Vaasan yliopisto
UNIVERSITY OF VAASA

HAFIZ AHMED

AI-Enhanced Framework for Evaluating Bioenergy Material Characteristics Linked to Lifecycle Emissions

School of Technology and Innovation
Master of Science in Technology
Degree programme in Industrial System Analytics

Vaasa 2026

UNIVERSITY OF VAASA
School of Technology and Innovation
Author: HAFIZ AHMED
Title of the Thesis: AI-Enhanced Framework for Evaluating Bioenergy Material Characteristics Linked to Lifecycle Emissions
Degree: Teknologian maisteri
Programme: Teollinen järjestelmäanalyysi
Supervisor: Emmanuel Ndzibah
Year: 2026 **Sivumäärä:** 91

ABSTRACT :

Biomass has become a prominent renewable energy source that could play an important role in low carbon energy production, waste valorization and mitigation of greenhouse gases. This potential in Sustainable Development has been increasingly noticed, especially with global energy systems moving away from fossil fuels. But the performance and environmental impact of biomass energy is different for different feedstocks because of the small differences in composition and their nonlinear characteristics. In this work the question is: Is it possible to create a unifying framework for biomass that can be interpreted and used to assess biomass based on energy content and life cycle emissions?

Thematic integration of assessments of bioenergy, life cycle analysis and machine learning interpretability. The research shows a correlation between the characteristics of the feedstocks (carbon, lignin, xylan, moisture, ash) and the predicted energy properties (HHV, LHV, bioenergy potential) and estimated lifecycle GHGs. Core literature suggests that it is important to assess these interactions comprehensively and not separately to gain a deeper insight into biomass sustainability and the selection of feedstock. The importance of using the interpretable machine learning SHAP, permutation importance, and sensitivity analysis methods to identify important variables that affect bioenergy and emissions.

Research methodology adopts a systematic process involving data pre-processing, feature engineering, computing of the empirical baseline (HAL), and model development using machine learning models, namely, Support Vector Machine (SVM), Random Forest (RF), XGBoost and Artificial Neural Network (ANN). It is found that nonlinear learning approaches perform better than traditional empirical predictions with SVM model showing the best results for predicting HHV and LHV. Key variables were identified to influence energy and emissions and char from these feedstocks was ranked as the most preferable in this regard due to their high energy potential and low life cycle emissions. This coordinated framework is a clear and sustainability-oriented tool to direct biomass assessment and facilitate low carbon bio-energy development.

KEYWORDS: Biomass feedstock, bioenergy, lifecycle greenhouse gas emissions, machine learning, artificial intelligence, higher heating value, lower heating value

DISCLAIMER AND DECLARATION OF INDEPENDENT AUTHORSHIP

I hereby declare that this Master of Science (MSc) thesis has been written independently and represents my own original work carried out in accordance with the academic rules, ethical standards, and research integrity principles of University of Vaasa.

I confirm that, to the best of my knowledge and intention, I have not engaged in any form of academic misconduct, including but not limited to plagiarism, falsification, ghostwriting, unauthorized collaboration, or the misuse of Artificial Intelligence (AI), Large Language Models (LLMs), or other automated content-generation tools in a manner that violates university regulations or academic integrity standards.

Furthermore, I declare that this thesis has not been subcontracted, outsourced, purchased, or produced in whole or in part by any paid service provider, third party, commercial writing agency, or external individual. All analysis, interpretation, writing, and presentation contained in this thesis are the result of my own independent academic effort unless explicitly and properly referenced.

Where AI-assisted tools or digital technologies may have been used for limited support purposes permitted under university guidelines (such as language refinement, grammar checking, formatting assistance, or idea organization), such usage has been conducted responsibly, transparently, and without compromising the originality, intellectual ownership, or academic integrity of the work.

I fully acknowledge and accept responsibility for the contents of this thesis. I understand that if evidence of plagiarism, unauthorized AI misuse, academic dishonesty, or third-party authorship is discovered at any stage before or after submission, the University reserves the right to take appropriate disciplinary and academic actions in accordance with its regulations and policies. Such actions may include rejection of the thesis, annulment of the degree, disciplinary sanctions, or any other measures deemed necessary by the University.

By signing this declaration, I affirm my commitment to honesty, transparency, and ethical academic conduct.

Student Name: Hafiz Ahmed

Student Number: 2401924

Programme: Master of Science in Technology Industrial System Analytics

Title of Thesis: AI-Enhanced Framework for Evaluating Bioenergy Material Characteristics Linked to Lifecycle Emissions

Contents

List of Figures

List of Tables

1. Introduction

1.1 Background of the Study

There is a significant shift of the global energy system concerned with climate change, increasing energy demand and demand to minimize reliance on fossil fuels. It has been warned that since the state of the atmosphere, ocean, and land is clearly warmer due to human influence, which is mostly caused by greenhouse gas emissions because of the use of fossil fuels, deep mitigation needs to be implemented to stabilize the climatic conditions in the long term (IPCC, 2023). Meanwhile, the International Energy Agency states that renewable energy keeps increasing in terms of power, heat, and transport, whereas modern bioenergy is currently one of the primary pillars of the energy transition (IEA, 2024).

In this transition, biomass will have a unique role in being storable, transportable and convertible to heat, electricity, fuels, and other value products. Sustainable bioenergy has been highlighted by the International Renewable Energy Agency as a key to decarbonizing hard-to-abate sectors and to complementing a larger low-carbon energy system (IRENA, 2022). Biomass hence will not only be a renewable source, but also a strategic input to energy security, decarbonization of the industry, and valorization of waste. Nevertheless, biomass environmental value does not necessarily come free; it relies on the source of feedstock, efficiency in the conversion process, and the greenhouse emission in the lifecycle. This is why biomass should be evaluated as a compositionally diverse material that has different energy and sustainability performance such as feedstock type and supply chain (IRENA, 2022; U.S. Department of Energy, 2025).

Biomass proves to be relevant to climate, which is another source of a recent increase in interest in Biomass Carbon Removal and Storage (BiCRS). BiCRS pathways generate biomass to sequester carbon in the atmosphere and store it in stable forms, the climate value of the systems, however, relies on sustainable feedstock sourcing, lifecycle accounting, and storage integrity (U.S. Department of Energy, 2025). This renders the question of feedstock selection a key concern in biomass-based climate mitigation, since a feedstock which is viewed as potentially appealing in energy terms, can still be a feedstock with a poor sustainability score once its sourcing or conversion route is carbon-intensive. Lifecycle assessment is also, in this case, paramount in defining how the biomass systems genuinely cause net environmental benefit (U.S. Department of Energy, 2025).

Elemental composition, proximate analysis, and lignocellulosic structure are important factors that drive the quality of biomass feedstock (Güner et al., 2022). Biomass characterization as studies have indicated that feedstocks with high carbon and lignin contents and lower moisture and ash contents usually exhibit enhanced bioenergy performance whilst those with less favorable compositional balances undergo fewer heating values and diminished conversion performance (Sun et al., 2023). More recent research on biomass and machine learning has shown that the relationships are nonlinear and can only be effectively modeled using data-driven methods as opposed to single-equation approaches (Cacho et al., 2023; Z. Wang et al., 2022). This generates a high demand to have combined systems that would analyze biomass feedstocks, not solely upon the foundation of eradication of essence, but also concerning their lifecycle emissions and sustainability results. In this way, the present research is placed in the context of the overall world community to assist in low-carbon energy transition with improved assessment of biomass. The paper concentrates on the connection between the content of feedstock, bioenergy performance, and lifecycle emissions by utilizing an AI-based framework and determining biomass materials which are not only energetically efficient, but also environmentally desirable (Guldhe et al., 2017).

1.2 Research Gap, Question, and Objectives

There are several gaps that are significant in literature even though biomass is an energy source that has been studied extensively. Numerous studies concentrate on the biomass yield, heating value, fuel quality, or on the lifecycle assessment of greenhouse gas emissions and lifecycle sustainability, the literature does not always present the whole picture of the effects of the composition of feedstock on both, energy performance and environmental burden simultaneously (Cherubini et al., 2009). The second gap is a paucity in the utilization of combined AI-based models to assess biomass. Recent research revealed that machine-based learning has led to successful predictions of biomass characteristics and energy yields, yet numerous studies continue to focus on a single outcome (like biomass yield or heating value) but do not couple the predictions with life cycle emissions or sustainability index. Though it has also indicated that even wider integration on both energy and environmental levels is necessary (Cacho et al., 2023; Šurić et al., 2023).

Another gap is that interpretability of the model and importance of the features. Although machine learning models like ANN, RF, XGBoost and SVR can enhance the accuracy of the prediction, most past studies have failed to fully describe the biomass variables that are influencing the model outputs (Mari Selvam & Balasubramanian, 2023). It is becoming acknowledged in the literature that explainable AI is required to find out the most impactful feedstock properties and to create trust in machine learning-based biomass testing. Similary, traditional equations and emission-factor methods still are valuable as a baseline methodology, but they fail in most cases to provide nonlinear interactions between feedstock properties, energy content and lifecycle emissions. Current literature thus can present useful estimates on a screening level, yet they may not necessarily generate coupled and interpretable framework of comparing biomass feeds on energetic and environmental bases (Martins et al., 2019).

Despite being considered a promising renewable material to enter the energy systems with low carbon content, the evaluation of biomass has been a complex task due to the high heterogeneity of feedstock biomass, and the fact that energy and emission characteristics of biomass are linked to nonlinear relationships which we see in different existing literature based on databases and various of time period in table 1.

Table 1. Summary of gap analysis process and identified research gaps.

Dimension	**Description**
Databases searched	Web of Science, Scopus, Google Scholar, ScienceDirect, IEEE Xplore
Search period	2005–2025 (focus on recent developments in biomass sustainability, AI and machine learning, and lifecycle assessment)
Search keywords	Biomass feedstock, bioenergy, higher heating value, lower heating value, lifecycle greenhouse gas emissions, machine learning, artificial intelligence, biomass sustainability.
Gap type: Knowledge	Although biomass energy properties and biomass lifecycle emissions have been well discussed in the existing literature, few studies address the biomass composition, bioenergy prediction, and biomass lifecycle sustainability assessment in an integrated framework in a single analytical model.
Gap type: Evidence	The existing research primarily concentrates on the prediction of the energy parameters or the life cycle assessment of one independently rather than applying AI-based modeling for energy parameters prediction together with life cycle GHG emission estimation, especially for heterogeneous biomass feedstocks.
Gap type: Methodological	For most cases, traditional empirical equations or emission-factor approaches fail to provide the information needed to identify the nonlinear relationships among elemental composition, proximate properties, lignocellulosic structure and sustainability

	performance. Furthermore, several research studies do not have interpretable AI methods like SHAP, permutation importance, and sensitivity analysis methods that can be used to pinpoint influential biomass factors.
Gap type: Practical	Even with the present accessibility of biomass data and AI-based assessment methods, there is a limited set of frameworks that can support selecting biomass, predicting energy, and evaluating impact on the environment for low-carbon bioenergy decision making, while accounting for sustainability concerns.
Research contribution of my thesis	In this thesis, a comprehensive and explainable biomass integrated and AI-tailored framework is created, which integrates both biomass composition analysis and machine learning prediction of HHV and LHV, as well as estimated LCGHGs emissions to facilitate sustainable selection of biomass and assessment of low-carbon bioenergy.

Biomass materials vary in elemental content, proximate analysis, lignocellulosic structure, moisture, ashes, and feedstock type, and these variations can result in significant changes among higher heating value, lower heating value, and lifecycle greenhouse gas emissions. This leads to the fact that biomass cannot be evaluated using a single uniform rule, or a simple linear assumption, as even a category of the same feedstock can have various behaviors based on its compositional makeup and its processing history (Šurić et al., 2023; Wang et al., 2024). This heterogeneity poses a significant issue to the biomass characterization and sustainable feedstock choice (Ahmed et al., 2019). Simultaneously, a considerable amount of available research considers properties of the energy and emissions in distinct individual contexts as opposed to considering them as one whole framework. This imbalanced practice makes it hard to establish which biomass feedstocks offer the most ideal balance between the energy content and lifecycle emissions, especially when the dataset includes several different biomass classes like: agricultural residues, forest residues, woody biomass, pellets, chars, and urban biomass wastes (Martins et al., 2019; Šurić et al., 2023).

The issue that concerns this thesis is as follows: biomass analysis is not adequately integrated as frequently the feedstock constitution, energy probability, and the lifecycle emission analysis are independent analytical exercises. More specifically, the non-linear character of biomass characteristics poses a challenge to traditional approaches, and the lack of an interpretative AI-based framework prevents the prediction of bioenergy performance with reasonable precision and the comparison of feedstocks based on a shared sustainability metric (Liu et al., 2025).

The research is designed to offer a more precise, interpretable, and sustainability-focused method towards evaluating biomass feedstocks by incorporating feedstock composition, machine learning prediction, and emission estimation into one analysis (Sharmila et al., 2024). Informing the research, the first aim of the research is to design AI-driven predictive models to assess bioenergy properties, such as higher heating value (HHV), lower heating value (LHV), and bioenergy potential, based on the input feedstock composition and proximate analysis.

The last objective deals with establishing how well biomass composition can be utilized to carry predictions related to energy-related outputs using machine learning models. The second aim is to see the correlation between feedstock properties and the estimated lifecycle greenhouse gas (GHG) emissions. This goal also deals with the environmental aspect of biomass assessment, by investigating the relationship between various feedstock characteristics and type and the estimated lifecycle emissions. The most important factors determine energy content and emissions to facilitate sustainable biomass selection. This goal will identify the most significant biomass attributes in creating bioenergy performance and environmental performance and will apply the results to rank feedstocks in sustainable bioenergy uses that integrate energy forecast, emission projections as well as interpretation of variables to make sustainable decisions.

1.3 Definitions and Scope of the Study

Biomass feedstock which is used and implies any organic biological materials used for bioenergy production or bioenergy sustainability assessment such as agricultural residue, forest residue, woody biomass, biomass pellets, chars and urban biomass wastes. The elemental composition, proximate properties, and lignocellulosic structure of these feedstocks vary widely, and can have direct impact on how they will perform as an energy resource based on their lifecycle greenhouse gas emissions. These are very different materials when it comes to elemental composition and proximate properties which affect their energy performance, and that also directly affect their lifecycle greenhouse gas emissions.

Bioenergy is a renewable energy obtained from thermal and thermochemical process routes of biomass; the energy quality criteria are included as higher heating value (HHV), lower heating value (LHV) and bioenergy potential. The predominant environmental sustainability indicator by which biomass assessed in this research is the lifecycle greenhouse gas emissions, which encompass the approximate amount of greenhouse gas emissions associated with biomass-related activities across the whole products life cycle, including the collection of feedstocks, the preprocessing, transport, conversion and utilization.

Machine learning is a data-driven computational approach that can determine these highly nonlinear relationships between biomass feedstock properties and the resulting energy and emission predictions without completely depending on fixed empirical relationships.

Artificial intelligence (AI) is the use of predictive machine learning model Support Vector Regression (SVR), Random Forest (RF), XGBoost and Artificial Neural Network (ANN) for biomass energy prediction and sustainability assessment.

Lower Heating Value (LHV) is the amount of energy recovered from the biomass combustion excluding the latent heat due to the presence of water vapor in the biomass after combustion, but which is not recovered during the normal combustion process.

Higher Heating Value (HHV) is the amount of energy recovered from the combustion process including the latent heat recovered from water vapor. This is why these concepts make up theoretical and analytical background of the study as they provide a holistic perspective of the relationships between biomass composition, energy quality, lifecycle emissions and AI-based sustainability assessment.

Our aim at understanding this relationship, based on the composition of biomass feedstock, across the predicted bioenergy properties and the estimated GHGs over the life cycle of biofuels via an integrated interpretable AI framework. Biomass feedstocks included in the study were selected from the NREL biomass dataset(Harman-Ware & Paeper, n.d.), and biomass elemental composition variables, proximate analysis variables, and biomass lignocellulosic composition variables, as well as biomass feedstock categories, were used as prediction variables for HHV, LHV, and estimated lifecycle emissions.

We use machine learning models such as SVR, RF, XGBoost and ANN and compare their prediction performance with the traditional biomass energy estimation, which are based on empirical models, to determine if the machine learning models could offer a better predictive power for biomass energy and sustainability assessment. The study also performs feature engineering, SHAP analysis, permutation importance, and sensitivity analysis to realize the most significant biomass variables that impact energy performance and environment outcomes.

A detailed techno-economic analysis, policy evaluation, optimization of industrial-scale processes, laboratory experimentation, and a complete cradle-to-grave lifecycle inventory assessment. Rather than conducting a full process-based lifecycle assessment, the lifecycle emissions are estimated by using a comparative emission-factor mapping approach, which is based on references by IPCC and GREET. The boundaries are set with the aim to make sure the research tackles the key gap in biomass literature – the need for an integrated, interpretable and sustainability-based framework that can simultaneously link biomass composition, predict bioenergy, and account for lifecycle GHG to enable sustainable biomass selection and low carbon bioenergy assessment.

1.4 Structure of the thesis

This thesis has organized the five chapters that together explore the composition of biomass feedstock, their bioenergy properties and estimated life cycle GHG emissions with an interpretable AI framework.

Chapter 1, the background of biomass research, covering the increasing role of biomass in low carbon energy systems and sustainable bioenergy development. The research problems, research gaps, objectives, scope of the study, definition of the core concepts, and the direction of the thesis are discussed in the chapter. It also creates the motivation to combine biomass composition, prediction of bioenergy production.

Chapter 2 summarizes literature review findings concerning biomass feedstocks, biomass composition, higher heating value (HHV), lower heating value (LHV), life cycle greenhouse gas emissions, biomass carbon removal and storage (BiCRS), empirical biomass estimation approaches and bioenergy and biomass research applications that utilize machine learning. The chapter also looks at the comparative studies carried out between AI based and empirical biomass assessment methods, and the gaps found both from the theoretical, methodological, and practical point of view that allows the integrated interpretable biomass assessment framework to be developed.

Chapter 3 describes Materials and Methods used in the research. The chapter presents the dataset from the National Renewable Energy Laboratory (NREL), research design, variable selection, data preprocessing methods, feature engineering techniques, lifecycle emission estimation methods, and machine learning models employed in the study, such as Support Vector Regression (SVR), Random Forest (RF), XGBoost, and Artificial Neural Network (ANN). This chapter also introduces the empirical baseline framework, model validation procedures, SHAP analysis, permutation importance, sensitivity analysis and the comparative AI-based and empirical assessment framework. **Chapter 4**. presents results and discussion. To extend, it is analyzed with Descriptive statistics, distribution analysis of the biomass feedstock properties, correlation analysis, feature engineering results, machine learning model performance, comparison between actual and predicted HHV and LHV, estimated life cycle emissions, and comparative results between AI based approach and empirical approach are also provided in the chapter. SHAP analysis, feature importance, sensitivity analysis, feedstock ranking and applications of the results for sustainable biomass selection and low-carbon bioenergy assessment are also introduced in the chapter.

