Modelling Canopy Structure of Forest using Big Geospatial Data and Deep Learning

Sunil S Fatehpur*

College of Military Engineering, Pune, India
*Corresponding author's email: sunilsfathepur@rediffmail.com

Abstract

The forestry sector cannot only sustain its carbon but also has the potential to absorb carbon from the atmosphere. To ascertain the carbon potential of the forest, analyzing the forest structural properties is very important because its structure is essential for comprehending and measuring forest biophysical functions. Canopy Height is an important parameter for measuring the above ground plant biomass accurately. At the local scale, forest stand attributes, such as tree girth, the shape of a crown, height, and tree architecture, are the defining characteristics of forest canopy structure. And Ground based methods are used to collect data on local forest canopy structure. At a regional scale, the number of trees, shrubs, herbs, their species, and their arrangements give the forest canopy structure. At this large scale, forest canopy structure defines the biome and its conditions, and it tells us whether the forest is Continuous/patchy or pristine/disturbed. Coniferous, scrub, grassland, broadleaved, etc. For which Remote sensing is the preferred method of collecting data about forest canopy structure at large scales. This study explores methodology for deriving canopy height on large scale using Light Detection and Ranging (LiDAR) technology using Global Ecosystems Dynamics (GEDI) sensor for the forest stands of Tadoba Tiger Reserve Chandrapur, Maharashtra. To achieve the objective different machine learning models would be prepared to ascertain the canopy height, PAI (Plant Area Index) and PAVD (Plant Area Volume Density) using the regression methods and ascertain the suitable model. The output from the models which will be taken as target variable and Vegetation indices such as NDVI (Normalized Difference Vegetation Index), NDWI (Normalized Difference Water Index) and GCVI (Green Chlorophyll Vegetation Index), Slope, Precipitation will be taken as predictor variables to calculate the carbon sequestration. This study will be useful to estimate carbon sequestration potential of Tadoba forest. This inturn will also help us to address the parameters of UN REDD+ (United Nation Reduction Emission from Deforestation in developing countries).

Keywords Canopy Height, PAI, PAVD, GEDI LiDAR, Regression, Machine Learning, carbon sequestration

1. Introduction

1.1 Tadoba Tiger Reserve

Tadoba Tiger Reserve is a wildlife sanctuary located in Chandrapur district of Maharashtra. Tadoba is Maharashtra's oldest and largest national park. The total area of the reserve is 625.4 square kilometers (241.5 sq mi). This includes Tadoba National Park, with an area of 116.55 square kilometers (45.00 sq mi) and Andhari Wildlife Sanctuary with an area of

508.85 square kilometers (196.47 sq mi). The reserve also includes 32.51 square kilometers (12.55 sq mi) of protected forest and 14.93 square kilometers (5.76 sq mi) of uncategorized land.

1.1.2 Flora

Tadoba Reserve is a predominantly southern tropical dry deciduous forest with dense woodlands comprising about eighty-seven per cent of the protected area. Teak is the predominant tree species. Other deciduous trees found in this area include ain (crocodile bark), bija, dhauda, hald, salai, semal and tendu. Beheda, hirda, karaya gum, mahua madhuca (crepe myrtle), palas (flame-of-the-forest, Butea monosperma) and Lannea coromandelica (wodier tree). Axlewood (Anogeissus latifolia, a fire-resistant species), black plum and arjun are some of the other tropical trees that grow in this reserve.Patches of grasses are found throughout the reserve. Bamboo thickets grow throughout the reserve in abundance. The climber kach kujali (velvet bean) found here is a medicinal plant used to treat Parkinson's disease. The leaves of bheria are used as an insect repellent and bija is a medicinal gum. Beheda is also an important medicine found here.

1.1.3 Fauna

Tadoba consists of a wide variety of fauna. Bengal Tiger is the keystone species in Tadoba tiger reserve. Apart from keystone species several other varieties of mammals are also present including Indian leopards,[8] sloth bears, gaur, nilgai, dhole, small Indian civet, jungle cats, sambar, barking deer, chital, chausingha and honey badger.

The estimation of canopy height has always been a topic of research in forest mapping, quality of habitat, biodiversity etc. (1). Forests play a major role in the ecosystem. A major amount of carbon stocks is found in the forests which are decreasing rapidly. Now the question arises what are carbon stocks? The quantity of carbon stored in the trees whether it is in the form of biomass, litter etc. is referred to as carbon stocks. This concern about decreasing amount of carbon stocks and the global climate change have emphasized the importance of finding out the systematic ways. Estimation of forest carbon stock depends on the accurate estimation of the forest structure of forest. Remote sensing technology is being used widely as it is synoptic nature (2). Remote sensing can capture the data very easily at the large spatial scales which is why the traditional field survey methods are replaced with it (3). Due to the simplicity of collecting data at very wide spatial scales, Remote sensing has largely replaced the field survey methods for several applications (3).

LiDAR remote sensing provides the vegetation structure in three dimensions as it captures spatial patterns of canopy height accurately (1). Previously canopy height was used to be measured in the field but the traditional methods are expensive so they cannot be used widely. LiDAR is far better as compared to that of the optical and microwave remote sensing at predicting the canopy height or biomass (1). It is an active remote sensing technology that emits a laser light pulse in the direction of surface (1).

The Global Ecosystem Dynamics Investigation LiDAR was designed to estimate the canopy height, branches etc. It was launched on 5th of September 2018. It is a high-resolution full waveform space borne LiDAR. GEDI comprises of 3 lasers with 1 splitting into 2

producing 4 beams. 8 beams are formed out of which 4 are Full Power Beams and 4 are Coverage Beams.

- Coverage Beams: Coverage Beams can penetrate canopies with up to 95 % canopy coverage.
- Full Power Beams: They can be used in the dense forests.

The diameter of the GEDI footprint is 25 m and the distance between two tracks is 600 m. GEDI provides the 3D map of the canopies.

1.1.4 Study Area

Tadoba Tiger Reserve is located in Chandrapur district of Maharashtra. It lies in 20.2484° N, 79.3607° E. The reserve consists of 577.96 square kilometers of reserved forest and 32.51 square kilometers of protected forest. The most popular species of the trees are Teak and Bamboo in this forest. Other trees include Ain (crocodile bark), Bija, Dhauda, Hald, Salai, Semal and Tendu. Beheda, Hirda, Karaya Gum, Mahua Madhuca (crepe myrtle) and Lannea coromandelica are other common species. Bengal Tiger is the keystone species in the Tadoba Tiger Reserve and apart from that other mammal species are also there.

