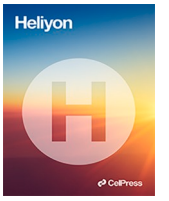
Heliyon 9 (2023) e21485

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon



Above ground tree biomass modeling using machine learning algorithms in western Terai Sal Forest of Nepal

Bikram Singh [a,*], Amit Kumar Verma [a], Kasip Tiwari [b], Rajeev Joshi [c]

[a] Forest Research Institute (Deemed to be) University, Dehradun-248195, Uttarakhand, India
[b] TTG Forestry LLC, Atlanta, GA, USA
[c] College of Natural Resource Management, Faculty of Forestry, Agriculture and Forestry University, Katari, 56310, Udayapur, Nepal

ARTICLE INFO

Keywords:
Machine learning algorithms
Above ground tree biomass
Sentinel-2A
Sustainable

ABSTRACT

The monitoring of forest biomass is a crucial biophysical parameter in forest ecosystems, as it provides valuable information for managing forests sustainably and tracking carbon circulation statistics. To achieve sustainable forest management, it is essential to monitor and study forest resources, particularly biomass. This study aimed to model above ground tree biomass (AGTB) using Machine Learning Algorithms (MLAs) in the western terai Sal forest of Nepal. AGTB was calculated using a systematic inventory sample plot, while spectral and textural variables were processed and masked for the study area using Sentinel-2A satellite imagery. Three MLAs namely support vector machine (SVM), random forest (RF), and stochastic gradient boosting (SGB), were employed for modeling with eight categorized variable datasets. Among the MLAs, the RF algorithm with a combination of gray-level co-occurrence matrix (GLCM) and raw bands (RB) dataset variable demonstrated the best performance, with a low RMSE value of 78.81 t ha^{-1} in the test data. However, the AGTB range from this model ranged from 118.34 to 425.97 t ha^{-1}. The study found that traditional indices, raw bands, and GLCM texture from near-infrared were important variables for AGTB. Nevertheless, the RF algorithm and the dataset combination of GLCM plus raw bands (RB) exhibited excellent performance in all model runs. Thus, this pioneering study on comparative MLAs-based AGTB assessment with multiple datasets variables can provide valuable insights for new researchers and the development of novel approaches for biomass/carbon estimation techniques in Nepal.

1. Introduction

Forest measurement is essential for national and international forest resource assessment report. Thus the report could support for biodiversity monitoring and management prospects at regional/local level working plan cum decision making process [1]. Growing stock estimation, carbon stock, above ground biomass of forest ecosystem are main variables which are directly related to sustainable forest management strategies and climate change mitigation policies development [2,3]. Likewise forest biomass quantification is crucial for economic valuation of forest and its supply in form of fuel wood, timber and so on [4,5]. Also carbon mapping in spatio-temporal way is important for its sustainability as not only for how much carbon reduced [6]. Forest ecosystem comprise 30 %

* Corresponding author.
E-mail addresses: bikramsingh2051@gmail.com (B. Singh), amitvermafri@gmail.com (A.K. Verma), kasiptiwari@gmail.com (K. Tiwari), joshi.rajeev20@gmail.com (R. Joshi).

https://doi.org/10.1016/j.heliyon.2023.e21485
Received 8 July 2023; Received in revised form 6 October 2023; Accepted 22 October 2023
Available online 2 November 2023

of global land surface, plays vital role in carbon circulation which help in mitigation of human induced emission effects [7]. However World Forest growing stock is declining, about 3 billion cubic meter has lost from 1990 to till [8]. Forest growing stock volume (FGSV) is highly correlated with above ground biomass [9,10]. Studies evidenced that Above ground biomass is significant predictive variables of carbon storage and half of total biomass would be carbon storage as per thumb rule [11,12].

Forest biomass estimation is used to be tough but accurate through field based destructive solely methods which cost much manpower, finance and time for local level assessment [10,13,14]. Nowadays, remote sensing and geographical information system (GIS) integration with plot based data on geo-statistical analysis enables the modeling of Above Ground Tree Biomass(AGTB) with reliable results in worldwide [15–17]. Widely utilized remote sensing satellite data from sources such as Landsat 8, MODIS, QuickBird, SPOT, as well as Light Detection and Ranging (LiDAR) and Sentinel 1 or 2 images, have proven instrumental in achieving highly accurate AGTB estimations across diverse regions [18–21]. However, the application of spatial-level forest AGTB estimation is not yet widespread in developing countries due to a dearth of innovative estimation methodologies. Spectral bands of satellite images based rationing calculated variable layers like Normalized Different Vegetation Index (NDVI), Ratio vegetation index (RDVI), Enhanced vegetation index (EVI) etc. have performed better relationship with biomass and their modeling based output improves the estimation accuracy [20,22]. More recently, researchers have begun harnessing various novel machine learning algorithms (MLA) to further elevate the precision of AGB estimation. These algorithms combine field survey biomass data with multiple derivatives of remote sensing data, ushering in a new era of improved estimation accuracy [23]. These MLAs have the classification and regression tree (CART) learner and its extensions using bagging or boosting techniques, followed by the development of random forests (RFs) and the gradient-boosted regression tree (GBRT) algorithms [23–25]. It is evidenced that different non-parametric machine learning algorithms (MLAs) like random forest (RF), support vector machine (SVM), maximum entropy (MaxEnt) algorithm, k-nearest neighbors (kNN) method, artificial neural network (ANN) and cubist (CB) algorithms have performed better than traditional linear regression models for biomass estimation in recent research [21,26,27]. Researches were carried out in comparison of performance output of these algorithms in estimation forest AGB i.e. [28] compared several techniques like stepwise linear regression, kNN, SVR, RF, and stochastic gradient boosting (SGB) methods for estimating forest AGB from Landsat series images in China, and concluded that the RF revealed the best performance among all. Similarly recent done by using Random Forest (RF), Gradient Boosted Model (SGB), and Extreme Gradient Boosting (XGB) for AGB estimation of Mangrove site forest in India resulted that RF gave good results among all however ensemble prediction exhibited more improvement in accuracy for estimation [12]. Though medium resolution satellites

