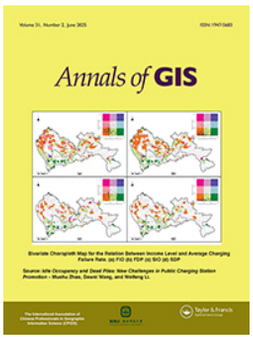


Annals of GIS

ISSN: 1947-5683 (Print) 1947-5691 (Online) Journal homepage: www.tandfonline.com/journals/tagi20

# Evaluating organic carbon in living and dead trees using GLCM features and explainable machine learning: insights from Italian national forest

Mehdi Fasihi, Alex Falcon, Giorgio Alberti, Luca Cadez, Francesca Giannetti, Antonio Tomao & Giuseppe Serra

**To cite this article:** Mehdi Fasihi, Alex Falcon, Giorgio Alberti, Luca Cadez, Francesca Giannetti, Antonio Tomao & Giuseppe Serra (27 Jun 2025): Evaluating organic carbon in living and dead trees using GLCM features and explainable machine learning: insights from Italian national forest, Annals of GIS, DOI: 10.1080/19475683.2025.2523737

**To link to this article:** https://doi.org/10.1080/19475683.2025.2523737



Published online: 27 Jun 2025.

ANNALS OF GIS
https://doi.org/10.1080/19475683.2025.2523737

Taylor & Francis
Taylor & Francis Group



# Evaluating organic carbon in living and dead trees using GLCM features and explainable machine learning: insights from Italian national forest

Mehdi Fasihi[a], Alex Falcon[a], Giorgio Alberti[b], Luca Cadez[b], Francesca Giannetti[c], Antonio Tomao[b] and Giuseppe Serra[a]

[a]Department of Mathematics, Computer Science and Physics, University of Udine, Udine, Italy; [b]Department of Agricultural, Food, Environmental and Animal Sciences, University of Udine, Udine, Italy; [c]Department of Agricultural, Food, Environmental and Forestry Sciences and Technologies, University of Florence, Florence, Italy

**ABSTRACT**

An accurate estimation of organic carbon (OC) in forest ecosystems is essential for understanding carbon dynamics and informing climate change mitigation strategies. This study presents a novel, explainable machine learning framework to estimate two key carbon pools: carbon sequestration in living trees (CSE) and carbon storage in standing deadwood (SDC). The methodology is structured into five key steps. First, we extract Gray-Level Co-occurrence Matrix (GLCM) texture features from LiDAR-derived canopy height models to quantify spatial heterogeneity in forest structure. Second, we integrate these GLCM metrics with vegetation indices (VIs), geomorphological variables, and weather data to create six distinct input configurations. Third, we train and evaluate teen models on each configuration to assess model performance and feature synergy. Fourth, we apply SHapley Additive exPlanations (SHAP) to the three models to transform them into an interpretable white-box model, identifying key predictors such as AVG_mean, SD_entropy, and SD_homogeneity. Finally, we assess model uncertainty using jackknife resampling and error bar analysis. The results indicate that CatBoost and Random Forest models deliver the highest performance for OC estimation. This study is the first to apply GLCM features for the joint estimation of CSE and SDC at a regional scale and to integrate explainable AI into forest carbon modelling. The framework provides a practical, transparent tool for forest managers, policymakers, and carbon monitoring systems, supporting high-resolution, scalable, and interpretable OC assessments.

**ARTICLE HISTORY**
Received 27 December 2024
Accepted 17 June 2025

**KEYWORDS**
Organic carbon; carbon sequestration; standing deadwood; explainable machine learning; remote sensing

## 1. Introduction

Forests are essential to the global carbon cycle, playing a crucial role in climate regulation by absorbing atmospheric carbon dioxide ($CO_2$) and storing it as organic carbon (OC). Organic carbon is held in various forest pools, including live trees, standing deadwood, litter, and soil. Globally, forests store approximately 861 gigatons (Gt) of carbon—42% in live biomass, 44% in soil, 8% in deadwood, and 5% in litter (Pan et al. 2011). This carbon sink absorbs an estimated 3.6 Gt C $year^{-1}$, offsetting a significant portion of annual anthropogenic emissions, which reached 37 Gt $CO_2$ in 2022 (Pan et al. 2024; Rehman and Lal 2023).

The OC in trees can be categorized into two primary parts. First, trees absorb atmospheric carbon dioxide ($CO_2$) through photosynthesis and convert it into organic compounds used to build their biomass (Schlesinger 1997). This process, known as carbon sequestration or Net Primary Production (NPP), effectively removes $CO_2$ from the atmosphere and stores it in wood, leaves, and roots of living plants, reducing atmospheric $CO_2$ concentration (Pan et al. 2011). Second, OC includes the carbon stored in lying and standing deadwood (Harmon et al. 1986). Even after a tree has died, it continues to store the carbon it accumulated during its lifetime (Harmon et al. 1986). The carbon stored in deadwood remains sequestered until the tree fully decays, continuing to influence the carbon balance of the forest (Lindenmayer, Laurance, and Franklin 2012). In particular, broken and unbroken snags (i.e. standing deadwood) play a crucial role in forest ecosystems not only in the carbon cycle, but also in biodiversity conservation, by providing important micro-habitats for wildlife, and in nutrient cycling, as they gradually decompose (Lindenmayer, Laurance, and Franklin 2012). Therefore, accurate estimation of OC sequestered yearly by live trees (CSE) as well as carbon stored in standing deadwood (SDC) in forests can provide policymakers with valuable insights

**CONTACT** Mehdi Fasihi Fasihi.mehdi@spes.uniud.it Department of Mathematics, Computer Science and Physics, University of Udine, Udine, Italy

into the best strategies to address greenhouse gas emissions, contributing to current and future climate change mitigation (Dai et al. 2021; Konda, Giri, and Mandla 2017).

Estimating OC in forest ecosystems has traditionally relied on two primary approaches: field-based measurements and ecosystem modelling. Field methods, such as eddy covariance techniques (Konda, Giri, and Mandla 2017), provide accurate, direct measurements of carbon stocks and fluxes. These approaches are valuable for calibration and validation but are often limited by their high cost, logistical complexity, and restricted spatial and temporal coverage (Masera et al. 2003). On the other hand, ecosystem models – such as the CENTURY model (Yona et al. 2020) and RothC model (Achard et al. 2012) – simulate carbon cycling processes over time, incorporating factors like decomposition, respiration, and biomass accumulation. While these models offer insights into long-term carbon dynamics, they typically require extensive input data and are highly sensitive to parameter uncertainties and assumptions about forest processes (Masera et al. 2003). As a result, despite their scientific robustness, both field-based and modelling approaches face practical challenges when applied to large-scale or repeated OC assessments, especially under varying environmental and management conditions.

Recently, remote sensing has gained significant importance as a tool for estimating and mapping forest biomass, and it is also widely acknowledged as an effective means to monitor sustainability in forestry (Estoque 2020; Fardusi, Chianucci, and Barbati 2017). This technique involves the use of satellites, aircraft, or drones to observe and analyse qualitative and quantitative characteristics of different ecosystems from a distance (U.S. Geological Survey 2023). Optical sensors, radar, and LiDAR systems are the main technologies employed for this purpose (Zhao et al. 2016). Additionally, remote sensing texture features, such as those derived from image texture analysis techniques, contribute to the comprehensive understanding of forest ecosystems and aid in tasks including species classification and habitat mapping (Zhou and Feng 2023). Among the established techniques for remote sensing texture feature processing, the Gray-Level Co-occurrence Matrix (GLCM) (Haralick, Shanmugam, and Dinstein 1973) stands as a cornerstone. In essence, GLCM quantifies spatial relationships between pixel values in an image, providing insights into texture characteristics (Haralick, Shanmugam, and Dinstein 1973). By analysing the frequency of pixel pairs with specific intensity values occurring at specified spatial relationships, GLCM unveils intricate texture patterns within remote sensing imagery. This method finds extensive application in diverse fields involving remote sensing data, including the estimation of growing stock (Li et al. 2019), total aboveground biomass (Charoenjit et al. 2015) in ecosystems, vegetation structure, and biomass distribution (Lu and Weng 2007). However, there are still two major limitations in the existing literature. First, there is a lack of research on methodologies that apply GLCM in combination with other remote sensing data, such as vegetation indices, geomorphological features, and weather parameters, for estimating OC stored in standing deadwood as well as OC sequestered yearly by trees. Second, black-box models limit their interpretability and thus their trustworthiness in real-world scenarios.

To address the first research gap, this study evaluates the integration of GLCM texture features within a machine learning framework, assessing their contribution to OC estimation performance across diverse algorithm families and data configurations. This research explores the integration of GLCM features into a ML framework and validate their usefulness with three different families of machine learning algorithms: (1) traditional models, including Support Vector Regression (SVR), K-nearest Neighbors (KNN), Multilayer Perceptron (MLP), and Linear Regression (LR); (2) tree-based models, such as Random Forest (RF), Gradient Boosting Decision Tree (GBDT), Extreme Gradient Boosting (XGBoost), and Categorical Boosting (CatBoost); and (3) ensemble-based models, namely Stack Ensemble (StackEns) and an AVG-Ensemble model. The AVG-Ensemble combines predictions from the top three individual performers – RF, CatBoost, and StackEns – by averaging their outputs to enhance robustness and predictive accuracy. Meanwhile, the StackEns model integrates predictions from multiple base learners using a meta-learner, allowing the framework to leverage the complementary strengths of diverse algorithms. These models are tested across six distinct input data configurations, representing unique combinations of variables, including: (1) Vegetation Indices (VIs) alone, (2) GLCM metrics alone, (3) a fusion of VIs and GLCM metrics, (4) VIs combined with geomorphological and weather parameters, (5) GLCM metrics combined with geomorphological and weather parameters, and (6) a comprehensive integration of VIs, geomorphological parameters, weather parameters, and GLCM metrics.

To overcome the second limitation concerning model interpretability, this study incorporates SHapley Additive exPlanations (SHAP) (Lundberg et al. 2020) to provide transparent insights into feature contributions and enhance trust in OC estimation results. SHAP values quantify the contribution of each input feature to individual predictions, thus enabling us to transform our model from a black-box to a white-box approach

(Lundberg et al. 2020). This shift enhances our ability to understand the influence of GLCM metrics on OC estimation and fosters greater trust in the model's predictions, aligning with recent efforts by Saleem et al. (Abdelhakim et al. 2024) and Abdelhakim et al. (Al Saleem, Harrou, and Sun 2024), who demonstrate the effectiveness of explainable machine learning in improving model transparency and predictive accuracy in industrial and environmental contexts. Table 1 provides a comparative summary of existing OC estimation approaches and highlights how the proposed solution addresses key gaps in the literature.

Furthermore, several practical pathways can directly integrate our findings into real-world forest monitoring systems. National forest agencies can adopt this framework to upscale plot-level carbon inventories by leveraging remote sensing data, which allows for cost-effective, high-resolution, and repeatable mapping of both CSE and SDC (Achard et al. 2012). Integration into operational workflows can be achieved through the deployment of these models within Geographic Information System (GIS) platforms, where spatial outputs such as predicted carbon densities and uncertainty maps can be overlaid with land management or biodiversity data layers (Yona et al. 2020). The explainability of the models, enabled through SHAP-based feature attribution, makes them particularly valuable for climate policy reporting, carbon credit verification mechanisms such as REDD+ programmes, and environmental compliance auditing (Asner et al. 2010). Practitioners can also use identified key predictors (e.g. AVG_mean, SD_entropy) to guide more efficient field sampling and sensor deployment, optimizing resource allocation towards high-priority areas. Furthermore, the incorporation of environmental covariates like geomorphology and weather variables ensures that the model remains robust across diverse forest types, management regimes, and climatic zones, making it adaptable to both temperate and tropical contexts. These capabilities make the framework a scalable, transparent decision-support tool that can significantly improve data-driven forest planning, conservation prioritization, and sustainable carbon monitoring.

The primary objective of this investigation, along with analysing distinct configurations, is to test and validate three key hypotheses.

**Hypothesis 1 (H1):**

Including GLCM metrics, which capture local variation in tree height, will improve the estimation of both CSE and SDC. GLCM metrics may help to better characterize local differences in tree size, thus improving tree growth estimation in forests with complex stand structures. Additionally, both broken and intact dead trees contribute to localized variations in tree height, indicating that incorporating GLCM as covariates could potentially enhance SDC estimates.

**Hypothesis 2 (H2):**

The integration of GLCM metrics and VIs is expected to yield more accurate estimates compared to models that rely solely on either GLCM or VIs. This enhancement stems from the complementary relationship between the spatial texture information provided by GLCM and the spectral insights offered by VIs. The combined effect is likely to be particularly significant for standing deadwood.

**Hypothesis 3 (H3):**

Incorporating geomorphological and weather parameters alongside GLCM metrics will further enhance the accuracy of OC estimation models, as these environmental factors are known to significantly influence tree growth and, therefore, carbon sequestration in forest ecosystems.

This study makes three key contributions:

**Table 1.** Comparative summary of existing OC estimation approaches.

| Approach | Data Used | Model Type | Advantages | Limitations |
|---|---|---|---|---|
| Field-based Sampling (Konda, Giri, and Mandla 2017) | On-site biomass, soil samples | Empirical | High accuracy, ground-truth | Labor-intensive, limited spatial coverage |
| Ecosystem Models (e.g. CENTURY, RothC) (Achard et al. 2012) (Masera et al. 2003) | Environmental variables | Process-based simulation | Captures biogeochemical dynamics | Requires extensive inputs, sensitive to assumptions |
| Satellite-based VIs (Cheng, Doosthosseini, and Kunkel 2022) (Uniyal et al. 2022), | NDVI, EVI, etc. | ML or regression | Wide coverage, temporal monitoring | Limited structural insight, poor under canopy |
| LiDAR-derived metrics (Csillik et al. 2019) (Vangi et al. 2023), | Tree height, canopy structure | ML (e.g. Random Forest) | Captures 3D forest structure | High cost, interpretability varies |
| GLCM features (limited studies) (Uniyal et al. 2022) (Csillik et al. 2019) (Jay Labadisos Argamosa et al. 2018), | Texture from imagery | ML (black-box) | Spatial heterogeneity insight | Not yet used for SDC; lacks transparency |
| Proposed (This Study) | GLCM, VIs, Geomorphology, Weather | Explainable ML | Multi-source fusion, interpretable, first to model CSE + SDC regionally | Requires complex preprocessing |

(1) **Novel integration of GLCM features for both CSE and SDC estimation**: This is the first known study to apply GLCM metrics for estimating both live and dead carbon pools across a regional forested landscape, expanding beyond the typical focus on live biomass or AGB.
(2) **Development of an explainable ML framework**: By using SHAP values, we provide transparency into feature importance and model behaviour, bridging the gap between predictive power and interpretability.
(3) **Comprehensive comparison of data configurations and algorithms**: Among the ten models evaluated, RF and CatBoost achieved the best performance. For CSE, RF attained an $R^2$ of 0.67 and RMSE of 0.85 (tC ha $-1$ yr $-1$), closely followed by CatBoost with an $R^2$ of 0.66 and RMSE of 0.86 (tC ha $-1$ yr $-1$). In the case of SDC, both models also performed well, with RF reaching an $R^2$ of 0.67 and RMSE of 2.73 (tC ha $-1$), while CatBoost achieved an $R^2$ of 0.65 and RMSE of 2.77 (tC ha $-1$). These results indicate that the choice of model plays a crucial role, with RF and CatBoost consistently delivering superior predictive accuracy across different configurations.

## 2. Background

### 2.1. Organic carbon estimation techniques

Accurately estimating OC levels is crucial for understanding soil health, ecosystem dynamics, and the carbon cycle and informing climate change mitigation policies. Various techniques have been developed to estimate OC across ecosystems, each with strengths, limitations, and specific applications (Brown 2002; Sun and Liu 2020).

Field-based approaches are widely employed to estimate OC in target ecosystems. These methods generally involve direct measurements of OC stocks, such as through soil sampling or biomass inventories, and may also include flux measurements, such as those obtained via Eddy covariance techniques (Konda, Giri, and Mandla 2017). Field-based techniques provide essential ground-truth data for validating model predictions and remote sensing estimates (Konda, Giri, and Mandla 2017). However, these measurements are often labour-intensive and time-consuming, and they may be limited in spatial and temporal coverage or require specific assumptions or conditions (e.g. the eddy covariance technique).

Another widely used approach for CSE estimation involves the use of ecosystem models. These models simulate the biophysical processes governing carbon cycling within ecosystems, incorporating factors such as organic matter decomposition, soil respiration, and biomass accumulation (Achard et al. 2012). Ecosystem models range from simple empirical models to complex process-based models that account for detailed physiological and ecological processes. Commonly used models include the CENTURY model for soil OC (Masera et al. 2003) and the RothC model for long-term carbon turnover in soils (Achard et al. 2012). While these models offer valuable insights into OC dynamics, they often require extensive data inputs and can be sensitive to uncertainties in parameters and assumptions (Masera et al. 2003).

Remote sensing technologies provide another powerful tool for estimating OC at large spatial scales. Satellite imagery and LiDAR (Light Detection and Ranging) products enable the mapping and monitoring of vegetation properties and soil physical characteristics, which are key indicators of OC stocks (Csillik et al. 2019). These data can be integrated with field measurements to produce spatially explicit maps of OC stocks and fluxes (Vangi et al. 2023). Advances in remote sensing, including higher spatial and spectral resolution sensors, have improved the accuracy of OC estimates (Uniyal et al. 2022). However, challenges remain in data interpretation, sensor calibration, and the integration of remote sensing data with ground-based measurements.

In summary, assessing OC is a complex task that involves integrating diverse data streams, methodologies, and tools. While ecosystem models, field-based measurements, and emerging technologies provide valuable insights, remote sensing stands out for its exceptional spatial coverage and efficiency in monitoring large areas repeatedly. By combining these varied approaches, we can deepen our understanding of OC dynamics within ecosystems and develop more robust strategies for climate change mitigation.