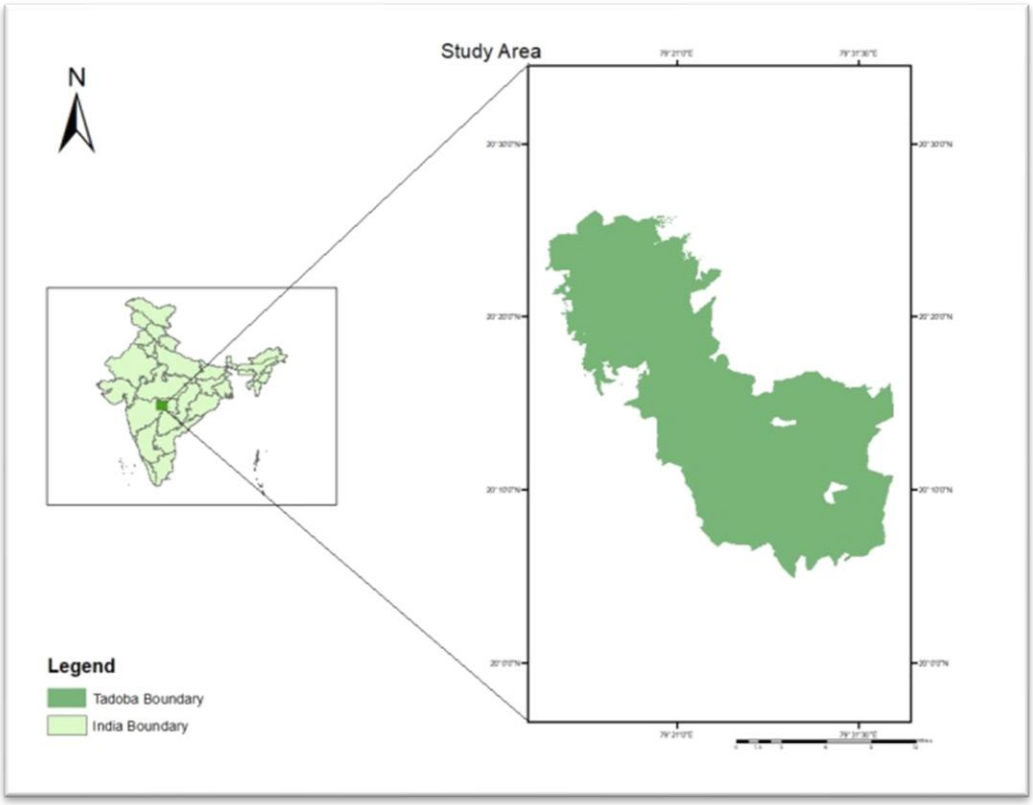


Fig.1.1 Location map Of Tadoba.

2. Materials and Methods

2.1 Datasets

2.1.1 Global Ecosystem Dynamics Investigation (GEDI)

GEDI stands for Global Ecosystem Dynamics Investigation. It is LiDAR mission which was launched by NASA on 5 December 2018. It is a high-resolution full waveform space borne LiDAR. GEDI comprises of 3 lasers which emits the wavelength of 1064 nm.1 laser split into 2 producing 4 beams. Beams form the eight ground tracks out of which 4 are full power beams

and 4 are coverage beams. Coverage beams were made to penetrate the canopies up to 95 % under typical conditions. It is advised to use GEDI full power beams in dense forests. The coverage lasers are also affected by the solar noise effects. The diameter of the GEDI footprint is 25 m and the distance between two tracks is 600 m.

Fig.2.1 GEDI.

Table 2.1 *Specifications of GEDI.*

Characteristic	Description
Platform	International Space Station
Type	Full waveform LiDAR
Launched on	5 December 2018
Coverage extent	51.6 N and S Latitude
Resolution	High Resolution
Consists of	lasers with 1 splitting into 2 producing beams
Pulse	60 m
Swath width	600 m
Diameter of footprint	25 m

The Land Processes Distributed Active Archive Centre of NASA provides the GEDI-collected waveforms (Level 1) and their processed data product (Level 2) (13). The geolocated waveforms as gathered by the GEDI system are included in the Level 1 data product, specifically the L1B data product (13). The Level 2 data is classified into two further levels i.e., Level 2A and Level 2B. Level 2A data product includes the ground elevation, canopy top height, relative return energy metrics and many other interpreted products from the return waveforms. The main purpose of L2A dataset is to provide the waveform interpretation and extracted products from each GEDI waveform. Level 2B data product includes footprint level canopy cover and vertical profile metrics. The L2B dataset's purpose is to take biophysical measurements from each GEDI waveform. These measurements, which include canopy cover, Plant Area Index (PAI), Plant Area Volume Density (PAVD), and Foliage Height Diversity, are based on the directional gap probability profile obtained from the L1B waveform. Level 3 dataset of Global Ecosystem Dynamics Investigation offers counts of laser footprints per 1 km x 1 km grid cell globally between -52- and 52-degrees latitude. Level 3 dataset also provides the gridded mean canopy height, standard deviation of canopy height, mean ground elevation, and ground elevation. L3 gridded products have enormous significance for climate modelling, forest management, snow and glacier monitoring, and the

creation of DEMs. They can also be used to quantify significant carbon and water cycling processes, biodiversity, habitat, and other critical activities. Level 4A dataset provides the estimates of Above Ground Biomass density (AGBD in Mg/ha) and it also estimates the prediction standard error laser footprint. 1km*1km estimates of mean above ground biomass estimates are provided by GEDI L4B data product.

Table 2.2 GEDI Geolocated Waveforms Product (L1B).

Characteristic	Description
Collection	GEDI
Short Name	GEDI01_B
DOI	10.5067/GEDI/GEDI01_B.002
Temporal Resolution	Varies
Temporal Extent	2019-04-18 – Present
Spatial Extent	Global
Coordinate System	51.6 N and S Latitude
Datum	Geographic (lat/long)
Geographic Dimensions	4.2 km cross track by one-fourth of an ISS orbit along track
File Size	~2GB
File Format	HDF5

Table 2.3 GEDI Elevation and Height Metrics Data Global Footprint Level (L2A).

Characteristic	Description
Collection	GEDI
Short Name	GEDI02_A
DOI	10.5067/GEDI/GEDI02_A.001
Temporal Resolution	Varies
Temporal Extent	2019-04-18 – Present
Spatial Extent	Global
Coordinate System	51.6 N and S Latitude
Datum	Geographic (lat/lon)
Geographic Dimensions	4.2 km cross track by one-fourth of an ISS orbit along track
File Size	~5GB
File Format	HDF5

Table 2.4 GEDI Canopy Cover and Vertical Profile Metrics Data Global Footprint Level (L2B).

Characteristic	Description
Collection	GEDI
Short Name	GEDI02_B
DOI	10.5067/GEDI/GEDI02_B.001
Temporal Resolution	Varies
Temporal Extent	2019-04-18 – Present
Spatial Extent	Global
Coordinate System	51.6 N and S Latitude
Datum	Geographic (lat/lon)
Geographic Dimensions	4.2 km cross track by one-fourth of an ISS orbit along track
File Size	~1GB
File Format	HDF5

2.1.2 Landsat 8

The visible, NIR, and SWIR portions of the spectrum are measured by Landsat 8 OLI. Landsat 8 images have 15-meter panchromatic and 30-meter multi-spectral spatial resolutions along a 185 km (115 mi) swath.

Fig.2.2 Landsat 8.

Table 2.5 Specifications of Landsat 8.

Band	Wavelength
Band 1 (Coastal Aerosol)	0.43 - 0.45 µm
Band 2 (Blue)	0.450 - 0.51 µm
Band 3 (Green)	0.53 - 0.59 µm
Band 4 (Red)	0.64 - 0.67 µm
Band 5 (NIR) Band 6 (SWIR 1)	0.85 - 0.88 µm 1.57 - 1.65 µm
Band 7 (SWIR 2)	2.11 - 2.29 µm
Band 8 (Panchromatic)	0.50 - 0.68 µm
Band 9 (Cirrus)	a. - 1.38 µm

2.1.3 CHIRPS

CHIRPS is an acronym for Climate Hazards Group InfraRed Precipitation with Station data. It is a 30+ year quasi global rainfall dataset. CHIRPS uses 0.05° resolution satellite imagery and in situ station data in order to build gridded rainfall time series for the purpose of trend analysis and seasonal monitoring.