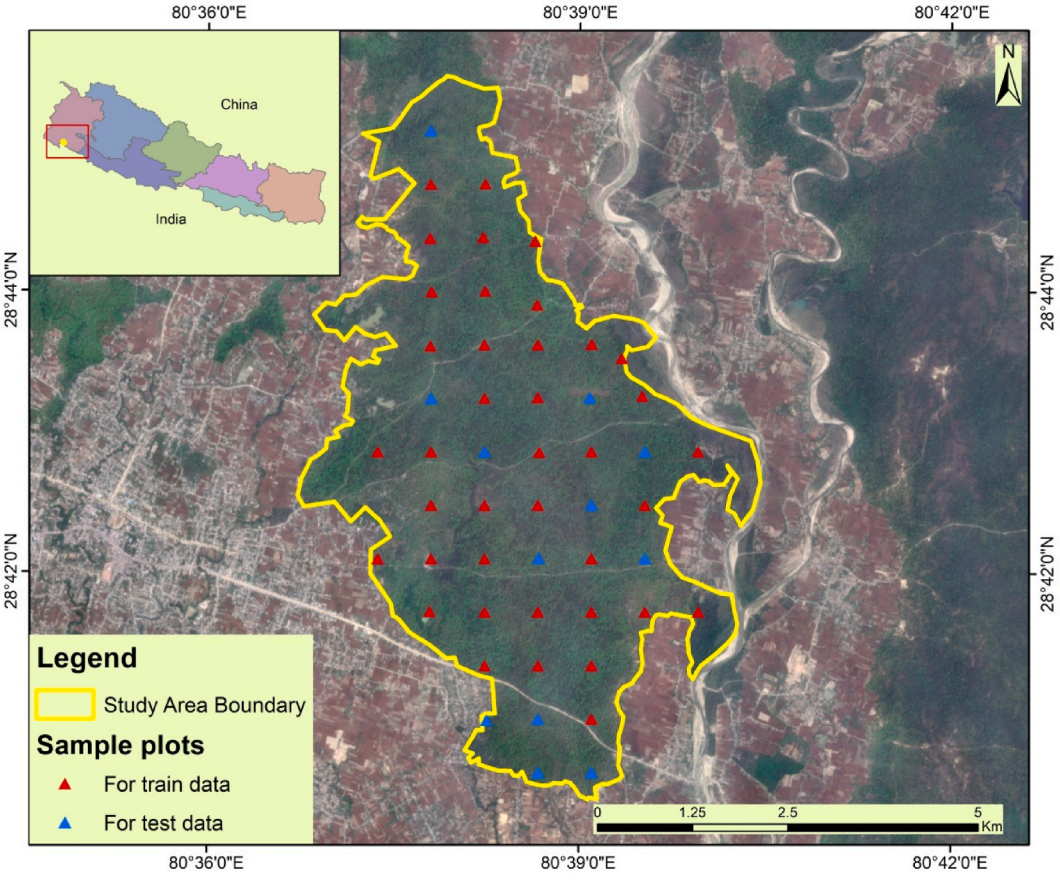


Fig. 1. Study area map.

images are available free of cost i.e. Sentinel 1 or 2 and Landsat 8 OLI/TIRS, the studies are very limited in Nepal to estimate AGB.

Sound methods to assess the actual forest resources are one of the important aspects of sustainable forest management principles. Above ground tree biomass is the living or dead standing tree weight expressed as mass per unit area which is important variable to assess carbon stock. Nepal committed at the national and international level to reduce emissions by addressing drivers of deforestation and forest degradation, and therefore it is actively engaged in the REDD readiness and implementation process. Nepal National REDD + Strategy had objective to improve National Forest Monitoring System with robust measurement, monitoring, reporting and verification mechanisms [29]. Department of forest and soil conservation has been followed point based field data to project overall country forest growing stock [30]. Furthermore, the community forest inventory guideline is also based on plot based for growing stock estimation. However, there are more than 22000 community forests. In order to develop new method/guideline which is based on geospatial machine learning modeling to project forest tree growing stock of spatial boundary, ample scientific research should carry in different part of country.