### 2.2. GLCM metrics in remote sensing and forestry applications

GLCM metrics have been employed across various domains, including assessing growing stock and total above-ground biomass in forest ecosystems. Uniyal et al. (2022) emphasized the quantification of above-ground biomass (AGB) in urban forests of Jodhpur, Rajasthan, India, to evaluate their significance in carbon sequestration. By employing machine learning (ML) regression algorithms, the authors effectively enhanced the predictive accuracy of AGB estimation, illustrating the potential of remote sensing and texture analysis in understanding urban forest dynamics. Csillik et al. (2019) underscore the critical role of tropical forests in

mitigating climate change and the challenges presented by human activities that convert these ecosystems from carbon sinks to sources. They introduced a novel method to create the first large-scale, high-resolution map of aboveground carbon stocks and emissions in Peru. This was achieved by integrating airborne LiDAR canopy height data with satellite imagery, utilizing a random forest regression model enhanced by GLCM texture features. Jay Labadisos Argamosa et al. (2018) investigated the application of the GLCM calculated from Sentinel-1 C-band data to estimate AGB in mangrove forests, a field that has been less explored compared to L-band synthetic aperture radar (SAR) data. Their research employed Random Forest regression and demonstrated that incorporating specific polarization features and principal components significantly enhanced the model's performance. Wijaya et al. (2010) focused on estimating AGB and stem volume in East Kalimantan's tropical forests using remote sensing (RS), GIS, and field data. A key contribution of their research was the integration of GLCM texture features with other predictors, such as Enhanced Thematic Mapper (ETM) band reflectance and VIs. This integration enabled a more nuanced analysis of forest structure, significantly enhancing the predictive models for AGB estimation. The analysis involved 1,460 sampling plots, revealing a slight decline in biomass over time. However, the study noted that RS estimates tended to be lower than those derived from GIS and field measurements, highlighting the complexities and uncertainties in modelling forest properties.

This literature review highlights the potential for improving AGB estimation accuracy by effectively capturing texture information while also emphasizing the need for further research. Notably, none of the previous studies have utilized GLCM features to estimate forest growth and, consequently, CSE. Furthermore, the existing literature primarily focuses on assessing living growing stock and biomass, with no reported results for standing deadwood in forests, which, as mentioned, is also important for biodiversity conservation.

### 2.3. Explainable models in forestry applications

To date, no research has utilized an explainable model to estimate OC while highlighting the interaction effects of variables like VIs, geomorphological parameters, weather factors, and GLCM. The most relevant studies in this area are represented by Ghafarian et al. (2022), Cheng et al. (Vangi et al. 2023), Li et al. (Jay Labadisos Argamosa et al. 2018), and Nguyen et al. (Asner et al. 2010). Ghafarian et al. (2022) utilized XGBoost and SHAP to predict forest microclimate in Brandenburg, Germany. Using standard meteorological data, they demonstrated XGBoost's effectiveness in predicting sub-canopy temperatures, crucial for ecological research. Their work highlights SHAP's role in interpreting complex models and aiding in landscape planning and forest management. Cheng, Doosthosseini, and Kunkel (2022) explored explainable AI (XAI) and feature unlearning (FUL) to improve the reliability of deep learning models in forestry. By integrating domain knowledge into model training, they improved classification accuracy and emphasized the importance of transparency in artificial intelligence (AI) models for forestry applications. Li et al. (2023) utilized a decision tree algorithm for modelling forest health due to its accuracy and interpretability. An objective weighted method based on the CRITIC criteria was applied to classify forest health using data from 132 forest samples. The findings indicated that species diversity is the key metric for assessing forest health and with their model achieving up to 90% accuracy, underscoring the value of interpretable machine learning in sustainable forest management. Nguyen, Kellenberger, and Tuia (2022) developed an explainable deep learning model for forest mapping, incorporating prior knowledge about forest definitions. Their model fosters trust by uncovering patterns in the data and allowing for manual adjustments, although the intermediate predictions continue to depend on black-box methods. This strategy enhances interpretability and holds the potential for adaptation to a range of land cover mapping tasks.

### 2.4. Identification of gaps

The investigation of OC sequestered by forests using ML models reveals a significant gap in the adoption of explainable models. While ML shows great potential in this area, the lack of transparency and interpretability has limited progress. Most research has overlooked the use of explainable models, particularly for understanding variable interactions in carbon estimation. This gap is crucial because understanding the complex relationships that influence OC is essential. Explainable models improve predictive accuracy and provide insights into the mechanisms behind predictions, enhancing interpretability and trustworthiness. In forestry, factors like VIs, geomorphological parameters, weather variables, and textural features such as GLCM metrics are influential, and integrating explainable models could be highly beneficial. These models allow researchers to explore the intricate interactions between variables and gain a deeper understanding of their contributions to OC. In addition, such models can be used to spatialize the aboveground carbon to produce wall-to-wall maps of

the target variables. Despite previous studies examining various factors affecting carbon sequestration, the full potential of explainable models remains underutilized. Bridging this gap by adopting advanced explainable models is imperative to improve prediction accuracy, reliability, and informed decision-making in forestry management.

## 3. Methodology

### 3.1. Dataset description

The study was carried out in the Autonomous Region of Friuli Venezia Giulia, located in northeastern Italy (Figure 1). It utilized a variety of data sources, including forest inventory records, digital elevation models, satellite imagery, and GLCM data, to examine OC within the region. By combining these diverse datasets, the research aimed to provide a comprehensive understanding of the factors influencing forest health and structure.

#### 3.1.1. Third Italian national forest inventory

The Third Italian National Forest Inventory (NFI) serves as a foundational dataset for this research. Established in 2015, the NFI employed a systematic sampling design across three phases, with data collection conducted between 2018 and 2019 (Gasparini and Papitto 2022). The first phase involved classifying land use and land cover through photointerpretation of high-resolution aerial images on a 1 km × 1 km grid that spans the entire country. In the second phase, a subset of these points was selected to assess qualitative forest characteristics (such as composition, forest category, management, etc.). The third phase involved conducting dendrometric field measurements on a subset of points categorized during the second phase. Data collection took place within circular plots with radii of 4 m and 13 m, determined based on specific tree diameter thresholds at 1.30 m height (Tabacchi, Di Cosmo, and Gasparini 2011). Diameters at breast height (DBH) were measured using dendrometric calipers, and tree heights were recorded with hypsometers. Additionally, the decay class for both standing and lying deadwood was documented. These measurements enabled the calculation of key attributes, including organic carbon (OC). For standing deadwood, distinct approaches were used to estimate biomass for unbroken and broken (truncated) trees. Biomass for unbroken dead trees was determined using the same method applied to living trees. For broken trees, volume was estimated using DBH and field-measured height, assuming a cylindrical shape. This volume was then converted to biomass using species

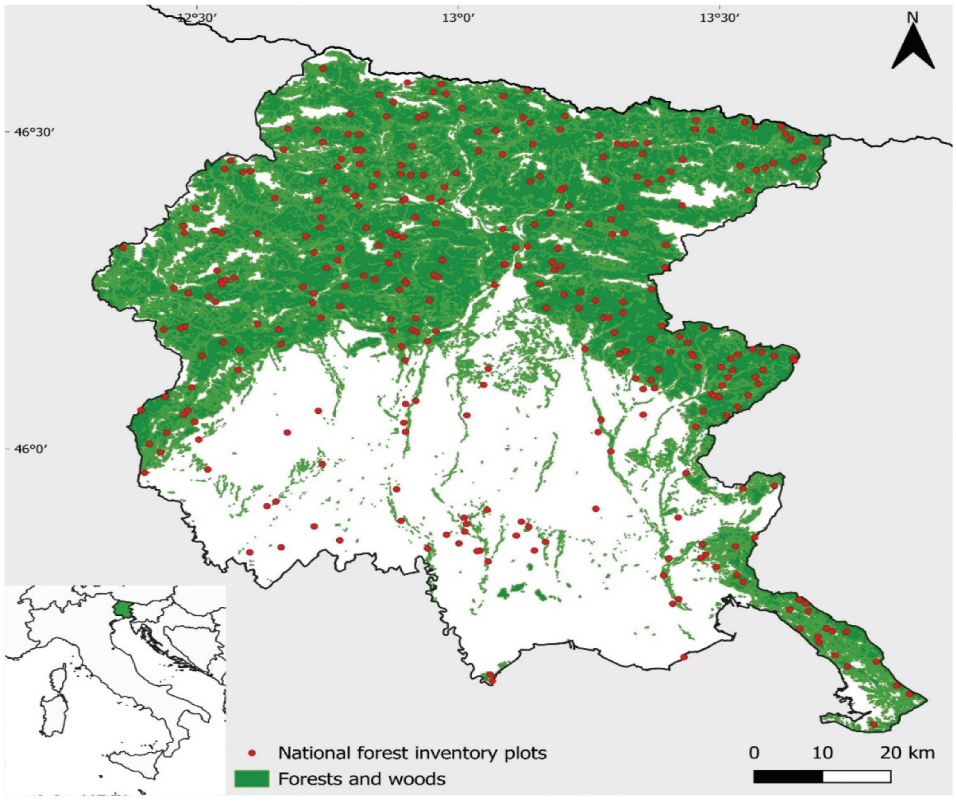



**Figure 1.** The Friuli Venezia Giulia region has wooded areas and national forest inventory plots. Reference system: WGS84.

group-specific conversion factors (conifers and broadleaves) and decay classes provided by Di Cosmo et al. (2013). The biomass of living and standing deadwood was converted into carbon stock by assuming a 50% carbon content in woody biomass (Krug et al. 2006). Carbon amounts, calculated at the plot level, were then standardized to tons per hectare (t $ha^{-1}$), based on the size of the survey plots (531 $m^2$ or 50 $m^2$, depending on the DBH threshold). Carbon sequestration efficiency (CSE) was derived from the annual volume increment, measured on a sub-sample of trees within each survey plot. These trees were scored at 1.30 m above the ground using an increment borer to determine the diameter increment over the last five annual rings (excluding the current year's ring). The annual volume increment was then calculated using allometric equations (Gasparini and Papitto 2022) and converted into carbon uptake following IPCC guidelines (Krug et al. 2006).

Our analysis was conducted on 279 inventory plots across the Friuli Venezia Giulia region (Italy), excluding 6 plots impacted by the 2018 Vaia storm. These plots were also used to derive predictor-related variables. By utilizing the recorded coordinates of the plot centres and the plot area (530 $m^2$), relevant variables were extracted through a geographic information system (GIS), as detailed in the subsequent sections. This comprehensive dataset offers essential measurements for evaluating forest biomass and understanding carbon dynamics. For additional details, please refer to Cadez et al. (2024; Fasihi et al. 2024).

### 3.1.2. Digital elevation models and derivatives

The data acquired from Airborne Laser Scanning (ALS) provides a crucial source of high-resolution information on forest tree structure. For this study, ALS data were collected by the Autonomous Region of Friuli Venezia Giulia. Specifically, the ALS data achieved an average point density of 16 points per square metre ($m^2$), with a reduced density of 10 points per $m^2$ observed above 1000 metres above sea level (a.s.l.) and a vertical accuracy of ±15 cm. The ALS data facilitated the generation of a Digital Terrain Model (DTM) with a resolution of 10 × 10meters, which served as the foundation for deriving key geomorphological predictors: elevation (ELE), slope (SLO), and aspect (ASP). The maximum, median, average, and standard deviation values for each predictor were calculated based on the plot area. Furthermore, high-resolution raster grid models derived from the point cloud – the Digital Surface Model (DSMf, first pulse) and the Digital Terrain Model (DTM, ground points) – were used to create a Canopy Height Model (CHM) by calculating the difference between them. The resulting CHM, with a spatial resolution of 0.5 metres, was subsequently used to extract the predictor variables described in Section 3.1.5

### 3.1.3. Satellite data

Satellite data play a critical role in understanding vegetation dynamics and forest health over time. In this study, we utilized Sentinel-2 Level 2A data collected from June to August during the tree-growing seasons aligned with the NFI sampling period. Data were processed using Google Earth Engine (Gorelick et al. 2017), and a median composite was generated for each spectral band after filtering for cloud presence. Four key VIs were derived from this image collection:

(1) Normalized Difference Vegetation Index (NDVI): widely used in forestry, NDVI estimates various vegetation properties, including leaf area index (LAI), plant productivity, aboveground biomass, and carbon stock (Huang et al. 2021), (Liang et al. 2022).
(2) Normalized Difference Infrared Index (NDII): a reflectance measurement sensitive to changes in tree water content, applicable in agriculture, forest canopy monitoring, and the detection of stressed vegetation (Buschmann 1993).
(3) Enhanced Vegetation Index (EVI): closely associated with Gross Primary Production (GPP), EVI measures vegetation greenness while accounting for atmospheric effects and canopy background interference. This makes it especially useful in areas with dense vegetation (Sims et al. 2006), (Huete et al. 2025).
(4) Green Normalized Difference Vegetation Index (GNDVI): a modified version of NDVI that is more sensitive to variations in chlorophyll content (Gitelson and Merzlyak 1998).

Also, for the VIs (i.e. NDVI, NDII, EVI, and GNDVI), we extracted the median values for each plot.

### 3.1.4. Weather data

To account for the influence of weather conditions on forest dynamics and growth, we incorporated temperature and rainfall data generated through the EURO-CORDEX project, representing the latest advancements in regional climate modelling at the European scale and providing high spatial resolution. The weather data were supplied in the form of NetCDF files, with temperature represented at a grid resolution of 500 × 500 m and rainfall at 5 × 5 km. The temperature data included average soil temperatures for both the summer (TEMP summer: June, July, August) and spring (TEMP spring: March,

April, May) seasons, with values assessed for maximum, minimum, and average were extracted as predictor based on the area of the plots. Precipitation data were also considered for the same seasonal periods (PREC summer and PREC spring), focusing on the average rainfall during these months that was extracted for each plot as predictor variables. Weather data were essential for understanding the temporal variations that affect forest health and tree growth, allowing us to contextualize our findings within changing environmental conditions.

### 3.1.5. GLCM metrics selection

The Gray Level Co-occurrence Matrix (GLCM) was developed by Haralick, Dinstein, and Shanmugam (1973) and provides a statistical framework for capturing the joint distribution of grey levels among pixels with specified spatial relationships (Li et al. 2021). This method is widely used for texture feature extraction, enabling the calculation of key image attributes. By leveraging GLCM, we can enhance our ability to identify and quantify image textural characteristics, forming a solid foundation for estimating OC and conducting other environmental assessments (Zhou and Feng 2023). Specifically, GLCM (Haralick, Dinstein, and Shanmugam 1973) was computed from the CHM using the GLCM R package, which uses a $3 \times 3$ processing window in all directions (Huang et al. 2021). For each inventory plot, we calculated a set of textural variables, i.e mean, variance, homogeneity, contrast, dissimilarity, entropy, and second moment. Then, average and standard deviation at the plot level were calculated for each textural variable. The full list of GLCM-related variables is reported in Table 2. This approach provides a comprehensive understanding of image texture and simplifies the extraction of relevant features necessary for further analysis (X. Li et al. 2021). Table 2 presents the formulas for GLCM attributes, where $i, j$ are row and column numbers, N is the number of gray levels in the image, $P(i,j)$represents the normalized GLCM value at position $(i,)$, $\mu$ is the mean intensity of the image. Normalization was deliberately not applied to the GLCM features to preserve their physical interpretability and maintain consistency with the original CHM-derived texture metrics. Since the analysis utilized tree-based algorithms such as XGBoost and CatBoost, which are inherently insensitive to feature scaling, normalization was neither necessary nor beneficial for predictive performance. Furthermore, SHAP values were employed to interpret model predictions, offering insights into feature importance while retaining features in their original scale. This approach, based on the method introduced by Lundberg, Allen, and Lee (2022), reinforced the decision to forgo normalization, as it allowed for both accurate modelling and meaningful interpretation of texture metrics.

### 3.1.6. Statistical analysis of data

This section provides an in-depth statistical analysis of the two primary target variables: CSE and SDC. Understanding their distributional properties is essential for guiding appropriate modelling strategies and ensuring a valid interpretation of results. Table 3 summarizes the descriptive statistics for both variables. The mean value for CSE is relatively low at 1.68 tC $ha^{-1}$ $yr^{-1}$, with less variability compared to SDC, which has a higher mean of 5.07 tC $ha^{-1}$. Both variables exhibit positive skewness and moderate kurtosis, indicating right-skewed distributions with a tendency towards extremely high values. The coefficients of variation—0.72 for CSE and 0.78 for SDC – suggest moderate relative variability in each dataset. Notably, SDC demonstrates a considerably wider overall range and interquartile range, reflecting greater dispersion in values.