Chapter 5 presents a summary of the key findings, theoretical insights, implications for practice, as well as limitations and recommendations for future studies. Finally, the overall contribution of the research towards the development of an integrated, interpretable and sustainability-oriented framework for biomass feedstock assessment and bioenergy decision making is discussed.

2. Literature review

2.1 Introduction

Biomass has emerged as one of the most valuable renewable sources during the shift towards sustainable energy systems since it can be turned into heat, electricity, fuels and value-added productions as well as reducing the reliance on fossil-based energy sources. Against a background of growing interest in climate change, energy security, and carbon emissions bioenergy based on biomass is seen as viable alternative to aid low-carbon development. The feasibility and sustainability of biomass usage however is heavily dependent on the characteristics of the feedstock, the characteristics of the conversion pathway and the environmental cost of production and utilization. Consequently, biomass cannot be considered as a homogeneous resource; instead, it needs to be considered by its make-up, power quality and life cycle influence. The supply chains and bioenergy systems of biomass are very elaborate as biomass must undergo several processes such as production, collection, storage, transportation, preprocessing, conversion, and final consumption. This necessitates greater insights into these intertwined phases, according to Martins et al. (2019), biomass-to-energy supply chain management, because any inefficiency or loss at one step will affect the overall sustainability and economic viability of the system. Their bibliometric assessment based on which revealed that the research in biomass supply chain management has grown very substantially; the area, however, requires greater assimilation of logistics, sustainability, and decision-making instruments.

Biomass materials vary in terms of the elemental composition, the proximate properties and the lignocellulosic structure, which in turn directly affect the energy contents and the conversion behavior. The bioenergy crops like switchgrass, Miscanthus, and Virginia

Mallow have been studied with the result that properties like carbon, hydrogen, ash, lignin, cellulose and hemicellulose, moisture, fixed carbon, and volatile matter have a close relation to yield, heating value, and fuel quality. As Cacho et al. (2023) have identified the climate, the soil, topography and management factors to have a strong impact on the biomass yield in advanced switch grass cultivars, Šurić et al. (2023) have determined that harvest season can influence biomass yield and heating value in Miscanthus and Virginia Mallow significantly. The need-to-know feedstock composition as well as environmental factors during the assessment of biomass as a feedstock to be used in bioenergy.

Higher heating value (**HHV**) and lower heating value (**LHV**) are some of the most essential energy parameters that are used in researching biomass as they are the most common parameters used to measure the quality of biomass fuels. HHV and LHV depend on both the elemental composition and structure, implying that feedstocks which contain higher carbon and lignin usually perform better in energy terms compared to feedstocks with higher moisture or ash levels. Meanwhile, these outputs cannot be predicted by only one variable, but rather by a mixture of interacting characteristics; which make possible prediction of them rather challenging than usual, using only traditional linear methods or methods that utilize equations. The literature has thus come to appreciate that characterization of biomass energy is more of a multivariate issue than a one fact computation.

Lifecycle greenhouse gas emissions are another highly important evaluation criterion related to biomass sustainability along with energy content. The biomass industry is now being considered not just in terms of its capacity to generate energy but also in terms of the possibility of decreasing emissions along the chain. Martins et al. (2019) identified that the environmental profile of biomass systems could be impacted by the decisions related to the supply chain, transport, logistics, and processing, whereas Wang et al. (2024) emphasized that contemporary bioenergy systems had to have instruments that would observe even the subtle aspects of the processes and connect them with sustainability outcomes. It is especially essential as a high-HHV feedstock can be considered environmentally undesirable, even though it must undergo a lot of processing or travel long distances or even go through conversion routes with lots of emissions.

Machine learning has thus become an encouraging modality of research on biomass and bioenergy as it can simulate non-linear links between the properties of feedstock, energy performance and their environmental performance. Wang et al. (2024) described that machine learning comes in particularly handy when theory-informed models have become so inflexible, narrow-minded or reliant on simplistic assumptions. Similarly, research on the improved switching grass yield and biomass characteristics has demonstrated that the ensemble models and neural networks can perform better compared to conventional statistical techniques when the two variables in relation are complicated and the data type is heterogeneous. This renders machine learning suitable for biomass characterization, energy forecasting, and sustainability evaluation, especially in cases where one wishes to identify hard-to-define trends that cannot be defined through normal empirical equations.

2.2 Biomass Carbon Removal and Storage (BiCRS) and Sustainable Feedstock Selection

The concept of Biomass Carbon Removal and Storage (BiCRS) represents a new pathway which can capture the carbon dioxide in the atmosphere and sequester it within biomass in some durable form, whether as a permanent underground storage deposit or in a long-lasting product. The term is explained in the BiCRS roadmap literature as a process that utilizes biomass to extract CO_2 out of the air, sequesters that carbon, and is not harmful to the food security, rural livelihoods, biodiversity conservation, and other valuable social and environmental values (Sandalow et al., 2021; U.S. Department of Energy, 2025).

Biomass feedstocks have a role to play in BiCRS beyond energy production. The carbon-removal pathway begins with feedstocks, the source of which must in turn be sourced, processed, and converted into a useful form, which in turn dictates whether the system has provided any real net climate benefits. In the literature, it is noted that feedstock logistics are important since the biomass needs to be harvested or collected, transported and delivered to the conversion facility in a conversion-ready form (U.S. Department of Energy, 2025). The supply of biomass is dispersive and extremely variable, as well, a factor that makes the quality of feedstock, cost, and sustainability key concerns in large-scale bioenergy systems.

2.3 Biomass Feedstocks for Bioenergy Production

Biomass feedstocks are the cornerstone of bioenergy industries as they are used as the raw material, which is transformed into heat, power, fuels, and value-added products, bio-mass is not a homogenous resource, but rather a general group of resources that encompass agricultural residues, forest residues, and woody biomass, pellets, chars, and urban biomass wastes. The composition of these feedstocks varies, as well as their moisture content, ash content, structural properties and processing history, and as a result, their applicability to bioenergy production also differs significantly. The choice of feedstock is, in turn, one of the most significant stages of biomass-based energy systems due to its impact on energy efficacy, the conversion dynamics, logistics, and lifecycle sustainability (Martins et al., 2019; Wang et al., 2024).

One of the best-known biomass feedstocks that are broadly debated in literature is agricultural residues due to their abundance, renewability, and accessibility as a by-product of food and crop production. Energy can be generated using materials like corn stover, cereal straw, rice husk, and sugarcane bagasse and they do not need special land to be grown. This renders them appealing to the resource-efficiency lens, especially in systems where resource-critical objectives are waste minimization and the use of both new and old resources in a circle. Meanwhile, agricultural residues tend to be both heterogeneous and seasonal in nature, and can be influenced by other competing uses, (such as soil cover, bedding, and industrial processing). Consequently, the agricultural residues must be considered not just based on their abundance, but also regarding their collection ability, quality reliability, and sustainability considerations (Martins et al., 2019; Wang et al., 2024).

Another large category of feedstocks in bioenergy studies is forest residues and woody biomass. These comprise branches, bark, thinning residues, logging residues, chips and the rest of the products of forestry work or other wood processing work. Woody biomass is often more structurally sound and potentially offers good combustion or conversion properties when dried and properly processed compared to many of herbaceous biomass. Another common form of conversion of woody biomass is densified, like pellet or char products, which enhance handling, storage, and transportation efficiency, which woody feedstocks are also significant in terms of contribution to sustainable bioenergy systems due to their large quantities in many cases as well as the ability to be used both in the combustion and thermochemical conversion pathways (Martins et al., 2019).

The processing of biomass products like pellets and chars are gaining importance, as they make biomass much more viable and denser in terms of its energy content. Biomass products in the form of pellets are highly utilized in combustion and co-firing systems, whereas char is solid products of thermochemical conversion (pyrolysis) that consist predominantly of carbon. Chars typically have more fixed carbon and less volatile matter than raw biomass and are thus usually more useful in energy-dense applications. Meanwhile, pellets and chars add further steps that could add to the lifecycle emissions or the cost of production that should be measured after not only their heating capacity, but also their environmental and supply-chain performance. Nonetheless, they tend to be more heterogeneous than agricultural or forest residues and are sometimes contaminated with ash-rich fraction, or mixed streams of material that can make processing more challenging and lower fuel quality. As such, focuses on urban biomass wastes as potentially promising yet more challenging feedstocks to be carefully characterized prior to effective energy or emissions measurements (Martins et al., 2019; Wang et al., 2024).

The biomass quality and conversion performance are also intimately connected to the choice of feedstock. Experiments involving energy crops like switchgrass, Miscanthus and Virginia Mallow indicate that potential energy and fuel quality may be highly affected by the biomass characteristics, harvest time, climatic conditions, and management activities. Cacho et al. (2023) proved the relationship between yield in advanced cultivars and switchgrass with environmental and agronomical variables, and Šurić et al. (2023) the time of harvest varies the yield of biomass in Miscanthus and Virginia Mallow and their energy properties. These results add to the larger body of work that biomass feedstocks cannot be viewed as static materials and that their energy activity relies on composition as well as growing conditions instead of feedstock.

2.4 Biomass Composition and Its Influence on Energy Properties

The chemical and structural make up of biomass dictate greatly the energy potential of its materials. The biomass feedstocks vary in terms of elemental composition, proximate analysis, and lignocellulosic structure and these variations directly affect their utility in bioenergy generation. This is why biomass characterization is an essential step in the assessment of the using feedstock, as the ratios of carbon, hydrogen, nitrogen, oxygen, moisture, volatile matter, fixed carbon, ash, cellulose, hemicellulose, and lignin can determine the ultimate energy behavior of a given feedstock, which cannot be characterized using the category of feedstock alone as, in many instances, the compositional differences form most.

One of the most significant measures of the quality of biomass energy is elemental composition. The primary element in biomass that contains energy is carbon, with greater function of carbon content in general linked to greater heating value. Energy release during combustion is also partly due to or contributed by hydrogen although it is not considered independently but usually in association with carbon and oxygen. Conversely, the presence of oxygen typically relates to lower energy density since oxygen-rich biomass has lower accessible chemical energy per unit mass. It is also common to avoid nitrogen and sulfur in biomass fuels as they are neutral or harmful to energy production and could cause pollutants to form during the conversion process. Consequently, biomass materials that have more carbon and hydrogen and less oxygen, nitrogen, and sulfur are generally more preferable to use in energy (Wang et al., 2024).

There is also a need to explain the properties of biomass energy by using proximate analysis. The typical proximate variables that are used when assessing the quality of biomass fuel include fixed carbon, volatile matter, ash and moisture. The fixed carbon is typically linked to the increased stability of solid fuel and the increased energy density, whereas the volatile matter influences both the process of ignition and combustion. Moisture reduces the effective energy content due to evaporation of water to result in combustion energy. Ash cannot be burnt, and thus, a high content of ash tends to diminish fuel quality and potentially cause operational issues, such as slagging, fouling, or reduced conversion efficiency enhancing moisture and ash contents subsequently build up to richer levels, while fixed carbon tends to increase the HHV and fuel performance (Šurić et al., 2023; Wang et al., 2024).

Recent studies in machine learning affirm these chemical interpretations by demonstrating that biomass energy properties can be governed by a condensation of elemental, proximate, and structural variables. Variables that come out prominently in explainable models are; fixed carbon, carbon content, lignin, xylan, moisture, and ash among others that are likely to predict higher heating value and lower heating value.

Usually, the heating value is generally increased by the presence of fixed carbon and carbon content whereas moisture tends to decrease the heating value. Lignin and xylan are also significant structural characteristics that point to the idea that energy properties are shaped by a complex of traits of biomass and not specific chemical factor (Hosseinzadeh-Bandbafha et al., 2021). Attention can be paid to higher heating value (HHV) and lower heating value (LHV) as they are the main parameters of the quality of biomass energy. HHV reflects all energy that is released during combustion, plus the latent heat available in the form of the water condensed, and LHV indicates the actual usable energy by ignoring the latent heat. A high moisture or ash content in biomass can lead to poor energy value despite a good composition on the raw biomass. That is why both HHV and LHV are affected by the actual chemical composition of the biomass and by the physical state of the biomass during the analysis, rendering these energy parameters integrated measures of the feedstock chemistry instead of the single-variable characteristics (Šurić et al., 2023; Wang et al., 2024).

2.5 Higher Heating Value and Lower Heating Value of Biomass

Two of the most useful indicators in biomass characterization are higher heating value (HHV) and lower heating value (LHV) since they indicate the amount of energy that a feedstock can provide and its ability to be used in bioenergy applications. These are parameters commonly employed throughout literature to make comparisons between biomass materials, assess the quality of fuels and to show whether a feedstock would be better eluted to combustion, gasification, pyrolysis, or other forms of thermal conversion. Due to the significant variability in biomass feeds composition and structure.

HHV and LHV offers an effective means to convert those variations into an energy-based metric to aid in feedstock selections and bioenergy planning (Friedl et al., 2005). The chemical composition of biomass is very important to HHV and LHV. Feedstocks that are richer in carbon and lignin are more likely to be high heating value whereas more moist, richer in ash, oxygen, and nitrogen feedstocks tend to be lower in effective energy content. Biomass with higher carbon content has higher chemistry energy-per-unit-mass energy and the importance of lignin is higher due to its greater carbon concentration and thermal resistivity as compared to cellulose and hemicellulose. This means that feedstocks that are rich in lignin will usually be more favored in terms of direct combustion, and biomasses with more carbohydrate fractions may be more suitable in biochemical conversion paths. The following compositional effects explain why HHV and LHV are not fixed properties, but the results of the elemental and structural composition of feedstock (Maksimuk et al., 2021). The properties of energy are the result of a combination of multiple feedstocks features instead of being in the effect of one overpowering driver (Demirbas, 2002).

HHV and LHV can also be important in terms of lifecycle sustainability. The amount of usable heat generated per unit mass can be increased by higher heating values and this could lessen the quantity of feedstock needed to generate a particular amount of energy. Nevertheless, the relationship between energy content and sustainability is not as direct as one could think since a biomass pathway with a desirable HHV or LHV could still contain a large amount of lifecycle emissions by having to undergo an intensive process, cover a long transport distance, or convert with a low efficiency. That is why literature tends to suggest even more to correlate heating values with a lifecycle evaluation of greenhouse gas to compare feedstocks available on both bases, energetic and environmental. This holistic view is particularly relevant in evaluating sustainable bioenergy, in the context of which the goal should not be solely to maximize the energy content, but also reduce environmental burden (Cherubini et al., 2009).

2.6 Lifecycle Greenhouse Gas Emissions of Biomass-Based Bioenergy

One of the key criteria applied to evaluate biomass-based bioenergy is lifecycle greenhouse gas (GHG) emissions since renewable origin is not a sufficient factor to assure of climate benefit.

Despite the common discussion of biomass as an example of a low-carbon option as fossil fuels, the overall performance of a biomass pathway in environmental terms depends on the overall sequence of operations involved in the production and harvesting of the feedstock, collection, preprocessing, transportation, conversion and later utilization of the final product. It is due to this that biomass systems should be evaluated on a lifecycle basis, other than based on energy content, because a feedstock that has a good heating value could still have an undesirable emission profile, in the case where its supply chain is energy-intensive or inefficient (Thornley et al., 2015). There can be GHG emissions occurring at multiple levels of bioenergy chain in biomass systems. Such emissions can be related to the cultivation or collection processes, usage of machinery, fertilizers, drying, storage, transportation, conversion losses and after processing needs. Moreover, indirect effects like land-use change, soil carbon variation, and variation in conversion efficiency can be experienced in biomass systems. Cherubini et al. (2009) demonstrated that they can considerably change the final climate profile of bioenergy systems, and that is why the same biomass pathway can seem to be advantageous in one research and less advantageous in another in case of various assumptions. It means lifecycle emissions are never some kinds of feedstock, but rather the outcome of the system where it is utilized and the land itself.

The most recognized method of assessing the environmental sustainability of the system of biomass-based energy is lifecycle assessment (LCA). LCA enables scientists to approximate the GHG emission throughout the complete life cycle of feedstock and calculate the outcomes and contrast the findings with fossil-centered reference systems. Emission-factor approaches are a common alternative in biomass research in cases where a complete process-based LCA could be lacking. Each group of biomass or feedstock is assigned a standard emission factor that typically relates in kg CO_2 - equivalent to unit of mass or energy. This allows an approximate calculation of emissions in the lifecycle in a short period of time, and a broad comparison of a variety of feedstocks in a similar framework. Both the article by Martins et al. (2019) and Wang et al. (2024) emphasize the relevance of simplified sustainability tools in case of biomass decision-making when screening feedstocks or incorporating emissions into a predictive model is the goal. In contrast, a feedstock with a moderate energy content can also be a preferable sustainability option as its lifecycle emissions can be lower which integrated assessment methods that can take into account the bioenergy quality, as well as lifecycle emissions (Johnson et al., 2011).

2.7 Empirical and Conventional Methods for Biomass Energy and Emission Estimation

Prior to machine learning being employed extensively in biomass research, biomass energy traits, and lifecycle emissions were generally determined by means of anecdotal equations, traditional statistical analyses, and life cycle evaluation techniques relying on compiled emission factors. These simplistic approaches are still significant due to their simplicity and transparency and their extensive use concerning screening biomass feedstocks, comparing the performance of energy, and making predictions about environmental impacts in circumstances where only partial information is known. In biomass and bioenergy research, the empirical methods have frequently been the benchmark on which more sophisticated models of data-driven models can be tested (Cherubini et al., 2009; Wang et al., 2024).

Given the easy to measure feedstock characteristics, higher heating value (HHV) and lower heating value (LHV) are often estimated using empirical equations with feedstock elemental composition, proximate analysis or lignocellulosic fractions as inputs. These models are appealing as they do not need too much calculation and can be directly applied to a bio sample under conditions that only simple compositional data are known (Dogan & Inglesi-Lotz, 2017).

Traditional emission estimation of biomass research is mostly dependent on emission-factor mapping. All groups of biomass or feedstocks have a common standard emission factor typically presented in terms of kg CO_2-equivalent divided by unit mass or unit energy. This is useful in allowing quick estimation of life cycle greenhouse gas emissions and comparison between a large range of biomass types is of particular interest when a screen level of sustainability is needed or when process information is not available at the requested detail. Meanwhile, the process of emission-factor mapping can be simplified by definition, as it lacks the full scope of feedstock peculiarities in logistics, transport intensity, processing conditions, or regional supply-chain variations (Cherubini et al., 2009; Martins et al., 2019).

The best-known assessment of biomass-associated greenhouse gas emission is conventional lifecycle assessment (LCA). Cherubini et al. (2009), demonstrated that the LCA outcomes of biomass and bioenergy systems are highly contingent on decisions related to the methodology including system boundary, allocation rules, fossil reference system, functional unit, and accounting stock of carbon. Such options may result in highly disparate emissions outputs when it comes to the evaluation of the same biomass feedstock. Traditional LCA, as such, is viable and well regarded, but needs broad background knowledge and methodological coherence in such a way that feasible results can be obtained. This makes LCA an effective sustainability model.