In Nepal, research utilizing spatial modeling approaches is scarce. Notably, a study in the Chitwan district of central Nepal employed object-based image analysis and supervised nearest neighbor classification methods, drawing on WorldView-2 satellite and LiDAR data, to quantify and map carbon stocks for predominant tree species [31]. Similarly, another investigation focused on a hilly watershed in the Gorkha district of Nepal, leveraging high-resolution GeoEye data to estimate carbon stocks [32]. However, these studies are more applicable to small areas and come with substantial costs. A recent study by Ref. [26] utilized medium-resolution Sentinel-2 data in the buffer zone community forest of Parsa National Park, Nepal, employing the random forest model algorithm for above-ground biomass assessment. Despite these efforts, there remains a dearth of multiple machine learning ensemble models for predicting forest tree growing stock and above-ground tree biomass in Nepal. Thus, the introduction of multiple machine learning model applications in forest resources promises a groundbreaking approach to spatially mapping biomass stock. Consequently, the outcomes of such endeavors could serve as essential references for developing technology-based guidelines for forest resource assessment across various management models in Nepal. The primary objective of this study is to estimate spatial Above Ground Tree Biomass (AGTB) by employing Machine Learning Algorithms (MLAs), encompassing an assessment of MLA performance and environmental variables in estimating AGTB in the Western Terai Sal Forest of Nepal.

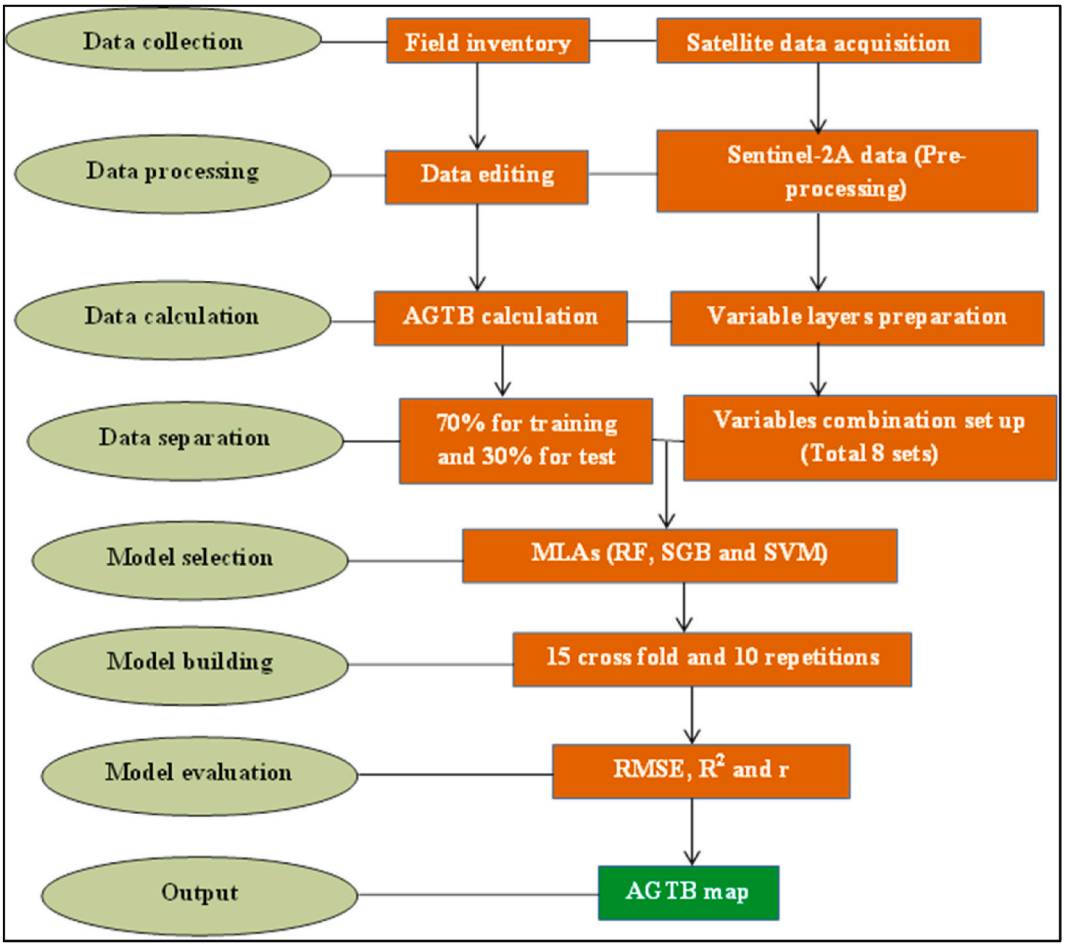


Fig. 2. Work flow of methodology.

