Human-centric ensemble AI for hydrothermal carbonization modeling and hydrochar properties prediction

Abdeldayem, Omar M.; Al-Sakkari, Eslam G.; Ortiz, Darwin; Dupont, Capucine; Ferras, David; Ouali, Mohamed Salah; Ragab, Ahmed; Kennedy, Maria

DOI
10.1016/j.jece.2025.117826

Publication date
2025

Document Version
Final published version

Published in
Journal of Environmental Chemical Engineering

Journal of Environmental Chemical Engineering 13 (2025) 117826

Contents lists available at ScienceDirect

Journal of Environmental Chemical Engineering

journal homepage: www.elsevier.com/locate/jece

Human-centric ensemble AI for hydrothermal carbonization modeling and hydrochar properties prediction

Omar M. Abdeldayem [a,b,c,*,1], Eslam G. Al-Sakkari [c,d,1], Darwin Ortiz [a,1], Capucine Dupont [a], David Ferras [e], Mohamed-Salah Ouali [c], Ahmed Ragab [c,d], Maria Kennedy [a,b]

[a] Department of Water Supply, Sanitation and Environmental Engineering, IHE Delft Institute for Water Education, Westvest 7, Delft 2611AX, the Netherlands
[b] Department of Water Management, Faculty of Civil Engineering and Geosciences, Delft University of Technology, Stevinweg 1, Delft 2628 CN, the Netherlands
[c] Department of Mathematics and Industrial Engineering, Polytechnique Montréal, 2500 Chemin de Polytechnique, Montréal, Québec H3T 1J4, Canada
[d] Natural Resources Canada - CanmetENERGY, 1615 Lionel-Boulet Blvd, P.O. Box 4800, Varennes, Québec J3X 1P7, Canada
[e] Department of Civil and Mining Engineering, University Polytecnic of Cartagena, Pza. del Cronista Isidoro Valverde Edif. La Milagrosa C.P., Cartagena 30202, Spain

ARTICLE INFO

Keywords:
Biomass
Hydrothermal carbonization
Hydrochar
Machine learning
Ensemble learning
Decision fusion

ABSTRACT

Hydrothermal carbonization (HTC) is a promising process for biomass valorization; however, optimizing HTC conditions and characterizing hydrochar through experimental methods remain costly and time-consuming. Artificial intelligence (AI) and its related machine learning (ML) techniques provide an efficient and cost-effective alternative, enabling efficient optimization and predictive analysis without extensive experimental tests. In this study, an analysis-ready database (ARD) comprising 544 data points was constructed using the authors' previous research (41) and scattered data compiled from the literature (503). An ensemble of eight diversified machine learning (ML) models was developed using biomass-agnostic properties and process conditions to predict hydrochar properties including elemental analysis, proximate analysis, and hydrochar yield. Tailor-made decision fusion models were developed for each target by merging the outputs of the best-performing models. Furthermore, a set of interpretable machine learning (IML) and explainable AI (XAI) techniques, leveraging feature importance analysis and SHapley Additive exPlanations (SHAP) values, indicated that biomass fixed carbon (FC) content and process temperature are the most influencing features, which were then considered as inputs for the ensemble models. The decision fusion models achieved high accuracy for target prediction, surpassing the currently used models in the literature, with adjusted R^2 values ranging from 0.98 to 1.0. For outputs that are typically difficult to predict, such as hydrochar yield, the model achieved an adjusted R^2 of 0.98, representing over a 5 % improvement compared to the best-performing model reported in the literature. All predictions were derived from white-box modeling by incorporating XAI into the ensemble-based learning, ensuring greater model transparency and interpretability.

1. Introduction

The generation of large quantities of wet biomass waste produced by various activities has gained attention for use as feedstock in various sustainable valorization processes. Dry-based thermochemical processes, such as gasification, pyrolysis, and torrefaction, have been commonly used to valorize low moisture-content biomass. However, these processes are not economically profitable for high moisture-content biomass (>50 wt%) due to the energy-intensive drying step [1].

Hydrothermal carbonization (HTC) is a thermochemical process widely used for valorization of wet waste biomass since it does not require any drying step. It operates in a water medium at temperatures ranging from 180 to 250°C, under autogenic pressure (2–10 MPa), with residence times (RT) that range typically between 1 and 72 h [2,3]. The HTC conditions allow water to act as a solvent and catalyst, facilitating reactions like hydrolysis, dehydration, decarboxylation, aromatization, and polymerization [2]. The main HTC products are solid hydrochar, a liquid fraction, and a (minor) gaseous fraction composed mainly of CO_2

* Corresponding author at: Department of Water Supply, Sanitation and Environmental Engineering, IHE Delft Institute for Water Education, Westvest 7, Delft 2611AX, the Netherlands.
E-mail addresses: o.m.h.m.abdeldayem@tudelft.nl, o.abdeldayem@un-ihe.org, omar.abdeldayem94@gmail.com (O.M. Abdeldayem).
[1] These authors contributed equally to the work

https://doi.org/10.1016/j.jece.2025.117826
Received 1 May 2025; Received in revised form 23 June 2025; Accepted 30 June 2025
Available online 30 June 2025

[4–6].

Hydrochar has the potential for multiple end-uses thanks to its distinctive properties, low inorganic content compared to raw biomass and biochar, presence of microspheres, high heating value (HHV), thermal stability, and hydrophobicity [7,8]. Initially, its applications were primarily limited to soil amendment and solid fuels. However, current advancements have expanded to end used such as adsorbent applications, additives for anaerobic digestion, catalysts, carbon sequestration, electrodes for supercapacitors and sodium-ion batteries [7,9,10].

The diverse applications of hydrochar, which each require specific properties and therefore process conditions, have pushed researchers to find approaches to predict the properties of hydrochar in specific applications. Several prediction approaches have been used in the HTC literature. Statistical approaches are the most common ones; linear and non-linear regression models, design of experiments (DOE) and response surface methodology (RSM), or other statistical methods such as the Taguchi statistical method have been used for the prediction of hydrochar properties [11]. However, the majority of statistical models are case-specific and highly dependent on the experimental data they build on, such as the used feedstock, process conditions (temperature, RT, and biomass to water (B/W) ratio), type and design of the used reactor, heating rate, and cooling rate. Hence, this dependency restricts the predictive ability of these models to a limited range of experimental conditions and hinders their broader applicability [11]. Kinetic and computational fluid dynamics (CFD) models have been used primarily to simulate the phenomena associated with the HTC process, with limitations on the prediction of hydrochar properties. Generally, their prediction is mainly limited to the hydrochar yield from a specific feedstock and reactor (for the case of CFD models). Machine learning (ML) has been adopted to overcome the previously mentioned limitations, relying mainly on a large dataset from experiments to predict various hydrochar properties. Hence, ML enables the development of generalized, biomass-agnostic predictive models that are not constrained by specific feedstocks, process conditions, or reactor configurations, allowing accurate predictions across diverse scenarios. This is closely tied to the concept of generalization, which refers to a model's ability to perform well on unseen or external data beyond the conditions it was trained on. Generalization is critical for ensuring practical applicability across diverse HTC setups [12].

As shown in Table 1, single and ensemble ML algorithms have been used to predict hydrochar properties using various variables, outputs, data points, and algorithms. Some studies have used single learners, such as decision trees, Artificial Neural Networks (ANN), *k*-Nearest *Neighbours* (*K*NN), and Support Vector Regression (SVR). Other studies have used an ensemble ML approach, which combines several weak learners to achieve enhanced prediction with respect to accuracy and stability using models like Random Forest (RF), eXtreme Boost Gradient (XGBoost), Gradient Boost Tree (GBT), Gradient Boost Regression (GBR) as shown in Table 1 [13]. The majority of ensemble models used in the HTC literature are tree-based and can be divided into bagging and boosting techniques [14]. Among those, RF is the most commonly used ensemble learning method using 'bagging', while XGBoost, GBT, and GBR are the commonly used 'boosting models', as shown in Table 1.

Table 1 presents the test R^2 values reported for various models used in the literature. Overall, model accuracy varies considerably depending on the type of model employed and the specific output variable being predicted. Notably, hydrochar yield emerged as one of the most challenging parameters to predict, with the highest reported R^2 in previous studies reaching only 0.93. However, RMSE values are often omitted, limiting a comprehensive understanding of model performance. In the few cases where RMSE was reported, it typically ranged between 4.5 % and 8 %, depending on the dataset and modeling approach. This relatively modest performance compared to other studied properties is reflected in both R^2 and RMSE, and is largely due to the high uncertainty inherent in experimental yield data and the limited generalization ability of single-model approaches across diverse biomass types and process conditions. To address this limitation, our study reports both R^2 and RMSE for hydrochar yield, enabling a more transparent and quantitative assessment of predictive accuracy.

While ensemble techniques such as RF and boosting methods have been applied in some HTC prediction tasks, these approaches are typically limited to homogeneous model families (e.g., tree-based ensembles). To the best of the authors' current knowledge, no diversified ensemble models have been used for predicting hydrochar properties from HTC, which is considered an area for improvement in predicting HTC outputs. Using ensemble diversified models, known as decision fusion models, offers several advantages. Firstly, it combines predictions from multiple models to enhance accuracy, as each model can identify different patterns in the data, leading to better overall results when their predictions are aggregated. Additionally, using a diverse set of models helps to reduce overfitting [30], as each model comes with its own unique bias and error. This diversity helps to balance out individual biases, resulting in more reliable outcomes. Moreover, a single model may struggle to perform well across noisy data, various data types, and distributions. By employing a range of models, different aspects of the data are addressed, making the system more resilient to data variations and uncertainties within it [31].

While ML techniques are capable of predicting the outputs of HTC, their black box nature often poses difficulties in interpreting the underlying processes and results. Given that HTC involves complex interactions between variables such as temperature, residence time, and feedstock characteristics, understanding how each input influences the output is critical for process optimization and reproducibility. Therefore, white box modeling using interpretable machine learning (IML) and explainable AI (XAI) approaches is increasingly important. White-box modeling refers to transparent models whose internal structures and mechanisms are explicitly explainable. Unlike black-box models, white-box models allow users to interpret how input variables directly influence outcomes, making them ideal for applications requiring transparency and explainability (e.g., regulatory compliance or scientific analysis) (Loyola-González, 2019). Explainable AI (XAI) and IML are related concepts, but they have distinct focuses. Interpretable ML aims to understand the inner workings of the model itself, while explainable AI focuses on providing understandable reasons for the model's decisions or predictions often after modeling. Hence, there is a need to add interpretability and explainability to the current models for a better understanding of the HTC process and the variables of interest [32].

In the context of HTC, several studies have employed both IML models and XAI techniques to assess feature importance after developing predictive models [21,22,27]. However, none of these studies have integrated IML and XAI techniques in a combined framework for systematic feature selection. In particular, SHAP (SHapley Additive Explanations) has been widely applied as a post-hoc method to explain the outputs of complex, black-box models. Nonetheless, the terminology is often misused in the HTC literature, where SHAP-based methods are incorrectly described as tools of interpretability rather than explainability, despite their intended role in explaining the predictions of non-transparent models.

Addressing the current knowledge gaps and areas for improvement, the overall objective of this study is to develop a biomass-agnostic model to predict hydrochar properties, while overcoming the limitations of previous research in terms of accuracy and lack of models' transparency and explainability. The specific objectives are: (i) to develop an analysis-ready database (ARD) by compiling experimental data from the authors' research and from the literature, (ii) to integrate combined IML with XAI with human expertise, providing a comprehensive understanding of how process variables influence the HTC process and robust features selection, (iii) to design and compare the performance of eight different single-output ML models for predicting hydrochar properties, and (iv) to create tailored decision fusion models for each specific output based on

Table 1
Overview of various MLAs-based studies for hydrochar prediction.