Conventional methods of biomass energy and emission estimation can thus be considered as the baseline method. They can be used as a preliminary screen, a place to make fast comparisons, and a general sustainability assessment, but are less precise at capturing the complexity of interactions between the feedstock composition, energy content, and lifecycle emissions. They are not as good at capturing nonlinear correlations or a combination of several compositional factors. That is why numerous recent papers started involving and integrating the classic approaches with machine learning models that are more competent at working with heterogeneous biomass data and operating with many variables simultaneously (Dogan & Inglesi-Lotz, 2017).

2.8 Machine Learning in Biomass and Bioenergy

There is a growing interest in applying machine learning in biomass and bioenergy research as biomass systems are highly variable, nonlinear, and affected by numerous interacting factors. The bioenergy results of feedstock stand, moisture content, lignocellulosic structure, growing conditions, harvest time, conversion technology, and environmental variables influence the bioenergy outcomes, typically imprecisely modeled by simple linear equations that have since evolved to machine learning to enhance better prediction and classification, optimization, and interpretation of biomass related research (Wang et al., 2022). Prediction of biomass yield is also one of the major applications of machine learning in the field of biomass research. The yield is a multivariate problem since biomass production is based on climate, soil, topography, management practices, and the type of crop. Cacho et al. (2023) revealed that machine learning models can effectively determine the biomass yield of advanced cultivars of switch master that are planted in marginal agricultural fields by combining the environmental, soil, topographic and agronomic factors. Their experiment revealed that ensemble models like a random forest and gradient boosting machine had particularly reliable performance when it comes to heterogeneous data and determining the most influential predictor of yield.

Prediction does not always work in biomass and bioenergy studies; researchers also want to know what variables contribute to the model. Wang et al.,(2024) underlined that feature selection and interpretability tools should be used in combination with machine learning to be able to find the most influential variables. This can be particularly handy in biomass studies since composition-based variables e.g., carbon, hydrogen, lignin, xylan, moisture, and ash as well as fixed carbon tend to have different impacts on energy, and environmental performance. Machine learning can be used to assist in scientific interpretation and practical feedstock selection by discovering the most significant predictors.

The second significant implication of machine learning in biomass studies is that it can help to analyze complicated bioenergy conversion routes. Bioenergy systems entail biological conversion, thermochemical conversion and complete supply-chain operations which tend to be undesignable through theory-based approaches alone. As stated by Wang et al. (2024), data-driven methods can be instrumental when the model in question is too complicated, when the empirical data come in a heterogeneous form, or when conventional methods are too inflexible to reveal nonlinear behavior.

Machine learning thus can be deployed to supplement traditional bioenergy modeling by learning on the data at hand and learning patterns that are not readily expressed in closed-form equations. Advancing applications of machine learning in biomass studies is also connected to the rise in the number of different data sources. Biological experiments and literature-based data, industrial processes, sensor signals, and even pictures or structural data are now taken advantage of in biomass studies. Even though, predictive performance can be very high when using models like artificial neural networks, boosting techniques, as well as, ensemble learners, they tend to appear as a black-box system without the use of explainability tools (Wang et al., 2022).

The new biomass literature thus stresses the application of SHAP, permutation importance and partial dependence to determine the role of individual variables in outputs that include HHV, LHV and biomass yield. In this regard, machine learning finds more applications beyond prediction and the ability to indicate which biomass features are the most important.

2.9 Comparative Studies Between AI-Based and Empirical Approaches

The evaluation of biomass and bioenergy is beginning to compare artificial intelligence (AI)-based methods with the traditional empirically based methods since the two are applied to approximate biomass energy characteristics and environmental footprint, although in entirely different mechanisms. Simplicity, transparency, and ease of use of the former methods have made empirical methods more cherished, whereas the methods of the latter are aimed at learning complex patterns directly through data. Traditional methods of empirical studies are still significant in the study of biomass since they can supply a precedent to assess higher heating value (HHV), lower heating value (LHV), and greenhouse gas (GHG) emission. These techniques usually assume the use of common equations, emission-factor mapping and the practice of lifecycle assessment. Liao & Yao (2021) have highlighted that the traditional lifecycle assessment is still an effective model to analyze bioenergy, as it has offered a systematic methodology to compare biomass options to fossil-based ones. This implies that although empirical approaches work well regarding screening and comparison, they may not be descriptive enough of biomass systems.

Conversely, AI-based approaches are becoming common since they can form nonlinear associations and factor several feedstock characteristics simultaneously. This is particularly true when using heterogeneous biomass data, where the type of feedstock, elemental composition, proximate analysis, and structural composition all play an interactive role in energy production and the environment. One of the major strengths of empirical methods is their interpretability. Standard equations and factors of emission are simple to grasp and use and thus they are applicable during the stage of screening at the initial stages. Their primary shortcoming, however, is that they tend to make simplified assumptions about relationships and might not scale well to feedstock-specific variance. The same empirical formula is used on highly dissimilar biomass materials in most instances, although those feedstocks can be chemically different, differ in moisture or processing history. This may cause error of prediction or over-simplification (Hassan et al., 2025).

Instead, AI-based approaches are more flexible and have improved predictive capabilities, albeit with new interpretability and data quality concerns. Great care is needed when modeling a nonlinear structure using models like random forest, XGBoost, support vector regression, or artificial neural networks, as they also can be black-box systems, unless explainability tools are employed. Wang et al. (2024) emphasized the role of SHAP, feature importance, and partial dependence plots in increasing the transparency of AI models.

Most of the AI-based and empirical methods are also complementary to each other as opposed to being mutually exclusive. Empirical methods offer a clear standard and a sensible starting point, whereas AI-based methods enhance flexibility, nonlinear modeling, and feature discovery. This complementarity has been especially useful in the case of biomass where lifecycle emissions and energy properties can be best understood through the lenses of both direct calculation methods and data-driven prediction methods (Cherubini et al., 2009; Wang et al., 2024).

2.10 Research Gaps in Biomass Energy and Lifecycle Emission Modeling

Biomass feeds are variable in nature and gathering suitable data is most of the time expensive, time-consuming, and subject to weather in the field, laboratory tests. The second major weakness is the relative paucity of lifecycle greenhouse gas emissions incorporated in biomass energy prediction studies. Numerous studies that deal with biomass yield, heating value, or conversion performance have been conducted to date, but they fail to include emissions as a component of the predictive framework. This presents a rift between energy characterization and sustainability assessment even though both need to appropriately assess biomass in low-carbon energy systems.

Cherubini et al. (2009) demonstrated that lifecycle assessment is important in terms of comprehending the climate performance of bioenergy pathways. The other crucial weakness of literature is lack of interpretability of models. Although more sophisticated machine learning algorithms like artificial neural network, gradient boosting, and ensemble models can lead to high prediction accuracy, they tend to function as a black box without any explanation mechanism unless they are used with explanation tools. This can be a severe drawback of biomass research in that decision-makers must have an idea of not only which feedstock excel but why feedstock excels. The interplay between feedstock composition, energy prediction, and emission estimation is not a common practice in many studies. The properties of biomass energy can be investigated outside the context of lifecycle emissions, and characterization of feedstock is viewed as an independent field of study as compared to that of sustainability. Overall, this fragmentation complicates the process of comparing feedstocks and working with a common basis and undermines the practical usefulness of the results (Aalto et al., 2019).

3. Materials and Methods

3.1 Research Philosophy and Methodological Framework

This study was developed as quantitative and comparative research with the objective of a study on the relationship between the properties of biomass feedstock, bioenergy performance and estimated lifecycle GHG emission. Measurable variables in the research are carbon (C), hydrogen (H), nitrogen (N), moisture (Mo), ash (Ash), volatile matter (VM), fixed carbon (FC), glucan, xylan, lignin, high heating value (HHV), low heating value (LHV) and estimated emissions. A positivist position in the study was used because the study analyzed the numerical data and compared the objective model outputs.

The research adopted the deductive approach since it was based on the existing understanding of relationship between biomass composition and heating value, fuel quality and environmental performance. With this knowledge, machine learning models were used to verify if the properties of the feedstocks could be used to estimate HHV/LHV and lifecycle emissions better than existing empirical methods. Thus, the step from now existing biomass energy theory to model-based testing and comparison was made.

This methodological design was mono-method quantitative since the study relied on secondary biomass data and computational analysis and not interviews, surveys, or field observations. The time horizon for this dataset was cross-sectional because one snapshot in time (the point of this study) was used, and not multiple snapshots collected over a long timeframe. This design was appropriate as the goal was the comparison of groups of biomass feedstock and the relationship between their differences in composition and energy and emission results.

The study has been carried out in a systematic approach with data preprocessing, feature engineering, lifecycle emission estimation, empirical baseline calculation, machine learning model development, model validation, and model interpretation. Using R2, MSE and RMSE, the performance of the models was evaluated while SHAP analysis, permutation importance and sensitivity analysis were used to explain the influence of the main biomass variables. The research was made transparent and comparable through this process and aligned with the main gap in the research: develop an integrated framework between biomass composition and bioenergy prediction and lifecycle greenhouse gas emissions.

3.2 Data

The collection of the data was done at the National Laboratory of the Rockies (NREL), Colorado, USA, on the project titled as, Addressing Critical Measuring, Reporting, and Verification Challenges of Durability and Sustainable Sourcing of Feedstocks to Biomass Carbon Removal and Storage (BiCRS) Pathways. It is a Biomass data on Carbon Removal and storage, Biomass and Bioproduct Data Table maintained by (Harman-Ware & Paeper, 2025). The reason behind the selection of the dataset is that it offers feedstock level data applicable to both bioenergy characterization and lifecycle emission assessment which directly corresponds to my goals. The types of feedstocks are agricultural residues, forest residues, woody biomass, pellets, chars, and urban biomass wastes. And dataset includes elements composition variables, proximate analysis variables, lignocellulosic composition variables, and energy property.

The key variables are C%, H%, N, moisture, volatile matter, fixed carbon, ash, glucan, xylan, acid-insoluble residue, lignin, HHV, LHV, and the estimated lifecycle greenhouse gas emissions. Its use in machine learning analysis is appropriate as it includes compositional predictors, energy- and emission-related responses variables where one can evaluate the biomass quality and sustainability.

3.3 Research Design

The research design was quantitative, comparative, and predictive to investigate the effect of biomass feedstock composition on bioenergy properties and greenhouse gases lifecycle emissions when feedstock characteristics were taken as the predictor variables and higher heating value (HHV), lower heating value (LHV) and estimated lifecycle emissions were taken as the response variables. This method follows the recent biomass modeling literature that focuses on applying data-driven frameworks to model nonlinear dependencies between feedstock characteristics and bioenergy production (Wang et al., 2024).

We pursued a relative analytical model where the required AI-based prediction was compared with traditional empirical estimation. Empirically based HHV, LHV and emission estimation tools are prone to simplify assumptions (Cherubini et al., 2009). This was the motivation behind the current design to compare predictive capability of machine learning models to: empirical calculations and whether AI-based techniques have greater accuracy and interpretability to biomass assessments. Recent research on biomass has revealed that even in cases where the goal is to relate biomass composition, energy quality and sustainability analysis in single framework, such integrated designs (Cacho et al., 2023; Šurić et al., 2023).

Figure 1, an AI-based model of evaluating the properties of biomass feeds and connecting them to lifecycle assessment of emission. During the preliminary phase, the biomass feedstock characteristics are taken as the primary input of the scheme. They are agricultural residues, forest residues, wood pellets, urban biomass wastes, and lignocellulosic constituent's glucan, xylan and lignin. These feedstock properties give the fundamental data to assess bioenergy potential. One then goes to preprocessing and feature engineering to prepare the dataset to be analyzed by machine learning. This step involves data cleaning, normalization, and development of features ratio like C/N ratio and fixed carbon to volatile matter ratio. SHAP interpretability is also added so that each variable is told to impact the model output.

Machine learning models are then applied to make predictions on bioenergy properties. They are the Random Forest model, the XGBoost model, the Support Vector Regression model and the Artificial Neural Network model. The models apply due to the ability to obtain complicated relations between feedstock composition and energy performance. Then, the framework forecasts HHV and LHV as the key energy output variables. HHV depicts higher heating value and LHV depicts lower heating value. The outputs play a role in defining the quality and applicability of various biomass feedstocks in regard to energy.

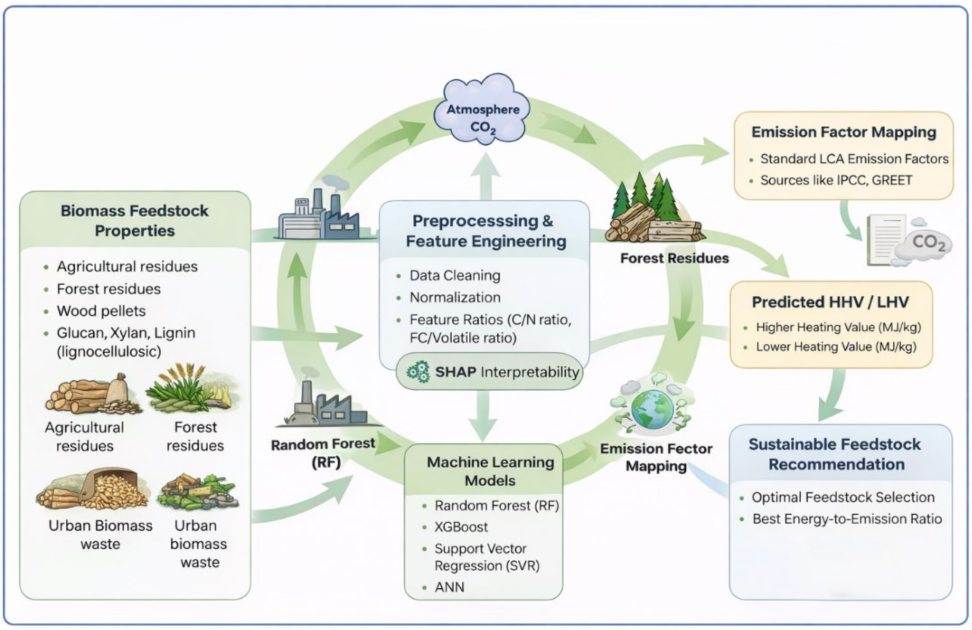

Figure 1. Biomass Feedstock Selection and Lifecycle Emission Assessment

This is followed by the inclusion of emission factors mapping in the structure. Lifecycle greenhouse gas emissions are estimated using mainstream LCA emission factors like IPCC (IPCC, 2023) and GREET sources (Long & Liu, 2023). The step connects the feedstock characteristics and estimated energy quantities to the environmental impact. This is then accompanied by sustainable feed stock recommendation.

Depending on the predicted HHV, LHV positions, and predicted lifecycle emissions, the framework determines the feedstock of biomass. The optimality of the feedstock considers the energy performance and the capacity to reduce the emissions. In the last phase, the structure aids in sustainable bioenergy decision-making. It can be used to determine the best feedstock which has the best energy-to-emission ratio. Thus, the diagram is a full-fledged AI-based biomass assessment, lifecycle emissions estimation, and sustainable feedstock selection.

3.4 Variables and Feature Definition

The variables in the study were construed in regard to their place in the predictive and comparative framework. The independent variables were feedstock composition and feedstock type, whereas the dependent variables included higher heating value (HHV), lower heating value (LHV) and estimated lifecycle of greenhouse gas emission. Variables of elemental composition, proximate analysis, lignocellulosic composition, and categorical feedstock were included in the dataset. The choice of these variables is due to the fact that according to the literature, chemical composition, structural composition, and the category feeding stock have a significant impact on the quality of biomass energy and environmental performance (Wang et al., 2024).

Variables of the elemental composition were carbon (C%), hydrogen (H%) and nitrogen (N%). These variables are the fundamental chemical composition of the biomass and are commonly used in biomass characterization as they are strongly related to fuel quality and combustion properties. Carbon has a positive effect on the energy content, hydrogen on combustion energy, and carbon dioxide is significant when considering the fuel composition and the impact that it may have on emissions. The analysis variables included the proximate variables moisture, volatile matter, fixed carbon, and ash. These were included because they characterize the thermal and combustive properties of the feedstock, and they are more generally used in biomass energy research to characterize differences in heating value and transformation performance. Moisture and ash tend to decrease the effective use of energy, but fixed carbon and volatile matter are the key points regarding the quality of fuel and thermal stability (Wang et al., 2024).

The variables were the lignocellulosic composition (glucan, xylan and acid-insoluble residue (Klason lignin)). These variables were chosen due to their expression of the structural fractions of biomass, and the importance in determining the impact of feedstock chemistry on bioenergy properties. Lignin is generally related to increased energy value whereas glucan and xylan are part of more extended carbohydrate system of biomass. Besides the compositional variables, feedstock type was a categorical variable since various biomass types, including agricultural residues, forest residues, woody biomass, pellets, chars, and urban biomass wastes, could have different profiles of their energy and emissions despite a chemical composition that might seem similar (Martins et al., 2019).

HHV, LHV and estimated lifecycle emissions were the dependent variables in this study. HHV and LHV were considered as the primary bioenergy performance indicators, lifecycle emissions as the primary sustainability indicator. These response variables have been selected since they enable the assessment of biomass feedstocks to be used in an energetic and environmental approach. Literature reports that GHG greenhouse gas emissions throughout lifecycle should be also used in the assessment of biomass, in addition to heating value, to establish a list of energy-rich and environmentally preferable feedstocks (Cherubini et al., 2009).

3.5 Data Preprocessing

Data screening and structuring were the initial stage in preprocessing. To make sure the identified variables were associated with the desired input and output categories of the study, the raw dataset was screened. The numerical variables included C%, H%, N%, moisture, volatile matter, fixed carbon, ash, glucan, xylan, HHV, LHV, estimated lifecycle emissions were maintained to be analyzed whereas the categorical variable which is the feedstock type was retained to be grouped and later converted to numerical values to be used by the machine learning model.

The second step involved dealing with missing or irregular values. As the data were collected based on various biomass types and sources, there might be incomplete or non-homogenous observations.

The third step comprised of examination of outliers and extreme values. We are aware that biomass data tend to have unusually high or low values due to variations in feedstock type, processing history or source of measurement. These patterns were identified using descriptive statistics and visualization prior to training a model.

The fourth step entailed modeling readiness of the variables by scaling and encoding. Continuous variables were maintained in a numerical form to have them analyzed, and after that, the feedstock type variable was converted into numerical indicator variables so that it could become a machine learning algorithm input at the same time.

The last preprocessing activity was to pre-process the data to feature engineer and build the model. With cleaning and ordering done, working with the dataset was prepared to make derived variables including C/N ratio and fixed carbon-to-volatile matter ratio. This was done to enable engineered variables to enhance the performance of the models by model compositional relationships through measures that are more efficient when only represented with engineered variables. The preprocessing step thus gave a basis on the dependable features of engineering, AI-based prediction, and comparative analysis against empirical ground.

3.6 Feature Engineering

Derived variables used in biomass modeling research tend to be more informative than measured ones since they reflect compositional relationships directly associated with fuel quality, thermal behavior, and sustainability results. Three variables derived in this study were developed: carbon-to-nitrogen ratio (C/N ratio), fixed carbon-to-volatile matter ratio (FC/Volatile ratio) and ash-corrected HHV. These characteristics were chosen as they capture key biomass properties in a move that is better suited to predictive modeling (Šurić et al., 2023; Wang et al., 2022).