**Table 2.** The formula for GLCM attributes, where $i, j$ are row and column numbers, N is the number of gray levels in the image, $P(i,j)$ represents the normalized GLCM value at position $(i.J)$, $\mu$ is the mean intensity of the image.

| GLCM attributes | Description | formula |
|---|---|---|
| Mean (MEAN) | The average grey level intensity of the image. | $\frac{1}{N^2}\sum_{i,j=0}^{N-1} P_{i,j}$ |
| Homogeneity (HOM) | Measures how uniform or regular the texture is. Higher values indicate a more uniform texture. | $\sum_{i,j=0}^{N-1} i\frac{P_{i,j}}{1+i-j^2}$ |
| Contrast | Measures local variations in intensity between neighbouring pixels. Higher values indicate greater differences. | $\sum_{i,j=0}^{N-1} iP_{i,j}(1-j)^2$ |
| Dissimilarity | Measures the average difference in intensity between neighbouring pixels. Reflects texture diversity. | $\sum_{i,j=0}^{N-1} iP_{i,j}[1-j]$ |
| Entropy | Quantifies the randomness or unpredictability of pixel intensities in the image. | $\sum_{i,j=0}^{N-1} iP_{i,j}\left(-lnP_{i,j}\right)$ |
| Variance | Measures the spread or dispersion of pixel intensities around the mean intensity value. | $\frac{\sum_{i,j}\left(x_{i,j}-\mu\right)^2}{n-1}$ |
| Second Moment | Quantifies the overall homogeneity or uniformity of the image texture. Higher values indicate a more uniform texture. | $\sum_{i,j=0}^{N-1} iP_{i,j}{}^2$ |

**Table 3.** Descriptive statistics of the target variables CSE and SDC. The table includes the average (avg), standard deviation (std), minimum (min), maximum (max), first quartile (Q1), median (Q2), third quartile (Q3), skewness, kurtosis, coefficient of variation (CV), range, and interquartile range (IQR), offering a comprehensive summary of central tendency, dispersion, and distribution characteristics.

| Statistic | CSE (tC ha-1 yr-1) | SDC (tC ha-1) |
|---|---|---|
| Avg | 1.682044775 | 5.067998254 |
| Std | 1.205472141 | 3.9685422 |
| Min | 0.011211573 | 0.02289625 |
| Max | 0.737296892 | 2.167823779 |
| Q1 | 1.438955245 | 4.012260174 |
| Q2 | 2.373536051 | 6.961895687 |
| Q3 | 6.469211801 | 20.57107738 |
| Skewness | 1.003962 | 1.379363 |
| Kurtosis | 0.88756 | 2.176127 |
| Coefficient of Variation (CV) | 0.716671 | 0.783059 |
| Range | 6.458 | 20.54818 |
| IQR | 1.636239 | 4.794072 |

To further investigate the presence of extreme values, an outlier analysis was conducted using the Interquartile Range (IQR) method (Figure 2). This approach identifies values lying beyond 1.5 times the IQR from either quartile as potential outliers. For CSE, approximately 3.23% of the data points were classified as outliers. For SDC, 3.58% were identified as outliers. These relatively low percentages indicate that both variables are largely free from extreme anomalies, contributing to the robustness of the dataset. Boxplots were employed to represent the data distribution and highlight potential outliers.

Additionally, histograms overlaid with Kernel Density Estimation (KDE) curves (Figure 3) were created to better understand the shape and spread of each distribution. The CSE distribution is strongly right-skewed, with a clear concentration of lower carbon stock values and a tapering tail extending towards higher values. This suggests that areas with high carbon stocks are less prevalent in the dataset. The KDE curve confirms this pattern and reveals a single dominant peak, implying a unimodal distribution. The SDC variable shows a similarly skewed distribution, albeit with a slightly broader spread and a more pronounced tail. This indicates higher variability in structural diversity across plots. The KDE visualization also reveals a clear mode, suggesting possible clustering of values potentially linked to different forest types or ecological conditions.

### 3.2. Machine learning models

This section presents the machine learning models explored in this study, grouped into four categories: Traditional Machine Learning Models, Tree-Based Models, and Ensemble Models. This structure reflects the diversity in modelling techniques and highlights the unique contributions of each category to regression tasks.

#### 3.2.1. Traditional machine learning models

This category comprises individual learners based on linear functions and distance metrics. These models are relatively simple and interpretable.

- **Support Vector Regression (SVR)**: Utilizes kernel functions to map inputs into higher-dimensional spaces, allowing it to model nonlinear patterns while controlling model complexity (Cortes, Vapnik, and Saitta 1995).

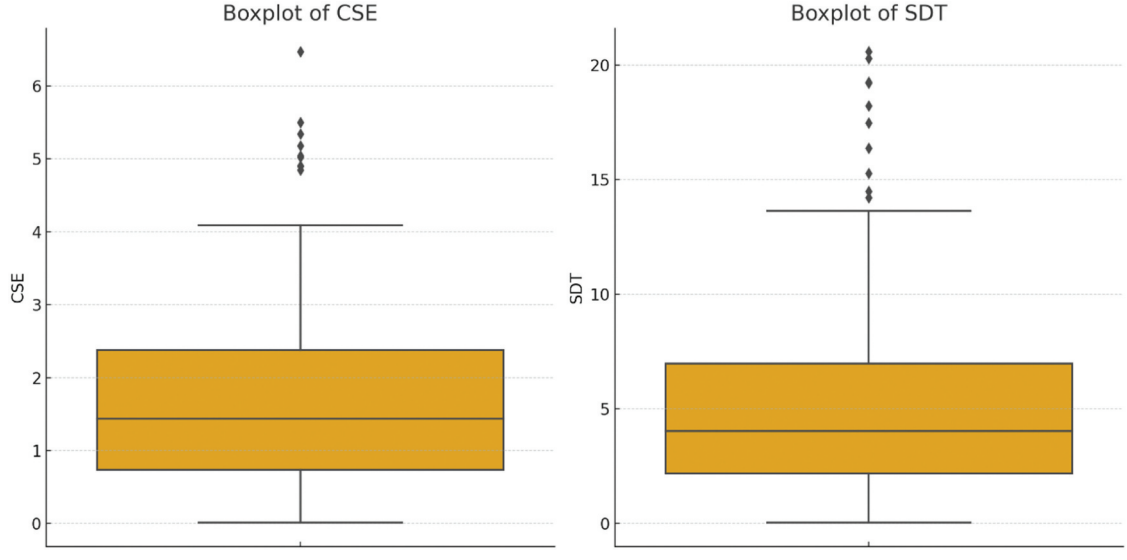



**Figure 2.** Boxplots of the target variables CSE and SDT showing the distribution and presence of outliers. Outliers are indicated by individual points beyond the whiskers, based on the IQR method.

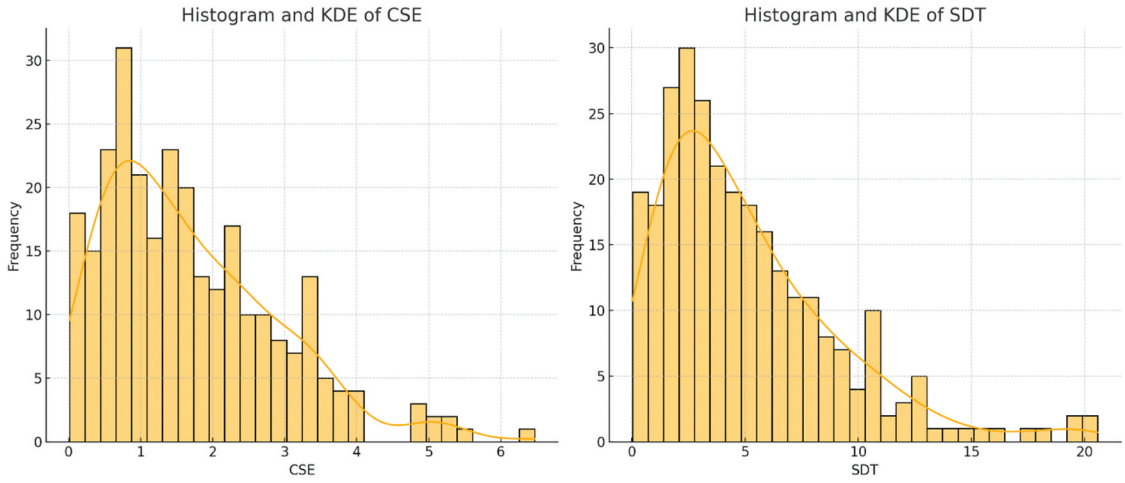



**Figure 3.** Distribution of target variables (CSE and SDC). Histograms with kernel density estimates (KDE) illustrate the distribution of CSE and SDC across the dataset. Both variables exhibit right-skewed distributions, indicating a higher concentration of lower values and a long tail towards higher values.

- **K-Nearest Neighbors (KNN)**: A non-parametric model that predicts the outcome based on the average of k nearest samples. The value of k is critical and affects model sensitivity and generalization (Xue and Su 2017).
- **Multilayer Perceptron (MLP)**: A feedforward neural network model with one or more hidden layers that uses activation functions and is trained via backpropagation to capture nonlinear relationships between inputs and outputs (Gao et al. 2018).
- **Linear Regression (LR)**: LR is a fundamental statistical method used to model the relationship between a dependent variable and one or more independent variables by fitting a straight line to the observed data. It is commonly used for predicting numerical outcomes and understanding the strength and direction of associations between variables (Montgomery, Peck, and Vining 2021).

### 3.2.2. Tree-based models

Tree-based models use decision tree structures to create an interpretable model for the data under analysis. These models are particularly effective in capturing nonlinear relationships, ranking feature importance, and handling mixed data types.

- **Random Forest (RF)**: An ensemble of decision trees built using bootstrap aggregating (bagging), reducing overfitting, and improving model stability and interpretability (Lundberg, Allen, and Lee 2022).
- **Gradient Boosting Decision Tree (GBDT)**: Constructs additive models in a sequential manner by minimizing residual errors at each stage through gradient descent (Cortes, Vapnik, and Saitta 1995).
- **Extreme Gradient Boosting (XGBoost)**: A scalable and regularized extension of GBDT that improves computation and generalization through efficient parallelism and regularization (Xue and Su 2017).
- **Categorical Boosting (CatBoost)**: A gradient boosting algorithm that handles categorical features using ordered boosting and target statistics, reducing overfitting and improving accuracy (Gao et al. 2018).

### 3.2.3. Ensemble models

Ensemble models integrate predictions from multiple base learners to enhance predictive accuracy and reduce generalization error.

- **Stack Ensemble (StackEns)**: A model that utilizes a shallow Decision Tree Regressor and a Bagging Regressor as base learners, with Ridge Regression serving as the meta-learner. This setup balances performance and interpretability while promoting model diversity. Compared to more complex learners such as CatBoost and XGBoost, this ensemble enhances generalization, which is particularly valuable in ecological modelling.
- **AVG-Ensemble Model**: In this study, three high-performing models – RF, CatBoost, XGBoost – were selected to form the ensemble by taking the average predictions of the constituent models. Their combined outputs aim to deliver more accurate and robust predictions than any single model alone. This strategy leverages the strengths of each model, capturing underlying data patterns

more effectively and improving overall prediction reliability. Mathematically, this can be expressed as follows (Equation 1), where $\tilde{y}_{ensemble,i}$ is the ensemble prediction for the $i-th$ data point, $N$ is the number of models, and $\tilde{y}_{j,i}$ is the prediction of $j-th$ model for the $i-th$ data point.

$$\tilde{y}_{ensemble,i} = \frac{1}{N}\sum_{j=1}^{N} \tilde{y}_{j,i} \quad (1)$$

### 3.3. Explainable model architecture

SHAP values provide a significant advantage in model interpretability by offering unique, consistent, and locally accurate attribution values (Ghafarian et al. 2022). Unlike traditional feature attribution methods, SHAP values are computed for every individual prediction, providing detailed insights into the model's behaviour with an unprecedented level of granularity (Ghafarian et al. 2022). This contrasts with feature attribution techniques used with ML algorithms, which typically assign importance to input features without considering the specifics of individual predictions. By integrating SHAP with ML algorithms, we comprehensively understand the complex relationship between GLCM metrics and OC. The key advantage of SHAP values lies in their ability to generate a richer visual representation, such as SHAP summary plots, which depict individualized feature attributions (Ghafarian et al. 2022). These plots highlight the range and distribution of a feature's impact on the model's output, offering valuable insights into how a feature's value correlates with its influence on predictions (Ghafarian et al. 2022). Figure 4 shows the pipeline named 'GLCM-Based Organic Carbon Estimation and Explainable ML Pipeline for Forest Inventory Data', developed as part of this research. The pipeline integrates texture analysis using GLCM and machine learning models to estimate organic carbon from forest inventory data, providing accurate predictions and model interpretability.

## 4. Experimental setup

### 4.1. Training and testing procedure

To enhance the learning of the model and ensure robust performance evaluation, we adopted a 5-fold cross-validation technique. This method minimizes the influence of data variability and promotes greater diversity in the training sets. The dataset was divided into five equal parts or 'folds', with each cycle comprising training on four folds (80% of the data) and testing on the remaining one-fold (20% of the data). This procedure was repeated five times, with each fold serving as the test set once. By computing the average performance metrics across all iterations and runs, we obtained a more dependable assessment of the model's efficacy. Moreover, in terms of tackling overfitting, 5-fold cross-validation indirectly addresses this concern by evaluating the model's performance across diverse data subsets. This ensures that the model's effectiveness is not overly tailored to specific training data, thereby aiding in the detection and prevention of overfitting issues.

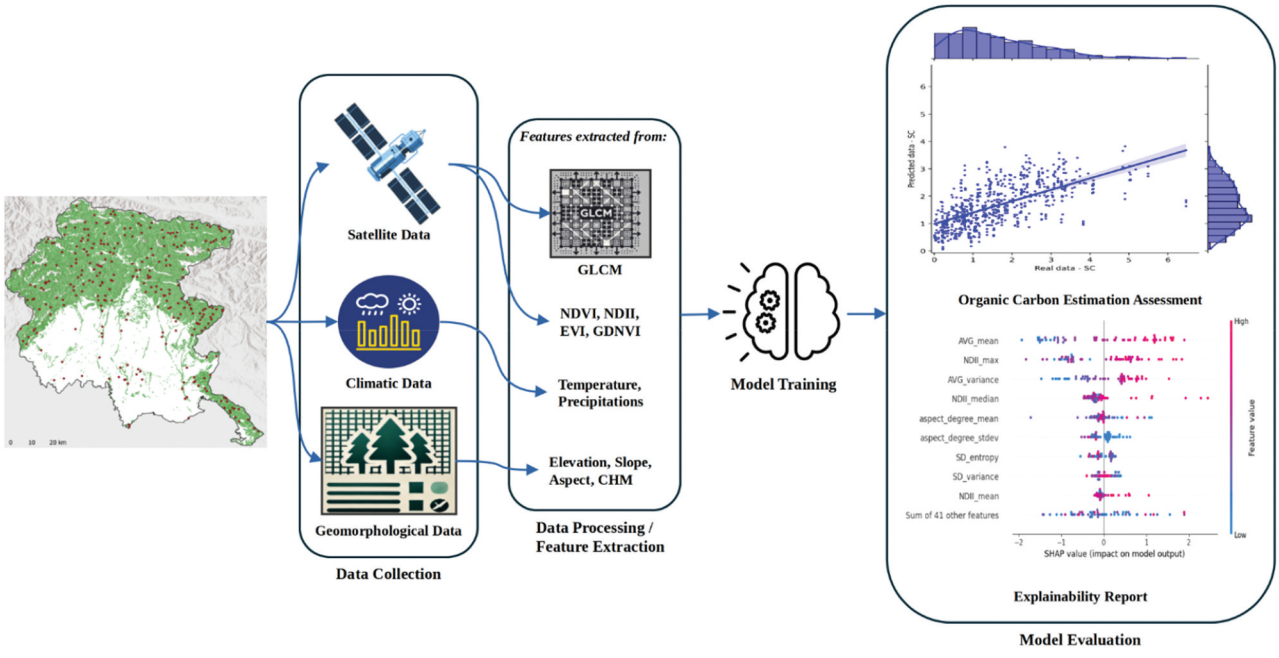



Figure 4. GLCM-Based organic carbon estimation and explainable ML pipeline for forest inventory data.

### 4.2. Assessment of model predictions using jackknife resampling and error bars

The jackknife resampling technique is a well-established method for evaluating the robustness and reliability of prediction models, particularly in complex environmental studies such as ours, which estimate OC in trees using GLCM features and explainable machine learning (Montgomery, Peck, and Vining 2021). In our study, jackknife resampling was applied by systematically leaving out one observation at a time from the training set and retraining the model on the remaining *n*-1 samples. Each of these jackknife models was then used to make predictions on a fixed test set. For each test sample, the standard deviation across predictions from the n jackknife models was computed, quantifying the variability introduced by perturbations in the training data (Breiman 2001). This uncertainty estimation is based on model output resampling via jackknife and does not rely on cross-validation folds. These standard deviations were used as error bars to visualize prediction uncertainty. This approach is especially valuable for small datasets or models that may be sensitive to individual data points, offering a practical and interpretable way to assess confidence in the model's predictions.

Error bars are a critical aspect of this research as they provide a visual representation of the uncertainty associated with each prediction (Friedman 2001). In the context of environmental modelling, where predictions can have significant implications for policy and management decisions, understanding the range of possible outcomes is essential. The jackknife-derived error bars in these figures reflect not only the variability in the predictions but also the model's sensitivity to individual data points in the training set (Shao and Tu 2012). Larger error bars suggest areas where the model is less certain, potentially due to factors like limited data, high variability in the input features, or the inherent complexity of the phenomena being modelled (Hesterberg 2015). These error bars visually indicate the confidence intervals around each prediction, helping us understand where our model performs well and where it struggles.

### 4.3. Evaluation metrics

We examine three metrics to assess the models' performance: coefficient of determination ($R^2$) (Equation 2), root-mean-square error (RMSE) (Equation 3), and root-mean-square error percentage (%RMSE) (Equation 4). Typically, lower RMSE and %RMSE values indicate superior performance, whereas a higher $R^2$ signifies better estimation accuracy. In these Equations, $\hat{y}$ refers to the predicted value, $y_i$ is the measured observed value, $\bar{y}$ represents the mean of the observed values, and n is the test sample size.

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

$$\mathrm{RMSE} = \sqrt{\frac{1}{\mathrm{n}} \sum_{\mathrm{i}=1}^{\mathrm{n}} (\mathrm{y_i} - \hat{y})^2} \tag{3}$$

$$\%RMSE = \frac{\sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y})^2}}{\bar{y}} \times 100 \tag{4}$$

### 4.4. Parameter tuning

To optimize the performance of our models and refine our estimates for OC estimation, we utilized the grid search technique (Feurer and Hutter 2019). This method entails generating a grid of hyperparameter values and training the models with each possible combination (Feurer and Hutter 2019). This systematic approach enables a thorough exploration of different hyperparameter configurations, aiding in the identification of the most effective setup for our models. Tables 4 and 5 provide comprehensive information on the range of hyperparameter values investigated, along with the optimal hyperparameters identified for the three top models as examples.