2. Materials and method

2.1. Study area

The study was carried out in the Terai Sal Forest patch of Far Western, Nepal that covers an area of 2786 ha (Fig. 1). The forest area is located in 28°43′19.56"N latitude and 80°38′10.39"E longitude with elevation approximate 180 m from mean sea level. The climate of study area consists of hot monsoon tropical climate with winter cold and summer hot. The forest patch falls under tropical Sal Forest of Nepal forest classification and managed by community forest users' group. Sal (*Shorea robusta*) is dominant species in study area with some associates species like *Terminalia tomentosa, Adina cordifolia, Syzigium cuminii* etc. The study area covers Jali community forest, Dhanchaudi community forest, Devhariya community forest, Siddanath community forest, Beli community forest, Patela community forest and Manehara community forest.

2.2. Sampling design and data collection

The overall methodology process completed in five steps as sampling design and data, plot-based biomass calculation, satellite data download, variable layer preparation and stacking, model run and model evaluation (Fig. 2). The sample design is based on Community forest inventory guideline 2005 [33]. According to these guidelines, the sampling intensity for any homogenous community based managed forest would be 0.1 %. Hence the number of sample plot required for study area calculated by following formula.

$$n = \frac{A \times I}{a \times 100\%} \tag{1}$$

where, A = Study area (Ha), I = Sampling Intensity, a = Sample plot area and n = Sample plot numbers required.

*Sample plot area would be 500 m^2 for tree according to CF inventory guideline 2005.

The total 51 numbers of sample plots were measured for this study area. Furthermore, required numbers of sample plot was layout on systematic sampling basis throughout study area. Hence systematic distance between one sample plot to another can be calculated by following formula.

$$d = \sqrt{\frac{A}{n+1}} \tag{2}$$

where, A = Study area (m^2), n = Sample plots number required and d = Distance between two sample plots.

Once distance is calculated, sample plots have been laid out in study spatially by using Fishnet tool in Arc map 10.5 versions (Fig. 3.1). The plots have been navigated by Global Positioning System (GPS) tools. At field, sample plots were measured in circular shape. The tree diameter at breast height (dbh) 1.3 m (>10 cm dbh only) and height were measured within each plot. The diameter at breast height was measured by diameter tape and height by Range finder.

2.3. Plot based above ground tree biomass (AGTB) calculation

The above ground tree biomass is calculated at plot level by improved allometric equation based on diameter and height developed by Ref. [34] at plot level as:

$$AGTB = 0.0673 \times \left(\rho \times D^2 \times H\right)^{0.976} \tag{3}$$

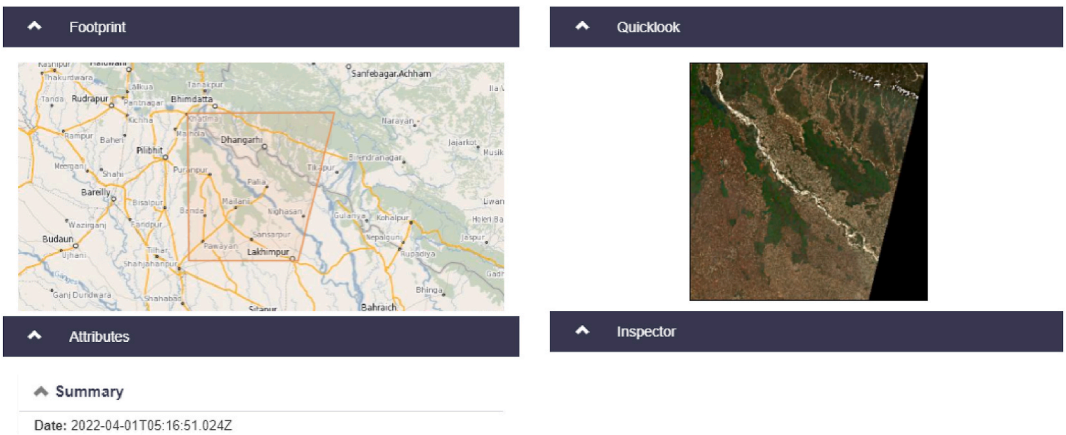


Fig. 3. Sentinel-2A imagery scene.

Where, AGTB = Above Ground Tree Biomass (Kg), ρ = Wood density$\left(\frac{gm}{cm^3}\right)$, H = Height of tree(m), D = Diameter at breast height(cm).

2.4. Satellite data processing and variable layers

For this study, we used the Sentinel-2A MSI Level-1C (L1C) products collected from the Copernicus Open Access Hub (https://scihub.copernicus.eu/, accessed in 2022 April 01) (Fig. 3). The satellite is rich in spectral information covering total 13 bands i.e., visible, red-edge, near-infrared and shortwave infrared wavelengths with varied spatial resolution. In this study, total 10 bands (Table 1) were used out of 13 in which 1, 9 and 10 bands were excluded. In order to resemble in spatial resolution, all bands were resampled to 10 m resolution. To run the model, raw bands (RB), traditional spectral indices(TSI), red edge spectral indices (RESI) and gray level co-occurrence matrix (GLCM). TSI, and RESI layer were prepared by using the mathematical formula through raster calculation (Table 1). And for GLCM layer, optimum window size 7 x 7 was selected to preparation using the near infrared band.