The C/N ratio was determined to indicate the carbon to nitrogen ratio in each feedstock. Although energy content is commonly related to carbon, the fuel composition and implications associated with the emissions are more relevant to nitrogen. The ratio had been obtained as:

$$\text{C/N ratio} = \frac{\mathrm{C}\ (\%)}{\mathrm{N}(\%)} \tag{1}$$

The proportion of carbon to volatile matter was estimated as fixed and represented the amount of solid carbon content in the biomass divided by the amount of volatile biomass content. Fixed carbon is linked with more steady combustion process whereas volatile matter is linked with ignition and release of combustible gases. This ratio was calculated as:

$$\text{FC/Volatile ratio} = \frac{\text{Fixed Carbon (\%)}}{\text{Volatile Matter (\%)}} \tag{2}$$

This characteristic was added as this is a characteristic of the structural balance of the biomass and tends to be associated with heating value and fuel behavior. Research on biomass characterization indicates that these ratios have the potential to enhance a better understanding of the model as they can discern a better view of the interaction between proximate properties than the variables by themselves (Wang et al., 2024).

The HHV was also adjusted to result in the ash-corrected value by the effects of ash on the effective energy content of biomass. The fraction of biomass that does not burn, which is called ash, is non-combustible and thus more ash in a biomass means less usable energy delivery. Ash content was used to correct HHV as follows to get a more realistic measure of energy:

$$\text{Ash-corrected HHV} = \frac{\mathrm{HHV}}{1-\frac{\mathrm{Ash}(\%)}{100}} \tag{3}$$

The correction can give an indication of heating value that is free of ash and can be useful in comparing feedstocks with varying mineral content on a more comparable basis. These corrected energy indicators are useful in biomass studies, as it is less prone to be skewed by non-combustible ash content and represents a more accurate picture of the intrinsic fuel quality of the organic fraction (Wang et al., 2024).

3.7 Lifecycle Emission Estimation

Lifetime greenhouse gas (GHG) emissions were calculated based on a feedstock-group emission-factor approach. The approach was chosen due to the ability to offer a viable approach to bridge the gap between the biomass feedstock properties and the environmental performance and maintain the emission performance between various biomass properties. Emission-factor mapping is widely applied in the literature on biomass and bioenergy where the goal is to compare feedstocks based on a shared sustainability foundation, in those cases where a complete process-based lifecycle inventory is not available to cover each sample (Cherubini et al., 2009; Wang et al., 2022).

They were given an emission factor according to the biomass group associated with each type of feedstock. Emission factor is the level of production of greenhouse gas that is produced per unit of bioenergy consumed expressed as kg CO_2-equivalent/MJ. By doing this, the predicted energy output of each sample could have been transformed into an estimated lifecycle emission value. The overall computation may be given as:

Estimated Lifecycle GHG Emissions = Emission Factor x Energy Output (4)

It also allowed comparative analysis of the AI-based framework and the empirical baseline, which is the focus of the current research. With the emissions fitted into the same model as HHV and LHV forecasting, the study could assess biomass feedstocks both energetically and environmentally.

3.8 Machine Learning Models

This research used machine learning models (supervised regression) to forecast higher heating value (HHV), lower heating value (LHV), and actual estimates of the lifecycle greenhouse gas emissions based on biomass feedstock characteristics. The models chosen were Support Vector Regression (SVR), Random Forest (RF), XGBoost, and Artificial Neural Network (ANN). The models tackled these issues because the problem of biomass prediction is nonlinear and multivariate and is also dependent on interactions among the elemental, proximate, and lignocellulosic variables (Cacho et al., 2023).

The overall machine learning prediction model can be presented as:

$$\hat{y} = f(x) \quad (5)$$

and where (X) denotes the matrix of predictor variables, (f) denotes the learning function that is estimated by using data.

In the case of Support Vector Regression, the optimization problem can be described as:

$$min\frac{1}{2} \parallel \omega \parallel^2 + \mathrm{C}\sum_{i=1}^{n}(\varepsilon_i + \varepsilon_i^*) \quad (6)$$

subjects to:

$$\mathrm{y_i} - (\omega^{\mathrm{T}}\emptyset(\mathrm{x_i}) + \mathrm{b}) \leq \varepsilon + \varepsilon_{\mathrm{i}}$$
$$(\omega^{\mathrm{T}}\emptyset(\mathrm{x_i}) + \mathrm{b}) - \mathrm{y_i} \leq \varepsilon + \varepsilon_{\mathrm{i}}^*$$
$$\varepsilon_i, \varepsilon_i^* \geq 0 \quad (7)$$

in which (w) is the weight vector, (C) is the penalty parameter, (ε) is the insensitive loss margin, ε(xi) = kernel-transformed input, and εi, ε_i^* introduced are slack variables. The biomass prediction is appropriate in estimation using SVR as it can model nonlinear relationships, particularly when the dataset is small (Wang et al., 2024).

In the case of Random Forest, the prediction will be achieved by averaging the results of several decision trees:

$$\hat{\mathrm{y}} = \frac{1}{\mathrm{T}}\sum_{\mathrm{t-1}}^{\mathrm{T}} \mathrm{h_t(x)} \quad (8)$$

ŷ (T) is the number of trees and ht(x) is prediction of the (t)-th tree. RF is also resilient to noise and multicollinearity and thus works better with non-homogeneous biomasses (Cacho et al., 2023).

In the case of XGBoost, the model is trained in stages, by the addition of weak learners to minimize prediction error:

$$\hat{y}_i^{(t)} = \hat{y}_i^{(t-1)} + nft(\mathrm{x_i}) \tag{9}$$

with n the learning rate and (ft(xi)) the new tree that has been added at iteration t. The whole objective function is:

$$Obj = \sum_{i=1}^{n} \iota\,(y_{i,}\hat{y}_i) + \quad \sum_{k=1}^{k} \Omega\,(fk) \tag{10}$$

where (ι) is loss function and (Omega (fk) is the regularization term. XGBoost finds broad application in biomass research due to its high effectiveness in highly structured data (Wang et al., 2022).

In the case of Artificial Neural Network, the result of every layer is calculated as:

$$\alpha^{(\iota)} = \sigma(w^{(\iota)}\alpha^{(\iota-1)} + b^{(\iota)}) \tag{11}$$

and is the activation of a single layer, (w(ι)), is the weight matrix, b is the bias vector and is the activation function used. In the case of the output layer, the prediction is provided by:

$$\hat{y} = w_2\sigma(w_1 x + b_1) + b_2 \tag{12}$$

ANN would be a good modeling tool to describe the properties of biomass energy due to the ability to focus the complex nonlinear interactions between the compositional variables (Šurić et al., 2023).

3.9 Empirical Baseline Calculation

The empirical underline was created and used as a traditional reference point of the comparison of the AI-based predictions. The traditional method of biomass assessment, where outputs are determined based not on learned data but on simple equations and empirical relationships based on the specific emission factor, is this baseline. Including this baseline was aimed at establishing whether AI-based framework can offer a better predictive accuracy and a more realistic sustainability assessment than traditional methods of calculations.

Conventionally, in the case of energy properties, the empirical baseline was given in the form of a standardized estimation function of the feedstock variables:

$$\hat{y}_{emp} = f(X) \tag{13}$$

empirical estimation of the target variable ($\hat{y}_{emp}$), (X) is the biomass feedstock properties. The empirical one was applied pratically as the standard non-AI comparison of the machine learning estimations of HHV and LHV. Such simplified estimation logic is prevalent in conventional biomass research since it is straightforward and simple to implement when one does not have a complete predictive framework (Cherubini et al., 2009).

On the environmental aspect, an empirical baseline had an approach of standard emission-factor. The greenhouse gas emission per part of the lifecycle was estimated to be:

$$\mathrm{GHG}_{emp} = EF\ X\ E \tag{14}$$

The empirical estimate of the lifecycle GHG emission is (GHG_{emp}), the emission factor of that feedstock group is (EF), and the biomass energy is (E) which could be either HHV or

LHV, depending on the comparison made. The approach is common in the biomass literature, as it offers a simple approach of estimating the emissions of various categories of feedstock, as well as comparing them on a level playing field (Cherubini et al., 2009; Martins et al., 2019).

3.10 Model Training and Validation

The different algorithms were trained on the chosen input features as elemental composition, proximate analysis, lignocellulosic composition, and the type of feedstock. The training phase sought to reduce the prediction error through an attempt to tune the model parameters to the most appropriate level of prediction. There were distinct models built of each of the target variables in such a way that the predictive behavior of the algorithms could be individually discussed in regard to HHV and LHV as well as lifecycle emissions. Normal performance was measured by normal regression parameters. The coefficient of determination was obtained:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

where (y_i) denotes the observed value, $(\hat{y}_i)$ denotes the predicted value and $(\bar{y})$ denotes the meaning of the observed values.

The mean squared error (MSE) is given as:

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

and the root mean squared error (RMSE) was formed as:

$$\text{RMSE} = \sqrt{MSE}$$

The predictive accuracy of SVR, RF, XGBoost, and ANN was compared using these metrics. An increased value of (R^2) and decreased value of MSE and RMSE showed a better performance of the model. Validation confirmed that SVR performed optimally overall with HHV and LHV prediction, which warrants its appropriateness with respect to modeling nonlinear relationships in the data.

3.11 Sensitivity Analysis

SHAP was analyzed to estimate the contribution of each input variable to the model output. The SHAP model breaks down the prediction into a base value and addition of feature contributions. SHAP was utilized to look at both how significant the variables are on a global scale over the dataset and the contribution of variables on single predictions. This method proved to be of great assistance in determining the effect of carbon, hydrogen, lignin, xylan, moisture, ash, fixed carbon, and engineered ratios on the bioenergy and emission performance (Wang et al., 2024).

Alternative methods of interpretation were also used, specifically permutation importance. Under this strategy, the significance of a variable can be calculated through random permutation of values and the consequent decrease in model performance. The general permutation importance can be defined as.

Sensitivity analysis was conducted whereby the alterations affected the HHV prediction when small manipulations were made in the most significant feedstock variables and lifecycle emissions were determined. In the experiment, the most important variables were varied by 5 percent and changes in the output values were measured against the prediction at a baseline.

3.12 Comparative AI-Based and Empirical Framework

The aim of this comparison was to identify whether machine-learning models are more accurate and informative estimators. Comparative analysis plays a role in biomass studies since empirical techniques continue to be prevalent as a starting point in this field, yet they may be constrained by simplifying assumptions and less flexibility when allowed to operate on heterogeneous feedstock data (Cherubini et al., 2009; Wang et al., 2024).

Under the empirical branch of the framework, the properties of biomass were employed to determine energy and emission performance via standard equations and mapping of the emission factors. The difference corresponds to the larger body of literature, where it is demonstrated that empirical and AI-based methods are applicable to different situations, respectively, namely, that is, to a screening of the basis or to a more complex nonlinear system with a multitude of interacting biomass properties (Šurić et al., 2023). Figure 2 a comparative framework of an AI-based method and an empirical baseline method in assessing the bioenergy feedstocks and their lifecycle emissions.

It demonstrates that this study is not purely prediction oriented, but also comparative, as it compares a data-driven AI approach with a traditional calculation-based approach, to find out which can provide more accurate answers to the question and make a more reliable assessment of sustainability.

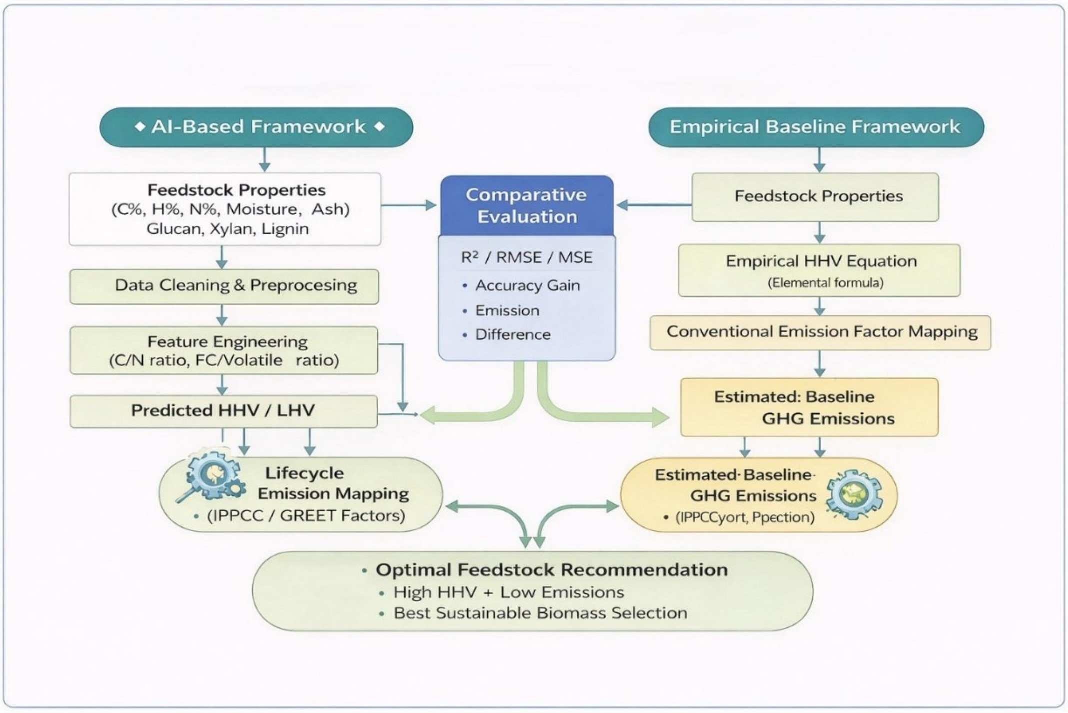

Figure 2. Comparative AI-based and Empirical Framework for bioenergy lifecycle emission assessment

On the AI-based framework side, the feedstock properties of C%, H, N, moisture, ash, glucan, xylan and lignin are taken. These input variables are first cleaned and preprocessed later transformed via the means of feature engineering, with such ratios as C/N ratio and fixed carbon to volatile matter ratio. Subsequently, this is followed by the prediction of HHV and LHV, which is an energy potential of biomass. This network has shown that AI framework ought to learn complicated relationships right based on feedstock composition and then transform them into energy predictions. The AI-powered pathway is then mapped through lifecycle emission to conventional parameters of IPCC and GREET (IPCC, 2023; Long & Liu, 2023). This implies that the framework has the potential to emit more greenhouse gases in addition to forecasting the energy properties of biomass once it has predicted the energy properties. By so doing, the AI system will interconnect feedstock quality, energy performance, and environmental impact into a single pipeline.

The reason is to find out about feedstocks with a high amount of energy and low emissions. Under the empirical baseline framework, the identical properties of feedstock are utilized, yet the procedure of prediction is based on a more conventional approach. In this case, an empirical expression of HHV is initially applied to feedstock characteristics then customary emission factor mapping is applied to estimate a baseline level of GHG emissions. This method is the ordinary, non-AI technique that is deployed in biomass and lifecycle assessment research works. It is a point with the help of which the AI-based framework can be evaluated.

Comparative evaluation is the most significant analysis section of the structure we followed (found in the center of the diagram). Measures like R 2, RMSE and MSE are used to compare the AI -based results and the empirical baseline. This comparison is required to compare the accuracy gains and differences in emission and estimate the difference to the traditional empirical calculation to argue whether AI could better perform energy prediction and lifecycle emissions assessment compared to conventional empirical calculation.

The final output of the decision-making is the optimum recommendation on feedstock at the bottom of the framework. This is the applied aim of this work. The framework finds the biomass feedstock that offers a desirable balance of high-energy value and low-environmental burden by summing the predicted HHV, LHV, and estimated emissions.

4. Results and Discussion

4.1 Descriptive Statistics

The descriptive statistics give an overall idea about the biomass feedstock characteristics and instruments of interest in the current study. According to table 1 there is a significant deviation among the feedstock properties that are diverse enough to allow machine learning analysis. This variability is critical since the chemical compositions, structural compositions and energy qualities of biomass feedstocks vary significantly, and all these factors will affect the performance of bioenergy and life-cycle emissions.

Table 2. Descriptive Statistics

	mean	**median**	**std**	**Min**	**max**
C	51.621	47.619	12.963	2.456	83.567
H	5.759	6.292	1.519	0.078	8.057
N	0.314	0.241	0.272	0	1.036
Moisture	6.011	5.137	4.741	0.057	31.183
Volatile	66.117	77.555	23.089	4.783	83.373
Fixed C	23.850	14.5	19.532	4.2	76.127
Ash proximate	4.029	0.963	11.559	0	90.963
Glucan	27.999	29.059	10.161	0	40.000
Xylan	7.718	7.143	4.025	0	22.015
Lignin	38.032	35.905	16.066	16.439	99.799
FC to Volatile ratio	0.724	0.190	1.158	0.135	4.030
HHV MJkg	20.362	20.36	2.818	16.169	32.942
LHV MJkg	18.428	18.43	2.152	14.825	32.018
Ash corrected HHV	23.989	20.443	24.638	17.232	219.490

In table 2, the elemental variables show the mean value of carbon content is 51.62 percent and median is 47.62 percent, showing a moderately skewed distribution, with some carbon-rich feedstocks at high end of the range. Hydrogen content is constant, with a mean of 5.76% and a median of 6.29, and the nitrogen content is low, with a mean of 0.31, which can be compared to normal lignocellulosic biomass. The less nitrogen content implies that most of the feedstocks in the data can be used in biomass energy production without excessive worries about nitrogen.

There is also evident variation in the proximate analysis variables across the feedstocks. The mean moisture content is 6.01, but the highest values are more than 31, which implies that some of the samples have a lot more water in them. The mean and median of volatile matter is 66.12 and 77.56 respectively and the mean and median of fixed carbon stands at 23.85 and 14.5 respectively, show a great variance in thermal and combustive properties. The mean value of the ash content is 4.03 although the maximum value is unusually large, which implies that it has extreme observations or unusually mineral rich feedstocks.

The lignocellulosic make-up too exhibits a significant diversity. The mean glucan is 28.00, the mean xylan is 7.72 and lignin is 38.03 having shown that the structural composition of the biomass samples is quite different. Since lignin is typically related to a greater heating value, and glucan and xylan define carbohydrate fractions, these variations are likely to affect both the HHV and LHV. Lignocellulosic fractions vary widely, which justifies applying machine learning models, as this type of model particularly has the advantage of being designed to manage nonlinear interactions between structural biomass components and energy output. There is also significant dispersion in the target variables. HHV mean value is 20.36 MJ/kg and LHV mean value is 18.43 MJ/kg and this means that most of the feedstocks would lie in a realistic energy range where bioenergy could be used. The distribution spread in these values indicates that the data have both low-energy uncooked residues and high-energy processed or carbon-intensive feeds. The estimates of lifecycle emissions also differ among the feedstocks, indicating that the performance of the biomass samples in the environment is not homogenous. This diversity is significant as it enables the investigation to not only explore the potential of bioenergy but also study the sustainability profile of various feedstock types.