**Table 4.** List of the hyperparameters of the algorithms, their meaning, and the range explored during grid search.

| Models | Parameter name | Meaning | Range |
|---|---|---|---|
| CatBoost | iterations | Max count of trees. | [100, 150, 200, 250] |
| | learning_rate | Boosting learning rate. | [0.03, 0.1] |
| | depth | Depth of a tree. | [2, 4, 6, 8, 10] |
| | l2_leaf_reg | Coefficient at the L2 regularization term of the cost function. | [0.2, 0.5, 1, 3, 4] |
| GBDT | n_estimators | The number of boosting stages to perform. | [1, 2, 5, 10, 20, 50, 100, 200, 500, 1000, 2000] |
| | max_leaf_nodes | Grow trees with max_leaf_nodes in the best-first fashion. | [2, 5, 10, 20, 50, 100] |
| | learning_rate | Boosting learning rate. | [1, 0.1, 0.01,0.001] |
| | max_depth | Maximum depth of the individual regression estimators. | [1,2,4] |
| | subsample | The fraction of samples to be used for fitting the individual base learners. | [0.5, 0.75, 1] |
| XGBoost | max_depth | Maximum tree depth for base learners. | [4, 5, 10] |
| | n_estimators | The number of boosting stages to perform. | [500, 600, 700, 800] |
| | learning_rate | Boosting learning rate. | [0.01, 0.015, 0.02] |

**Table 5.** Best hyperparameters found for all the models, for all configurations, for the target variable CSE.

| Model | Conf1 | Conf2 | Conf3 | Conf4 | Conf5 | Conf6 |
|---|---|---|---|---|---|---|
| **CatBoost** | learning_rate: 0.03<br>l2_leaf_reg: 0.5<br>iterations: 150<br>depth: 2 | learning_rate: 0.1<br>l2_leaf_reg: 1<br>iterations: 100<br>depth: 2 | learning_rate: 0.03<br>l2_leaf_reg: 0.5<br>iterations: 150<br>depth: 2 | learning_rate: 0.1<br>l2_leaf_reg: 3<br>iterations: 200<br>depth: 6 | learning_rate: 0.1<br>l2_leaf_reg: 3<br>iterations: 200<br>depth: 6 | learning_rate: 0.1<br>l2_leaf_reg: 3<br>iterations: 200<br>depth: 6 |
| **GBDT** | subsample: 0.75<br>n_estimators: 100<br>max_leaf_nodes: 100<br>max_depth: 2<br>learning_rate: 0.1 | subsample: 0.5<br>n_estimators: 100<br>max_leaf_nodes: 20<br>max_depth: 4<br>learning_rate: 0.01 | subsample: 1<br>n_estimators: 1000<br>max_leaf_nodes: 100<br>max_depth: 4<br>learning_rate: 0.01 | subsample: 1<br>n_estimators: 1000<br>max_leaf_nodes: 100<br>max_depth: 4<br>learning_rate: 0.01 | subsample: 1<br>n_estimators: 1000<br>max_leaf_nodes: 100<br>max_depth: 4<br>learning_rate: 0.01 | subsample: 1<br>n_estimators: 1000<br>max_leaf_nodes: 100<br>max_depth: 4<br>learning_rate: 0.01 |
| **XGBoost** | n_estimators: 700<br>max_depth: 4<br>learning_rate: 0.01 | n_estimators: 700<br>max_depth: 4<br>learning_rate: 0.01 | n_estimators: 700<br>max_depth: 4<br>learning_rate: 0.01 | n_estimators: 600<br>max_depth: 5<br>learning_rate: 0.02 | n_estimators: 600<br>max_depth: 5<br>learning_rate: 0.02 | n_estimators: 600<br>max_depth: 5<br>learning_rate: 0.02 |

## 5. Results

The performance of the models in estimating OC was evaluated using six distinct configurations of input variables. These configurations included: (1) solely VIs, (2) solely GLCM Metrics, (3) a fusion of VIs and GLCM Metrics, (4) a combination of VIs, geomorphological parameters, and weather parameters, (5) GLCM metrics along with geomorphological parameters and weather parameters, and (6) a holistic approach integrating all variables.

The incorporation of GLCM metrics significantly improved model performance across all configurations for CSE estimation (Table 6). Conf 2, which used only GLCM for CS estimation, consistently outperformed Conf 1, which relied solely on vegetation indices (VIs), across nearly all models. This highlights the strong predictive power of texture-based features in modelling CSE. For example, CatBoost in Conf 4, which includes VIs, geomorphological parameters, and weather data, achieved an $R^2$ of 0.62, an RMSE of 0.89 tC $ha^{-1}$ $yr^{-1}$, and a %RMSE of 52.80%. This was further improved in Conf 5, where adding GLCM metrics resulted in an $R^2$ of 0.62 and a reduced %RMSE of 52.80%. The best CatBoost performance was seen in Conf (integrating VIs, GLCM, geomorphological, and weather variables), achieving an $R^2$ of 0.66, an RMSE of 0.86 tC $ha^{-1}$ $yr^{-1}$, and a %RMSE of 50.99%. These results show that including GLCM features adds value beyond traditional spectral indices. A similar trend is observed with XGBoost and GBDT models. In both cases, Conf 6 consistently delivers the best performance, indicating that the integration of all feature types yields the most accurate predictions. For instance, XGBoost achieved its highest $R^2$ of 0.59 in Conf 5 and 6, coupled with the lowest RMSE and %RMSE values (0.92 tC $ha^{-1}$ $yr^{-1}$ and ~ 54.2%, respectively). GBDT followed this pattern, with its best performance in Conf 6.

However, the RF model outperformed all models in the study, particularly in Conf 6. RF achieved the highest overall $R^2$ of 0.67, the lowest RMSE of 0.85 tC $ha^{-1}$ $yr^{-1}$, and the lowest %RMSE of 50.22%, indicating its superior capability in modelling CSE when provided with a comprehensive feature set. The consistent improvement from Conf 1 through Conf 6 in RF performance further underscores the value of using a multimodal feature integration strategy. For example, RF improved from an $R^2$ of 0.24 in Conf 1 to 0.65 in Conf 5 and finally to 0.67 in Conf 6.

This trend was mirrored, albeit to a lesser extent, in other models such as MLR, MLP, KNN, SVR, and ensemble approaches. While these models showed noticeable improvements with the inclusion of GLCM metrics (especially in Conf 2 and 5), their predictive performance generally lagged behind that of CatBoost, XGBoost, GBDT, and RF. For instance, MLP and KNN showed significant performance drops in Conf 4–6, highlighting their sensitivity to input feature complexity and model structure. Stackens and AVG-Ens approaches yielded moderate improvements over simpler configurations but did not surpass the top-performing individual models. In Conf 6, Stackens achieved an $R^2$ of 0.49 and an RMSE of 0.92 tC $ha^{-1}$ $yr^{-1}$, while AVG-Ens recorded an $R^2$ of 0.42.

A similar pattern is observed in the estimation of SDC, where incorporating GLCM metrics consistently improved model performance across all models and configurations (Table 7), highlighting the added value of texture-based information from remote sensing. Conf 2, which incorporated only GLCM metrics alongside VIs, consistently outperformed Conf 1 (VIs only), indicating the independent predictive power of GLCM features. For instance, CatBoost exhibited a marked improvement in $R^2$ from 0.18 to 0.51, with a corresponding decrease in RMSE from 3.33 to 3.16 tC $ha^{-1}$). RF followed a similar trend, with $R^2$ increasing from 0.17 to 0.50. These improvements underscore the ability of GLCM metrics to capture spatial texture information that complements spectral data in modelling soil development classes.

**Table 6.** Model performance for CS predicting across six input configurations using 10 ML models. Metrics include $R^2$ (higher is better), RMSE (tC ha −1 yr − 1), and %RMSE (both of which are better when lower). CatBoost and RF consistently demonstrate superior performance, yielding the best results in configuration 6, with the highest $R^2$, lowest RMSE, and lowest %RMSE. Bold values indicate the best performance per metric. Bold values indicate the best performance per metric.

| Model | Configs | $R^2$ | RMSE (tC ha − 1) | %RMSE |
|---|---|---|---|---|
| CatBoost | Conf (1) | 0.25 ± 0.10 | 1.00 ± 0.13 | 59.11 ± 8.82 |
| | Conf (2) | 0.54 ± 0.08 | 0.96 ± 0.09 | 56.88 ± 6.12 |
| | Conf (3) | 0.63 ± 0.08 | 0.89 ± 0.12 | 52.80 ± 8.18 |
| | Conf (4) | 0.25 ± 0.13 | 1.00 ± 0.14 | 59.21 ± 9.36 |
| | Conf (5) | 0.62 ± 0.11 | 0.89 ± 0.11 | 52.80 ± 7.62 |
| | **Conf (6)** | **0.66 ± 0.09** | **0.86 ± 0.12** | **50.99 ± 8.01** |
| GBDT | Conf (1) | 0.21 ± 0.08 | 1.03 ± 0.13 | 60.60 ± 8.76 |
| | Conf (2) | 0.49 ± 0.12 | 0.99 ± 0.08 | 58.51 ± 5.63 |
| | Conf (3) | 0.55 ± 0.12 | 0.95 ± 0.11 | 56.22 ± 7.28 |
| | Conf (4) | 0.24 ± 0.08 | 1.01 ± 0.13 | 59.62 ± 8.49 |
| | Conf (5) | 0.58 ± 0.15 | 0.92 ± 0.12 | 54.18 ± 7.99 |
| | Conf (6) | 0.58 ± 0.13 | 0.93 ± 0.13 | 54.85 ± 8.68 |
| KNN | Conf (1) | 0.20 ± 0.09 | 1.03 ± 0.13 | 61.03 ± 8.86 |
| | Conf (2) | 0.53 ± 0.06 | 0.97 ± 0.09 | 57.45 ± 6.22 |
| | Conf (3) | 0.53 ± 0.06 | 0.97 ± 0.09 | 57.45 ± 6.22 |
| | Conf (4) | 0.06 ± 0.07 | 1.12 ± 0.08 | 65.84 ± 5.78 |
| | Conf (5) | 0.43 ± 0.11 | 1.04 ± 0.07 | 61.31 ± 5.19 |
| | Conf (6) | 0.37 ± 0.15 | 1.06 ± 0.09 | 62.38 ± 6.34 |
| MLR | Conf (1) | 0.18 ± 0.08 | 1.04 ± 0.09 | 61.44 ± 6.23 |
| | Conf (2) | 0.56 ± 0.08 | 0.95 ± 0.09 | 56.26 ± 5.85 |
| | Conf (3) | 0.54 ± 0.08 | 0.97 ± 0.12 | 57.18 ± 7.49 |
| | Conf (4) | 0.14 ± 0.14 | 1.07 ± 0.08 | 62.88 ± 5.62 |
| | Conf (5) | 0.55 ± 0.12 | 0.95 ± 0.12 | 56.09 ± 7.62 |
| | Conf (6) | 0.49 ± 0.16 | 0.99 ± 0.13 | 58.55 ± 8.36 |
| MLP | Conf (1) | 0.15 ± 0.06 | 1.07 ± 0.11 | 62.96 ± 7.58 |
| | Conf (2) | 0.52 ± 0.12 | 0.98 ± 0.09 | 58.01 ± 5.55 |
| | Conf (3) | 0.53 ± 0.09 | 0.97 ± 0.09 | 57.14 ± 6.01 |
| | Conf (4) | 0.02 ± 0.04 | 1.17 ± 0.10 | 68.95 ± 6.90 |
| | Conf (5) | 0.11 ± 0.10 | 1.17 ± 0.10 | 68.92 ± 7.03 |
| | Conf (6) | 0.13 ± 0.13 | 1.17 ± 0.10 | 69.14 ± 6.67 |
| RF | Conf (1) | 0.24 ± 0.09 | 1.01 ± 0.12 | 59.53 ± 8.17 |
| | Conf (2) | 0.53 ± 0.09 | 0.97 ± 0.09 | 57.16 ± 6.28 |
| | Conf (3) | 0.63 ± 0.07 | 0.89 ± 0.12 | 52.63 ± 7.87 |
| | Conf (4) | 0.29 ± 0.08 | 0.98 ± 0.11 | 57.63 ± 7.55 |
| | Conf (5) | 0.65 ± 0.09 | 0.87 ± 0.11 | 51.40 ± 7.53 |
| | **Conf (6)** | **0.67 ± 0.08** | **0.85 ± 0.13** | **50.22 ± 8.27** |
| SVR | Conf (1) | 0.24 ± 0.09 | 1.01 ± 0.12 | 59.53 ± 8.17 |
| | Conf (2) | 0.54 ± 0.11 | 0.96 ± 0.10 | 56.93 ± 6.56 |
| | Conf (3) | 0.54 ± 0.11 | 0.96 ± 0.10 | 56.93 ± 6.56 |
| | Conf (4) | 0.17 ± 0.13 | 1.05 ± 0.11 | 62.02 ± 7.77 |
| | Conf (5) | 0.42 ± 0.19 | 1.02 ± 0.09 | 60.38 ± 6.40 |
| | Conf (6) | 0.41 ± 0.12 | 1.06 ± 0.09 | 62.26 ± 6.39 |
| XGBoost | Conf (1) | 0.24 ± 0.07 | 1.01 ± 0.13 | 59.41 ± 8.46 |
| | Conf (2) | 0.39 ± 0.15 | 1.06 ± 0.11 | 62.74 ± 7.27 |
| | Conf (3) | 0.57 ± 0.11 | 0.94 ± 0.12 | 55.27 ± 7.91 |
| | Conf (4) | 0.17 ± 0.13 | 1.05 ± 0.11 | 62.02 ± 7.77 |
| | Conf (5) | 0.59 ± 0.13 | 0.92 ± 0.12 | 54.18 ± 7.64 |
| | Conf (6) | 0.59 ± 0.14 | 0.92 ± 0.14 | 54.22 ± 9.09 |
| Stackens | Conf (1) | 0.17 ± 0.11 | 1.05 ± 0.12 | 67.89 ± 7.91 |
| | Conf (2) | 0.46 ± 0.13 | 0.94 ± 0.14 | 61.03 ± 7.80 |
| | Conf (3) | 0.46 ± 0.12 | 0.94 ± 0.15 | 61.12 ± 6.90 |
| | Conf (4) | 0.24 ± 0.13 | 1.03 ± 0.13 | 66.90 ± 5.81 |
| | Conf (5) | 0.54 ± 0.11 | 0.89 ± 0.10 | 57.68 ± 7.70 |
| | Conf (6) | 0.49 ± 0.12 | 0.92 ± 0.11 | 59.82 ± 5.61 |
| AVG-Ens | Conf (1) | 0.12 ± 0.15 | 1.08 ± 0.10 | 69.55 ± 6.93 |
| | Conf (2) | 0.42 ± 0.16 | 0.96 ± 0.13 | 62.33 ± 6.70 |
| | Conf (3) | 0.49 ± 0.12 | 0.93 ± 0.14 | 60.01 ± 7.91 |
| | Conf (4) | 0.26 ± 0.15 | 1.03 ± 0.16 | 66.43 ± 8.40 |
| | Conf (5) | 0.55 ± 0.13 | 0.89 ± 0.10 | 57.57 ± 6.75 |
| | Conf (6) | 0.42 ± 0.17 | 0.97 ± 0.11 | 62.46 ± 7.91 |

**Table 7.** Model performance for SDC predicting across six input configurations using 10 ML models. Metrics include $R^2$ (higher is better), RMSE (tC ha − 1), and %RMSE (both lower is better). CatBoost and RF consistently show superior performance, achieving the best results in configuration 6 with the highest $R^2$ and lowest RMSE and %RMSE. Bold values indicate the best performance per metric.