2.5. Machine learning algorithm for AGTB modeling

Machine learning (ML) is the scientific study of algorithms and statistical models that computer systems use to perform a specific

Table 1
Variable description used in AGTB prediction modeling.

Variable layer	Formula	Reference
Raw bands (RB)		[35]
Blue	B2	
Green	B3	
Red	B4	
Red edge 1	B5	
Red edge 2	B6	
Red edge 3	B7	
NIR	B8	
NIR narrow	B8A	
SWIR 1	B11	
SWIR 2	B12	
Traditional spectral indices (TSI)		
NDVI	$(\rho NIR - \rho Red)/(\rho NIR + \rho Red)$	[36]
CLG	$\rho NIR/\rho Green - 1$	[37]
VARIg	$(\rho Green - \rho Red)/(\rho Green + \rho Red - \rho Blue)$	[38]
NDII	$(\rho NIR - \rho SW\ IR1)/(\rho NIR + \rho SW\ IR1$	[39]
MSAVI	$[2\,\rho NIR + 1 - \sqrt{(2\,\rho NIR+1)^2 - 8(\rho NIR - \rho Red)}]/2$	[40]
SR	$\rho NIR/\rho Red$	[41]
DVI	$\rho NIR - \rho Red$	[42]
ARVI	$(\rho NIR - (2\rho Red - \rho Blue))/(\rho NIR + (2\rho Red - \rho Blue))$	[43]
Red edge spectral indices (RESI)		
MCARI	$[(\rho RE1 - \rho Red) - 0.2(\rho RE1 - \rho Green)](\rho RE1/\rho Red)$	[44]
NDVIre1	$(\rho NIR - \rho RE1)/(\rho NIR + \rho RE1)$	[45]
NDVIre2	$(\rho NIR - \rho RE2)/(\rho NIR + \rho RE2)$	[46]
SRre	$\rho NIR/\rho RE1$	[47]
NDVIre3	$(\rho NIR - \rho RE3)/(\rho NIR + \rho RE3)$	[46]
CLre	$\rho RE3/\rho RE1 - 1$	[48]
NDre2	$(\rho RE3 - \rho RE1)/(\rho RE3 + \rho RE1)$	[49]
Gray-level co-occurrence matrix (GLCM)		
Mean	$\sum_{i,j=0}^{N-1} iP_{i,j}$	[50]
Variance	$\sum_{i,j=0}^{N-1} iP_{i,j}(1-\mu_i)$	
Homogeneity	$\sum_{i,j=0}^{N-1} iP_{i,j}/(1+(i-j)^2)$	
Contrast	$\sum_{i,j=0}^{N-1} iP_{i,j}(i-j)$	
Dissimilarity	$\sum_{i,j=0}^{N-1} iP_{i,j}\lvert i-j\rvert$	
Entropy	$\sum_{i,j=0}^{N-1} P_{i,j} \ln P_{i,j}$	
Second moment	$\sum_{i,j=0}^{N-1} iP_{i,j}^2$	
Correlation	$\sum_{i,j=0}^{N-1} (i(\sum_{i,j=0}^{N-1} ijP_{i,j} - \mu_i\mu_i)/\sigma_i^2\sigma_j^2)$	

*ARVI: Atmospherically Resistant Vegetation Index; CLg: Green Chlorophyll Index; DVI: Difference Vegetation Index; MSAVI: Modified Soil Adjusted Vegetation Index; NDII: Normalized Difference Infrared Index; NDVI: Normalized Difference Vegetation Index; SR: Simple Ratio; VARIg: Visible Atmospherically Resistant Index green; CLre: Chlorophyll Index red-edge; MCARI: Modified Chlorophyll Absorption Ratio Index; NDVIre1: Normalized Difference Vegetation Index Red-edge 1; NDVIre2: Normalized Difference Vegetation Index Red-edge 2; NDVIre3: Normalized Difference Vegetation Index Red-edge 3; NDre2: Normalized Difference Red-edge 2; SRre: Simple Ratio Red-edge.

task without being explicitly programmed. Based on Machine learning, many algorithms were developed for geospatial predictive modeling for environmental research [51]. CARET (Classification and Regression Training) package based machine learning algorithms (Random Forest, Support Vector Machine and stochastic gradient boosting.) was selected for AGTB estimation in R statistical software [12]. Furthermore the CARET package had calculated environment variable importance for each model run.

2.6. Model evaluation

A total of 51 sampling plots based AGTB value were randomly partition into two subsets: 70 % was used for the training data set for model learning and the remaining 30 % as the testing the model built. Cross-validation was applied to assess the generalization capability of the model when there are an inadequate number of sample points. In addition, we selected three criteria for evaluating the RF model effectiveness, namely, the RMSE (Equation (4)), the coefficient of determination (R^2, Equation (5)), and Pearson's correlation coefficient of the root mean square error (r, Equation (6)). These were calculated using fifteen-fold cross-validation. Generally, the higher R^2 and r and the lower RMSE represents, the model fits better.