Feedstock	Input variable	Output variable (hydrochar related)	Number of Data points	Type of MLA	Test R^2	Test hydrochar yield R^2	Test hydrochar yield RMSE	Reference
Sewage sludge and lignocellulosic biomass	C, H, N, O, S, VM, FC, Ash, T, RT, SL, SSR	C, H/C, O/C, N/C, HHV, FR, HY, EY	221	XGB[a], RF[a]	0.83–0.95	0.83	6.55	[13]
Woody biomass, herbaceous biomass, and food wastes	C, H, N, O, S, VM, FC, Ash, t	HY, N/C, HHV, DHD, DCD, Ash	296	ANN-PSO	0.84–0.98	0.86	5.46	[15]
Sewage sludge	C, O, T, RT,	HHV, C, H	-	ANN	0.94–0.97	-	-	[16]
Municipal solid waste	C, H, N, O, VM, FC, Ash, T, RT, WC.	CCS, C, N/C, O/C, H/C, HHV, HY, CR, ER	248	RF, SVM, DNN[a]	0.55–0.91	0.90	7.05	[17]
Sewage sludge	N, C, VM, FC, T	N	138	ANN.	0.88–0.99	-	-	[18]
Cellulose, poplar, and wheat straw	C, N, S, H, T, RT	C	132	ANN, RF[a]-, SVR[a]-, KNN	0.70–0.82	-	-	[19]
Poultry litter	T, RT	C, iP	-	ANN	0.83–0.91	-	-	[20]
Biomass, manure, agricultural waste, food waste, sludge, algae and wood waste	C, H, O, N, VM, Ash, FC, HHV; T, RT, SLR.	HY, Ash, C, HHV, EY efficiency	333	RF, SVM, XGB[a]	0.83–0.99	0.88	4.47	[21]
Agricultural and forestry biomass, manure, sewage sludge, and food waste	C, H, N, O, S, VM, FC, Ash, T, RT, WC	C, H, N, O, S, Ash, HY	716	GBR[a], RF	0.78–0.98	0.91	5.5	[22]
Municipal sludge	C, H, N, O, S, VM, FC, Ash, HHV.	HHV, CR, ER	246	RF[a], GBT, ANN	0.65–0.98	-	-	[23]
Lignin, cellulose, food waste, sludge, and manure	C, H, O, N content, VM, FC, WC, T, HR, Ht	C, H, N, O, HY, C recovery, ER, HHV	497	DNN, RF, XGB[a]	0.75–0.95	0.93	5.68	[24]
Sewage sludge	C, H, N, O, VM, FC, Ash, TP of SS, T, RT, pH of the feedwater, DM,	TP	185	RF	0.92–0.95	-	-	[25]
Sewage sludge, food waste and manure	C, H, N, O, FC, Ash, VM, T, RT, WC (all data dry-based values)	HY, HHV, ED, ER efficiency	248	SVR[a], RF	0.88–0.96	0.88	7.83	[26]
Various biomass types	C, H, N, O, S, FC, Ash, VM, T, RT, SLR	C, H, N, O, S, Ash VM, FC HY, HHV	536	DTR[a], SVR	0.13–0.99	0.88	6.85	[27]
Various biomass types	T, RT, SLR	HHV	521	BO-GPR[a], BO-Ensemble, BO- DT	0.82–0.97	-	-	[28]
Sludge, rice straw, distiller's grains, rice straw, straw, and microalgae	C, H, O, TN, Ash, T, t, P, HR	PY, HY, Ash, pH, TN, TP	226	SVM, ANN, RF[a]	0.82–0.96	0.90	-	[29]

Properties: Volatile matter (VM), fixed carbon (FC), ash content (Ash), carbon (C), hydrogen (H), oxygen (O), nitrogen (N), sulfur (S), solid to liquid ratio (SLR), dry matter content (DM), nitrogen content (N); total phosphorus content (TP), total nitrogen content (TN), inorganic phosphorus (iP), SSR (ratio of sewage sludge in the mixture). Hydrochar yield (HY), Higher heating value (HHV), Fuel Ratio (FR), energy yield (EY), energy densification (ED), energy recovery (ER) process yield (PY), dehydration degree (DHD), decarboxylation degree (DCD), carbon capture and storage (CCS), carbon recovery (CR). **HTC conditions:** Water biomass ratio (W/B), Reaction temperature (T), reaction pressure (Pr), residence time (RT), heating rate (HR), heating time (Ht), SL (solid loading), water content (WC). **MLAs:** Support Vector Machine (SVM), Random Forest (RF), Artificial Neural Network (ANN), Artificial Neural Network combined with Particle Swarm Optimization (ANN-PSO), Gradient Boosting regression (GBR), eXtreme Gradient Boosting (XGB), Deep Neural Networks (DNN), Support Vector Machine regression (SVR), K nearest neighbors (KNN), Gradient Boosting Tree (GBT), Decision Tree Regression (DTR), Bayesian Optimization (BO), Decision Tree (DT), Gaussian Process Regression (GPR).

[a] Best performance model

the eight different single-output ML models while considering the integration of XAI for white-box-based learning (v) evaluate the models using Root Mean Square Error (RMSE), the Coefficient of Determination (R^2), and the Adjusted R^2.

2. Methodology

2.1. General approach

As depicted in Fig. 1, this study follows a human-centric approach, characterized by the active involvement of a human expert throughout all stages. Initially, data were collected from various sources and prepared for analysis. These prepared data were subsequently modeled and validated to ensure accuracy. Throughout each stage, collection, preparation, modeling, and validation, the human expert interacts closely with the data and modeling processes, providing mentorship, oversight, and domain-specific insights to achieve accurate predictions.

2.2. Database development and construction

Fig. 2 outlines the process for developing and constructing an ARD. The first step involved collecting data from various literature sources. Next, the constraints and assumptions within these data were thoroughly analyzed. After that, various visualization techniques (simple statistics), such as Pearson correlation and boxplots were applied to provide insights about the data. The data were then cleaned by removing voids and outliers. Important features were selected, and the final step involved splitting the data into training and testing sets for further analysis.

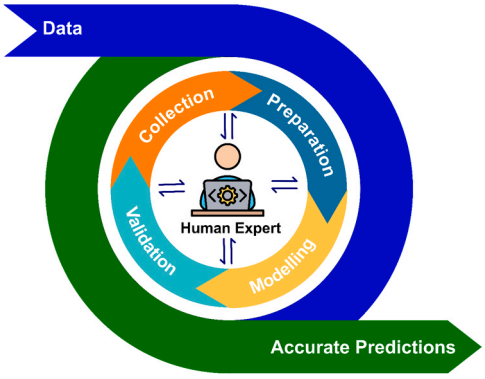

Fig. 1. General approach.

2.2.1. Database description

One of the key requirements for ML modeling is having a representative dataset in terms of both data quality and coverage across the studied range. Although certain properties of biomass and hydrochar, such as macromolecules and inorganic elements, are relevant for predicting hydrochar characteristics, they were excluded from this study because they are not frequently measured in HTC studies, resulting in limited data availability for a complete database construction for ML modeling. Therefore, this study used the most commonly reported biomass properties and process parameters. The main properties of the considered raw biomass and hydrochar were the elemental analysis (carbon (C), hydrogen (H), nitrogen (N), and oxygen content (O)), proximate analysis (volatile matter (VM), fixed carbon (FC), and ash content), and hydrochar yield. The considered operational parameters for the dataset were reactor volume, water volume, temperature, RT, stirring rate, biomass-to-water (B/W) ratio, and moisture content. The full dataset can be visualized in supplementary data 2.

Typical uncertainty ranges for elemental analysis of CHN lie between 0.05 % and 0.5 %, while uncertainty for oxygen (O) is generally around 1 %. As for proximate analysis, the typical uncertainty is around 0.2–0.5 % for ash and ranges between 0.5 % and 1.5 % for VM and FC. As for hydrochar yield, the uncertainty typically ranges from 1 % to 3 % under controlled laboratory conditions, but this value may increase when using heterogeneous biomass. The uncertainty is relatively high due to mass losses that frequently occur during hydrochar yield measurements [9,33].

The complete dataset is composed of 544 data points and was compiled from 43 publications, including previous experimental work from the authors of this work. These publications were considered because they provide the complete analysis required for the database. The keywords for collecting the dataset included hydrothermal carbonization, lignocellulosic biomass, food waste, agricultural waste, manure, and sewage sludge. To account for potential inconsistencies between sources, only studies with well-documented process parameters and complete input-output characterizations were included. These included clearly reported values for HTC conditions, elemental and proximate analyses, and hydrochar yield. This ensured a baseline level of methodological comparability across studies.

The compiled database relies on published HTC studies and experimental results that often employ pre-dried lignocellulosic biomass.

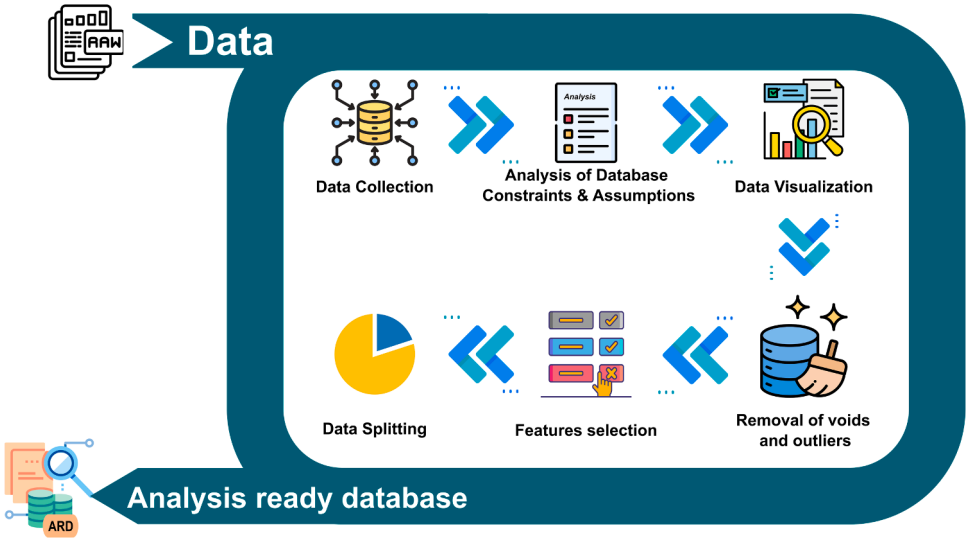

Fig. 2. Construction of analysis ready database.

Although it would have been more desirable to include a higher proportion of wet biomass to better reflect industrial conditions, pre-drying is commonly applied at the laboratory scale for practical reasons, primarily to reduce variability, simplify handling, and control moisture-related effects. This trade-off is widely adopted by researchers to ensure reproducibility and data comparability. Nonetheless, approximately 10 % of the compiled dataset includes feedstocks such as sewage sludge and manure, which were processed using their original moisture content and thus more closely represent industrially relevant scenarios where drying is not viable. The current database represents the most comprehensive set of publicly available and consistently reported input–output parameters to date. As the field evolves, the database remains open to expansion to incorporate additional datasets involving raw biomass with its original moisture content.

2.2.2. Database considerations

In order to maintain consistency and enhance the comparability of the dataset, all data were gathered and processed on a dry basis. Additionally, the elemental and proximate analysis of the biomass and hydrochar in the datasets were unified based on Eq. (1) and Eq (2), respectively, where O and FC are calculated by difference (EN 15104:2011, ASTM D Standard 3172–73).

$$O = 100\% - C - H - N - Ash \tag{1}$$

$$FC = 100\% - VM - Ash \tag{2}$$

The following assumptions were also considered in the dataset construction:

a) If the B/W ratio wasn't provided in a study, it was calculated based on the solid content and the water volume used. Additionally, for experiments that used wet biomass without pre-drying, the B/W ratio was determined using moisture content, added water volume, and solids content.
b) If the biomass was dried as a pretreatment before the HTC process, the moisture content was assumed to zero.
c) If the hydrochar yield wasn't directly reported, the value was extracted from the figures presented in the studies.
d) Sulfur content was not included in the dataset since it is often not measured for biomass. However, new component percentages were recalculated if sulfur content was provided, assuming the absence of sulfur, as per Eq. 1.
e) Classification of biomass is done according to Section 1 in the supplementary file.

2.2.3. Data visualization

A statistical analysis was performed to identify and categorize the different types of biomass in the dataset. A boxplot combined with a kernel density estimation was generated to visualize the distribution and variability of different operating conditions, biomass properties, and hydrochar properties. Moreover, Pearson correlation was calculated using Eq. (3) to preliminarily describe the linear relationship between input and output variables. Further, the *p-value* of the *Pearson* correlation was also calculated to measure the statistical significance of the obtained correlations [34].