Table 2.6 Specifications of CHIRPS

Name	Rainfall - CHIRPS daily	Rainfall-CHIRPS monthly
Cell size - X (degrees)	0.05 (~5 km)	0.05 (~5 km)
Cell size - X (degrees)	0.05 (~5 km)	0.05 (~5 km)
Coordinate reference system	EPSG: 4326	EPSG: 4326
Temporal resolution	1 day	1 month
Temporal range	1981-present	1981-present
Update frequency	Daily	Monthly

2.1.4 SRTM (Shuttle Radar Topography Mission): Shuttle Radar Topography Mission obtained the DEMs from 56°S to 60°N. It generates the highest resolution digital topographic database of Earth.

Table 2.7 Specifications of SRTM.

Dataset	Specification
Projection	Geographic
Horizontal Datum	WGS84
Vertical Datum	EGM96
Vertical Units	meters
Spatial Resolution	1 arc-second for global coverage (~30 meters) 3 arc-seconds for global coverage (~90 meters)
Raster size	1 degree tiles
C-band wavelength	a. cm

2.2 Methodology

For this study initially the acquisition and processing of GEDI data was performed in order to derive canopy height metrics, PAI, PAVD and elevation. GEDI data is processed in python using the Jupyter Notebook, Google Colab and Pycharm. The overall methodology is shown in fig.2.4.

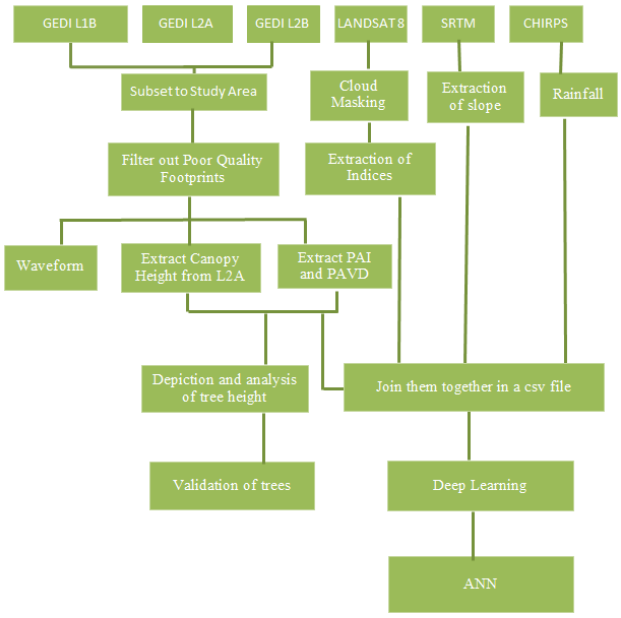


Fig.2.4 Methodology Flowchart.

2.2.1 GEDI Data Processing

The GEDI data which is used in this study is downloaded from the NASA data portal. Version 2 data, which was made available in April 2021 (https://lpdaac.usgs.gov/news/release-gedi-version-2-data-products/ (accessed on May 11, 2021)), was chosen since it has better geolocation accuracy. The footprints from two tracks of GEDI footprints acquired in March of 2020 respectively, are used for this study. Every GEDI track has 8 subtracts, 4 from full power beam and 4 from coverage beam. Coverage beams

are generally not recommended for dense vegetation as they cannot penetrate the dense vegetation. GEDI data is processed in Python. The version of Python which is used is 3.9. Data is processed in Jupyter notebook, Google Colab, Pycharm. The footprints were first sub-setted to the study area and then filtered to remove the poor-quality LiDAR shots using the available quality assurance flags supplied with the GEDI data (quality_flag = 1, degrade_flag = 0), and a sensitivity threshold of >0.95 was applied. These quality control criteria resulted in the deduction of the number of useable GEDI LiDAR footprints.

2.2.1.2 GEDI Shots
2.2.1.2.1 GEDI L2A

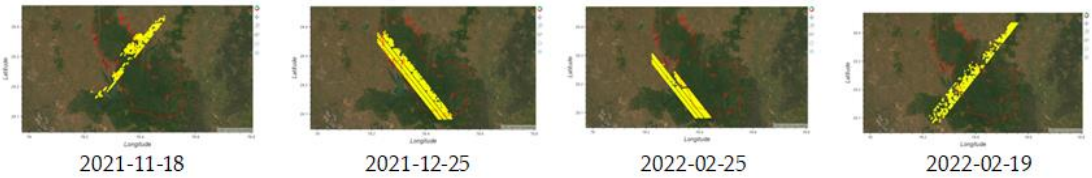


Fig.2.5 GEDI Shots with their date of acquisition

2.2.1.2.2 GEDI L2B

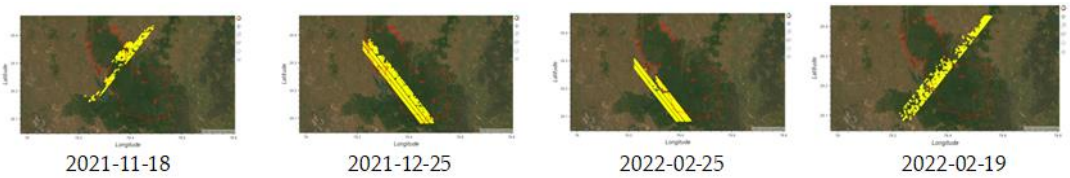


Fig.2.6 GEDI Shots with their date of acquisition.

2.2.1.2.3 Geolocated Waveform Data from the Level 1B data: L1B data processing determines the waveform over the forest stand.

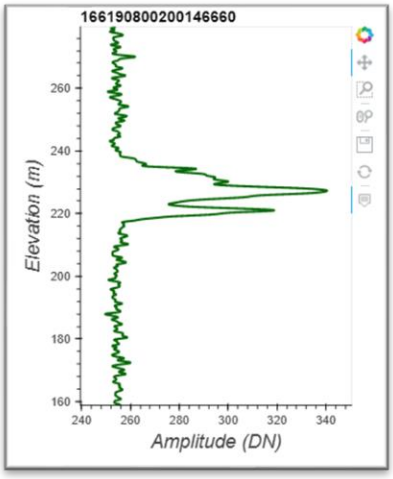


Fig.2.7 GEDI waveform over a forest stand.

2.2.1.2.4 Elevation and Canopy Height from the Level 2A data: L2A processing determines ground elevation and canopy heights at the footprint level using the geolcated received waveform (L1B product). The elevation value used from this product is the "elev_lowestmode" value. elev_lowestmode refers to the elevation of the centre of lowest mode relative to reference ellipsoid. The L2A product also includes the height above the

ground of each energy quantile in the received waveform and these are expressed as a height above the ground.

2.2.1.2.5 Total PAI and PAVD from the Level 2B data: The GEDI Level 2B data product contains the total PAI and footprint-level vertical profile metrics of PAVD, which are evaluated in this study. A L2A and L2B GEDI granule is processed in python for estimating the canopy height at a certain latitude and longitude. The height obtained from GEDI shot is then compared to ALS data of Goa in order to assess the accuracy of GEDI estimates.

2.2.1.2.6 LANDSAT8 Data Processing: Landsat 8 spectral reflectance data is processed in Google Earth Engine to derive the Maximum, Mean and Median values of vegetation Indices viz., NDVI, NDWI and GCVI and then sample those values for the footprints which were obtained from GEDI L2A and L2B data.