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{(\widehat{y}_i - y_i)^2}{n}} \tag{4}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \widehat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \overline{y}_i)^2} \tag{5}$$

$$r = n\left(\frac{\sum y_i \widehat{y}_i) - (\sum y_i)(\sum \widehat{y}_i)}{\sqrt{\left[n(\sum y_i{}^2) - (\sum y_i)^2\right]\left[(n\sum \widehat{y}_i{}^2) - (\sum \widehat{y}_i)^2\right]}}\right. \tag{6}$$

Where, y_i and $\widehat{y}_i$ represents the field measured and predicted AGTB values in the ith sample respectively, $\overline{y}_i$ is the measured AGTB averaged over all samples, and n represents the size of the samples in different data set.

3. Results

Total eight set of variable combinations as per mentioned in methodology section were calibrated for each model i. e., SVM, RF and SGB. In order to optimize the values of hyper parameter according to the minimum RMSE, the model run for every replication was based on 15 cross fold and 10 repetitions. The results obtained and model performance is explained separately as follows.

3.1. Support vector machine

The model performance for each variable combination in model calibration, training data and test data is illustrated (Table 2). For SVM algorithm, eight categories of variable combination reflect varied performance for model calibration, train data and test data. Model performance in calibration is found less RMSE value (118.11 tha^{-1}) in case of all variables stacked (RB, TSI, RESI, GLCM) for AGTB estimation with R^2 (0.77). Thus, the AGTB range given by this model based on all variables is 159.72–365.96 t ha^{-1} (Fig. 5(h)). Similarly, the algorithms had responded differently for each variable combination for train data on field measured data versus predicted data and thus all variable based model has showed good among all with RMSE (112.13 t ha^{-1}), R^2 (0.55) and r (0.74) (Table 2). And test data partition before model build was compared with predicted biomass data by model run for each variable combination, we found best statistical response on traditional spectral indices-based model with RMSE (83.36 t ha^{-1}), R^2 (0.26) and r (0.51) (Table 2) which shows the AGTB of range 171.21–356.09 t ha^{-1}(Fig. 5(b)). The scatter plot is clearly shown (Fig. 4) for observed and predicted AGTB for eight combinations for test data. The predicted spatial AGTB map for each variable combination is presented in Fig. 5.

Table 2
Model performances on train and test data for SVM algorithm.

S.N.	Variables	Model performance		Training data			Test data		
		R^2	RMSE	R^2	RMSE	R	R^2	RMSE	R
1	Raw bands	0.74	122.23	0.58	127.98	0.70	0.13	90.31	0.36
2	Traditional spectral indices	0.74	121.41	0.38	127.58	0.62	0.26	83.36	0.51
3	Red edge spectral indices	0.75	120.55	0.24	127	0.49	0.02	105.08	−0.13
4	Gray-level co-occurrence matrix (GLCM)	0.75	124.52	0.63	126.32	0.70	0.05	93.19	0.22
5	Traditional spectral indices and raw bands	0.72	119.78	0.42	126.06	0.65	0.16	90.75	0.40
6	Red edge spectral indices and raw bands	0.72	119.79	0.36	127.54	0.60	0.00	97.04	0.01
7	Gray-level co-occurrence matrix (GLCM) and raw bands	0.75	120.52	0.47	123.67	0.68	0.03	94.97	0.18
8	All	0.77	118.11	0.55	112.13	0.74	0.00	103.43	−0.03