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2}\sqrt{\sum (y_i - \bar{y})^2}} \tag{3}$$

Where r represents the *Pearson* correlation coefficient, x and y correspond to the selected variables to be examined, $\bar{x}$ and $\bar{y}$ represent their corresponding means, and x_i and y_i denote the individual values of the studied variables.

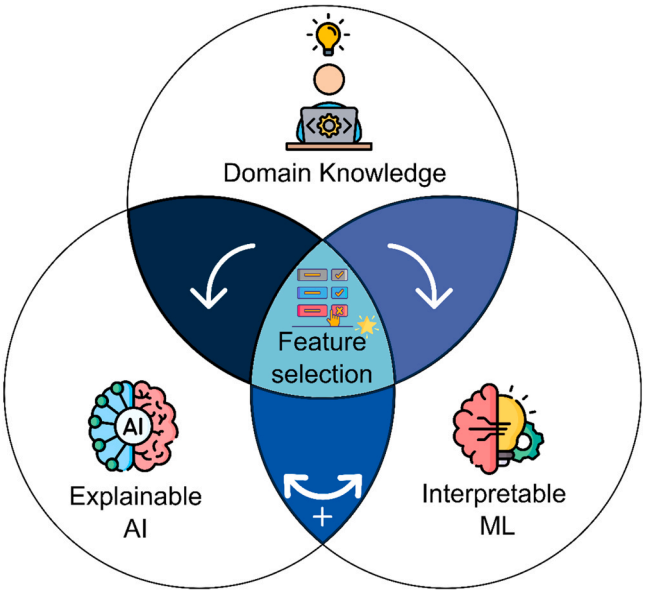

Fig. 3. Feature selection process followed in this study.

2.3. Database preparation for AI modeling

2.3.1. Pre-processing

Several preprocessing steps should be applied to prepare an ARD, as shown in Fig. 2. First of all, empty spaces and voids are removed to ensure data cleanliness and accuracy. This step enables a fair comparison of models by ensuring that all models are evaluated on the same dataset without giving an advantage to models that can handle missing data. Secondly, outliers are removed based on data analysis and human expertise. Moreover, feature importance is done using a weak learner to define the important variables. Additionally, binning and stratification of the dataset were used to ensure that different subgroups within a dataset are proportionally represented in both the training and testing sets; hence reducing bias and enhancing generalization. For certain models, such as ANN, GPR, KNN, and SVR, standardization was applied. Finally, the dataset was randomly split into training and testing sets, 80 % and 20 %, respectively.

2.3.2. Features selection

Conventional feature selection methods often rely solely on human expertise, without incorporating explainability or interpretability. Other conventional approaches include mutual information and feature importance methods that do not involve permutation-based techniques. In this study, a novel methodology was used to perform feature selection based on three main pillars, domain knowledge, XAI and IML as illustrated in Fig. 3. At the beginning of the study, the human expert, with their domain knowledge, can define the possible important features to be studied. Then, with the help of XAI and IML, the most effective features can be selected under the supervision of human experts, *i.e.*, process experts and data analysts.

Tree-based techniques used for feature selection are regarded as semi-explainable methods due to only giving the relative importance of each variable; therefore, an explainable technique was employed in the current study to overcome this challenge [32]. In this study, SHapley Additive exPlanations (SHAP) values were calculated to provide numerical representations of the influence of each controlled variable on the measured ones, including the interactions between different variables. Overall, the combination of interpretable and explainable techniques merges their advantages and avoids their limitations.

The role of IML is to perform feature importance in order to make a well-informed selection. Hence, a weak learner at the start of the study is built to map the possible output variables to all the possible or available

inputs. In the current study, random forest (RF) was developed to perform this mapping. Afterwards, the relative importance of all the input variables is evaluated via the embedded screening ability of RF. These results give only the relative importance without highlighting the sign or the direction of the variables' impacts on the desired outputs. Hence, the combination of IML with XAI can guarantee the best performance for feature selection by providing both the magnitude and direction of all the impacts [32]. As a way to develop the XAI model, SHAP values (as one of the model agnostic methods) were evaluated after modeling using RF. The SHAP Python package was adopted to facilitate the evaluation. Readers interested in more explanation about the mathematical background of feature importance evaluation using tree-based models such as RF and the methodology to determine SHAP values and its integration with ML models can refer to these studies [35–41]. The results give the average SHAP values for the whole dataset after being evaluated locally for each observation/data point. These values give precise estimation of each variable's impact magnitude and direction.

2.4. Modeling

2.4.1. Modeling overview

Hydrothermal carbonization is a complex process that involves multiple interconnected physical and chemical phenomena with non-linear relationships. Therefore, some models might be more suitable than others in predicting specific properties of hydrochar.

Ensemble modeling and decision fusion have several advantages in the context of HTC simulations. Ensemble methods can reduce the risk of overfitting through prediction aggregation from multiple models, resulting in accurate and reliable results [42]. Hence, this aggregation of models improves robustness by mitigating the impact of errors from any single model, ensuring more reliable results and the elimination of individual biases and errors, further improving performance [42]. Furthermore, ensemble modeling is particularly effective in dealing with

Table 2
Comparison between different machine learning models.

Model	Architecture and Mechanism	Advantages	Limitations	Reference
Decision tree	• Single learner • Root, internal, and leaves nodes • Gini impurity/information gain to manage splitting • Pruning to avoid overfitting	• Resultant tree is understandable and interpretable • Data preparation is easier • Can use multiple data types such as numeric or categorical	• Cannot handle missing data • Required class must be mutually exclusive • Prone to overfitting	[44,45]
Random forest	• Ensemble of decision trees • Bagging method (bootstrapping+aggregation) • Parallel ensemble algorithm	• Easy to implement • Few hyperparameters • Flexible and scale well for large datasets	• Prone to overfitting • Not suitable for small data sets • Computationally expensive than DT	[46–48]
XGBoost	• Ensemble of decision trees • Depthwise (level wise) growth of decision trees • Gradient boosting framework • Sequential ensemble algorithm • Loss function contains regularization function that prevents overfitting	• Does not need data normalization or feature scaling • Capable of handling missing data • Can output feature importance • Can handle large datasets • Less prone to overfitting	• Requires careful tuning of its numerous hyperparameters • Computationally expensive compared to DT and random forest	[48,49]
CatBoost	• Ensemble of decision trees • Symmetric growth of decision trees (balanced trees) • Gradient boosting • Ordered boosting) • Ordered target statistics encoding	• Excellent performance with categorical input data • Capable of handling missing data	• Complex in implementation and tuning of its numerous hyperparameters • Computationally expensive compared to DT and random forest	[46,48, 50]
LightGBM	• Ensemble of decision trees • Leafwise growth of decision trees • Sequential ensemble algorithm • Gradient one side sampling (GOSS) • Exclusive feature bundling • Histogram binning	• Fast compared to other ensemble learners • Low memory consumption • Higher accuracy than most boosting methods	• Can overfit with small training datasets • Leafwise splitting can lead to overfitting	[48,49,51, 52]
Support vector regression	• Single learner • Support vectors • Non-linear regression technique • Kernel functions handle non-linearity • Regularization parameter to control trade-off between achieving a low error on training data and minimizing the model complexity	• Less risk of overfitting • Able to handle multiple feature spaces	• Slow convergence for large datasets • Sensitive to data noise • The resulting model, weight, and impact of the variables are generally challenging to comprehend.	[45,47,53, 54]
K nearest neighbor (KNN) regression	• Single learner • Instant-based learning algorithm (lazy learner) that doesn't build a model during training. Instead, it stores all the training instances and performs computation during prediction. • It determines the similarity between instances using a distance metric • It uses majority voting or averaging of the K nearest neighbors • Non-parametric method, it makes no assumptions about the underlying data distribution	• Simple and fast algorithm • Handles noisy data • Handles missing data	• Computationally expensive for large datasets • Features are given the same importance • Oftenly requires standardization • Requires large memory as it stores all training data	[45,55, 56]
Gaussian process regression (GPR)	• Single learner • Probalistic model • Bayesian non-parametric approach • Incorporates uncertainty in prediction • Hyperparameters learned from data using marginal likelihood maximization	• High flexibility to model functions of any shape. • Powerful tool to model, explore, and exploit unknown functions	• Choosing the right kernel function might be complex and requires domain knowledge • Computationally expensive for large datasets	[57–59]
ANN	• Input layer, hidden layer, output layer • Learning rate • Forward and backpropagation • Activation and loss functions • Regularization	• Determine complex non-linear relationships between dependent and independent variables • Can be used for large datasets • Handles noisy data	• Computationally expensive • Black box characteristics	[45, 60–62]

biases that arise from heterogeneous data sources. In the context of HTC, where data originate from numerous studies with varying experimental protocols, the combined predictions from diverse models help smooth over inconsistencies and capture patterns, thus improving reliability across diverse conditions [43].

In this study, 8 different machine learning models (DT, RF, XGBoost, CatBoost, LightGBM, SVR, KNN, and ANN) were selected to evaluate their performance in predicting hydrochar properties from HTC. These models were chosen based on their wide and successful application in various fields where tabular data are modeled. Additionally, the majority of selected tree-based models were chosen specifically for their ability to handle experimental uncertainties through robustness to noisy or inconsistent data, uncertainty quantification via probability estimation, and ensemble-based predictions that effectively reduce variance. Table 2 presents the 8 different machine learning models, detailing their architecture, mechanisms, advantages, and disadvantages.

The objective is to assess the performance of each model individually, identifying which models could most accurately predict the desired hydrochar properties. Once the best-performing models are identified, they will be further utilized in ensemble learning and decision fusion approaches. This strategy aims to combine the strengths of the top models, enhancing predictive accuracy and robustness while providing a more comprehensive understanding of the hydrochar properties.

To manage model complexity, only the best-performing individual models for each target variable were selected for the ensemble decision fusion step. In this framework, the machine learning algorithms are trained independently, as there is no interaction between models during the training phase. Instead, their outputs are aggregated post-training through a decision fusion strategy. This avoided unnecessary overfitting and reduced computational error redundancy. Hyperparameter tuning was performed using cross-validation, and RMSE was the optimization criterion.

2.4.2. Single model hyperparameters optimization

For each single model, hyperparameter optimization was carried out based on data science best practices. Table S-1 in the supplementary materials illustrates the used models with their hyperparameters and ranges. This study evaluated both grid search [63] and Optuna-based Bayesian optimization [64] for hyperparameter tuning, using 5-fold cross-validation and minimizing the RMSE as the optimization objective.

Bayesian optimization methods optimize complex functions by building a probabilistic model that efficiently explores the search space. They balance exploration and exploitation, enabling the identification of optimal solutions with fewer steps compared to traditional methods [65]. In this study, Optuna's default Tree-structured Parzen Estimator (TPE) sampler was used. Each study was conducted for 50 trials, and no pruning or early stopping was applied. Optimization was executed on a personal HP ProBook laptop equipped with an Intel Core i7–1165G7 processor (2.80 GHz), 16 GB RAM, and integrated Intel Iris Xe graphics.

2.4.3. Ensemble learning and decision fusion

After simulating the individual models, the most accurate ones (determined by their performance on test data) were selected for ensemble learning and decision fusion. Fig. 4 outlines the multi-step process used to enhance predictive accuracy. Initially, in step 1, predictions from the most accurate *N* models are aggregated through averaging, a technique that leverages the strengths of each model while mitigating individual biases [31]. In step 2, these averaged predictions are combined with the original predictions from the best-performing models and fed into a stacked learner. The same eight models were evaluated as stacked learners, and the best-performing learner was selected for the second step. This stacked learner, which acts as a meta-model, synthesizes the information from multiple models, learning from previous prediction errors and leading to significantly improved accuracy and robustness in the final predictions. This process is iterative for every single output.

To ensure that the ensemble complexity was appropriate for the dataset size, model inclusion in each fusion stage was determined by empirical performance. Specifically, model selection was based on test set evaluation metrics (e.g., RMSE and adjusted R^2), and only the top models were included in the averaging step. This data-driven selection process avoided fixed or rule-based inclusion criteria and helped maintain methodological flexibility across output variables.