| Model | Configs | $R^2$ | RMSE (tC ha − 1) | %RMSE |
|---|---|---|---|---|
| CatBoost | Conf (1) | 0.18 ± 0.13 | 3.33 ± 0.45 | 65.10 ± 9.85 |
| | Conf (2) | 0.51 ± 0.08 | 3.16 ± 0.33 | 61.60 ± 7.41 |
| | Conf (3) | 0.62 ± 0.08 | 2.89 ± 0.42 | 56.49 ± 9.01 |
| | Conf (4) | 0.40 ± 0.12 | 2.85 ± 0.39 | 55.61 ± 8.44 |
| | Conf (5) | 0.63 ± 0.09 | 2.84 ± 0.33 | 55.38 ± 7.14 |
| | **Conf (6)** | **0.65 ± 0.11** | **2.77 ± 0.41** | **53.99 ± 8.62** |
| GBDT | Conf (1) | 0.14 ± 0.09 | 3.43 ± 0.38 | 66.90 ± 8.48 |
| | Conf (2) | 0.47 ± 0.05 | 3.27 ± 0.33 | 63.74 ± 7.55 |
| | Conf (3) | 0.48 ± 0.15 | 3.18 ± 0.40 | 62.09 ± 8.68 |
| | Conf (4) | 0.32 ± 0.13 | 3.04 ± 0.41 | 59.28 ± 8.78 |
| | Conf (5) | 0.58 ± 0.11 | 2.95 ± 0.37 | 57.54 ± 7.80 |
| | Conf (6) | 0.57 ± 0.13 | 2.99 ± 0.40 | 58.36 ± 8.58 |
| KNN | Conf (1) | 0.16 ± 0.09 | 3.39 ± 0.43 | 66.28 ± 9.50 |
| | Conf (2) | 0.48 ± 0.08 | 3.23 ± 0.32 | 63.09 ± 7.32 |
| | Conf (3) | 0.48 ± 0.08 | 3.23 ± 0.32 | 63.09 ± 7.32 |
| | Conf (4) | 0.04 ± 0.07 | 3.61 ± 0.27 | 70.45 ± 6.47 |
| | Conf (5) | 0.41 ± 0.12 | 3.34 ± 0.29 | 65.14 ± 6.98 |
| | Conf (6) | 0.41 ± 0.12 | 3.34 ± 0.29 | 65.14 ± 6.98 |
| MLR | Conf (1) | 0.12 ± 0.09 | 3.46 ± 0.32 | 67.52 ± 7.08 |
| | Conf (2) | 0.51 ± 0.09 | 3.15 ± 0.32 | 61.40 ± 7.14 |
| | Conf (3) | 0.48 ± 0.09 | 3.22 ± 0.37 | 62.79 ± 7.93 |
| | Conf (4) | 0.30 ± 0.11 | 3.09 ± 0.34 | 60.20 ± 7.39 |
| | Conf (5) | 0.51 ± 0.17 | 3.10 ± 0.40 | 60.56 ± 8.59 |
| | Conf (6) | 0.47 ± 0.15 | 3.20 ± 0.37 | 62.50 ± 7.94 |
| MLP | Conf (1) | 0.15 ± 0.12 | 3.41 ± 0.50 | 66.65 ± 10.69 |
| | Conf (2) | 0.55 ± 0.16 | 3.33 ± 0.51 | 64.82 ± 9.41 |
| | Conf (3) | 0.49 ± 0.16 | 3.16 ± 0.36 | 61.72 ± 7.88 |
| | Conf (4) | 0.03 ± 0.05 | 3.74 ± 0.33 | 73.06 ± 7.62 |
| | Conf (5) | 0.11 ± 0.11 | 3.74 ± 0.33 | 72.94 ± 7.77 |
| | Conf (6) | 0.11 ± 0.11 | 3.74 ± 0.34 | 72.95 ± 7.86 |
| RF | Conf (1) | 0.17 ± 0.12 | 3.36 ± 0.41 | 65.51 ± 8.95 |
| | Conf (2) | 0.50 ± 0.10 | 3.17 ± 0.33 | 61.84 ± 7.30 |
| | Conf (3) | 0.62 ± 0.08 | 2.90 ± 0.43 | 56.59 ± 9.12 |
| | Conf (4) | 0.40 ± 0.12 | 2.85 ± 0.42 | 55.68 ± 9.03 |
| | Conf (5) | 0.64 ± 0.09 | 2.80 ± 0.37 | 54.59 ± 8.04 |
| | **Conf (6)** | **0.67 ± 0.09** | **2.73 ± 0.42** | **53.32 ± 9.04** |
| SVR | Conf (1) | 0.17 ± 0.09 | 3.37 ± 0.45 | 65.90 ± 10.00 |
| | Conf (2) | 0.50 ± 0.12 | 3.16 ± 0.33 | 61.69 ± 7.43 |
| | Conf (3) | 0.50 ± 0.12 | 3.16 ± 0.33 | 61.69 ± 7.44 |
| | Conf (4) | 0.00 ± 0.11 | 3.67 ± 0.26 | 71.58 ± 6.23 |
| | Conf (5) | 0.43 ± 0.08 | 3.33 ± 0.38 | 65.09 ± 8.56 |
| | Conf (6) | 0.43 ± 0.08 | 3.33 ± 0.38 | 65.09 ± 8.56 |
| XGBoost | Conf (1) | 0.03 ± 0.26 | 3.73 ± 0.62 | 72.77 ± 13.14 |
| | Conf (2) | 0.38 ± 0.17 | 3.40 ± 0.36 | 66.39 ± 7.86 |
| | Conf (3) | 0.52 ± 0.12 | 3.12 ± 0.43 | 60.85 ± 9.19 |
| | Conf (4) | 0.34 ± 0.11 | 3.00 ± 0.39 | 58.57 ± 8.61 |
| | Conf (5) | 0.58 ± 0.10 | 2.97 ± 0.32 | 57.86 ± 7.08 |
| | Conf (6) | 0.57 ± 0.11 | 3.00 ± 0.40 | 58.67 ± 6.40 |
| Stackens | Conf (1) | 0.37 ± 0.10 | 3.17 ± 0.35 | 68.11 ± 8.50 |
| | Conf (2) | 0.43 ± 0.12 | 3.08 ± 0.43 | 66.36 ± 6.70 |
| | Conf (3) | 0.48 ± 0.09 | 3.00 ± 0.33 | 64.52 ± 8.86 |
| | Conf (4) | 0.24 ± 0.17 | 1.03 ± 0.40 | 66.90 ± 7.50 |
| | Conf (5) | 0.52 ± 0.11 | 2.92 ± 0.36 | 62.83 ± 8.86 |
| | Conf (6) | 0.45 ± 0.15 | 3.03 ± 0.43 | 65.66 ± 6.80 |
| AVG-Ens | Conf (1) | 0.14 ± 0.06 | 3.39 ± 0.26 | 72.88 ± 7.65 |
| | Conf (2) | 0.33 ± 0.05 | 3.23 ± 0.40 | 69.40 ± 8.45 |
| | Conf (3) | 0.52 ± 0.15 | 2.92 ± 0.37 | 62.86 ± 7.47 |
| | Conf (4) | 0.35 ± 0.17 | 3.20 ± 0.34 | 68.89 ± 6.75 |
| | Conf (5) | 0.48 ± 0.05 | 2.99 ± 0.28 | 64.39 ± 8.05 |
| | Conf (6) | 0.48 ± 0.08 | 2.99 ± 0.32 | 64.37 ± 6.25 |

Further gains were observed when GLCM metrics were combined with additional predictors. Conf 5 (VIs + GLCM + geomorphology) and Conf 6 (all inputs) generally yielded the best results across models. CatBoost achieved its highest performance in Conf 6 ($R^2 = 0.65$, RMSE = 2.77 (tC $ha^{-1}$), %RMSE = 53.99%), while RF outperformed all other models in the same configuration ($R^2 = 0.67$, RMSE = 2.73 tC $ha^{-1}$, %RMSE = 53.32%). This

highlights the strength of these models in leveraging diverse feature sets to enhance prediction accuracy. XGBoost and GBDT also benefited from GLCM integration, achieving their highest $R^2$ in Conf 5 and 6, although XGBoost showed a slight decline in Conf 6, suggesting sensitivity to high-dimensional inputs.

Other models, including KNN, SVR, and MLR, showed moderate improvement with adding GLCM metrics, particularly in Conf 2. However, gains plateaued or declined in more complex configurations, reflecting their limited capacity to exploit high-dimensional data fully. MLP, in contrast, showed inconsistent results, with only marginal improvement in Conf 2 and poor performance in subsequent configurations, possibly due to overfitting or inadequate model tuning. Ensemble approaches also benefited from the inclusion of GLCM features. The stacking ensemble showed progressive improvement up to Conf 5 ($R^2 = 0.52$), while the averaging ensemble peaked in Conf 3 ($R^2 = 0.52$), indicating that the combined strength of multiple base learners can effectively utilize GLCM-derived texture information.

Overall, these findings emphasize the critical role of GLCM metrics, particularly when combined with geomorphological and weather variables, in enhancing model performance for CSE and SDC estimation. Conf 6 consistently emerges as the most effective setup across all models, with RF leading the performance rankings. This suggests that ensemble-based tree models, especially RF and CatBoost, are well-suited to handle complex, heterogeneous feature sets for environmental prediction tasks.

To further evaluate the performance, we implemented a regression plot to show the relationship between actual data and model predictions. Figures 5 and 6 illustrate this relationship for the target variables using the CatBoost model for Conf 5 and 6. These figures include marginal distribution bar charts to provide additional context. The thin diagonal line represents a perfect fit, while the thicker line indicates the regression line from the model. As demonstrated in the figures, our results support the accuracy and reliability of the model predictions.

Figure 7 shows Jackknife Resampling and Error Bars, in which the predicted versus reported values for CSE and SDC generally align with the perfect prediction line, demonstrating the model's accuracy. However, the observed deviations, particularly at higher reported OC values, accompanied by larger error bars, suggest increased uncertainty and potential areas for model refinement. This interpretation underscores the importance of considering prediction uncertainty in environmental modelling, guiding future improvements to enhance model reliability and performance (Beven 2018).

### 5.1. Comparative performance analysis

An ANOVA test was performed to evaluate the statistical significance of the differences between configurations, followed by Tukey's HSD analysis (Table 8), using the CatBoost model as an example to estimate CSE across various configurations.

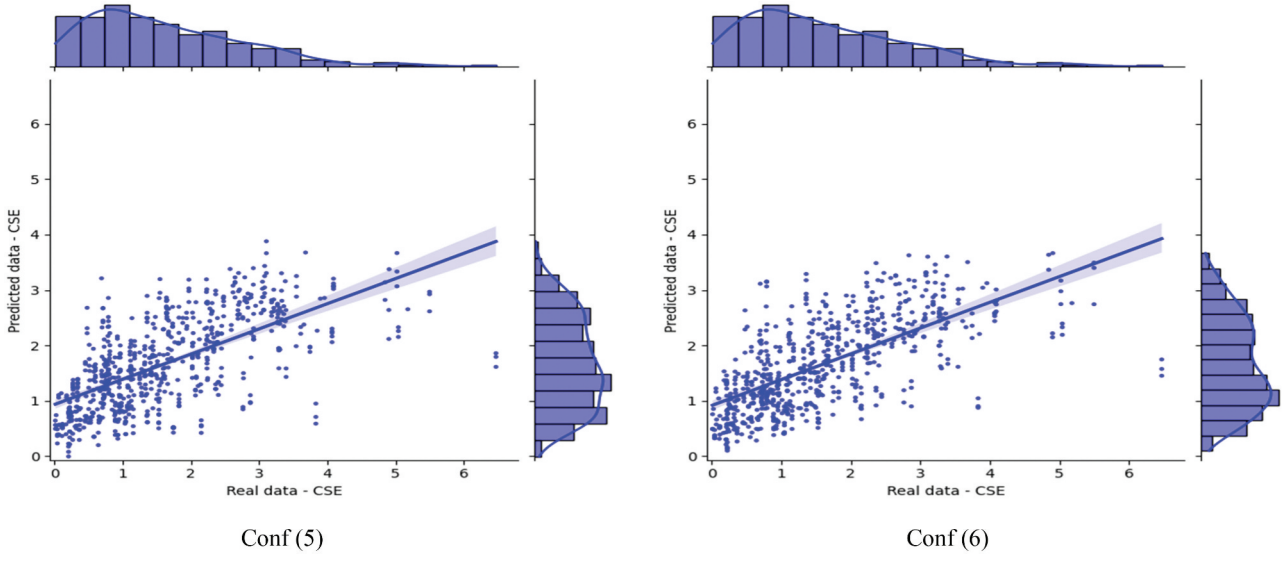



**Figure 5.** Scatter plots with marginal histograms showing the relationship between real and predicted values for the target variable CSE (tC $ha^{-1}$ $yr^{-1}$) using the CatBoost model with two different feature configurations: conf (5) (left) and Conf (6) (right). Both models exhibit a positive correlation, with Conf (6) showing slightly better alignment and narrower prediction intervals, suggesting improved predictive performance.

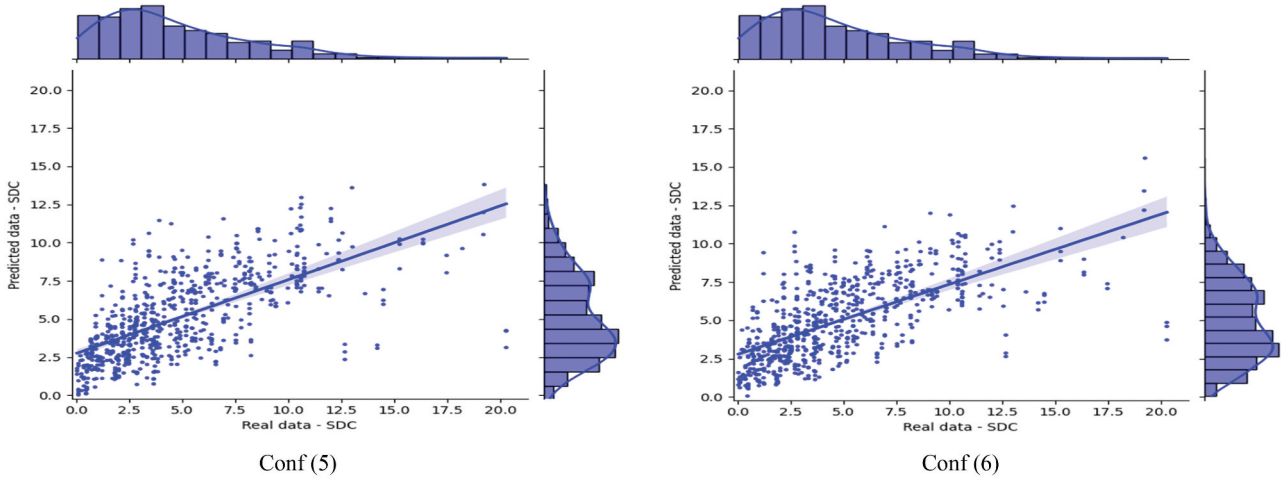



**Figure 6.** Scatter plots with marginal histograms showing the relationship between real and predicted values for the target variable SDC (**tC ha−1**) using the CatBoost model with two different feature configurations: conf (5) (left) and Conf (6) (right). Both models exhibit a positive correlation, with Conf (6) showing slightly better alignment and narrower prediction intervals, suggesting improved predictive performance.

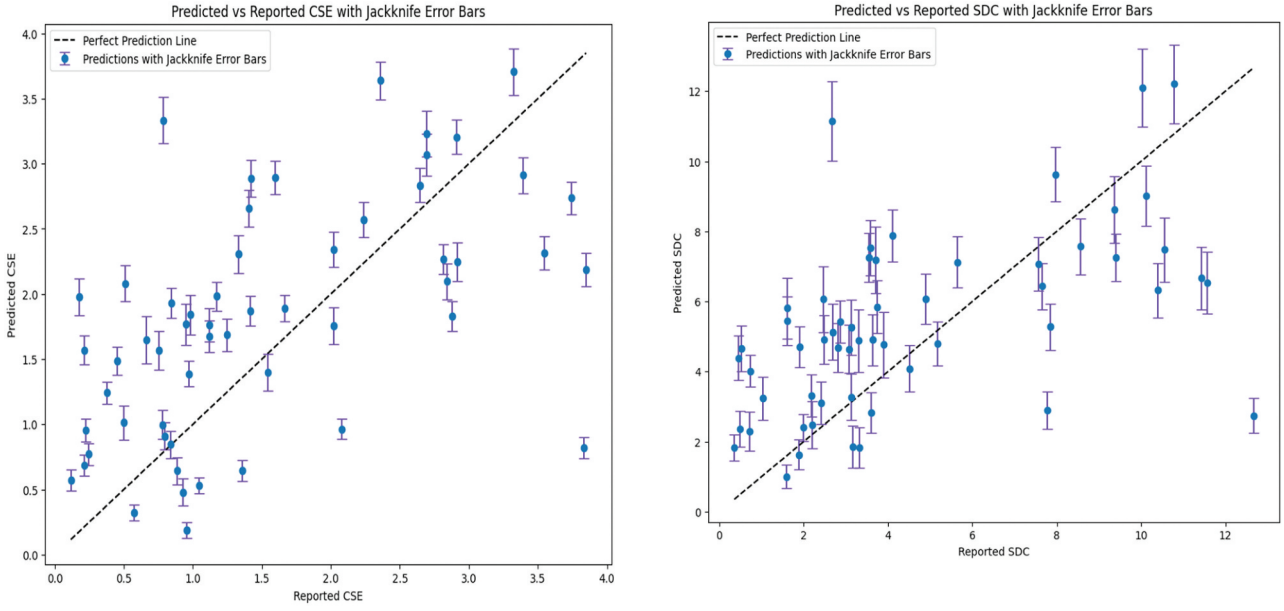



**Figure 7.** Predicted vs. Reported CSE (left) and SDC (right) with jackknife error bars. Scatter plots show model predictions against self-reported scores for CSE and SDC. Vertical error bars represent jackknife-estimated uncertainty. The dashed diagonal line indicates perfect prediction. Error bars illustrate confidence in predictions and highlight variability across individuals.

For $R^2$, the ANOVA test revealed significant differences across the configurations ($F(5, 12) = 24.303$, $p < 0.001$). When comparing Conf 1 (VIs only) to Conf 3 (VIs + GLCM Metrics), Tukey's HSD test indicated a significant increase in $R^2$, with a mean difference of 0.33, suggesting that the inclusion of GLCM metrics significantly enhances the model's explanatory power. Similarly, when comparing Conf 4 (VIs + geomorphological + weather parameters) with Conf 6 (which additionally includes GLCM metrics), there was a significant improvement in $R^2$, with a mean difference of 0.03, indicating the positive impact of GLCM metrics on model performance. However, the comparison between Conf 4 and Conf 5 (GLCM Metrics + geomorphological + weather parameters) showed no significant difference in $R^2$. For RMSE, significant differences were also observed ($F(5, 12) = 7.980$, $p < 0.002$). Tukey's HSD test revealed that adding GLCM metrics in Conf 3 reduced RMSE compared to Conf 1, with a mean difference of −0.17, demonstrating improved predictive accuracy. Similarly, comparing Conf 4 with Conf 6, including GLCM metrics,

**Table 8.** Tukey's HSD post-hoc analysis for model configurations.

| Metric | Configuration Comparison | Mean Difference | p-value | Significant ($p < 0.05$) |
|---|---|---|---|---|
| $R^2$ | Conf (1) vs Conf (3) | 0.33 | 0.018 | Yes |
| $R^2$ | Conf (4) vs Conf (5) | 0.01 | 0.884 | No |
| $R^2$ | Conf (4) vs Conf (6) | 0.03 | 0.041 | Yes |
| RMSE | Conf (1) vs Conf (3) | −0.17 | 0.039 | Yes |
| RMSE | Conf (4) vs Conf (5) | −0.05 | 0.732 | No |
| RMSE | Conf (4) vs Conf (6) | −0.12 | 0.049 | Yes |

led to a further reduction in RMSE, with a mean difference of −0.12. However, there was no significant difference between Conf 4 and Conf 5 for RMSE.

These findings confirm that the integration of GLCM metrics, particularly in Conf 3 and Conf 6, provides superior performance in estimating carbon stock, as evidenced by improved $R^2$ and reduced RMSE. Therefore, these configurations should be preferred for more accurate carbon stock estimation.