4.2 Distribution of Biomass Feedstock Properties

The biomass is compositionally varied, and this is a key requirement in the effort of developing powerful predictive models. Figure 3 shows that the bulk of the carbon content is concentrated with the values of about 45 -52 with some samples going much higher. It is skewed to the right and it can be explained by the fact that feedstocks contain modest amounts of carbon; however, there are few samples rich in carbon, and are placed at the right end of the scale. This trend is significant since carbon is a significant component of biomass energy, and the availability of high-carbon feedstocks is likely to affect both HHV and lifecycle emissions.

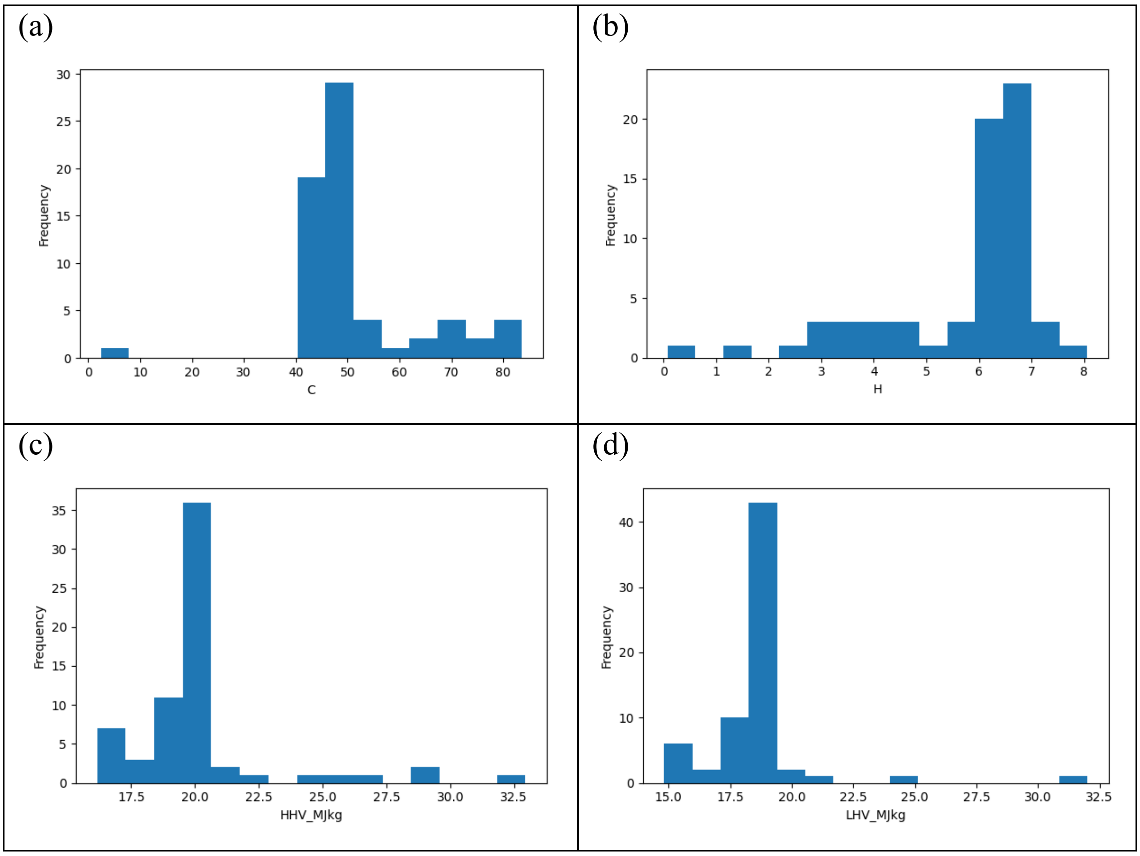

Figure 3. Distribution of (a) C, (b) H, (c) HHV Mjkg, (d) LHV Mjkg

Hydrogen is more stable than carbon. Most of the observations are concentrated between 6% and 7%, the distribution is biased slightly towards the left, since few samples are in the lower range of hydrogen. This restricted dispersion implies that hydrogen is not as dispersant as carbon is on the feedstocks. Though hydrogen is still relevant as an elemental variable, its variability is less than most of the other elemental variables, suggesting that it could be acting in a sub-ordinate role to carbon-rich indicators such as fixed carbon and lignin.

The higher heating value (HHV) is concentrated between 19.5 and 21 MJ/kg, indicating that most biomass samples are similar at a moderate range of energy. The distribution is skewed to the right since a few feedstocks achieve much higher values of HHV to the point of about 25 to 33 MJ/kg. These observations with high energy are related to char-based or otherwise carbon-rich materials, which are also expected to deliver higher quality fuel. The distribution thus proves that the feedstock type and composition have an extraordinarily strong effect on heating value.

4.2.1 Higher heating value (HHV) MJkg vs Carbon (C)

Figure 4 indicates that there is less a positive relationship between carbon content (C%) and HHV, that is, higher carbon content is usually related to a higher heating value. But it is not strictly linear, since most of the samples fall in the range C = 42 -52 percent and HHV = 16-21 MJ/kg, with small groups of samples with high carbon concentration, around 70 -80 percent, having much higher values of HHV than their counterparts, up to around 25-33 MJ/kg.

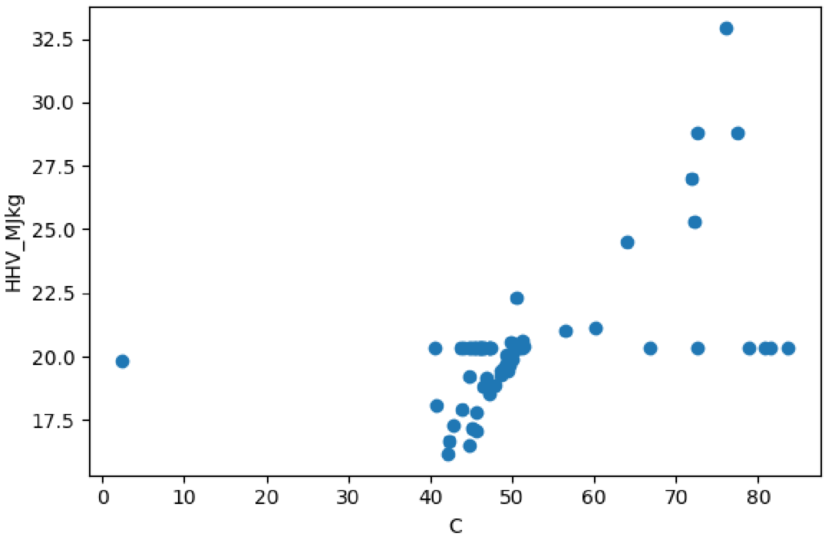

Figure 4. Higher heating value (HHV) MJkg vs Carbon (C)

4.2.2 Higher heating value (HHV) MJkg vs Hydrogen (H)

Figure 5 indicates that there is no significant simple linear increase in HHV with hydrogen (H%) present. Majority of the samples have a high concentration around H = 67% and HHV = 18.520.5 MJ/kg, indicating that most feedstocks have equal amounts of hydrogen and equal heating values. Some of the points with lower hydrogen content, say 1-4 percent, are given much higher values of HHV with a range of about 25-33 MJ/kg. This means that hydrogen per se is not strong. This implies that the joint effect of a combination of various feedstock properties, including carbon, fixed carbon, and ash, volatile matter, lignin, and structural composition, causes a greater effect on HHV than the impact of hydrogen content alone.

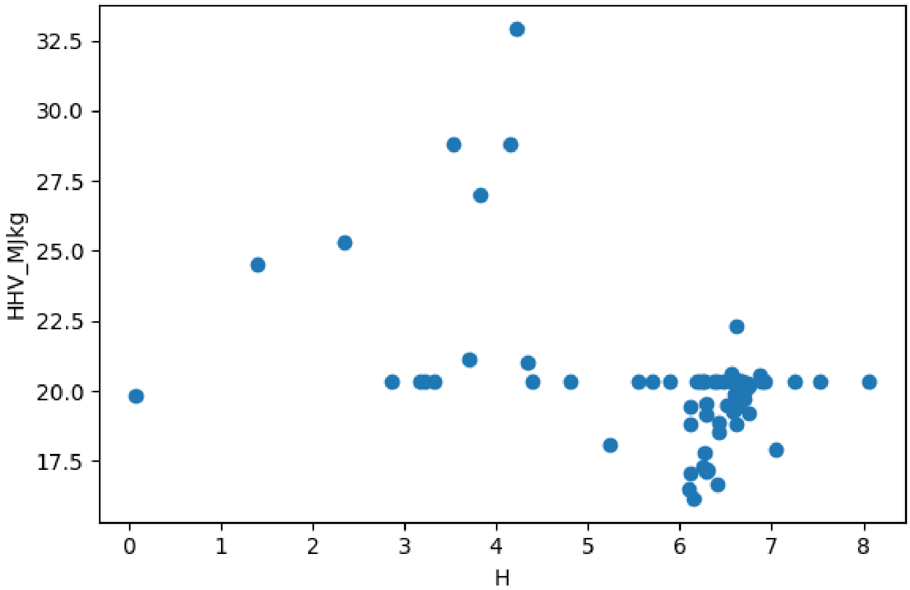

Figure 5. Higher heating value (HHV) MJkg vs hydrogen (H)

4.2.3 Higher heating value (HHV) MJkg vs Lignin

Figure 6 indicates that there is no strong linear relation of HHV to lignin content through the entire data, however the major cluster indicates the existence of a liaison between middle lignin amounts and middle HHV amounts. The main samples are clumped in the range of lignin = 38% and HHV = 18-21 MJ/kg, which means that most feedstocks can belong to a similar range of compositions and energy. The mean location of this central cluster raises an inference that lignin is an influential characteristic, although not the sole one that dictates heating value. Some few lignin observations at 35 -40 percent indicate a much bigger HHV, of about 24-33 MJ/kg.

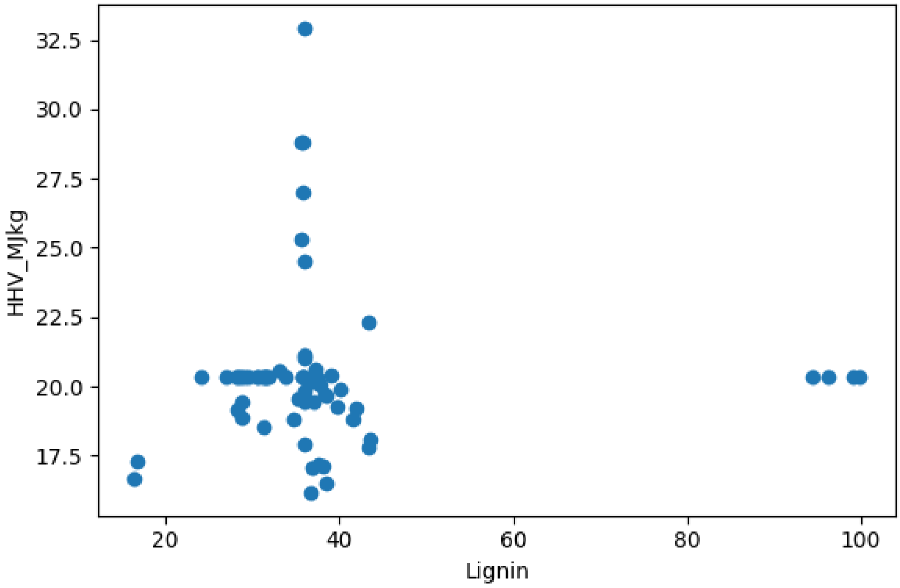

Figure 6. Higher heating value (HHV) MJkg vs Lignin

4.3 Correlation Analysis

The elemental and proximate variables depicting a positive correlation with HHV with a correlation of about 0.91 have the fixed carbon variable and the other ones in figure 7 with lower fixed carbon content have a greater heating value. There is also a close positive relationship between carbon and fixed carbon where carbon is approximately 0.89 but this is not surprising since these two variables indicate the portion of biomass with a substantial proportion of carbon. Fixed carbon is also highly correlated with hydrogen with a coefficient of approximately 0.92, although because of the negative relationship between hydrogen and HHV, hydrogen is not structurally associated with the makeup of the feedstocks like carbon-rich fractions.

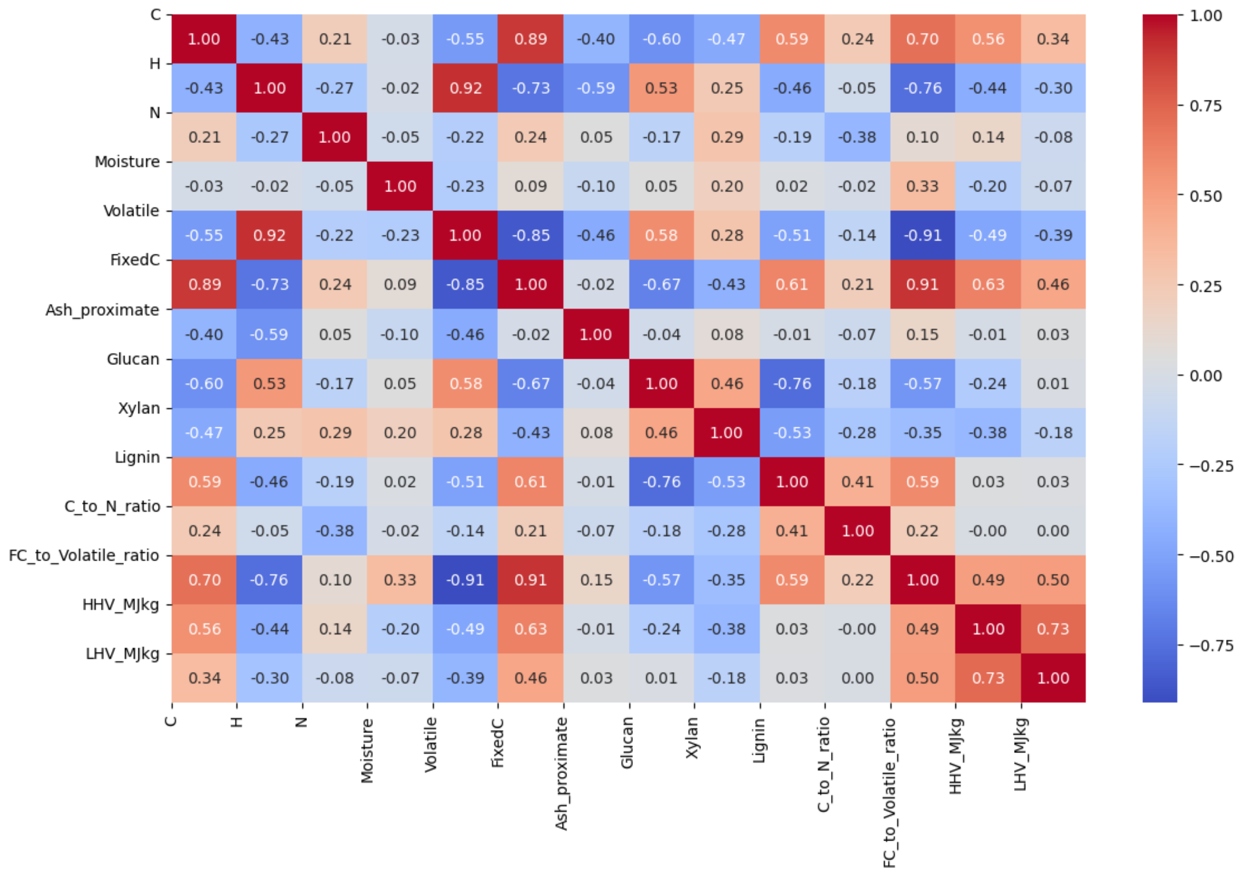

Figure 7. Correlation among the biomass composition variables

Conversely, figure 7, volatile matter indicates a remarkably close negative correlation between Fc volatile ratio and volatile matter, with an approximate proportion of -0.91, implying that the more volatile matter is found in feedstocks, the lower the heating value in such data. Hydrogen and HHV have a negative correlation as well; that is, it is approximately -0.76 but moisture has a weaker correlation. These trends demonstrate that feedstocks with greater volatile fractions and higher moisture have lower usable energy content, with feedstocks containing greater fixed carbon content being more energy dense. The inverse nature of the volatile matter and the HHV is especially significant since it demonstrates a significant impact of proximate analysis variables on biomass energy performance.

There are significant associations between the lignocellulosic and HHV and LHV. The relationship between Lignin and HHV is moderate and positive (around 0.59) as it was expected that the biomass rich in lignin has greater energy potential. Xylan exhibits a less consistent and sharply fluctuating correlation with the energy outputs, although the correlation between Xylan and the C/N ratio as well as fixed carbon is extremely negative, implying that the structure of carbohydrates in the feedstock interacts with the overall energy balance of the feedstock. These correlations demonstrate that lignocellulosic composition does not only plays a role in prediction of energy, but it is more indirect and relies more on the relationship with the elemental and proximate variables.

The association between HHV and LHV is also positive and significant, and it is approximately 0.73. This is agreeable since the two variables are indications of biomass energy content, and they are affected by alike feedstock characteristics. Nonetheless, the correlation is not ideal which implies that HHV and LHV possess different information and they are to be considered as separate target variables in the model. Their distinction is noteworthy since LHV indicates the amount of usable energy calculated by subtracting the energy losses associated with moisture, and HHV is the total energy that can be combusted. Weak or less precise relationships are manifested in some features. There is an average, positive correlation between ash content and HHV and LHV in the matrix, but this could have been caused by the other feedstock features. There are relatively weak correlations between feedstock type indicators and the energy yields, indicating that the categorical nature of the biomass is not as informative as the chemical and structural nature of the biomass. This strengthens the usefulness of composition-based predictors over using feedstock labels solely.

4.4 Feature Engineering

About three quarters of the feedstocks in table 3 contain a lot of carbon relative to nitrogen, as seen by the C/N ratio being exceedingly high values. The advantage of this is that it measures this equilibrium between energy carrying carbon and low nitrogen, and it might be applicable both to characterize energy and provide sustainability. The

C/N ratio is also heterogeneously distributed, as indicated by the widespread, which indicates the heterogeneity of the biomass samples and supports.

The values of the FC-to-volatile ratio are more moderate and variation in the ratio of solid combustible fraction and the volatile portion of the biomass in table 3. This ratio is significant as it is an argument of thermal behavior and quality of the fuel in scaled form. Feedstocks containing a substantial proportion of fixed carbon as compared to volatile matter are usually characterized by more predictable energy behavior, and as such this characteristic can be of value when predicting bioenergy.

Table 3. Feature Engineering

C to N ratio	FC to Volatile ratio	Ash corrected HHV
199.774	0.161	19.031
146.976	0.248	22.474
207.0449	0.176	19.129
73.656	0.191	19.639
212.022	0.174	19.474

To account the heating value as affected by the ash content, the ash-corrected HHV was added. As ash is non-combustible matter, the corrected figure makes the best approximation to the available energy potential. The resultant figures indicated that the values of the adjusted heating were mostly elevated compared to the crude estimated HHV values, hence indicating that ash adjustment is applicable when comparing feedstocks with varying mineral composition.