Additionally, hyperparameter tuning for all models was conducted using cross-validation embedded in Optuna-based Bayesian optimization or grid search. Moreover, stratified binning was used for train-test splitting to preserve the distribution of output variables across subsets, minimizing sampling bias during evaluation.

2.4.4. Evaluation metrics

Several evaluation metrics were employed in this study to assess the accuracy and variance of individual models and decision fusion models. The RMSE was used to evaluate the accuracy of the models, with RMSE offering an interpretable measure of prediction error in the same units as the target variable. Additionally, the R^2 and adjusted R^2 were used to assess the proportion of variance explained by the models. These metrics together provide a robust evaluation of both the precision and generalization of the models used in this study.

2.4.5. Model explainability

The optimized models are then integrated with the XAI method mentioned earlier in Section 2.3.2, *i.e.,* SHAP values determination. This action converts the black-box models to white-box ones that give a clearer picture of the modeling process by providing the weights of each input variable regarding the prediction of different output variables. As the modeling in this study is based on the fusion of various ML models, therefore, SHAP values are determined for each individual model, and then their average values are determined for the whole developed fused/ensemble model. In addition, these values were evaluated for the final fusion in order to determine the effect of models' combination on the final results.

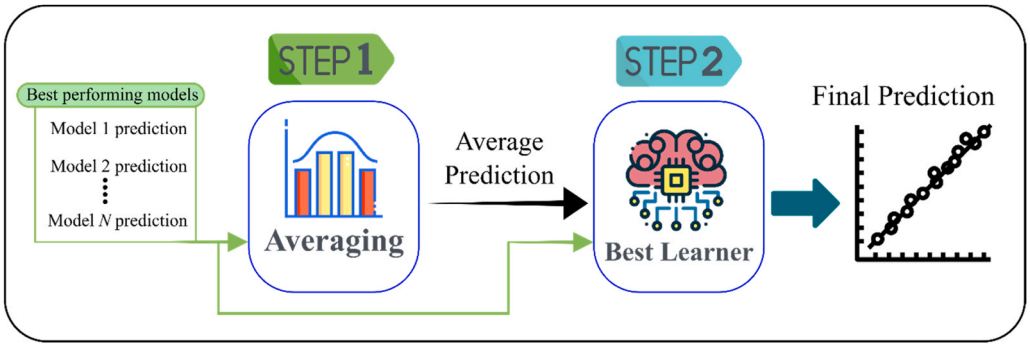

Fig. 4. Ensemble learning and decision fusion steps.

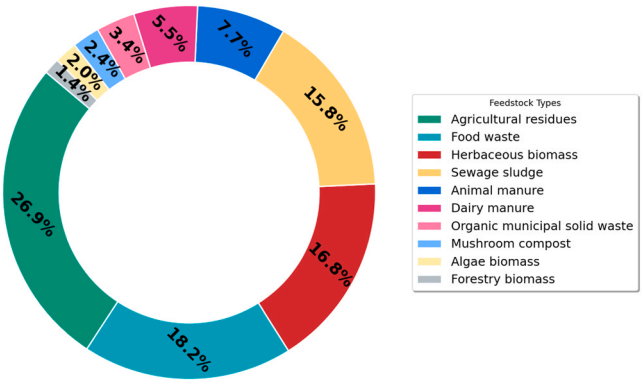

Fig. 5. Composition of the database.

2.5. Validation

In addition to the testing dataset, which represents 20 % of the full database, an independent validation set was used to further validate the developed ensemble model. This external validation dataset, consisting of 27 experimental data points collected from various literature sources and previous experiments done by the authors of this study, was not seen by the model during training or testing. The validation dataset and the corresponding results are provided in the supplementary materials.

2.6. Modeling aptitudes

The proposed modeling approach is highly feasible and reproducible using standard computational resources (HP ProBook laptop equipped with an Intel Core i7–1165G7 processor (2.80 GHz), 16 GB RAM, and integrated Intel Iris Xe graphics) and open-source tools such as Python, Optuna, and SHAP. Repeatability is ensured by a transparent methodology, including clearly defined preprocessing, feature selection, and model evaluation steps. Reliability is strengthened through the use of ensemble modeling, which reduces overfitting risks, and through validation with both internal testing and an external dataset of 27 unseen data points. The explainable AI (e.g., SHAP values) enhances transparency, making the approach suitable for scientific and practical use.

3. Results and discussions

3.1. Statistical analysis of the database

An analysis of the collected dataset was carried out to identify the categories of biomass within it. Fig. 5 illustrates the distribution of biomass categories within the database. The largest category was lignocellulosic biomass, constituting almost half of the dataset, which included agricultural waste, herbaceous biomass, and forestry biomass. The rest was constituted of non-lignocellulosic biomass, mainly sewage sludge and food waste, making almost 34 % together. Biomass such as animal manure, dairy manure, organic municipal solid waste, mushroom compost, and algae biomass had smaller proportions of the dataset, contributing to the remaining portion of the dataset.

The data distribution was also analyzed using a boxplot combined with a kernel density estimation for process conditions and elemental and proximate analyses of biomass and hydrochar, as illustrated in Fig. 6.

Fig. 6a presents the distribution of HTC operational parameters within the database. The reactor volume ranges from 30 to 500 mL, the water volume from 20 to 300 mL, and the stirring rate from 0 to 300 rpm. The B/W ratio varies between 0.04 and 0.25, while the moisture content is mainly at 0 %, indicating that the majority of HTC experiments are done in pre-dried conditions. The temperature varies from 150 to 300 °C, and the residence time is from 4 to 250 mins. It is also important to note that the outliers in these operational parameters are associated with specific experiments conducted under specific lab-scale conditions, which are not in the typical range of HTC. However, these outliers are not significant when considering the overall size of the database.

For the elemental analysis of biomass in Fig. 6b, carbon content on a dry basis ranges from 32 % to 55 %, hydrogen from 3.5 % to 7.5 %, nitrogen from 0.2 % to 7.5 %, and oxygen from 22 % to 50 %. These ranges align with those found in commonly used biomass types, as reported by Vassilev et al. [66,67]. Regarding proximate analysis, volatile matter content ranges from 60 % to 90 %, fixed carbon content from 0.2 % to 25 %, and ash content from 0.2 % to 32 %, all within the typical ranges for biomass, as also reported by Vassilev et al. [66,67]. It is important to note that elemental and proximate analysis distribution and their outliers reflect the diversity of biomass types rather than discrepancies among experiments in the literature. The majority of the outliers across the elemental and proximate analysis of biomass corresponded to sewage sludge, as illustrated in Figure S-1 in the supplementary file.

Fig. 6c shows the distribution of hydrochar properties within the database. For elemental analysis, hydrochar carbon content on dry basis ranges from 27 % to 76 %, hydrogen content from 3 % to 8 %, nitrogen content from 0.2 % to 6 %, and oxygen content from 2 % to 50 %. In the proximate analysis, volatile matter ranges from 24 % to 90 %, fixed carbon from 2 % to 57 %, and ash content from 1 % to 50 %. Hydrochar yield varies between 19 % and 95 %. The wide ranges observed in the hydrochar properties are attributed to significant variations in feedstock characteristics and HTC operational parameters included in the database. Other researchers have also noted a similar pattern in hydrochar properties distribution [21,27]. Additionally, the presence of outliers is primarily due to the inclusion of sludge, as shown in Figure S-1 in the supplementary file.

The Pearson correlation between input features and output targets was calculated as illustrated in Fig. 7. To improve clarity and brevity in this figure and later figures, parameter names were annotated with the suffix _B for features related to the biomass feedstock and _H for those related to the hydrochar, allowing for more concise labeling without loss of contextual meaning. Regarding the biomass and hydrochar relationship, the ash content is the most distinctive, as it had a significant negative correlation with most of the other elemental and proximate analyses. This is due to the fact that Eqs. (1) and (2) were utilized in the construction of the database, and the increase in the ash content leads to a decrease in the other contents. It is also important to mention that ash content differs significantly from one biomass to another, as the inorganic elements that constitute ash are different. However, ash is used instead of inorganic elements, as data on these inorganic elements is not available in most studies.

For process parameters, the temperature was the most influential parameter on most hydrochar properties, which is consistent with the literature [68]. When temperature increases, hydrochar yield, volatile matter, oxygen, and hydrogen significantly decrease as volatile matters are released, increasing dehydration reactions and lowering the hydrochar yield [69]. The RT did not have a strong correlation with hydrochar parameters within the used data range, similar to the findings of Zhu et al. [24]. This can be attributed to the fact that the majority of experiments were conducted in a relatively narrow range (4–250 mins) with few experiments conducted at much longer RT. Another possibility might be that linear correlation does not present RT well due to process non-linearity. The other process parameters, such as B/W ratio, reactor volume, water volume, and stirring rate, did not provide strong correlations with the studied hydrochar properties. It is worth noting that this step is only a preliminary one in order to provide a simple overview of possible relationships and dependencies between different variables in the available data. It is done to detect and remove any highly correlated input variables to ease the process of robust feature selection afterwards.

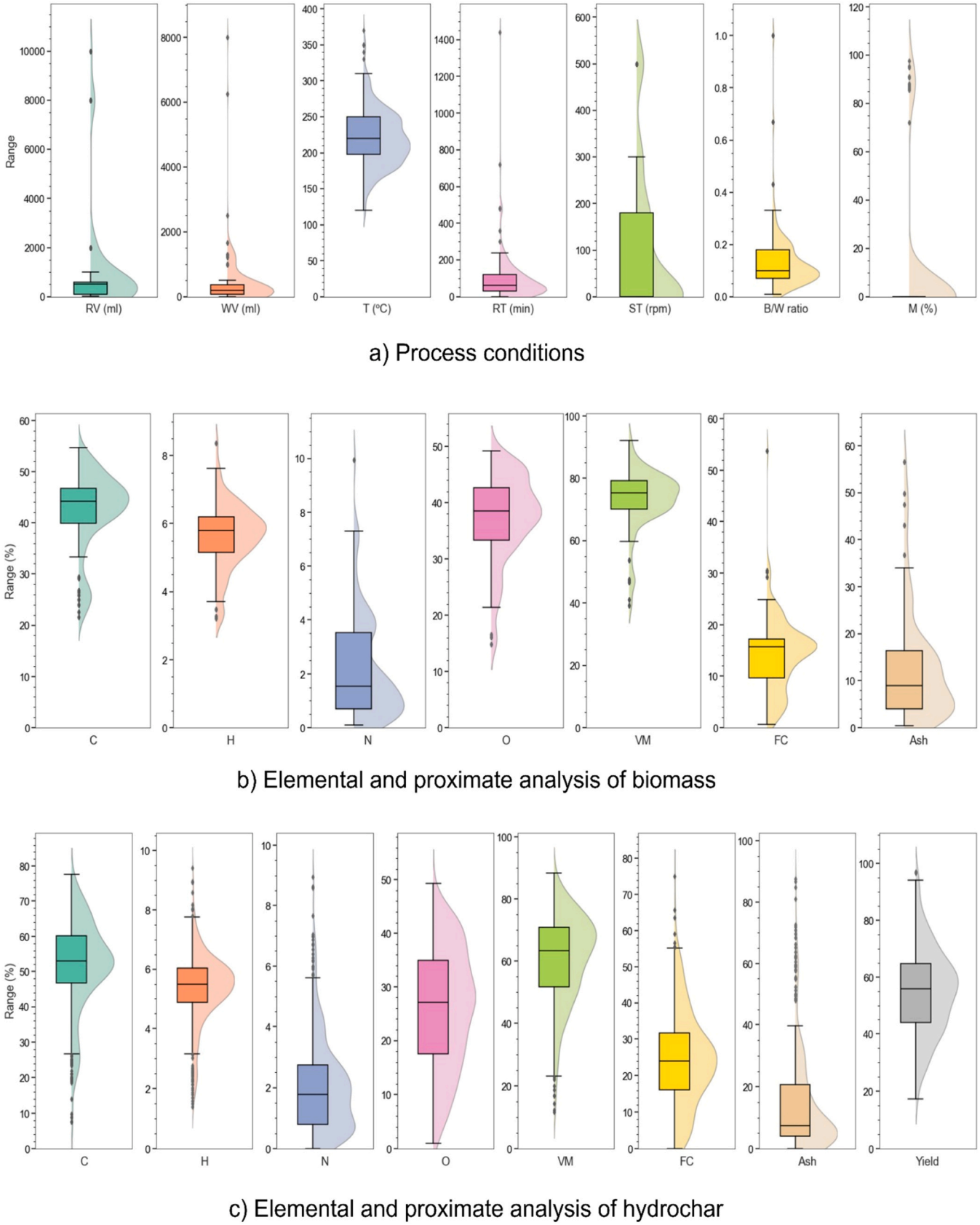

Fig. 6. Boxplot combined with a kernel density estimation for (a) process conditions (b) Elemental and proximate analysis of biomass (c) Elemental and proximate analysis of hydrochar.

