Machine Learning Method for Carbon Stock Classification with Drone and GEE Data

1st Mohammad Daffa Setiawan
School of Computing
Telkom University
Bandung, Indonesia
daffasetiawan@student.telkomuniversity.ac.id

2nd Erwin Budi Setiawan
School of Computing
Telkom University
Bandung, Indonesia
erwinbudisetiawan@telkomuniversity.ac.id

Abstract*— Accurately classifying carbon stock is essential for tackling climate change, as it helps improve forest management and carbon storage efforts. However, traditional measurement methods are often costly, time-consuming, and require extensive fieldwork. To address these challenges, remote sensing combined with machine learning offers a more efficient and scalable solution. This study explores the use of XGBoost and Random Forest classifiers to classify carbon stock levels using drone and Google Earth Engine (GEE) imagery, with VGG16 extracting features from the images. The dataset, collected from field plots at Telkom University in Bandung, Indonesia, consists of 2,114 drone images and 2,526 GEE images, labeled into three categories: low, medium, and high carbon stock. The results show that XGBoost applied to drone imagery achieves the highest accuracy of 90.79%, outperforming Random Forest and GEE-based models. This study underscores the potential of deep learning and ensemble methods in improving carbon stock estimation, supporting better environmental conservation.

***Keywords*— Carbon Stock, Drone Imagery, GEE, Random Forest, VCG16, XGBoost.**

I. INTRODUCTION

Carbon stock refers to the amount of carbon accumulated in a specific ecosystem or reservoir, such as forests, soil, oceans, and the atmosphere [1]. Carbon can exist in various forms, including biomass (trees, plants, and animals), soil organic matter, and sediments [2]. The amount of carbon stored in an ecosystem has a significant impact on our planet. Fluctuations in carbon stock can lead to various negative consequences, such as global temperature rise and land degradation [3]. Therefore, measuring the amount of carbon in natural ecosystems is crucial for determining appropriate actions to mitigate the harmful effects caused by carbon such as High Carbon Stock Approach, which is practical tool to identify and protect tropical forests under threat from agricultural expansion.

To measure the carbon stock in a soil, two of which are direct soil sampling [4] and remote sensing techniques, which involve collecting and analyzing information about objects or areas from a distance [5]. However, direct soil sampling is considered a high-cost traditional method and is also regarded as ineffective because calculation in traditional method requires to measure vegetative biomass such as species type, tree diameter, and tree height [6]. It is considered ineffective for several reasons, including the challenges of accessing difficult terrain, the significant effort required to reach remote areas, and the necessity of advanced tools to conduct sampling at the site [7].

To address those limitations, remote sensing technology offers a scalable alternative to traditional methods for carbon stock estimation using aerial or satellite imagery. However, extracting meaningful insights from raw data requires advanced machine learning techniques, which have proven effective in processing aerial imagery for accurate large-scale environmental analysis. Several studies have utilized remote sensing technology to measure carbon stock, employing various machine learning methods with aerial imagery datasets, demonstrating its effectiveness in capturing large-scale environmental data for accurate carbon stock estimation [8][9]. A study compared several machine learning algorithms to predict carbon stock, including Bagging (Bootstrap Aggregating), AdaBoost (Adaptive Boosting), XGBoost (Extreme Gradient Boosting), and Random Forest, concluding that XGBoost was the best-performing method [5]. Other studies also compared Random Forest (RF), Classification and Regression Trees (CART), Gradient Boosting Trees (GBT), and Support Vector Machine (SVM) for carbon stock estimation with remote sensing method, concluding that Random Forest is the best-performing algorithm compared to other machine learning methods using USGS Landsat 8 Level 2 [10]. Also, research using the XGBoost model to predict SOCS (Soil Carbon Stock) resulting a satisfactory outcome, which conclude that the XGBooost model has been very efficient to predict the SOCS using the combination of Sentinel-1 and Sentinel-2 images [11].

Based on a previous study, a remote sensing method often use the aerial imagery datasets, normally using the satellite images. To access those images, Google Earth Engine or often called GEE provides access to a vast repository of satellite imagery, enabling the analysis of land cover, which is critical for carbon stock measurements [12]. GEE offers free and open access to a remote sensing data with various of datasets such as MODIS, Sentinel, and Landsat that has been used in the previous study [10]. This makes it easier to obtain data from various satellites for remote sensing purposes especially in carbon stock measurements aspect. Other aerial imagery datasets can also obtain with drone or UAV that offers high spatial resolution and flexibility, making it well-suited for capturing detailed, localized data on vegetation structure and soil properties, which are essential for more precise carbon stock estimation [13]. Both datasets are capable of capturing key environmental parameters, such as land cover types, vegetation indices, and biomass density, which play a crucial role in monitoring carbon storage.

Machine learning models that excel in terms of predicting a soil carbon stock with a remote sensing technology are XGBoost and Random Forest based on the previous study [10] [11]. Both models perform a better result compared to other machine learning models such as AdaBoost and SVM. The XGBoost model has been used in various field due to its high accuracy, stability, and lack of overfitting [14]. The Random