2.2.1.2.7 Shuttle radar Topography Mission: SRTM data is also processed in Google Earth Engine to obtain the slope values for the GEDI footprints.

2.2.1.2.8 CHIRPS Data Processing: CHIRPS data is processed in Google Earth Engine in order to obtain the sum and mean of rainfall for the GEDI footprints.

2.2.1.2.9 Machine Learning Modelling: Four machine learning models (Ordinary Least Square Regression, Multilayer Perceptron, Random Forest, Cat boost) were used to model the canopy height.

Linear Regression: Linear regression is one of the easiest and most popular Machine Learning algorithms. It is a statistical method that is used for predictive analysis. Linear regression makes predictions for continuous/real or numeric variables such as sales, salary, age, product price, etc. This algorithm shows a linear relationship between a dependent (y) and one or more independent (y) variables, hence called as linear regression. Since linear regression shows the linear relationship, which means it finds how the value of the dependent variable is changing according to the value of the independent variable.

2.2.1.2.10 Multi-Layer Perceptron: It is an Artificial Neural network which consists of 3 or more layers of perceptrons. The layers are:

1. A single input layer
2. 1 or more hidden layers
3. A single output layer of perceptrons

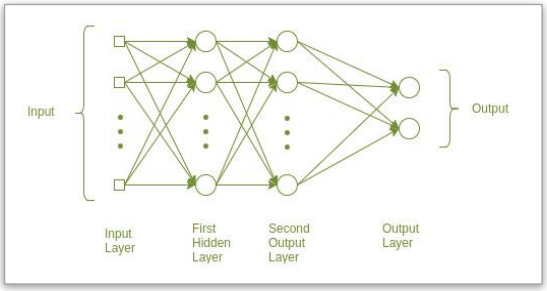


Fig.2.8 Multi-Layer Perceptron Model.

The data flows from input layer to hidden layers and then the output layer. The number of hidden layers can be increased as much as we want in order to increase the complexity of the model. MLP can be used in classification as well as regression problems. MLP gives the most accurate results for classification problems.

2.2.1.2.11 Random Forest: Random Forest is a machine learning algorithm that uses combined learning method for regression. Ensemble learning method is a technique which combines the predictions from multiple machine learning algorithms to prepare a more accurate model.

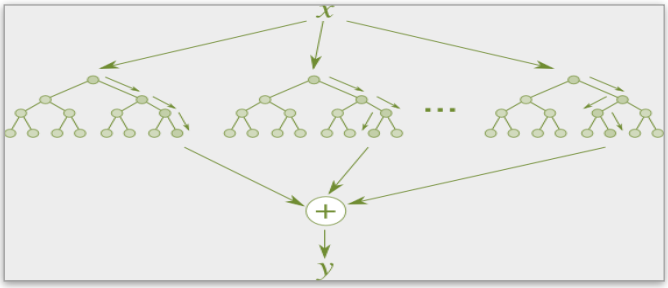


Fig.2.9 Random Forest Model.

2.2.1.2.12 Catboost: Catboost name comes from Category and Boosting. Catboost is built upon the principles of Decision trees and gradient boosting. The basic goal of boosting is to combine multiple weak models, or models that just slightly outperform chance, in order to produce a strong, competitive predictive model through greedy search. Since gradient boosting fits the decision trees sequentially, the fitted trees can learn from the mistake of former trees and reduce the errors. Catboost achieves the best results on the benchmark

Fig.2.10 Catboost Model.

The canopy height estimated from GEDI was used as dependent or target variable and Band values ranging from B2 to B7, Maximum, Minimum and Median values of NDVI, NDWI, GCVI, Sum and Mean of Rainfall, slope and Tandem-X were taken as independent or predictor variables.

Duplicate, null values and outliers were removed from Dataframe. Dataset was then normalized and then the outliers are removed from the normalized dataset. The filtered dataset was then splitted into 80 % training and 20 % testing. These training and testing datasets are then used to build and compile the model and assess the accuracy.

For MLP there were 20 hidden layers and 5 neurons for each layer were built after tuning of model hyperparameters. A learning rate of 0.01 was applied. The MLP regression model was conducted in Google Colab software by using the sklearn library.

For Random Forest Grid Search CV is applied in order to find out the best set of hyperparameters while running the model.

For Catboost the K Fold Cross Validation method is used to prevent the overfitting. Normalization was performed using Standard Scalar method for Catboost model. Bernoulli bootstrap type parameter is used. It corresponds Stochastic Gradient Boosting. In Bernoulli every example is sampled independently for choosing the current split with the probability indicated by the subsample parameter. R2 score is used as a measure of accuracy.

Table 2.8 Input variables used for training and validation.

Bands/Indexes	Wavelength/Equation	Reference
Band 1	0.43 - 0.45 µm	Landsat 8 Band 1
Band 2	0.450 - 0.51 µm	Landsat 8 Band 2
Band 3	0.53 - 0.59 µm	Landsat 8 Band 3
Band 4	0.64 - 0.67 µm	Landsat 8 Band 4
Band 5	0.85 - 0.88 µm	Landsat 8 Band 5
Band 6	1.57 - 1.65 µm	Landsat 8 Band 6
Band 7	2.11 - 2.29 µm	Landsat 8 Band 7
Normalized Difference Vegetation Index	(NIR - Red) / (NIR + Red)	
Normalized Difference Water Index	(Green – SWIR1) / (Green + SWIR1)	
Green Chlorophyll Vegetation Index	(NIR / Green) - 1	
Precipitation		
Slope		

3. Results

3.1 GEDI L2A

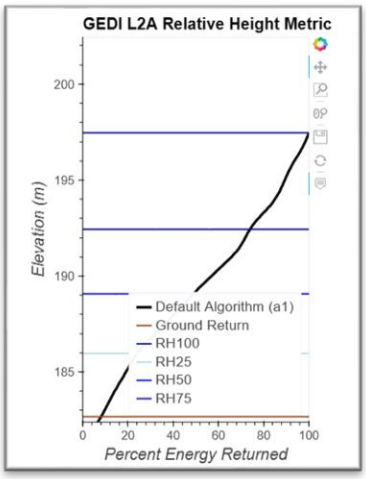


Fig. 3.1 Relative Height Metrics.

Fig.3.1 represents the plot of Relative Height Metrics. The height at which a specific quantile energy is reached in relation to the ground is shown by the relative height metrics. This figure shows the elevation at a certain percentage of energy returned. The black line in the graph represents the total amount of energy returned starting from the ground return to the top of canopy.

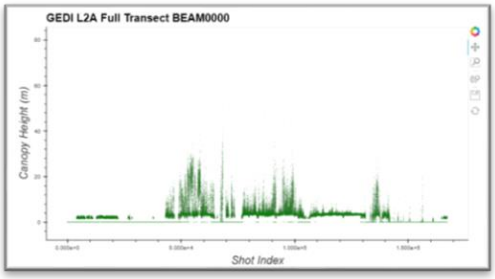


Fig.3.2 Plot of Canopy Height.

Fig.3.2 shows the canopy height for each shot of a beam at 100 percentage of energy returned. This figure is showing the height of every shot.