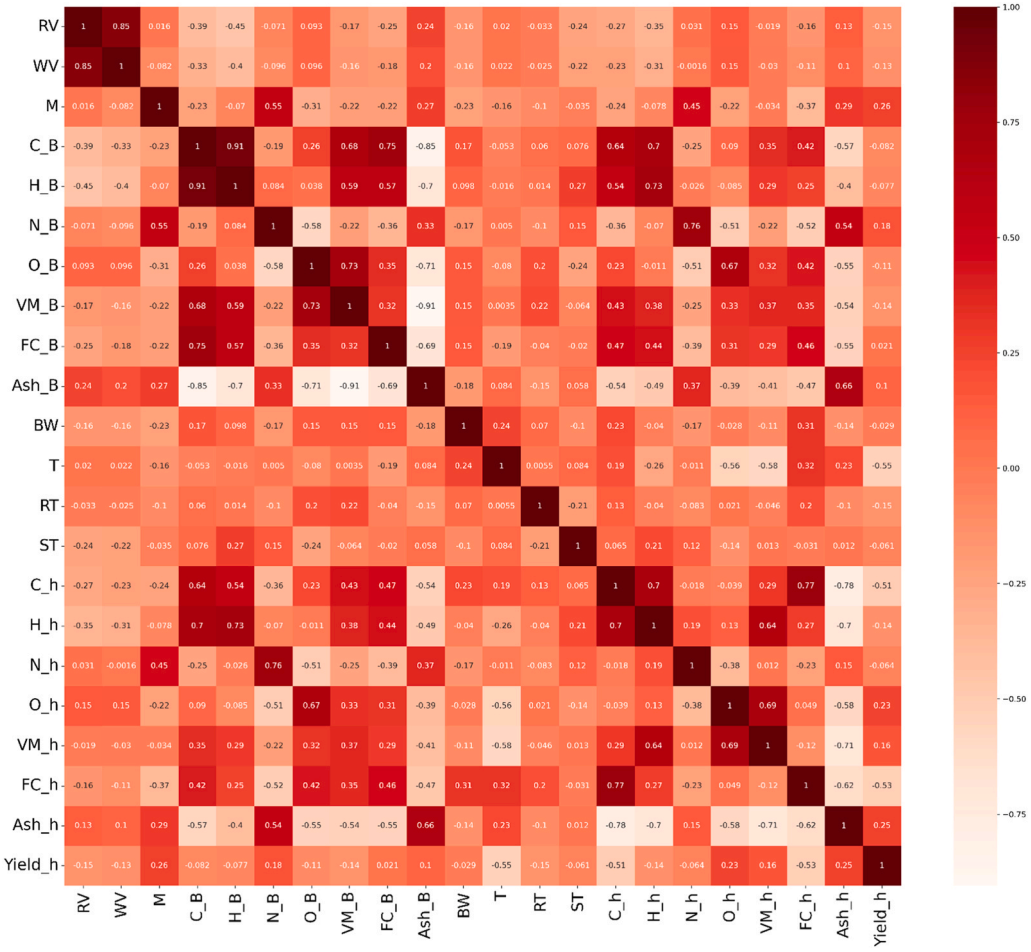

Fig. 7. Pearson correlation of the input and output parameters.

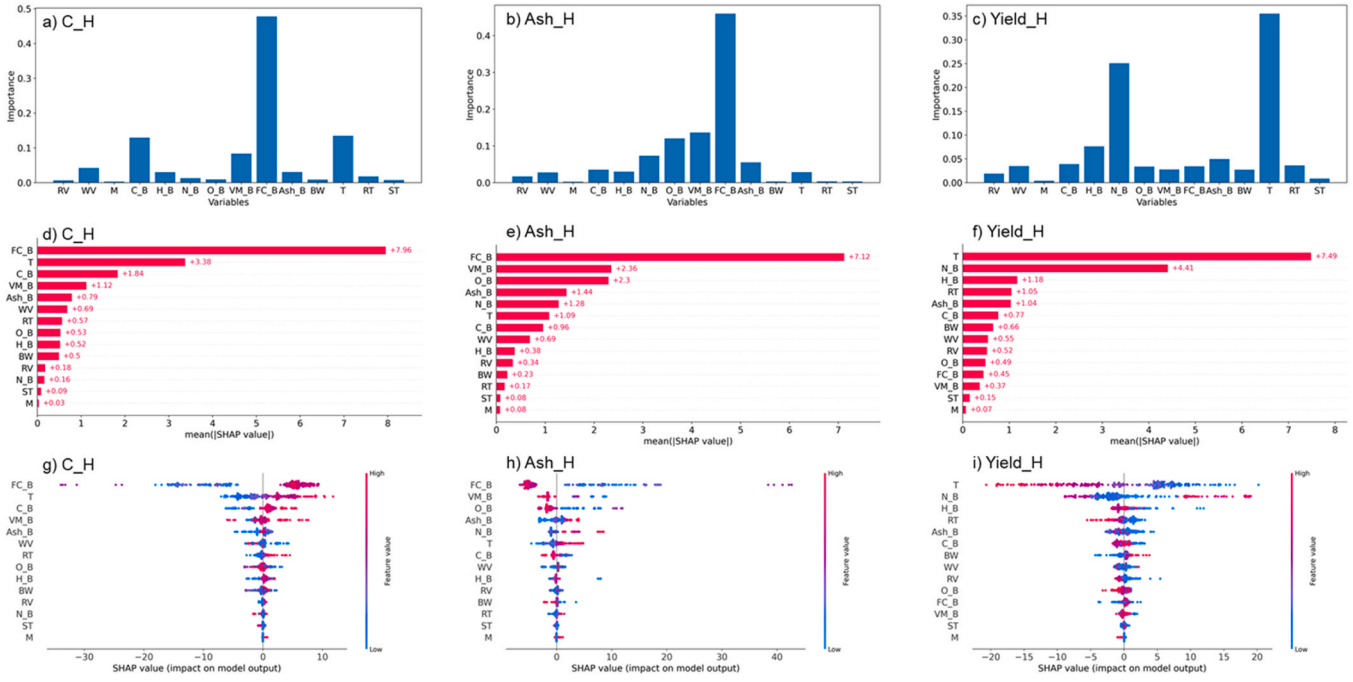

Fig. 8. (a,b,c) Feature importance (d,f,g) Average SHAP magnitude (h,i,j) SHAP influence.

3.2. Feature importance

Fig. 8 provides a comprehensive analysis of feature importance and explainability for C content, ash, and hydrochar yield, using feature importance plots (Fig. 8a, b, c) and SHAP values (Fig. 8d, f, g). Additional variables are presented in Figure S-2 in the supplementary file. From Fig. 8a, it is evident that the fixed carbon (FC) content in the biomass has the highest feature importance for the C content in hydrochar, followed by process temperature and the biomass C content. Fig. 8d and g together show the SHAP values (Fig. 8d) and their relative influence (Fig. 8g). Based on the combined assessment of these figures, biomass with higher FC and C content is more likely to produce hydrochar with higher carbon content. Additionally, higher process temperatures were observed to result in higher carbon content in the hydrochar, aligning with previous experimental studies [9,70,71].

From Fig. 8(b, e, h), it is clear that the ash content in the biomass has the highest feature importance and influence on the ash content in the hydrochar. Temperature plays the most significant role in determining hydrochar yield, as shown in Fig. 8a and b. Fig. 8c further indicates that temperature negatively affects the hydrochar yield, which aligns with previous experimental findings [72]. In addition, based on studying the correlation between different input variables, some highly correlated variables were not considered to have accurate predictions afterward. For example, the reactor vessel (RV) is highly correlated with water volume (WV), the input variable FC_B (calculated by difference) is highly correlated with both VM_B and Ash_B, and the input variable H_B is highly correlated with C_B. Hence, among the previously mentioned variables, only RV, FC_B, Ash_B, and C_B were considered. Although FC_B is calculated by difference, it was included due to its high feature importance (as shown by Fig. 8a and b) and strong correlation with VM_B and Ash_B, which makes it a valid choice. Moreover, ST were removed due to having a high percentage of missing data in the raw dataset and its lack of influence on the HTC process, which was demonstrated by previous studies [73,74]. Based on the combination of all these rules, methods of features selection, and with the guidance of human expertise, the most effective/important input variables, in total, are T, FC_B, O_B, N_B, Ash_B, C_B, RV, RT and BW. In the case of studying hydrochar yield the most important ones are T, O_B, N_B, Ash_B, C_B, RV, RT, and BW. For the Ash_h output variable, the most important input variables are T, FC_B, O_B, N_B, Ash_B and C_B. For the prediction of carbon content in the produced hydrochar, T, FC_B, N_B, Ash_B, C_B and RT are the most important variables to be used.

3.3. Model development and hyperparameters optimization

A summary of test predictions of the 8 studied single models is presented in Table 3. Additionally, the scatter plots of the training and testing predictions are illustrated in the supplementary file from Figure S-2 to S-9. The hyperparameters of the optimized most accurate models are presented in the supplementary material, Tables S1-S8. As mentioned before, these values were achieved through Bayesian optimization.

For predicting CHON content, most models, such as KNN, SVR RF, XGBoost, Light GBM, and Catboost, showed high accuracy with a test R^2 varying between 0.87 and 0.97, which is similar to the findings obtained in literature by Liu et al. [21] and Leng et al. [22]. One of the reasons for the high predictability of these hydrochar properties is the low uncertainty associated with their measurements in the experimental database and the proper choice of ensemble models that can deal with uncertainties in the used database. The ANN and decision tree models provided poor prediction, especially for predicting H and O. The relatively lower performance of the ANN is primarily attributed to the relatively small size of the database used in this study, as ANN models typically perform better with larger datasets and big data. The decision tree model, which is a weak learner, tends to overfit the training data, leading to significant variance in the test data and reduced prediction accuracy. Additionally, the slightly lower R^2 and higher RMSE values in the O predictions may be due to the fact that the experimental O data were calculated by difference.

For proximate analysis predication, similar observations were noticed. KNN, SVR random forest, XGBoost, Light GBM, and Catboost showed high accuracy with a test R^2 varying between 0.82 and 0.98. ANN, decision tree, and GPR provided lower accuracy.

The prediction accuracy for the hydrochar yield is low compared to other parameters, mainly because of the relatively high uncertainty in the hydrochar yield measurements, as previously described in Section 2.2.1. These inaccuracies in the data lead to reduced accuracy in subsequent predictions by models. XGBoost, CATBoost, and Light GBM provided the best test R^2 of 0.90 and 0.89, which is slightly higher than the findings of Liu et al. [21] and Djandja et al. [13] of 0.87 and 0.83 using XGBoost. The RMSE values for XGBoost and CatBoost in the current study were 4.73 and 5.17, respectively, whereas the lowest RMSE for hydrochar yield reported in the literature using a single model was 4.47 with XGBoost as illustrated in Table 1 by Liu et al., [21].

Overall, boosting ensemble models like CATBoost, XGBoost, and Light GBM, outperformed the other models across most of the properties studied. These models demonstrated better accuracy and robustness in capturing complex patterns in the data, making them more reliable for predicting the various properties under investigation.

3.4. Decision fusion models

As shown earlier in Fig. 3, the most accurate machine learning

Table 4
Models used in decision fusion for each hydrochar property.