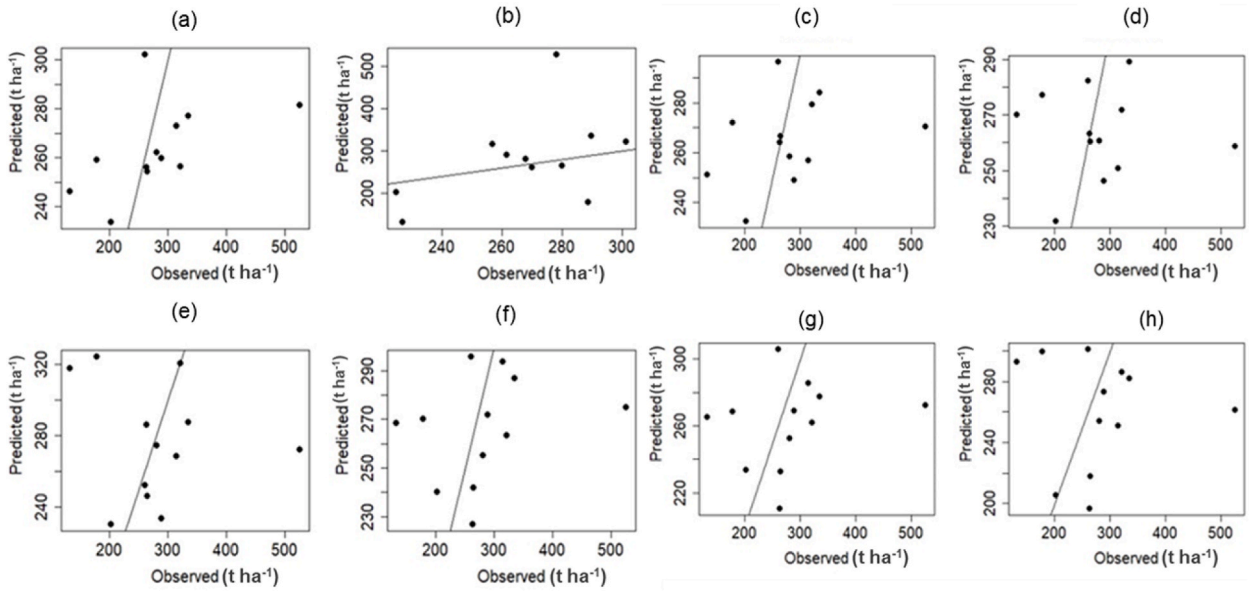


Fig. 4. Scatter plot for predicted and observed biomass in SVM algorithm for test data(a = Scatter plot for Red bands variable combination, b = Scatter plot for Traditional spectral indices variable combination, c = Scatter plot for Red edge spectral indices variable combination, d = Scatter plot for Gray-level co-occurrence matrix variable combination, e = Scatter plot for Red bands and Traditional spectral indices variable combination, f = Scatter plot for Red bands and Red edge spectral indices variable combination, g = Scatter plot for Red bands and Gray-level co-occurrence matrix variable combination and h = Scatter plot for Red bands, Traditional spectral indices, variable combination, Red edge spectral indices and Gray-level co-occurrence matrix**)**.

3.2. Random forest

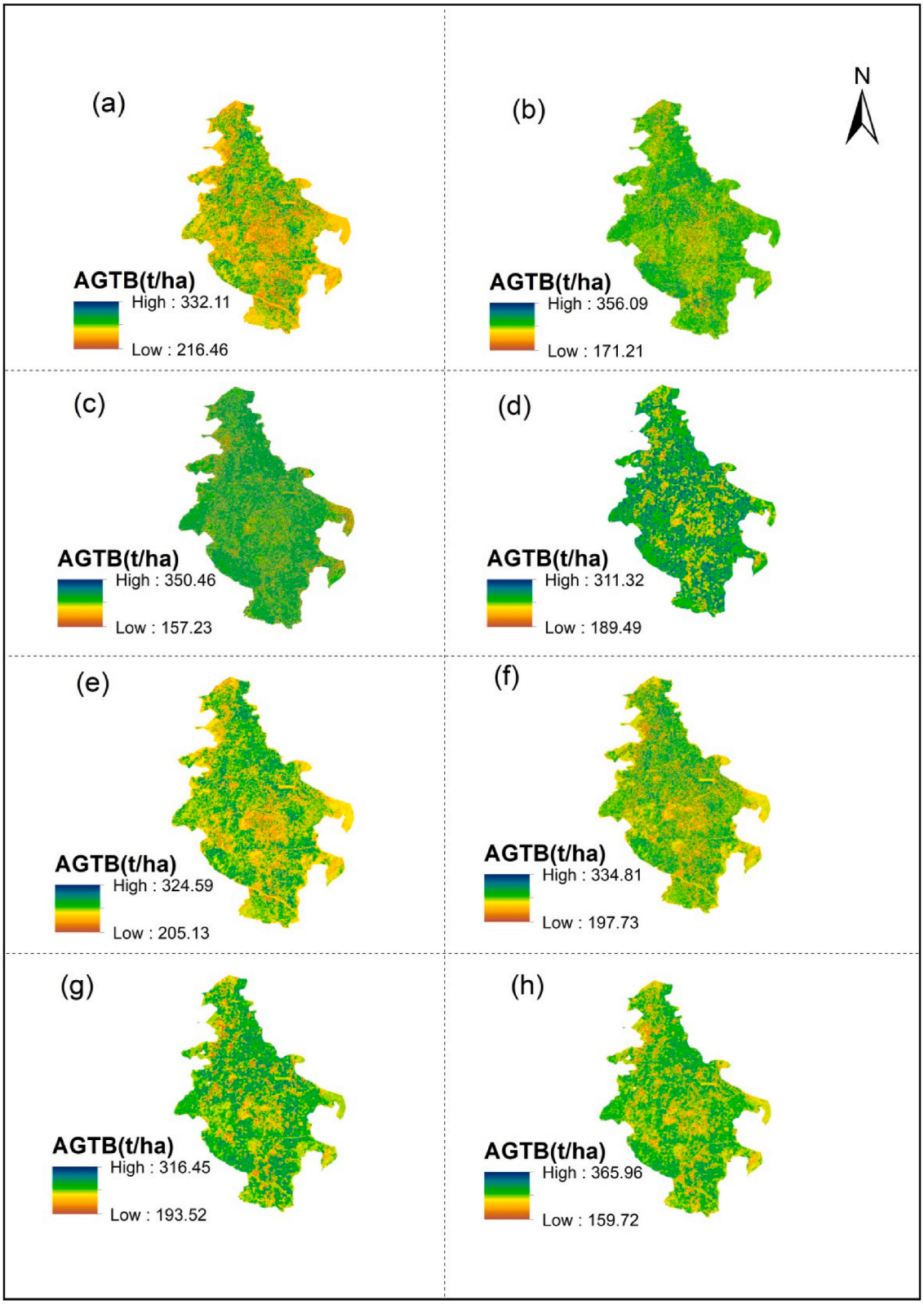
The model performance for each variable combination in model calibration, training data and test data is presented (Table 3). For RF algorithm, eight categories of variable combination reflect varied performance for model calibration, train data and test data. Model performance in calibration is found less RMSE value (122.45 t ha^{-1}) in case of all variables stacked (RB, TSI, RESI, GLCM) for AGTB estimation with R^2 (0.77). The range of AGTB based on all variables by using RF algorithm is found 115.92 to 448.83 t ha^{-1}(Fig. 7(h)) which is significantly different from biomass obtained based on SVM algorithm. Similarly, the algorithm had shown significantly different prediction of AGTB for each variable category combination as compared train data on field measured data versus predicted data and thus TSI plus RB based model has showed good among all with RMSE (55.56 t ha^{-1}), R^2 (0.92) and r (0.96) (Table 2) which ranges the AGTB 109.78 to 471.33 t ha^{-1}(Fig. 7(e)). And test data partition before model build was compared with predicted biomass data by model run for each variable combination, we found best statistical response on GLCM plus RB based model with RMSE (78.81 t ha^{-1}), R^2 (0.34) and r (0.58) (Table 3) and thus has given the AGTB map of range 118.34–425.97 t ha^{-1} (Fig. 7(g)).The scatter plot is clearly shown (Fig. 6) for observed and predicted AGTB (t ha^{-1}) for eight combinations. The predicted spatial AGTB map for each variable combination is visualized (Fig. 7).

