine learning methods to estimate productivity of harvesters: mechanized timber harvesting in Brazil. Forests. 13(7):1068. doi: 10.3390/f13071068.

Naqvi SR, Ullah Z, Taqvi SAA, Khan MNA, Farooq W, Mehran MT, Juchelková D, Štěpanec L. 2023. Applications of machine learning in thermochemical conversion of biomass-A review. Fuel. 332:126055. doi: 10.1016/j.fuel.2022.126055.

Neto T, Constantino M, Martins I, Pedroso JP. 2020. A multi-objective monte carlo tree search for forest harvest scheduling. Eur J Oper Res. 282(3):1115–1126. doi: 10.1016/j.ejor.2019.09.034.

Nieto PG, García–Gonzalo E, Lasheras FS, Paredes–Sánchez JP, Fernández PR. 2019. Forecast of the higher heating value in biomass torrefaction by means of machine learning techniques. J Comput Appl Math. 357:284–301. doi: 10.1016/j.cam.2019.03.009.

Noble WS. 2006. What is a support vector machine? Nat Biotechnol. 24(12):1565–1567. doi: 10.1038/nbt1206-1565.

Onsree T, Tippayawong N. 2021. Machine learning application to predict yields of solid products from biomass torrefaction. Renewable Energy. 167:425–432. doi: 10.1016/j.renene.2020.11.099.

Ozbas EE, Aksu D, Ongen A, Aydin MA, Ozcan HK. 2019. Hydrogen production via biomass gasification, and modeling by supervised machine learning algorithms. Int J Hydrogen Energy. 44 (32):17260–17268. doi: 10.1016/j.ijhydene.2019.02.108.

Palansooriya KN, Li J, Dissanayake PD, Suvarna M, Li L, Yuan X, Sarkar B, Tsang DC, Rinklebe J, Wang X, et al. 2022a. Prediction of soil heavy metal immobilization by biochar using machine learning. Environ Sci Technol. 56(7):4187–4198. doi: 10.1021/acs.est.1c08302.

Pinho TM, Coelho JP, Veiga G, Moreira AP, Boaventura-Cunha J. 2018 June. Soft computing optimization for the biomass supply chain operational planning. 2018 13th APCA International Conference on Automatic Control and Soft Computing (CONTROLO), Ponta Delgada, Portugal, IEEE. p. 259–264.

Pourshamsi M, Xia J, Yokoya N, Garcia M, Lavalle M, Pottier E, Balzter H. 2021. Tropical forest canopy height estimation from combined polarimetric SAR and LiDAR using machine-learning. ISPRS J Photogrammetry Remote Sens. 172:79–94. doi: 10.1016/j.isprsjprs.2020.11.008.

Rentizelas AA, Tolis AJ, Tatsiopoulos IP. 2009. Logistics issues of biomass: the storage problem and the multi-biomass supply chain. Renew Sustain Energy Rev. 13(4):887–894. doi: 10.1016/j.rser.2008.01.003.

Rex FE, Silva CA, Dalla Corte AP, Klauberg C, Mohan M, Cardil A, Silva VSD, Almeida DRAD, Garcia M, Broadbent EN, et al. 2020. Comparison of statistical modelling approaches for estimating tropical forest aboveground biomass stock and reporting their changes in low-intensity logging areas using multi-temporal LiDAR data. Remote Sens. 12(9):1498. doi: 10.3390/rs12091498.

Rodias E, Berruto R, Bochtis D, Sopegno A, Busato P. 2019. Green, yellow, and woody biomass supply-chain management: a review. Energies. 12 (15):3020. doi: 10.3390/en12153020.

Santi E, Paloscia S, Pettinato S, Cuozzo G, Padovano A, Notarnicola C, Albinet C. 2020. Machine-learning applications for the retrieval of forest biomass from airborne P-band SAR data. Remote Sens. 12 (5):804. doi: 10.3390/rs12050804.

Shabani S, Jaafari A, Bettinger P. 2021. Spatial modeling of forest stand susceptibility to logging operations. Environ Impact Assess Rev. 89:106601. doi: 10.1016/j.eiar.2021.106601.

Shabani S, Pourghasemi HR, Blaschke T. 2020. Forest stand susceptibility mapping during harvesting using logistic regression and boosted regression tree machine learning models. Global Ecol Conserv. 22: e00974. doi: 10.1016/j.gecco.2020.e00974.

Shadbahr J, Bensebaa F, Ebadian M. 2021. Impact of forest harvest intensity and transportation distance on biomass delivered costs within sustainable forest management-A case study in southeastern Canada. J Environ Manag. 284:112073. doi: 10.1016/j.jenvman.2021.112073.

Shataee S, Kalbi S, Fallah A, Pelz D. 2012. Forest attribute imputation using machine-learning methods and ASTER data: comparison of k-NN, SVR and random forest regression algorithms. Int J Remote Sens. 33(19):6254–6280. doi: 10.1080/01431161.2012.682661.