4.5 Model Performance for HHV and LHV Prediction

The prediction of higher heating value (HHV) indicates that there was a significant difference in the capacity of the machine learning models to model the full dependence on feedstock properties and energy content. The Support Vector Regression (SVR) using base feature set model exhibited the highest test R^2, test RMSE of 1.0673 and test MSE of 1.1390 to table 4. It means that the SVR model was able to account for a significant amount of the variance in HHV and yield small errors. The finding validates that HHV can be forecasted well based on the biomass composition variables that become available when the nonlinear learning approach is employed.

The models comparison also indicates that the base plus derived feature set performed poorly when compared to the base feature set in HHV prediction. However, feature engineering did not enhance the accuracy of HHV prediction in this instance despite its role in enhancing interpretability. XGBoost using the base feature set achieved the second-best result of 0.7000 on the test R^2, meaning that it achieved acceptable performance in predicting but lower than the SVR. Random Forest, Artificial Neural Network and linear regression, on the other hand, demonstrated poorer results.

Table 4. Model validation summary

Target	Feature_Set	Model	R2	RMSE	MSE	CV R2 Mean	CV RMSE Mean
HHV_MJkg	Base	SVR	0.7601	1.0673	1.1390	0.7414	1.3774
HHV_MJkg	Base	XGBoost	0.7000	1.1936	1.4246	0.6742	1.5302
HHV_MJkg	Base+Derived	SVR	0.6176	1.3476	1.8161	0.7183	1.4209
HHV_MJkg	Base+Derived	XGBoost	0.5151	1.5174	2.3025	0.4922	1.9203
HHV_MJkg	Base+Derived	ANN	0.1152	2.0498	4.2015	-0.0392	2.7487
HHV_MJkg	Base	ANN	-0.0325	2.2143	4.9030	-0.0650	2.7768
HHV_MJkg	Base+Derived	RandomForest	-0.2504	2.4367	5.9374	0.3484	2.1124
HHV_MJkg	Base	RandomForest	-0.2649	2.4508	6.0066	0.3573	2.0984
HHV_MJkg	Base	LinearRegression	-54.6875	16.2616	264.4391	-10.9145	5.4935
HHV_MJkg	Base+Derived	LinearRegression	-68.2084	18.1286	328.6448	-13.8227	6.1260
LHV_MJkg	Base+Derived	SVR	0.2170	0.6768	0.4581	0.2256	1.5483
LHV_MJkg	Base	SVR	0.2129	0.6786	0.4605	0.2101	1.5649
LHV_MJkg	Base+Derived	XGBoost	0.1590	0.7015	0.4921	0.2037	1.5864
LHV_MJkg	Base	XGBoost	-0.3453	0.8872	0.7871	0.0516	1.6543
LHV_MJkg	Base	ANN	-0.8206	1.0321	1.0652	-0.6094	1.9806
LHV_MJkg	Base+Derived	ANN	-0.8731	1.0469	1.0959	-0.6171	1.9819
LHV_MJkg	Base+Derived	RandomForest	-2.6756	1.4665	2.1506	-0.3850	1.7088
LHV_MJkg	Base	RandomForest	-3.1109	1.5509	2.4053	-0.3025	1.6796
LHV_MJkg	Base	LinearRegression	-322.1209	13.7498	189.0573	-69.4536	5.5958
LHV_MJkg	Base+Derived	LinearRegression	-422.1300	15.7344	247.5724	-90.6112	6.3771

Similarly, prediction of lower heating value (LHV) reveal that Support Vector Regression (SVR) showed optimal performance in terms of overall prediction, but the predictive capability was low. The optimal performance was found with SVR when using the base-plus-derived feature set, registering a test R^2 equal to 0.2170, RMSE equal to 0.6768 and MSE equal to 0.4581. The identical base-only SVR model gave a similar outcome with a test R^2 of 0.2129 and it means the addition of engineered variables to the model did not alter the predictive accuracy, although it marginally improved the outcome.

4.5.1 Actual vs Predicted HHV and LHV

It is in Figure 8 that the actual and predicted HHV are compared as it is worked out with the SVR model. The points are close to the diagonal reference line, and this suggests that the model can be used to reproduce the overall trend of the true HHV values with a fair degree of accuracy. Most of the observations lie near the mid-range values and in such samples, the observed and predicted values are near implying good consistency of the model in the central region. Meanwhile, some of the points do not follow the (horizontal) diagonal line, particularly at lower and higher ends of the range of HHV. The deviations were indicative of the fact that the model may be overestimating few individual samples, especially those ones, which might reflect some more odd feedstock characteristics. The general trend remains a good correlation between the real and forecasted values.

Conversely, the most points on the LHV curve are located rather close to the 45-degree reference line, which means that the model is able to explain the overall direction of LHV quite well. Most of the observations can be concentrated in the central range and predictions here are not very far apart to the actual values indicating a reliable performance of the models. Meanwhile, a few deviations of the diagonal line are visible. Others are slightly overpredicted whereas others are underpredicted, particularly at the mid-range values between 18.4 to 18.9 MJ/kg. This implies that not all the samples can be well predicted by the SVR and some samples of feedstock with varying composition properties may be more difficult to capture well by the model. It is also still moderate in its spread that is not as tight as one would have expected in a highly precise model.

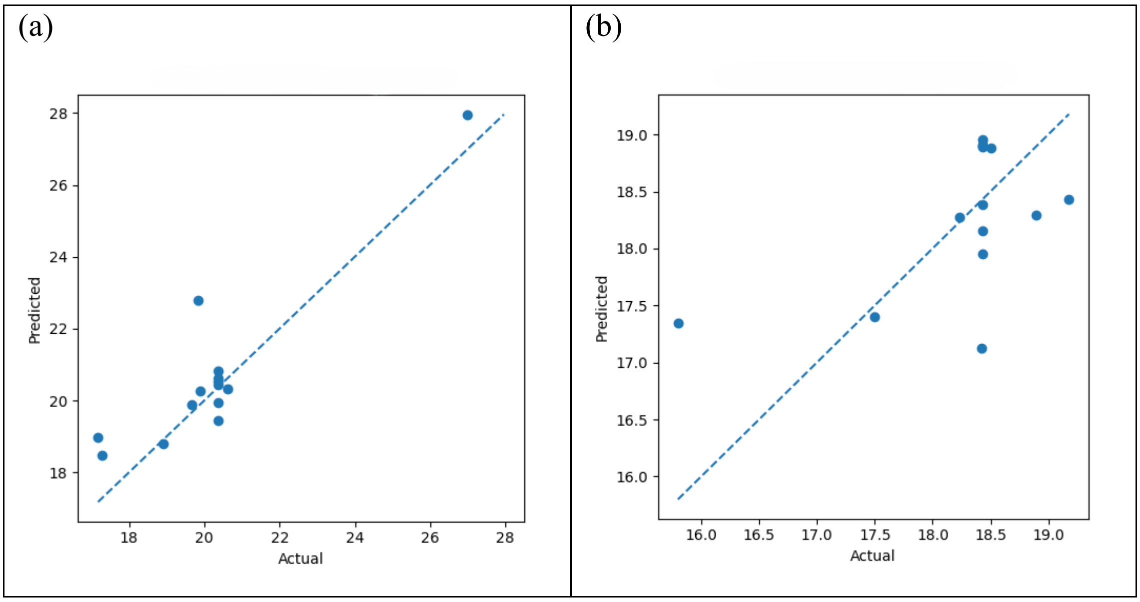

Figure 8. Actual vs Predicted (a) HHV and (b) LHV

4.6 Feedstock group emission factors

The table 5 feedstock group emission factors that are employed to calculate the lifecycle greenhouse gas emissions of several types of biomasses. All feedstock types belong to a larger group, and all groups would correspond to a particular emission factor in kg CO_2-eq per MJ. This forms the foundation for comparing the feedstocks on an equivalent environmental scale and allowing the step of lifecycle emission mapping in the thesis outline.

Table 5. Feedstock group emission factors

Type	Feedstock group	kgCO2eq_per_MJ
Agricultural waste, almond husk	Agricultural residue	0.012
Corn stover	Agricultural residue	0.012
Wood pellet ash	Ash-rich residue	0.02
Agricultural waste, almond husk char	Biochar/char	0.008
Corn pellet char	Biochar/char	0.008
Corn stover char	Biochar/char	0.008
Mixed wood forest residue char	Biochar/char	0.008
Mixed wood urban waste char	Biochar/char	0.008
Softwood char	Biochar/char	0.008
Wood pellet char	Biochar/char	0.008
Burned Jeffrey Pine wood	Burned/treated woody residue	0.011
Construction and Demolition Waste	Construction/Demolition waste	0.014
Jeffrey pine chips forest residue	Forest residue	0.009
Jeffrey pine slash forest residue	Forest residue	0.009
Lodgepole pine chips forest residue	Forest residue	0.009
Lodgepole slash forest residue	Forest residue	0.009
Mixed wood forest residue	Forest residue	0.009
Softwood forest residue	Forest residue	0.009
Corn stover pellets	Pelletized biomass	0.013
Wood pellet sourced outside of US	Pelletized biomass	0.013
Mixed wood bark urban waste	Urban wood/green waste	0.015
Mixed wood urban waste	Urban wood/green waste	0.015
Urban green waste	Urban wood/green waste	0.015
Agricultural waste, wood	Woody biomass	0.01
Screened softwood chips	Woody biomass	0.01

Woody biomass is the least emitting including agricultural waste wood and screened softwood chips with an emission factor of 0.01 kg CO_2 -eq/MJ. This implies that these feedstocks are viewed as comparatively low-emission feedstocks. The values of forest residues are also low 0.009 kg CO_2-eq/MJ, which means that the ecological performance is also quite good. Even lower is the factor of 0.008 kg CO_2 -eq/MJ of biochar/char materials, which indicates that such substances may be associated with a lower burden of performance in the lifecycle, particularly compared to raw residues, or processed biomass forms.

Granted, feedstocks that have been more processed (or those handling conditions are poorer) have higher emission factors. Farm waste has a rating of 0.012kg of CO_2-eq/MJ, whereas pelletized biomass generates 0.013 kg CO_2-eq/MJ. Urban wood/green waste has a factor of 0.015 kg CO_2 -equivalent/ MJ whereas construction/demolition waste has 0.014 kg CO_2 -equivalent/ MJ. The largest figure in the table 4 is ash rich residue, namely, wood pellet ash, of 0.02 kg CO_2-eq per MJ, the highest ranked among the listed residues.

4.7 Comparison of AI-Based and Empirical HHV and LHV Estimates

The discussion of the higher heating value HHV and LHV estimates based on AI and empirical data shows that there is an obvious advantage of the AI-based framework. The empirical baseline was quite poor with the R^2 value of -3.479 with HHV; and with LHV the R^2 value was negative and equal to -8.270 which shows that the traditional equation is explaining the values in HHV worse than a simple straight mean-based forecast. Moreover, the HHV and LHV Empirical model had RMSE of 5.916 and MSE of 35.029 and 6.503 and 42.289 respectively signifying that errors in prediction were high and the HHV and LHV were not using the appropriate method in table 6.

Table 6. HHV and LHV empirical vs AI comparison

Method		**R2**	**RMSE**	**MSE**
AI-based HHV	HHV	0.968	0.498	0.248
Empirical HHV baseline		-3.479	5.918	35.029
AI-based LHV	LHV	0.309	1.776	3.154

Empirical LHV baseline		-8.270	6.503	42.289

Conversely, HHV and LHV models based on AI demonstrated a good predictive performance with R^2 of 0.968 and 0.309, RMSE of 0.498 and 1.776, and MSE of 0.248 and 3.154 respectively in table 6. These findings indicate that the AI model managed to capture all the variability in HHV and LHV and gave predictions whose value were nearly identical to the observed values. The empirical enhancements over empirical baseline are dramatic evidence of the fact that the biomass-energy correlation is non-linear and the data-driven model best describes the relationship as compared to a traditional formula. Therefore, the AI-driven construct provides a far more accurate approach to HHV and LHV estimation as compared to the empirical baseline. The superior performance of the AI model proves that there is enough information in feedstock composition to get a good forecast of heating value by considering nonlinear relationship well modeled.

4.8 Comparison of AI-Based and Empirical Lifecycle Emission Estimates

Results of AI-based and empirical lifecycle greenhouse gas emissions estimates indicate that the AI-based approach gives a more consistent and accurate estimate of the greenhouse gas emissions from different feedstocks. In most cases, the empirical emissions estimates show a higher emission load than the AI-based estimates, suggesting that the empirical lifecycle analysis approach may overestimate the emission burden compared to the data-driven, AI-based method. This insight suggests that the AI-based model is more capable of capturing the differences between biomass feedstocks (in terms of their composition) and of capturing impact differences in terms of emissions. We also find that the difference between the two approaches varies among feedstocks. The absolute difference between the empirical and AI-based emissions is small for several char-based feedstocks such as corn stover char, corn pellet char, mixed wood urban waste char, and mixed wood forest residue char. This suggests that the AI and empirical approaches are similar for the feedstocks with more homogeneous, carbon-rich characteristics. However, feedstocks like wood pellet ash display an exceedingly high difference, which implies that empirical methods are not accurate for feedstocks that have extreme characteristics and have been processed or manufactured.

The feedstocks with higher energy efficiency and lower emission intensity, such as char and woody feedstocks, continue to be considered as better feedstocks on the AI-based calculation. In parallel, ash- and urban waste-based feedstocks still have elevated lifecycle emissions, suggesting that feedstock properties have a key role in sustainability. But the AI-based method is more tempered and does not give the extreme results of the empirical analysis. Thus, the approach based on AI methods provides a better foundation for comparing feedstocks and assessing feedstocks with the most promising balance between bioenergy potential and environmental sustainability.

4.8.1 Estimated Lifecycle Emissions by Feedstock Type

Emissions from the estimated lifecycle greenhouse gas emissions vary between feedstock types, but the majority of feedstocks have values in a relatively narrow range of around 0.16 - 0.29 kg CO_2-eq/kg biomass, which implies that the estimated emissions are relatively similar for most feedstocks, with only a few higher and some moderate values in figure 9. These results indeed suggest that the type of feedstock has a noticeable impact on estimated lifecycle emissions, but that the variation is not as extreme as for all the feedstocks. The highest estimated emission value is seen for wood pellet ash with a value of approx. 0.45 kg CO_2-eq/kg biomass that stands out clearly from the rest. This could point to the fact that this feedstock has a much higher value in the applied emission-factor approach, perhaps due to intense processing, treatment of residues, or the mapping of the emission factor to this group. Other notable high-emission feedstocks include mixed wood bark urban waste, construction and demolition waste, mixed wood urban waste, corn stover char and mixed wood forest residue, which are all close to, or above the cluster upper bound.

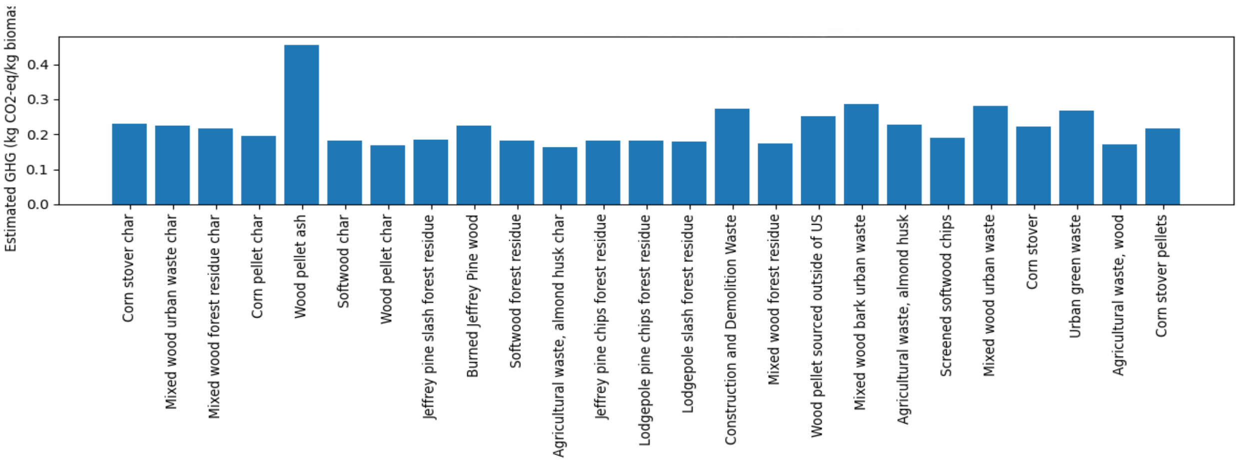

Figure 9. Estimated Lifecycle Emissions by Feedstock Type

Other feedstocks, such as wood pellet char, softwood char, mixed wood forest residue char and some of the residue-based feedstocks, are more in the middle of the distribution. This could be an indication that different classes of char-based and residue-based feedstocks do not have a consistent effect on lifecycle emissions and that the environmental performance of biomass is highly dependent on the type of feedstock and the applied emission factor. Agricultural and forest residues have lower emissions on average than the processed pellet and ash-fired feedstocks.

4.9 Feature Importance and SHAP analysis

In predicting HHV, the SHAP analysis revealed that hydrogen (H), fixed carbon, xylan, carbon (C) and moisture are the top contributors. The SHAP summary plot showed that fixed carbon and carbon demonstrated a positive influence on the predicted HHV, as would be expected from biomass that contains more carbon. On the other hand, increasing moisture appeared to decrease HHV, which is consistent with the negative influence of water on the fuel energy. Hydrogen also had an appreciable effect but was not as clearly related as the carbon-rich features. The feature that significantly influenced predicted LHV was the fixed carbon-to-volatiles ratio, followed by xylan, ash (proximate), volatile matter, fixed carbon, carbon, moisture, and C/N ratio.

4.9.1 HHV

Importance of the various features in predicting HHV. The strength of each input feature averaged over all trees is shown in decreasing order (top to bottom). In this example, H, FixedC, Xylan, C, and Moisture seem to be the most key features. The direction of influence is shown in Figure 10. For instance, higher values of FixedC and C tend to increase the predicted HHV, because there are more red dots on the positive side of the SHAP value. This suggests that feedstocks with greater fixed carbon content and C are likely to have a higher heating value. On the other hand, higher values of H and Moisture are likely to push down the predicted value of HHV because the red points for these variables are mostly located on the negative SHAP side.

Finally, Xylan has a mixed pattern, which would indicate a nonlinear interaction with the HHV. Some values of xylan increase the prediction, and some decrease it. This suggests that xylan might be affecting the prediction indirectly, through its interaction with other structural variables like lignin and glucan. Lignin also has a lesser influence with both negative and positive SHAP values, meaning that while lignin has a small influence, it is less influential than fixed carbon or carbon content. The feedstock-type variable is towards the bottom of figure 10 and is therefore less important in the model than the chemical measures. Their SHAP values are near zero, which means that feedstock category makes a smaller contribution to the prediction of HHV than the composition-based variables. This means the AI model is learning more from the biomass properties, rather than from the feedstock type.