### 5.2. Model performance and overfitting analysis

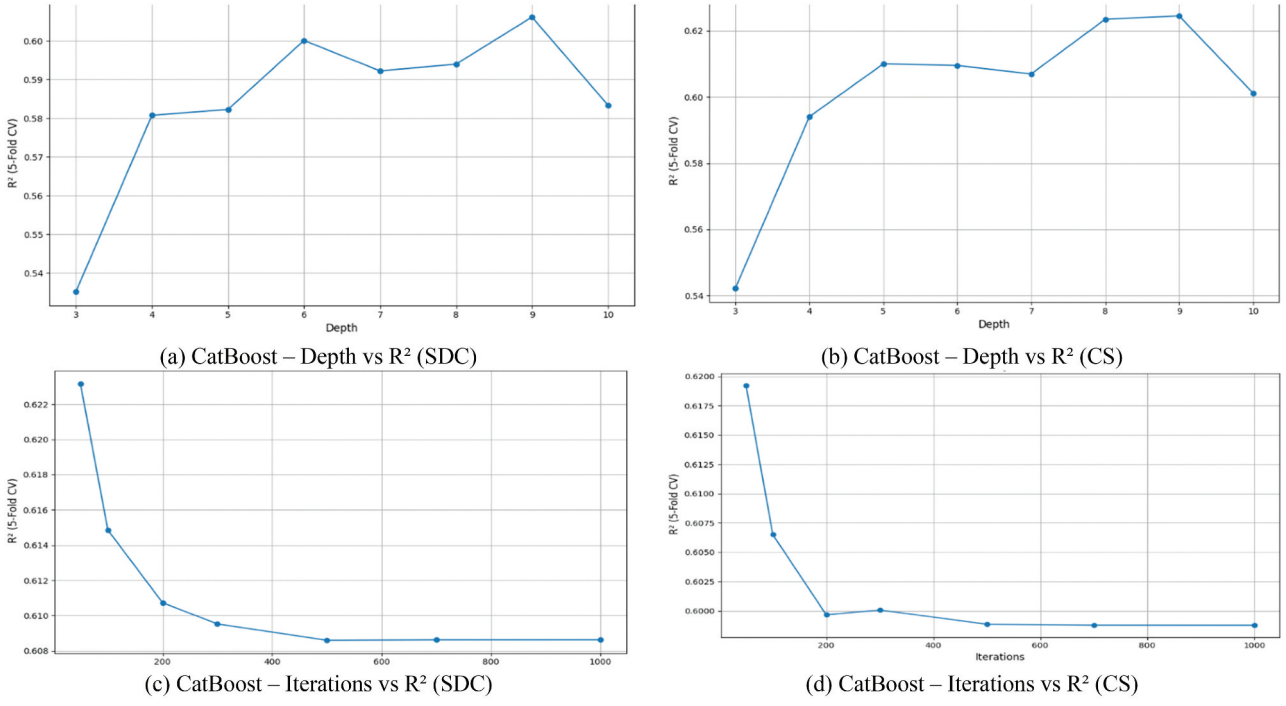
We applied thorough hyperparameter tuning and 5-fold cross-validation across all models to ensure robust and generalizable predictions. Figure 8 presents the sensitivity of CatBoost performance to tree depth and the number of boosting iterations for both prediction targets: CS (right) and SDC (left). In both cases, optimal $R^2$ values were achieved at moderate tree depths (around depth 9), with only a slight decline in performance beyond this point – a common indicator of potential overfitting. However, early stopping during training helped mitigate this risk, as seen in the controlled performance across iteration counts.

Figure 9 illustrates similar sensitivity analyses for RF models, focusing on tree depth and the number of estimators. These models displayed relatively stable performance, with $R^2$ values plateauing beyond certain depths and estimator thresholds. This consistency reflects RF's inherent robustness to overfitting, especially when used with cross-validation and limited maximum depth.

Overall, the combination of systematic hyperparameter optimization, early stopping, and cross-validation contributed to well-tuned models with strong generalization capacity, as reflected in the smooth performance trends shown in the figures.

### 5.3. SHAP analysis for feature importance and model interpretability

In this subsection, SHAP was applied to improve the interpretability of the XGBoost, CatBoost, and Gradient Boosting models, focusing on configuration Conf 6 for its




(a) CatBoost – Depth vs R² (SDC)
(b) CatBoost – Depth vs R² (CS)
(c) CatBoost – Iterations vs R² (SDC)
(d) CatBoost – Iterations vs R² (CS)

**Figure 8.** Sensitivity of CatBoost model performance to tree depth and number of iterations for SDC (left) and CSE (right) prediction. For both targets, peak performance is observed around depth 9, with $R^2$ declining beyond this point, suggesting potential overfitting. Iteration analysis shows rapid early improvement, followed by performance degradation after ~ 50–100 iterations.

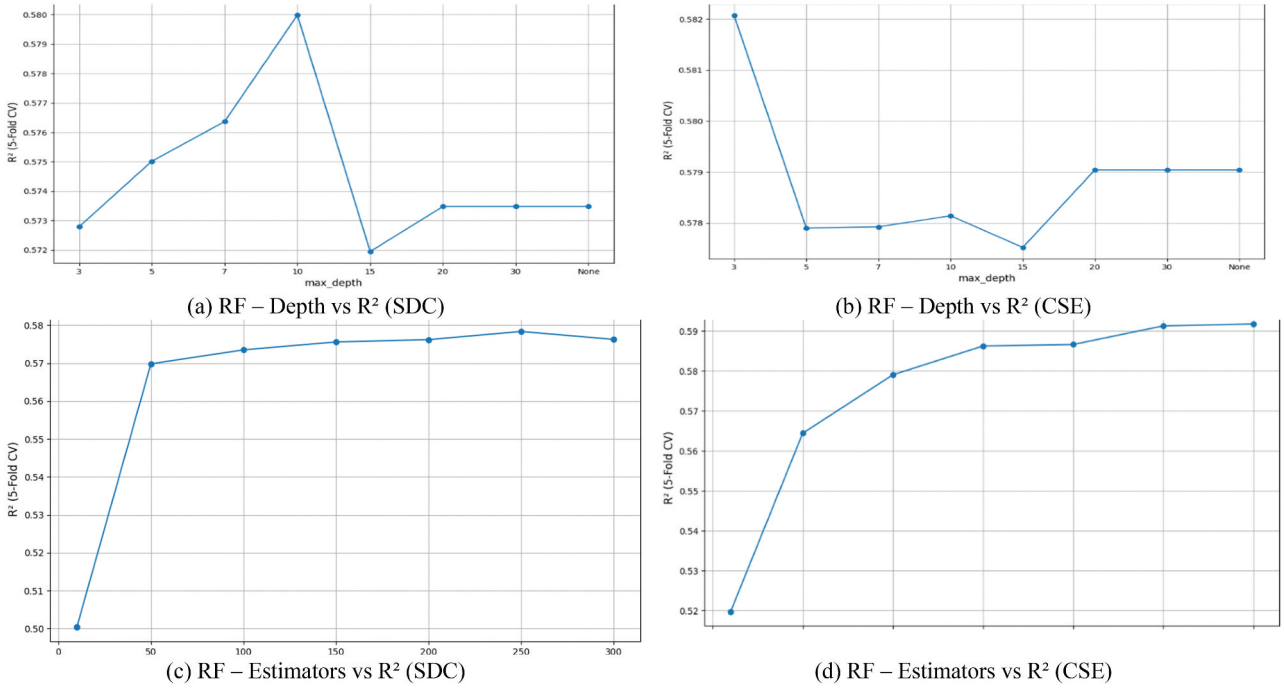



**Figure 9.** Sensitivity of Random forest model performance to tree depth and number of estimators for SDC (left) and CSE (right) prediction. Performance stabilizes around moderate tree depth, with marginal gains beyond ~ 100 estimators, indicating robustness and minimal risk of overfitting.

simplicity. To achieve a comprehensive understanding, three types of SHAP plots were used: the beeswarm plot, the force plot, and the heatmap. The SHAP beeswarm plot (Figure 10) visualizes the spread and impact of features across all predictions, highlighting which features have the most influence and how their values (from low to high) affect model outputs. The SHAP force plot (Figure 11) shows how individual features contribute to specific predictions, providing insights into the model's behaviour at the instance level. Meanwhile, the SHAP heatmap (Figure 12) aggregates these contributions across the dataset to reveal broader patterns and trends in feature importance.

In the SHAP beeswarm plots (Figure 10), feature importance was compared across two configurations, CSE and SDC, for all three models. For XGBoost, in the CSE configuration, features such as AVG_variance, AVG_mean, and NDII_median were the most impactful, with higher values generally leading to higher model outputs. In the SDC configuration, AVG_mean, NDII_max, and AVG_variance became more dominant. Similarly, the CatBoost model for CSE emphasized AVG_mean, AVG_variance, and NDII_median, while SDC focused on AVG_mean, AVG_variance, NDII_max, and SD_mean. The Gradient Boosting model showed that in CSE, features like AVG_mean, NDII_max, and GNDVI_max had a strong influence, whereas in SDC, AVG_mean, NDII_max, and SD_mean were particularly important.

Figure 11 presents SHAP force plots for both CSE and SDC configurations, using instances whose feature values are close to the dataset's average. These plots offer detailed insights into how individual features push or pull model predictions. In subplot (a), for the CSE configuration with the XGBoost model, GNDVI_max emerges as the most influential feature, positively pushing the prediction higher. Meanwhile, features like AVG_variance, AVG_mean, SD_entropy, and elevation_stdev contribute negatively, pulling the prediction downward. Subplot (b), corresponding to the SDC configuration, shows that features such as SD_homogeneity, SD_entropy, NDII_max, and GNDVI_max collectively push the prediction to a higher value, while AVG_mean, AVG_variance, and aspect_degree_stdev have a lowering effect. In subplot (c), similar to earlier observations, GNDVI_max remains the most impactful positive contributor for the CatBoost model in the CSE configuration, whereas AVG_variance, AVG_mean, SD_entropy, and elevation_stdev again act as suppressors. Subplot (d), representing CatBoost in the SDC setting, highlights SD_homogeneity, SD_entropy, NDII_max, and GNDVI_max as positive drivers of the prediction, while AVG_mean, AVG_variance, and aspect_degree_stdev work to decrease the output.

For the Gradient Boosting model, subplot (e) shows that features such as SD_dissimilarity, SD_mean, and NDII_max have strong positive impacts in the CSE

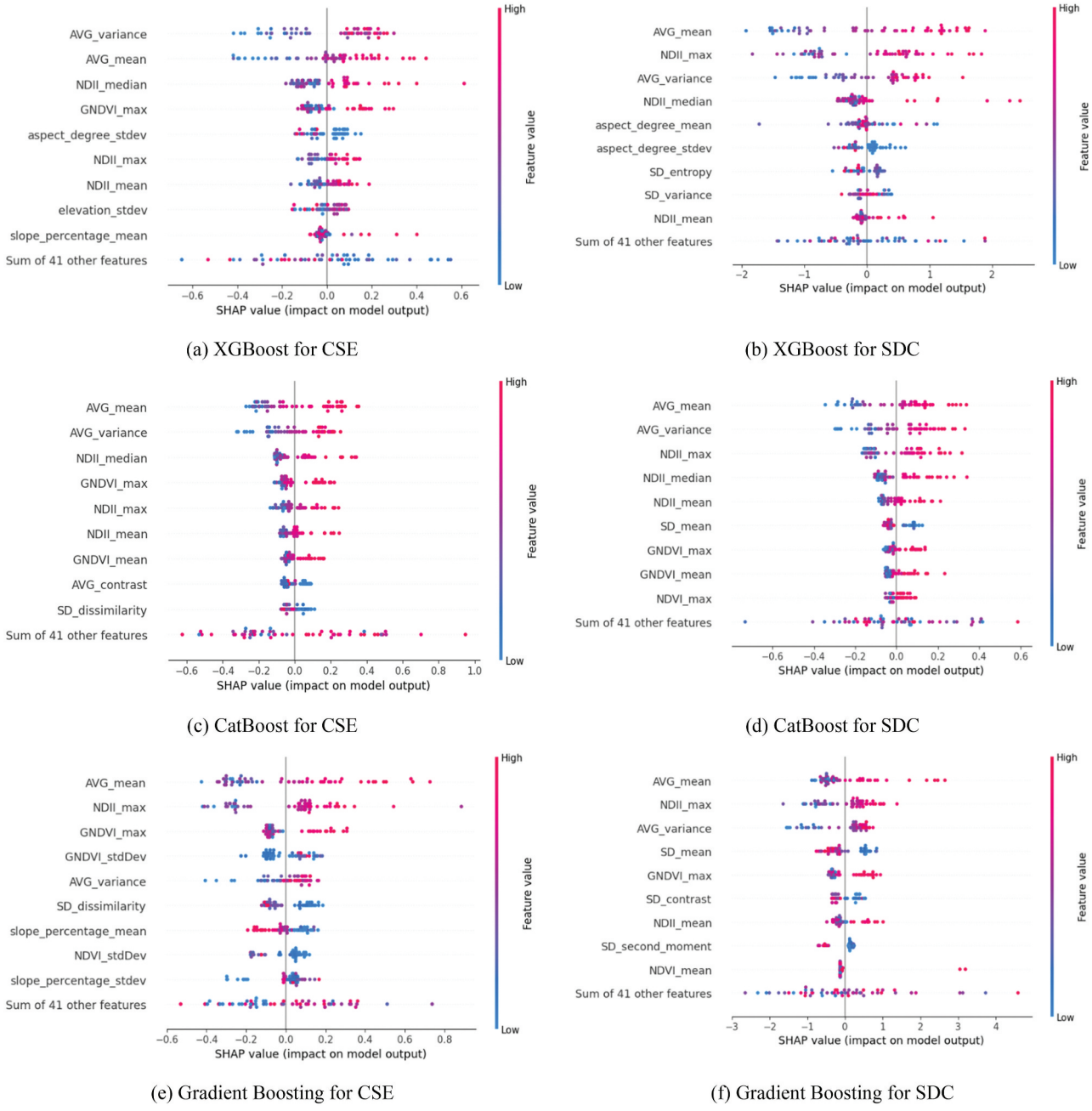



**Figure 10.** SHAP beeswarm plots for Conf 6 comparing the feature impacts on CSE (left) and SDC (right) across three different models. Subplots (a, b) correspond to XGBoost, (c, d) to CatBoost, and (e, f) to Gradient Boosting. SHAP values quantify the contribution of each feature to the model's output. The colour indicates the feature value (red = high, blue = low), and the horizontal spread reflects the magnitude and direction of each feature's impact.

configuration, significantly pushing the prediction upward. However, slope_percentage_mean and slope_-percentage_max counterbalance this by pulling the prediction down. Finally, in subplot (f) for the SDC configuration, features including SD_contrast, GNDVI_max, NDII_max, AVG_variance, and NDVI_mean prominently elevate the model's prediction. The only feature exerting a notable negative effect is AVG_mean, which pulls the prediction downward.

Figure 12 displays SHAP heatmaps for Conf 6, providing a detailed view of how feature impacts vary across individual instances for both the CSE and SDC configurations. In the CSE heatmap (left), features such as AVG_variance, AVG_mean, and NDII_median show substantial influence across the dataset, with some instances where these features consistently push the model output higher (shown in red) and others where they pull it lower (shown in blue). Additionally, features

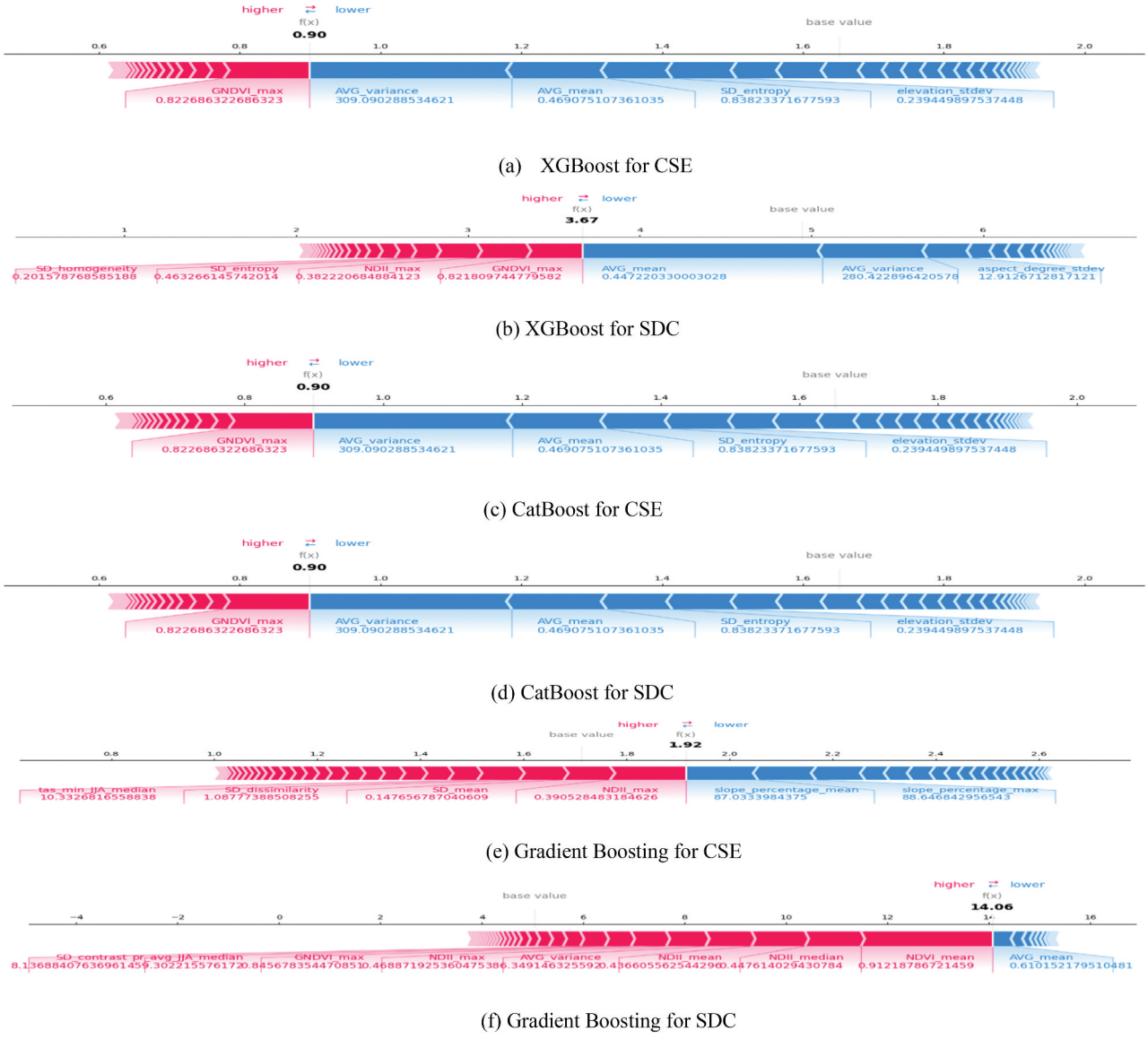



**Figure 11.** SHAP force plots for Conf 6, illustrating the SHAP values for CSE and SDC. Subplots (a, b) correspond to XGBoost, (c, d) to CatBoost, and (e, f) to Gradient Boosting. These plots show how individual features contribute to shifting the model's prediction from the base value to the final output. Red segments indicate features that increase the prediction, while blue segments represent features that decrease it, with segment width reflecting the magnitude of impact.

like GNDVI_max and NDII_max also play important but more variable roles, indicating that their contributions depend on the specific instance. In the SDC heatmap (right), AVG_mean emerges as a strong and consistent driver, often pushing the model's predictions upward. Other significant features include NDII_max, AVG_variance, and NDII_median, which, although influential, show more variability in their effects across different samples.