Shokirov S, Schaefer M, Levick SR, Jucker T, Borevitz J, Abdurahmanov I, Youngentob K. 2021. Multi-platform LiDAR approach for detecting coarse woody debris in a landscape with varied ground cover. Int J Remote Sens. 42(24):9324–9350. doi: 10.1080/01431161.2021.1995072.

Singh A, Thakur N, Sharma A. 2016 Mar. A review of supervised machine learning algorithms. 2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom), New Delhi, India, IEEE. p. 1310–1315.

Staben G, Lucieer A, Scarth P. 2018. Modelling LiDAR derived tree canopy height from Landsat TM, ETM+ and OLI satellite imagery—A machine learning approach. Int J Appl Earth Obs Geoinf. 73:666–681. doi: 10.1016/j.jag.2018.08.013.

Sujith AVLN, Swathi R, Venkatasubramanian R, Venu N, Hemalatha S, George T, Hemlathadhevi A, Madhu P, Karthick A, Muhibbullah M, et al. 2022. Integrating nanomaterial and high-performance fuzzy-based machine learning approach for green energy conversion. J Nanomater. 2022:1–11. doi: 10.1155/2022/5793978.

Tamiminia H, Salehi B, Mahdianpari M, Beier CM, Klimkowski DJ, Volk TA. 2021. comparison of machine and deep learning methods to estimate shrub willow biomass from UAS imagery. Can J Remote Sens. 47(2):209–227. doi: 10.1080/07038992.2021.1926952.

Tang Q, Chen Y, Yang H, Liu M, Xiao H, Wu Z, Chen H, Naqvi SR. 2020. Prediction of bio-oil yield and hydrogen contents based on machine learning method: effect of biomass compositions and pyrolysis conditions. Energy Fuels. 34(9):11050–11060. doi: 10.1021/acs.energyfuels.0c01893.

US Energy Information Administration. 2023. Monthly energy review. May 2023.

Walker TW, Chew AK, Van Lehn RC, Dumesic JA, Huber GW. 2020. Rational design of mixed solvent systems for acid-catalyzed biomass conversion processes using a combined experimental, molecular dynamics and machine learning approach. Top Catal. 63(7–8):649–663. doi: 10.1007/s11244-020-01260-9.

Wang HSH, Yao Y. 2023. Machine learning for sustainable development and applications of biomass and biomass-derived carbonaceous materials in water and agricultural systems: a review. Resour Conserv Recycl. 190:106847. doi: 10.1016/j.resconrec.2022.106847.

Wei H, Luo K, Xing J, Fan J. 2022. Predicting co-pyrolysis of coal and biomass using machine learning approaches. Fuel. 310:122248. doi: 10.1016/j.fuel.2021.122248.

Windrim L, Bryson M, McLean M, Randle J, Stone C. 2019. Automated mapping of woody debris over harvested forest plantations using UAVs, high-resolution imagery, and machine learning. Remote Sens. 11(6):733. doi: 10.3390/rs11060733.

Xing Y, Zheng Z, Sun Y, Agha Alikhani M, Baghban A. 2021. A review on machine learning application in biodiesel production studies. Int J Chem Eng. 2021:1–12. doi: 10.1155/2021/2154258.

Yue D, You F, Snyder SW. 2014. Biomass-to-bioenergy and biofuel supply chain optimization: overview, key issues and challenges. Comput Chem Eng. 66:36–56. doi: 10.1016/j.compchemeng.2013.11.016.

Zhang T, Cao D, Feng X, Zhu J, Lu X, Mu L, Qian H. 2022. Machine learning prediction of bio-oil characteristics quantitatively relating to biomass compositions and pyrolysis conditions. Fuel. 312:122812. doi: 10.1016/j.fuel.2021.122812.

Zhang X, Wang J, Vance J, Wang Y, Wu J, Hartley D. 2020a. Data analytics for enhancement of forest and biomass supply chain management. Curr Forestry Rep. 6(2):129–142. doi: 10.1007/s40725-020-00111-w.

Zhang Y, Ma J, Liang S, Li X, Li M. 2020b. An evaluation of eight machine learning regression algorithms for forest aboveground biomass estimation from multiple satellite data products. Remote Sens. 12(24):4015. doi: 10.3390/rs12244015.

Zhao Q, Yu S, Zhao F, Tian L, Zhao Z. 2019. Comparison of machine learning algorithms for forest parameter estimations and application for forest quality assessments. For Ecol Manag. 434:224–234. doi: 10.1016/j.foreco.2018.12.019.

Zhu X, Li Y, Wang X. 2019. Machine learning prediction of biochar yield and carbon contents in biochar based on biomass characteristics and pyrolysis conditions. Bioresource Technol. 288:121527. doi: 10.1016/j.biortech.2019.121527.

Zhu X, Wang X, Ok YS. 2019. The application of machine learning methods for prediction of metal sorption onto biochars. J Hazard Mater. 378:120727. doi: 10.1016/j.jhazmat.2019.06.004.