Hydrochar properties	Models used
C	CATBoost, XGBoost, Light GBM, SVR, KNN, Random forest
H	CATBoost, XGBoost, Light GBM, KNN
N	CATBoost, XGBoost, Light GBM, SVR, KNN, Random forest
O	CATBoost, XGBoost, Light GBM, SVR
VM	CATBoost, XGBoost, Light GBM
FC	CATBoost, XGBoost, Light GBM, SVR
Ash	CATBoost, XGBoost, Light GBM, SVR, Random forest
Hydrochar yield	CATBoost, XGBoost

Table 3
Comparison of the different used single models for test data for R^2 and RMSE.

Model	KNN		SVR		ANN		Decision tree		Random forest		Light GBM		XGBoost		CATBoost	
Properties	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE
C	0.96	2.76	0.96	2.80	0.90	4.20	0.91	4.22	0.97	2.45	0.96	2.80	0.96	2.69	0.96	2.72
H	0.95	0.30	0.92	0.37	0.85	0.49	0.82	0.54	0.92	0.36	0.96	0.26	0.96	0.26	0.95	0.28
N	0.97	0.29	0.97	0.28	0.92	0.48	0.96	0.33	0.97	0.29	0.97	0.29	0.97	0.29	0.98	0.26
O	0.87	3.96	0.94	2.75	0.73	5.75	0.68	6.28	0.91	3.35	0.93	2.96	0.94	2.71	0.95	2.47
VM	0.83	5.41	0.87	4.77	0.84	5.51	0.56	8.75	0.82	5.51	0.92	3.75	0.88	4.48	0.91	3.94
FC	0.93	3.47	0.92	3.81	0.80	5.88	0.81	5.94	0.83	5.61	0.89	4.41	0.91	4.05	0.93	3.65
Ash	0.97	2.79	0.98	2.69	0.98	2.52	0.98	2.50	0.97	2.74	0.98	2.48	0.98	2.60	0.98	2.37
Hydrochar yield	0.78	7.21	0.74	7.78	0.68	8.49	0.56	10.10	0.80	6.76	0.83	6.46	0.90	4.73	0.89	5.17

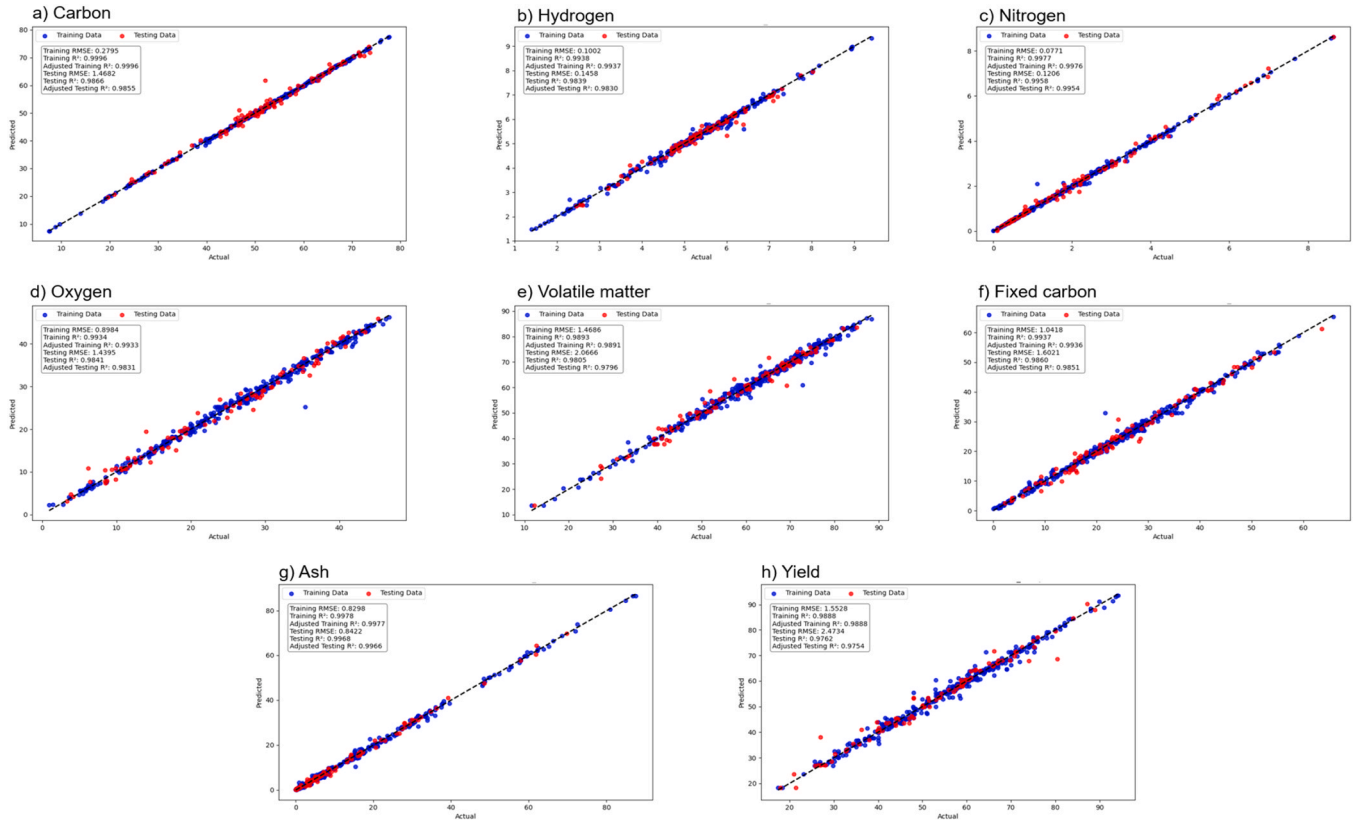

Fig. 9. Decision fusion models for different hydrochar properties.

models from the testing phase were selected for ensemble decision fusion learning. Table 4 lists the models used in the decision fusion process for each studied hydrochar property. Notably, CATBoost, XGBoost, and LightGBM were included in most decision fusion models due to their high predictive accuracy, as previously discussed in Section 3.3. Based on the results of the conducted trials using stacked learners, CATBoost was used as the stacked learner for N, O, VM, FC, and ash prediction, while XGBoost was used as the stacked learner for C, H and hydrochar yield.

As shown by Fig. 9, the decision fusion models demonstrated exceptionally high accuracy in predicting the test data, with adjusted test R^2 values ranging from 0.98 to 1.0 across all the parameters studied. To the best of the current author's knowledge, this level of predictive accuracy is significantly higher than that reported in other studies investigating similar output parameters, as illustrated by Table 1, particularly for challenging outputs, such as hydrochar yield. Previous studies predicting hydrochar yield have reported test R^2 values for their predictions ranging from 0.78 to 0.91, while in this study, the test R^2 was 0.98 [13,21,22,26,27]. Additionally, the RMSE for the hydrochar yield in this study was 2.47 using the decision fusion ensemble model, whereas RMSE values reported in the literature for individual models ranged from 4.47 to 7.83 [13,21,22,26,27].

The fusion of multiple ML models enhances prediction accuracy because, as previously mentioned, no single model can fully capture the

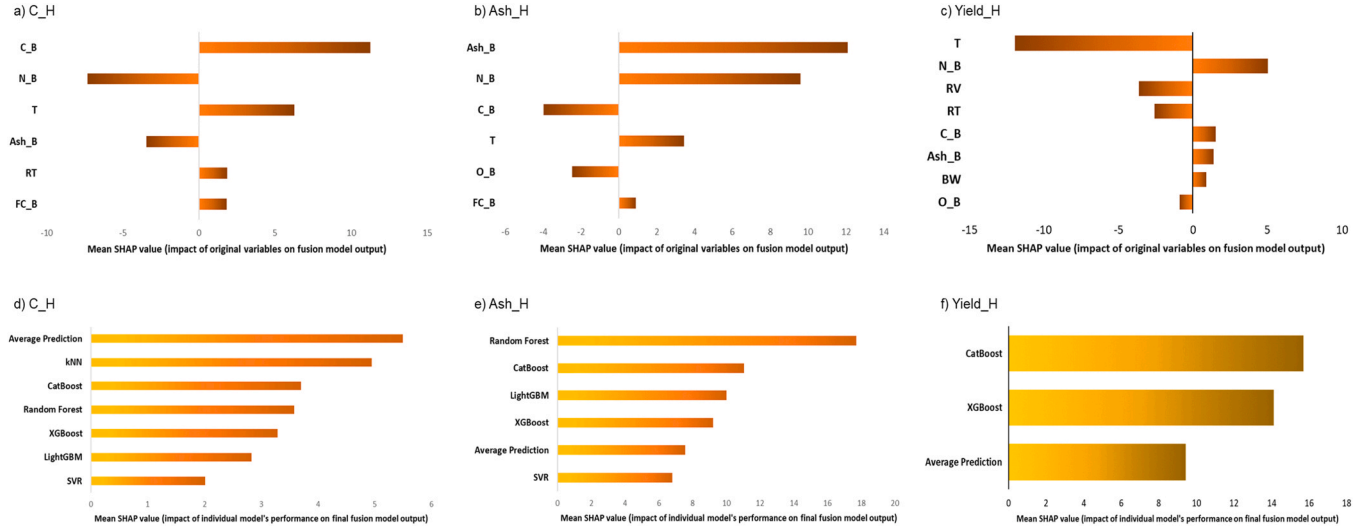

Fig. 10. Mean SHAP values for decision fusion.

complexity of the data [31]. Each ML model in this study is designed to handle different aspects of data complexity and identify underlying patterns within specific regions. By averaging and stacking the outputs of these models, a consensus is formed, allowing for a more comprehensive understanding of the dataset's overall pattern. This rationale served as the primary motivation for adopting ensemble learning in the present study.

The selection of specific models for each fusion configuration was entirely data-driven. Models were ranked based on their test set performance (adjusted R^2 and RMSE), and only the top performing models were retained for fusion. This ensured that the ensemble models remained compact and avoided unnecessary complexity, reducing the risk of overfitting.

To further validate the obtained results, an independent dataset compiled from both relevant literature and previous experimental work conducted by the authors was utilized. As seen in Figure S11, the developed models demonstrated high predictive accuracy, achieving R^2 values ranging from 0.90 to 0.99, thereby confirming the robustness and effectiveness of the proposed methodology.

While the currently developed decision fusion model is based on a dataset that predominantly includes pre-dried biomass, this limitation stems from the current practices in the existing HTC lab-scale literature. Nonetheless, the model architecture is designed to be modular and updatable. As more data become available, especially from experiments using raw biomass with its original moisture content, the model can be fine-tuned to improve its capabilities in the prediction of hydrochar properties from wet raw biomass. Additional relevant variables can also be incorporated into the feature space to enhance prediction accuracy under varying conditions. This flexibility ensures that the model remains robust and scalable as new insights and data emerge from both research and industrial applications.

3.5. Model explainability

The SHAP values resulting from the ensemble models based on averaging and final decision fusion are presented in Fig. 10. The selected outputs are the carbon and ash contents in the produced hydrochar, as well as the hydrochar yield. The carbon content in hydrochar increases when the carbon content in the feedstock increases. Moreover, the carbon content has an inverse relationship with the nitrogen and ash. The model's explainability further confirmed that temperature is the highest influencing operating condition, followed by the residence time. The carbon content tends to increase in the hydrochar at high temperature and residence time, mainly due to the dehydration reactions that increase the H/C and O/C ratios [75].

In the case of ash content in hydrochar, SHAP analysis revealed that the ash content in the feedstock is the most dominant predictor. The behavior of ash is highly dependent on the type and composition of the inorganic elements present, which can vary significantly based on factors such as biomass species, cultivation conditions, soil type, etc. In the case of the database used in this study, the observed SHAP patterns suggest that the inorganic elements tend to concentrate in the hydrochar across most of the studied samples; however, this result is totally dependent on the inorganic elements composition in the ash content in the specific samples used in the database and it might not reflect the case in all other biomass samples outside of the database.