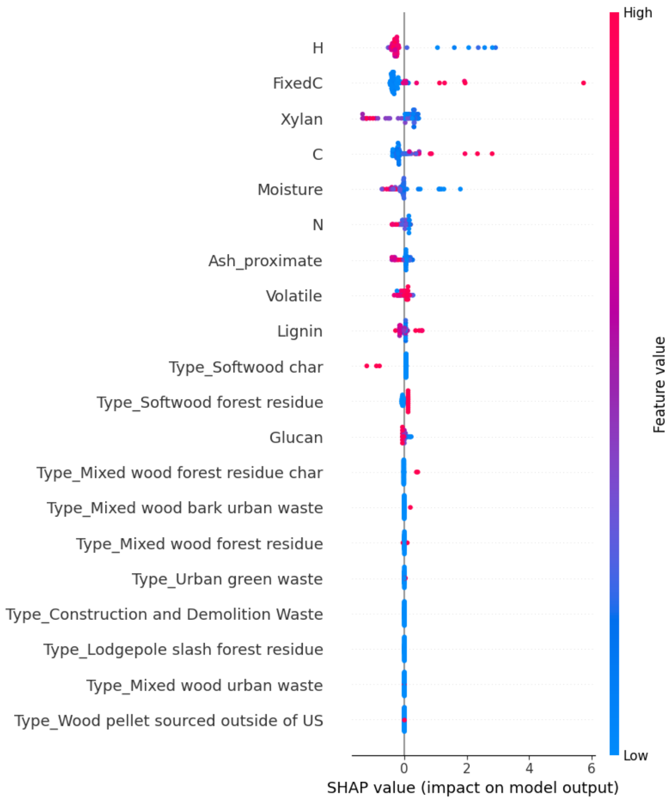

Figure 10. SHAP Summary plot – HHV

In summary, the SHAP plot confirms carbon content and fixed carbon to be the main positive predictors of HHV, while hydrogen and moisture have a decreasing effect on HHV. It also demonstrates that the predictions of the AI model can be explained using SHAP, thus improving the interpretability of the thesis framework.

In Figure 11, the average SHAP value of each variable is plotted. The top predictor variable is hydrogen (H), followed by fixed carbon (FixedC), xylan, carbon (C), and moisture. This suggests that the model strongly depends on elemental information and main structural variables in predicting higher heating value. Hydrogen and fixed carbon are the largest, average contributors and moisture is also significant and contributes less than the top five variables. They have an impact on HHV but it is smaller.

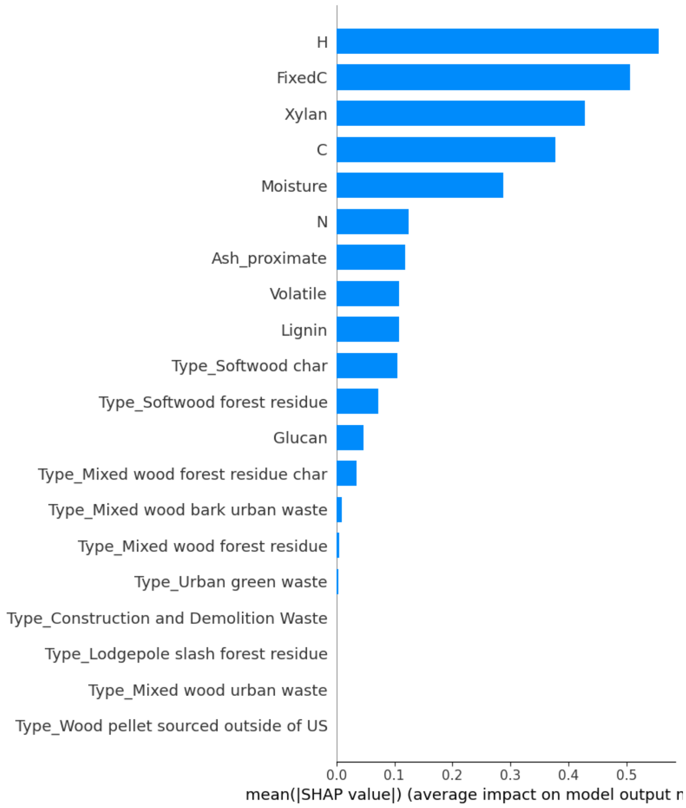

Figure 11. SHAP feature importance – HHV

4.9.2 LHV

Figure 12 LHV shows that the main contributors to the prediction are FC to volatile, xylan, ash proximate, volatile matter, fixed carbon, carbon, and moisture. This is because these predictors have the greatest average influence on the predicted LHV. This suggests a greater dependency of LHV prediction on combination of the compositional factors rather than on feedstock type. The FC to Volatile ratio is the most important predictor and higher values are positive for LHV prediction. This is as expected, as a higher FC/Vol ratio typically indicates higher-strength biomass. Fixed carbon and carbon content also show significant positive influence; therefore, higher carbon-rich fractions in feedstocks lead to higher predicted LHV values. By contrast, volatile and moisture content mostly have a negative effect, which is in line with their properties of diminishing effective energy content in biomass.

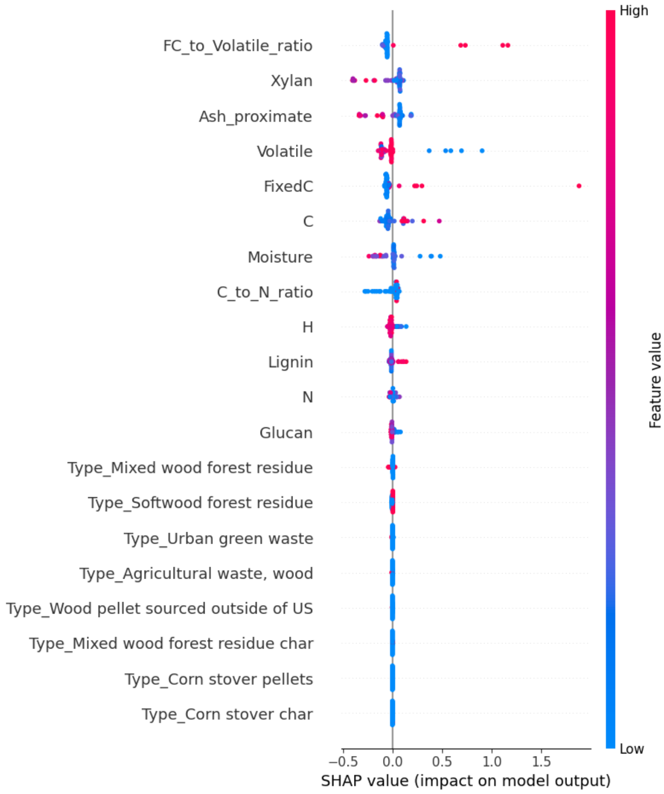

Figure 12. SHAP Summary plot – LHV

Xylan and ash proximate as shown in Figure 12 also have significant effects, but in a non-unidirectional way. This implies that the chemical structure and mineral content of biomass have a combined effect on LHV. Lignin, nitrogen, and glucan are second-order variables with a lower influence than the main predictors, but they still contribute to the overall predictive power of the model. Many of their SHAP values are near to zero, which explains that these are less important predictors than the leading variables.

Overall, figure 12 suggests that fixed carbon, volatiles, carbon, moisture, and xylan contributions dominate the prediction of the LHV, while the other variables effectively contribute less. This justifies the interpretation of the results using SHAP to rank the biomass properties in terms of their importance to energy assessment.
Figure 13 demonstrates that the model depends mostly on the FC to Volatile ratio, then xylan, ash proximate, volatile matter, fixed carbon, carbon, moisture, and carbon to nitrogen (C to N) ratio. This suggests that the lower heating value is driven most by the ratio of fixed carbon to volatile rather than sample types. The FC to Volatile ratio has the

largest average SHAP value, so it plays the most vital role in determining the model predictions of LHV. The following important variables are xylan, ash proximate, volatile matter, and fixed carbon, meaning the lignocellulosic structure and proximate analysis are a key factor in predicting the LHV. This is significant, as it demonstrates that lignocellulosic composition alone plays a strong role in predicting the LHV. Ash and volatile matter are also ranked as significant, which makes sense according to biomass energy theory because ash is inversely correlated with fuel quality and volatile matter influences combustible behavior and combustion energy.

Carbon and moisture are also moderately important. Carbon increases heating value, whereas moisture typically decreases the energy available (due to heat loss during water evaporation).

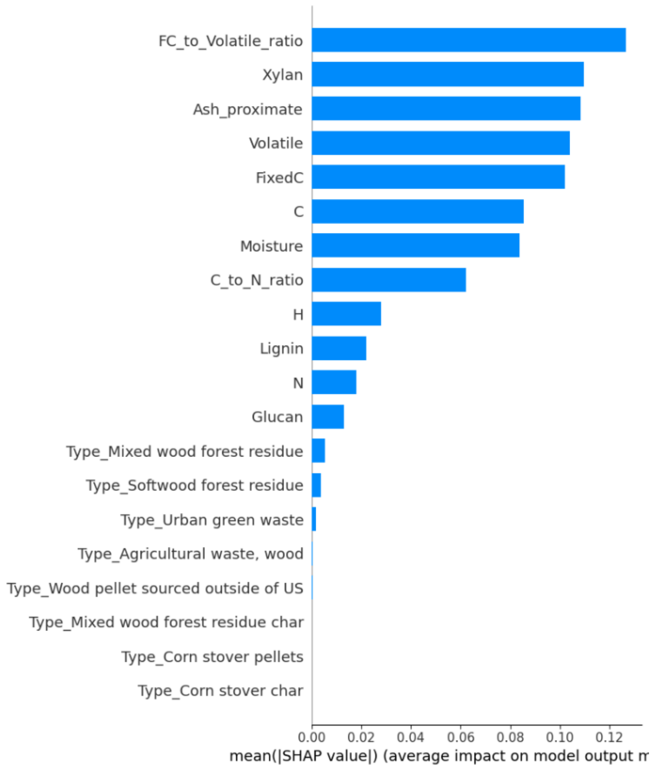

Figure 13. SHAP feature importance – LHV

The C to N ratio also has an important contribution, implying that adjusted ratios capture aspects of the composition that are not represented by original variables. On the other hand, the average effects of hydrogen, lignin, nitrogen, and glucan are smaller, but they still contribute valuable information.

4.10 Sensitivity Analysis

The variables carbon (C), hydrogen (H) and lignin concentrations were used for sensitivity analyses, where the variables were increased and decreased by 5% to test the model stability and to show which variables caused the largest change in the model results in table 7. This is an important test because its foes indicated if the model is responding in a consistent and sensible way as the features of the biomass are slightly changed.

Table 7. Sensitivity analysis of HHV and LHV

	Variable	Change	Predicted Value	Estimated GHG kgCO2 per kg
HHV	C	-5%	19.295341	0.173658
	C	5%	19.501651	0.175515
	H	-5%	19.432891	0.174896
	H	5%	19.354613	0.174192
	Lignin	-5%	19.309098	0.173782
	Lignin	5%	19.470832	0.175237
LHV	C	-5%	17.961686	0.161655
	C	5%	18.045197	0.162407
	H	-5%	18.019642	0.162177
	H	5%	17.987162	0.161884
	Lignin	-5%	17.977547	0.161798
	Lignin	5%	18.027704	0.162249

Results indicate that carbon positively affects the HHV and LHV where 5% reduction in carbon content resulted in a lowering of predicted HHV to 19.295341 MJ/kg and LHV to

17.961686 MJ/kg respectively, while a 5% increase in carbon increased the predicted HHV to 19.501651 MJ/kg and LHV to 18.045197 MJ.kg in table 7. This suggests that any variation in carbon results in variation of energy prediction. This may be because carbon-rich feedstocks typically have higher chemical energy content meaning higher heating values.

The influence of hydrogen was less significant, 5% decrease in hydrogen produced a predicted HHV of 19.432891 MJ/kg and LHV to 18.019642 MJ/kg, while a 5% increase resulted in 19.354613 MJ/kg and 17.987162 MJ/kg in table 7. The effect of hydrogen on HHV and LHV is smaller in the current model. The change in the output is small, and hydrogen is not the primary source of change in heating values. The overall value for the change is slightly negative and shows that hydrogen has less influence on the energy prediction value than carbon-rich structural variables.

The change in output also exhibits a positive response for lignin for HHV and LHV. A 5% reduction in lignin resulted in a predicted HHV of 19.309098 MJ/kg and LHV to 17.977547 MJ.kg, but a 5% increase in lignin led to a predicted HHV of 19.470832 MJ/kg and 18.027704 MJ/kg for LHV in table 7. That confirms the positive effect of lignin on energy, which is in good agreement with the existing biomass literature as lignin is one of the major and carbon-rich structural components that may indicate higher fuel quality. The emissions predicted over the biomass lifecycle also improved slightly when lignin content varied and show that better energy is associated with a slight increase in emissions under the emission-factor method adopted in this study.

4.11 Feedstock Ranking and Sustainable Selection

Char-based feedstocks are at the top of the list in table 8. The highest energy-to-emission ratios (approximately 125), and thus the best energy-production-emissions balance, are associated with corn stover char, mixed wood urban waste char, mixed wood forest residue char, corn pellet char, softwood char, and wood pellet char. These feedstocks were also predicted to have high HHV and LHV, which confirms that carbon-rich, thermally treated feedstocks typically perform better for bioenergy applications than residues or highly processed waste materials. Their value rating indicates that char-based biomass is a potentially ideal feedstock for sustainable bioenergy in this case.

Table 8. Feedstock recommendation table

Type	**Pred HHV MJkg**	**Pred LHV MJkg**	**Estimated GHG kgCO2eq per kg using HHV**	**Energy to Emission Ratio**
Corn stover char	28.7020	18.9301	0.2296	124.9995
Mixed wood urban waste char	27.9629	18.9102	0.2237	124.9994
Mixed wood forest residue char	26.9467	18.8130	0.2156	124.9994
Corn pellet char	24.4461	19.7057	0.1956	124.9994
Softwood char	22.6177	19.0046	0.1809	124.9993
Wood pellet char	21.1526	19.7297	0.1692	124.9993
Agricultural waste, almond husk char	20.2638	18.2163	0.1621	124.9992
Jeffrey pine slash forest residue	20.4837	18.1712	0.1844	111.1105
Softwood forest residue	20.2720	18.1990	0.1824	111.1105
Jeffrey pine chips forest residue	20.2494	18.3446	0.1822	111.1105
Lodgepole pine chips forest residue	20.0755	18.3297	0.1807	111.1105
Lodgepole slash forest residue	19.9142	18.2715	0.1792	111.1105
Mixed wood forest residue	19.4427	17.9177	0.1750	111.1105
Screened softwood chips	18.9074	18.1363	0.1891	99.9995
Agricultural waste, wood	17.2037	17.0637	0.1720	99.9994
Burned Jeffrey Pine wood	20.4162	18.4464	0.2246	90.9087
Agricultural waste, almond husk	18.9662	17.3445	0.2276	83.3330
Corn stover	18.4827	17.1223	0.2218	83.3330
Wood pellet sourced outside of US	19.2856	17.5587	0.2507	76.9228

Corn stover pellets	16.7469	16.7539	0.2177	76.9227
Construction and Demolition Waste	19.5336	18.4130	0.2735	71.4283
Mixed wood bark urban waste	19.1048	17.2286	0.2866	66.6664
Mixed wood urban waste	18.7979	18.1136	0.2820	66.6664
Urban green waste	17.7721	17.2848	0.2666	66.6664
Wood pellet ash	22.7920	18.8812	0.4558	49.9999

The second group of feedstocks, such as Jeffrey pine slash forest residue, softwood forest residue, Jeffrey pine chips forest residue, lodgepole pine chips forest residue, lodgepole slash forest residue, and mixed wood forest residue, exhibited the energy-to-emission ratios of around 111. These feedstocks also showed reliable performance and can be regarded as promising biomass options, but the balance of their energy-to-emission ratios was slightly worse than the char feedstocks. This ranking shows that forest residues continue to be promising biomass feedstock for bioenergy systems, particularly if they can be efficiently collected and used.

Feedstocks such as screened softwood chips and agricultural waste wood had low scores (around 100) in the ranking result, which could show a more balanced but less energy-efficient performance. In the middle of the list, we found burned Jeffrey pine wood, almond husk and corn stover with good, but not high, scores (83-91). All those can be used in bioenergy, but they don't have the same environment benefit and good energy yield as the previous ranking.

Feedstocks such as non-US source wood pellet, corn stover pellet, construction and demolition waste, mixed wood bark urban waste, mixed wood urban waste, urban green waste, and particularly wood pellet ash exhibited a lower level of sustainability. Wood pellet ash has the smallest energy-to-emission ratio (about 50), and thus, was the least sustainable. The findings indicate that multiple-processed (e.g. wood pellets) or high-ash feedstocks might have higher environmental impacts or lower usable energy content and therefore be less preferred in sustainable biomass feedstock selection for the current study.

4.12 Discussion of Major Findings

The machine learning algorithms were able to capture the relationship between inputs and the target variables, although the relationships between inputs and HHV being different from those between inputs and LHV. The highest overall accuracies for HHV and LHV predictions were obtained for SVR, which demonstrates that nonlinear kernel-based machine learning is a better approach to predicting bioenergy properties than the traditional empirical methods.

A key finding of the analysis is that the algorithm achieved better performance in predicting HHV than LHV. The AI models still showed much better performance than both empirical baselines which had proposed that AI models should better predict HHV, LHV, and GHG emissions from feedstock properties than linear regression. But also, bioenergy performance and emissions depend more on feedstock composition than feedstock label.

According to the feature importance, SHAP interpretation, and sensitivity test, the most notable features of the models are carbon, hydrogen, lignin, fixed carbon, xylan, moisture and ash. Specifically, carbon and lignin content have a positive influence on HHV, but moisture and volatile matter have a negative influence. As for LHV, the fixed carbon-to-volatiles ratio and xylan are the most important variables that proposed that some feedstock characteristics are more important than operational factors in determining energy content and emission.

The lifecycle emission assessment also demonstrated the impact of feedstock features on sustainability. The AI-based approach generated more consistent and plausible results for emissions than the empirical approach and the ranking of feedstocks indicated that char-based fuels and wood-based biomass were more likely to provide a high energy content with timely lifecycle emissions. On the other hand, feedstocks with high ash content and urban waste solid fuels exhibited poorer sustainability.

The evidence of the usefulness of the proposed approach is best seen in the results from the AI-based and empirical methods. For heating value (HHV and LHV), the empirical estimation was unsuitable while AI-based models produced more reliable values. This was also found for lifecycle emissions, where the AI-based values were more realistic than the empirical values. This verifies that the empirical method is not adequate for diverse biomass data and justifies the proposed approach of using machine learning for integrated biomass assessment, which suggested that biomass composition, combustible, lignocellulosic characteristics could be used to estimate bioenergy efficiency and lifecycle emissions with AI models.

This study also provides some support for combining AI predictions with LCA-derived emission factors that will yield more suitable estimates of GHG emissions than conventional empirical methods. The AI-based system gave better emission predictions than the empirical system and better maintenance of the rank of the feedstocks. However, the combination of AI models and emission factors enhanced the analytical power of the framework and enabled useful sustainability.