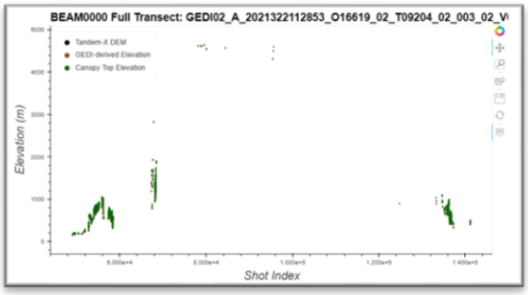


Fig.3.3 Combined plot for DEM, GEDI-derived elevation and Canopy Top Elevation for a beam.

The plot in Fig.3.3 represents all the elevations viz., Tandem-X, GEDI derived elevation and Canopy Top Elevation combined for every shot in a beam. In the graph the black, brown and green color represents the Tandem-X, GEDI derived Elevation and Canopy Top Elevations respectively.

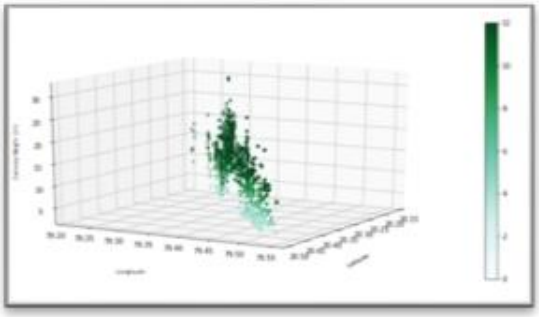


Fig. 3.4 3D plot of Canopy Height

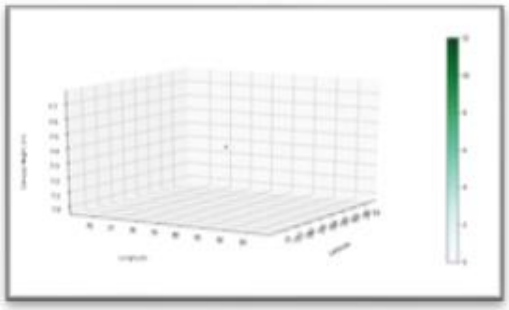


Fig.3.5 3D plot of canopy height of a shot

Fig.3.4, 3.5 represents canopy height in three dimensions. The height of a single shot which is present at a latitude and longitude of 20.37096614 and 79.386666339 is **7.36 meters.**

	Shot Number	Beam	Latitude	Longitude	Tandem-X DEM	Elevation (m)	Canopy Elevation (m)	Canopy Height (rh100)	RH 98	RH 25	Quality Flag	Degrade Flag	Sensitivity
49239	1.661900e+17	BEAM0000	20.181579	79.225858	165.688522	163.856384	173.214005	9.350000	8.34	-0.44	1.0	0.0	0.957679
49545	1.661900e+17	BEAM0000	20.305461	79.331020	188.979797	181.134979	197.155243	16.020000	15.38	3.70	1.0	30.0	0.961883
49557	1.661900e+17	BEAM0000	20.310289	79.335105	186.930435	185.730057	194.526230	8.790000	7.86	-1.60	1.0	30.0	0.957106
49583	1.661900e+17	BEAM0000	20.320743	79.343976	189.079498	182.982254	199.676270	16.690001	15.68	2.02	1.0	30.0	0.958706
49584	1.661900e+17	BEAM0000	20.321143	79.344319	189.079498	181.740356	199.594711	17.850000	16.84	2.80	1.0	30.0	0.951241
...	...	...	...	...	...	...	...	...	...	...	...	...	...
1220210	1.661911e+17	BEAM1011	20.420577	79.483996	150.275940	150.136536	158.741470	8.600000	3.44	-1.12	1.0	0.0	0.971969
1220211	1.661911e+17	BEAM1011	20.420981	79.484337	150.275940	150.429276	154.731750	4.300000	2.84	-1.19	1.0	0.0	0.976027
1220212	1.661911e+17	BEAM1011	20.421385	79.484678	149.668655	150.729279	162.065353	11.330000	9.12	-1.19	1.0	0.0	0.971395
1220213	1.661911e+17	BEAM1011	20.421788	79.485018	149.668655	148.803024	153.479614	4.670000	3.06	-1.12	1.0	0.0	0.970454
1220222	1.661911e+17	BEAM1011	20.425423	79.488086	150.799072	150.537155	154.839630	4.300000	2.99	-1.15	1.0	0.0	0.963309

928 rows × 13 columns

Fig.3.6 Dataframe obtained after processing the GEDI L2A data consisting of the canopy height.

Fig.3.6 consists of the dataframe which is obtained after the processing of L2A data. The dataframe consisted the Canopy Height in meters.

3.2 GEDI L2B

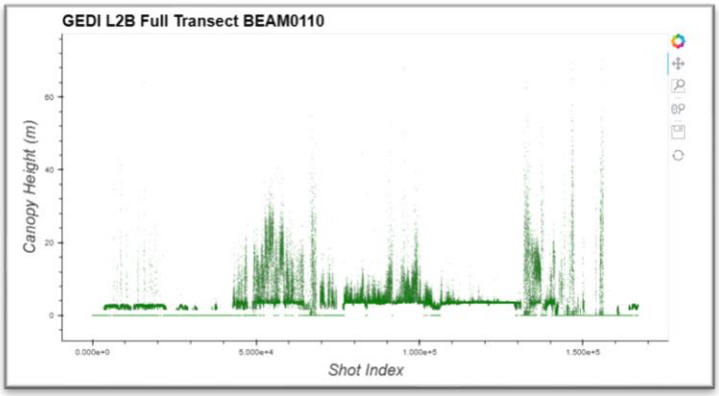
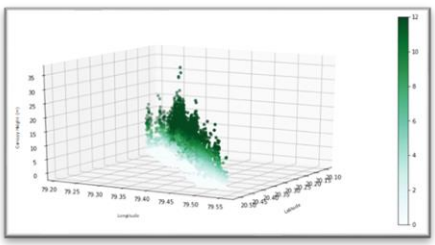
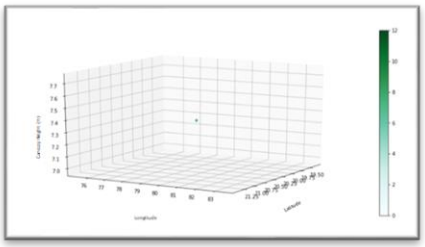


Fig.3.7 Plot of Canopy Height.

Fig.3.7 shows the canopy height for each shot of a beam at 100 percentage of energy returned. This figure is showing the height of every shot.



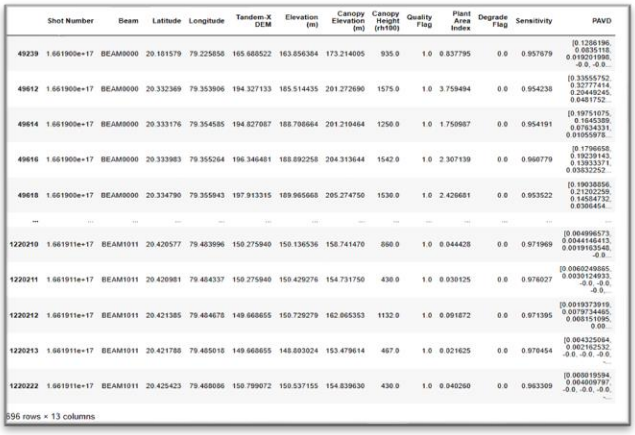


	Shot number	latitude	longitude	Height
0	1.66E+17	20.37095614	79.38666339	7.36

Fig.3.9 3D plot of Canopy Height.