In the CatBoost model's heatmaps, shown in subplots (c) and (d), a similar trend is observed. In CSE (subplot c), AVG_mean, AVG_variance, NDII_median, and GNDVI_max dominate the feature impacts, with AVG_mean and NDII_median particularly standing out for their strong positive influence in later parts of the dataset. For the SDC configuration (subplot d), AVG_mean and NDII_max again play leading roles, while features like SD_entropy and SD_variance contribute more sporadically, affecting specific instances rather than consistently across all cases. In the Gradient Boosting model, illustrated in subplots (e) and (f), AVG_mean, AVG_variance, and NDII_median remain consistently influential for CSE, while spectral indices like GNDVI_max and NDII_mean also contribute notably, especially in the latter instances. In the SDC configuration (subplot f), AVG_mean and NDII_max continue to dominate, and features such as GNDVI_stdDev, SD_dissimilarity, and terrain-related metrics like

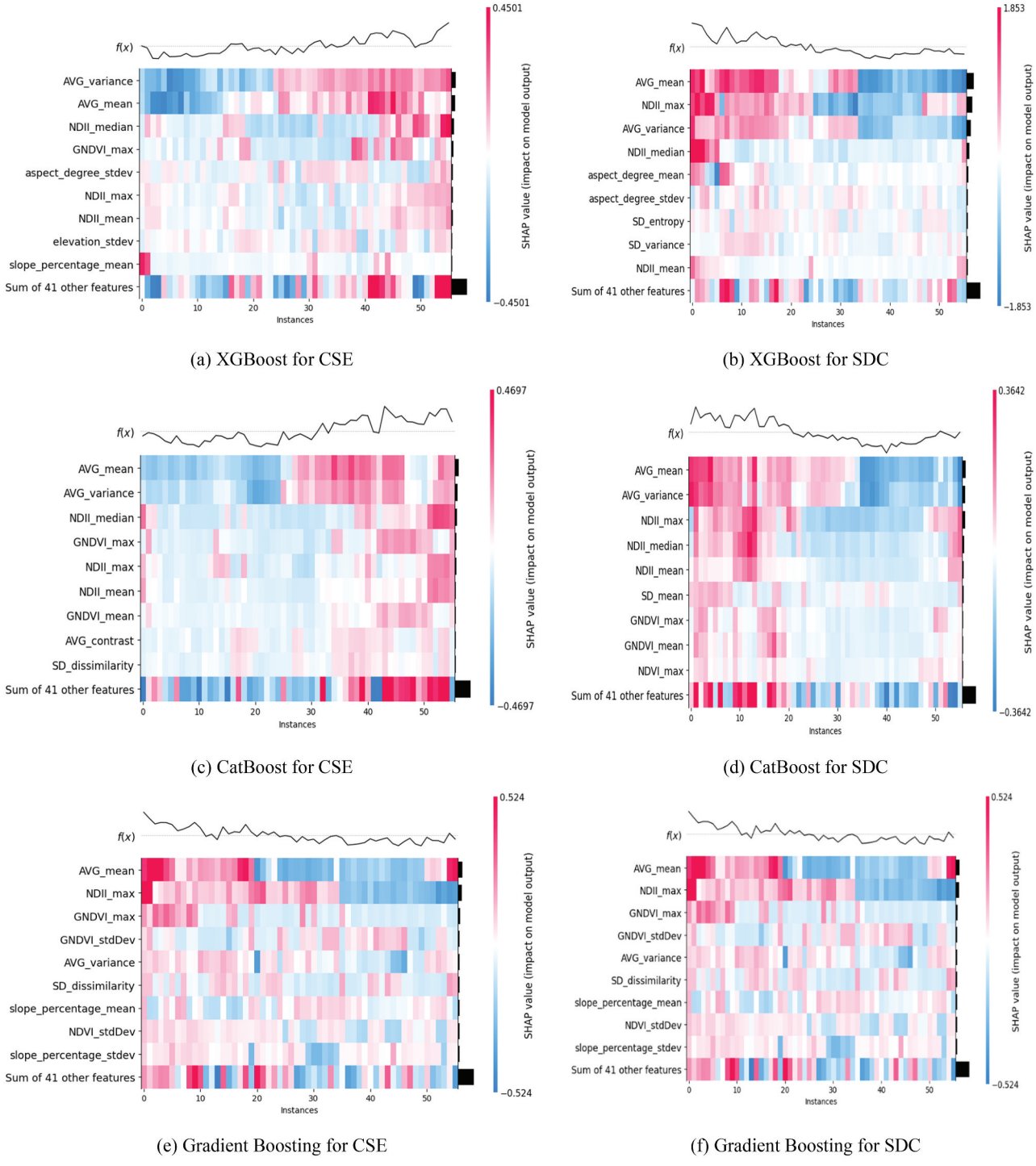



**Figure 12.** SHAP heatmaps for Conf 6, illustrating SHAP values for CSE (left) and SDC (right). Subplots (a, b) correspond to XGBoost, (c, d) to CatBoost, and (e, f) to Gradient Boosting. These heatmaps display the contribution of each feature across multiple samples, where colour intensity represents the magnitude and direction of the SHAP values, highlighting patterns of feature influence across the dataset.

slope_percentage_mean and slope_percentage_stdev offer additional explanatory power.

Overall, this SHAP analysis highlights the critical role of GLCM features, especially AVG_mean and AVG_variance, in accurately predicting OC in forest inventories. Offering an explainable framework deepens our understanding of how texture-based metrics from remote sensing data can guide sustainable forest management practices, with significant implications for carbon sequestration and climate change mitigation efforts

in Italy's forests. The most influential GLCM metrics in predicting OC include `AVG_mean`, `AVG_variance`, `SD_entropy`, and `SD_homogeneity`. These metrics are key to the model's predictions, reflecting vital aspects of forest structure and health, and their consistent importance underscores the model's effectiveness in leveraging these texture-based metrics for accurate forest inventory assessments.

## 6. Discussion

This study contributes to the growing body of research on the application of GLCM metrics in remote sensing and forestry, particularly for estimating CSE and SDC. Although GLCM metrics are well-established for evaluating the above-ground biomass of living trees, their application in estimating CSE and SDC in forested areas remains relatively underexplored. The RF model exhibited superior performance in estimating CSE ($R^2 = 0.67$, RMSE = 0.85 (tC ha −1 yr − 1), %RMSE = 53.22) and SDC ($R^2 = 0.67$, RMSE = 2.73 (tC ha − 1), %RMSE = 50.22), surpassing previous studies, including those by Fasihi et al. (2024), Safari and Sohrabi (2016), Uniyal et al. (2022), and others (Du et al. 2010; Frazier et al. 2014; Labrecque et al. 2006). These earlier studies employed machine learning models such as Random Forest and XGBoost, which generally reported higher RMSE values and comparable or lower $R^2$ scores. However, direct comparisons between studies remain challenging due to differences in forest conditions, sampling methods, and modelling approaches, as noted by prior research (Safari et al. 2017; Zandler, Brenning, and Samimi 2015).

In addition, the results support both Hypothesis 2 (H2) and Hypothesis 3 (H3). For H2, the fusion of VIs and GLCM metrics (Conf 3) consistently outperformed models using VIs alone (Conf 1) or GLCM metrics alone (Conf 2). This suggests that the complementary nature of spatial texture information from GLCM and spectral data from VIs enhances carbon estimation accuracy. Regarding H3, including geomorphological and weather parameters alongside GLCM metrics (Conf 5 and 6) further improved model performance compared to using GLCM metrics alone (Conf 2). This highlights the significant role of environmental factors, such as geomorphology and weather conditions, in carbon dynamics within forest ecosystems. Therefore, integrating these variables with GLCM metrics enhances the accuracy of OC estimations.

The discussion of key GLCM features highlights their critical role in estimating OC in forest inventories. The analysis identifies four primary features: AVG_mean, AVG_variance, SD_entropy, and SD_homogeneity, each of which provides insight into forest structure and its relationship to carbon sequestration. The metric AVG_mean represents the average intensity of pixel values, which correlates with denser forests that sequester more carbon due to greater biomass (Immitzer, Vuolo, and Atzberger 2016). Similarly, AVG_variance captures spatial variability, which reflects heterogeneity in vegetation structure, often an indicator of mature, carbon-rich forest stands (Immitzer, Vuolo, and Atzberger 2016). SD_entropy measures the randomness or disorder in texture, providing information on structural complexity, which is associated with diverse vegetation layers and increased carbon storage (Ranjan and Parida 2020). SD_homogeneity assesses the uniformity of pixel pairs; lower homogeneity suggests structural irregularity typically found in ecologically rich and carbon-dense forests with abundant standing deadwood (Ranjan and Parida 2020). Together, these features offer a nuanced view of forest structure that goes beyond spectral data alone. By explicitly linking each GLCM metric to specific forest characteristics influencing carbon dynamics, this deeper interpretation enhances the transparency and ecological relevance of the model outputs. These findings support Hypothesis 1 (H1), as they demonstrate that local variability in tree height and texture patterns, particularly as measured by GLCM metrics, provides valuable information for estimating CSE and SDC. The findings underscore the importance of these features in understanding forest ecosystems and their role in global climate change mitigation.

The study highlights the importance of integrating GLCM metrics into forestry management to enhance the accuracy and efficiency of OC estimation in forests. This approach enables more precise estimates of carbon sequestered by living trees and stored in dead trees, which is crucial for informed decision-making and sustainable forest management. By capturing complex spatial patterns linked to carbon dynamics, forestry managers can better allocate resources, focusing on areas with high OC potential or vulnerable ecosystems. In addition, an accurate estimation of the amount of standing deadwood can help forest managers improve efficiency in biodiversity conservation, given the crucial role of deadwood in supporting the diversity of multiple taxa (Parisi et al. 2018). Furthermore, the use of explainable models, like white box models, increases transparency, helping stakeholders understand key factors influencing carbon storage. This facilitates targeted interventions, such as reforestation or selective logging, to optimize carbon storage and maintain ecosystem health. While this study presents a novel and explainable machine learning approach to estimate OC in living and dead trees, several complexities and limitations must be acknowledged. First, the use of ten different machine

learning models and six data configurations introduces computational complexity. Model training, hyperparameter tuning, and validation across such a broad experimental design require time and resources. Second, although GLCM metrics offer valuable spatial texture information, they are sensitive to image resolution, noise, and preprocessing steps, which can affect the stability of feature extraction across different forest types or sensor platforms. Moreover, while SHAP values enhance model interpretability, they may not fully capture the interactions between highly correlated features, particularly in high-dimensional input spaces combining VIs, GLCM, and environmental parameters.

## 7. Conclusions

This study highlights the significant role of Gray-Level Co-occurrence Matrix (GLCM) metrics in enhancing the accuracy of organic carbon (OC) estimation through remote sensing. By integrating texture-based features with machine learning models, particularly CatBoost and RF, high levels of precision were achieved in predicting carbon storage, offering valuable insights for forest management and climate change mitigation efforts. A deeper analysis of individual GLCM features revealed their relevance to OC estimation: features such as contrast, homogeneity, and entropy effectively capture spatial heterogeneity, vegetation structure, and surface complexity – factors closely associated with organic matter distribution. These features enhance model sensitivity to ecological variations that influence carbon dynamics. The integration of SHAP values within a white-box modelling framework further strengthened the interpretability of results, offering transparency into the contributions of each variable. This is essential for building trust in data-driven decision-making processes. While the study successfully demonstrates the potential of combining remote sensing with explainable AI, it is limited by the lack of detailed soil data, which is critical for a comprehensive understanding of carbon sequestration. Future research should aim to incorporate soil characteristics to refine model performance and ecological accuracy. Moreover, future work could explore the use of Graph Neural Networks (GNNs) to model spatial and structural relationships in forest ecosystems. GNNs offer a promising approach to represent complex spatial dependencies and interactions within forest and soil environments, potentially improving the robustness of carbon estimation frameworks. In addition, we plan to design and evaluate a compact deep learning model tailored for OC estimation. This approach will prioritize simplicity, efficiency, and suitability for real-world, resource-constrained applications.

## Acknowledgements

Data on CS and SDC were obtained from the original dataset by the Carabinieri, the Council for Agricultural Research and Analysis of the Agricultural Economy – National Inventory of Forests and Forest Carbon Reservoirs (INFC), available at www.inventoryforest.org under the Creative Commons Attribution 4.0 license. Digital Elevation Model (DEM) and related data were provided by the Autonomous Region Friuli Venezia Giulia under the Italian Open Data License v2.0 (IODL 2.0). The dataset also includes modified Copernicus Sentinel data from 2019, 2020, and 2021. Climate data were sourced from the Climate Projections Platform for the northeast, produced by the Regional Agency for Environmental Prevention and Protection of Veneto and the Regional Agency for Environmental Prevention and Protection of Friuli Venezia Giulia, using data from the Royal Netherlands Meteorological Institute, Swedish Meteorological and Hydrological Institute, Max Planck Institute for Meteorology, CLM-Community, ARPAV, ARPA FVG, ARPAE Emilia-Romagna, the Autonomous Provinces of Trento and Bolzano, and the Italian Air Force. These data are available under the Creative Commons Attribution 3.0 Italy license (CC BY-SA 3.0 IT) at https://clima.arpa.veneto.it.

## Disclosure statement

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.

## Funding

This study was funded by the European Union – Next Generation EU, under Mission 4, Component 2, Inv. 1.1, as part of the PRIN 2022 project "REWILDFIRE" (CUP: G53D23001360006).

## CRediT authorship contribution statement

**Mehdi Fasihi**: Conceptualization, Methodology, Software, Validation, Writing-original draft. **Alex Falcon**: Writing – Review & Editing. **Francesca Giannetti**: Conceptualization. **Luca Cadez**: Methodology, Data curation, Writing – original draft, Writing – review & editing. **Antonio Tomao**: Conceptualization, Resources, Data Curation, Writing – Review & Editing. **Giorgio Alberti**: Conceptualization, Writing – Review & Editing, Supervision, Project administration, Funding acquisition. **Giuseppe Serra**: Conceptualization, Writing – Review & Editing, Supervision, Project administration, Funding acquisition.

## Data availability statement

The data supporting the findings of this study are available from the corresponding author upon reasonable request.

## References


Abdelhakim, D., F. Harrou, Y. Sun, S. Makhfi, and M. Habak. 2024. "Explainable Machine Learning for Enhancing Predictive Accuracy of Cutting Forces in Hard Turning Processes." *International Journal of Advanced Manufacturing Technology* 135 (1): 939–961. November. https://doi.org/10.1007/S00170-024-14470-2/FIGURES/20.

Achard, F., M. Brady, R. DeFries, G. Grassi, M. Herold, D. Mollicone, B. Mora, et al. 2012. *A Sourcebook of Methods and Procedures for Monitoring and Reporting Anthropogenic Greenhouse Gas Emissions and Removals Associated with deforestation, Gains and Losses of Carbon Stocks in Forests Remaining Forests, and Forestation: Gofc-Gold Report Version COP18-1*. Wageningen University. Accessed April 23, 2025. https://discovery.dundee.ac.uk/en/publications/a-sourcebook-of-methods-and-procedures-for-monitoring-and-reporti.

Al Saleem, M., F. Harrou, and Y. Sun. 2024. "Explainable Machine Learning Methods for Predicting Water Treatment Plant Features Under Varying Weather Conditions." *Results in Engineering* 21:101930. March. https://doi.org/10.1016/J.RINENG.2024.101930.

Asner, G. P., G. V. N. Powell, J. Mascaro, D. E. Knapp, J. K. Clark, J. Jacobson, T. Kennedy-Bowdoin. 2010. "High-Resolution Forest Carbon Stocks and Emissions in the Amazon." *Proceedings of the National Academy of Sciences of the USA* 107 (38): 16738–16742. September. https://doi.org/10.1073/PNAS.1004875107/SUPPL_FILE/PNAS.201004875SI.PDF.

Beven, K. 2018. *Environmental Modelling: An Uncertain Future?* London: CRC Press doi:https://doi.org/10.1201/9781482288575.

Blanco A. C., A. B. Baloloy, C. G. Candido, J. B. Dumalag, L. L. Dimapilis, and E. C. Paringit. 2018. "MODELLING Above GROUND BIOMASS of MANGROVE FOREST USING sentinel-1 IMAGERY Jie, Jiang." In *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 13–20. Göttingen, Germany: Copernicus GmbH. April. https://doi.org/10.5194/isprs-annals-IV-3-13-2018.

Breiman, L. 2001. "Random Forests." *Machine Learning* 45 (1): 5–32. https://doi.org/10.1023/A:1010933404324.

Brown, S. 2002. "Measuring Carbon in Forests: Current Status and Future Challenges." *Environmental Pollution* 116 (3): 363–372. March. https://doi.org/10.1016/S0269-7491(01)00212-3.