In the case of hydrochar yield prediction, based on the developed XAI model, temperature exhibited the highest SHAP value across most target variables, highlighting its dominant role in governing HTC reactions. Mechanistically, increasing temperature accelerates thermal degradation pathways, such as dehydration and decarboxylation, which leads to reduced hydrochar yield but increases FC and energy density [75,76]. In addition, nitrogen content in the original biomass has the second most significance on the hydrochar yield where biomass with higher nitrogen contents tend to produce hydrochar with high hydrochar yield. With respect to the effect of different developed models on the final prediction accuracy, CatBoost seems to have the biggest share, which means that in future modeling tasks or relearning this model, will need extra care during its hyperparameters re-optimization in order to have more accurate and reliable predictions.

3.6. Future work

Although several studies have applied ML in the HTC of biomass, several challenges remain that hinder its further development. A primary issue is the inadequate reporting of methods and results within the current HTC literature. Key operational conditions of experiments are often not fully disclosed, even when they are critical to understanding the outcomes. Hence, this lack of information limits the broader applicability and usability of the data provided.

Another challenge is the inconsistency in the used data across many databases developed by previous researchers for predicting hydrochar using ML. For instance, in previously used databases in literature, within a single database, some entries may report the sum of CHONS and ash as 100 %, while others include only CHONS summing to 100 %. Such inconsistencies and the lack of standardized databases hinder the accumulation of collective knowledge and collaboration among researchers, and potentially lead to inaccurate conclusions. To address this, it is advisable to follow uniform standards while compiling data in databases.

Furthermore, it is recommended that experimental studies include more complete and comprehensive characterization analyses of known properties influencing HTC behaviour, such as the presence of inorganic elements and macromolecular compounds. Once sufficient reliable data are available in the literature, ML modeling techniques can be more effectively applied to extract further meaningful insights and enhance predictive accuracy. Besides that, it is recommended that future studies prioritize the use of wet biomass consistent with industrial HTC practices instead of pre-dried feedstock. This will improve the reliability of experimental data and enhance the predictive performance of ML models.

4. Conclusions

The study demonstrates the high predictive accuracy of decision fusion via developing diversified ensemble models in predicting hydrochar properties. These ensemble models achieved exceptional adjusted R^2 values (0.98–1.0) across various parameters, outperforming single models by capturing complex data patterns in different regions. Decision fusion proved especially effective for challenging outputs like hydrochar yield. In addition to accuracy, the study emphasizes the importance of model explainability, using XAI to explain the influence of key features and providing insights into how these features impact predictions, enhancing model transparency. Additionally, by applying the developed novel initial features selection methodology combining IML and XAI with human expertise in the early stages, fixed carbon content and process temperature were identified among the most influential factors for hydrochar carbon content, while temperature was crucial for the hydrochar yield. This novel feature selection approach had a good impact on the ensemble modeling learning tasks afterwards. Overall, decision fusion combined with XAI not only improved predictive performance but also contributed to a better understanding and explainability of the machine learning models in the HTC process.

CRediT authorship contribution statement

Omar M. Abdeldayem: Visualization, Project administration, Formal analysis, Writing – review & editing, Validation, Methodology, Data curation, Writing – original draft, Software, Investigation, Conceptualization. **Eslam G. Al-Sakkari:** Supervision, Conceptualization, Methodology, Writing – review & editing, Investigation. **Darwin Ortiz:** Investigation, Writing – review & editing, Formal analysis,

Software. **Capucine Dupont:** Writing – review & editing, Project administration, Conceptualization, Supervision, Methodology, Resources, Funding acquisition. **David Ferras:** Project administration, Writing – review & editing, Formal analysis, Supervision, Conceptualization. **Mohamed-Salah Ouali:** Conceptualization, Supervision. **Ahmed Ragab:** Resources, Writing – review & editing, Investigation, Supervision, Conceptualization. **Maria Kennedy:** Writing – review & editing, Conceptualization, Supervision, Resources.

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work the authors used Grammarly and ChatGPT 4.0 in order to improve the *English* of the Manuscript. After using these tools, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This research has been funded by the Research and Innovation Action project BIO4AFRICA implemented under European Union HORIZON 2020 (Grant Agreement No 101000762). Omar M. Abdeldayem would like to thank the Lamminga Fonds Foundation for supporting his research stay at Polytechnique Montréal.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.jece.2025.117826.

Data availability

Data will be made available on request.

References

[1] M. Volpe, A. Picone, F.C. Luz, M.C. at Mosonik, R. Volpe, A. Messineo, Potential pitfalls on the scalability of laboratory-based research for hydrothermal carbonization, Fuel (2022), https://doi.org/10.1016/j.fuel.2022.123189.
[2] A. Funke, F. Ziegler, Hydrothermal carbonization of biomass: A summary and discussion of chemical mechanisms for process engineering, Biofuels Bioprod. Bioref. (2010), https://doi.org/10.1002/bbb.198.
[3] Y. Shen, A review on hydrothermal carbonization of biomass and plastic wastes to energy products, Biomass. Bioenergy (2020), https://doi.org/10.1016/j.biombioe.2020.105479.
[4] P.J. Arauzo, M.P. Olszewski, X. Wang, J. Pfersich, V. Sebastian, J. Manyà, N. Hedin, A. Kruse, Assessment of the effects of process water recirculation on the surface chemistry and morphology of hydrochar, Renew. Energy (2020), https://doi.org/10.1016/j.renene.2020.04.050.
[5] M. Heidari, A. Dutta, B. Acharya, S. Mahmud, A review of the current knowledge and challenges of hydrothermal carbonization for biomass conversion, J. Energy Inst. (2019), https://doi.org/10.1016/j.joei.2018.12.003.
[6] S.K. Hoekman, A. Broch, C. Robbins, Hydrothermal carbonization (HTC) of lignocellulosic biomass, Energy Fuels (2011), https://doi.org/10.1021/ef101745n.
[7] S. Masoumi, V.B. Borugadda, S. Nanda, A.K. Dalai, Hydrochar: a review on its production technologies and applications, Catalysts (2021), https://doi.org/10.3390/catal11080939.
[8] C. Park, E.J. Kim, Comparison of microalgal hydrochar and pyrochar: production, physicochemical properties, and environmental application, Environ. Sci. Pollut. Res. Int (2024), https://doi.org/10.1007/s11356-023-31317-7.
[9] A.F. Qatarneh, C. Dupont, J. Michel, L. Simonin, A. Beda, C. Matei Ghimbeu, V. Ruiz-Villanueva, D. da Silva, H. Piégay, M.J. Franca, River driftwood pretreated via hydrothermal carbonization as a sustainable source of hard carbon for Na-ion battery anodes, J. Environ. Chem. Eng. (2021), https://doi.org/10.1016/j.jece.2021.106604.
[10] Q. Zhu, X. Liu, T. Hao, M. Zeng, J. Shen, F. Zhang, W. De Vries, Modeling soil acidification in typical Chinese cropping systems, Sci. Total Environ. (2018), https://doi.org/10.1016/j.scitotenv.2017.06.257.
[11] G. Ischia, L. Fiori, Hydrothermal Carbonization of Organic Waste and Biomass: A Review on Process, Reactor, and Plant Modeling, Waste Biomass. Valoriz. (2021), https://doi.org/10.1007/s12649-020-01255-3.
[12] X. Ying, 2019, An Overview of Overfitting and its Solutions, in: Journal of Physics: Conference Series. https://doi.org/10.1088/1742-6596/1168/2/022022.
[13] O.S. Djandja, S. Kang, Z. Huang, J. Li, J. Feng, Z. Tan, A.A. Salami, B.G. Lougou, Machine learning prediction of fuel properties of hydrochar from co-hydrothermal carbonization of sewage sludge and lignocellulosic biomass, Energy (2023), https://doi.org/10.1016/j.energy.2023.126968.
[14] S. González, S. García, J. Del Ser, L. Rokach, F. Herrera, A practical tutorial on bagging and boosting based ensembles for machine learning: algorithms, software tools, performance study, practical perspectives and opportunities, Inf. Fusion (2020), https://doi.org/10.1016/j.inffus.2020.07.007.
[15] L. Mu, Z. Wang, D. Wu, L. Zhao, H. Yin, Prediction and evaluation of fuel properties of hydrochar from waste solid biomass: machine learning algorithm based on proposed PSO–NN model, Fuel (2022), https://doi.org/10.1016/j.fuel.2022.123644.
[16] T.N. Kapetanakis, I.O. Vardiambasis, C.D. Nikolopoulos, A.I. Konstantaras, T. K. Trang, D.A. Khuong, T. Tsubota, R. Keyikoglu, A. Khataee, D. Kalderis, Towards engineered hydrochars: application of artificial neural networks in the hydrothermal carbonization of sewage sludge, Energies (2021), https://doi.org/10.3390/en14113000.
[17] J. Li, X. Zhu, Y. Li, Y.W. Tong, Y.S. Ok, X. Wang, Multi-task prediction and optimization of hydrochar properties from high-moisture municipal solid waste: Application of machine learning on waste-to-resource, J. Clean. Prod. (2021), https://doi.org/10.1016/j.jclepro.2020.123928.
[18] O.S. Djandja, P.G. Duan, L.X. Yin, Z.C. Wang, J. Duo, A novel machine learning-based approach for prediction of nitrogen content in hydrochar from hydrothermal carbonization of sewage sludge, Energy (2021), https://doi.org/10.1016/j.energy.2021.121010.
[19] M. Aghaaminiha, R. Mehrani, T. Reza, S. Sharma, Comparison of machine learning methodologies for predicting kinetics of hydrothermal carbonization of selective biomass, Biomass. Convers. Biorefinery (2023), https://doi.org/10.1007/s13399-021-01858-3.
[20] H.Y. Ismail, S. Shirazian, I. Skoretska, O. Mynko, B. Ghanim, J.J. Leahy, G. M. Walker, W. Kwapinski, ANN-kriging hybrid model for predicting carbon and inorganic phosphorus recovery in hydrothermal carbonization, Waste Manag (2019), https://doi.org/10.1016/j.wasman.2018.12.044.
[21] Q. Liu, G. Zhang, J. Yu, G. Kong, T. Cao, G. Ji, X. Zhang, L. Han, Machine learning-aided hydrothermal carbonization of biomass for coal-like hydrochar production: parameters optimization and experimental verification, Bioresour. Technol. (2024), https://doi.org/10.1016/j.biortech.2023.130073.
[22] L. Leng, J. Zhou, W. Zhang, J. Chen, Z. Wu, D. Xu, H. Zhan, X. Yuan, Z. Xu, H. Peng, Z. Yang, H. Li, Machine-learning-aided hydrochar production through hydrothermal carbonization of biomass by engineering operating parameters and/or biomass mixture recipes, Energy (2024), https://doi.org/10.1016/j.energy.2023.129854.
[23] X. Zhu, B. Liu, L. Sun, R. Li, H. Deng, Xiefei Zhu, D.C.W. Tsang, Machine learning-assisted exploration for carbon neutrality potential of municipal sludge recycling via hydrothermal carbonization, Bioresour. Technol. (2023), https://doi.org/10.1016/j.biortech.2022.128454.
[24] S. Zhu, N. Preuss, F. You, Advancing sustainable development goals with machine learning and optimization for wet waste biomass to renewable energy conversion, J. Clean. Prod. (2023), https://doi.org/10.1016/j.jclepro.2023.138606.
[25] O.S. Djandja, A.A. Salami, Z.C. Wang, J. Duo, L.X. Yin, P.G. Duan, Random forest-based modeling for insights on phosphorus content in hydrochar produced from hydrothermal carbonization of sewage sludge, Energy (2022), https://doi.org/10.1016/j.energy.2022.123295.
[26] J. Li, L. Pan, M. Suvarna, Y.W. Tong, X. Wang, Fuel properties of hydrochar and pyrochar: Prediction and exploration with machine learning, Appl. Energy (2020), https://doi.org/10.1016/j.apenergy.2020.115166.
[27] A. Shafizadeh, H. Shahbeik, S. Rafiee, A. Moradi, M. Shahbaz, M. Madadi, C. Li, W. Peng, M. Tabatabaei, M. Aghbashlo, Machine learning-based characterization of hydrochar from biomass: Implications for sustainable energy and material production, Fuel (2023), https://doi.org/10.1016/j.fuel.2023.128467.
[28] E.Y. Kaya, I. Ali, Z. Ceylan, S. Ceylan, Prediction of higher heating value of hydrochars using Bayesian optimization tuned Gaussian process regression based on biomass characteristics and process conditions, Biomass. Bioenergy (2024), https://doi.org/10.1016/j.biombioe.2023.106993.
[29] C. Chen, Z. Wang, Y. Ge, R. Liang, D. Hou, J. Tao, B. Yan, W. Zheng, R. Velichkova, G. Chen, Characteristics prediction of hydrothermal biochar using data enhanced interpretable machine learning, Bioresour. Technol. (2023), https://doi.org/10.1016/j.biortech.2023.128893.
[30] K.T. Peterson, V. Sagan, P. Sidike, E.A. Hasenmueller, J.J. Sloan, J.H. Knouft, Machine learning-based ensemble prediction of water-quality variables using feature-level and decision-level fusion with proximal remote sensing, Photogramm. Eng. Remote Sens. (2019), https://doi.org/10.14358/PERS.85.4.269.
[31] E.G. Al-Sakkari, A. Ragab, M. Amer, O. Ajao, M. Benali, D.C. Boffito, H. Dagdougui, M. Amazouz, Ensemble machine learning to accelerate industrial decarbonization: prediction of Hansen solubility parameters for streamlined chemical solvent selection, Digit. Chem. Eng. 14 (2025) 100207, https://doi.org/10.1016/j.dche.2024.100207.