Fig.3.10 3D plot of canopy height of a shot.

Fig.3.9, 3.10 represents canopy height in three dimensions. The height of a single shot which is present at a latitude and longitude of 20.37096614 and 79.386666339 is **7.36 meters.**



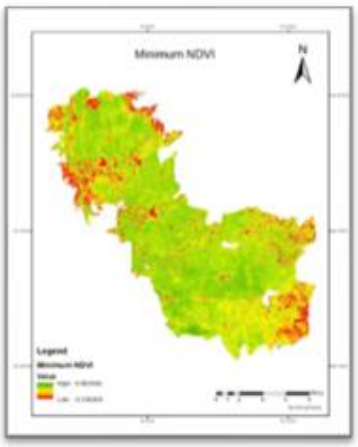
	Shot Number	Beam	Latitude	Longitude	Tandem-X DEM	Elevation (m)	Canopy Elevation (m)	Canopy Height (rh100)	Quality Flag	Plant Area Index	Degrade Flag	Sensitivity	PAVD
49239	1.661900e+17	BEAM0000	20.101579	79.225858	165.688522	163.856384	173.214005	935.0	1.0	0.837795	0.0	0.957679	[0.1286196, 0.0835118, 0.019201998, -0.0, -0.0...
49612	1.661900e+17	BEAM0000	20.332369	79.353906	194.327133	185.514435	201.272690	1575.0	1.0	3.759494	0.0	0.954238	[0.33555752, 0.32777414, 0.29449245, 0.0481752...
49614	1.661900e+17	BEAM0000	20.333176	79.354585	194.827087	188.708664	201.210464	1250.0	1.0	1.750987	0.0	0.954191	[0.19751075, 0.1645389, 0.07834331, 0.01055978...
49616	1.661900e+17	BEAM0000	20.333983	79.355264	196.346481	188.892258	204.313644	1542.0	1.0	2.307139	0.0	0.968779	[0.1796658, 0.19239143, 0.13933371, 0.03832252...
49618	1.661900e+17	BEAM0000	20.334790	79.355943	197.913315	189.965668	205.274750	1536.0	1.0	2.426681	0.0	0.953522	[0.19038856, 0.21202259, 0.14584732, 0.0306454...
...	...	...	...	...	...	...	...	...	...	...	...	...	...
1220210	1.661911e+17	BEAM1011	20.420577	79.403996	150.275940	150.136536	158.741470	860.0	1.0	0.044428	0.0	0.971969	[0.004996573, 0.0044146413, 0.0019163548, -0.0...
1220211	1.661911e+17	BEAM1011	20.420981	79.404337	150.275940	150.429276	154.731750	430.0	1.0	0.030125	0.0	0.976027	[0.0060249865, 0.0030124933, -0.0, -0.0, -0.0, 0...
1220212	1.661911e+17	BEAM1011	20.421385	79.404678	149.668655	150.729279	162.065353	1132.0	1.0	0.091872	0.0	0.971395	[0.0019373919, 0.0079734485, 0.008151095, 0.00...
1220213	1.661911e+17	BEAM1011	20.421788	79.405018	149.668655	148.803024	153.479614	467.0	1.0	0.021625	0.0	0.976454	[0.004325084, 0.002162532, -0.0, -0.0, -0.0, 0...
1220222	1.661911e+17	BEAM1011	20.425423	79.408086	150.799072	150.537155	154.839630	430.0	1.0	0.040260	0.0	0.963309	[0.000019594, 0.004009797, -0.0, -0.0, -0.0, -...

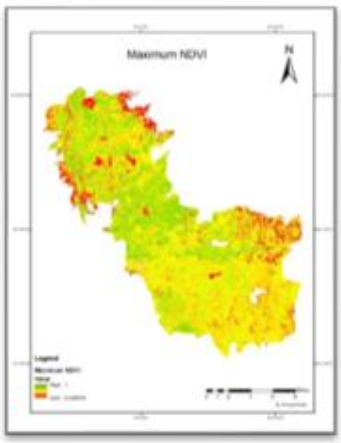
596 rows × 13 columns

Fig.3.11 Dataframe obtained after processing the GEDI L2B data consisting of the canopy height (2020).

Fig.3.11 consists of the dataframe which is obtained after the processing of L2B data. The dataframe consisted the Canopy Height, Canopy Elevation in meters, Plant Area Index, Plant Area Volume Density.

3.3 Modelling: For modelling canopy height, PAI and PAVD the values of maximum, minimum and median values of NDVI, NDWI, GCVI, Slope, Sum and Mean of rainfall were calculated.





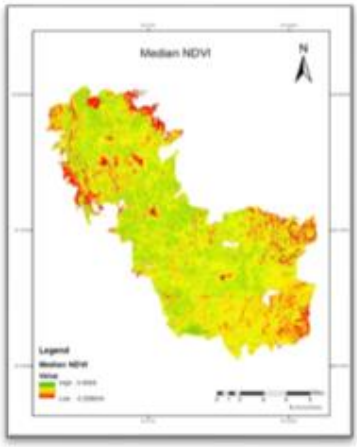


Fig.3.12 Maximum NDVI Map. **Fig.3.13** Minimum NDVI Map. **Fig.3.14** Median NDVI Map.

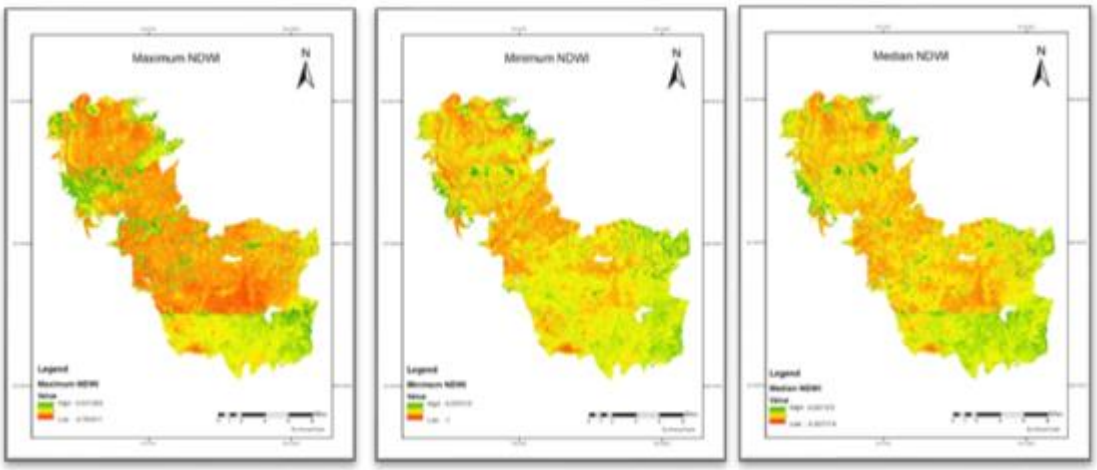


Fig.3.15 Maximum NDWI Map. **Fig.3.16** Minimum NDWI Map. **Fig.3.17** Median NDWI Map.