Buschmann, C. 1993. "NATUR WISSENSCHAFTE oo Fernerkundung von Pflanzen Ausbreitung, Gesundheitszustand und Produktivitfit doi:https://doi.org/10.1007/BF01136034."

Cadez, L., A. Tomao, F. Giannetti, G. Chirici, and G. Alberti. 2024. "Mapping Forest Growing Stock and Its Current Annual Increment Using Random Forest and Remote Sensing Data in Northeast Italy." *Forests* 15 (8): 1356. August. https://doi.org/10.3390/f15081356.

Charoenjit, K., P. Zuddas, P. Allemand, S. Pattanakiat, and K. Pachana. 2015. "Estimation of Biomass and Carbon Stock in Para Rubber Plantations Using Object-Based Classification from Thaichote Satellite Data in Eastern Thailand Journal of Applied Remote Sensing, 9(1) : 096072-–096072. doi:https://doi.org/10.1117/1.JRS.9.096072."

Cheng, X., A. Doosthosseini, and J. Kunkel. 2022. "Improve the Deep Learning Models in Forestry Based on Explanations and Expertise." *Frontiers in Plant Science* 13. May. https://doi.org/10.3389/fpls.2022.902105.

Cortes, C., V. Vapnik, and L. Saitta. 1995. "Support-Vector Networks." *Machine Learning* 20 (3): 273–297. September. https://doi.org/10.1007/BF00994018.

Csillik, O., P. Kumar, J. Mascaro, T. O'Shea, and G. P. Asner. 2019. "Monitoring Tropical Forest Carbon Stocks and Emissions Using Planet Satellite Data." *Scientific Reports* 9 (1). December. https://doi.org/10.1038/s41598-019-54386-6.

Dai, L., Y. Zhang, L. Wang, S. Zheng, and W. Xu. 2021. "Assessment of Carbon Density in Natural Mountain Forest Ecosystems at Northwest China." *International Journal of Environmental Research* 18 (4): 1–12. February. https://doi.org/10.3390/ijerph18042098.

Di Cosmo, L., P. Gasparini, A. Paletto, and M. Nocetti. 2013. "Deadwood Basic Density Values for National-Level Carbon Stock Estimates in Italy." *Forest Ecology & Management* 295:51–58. May. https://doi.org/10.1016/J.FORECO.2013.01.010.

Du, H., R. Cui, G. Zhou, Y. Shi, X. Xu, W. Fan, Y. Lü. 2010. "The Responses of Moso Bamboo (Phyllostachys Heterocycla Var. Pubescens) Forest Aboveground Biomass to Landsat TM Spectral Reflectance and NDVI." *Acta Ecologica Sinica* 30 (5): 257–263. October. https://doi.org/10.1016/J.CHNAES.2010.08.005.

Estoque, R. C. 2020. "A Review of the Sustainability Concept and the State of SDG Monitoring Using Remote Sensing." *MDPI AG* 12 (11): 1770. June 01. https://doi.org/10.3390/rs12111770.

Fardusi, M. J., F. Chianucci, and A. Barbati. 2017. "Concept to Practices of Geospatial Information Tools to Assist Forest Management and Planning Under Precision Forestry Framework: A Review." Ann. Silvic. Res, 41(1) : 3–14. d oi:http://dx.d oi.org/10.12899/asr-1354.

Fasihi, M., B. Portelli, L. Cadez, A. Tomao, A. Falcon, G. Alberti, G. Serra, et al. 2024. "Assessing Ensemble Models for Carbon Sequestration and Storage Estimation in Forests Using Remote Sensing Data." *Ecological Information* 83:102828. November. https://doi.org/10.1016/J.ECOINF.2024.102828.

Feurer, M., and F. Hutter. 2019. *Hyperparameter Optimization* (Cham, Switzerland: Springer International Publishing), 3–33. https://doi.org/10.1007/978-3-030-05318-5_1.

Frazier, R. J., N. C. Coops, M. A. Wulder, and R. Kennedy. 2014. "Characterization of Aboveground Biomass in an Unmanaged Boreal Forest Using Landsat Temporal Segmentation Metrics." *ISPRS Journal of Photogrammetry and Remote Sensing* 92:137–146. June. https://doi.org/10.1016/J.ISPRSJPRS.2014.03.003.

Friedman, J. H. 2001. "Greedy Function Approximation: A Gradient Boosting Machine." *Annals of Statistics*: 1189–1232. https://www.jstor.org/stable/2699986.

Gao, Y., D. Lu, G. Li, G. Wang, Q. Chen, L. Liu, D. Li, et al. 2018. "Comparative Analysis of Modeling Algorithms for Forest Aboveground Biomass Estimation in a Subtropical Region." *Remote Sensing (Basel)* 10 (4): 627. https://doi.org/10.3390/rs10040627.

Gasparini, P., and G. Papitto. 2022. "The Italian Forest Inventory in Brief: L'inventario forestale nazionale italiano in breve." *Springer Tracts in Civil Engineering*: 1–15. https://doi.org/10.1007/978-3-030-98678-0_1/TABLES/1.

Ghafarian, F., R. Wieland, D. Lüttschwager, and C. Nendel. 2022. "Application of Extreme Gradient Boosting and Shapley Additive Explanations to Predict Temperature Regimes Inside Forests from Standard Open-Field Meteorological Data." *Environmental Modelling and Software* 156:105466. October. https://doi.org/10.1016/j.envsoft.2022.105466.

Gitelson, A. A., and M. N. Merzlyak. 1998. "Remote Sensing of Chlorophyll Concentration in Higher Plant Leaves." *Advances in Space Research* 22 (5): 689–692. January. https://doi.org/10.1016/S0273-1177(97)01133-2.

Gorelick, N., M. Hancher, M. Dixon, S. Ilyushchenko, D. Thau, and R. Moore. 2017. "Google Earth Engine: Planetary-Scale Geospatial Analysis for Everyone." *Remote Sensing of Environment* 202:18–27. December. https://doi.org/10.1016/J.RSE.2017.06.031.

Haralick, R. M., I. Dinstein, and K. Shanmugam. 1973. "Textural Features for Image Classification." *IEEE Transactions on Systems, Man, and Cybernetics* SMC-3 (6): 610–621. https://doi.org/10.1109/TSMC.1973.4309314.

Haralick, R. M., K. Shanmugam, and I. H. Dinstein. 1973. "Textural_features_for_image_classification." *IEEE Transactions on Systems, Man, and Cybernetics* ".SMC-3(6) : 610–621 doi:https://doi.org/10.1109/TSMC.1973.4309314

Harmon, M. E., J. F. Franklin, F. J. Swanson, P. Sollins, S. V. Gregory, J. D. Lattin, N. H. Anderson, et al. 1986. "Ecology of Coarse Woody Debris in Temperate Ecosystems." *Advances in Ecological Research* 15:133–302. https://doi.org/10.1016/S0065-2504(08)60121-X

Hesterberg, T. C. 2015. "What Teachers Should Know About the Bootstrap: Resampling in the Undergraduate Statistics Curriculum." *American Statistician* 69 (4): 371–386. https://doi.org/10.1080/00031305.2015.1089789.

Huang, S., L. Tang, J. P. Hupy, Y. Wang, and G. Shao. 2021. "A Commentary Review on the Use of Normalized Difference Vegetation Index (NDVI) in the Era of Popular Remote Sensing." *Journal of Forestry* 32 (1): 1–6. February. https://doi.org/10.1007/s11676-020-01155-1.

Huete, A., K. Didan, T. Miura, E. P. Rodriguez, X. Gao, and L. G. Ferreira. 2025. "Overview of the Radiometric and Biophysical Performance of the MODIS Vegetation Indices." www.elsevier.com/locate/rse.

Immitzer, M., F. Vuolo, and C. Atzberger. 2016. "First Experience with Sentinel-2 Data for Crop and Tree Species Classifications in Central Europe." *Remote Sensing* 8 (3): 166. February. https://doi.org/10.3390/RS8030166.

Konda, R. K., V. Giri, and V. R. Mandla. 2017. "Study and Evaluation of Carbon Sequestration Using Remote Sensing and GIS: A Review on Various Techniques." *International Journal of Civil Engineering & Technology* 8 (4): 8–12. http://www.iaeme.com/IJCIET/issues.asp?JType=IJCIET&VType=8&IType=4 .

Krug, T., W. A. Kurz, R. D. Lasco, D. L. Martino, and B. G. McConkey. 2006 Generic Methodologies Applicable to Multiple Land-Use Categories. " 2006 IPCC Guidelines for National Greenhouse Gas Inventories, Volume 4: Agriculture, Forestry and Other Land Use 4."

Labrecque, S., R. A. Fournier, J. E. Luther, and D. Piercey. 2006. "A Comparison of Four Methods to Map Biomass from Landsat-Tm and Inventory Data in Western Newfoundland." *Forest Ecology & Management* 226 (1–3): 129–144. May. https://doi.org/10.1016/J.FORECO.2006.01.030.

Li, J., B. He, S. Ahmad, and W. Mao. 2023. "Leveraging Explainable Machine Learning Models to Assess Forest Health: A Case Study in Hainan, China." *Ecology Evolution* 13 (9). September. https://doi.org/10.1002/ece3.10558.

Li, X., J. Long, M. Zhang, Z. Liu, and H. Lin. 2021. "Coniferous Plantations Growing Stock Volume Estimation Using Advanced Remote Sensing Algorithms and Various Fused Data." *Remote Sensing (Basel)* 13 (17): 3468. September. https://doi.org/10.3390/rs13173468.

Li, Z., Q. Zan, Q. Yang, D. Zhu, Y. Chen, and S. Yu. 2019. "Remote Estimation of Mangrove Aboveground Carbon Stock at the Species Level Using a Low-Cost Unmanned Aerial Vehicle System." *Remote Sensing (Basel)* 11 (9): 1018. May. https://doi.org/10.3390/rs11091018.

Liang, J., J. G. P. Gamarra, N. Picard, M. Zhou, B. Pijanowski, D. F. Jacobs, P. B. Reich, et al. 2022. "Co-Limitation Towards Lower Latitudes Shapes Global Forest Diversity Gradients." *Nature Ecology & Evolution* 6 (10): 1423–1437. https://doi.org/10.1038/s41559-022-01831-x.

Lindenmayer, D. B., W. F. Laurance, and J. F. Franklin. 2012. "Global Decline in Large Old Trees." *Science* 338 (6112): 1305–1306. https://doi.org/10.1126/science.1231070.

Lu, D., and Q. Weng. 2007. "A Survey of Image Classification Methods and Techniques for Improving Classification Performance." *Taylor and Francis Ltd* 28 (5): 823–870. https://doi.org/10.1080/01431160600746456.

Lundberg, S. M., P. G. Allen, and S.-I. Lee. 2022. "A Unified Approach to Interpreting Model Predictions." https://github.com/slundberg/shap.

Lundberg, S. M., G. Erion, H. Chen, A. DeGrave, J. M. Prutkin, B. Nair, R. Katz, J. Himmelfarb, N. Bansal, and S.-I. Lee. 2020. "From Local Explanations to Global Understanding with Explainable AI for Trees." *Nature Machine Intelligence* 2 (1): 56–67. January. https://doi.org/10.1038/s42256-019-0138-9.

Masera, O. R., J. F. Garza-Caligaris, M. Kanninen, T. Karjalainen, J. Liski, G. J. Nabuurs, A. Pussinen, et al. 2003. "Modeling Carbon Sequestration in afforestation, Agroforestry and Forest Management Projects: The CO2FIX V.2 Approach." *Ecological Modelling* 164 (2–3): 177–199. June. https://doi.org/10.1016/S0304-3800(02)00419-2.

Montgomery, D. C., E. A. Peck, and G. G. Vining. 2021. *Introduction to Linear Regression Analysis*. 6th ed. Hoboken, New Jersey: John Wiley & Sons.

Nguyen, T. A., B. Kellenberger, and D. Tuia. 2022. "Mapping Forest in the Swiss Alps Treeline Ecotone with Explainable Deep Learning." *Remote Sensing of Environment* 281:113217. November. https://doi.org/10.1016/j.rse.2022.113217.

Pan, Y., R. A. Birdsey, J. Fang, R. Houghton, P. E. Kauppi, W. A. Kurz, O. L. Phillips, et al. 2011. "A Large and Persistent Carbon Sink in the world's Forests." *Science* 333 (6045): 988–993. August. https://doi.org/10.1126/SCIENCE.1201609/SUPPL_FILE/PAPV2.PDF.

Pan, Y., R. A. Birdsey, O. L. Phillips, R. A. Houghton, J. Fang, P. E. Kauppi, H. Keith, et al. 2024. "The Enduring World Forest Carbon Sink." *Nature* 631 (8021): 563–569. July. https://doi.org/10.1038/s41586-024-07602-x.

Parisi, F., S. Pioli, F. Lombardi, G. Fravolini, M. Marchetti, and R. Tognetti. 2018. "Linking Deadwood Traits with Saproxylic

Invertebrates and Fungi in European Forests - a Review." *iForest - Biogeosciences and Forestry* 11 (3): 423. June. https://doi.org/10.3832/IFOR2670-011.

Ranjan, A. K., and B. R. Parida. 2020. "Estimating Biochemical Parameters of Paddy Using Satellite and Near-Proximal Sensor Data in Sahibganj Province, Jharkhand (India)." *Remote Sensing Applications: Society & Environment* 18:100293. April. https://doi.org/10.1016/J.RSASE.2020.100293.

Rehman, A. N., and B. Lal. 2023. "Machine Learning in CO2 Sequestration Bhajan, Lal." In *Machine Learning and Flow Assurance in Oil and Gas Production*, 119–140. Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-24231-1_7.

Safari, A., and H. Sohrabi. 2016. "Ability of Landsat-8 OLI Derived Texture Metrics in Estimating Aboveground Carbon Stocks of Coppice Oak Forests." In *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, 751–754. Zürich, Switzerland: International Society for Photogrammetry and Remote Sensing. https://doi.org/10.5194/isprsarchives-XLI-B8-751-2016.

Safari, A., H. Sohrabi, S. Powell, and S. Shataee. 2017. "A Comparative Assessment of Multi-Temporal Landsat 8 and Machine Learning Algorithms for Estimating Aboveground Carbon Stock in Coppice Oak Forests." *International Journal of Remote Sensing* 38 (22): 6407–6432. November. https://doi.org/10.1080/01431161.2017.1356488.

Schlesinger, W. H. 1997. "Biogeochemistry: An Analysis of Global Change." *Biogeochemistry* 3:332–337.

Shao, J., and D. Tu. 2012. *The Jackknife and Bootstrap*. New York: Springer Science & Business Media.

Sims, D. A., A. F. Rahman, V. D. Cordova, B. Z. El-Masri, D. D. Baldocchi, L. B. Flanagan, A. H. Goldstein, et al. 2006. "On the Use of MODIS EVI to Assess Gross Primary Productivity of North American Ecosystems." *Journal of Geophysical Research Biogeosciences* 111 (4). December. https://doi.org/10.1029/2006JG000162.

Sun, W., and X. Liu. 2020. "Review on Carbon Storage Estimation of Forest Ecosystem and Applications in China." *Forest Ecosystems* 7 (1): 1–14. December. https://doi.org/10.1186/s40663-019-0210-2.

Tabacchi, G., L. Di Cosmo, and P. Gasparini. 2011. "Aboveground Tree Volume and Phytomass Prediction Equations for Forest Species in Italy." *European Journal of Forest Research* 130 (6): 911–934. November. https://doi.org/10.1007/s10342-011-0481-9.

Uniyal, S., S. Purohit, K. Chaurasia, S. S. Rao, and E. Amminedu. 2022. "Quantification of Carbon Sequestration by Urban Forest Using Landsat 8 OLI and Machine Learning Algorithms in Jodhpur, India." *Urban Forestry and Urban Greening* 67:127445. January. https://doi.org/10.1016/j.ufug.2021.127445.

U.S. Geological Survey. "What is Remote Sensing and What is it Used For?" Accessed April 21, 2023. https://www.usgs.gov/faqs/what-remote-sensing-and-what-it-used.

Vangi, E., G. D'Amico, S. Francini, C. Borghi, F. Giannetti, P. Corona, M. Marchetti, et al. 2023. "Large-Scale High-Resolution Yearly Modeling of Forest Growing Stock Volume and Above-Ground Carbon Pool." *Environmental Modelling & Software* 159:105580. January. https://doi.org/10.1016/J.ENVSOFT.2022.105580.

Wijaya, A., S. Kusnadi, R. Gloaguen, and H. Heilmeier. 2010. "Improved Strategy for Estimating Stem Volume and Forest Biomass Using Moderate Resolution Remote Sensing Data and GIS." *Journal of Forestry* 21 (1): 1–12. https://doi.org/10.1007/s11676-010-0001-7.

Xue, J., and B. Su. 2017. "Significant Remote Sensing Vegetation Indices: A Review of Developments and Applications." *Journal of Sensors* 2017:1–17. https://doi.org/10.1155/2017/1353691.

Yona, L., B. Cashore, R. B. Jackson, J. Ometto, and M. A. Bradford. 2020. "Refining National Greenhouse Gas Inventories." *AMBIO: A Journal of the Human Environment* 49 (10): 1581–1586. October. https://doi.org/10.1007/s13280-019-01312-9.

Zandler, H., A. Brenning, and C. Samimi. 2015. "Quantifying Dwarf Shrub Biomass in an Arid Environment: Comparing Empirical Methods in a High Dimensional Setting." *Remote Sensing of Environment* 158:140–155. March. https://doi.org/10.1016/J.RSE.2014.11.007.

Zhao, P., D. Lu, G. Wang, C. Wu, Y. Huang, and S. Yu. 2016. "Examining Spectral Reflectance Saturation in Landsat Imagery and Corresponding Solutions to Improve Forest Aboveground Biomass Estimation." *Remote Sensing (Basel)* 8 (6): 469. https://doi.org/10.3390/rs8060469.

Zhou, Y., and Z. Feng. 2023. "Estimation of Forest Stock Volume Using Sentinel-2 MSI, Landsat 8 OLI Imagery and Forest Inventory Data." *Forests* 14 (7): 1345. July. https://doi.org/10.3390/f14071345.