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

Forest model also proven to be an effective model in conjunction with advanced feature extraction methods for image classification tasks [15]. Both models have good potential to classify carbon stock with the aerial imagery datasets. However, most research has not used the classification method for carbon stock prediction, and the use of feature extraction in this domain remains unexplored.

In terms of carbon stock classification, for remote sensing algorithms, feature extraction is a crucial step for reducing redundancy and enhancing discriminative information [16]. The study [17] mentions that the use of feature extraction can improve model accuracy. One of the famous feature extractors is VGG16, a model that has been trained with a dataset consisting of thousands of image categories and has the ability to learn a feature from an image such as spatial, edges, shapes, etc [18]. This pre-trained model has been proven to provide an effective image feature representation for classification [19]. The use of this model is expected to have a significant impact on extracting aerial imagery datasets into features that can be used by the classifier later.

The study proposed identifying the most effective machine learning classifier between two widely recognized models, XGBoost and Random Forest, for remote sensing-based carbon stock classification, utilizing VGG16 as the feature extractor. Furthermore, this study examines the performance of two distinct datasets, which are GEE and Drone images, to determine the most suitable dataset for carbon stock classification in remote sensing applications. Both datasets have their respective strengths and limitations concerning data acquisition and image quality. To the best of the author's knowledge, a remote sensing carbon stock classification with machine learning method remains relatively unexplored and few studies utilized feature extractors in classification tasks. Comparison or combination of the GEE and Drone datasets are also still uncommon. Therefore, the main contribution of this study lies in the development of an efficient and high accuracy method for carbon stock classification by systematically evaluating the combination of machine learning models and datasets.

II. Proposed Method

This following Figure 1 is the proposed method for this study

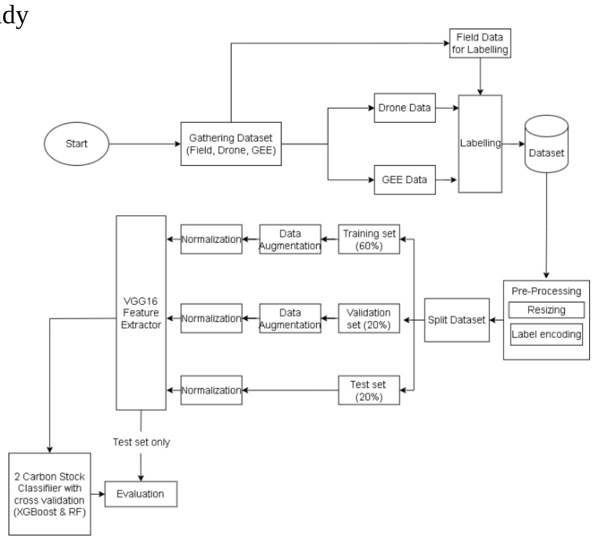

Fig. 1. Carbon classification flow

A. Dataset Preparation

The carbon stock data are collected directly from the field of Telkom University Area located in Bandung, Indonesia for dataset labeling, alongside aerial footage from both GEE and drones. The Indonesian National Standard (SNI) for Carbon Stock Measurement and Estimation serves as the primary guideline for data collection. This involves creating square plots with dimensions of 20 m x 20 m divide it into four sub-plot variations: Sub-plot A (1 m x 1 m), Sub-plot B (5 m x 5 m), Sub-plot C (10 m x 10 m), and Sub-plot D (20 m x 20 m) as shown on Figure 2. Each sub-plot contains distinct samples for measurement purposes [20]. A total of six plots were successfully acquired from the Telkom University area.

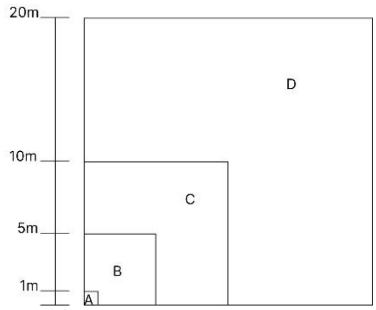

Fig. 2. SNI carbon measurement square plot

The coordinates of each plot were recorded during field sampling to assist the collection of aerial imagery from GEE and drones. Drone images were captured using a DJI Mavic 2 Pro with a Hasselblad L1D-20c camera at 100 meters altitude. This data acquisition provided localized, low-distortion imagery, enhancing VGG16's feature extraction and improving accuracy and model robustness which produces very high resolution images. Those images then labeled with the corresponding amount of carbon for each plot with the following format: DATA SOURCE - ZxHxPx - CARBON AMMOUNT

TABLE I. SAMPLE OF DATASET LABELLING

Data	Information		
	Aerial Imagery	Carbon Amount (Kg)	Labelled Dataset
GEE		2.826 Kg	 G-Z4H1P1-2826
Drone		2.826 Kg	 D-Z4H1P1-2826