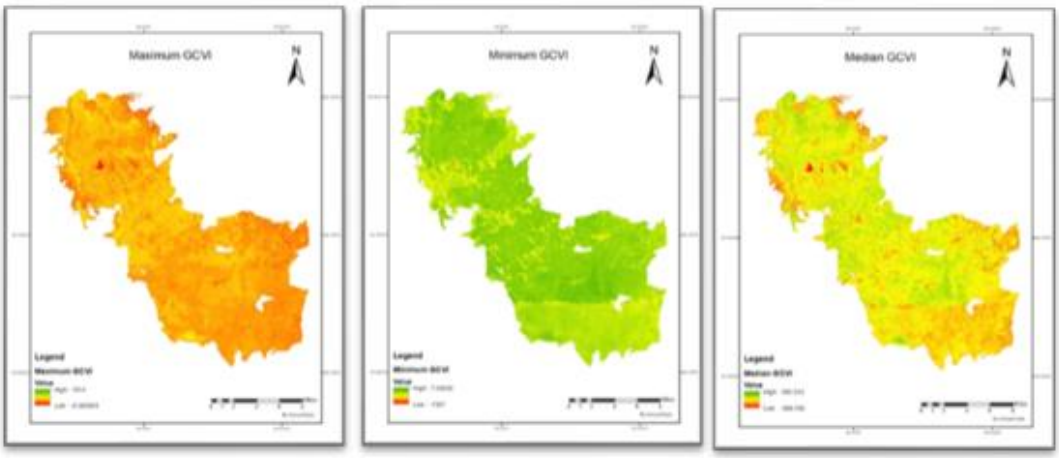


Fig.3.18 Maximum GCVI Map. **Fig.3.19** Minimum GCVI Map. **Fig.3.20** Median GCVI Map.

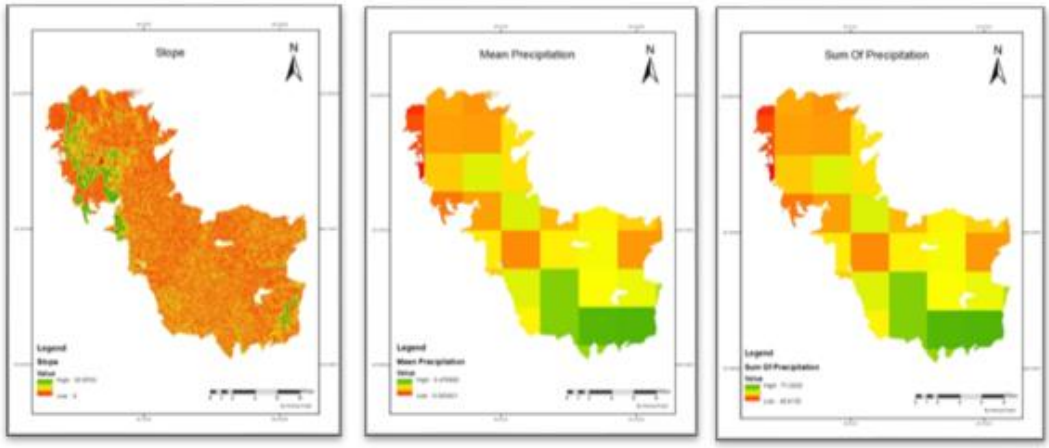


Fig.3.21 Slope. **Fig.3.22** Mean Precipitation. **Fig.3.23** Sum of Precipitation.

After the removal of duplicate, null values, outliers and normalization of the dataset, the GEDI footprints were used as the training and testing dataset for deep learning modelling. Canopy Height is used as the target variable and Maximum, Minimum and Median values of NDVI, NDWI and GCVI, Slope, Sum and Mean of rainfall were used as predictor variables.

A correlation matrix was plotted to identify the correlation between the dependent and independent variables.

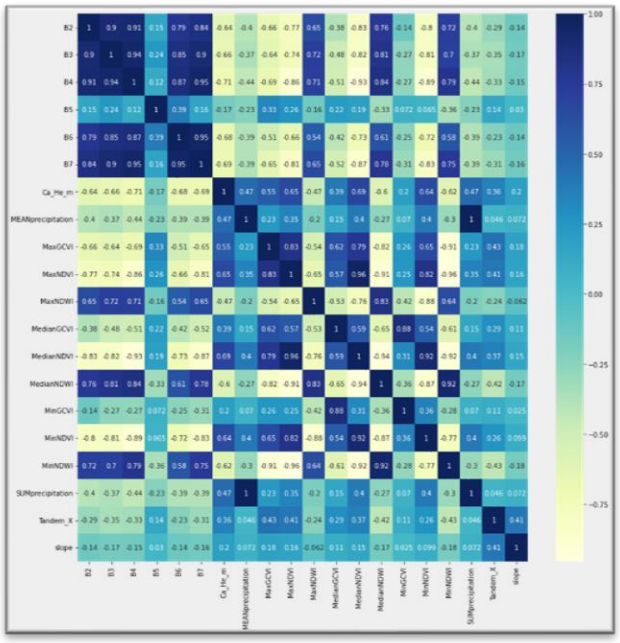


Fig.3.24 Correlation Matrix.

By looking at correlation matrix it was concluded that canopy height is correlated with most of the predictor variables except for that of Mean rainfall and minimum values of GCVI. Independent validation was performed using Linear Regression, Multi-Layer Perceptron Regressor, Catboost Regressor and Random Forest Regressor.

Model	Target Variable	Predictor Variables	R^2 Score
Linear Regression	Canopy Height	Max NDVI, Min NDVI,	0.65
	PAI	Median NDVI, Max	0.47
	PAVD	NDWI, Min NDWI,	0.35
		Median NDWI, Max GCVI, Min GCVI, Median GCVI, Mean Precipitation, Sum Precipitation, Slope	
Catboost	Canopy Height	Max NDVI, Min NDVI,	0.82
	PAI	Median NDVI, Max	0.71
	PAVD	NDWI, Min NDWI,	0.49
		Median NDWI, Max GCVI, Min GCVI, Median GCVI, Mean Precipitation, Sum Precipitation, Slope	
Random Forest	Canopy Height	Max NDVI, Min NDVI,	0.81
	PAI	Median NDVI, Max	0.69
	PAVD	NDWI, Min NDWI,	0.46
		Median NDWI, Max GCVI, Min GCVI, Median GCVI, Mean Precipitation, Sum Precipitation, Slope	
Multi Layer Perceptron	Canopy Height	Max NDVI, Min NDVI,	0.76
	PAI	Median NDVI, Max	0.58
	PAVD	NDWI, Min NDWI,	0.47
		Median NDWI, Max GCVI, Min GCVI, Median GCVI, Mean Precipitation, Sum Precipitation, Slope	

Table 3.1 Interpretation of R2 Score.

R^2 Score	Interpretation
0.75 - 1	Significant amount of variance explained
0.5 - 0.75	Good amount of variance explained
0.25 - 0.5	Small amount of variance explained
0 - 0.25	Little to no variance explained

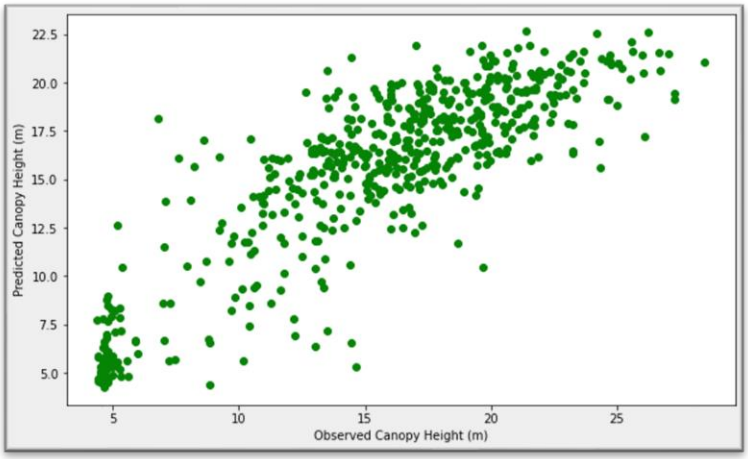


Fig.3.25 Scatterplot depicting the Observed and Predicted Canopy Height.