5. Conclusion and Recommendations

5.1 Introduction

The composition of biomass feedstock is a powerful predictor of the quality of bioenergy and the lifecycle of greenhouse gases emitted, and the relations can be approximated effectively with an AI-enhanced framework of analysis using biomass feedstock data. Machine learning offers a better way of predicting biomass energy, more reliable and more flexible as compared to the traditional method of using conventional empirical estimations. SVR was the best-performing algorithm overall (HHV and LHV) in the testing, which demonstrates that nonlinear data-driven modeling is particularly appropriate with heterogeneous biomass. When the AI-based and the empirical methods were compared, evidence of a strong improvement in the predictive accuracy was observed. Not only have we predicted energy properties, but we also combined feedstock characterization, machine learning prediction, empirical baseline comparison, and feedstock ranking in a single framework. This integration enhances the analytic worth of the research, and it is a difference with other traditional biomass prediction studies which would characteristically analyze either the energy content or emissions individually. The selected and char-based woody feedstocks have the preferred balance between bioenergy potential and lifecycle emissions, and the ash-rich feedstocks and urban waste feedstocks are disadvantaged.

The importance of this finding is at the local level as it helps make informed feedstock screening and bio-mass use decisions in BiCRS and bio-energy use. Regularly, the topic of low-carbon energy systems around the world is being discussed by demonstrating that AI can be applied to enhance biomass analysis, minimize uncertainty in feedstock analysis, and assist in allocating resources in a more sustainable way.

Concisely, the primary importance of the present study is in correlating the feedstock composition, bioenergy forecast, and lifecycle emissions in a single coherent analytical framework when combining predictive machine learning, empirical comparison, explainable AI, and sustainability ranking of biomass feedstocks. Lifecycle emissions depend on the type of feedstock and are affected by the composition. Feedstocks comprised of char and woody tended to have better energy-emission profiles than the ash-rich and urban waste materials. By combining emission-factor mapping and projected bioenergy properties, the comparison of feedstocks based on energy content, as well as, based on environmental burden, was possible. The analysis of feature importance, SHAP interpretation, correlation coefficient, and sensitivity analysis indicated consistently that carbon, hydrogen, lignin, fixed carbon, xylan, moisture and ash are the most significant variables to use in the model to derive variables like the fixed carbon to volatile ratio, which can be used to explain biomass quality and the C/N ratio to verify that sustainable biomass selection is based on both the compositional quality and the lifecycle emission performance.

5.2 Theoretical and Practical Contributions

The work will help develop a more comprehensive knowledge of the impact biomass feedstock composition has on both the performance of energy and the environmental impact. The three elements, which are elemental composition, proximate analysis and lignocellulosic structure are not independent markers of biomass quality, but rather, they interact to determine higher heating value, lower heating value, and approximate lifecycle greenhouse gas emissions of biomass. This confirms the opinion that the assessment of biomass is a multivariate and nonlinear issue as compared to a single-step situation of estimating the activity. Another theoretical contribution that can be seen is that machine learning may be applied not just to make predictions, but also interpretations and sustainability comparisons. Nonlinear models are more appropriate, compared to the conventional empirical methods in heterogeneous biomass data, which was demonstrated using SVR, RF, XGBoost, and ANN. Of these models, SVR proved to have the highest overall performance in both HHV and LHV; this validates the idea that kernel-based learning can be effective when it comes to biomass predications problems with more complex feedstock-output relationships than a linear or equation-based method, and where the data consists of a wide variety of biomass types and a large number of different compositional variables.

Conversely, the findings of the feedstock ranking show that char-based and selected woody feedstocks are the best to combine energy production and lifecycle emissions. It has direct applications to screening biomass, prioritization of feedstock, and low-carbon bioenergy planning. The findings are the ones that can assist the researchers, policymakers, and bioenergy developers in choosing the feedstocks, which offer them a more efficient balance of energy efficiency and environmental performance. Practically, the framework may be applied to reduce the number of feedstock options after which more specific experimental, techno-economic or lifecycle analysis is to be conducted. The work is also practical in that it demonstrates that AI-based forecasting is more stable compared to empirical prediction. In the case of HHV and LHV, the empirical baseline did not work well, whereas in the case of more heterogeneous feedstocks like biomass, the more heterogeneous the data, the stronger the AI-based framework prediction that can be useful in assessing biomass in practice.

5.3 Limitations

The research has several limitations, firstly, the analysis was performed with secondary data but not with primary experimental data. At the time of the publication, even though the dataset was appropriate to do machine learning analysis, it was tabulated based on an available source hence relied on the quality, completeness, and consistency of the initial records. Consequently, this gave constraints to the study due to the accessibility of the variables utilized in datasets and the inability to include other biomass properties that affect bioenergy performance and lifecycle emissions.

Secondly, the data was not exceptionally large, about 79 samples were used. Although it was an adequate size of a sample to undertake the analysis and derive useful patterns, it remains a limitation to the extrapolation of the models. Lastly, small datasets are more prone to overfitting and when models are used on a scale of larger or more varied biomass feedstocks, their performance can vary. It is due to this fact that the results can be seen as a solid framework which needs more refinement instead of a universal model that is applicable in all biomass types.

Thirdly, a feedstock-group emission-factor method was used to estimate lifecycle greenhouse gas emissions in this study as compared to a full experimental lifecycle assessment. This implies that the emission estimates are a simplified and comparative sustainability metric and not a full cradle-to-grave lifecycle inventory. This approach is efficient to screen and compare as well as not encompass all the site-related, logistical, and process-level variables affecting emissions in the real world. Thus, the environmental outcomes are to be considered as some estimations, not precise measurements of the lifecycle.

Fourthly, the framework was created in a single-dataset setting, even though various machine learning model usage was used in the study. This implies that no external validation of the models that were external to an independent dataset was done.

Consequently, the issue of model transferability concerning other biomass sets or other systems in relation to BiCRS is yet to be addressed. Furthermore, a hybrid machine learning model could have provided more precise predictions and, if externally validated, would have increased confidence in the framework's robustness and its broader applicability.

5.4 Recommendations for Future Research

The initial priority in future investigations is to broaden the data applied in the biomass energy and lifecycle emission modeling. The secondary data used in the present study was quite small and even though the findings were significant, more extensive, and diverse data would enhance robustness and generalizability of the models. Specifically, biomass samples should be increased, geographical location be observed in more categories, additional feedstock types to ensure that the models can capture the variation of reality biomass resources and help improve the prediction where we took the elemental composition, proximate analysis, lignocellulosic composition and feedstock type of biomass, but other variables including particle size, geographical origin, storage, pretreatment history, conversion technology and harvest These variables would be included, enabling greater explanation of the variability in HHV, LHV and lifecycle emissions and provide support in the full characterization of biomass. This would come in handy particularly in enhancing predictive abilities of the models of complex and heterogeneous feedstocks.

The other significant trend is full-lifecycle assessment, instead of the simplified emission-factor based estimation that is shifting towards full cradle-to-grave lifecycle inventory data. This would enable the evaluation of emissions to be considered more closely in cultivation, transport, preprocessing, conversion, and end use. Another more thorough LCA would be used to make the biomass feedstocks have a stronger base in terms of environmental interpretation and offer a more accurate determination of how thematic biomass is on climate impact.

Also, newer studies need to be done on more advanced and hybrid machine learning methods. Whereas SVR has been the most effective in our research, alternative options like hybrid ensemble models, stacked models or transfer learning approaches can be used to give an additional boost. Deep learning can also be applicable to larger data sets or to a combination of distinct types of biomasses. Meanwhile, the interpretability of the models ought to not be compromised to ensure that the formed predictions are still associated with the results of the feedstock chemistry and sustainability.

Lastly, a study into the future is advised to evaluate the framework on industrial or pilot scale environment where the tested model can be applied at work within a biomass system. This validation would demonstrate the effectiveness of the framework in real-world conditions and would assist in breaking the gap between laboratory-level data analysis and bigger bioenergy use.

5.5 Biomass Selection and Bioenergy Assessment

Selection of the feedstock should not rely on a sole characteristic like high HHV or availability of biomass. Rather, a composite perspective involving both compositional quality and predicted energy performance and lifecycle greenhouse gas emissions should be used to evaluate biomass. The importance of this is that a feedstock that contains a lot of energy might not be very desirable when its environmental burden is high. To conduct viable biomass screening, however, it is advisable to make use of an integrated framework prior to choosing feedstocks to bioenergy or BiCRS-related uses.

Secondly, selected wooden feedstocks and char-based feedstocks are to be preferred in instances where the aim is to generate the most energy possible by minimizing the lifecycle emissions. The results of the ranking revealed that the char of corn stover, mixed wood urban waste materials, mixed wood forest residue materials, corn pellet materials, softwood materials, and wood pellet materials have the best balance between bioenergy potential and emissions. Such feedstocks can hence be good subjects of low carbon bioenergy systems. Conversely, highly mixed, and ash-rich urban waste feedstocks are more to be treated with caution, due to poor predicted sustainability performance.

Thirdly, biomass assessment must emphasize more compositional variables which emerged to be most significant in this research. When the use of biomass in practical biomass assessment and screening, carbon, lignin, fixed carbon, xylan, moisture, ash, and volatile matter should be used as key screening variables. These variables have a better foundation to decide than feedstock label alone since the variables are more directly related to heating value and lifecycle emissions. Feedstock characterization should go as far as possible with elemental, proximate, and lignocellulosic analysis prior to making any conversion decisions.

Fourthly, biomass assessment needs to use machine learning models as decision-support tools. SVR worked most effectively in prediction for both HHV and LHV, showing that nonlinear AI-based techniques can enhance biomass assessment in this study. In practice, machine learning can be used to predict the energy content, compare feedstocks, and select the best biomass options prior to conducting costly laboratory or pilot-scale analyses. This will save time and effort when selecting feedstock at initial stages.

Fifthly, comparatively low-carbon bioenergy performance and lifecycle emissions should be considered in all the biomass evaluation studies. Easy access to energy-rich biomass does not necessarily imply that it is the most sustainable choice and analysis of emissions is required to ensure that a feedstock with a poor environmental footprint is not chosen. The practical biomass evaluation should thus integrate the HHV, LHV, and lifecycle GHG emissions under one model. This will assist researchers and practitioners to better trade-off their decisions regarding feedstock selection, conversion pathways, and in terms of sustainability.

References

Aalto, M., Kc, R., Korpinen, O.-J., Karttunen, K., & Ranta, T. (2019). Modeling of biomass supply system by combining computational methods – A review article. *Applied Energy*, *243*, 145–154. https://doi.org/10.1016/j.apenergy.2019.03.201

Ahmed, M. U., Andersson, P., Andersson, T., Aparicio, E. T., Baaz, H., Barua, S., Bergström, A., Bengtsson, D., Orisio, D., Skvaril, J., & Zambrano, J. (2019). A Machine Learning Approach for Biomass Characterization. *Energy Procedia*, *158*, 1279–1287. https://doi.org/10.1016/j.egypro.2019.01.316

Cacho, J. F., Feinstein, J., Zumpf, C. R., Hamada, Y., Lee, D. J., Namoi, N. L., Lee, D., Boersma, N. N., Heaton, E. A., Quinn, J. J., & Negri, C. (2023). Predicting Biomass Yields of Advanced Switchgrass Cultivars for Bioenergy and Ecosystem Services Using Machine Learning. *Energies*, *16*(10), 4168. https://doi.org/10.3390/en16104168

Cherubini, F., Bird, N. D., Cowie, A., Jungmeier, G., Schlamadinger, B., & Woess-Gallasch, S. (2009). Energy- and greenhouse gas-based LCA of biofuel and bioenergy systems: Key issues, ranges and recommendations. *Resources, Conservation and Recycling*, *53*(8), 434–447. https://doi.org/10.1016/j.resconrec.2009.03.013

Demirbas, A. (2002). Relationships between Heating Value and Lignin, Moisture, Ash and Extractive Contents of Biomass Fuels. *Energy Exploration & Exploitation*, *20*(1), 105–111. https://doi.org/10.1260/014459802760170420

Dogan, E., & Inglesi-Lotz, R. (2017). Analyzing the effects of real income and biomass energy consumption on carbon dioxide (CO2) emissions: Empirical evidence from the panel of biomass-consuming countries. *Energy*, *138*, 721–727. https://doi.org/10.1016/j.energy.2017.07.136

Friedl, A., Padouvas, E., Rotter, H., & Varmuza, K. (2005). Prediction of heating values of biomass fuel from elemental composition. *Analytica Chimica Acta*, *544*(1–2), 191–198. https://doi.org/10.1016/j.aca.2005.01.041

Guldhe, A., Singh, B., Renuka, N., Singh, P., Misra, R., & Bux, F. (2017). Bioenergy: A Sustainable Approach for Cleaner Environment. In K. Bauddh, B. Singh, & J.

Korstad (Eds.), *Phytoremediation Potential of Bioenergy Plants* (pp. 47–62). Springer Singapore. https://doi.org/10.1007/978-981-10-3084-0_2

Güner, Ş. T., Diamantopoulou, M. J., Poudel, K. P., Çömez, A., & Özçelik, R. (2022). Employing artificial neural network for effective biomass prediction: An alternative approach. *Computers and Electronics in Agriculture*, *192*, 106596. https://doi.org/10.1016/j.compag.2021.106596

Harman-Ware, A. E., & Paeper, C. (n.d.). *Biomass for Carbon Removal and Storage (BiCRS) Biomass and Bioproduct Data Table* (p. 1 files.) [Dataset]. National Laboratory of the Rockies. https://doi.org/10.7799/3009155

Harman-Ware, A. E., & Paeper, C. (2025). *Biomass for Carbon Removal and Storage (BiCRS) Biomass and Bioproduct Data Table* (p. 1 files.) [Dataset]. National Laboratory of the Rockies. https://doi.org/10.7799/3009155

Hassan, R., sharifzadegan, A., & Baghban, A. (2025). AI-based algorithms for estimating hydrochar properties in terms of biomass ultimate analysis. *Results in Engineering*, *26*, 105141. https://doi.org/10.1016/j.rineng.2025.105141

Hosseinzadeh-Bandbafha, H., Aghbashlo, M., & Tabatabaei, M. (2021). Life cycle assessment of bioenergy product systems: A critical review. *E-Prime - Advances in Electrical Engineering, Electronics and Energy*, *1*, 100015. https://doi.org/10.1016/j.prime.2021.100015

IEA. (2024). *Bioenergy*. IEA. https://www.iea.org/energy-system/renewables/bioenergy

IPCC (Ed.). (2023). *IPCC, 2023: Climate Change 2023: Synthesis Report.* (First). Intergovernmental Panel on Climate Change (IPCC). https://doi.org/10.59327/IPCC/AR6-9789291691647

IRENA. (2022). *Bioenergy for the energy transition: Ensuring sustainability and overcoming barriers*.

Johnson, D. R., Willis, H. H., Curtright, A. E., Samaras, C., & Skone, T. (2011). Incorporating uncertainty analysis into life cycle estimates of greenhouse gas emissions from biomass production. *Biomass and Bioenergy*, *35*(7), 2619–2626. https://doi.org/10.1016/j.biombioe.2011.02.046

Liao, M., & Yao, Y. (2021). Applications of artificial intelligence-based modeling for bioenergy systems: A review. *GCB Bioenergy*, *13*(5), 774–802. https://doi.org/10.1111/gcbb.12816

Liu, M., Tao, J., Mu, L., Su, H., Peng, H., Cheng, Z., & Chen, G. (2025). Machine learning-driven predictions of biochar yield and NPK composition: Insights into biomass pyrolysis with data augmentation and model interpretability. *Carbon Research*, *4*(1), 62. https://doi.org/10.1007/s44246-025-00229-1

Long, F., & Liu, H. (2023). An integration of machine learning models and life cycle assessment for lignocellulosic bioethanol platforms. *Energy Conversion and Management*, *292*, 117379. https://doi.org/10.1016/j.enconman.2023.117379

Maksimuk, Y., Antonava, Z., Krouk, V., Korsakova, A., & Kursevich, V. (2021). Prediction of higher heating value (HHV) based on the structural composition for biomass. *Fuel*, *299*, 120860. https://doi.org/10.1016/j.fuel.2021.120860

Mari Selvam, S., & Balasubramanian, P. (2023). Influence of Biomass Composition and Microwave Pyrolysis Conditions on Biochar Yield and its Properties: A Machine Learning Approach. *BioEnergy Research*, *16*(1), 138–150. https://doi.org/10.1007/s12155-022-10447-9

Martins, L. O. S., Carneiro, R. A. F., Torres, E. A., Silva, M. S., Iacovidou, E., Fernades, F. M., & Freires, G. M. (2019). Supply Chain Management of Biomass for Energy Generation: A Critical Analysis of Main Trends. *Journal of Agricultural Science*, *11*(13), 253. https://doi.org/10.5539/jas.v11n13p253

Sandalow, D., Aines, R., Friedmann, J., McCormick, C., & Sanchez, D. (2021). *Biomass Carbon Removal and Storage (BiRCS) Roadmap* (LLNL-TR--815200, 1763937, 1024342; p. LLNL-TR--815200, 1763937, 1024342). https://doi.org/10.2172/1763937

Sharmila, V. G., Shanmugavel, S. P., & Banu, J. R. (2024). A review on emerging technologies and machine learning approaches for sustainable production of biofuel from biomass waste. *Biomass and Bioenergy*, *180*, 106997. https://doi.org/10.1016/j.biombioe.2023.106997

Sun, C., Yang, Z., Zheng, Z., Li, W., Tan, H., Huang, Y., & Zhang, Y. (2023). Exploring how lignin promoting the co-pyrolysis with polylactic acid: Artificial neural network modeling, kinetic analysis and product distribution. *Sustainable Materials and Technologies*, *35*, e00549. https://doi.org/10.1016/j.susmat.2022.e00549

Šurić, J., Voća, N., Peter, A., Bilandžija, N., Brandić, I., Pezo, L., & Leto, J. (2023). Use of Artificial Neural Networks to Model Biomass Properties of Miscanthus (Miscanthus × giganteus) and Virginia Mallow (Sida hermaphrodita L.) in View of Harvest Season. *Energies*, *16*(11), 4312. https://doi.org/10.3390/en16114312

Thornley, P., Gilbert, P., Shackley, S., & Hammond, J. (2015). Maximizing the greenhouse gas reductions from biomass: The role of life cycle assessment. *Biomass and Bioenergy*, *81*, 35–43. https://doi.org/10.1016/j.biombioe.2015.05.002

U.S. Department of Energy. (2025). *Best Practices for Life Cycle Assessment (LCA) of Biomass Carbon Removal and Storage (BiCRS) Technologies*. https://www.energy.gov/sites/default/files/2025-01/BiCRS%20Best%20Practices%20for%20LCA.pdf

Wang, R., He, Z., Chen, H., Guo, S., Zhang, S., Wang, K., Wang, M., & Ho, S.-H. (2024). Enhancing biomass conversion to bioenergy with machine learning: Gains and problems. *Science of The Total Environment*, *927*, 172310. https://doi.org/10.1016/j.scitotenv.2024.172310

Wang, Z., Peng, X., Xia, A., Shah, A. A., Huang, Y., Zhu, X., Zhu, X., & Liao, Q. (2022). The role of machine learning to boost the bioenergy and biofuels conversion. *Bioresource Technology*, *343*, 126099. https://doi.org/10.1016/j.biortech.2021.126099