[32] E.G. Al-Sakkari, A. Ragab, T.M.Y. So, M. Shokrollahi, H. Dagdougui, P. Navarri, A. Elkamel, M. Amazouz, Machine learning-assisted selection of adsorption-based carbon dioxide capture materials, J. Environ. Chem. Eng. (2023), https://doi.org/10.1016/j.jece.2023.110732.
[33] O.M. Abdeldayem, C. Dupont, D. Ferras, M. Kennedy, An experimental and numerical investigation of secondary char formation in hydrothermal carbonization: revealing morphological changes via hydrodynamics, RSC Adv. 15 (2025) 12723–12738, https://doi.org/10.1039/d4ra08995b.
[34] Berman, J.J., 2016. Understanding Your Data, in: Data Simplification. https://doi.org/10.1016/b978-0-12-803781-2.00004-7.
[35] A. Barredo Arrieta, N. Díaz-Rodríguez, J. Del Ser, A. Bennetot, S. Tabik, A. Barbado, S. Garcia, S. Gil-Lopez, D. Molina, R. Benjamins, R. Chatila, F. Herrera, Explainable Artificial Intelligence (XAI): concepts, taxonomies, opportunities and challenges toward responsible AI, Inf. Fusion (2020), https://doi.org/10.1016/j.inffus.2019.12.012.
[36] J. Cai, J. Luo, S. Wang, S. Yang, Feature selection in machine learning: a new perspective, Neurocomputing (2018), https://doi.org/10.1016/j.neucom.2017.11.077.
[37] R.C. Chen, C. Dewi, S.W. Huang, R.E. Caraka, Selecting critical features for data classification based on machine learning methods, J. Big Data (2020), https://doi.org/10.1186/s40537-020-00327-4.
[38] B.H. Menze, B.M. Kelm, R. Masuch, U. Himmelreich, P. Bachert, W. Petrich, F. A. Hamprecht, A comparison of random forest and its gini importance with standard chemometric methods for the feature selection and classification of spectral data, BMC Bioinforma. (2009), https://doi.org/10.1186/1471-2105-10-213.
[39] A.M. Musolf, E.R. Holzinger, J.D. Malley, J.E. Bailey-Wilson, What makes a good prediction? Feature importance and beginning to open the black box of machine learning in genetics, Hum. Genet (2022), https://doi.org/10.1007/s00439-021-02402-z.
[40] K. Roshan, A. Zafar, Utilizing Xai technique to improve autoencoder based model for computer network anomaly detection with shapley additive explanation (SHAP), Int. J. Comput. Netw. Commun. (2021), https://doi.org/10.5121/ijcnc.2021.13607.
[41] A.M. Salih, Z. Raisi-Estabragh, I.B. Galazzo, P. Radeva, S.E. Petersen, K. Lekadir, G. Menegaz, A perspective on explainable artificial intelligence methods: SHAP and LIME, Adv. Intell. Syst. 2400304 (2024) 1–8, https://doi.org/10.1002/aisy.202400304.
[42] O. Sagi, L. Rokach, Ensemble learning: a survey, Wiley Interdiscip. Rev. Data Min. Knowl. Discov. (2018), https://doi.org/10.1002/widm.1249.
[43] R. Polikar, Ensemble learning, Ensemble Mach. Learn. Methods Appl. (2012) 1–34.
[44] S. Tangirala, Evaluating the impact of GINI index and information gain on classification using decision tree classifier algorithm, Int. J. Adv. Comput. Sci. Appl. (2020), https://doi.org/10.14569/ijacsa.2020.0110277.
[45] S. Uddin, A. Khan, M.E. Hossain, M.A. Moni, Comparing different supervised machine learning algorithms for disease prediction, BMC Med. Inform. Decis. Mak. (2019), https://doi.org/10.1186/s12911-019-1004-8.
[46] C. Bentéjac, A. Csörgő, G. Martínez, A comparative analysis of gradient boosting algorithms. Artificial Intelligence Review, Springer Netherlands, 2021, https://doi.org/10.1007/s10462-020-09896-5.
[47] M. Liao, Y. Yao, Applications of artificial intelligence-based modeling for bioenergy systems: a review, GCB Bioenergy 13 (2021) 774–802, https://doi.org/10.1111/gcbb.12816.
[48] I.D. Mienye, Y. Sun, A survey of ensemble learning: concepts, algorithms, applications, and prospects, IEEE Access (2022), https://doi.org/10.1109/ACCESS.2022.3207287.
[49] J. Korstanje, Advanced forecasting with python, Adv. Forecast. Python (2021), https://doi.org/10.1007/978-1-4842-7150-6.
[50] Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A.V., Gulin, A., 2018. Catboost: Unbiased boosting with categorical features, in: Advances in Neural Information Processing Systems.
[51] Hancock, J., Khoshgoftaar, T.M., 2020. Performance of CatBoost and XGBoost in Medicare Fraud Detection, in: Proceedings - 19th IEEE International Conference on Machine Learning and Applications, ICMLA 2020. https://doi.org/10.1109/ICMLA51294.2020.00095.
[52] Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., Liu, T.Y., 2017. LightGBM: A highly efficient gradient boosting decision tree, in: Advances in Neural Information Processing Systems.
[53] D. Basak, S. Pal, D. Chanda Patranabis, Support vector regression, Neural Inf. Process. Lett. Rev. (2007) 11.
[54] B. Schölkopf, An introduction to support vector machines, Recent Adv. Trends Nonparametr. Stat. (2003) 3–17, https://doi.org/10.1016/B978-044451378-6/50001-6.
[55] F. Acito, in: F. Acito (Ed.), k Nearest Neighbors BT - Predictive Analytics with KNIME: Analytics for Citizen Data Scientists, Springer Nature Switzerland, Cham, 2023, pp. 209–227, https://doi.org/10.1007/978-3-031-45630-5_10.
[56] G. Guo, H. Wang, D. Bell, Y. Bi, K. Greer, KNN model-based approach in classification (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), Lect. Notes Comput. Sci. (2003), https://doi.org/10.1007/978-3-540-39964-3_62.
[57] Lindsten, F., Schön, T.B., Svensson, A., Wahlström, N., 2017. Probabilistic modeling – linear regression & Gaussian processes 29.
[58] R. Ranjan, B. Huang, A. Fatehi, Robust Gaussian process modeling using em algorithm, J. Process Control (2016), https://doi.org/10.1016/j.jprocont.2016.04.003.
[59] E. Schulz, M. Speekenbrink, A. Krause, A tutorial on Gaussian process regression: modelling, exploring, and exploiting functions, J. Math. Psychol. (2018), https://doi.org/10.1016/j.jmp.2018.03.001.
[60] O.I. Abiodun, A. Jantan, A.E. Omolara, K.V. Dada, N.A.E. Mohamed, H. Arshad, State-of-the-art in artificial neural network applications: A survey, Heliyon (2018), https://doi.org/10.1016/j.heliyon.2018.e00938.
[61] C.M. Bishop, Training with noise is equivalent to tikhonov regularization, Neural Comput. (1995), https://doi.org/10.1162/neco.1995.7.1.108.
[62] J. Schmidhuber, Deep Learning in neural networks: an overview, Neural Netw. (2015), https://doi.org/10.1016/j.neunet.2014.09.003.
[63] Yu, T., & Zhu, H. (2020). Hyper-parameter optimization: A review of algorithms and applications. *arXiv preprint arXiv:2003.05689*.
[64] T. Akiba, S. Sano, T. Yanase, T. Ohta, M. Koyama, Optuna: A next-generation hyperparameter optimization framework. Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining, 2019, pp. 2623–2631.
[65] B. Shahriari, K. Swersky, Z. Wang, R.P. Adams, N. De Freitas, Taking the human out of the loop: a review of Bayesian optimization, Proc. IEEE 104 (1) (2015) 148–175.
[66] S.V. Vassilev, D. Baxter, L.K. Andersen, C.G. Vassileva, An overview of the chemical composition of biomass, Fuel (2010), https://doi.org/10.1016/j.fuel.2009.10.022.
[67] S.V. Vassilev, D. Baxter, L.K. Andersen, C.G. Vassileva, T.J. Morgan, An overview of the organic and inorganic phase composition of biomass, Fuel (2012), https://doi.org/10.1016/j.fuel.2011.09.030.
[68] H.B. Sharma, A.K. Sarmah, B. Dubey, Hydrothermal carbonization of renewable waste biomass for solid biofuel production: a discussion on process mechanism, the influence of process parameters, environmental performance and fuel properties of hydrochar, Renew. Sustain. Energy Rev. (2020), https://doi.org/10.1016/j.rser.2020.109761.
[69] R.V.P. Antero, A.C.F. Alves, S.B. de Oliveira, S.A. Ojala, S.S. Brum, Challenges and alternatives for the adequacy of hydrothermal carbonization of lignocellulosic biomass in cleaner production systems: a review, J. Clean. Prod. (2020), https://doi.org/10.1016/j.jclepro.2019.119899.
[70] X. Chen, X. Ma, X. Peng, Y. Lin, Z. Yao, Conversion of sweet potato waste to solid fuel via hydrothermal carbonization, Bioresour. Technol. (2018), https://doi.org/10.1016/j.biortech.2017.10.096.
[71] H. Li, S. Wang, X. Yuan, Y. Xi, Z. Huang, M. Tan, C. Li, The effects of temperature and color value on hydrochars' properties in hydrothermal carbonization, Bioresour. Technol. (2018), https://doi.org/10.1016/j.biortech.2017.10.046.
[72] A.L. Pauline, K. Joseph, Hydrothermal carbonization of organic wastes to carbonaceous solid fuel – a review of mechanisms and process parameters, Fuel (2020), https://doi.org/10.1016/j.fuel.2020.118472.
[73] O. Abdeldayem, M. Al Noman, C. Dupont, D. Ferras, L. Grand Ndiaye, M. Kennedy, Hydrothermal carbonization of Typha australis: influence of stirring rate, Environ. Res. (2023).
[74] O.M. Abdeldayem, C. Dupont, D. Ferras, L.G. Ndiaye, M. Kennedy, Reconsidering lab procedures for hydrothermal carbonization of biomass: the impact of pre-drying and stirring, J. Anal. Appl. Pyrolysis 179 (2024) 106459, https://doi.org/10.1016/j.jaap.2024.106459.
[75] J.A. Libra, K.S. Ro, C. Kammann, A. Funke, N.D. Berge, Y. Neubauer, M.M. Titirici, C. Fühner, O. Bens, J. Kern, K.H. Emmerich, Hydrothermal carbonization of biomass residuals: a comparative review of the chemistry, processes and applications of wet and dry pyrolysis, Biofuels (2011), https://doi.org/10.4155/bfs.10.81.
[76] H.S. Kambo, A. Dutta, Comparative evaluation of torrefaction and hydrothermal carbonization of lignocellulosic biomass for the production of solid biofuel, Energy Convers. Manag (2015), https://doi.org/10.1016/j.enconman.2015.08.031.