Fig.3.25 represents the true and predicted values of canopy height in meters. The graph is showing a positive correlation between the true values and predicted values. Fig.3.26, 3.27 and 3.28 represents the predicted maps of Canopy Height, PAI and PAVD.

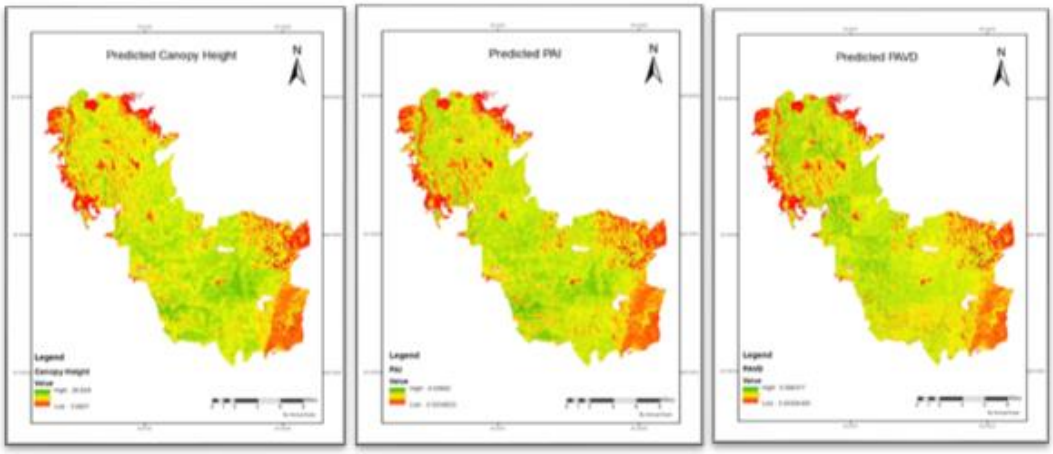


Fig.3.26 Predicted Map of Canopy Height. **Fig.3.27** Predicted Map Of PAI. **Fig.3.28** Predicted Map Of PAVD.

3.4 Discussion

In this study, GEDI LiDAR data is used to extract the canopy height, PAI and PAVD of forest stands for Tadoba Tiger Reserve. vTo my knowledge, this study demonstrated the first attempt to apply the regression models to model the forest canopy height over the region of

Tadoba Tiger Reserve. The GEDI canopy height product provided a very important input dataset for the deep-learning models for large-scale and spatially contiguous mapping of canopy height, which is important for the models that require for big training data to assist the regression model construction. Out of 4 models which were performed Catboost and Random Forest performed satisfactorily. The Catboost and Random Forest Model obtained the satisfactory accuracy in the prediction of canopy height, PAI and PAVD which implies that the true values are close to the predicted values. It suggests the high potential of Catboost regression and Random Forest regression in the modelling of canopy height, PAI and PAVD. For the predictor variables Landsat 8 band spectral reflectance values, vegetation indices viz., NDVI, NDWI, GCVI, Slope, Precipitaion and Tandem-X. Future efforts should be made to add more environmental factors to model the canopy height, PAI and PAVD.

3.5 Conclusion

This study estimated the canopy height, PAI and PAVD for the region of Tadoba Tiger Reserve. Several regression models were also prepared to model the canopy height namely Linear Regression, Multi-Layer Perceptron, Random Forest and Catboost Regression. Out of all these models Random Forest and Catboost regression performed satisfactorily for canopy height and catboost for PAI and PAVD. Accurate prediction of forest carbon sequestration potential requires a comprehensive understanding of tree growth relationships. The forest carbon sequestration potential is not only influenced by forest growth but also by climatic factors, topographic factors, land use change, management measures, etc Future efforts should be made to add more environmental factors. Here the estimation of forest growth can be obtained by this project and by applying the models developed under this project and other factors collected as mentioned above, will help to calculate the carbon sequestration of the forest.

Acknowledgements

My foremost appreciation to my mentor and supervisor, Col K Joshil Raj for his constant supervision, guidance, and encouragement throughout this project. I sincerely thank him for his generous involvement in advising, providing constructive feedback and full support. Lastly, I would also like to thank all my friends and family members for encouraging and supporting me whenever I needed them.

References

1. Dhargay S, Lyell CS, Brown TP, Inbar A, Sheridan GJ, Lane PNJ. Performance of GEDI Space-Borne LiDAR for Quantifying Structural Variation in the Temperate Forests of South-Eastern Australia. Remote Sens. 2022 Jul 28;14(15):3615.
2. Dorado-Roda I, Pascual A, Godinho S, Silva C, Botequim B, Rodríguez-Gonzálvez P, et al. Assessing the Accuracy of GEDI Data for Canopy Height and Aboveground Biomass Estimates in Mediterranean Forests. Remote Sens. 2021 Jun 10;13(12):2279.
3. Guerra-Hernández J, Pascual A. Using GEDI lidar data and airborne laser scanning to assess height growth dynamics in fast-growing species: a showcase in Spain. For Ecosyst. 2021 Dec;8(1):14.
4. Ghosh SM, Behera MD. Forest canopy height estimation using satellite laser altimetry: a case study in the Western Ghats, India. Appl Geomat. 2017 Sep;9(3):159–66.
5. C. KUNHIKANNAN, N. RAMA RAO. PHENOLOGICAL STUDIES OF TREES OF TADOBA NATIONAL PARK, CHANDRAPUR, MAHARASHTRA, INDIA. Res Gate.

6. C. Kunhikannan, N. Rama Rao, S.S. Bisen. VEGETATION ECOLOGY OF TADOBA NATIONAL PARK, CHANDRAPUR, MAHARASHTRA. Res Gate. :27.
7. Lee WJ, Lee CW. Forest Canopy Height Estimation Using Multiplatform Remote Sensing Dataset. J Sens. 2018;2018:1–9.
8. Marselis SM, Keil P, Chase JM, Dubayah R. The use of GEDI canopy structure for explaining variation in tree species richness in natural forests. Environ Res Lett. 2022 Apr 1;17(4):045003.
9. Michelle Hofton PP. Mapping global forest canopy height through integration of GEDI and Landsat data. Elsevier. 2020 Nov 7;11.
10. Poonam Tripathi PT. Plant height profiling in western India using LiDAR data. Res Gate. 2013 Oct 10;105:9.
11. Paliwal A, Mathur VB. Spatial pattern analysis for quantification of landscape structure of Tadoba-Andhari Tiger Reserve, Central India. J For Res. 2014 Mar;25(1):185–92.

Citation

Fatehpur, S.S. (2024). Modelling Canopy Structure of Forest using Big Geospatial Data and Deep Learning. In: Dandabathula, G., Bera, A.K., Rao, S.S., Srivastav, S.K. (Eds.), Proceedings of the 43rd INCA International Conference, Jodhpur, 06–08 November 2023, pp. 399–417, ISBN 978-93-341-2277-0.