3.3. Stochastic gradient boosting

The model performance for each variable combination in model calibration, training data and test data is presented (Table 4). For SGB algorithm; eight categories of variable combination reflect varied performance for model calibration, train data and test data. Model performance in calibration is found less RMSE value (119.84 t ha^{-1}) in case of all variables stacked (RB, TSI, RESI, GLCM) for AGTB estimation with R^2 (0.77) (Table 4). Thus, the SGB algorithm based on all variables has predicted the AGTB map of range 196.65–367.28 t ha^{-1}(Fig. 9(h)). Similarly, the algorithm had shown significantly different prediction of AGTB for each variable category combination as compared train data on field measured data versus predicted data and thus all variable based model has showed good among all with RMSE (111.32 t ha^{-1}), R^2 (0.64) and r (0.80) (Table 4). And test data partition before model build was compared with predicted biomass data by model run for each variable combination, we found best statistical response on all variable based model with RMSE(83.42 t ha^{-1}), R^2 (0.27) and r (0.52) (Table 4).The scatter plot is clearly shown (Fig. 8) for observed and predicted AGTB for eight combinations for test data. The predicted spatial AGTB map for each variable combination is presented in Fig. 9.

3.4. Variable importance

The important of the input variables for the model based on all variables were determined for all algorithms i.e., RF (Fig. 10(a)),



(caption on next page)

Fig. 5. AGTB predicted map by SVM algorithm for each variable combination (a = AGTB predicted map for Red bands variable combination, b = AGTB predicted map for Traditional spectral indices variable combination, c = AGTB predicted map for Red edge spectral indices variable combination, d = AGTB predicted map for Gray-level co-occurrence matrix variable combination, e = AGTB predicted map for Red bands and Traditional spectral indices variable combination, f = AGTB predicted map for Red bands and Red edge spectral indices variable combination, g = AGTB predicted map for Red bands and Gray-level co-occurrence matrix variable combination and h = AGTB predicted map for Red bands, Traditional spectral indices, variable combination, Red edge spectral indices and Gray-level co-occurrence matrix). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

Table 3
Model performances on train and test data for RF algorithm.

S.N.	Variables	Model performance		Train data			Test data		
		R^2	RMSE	R^2	RMSE	r	R^2	RMSE	R
1	Raw bands	0.76	133.37	0.89	67.63	0.94	0.15	89.55	0.39
2	Traditional spectral indices	0.77	123.10	0.91	59.57	0.95	0.34	79.01	0.58
3	Red edge spectral indices	0.73	131.72	0.84	70.15	0.92	0.04	100.69	0.20
4	Gray-level co-occurrence matrix (GLCM)	0.73	126.48	0.86	61.67	0.93	0.32	82.44	0.56
5	Traditional spectral indices and raw bands	0.74	123.70	0.92	55.56	0.96	0.13	92.29	0.36
6	Red edge spectral indices and raw bands	0.71	127.05	0.88	67.87	0.94	0.14	90.6	0.37
7	Gray-level co-occurrence matrix (GLCM) and raw bands	0.73	125.80	0.89	66.55	0.94	0.34	78.81	0.58
8	All	0.77	122.45	0.91	65.21	0.95	0.30	80.78	0.55

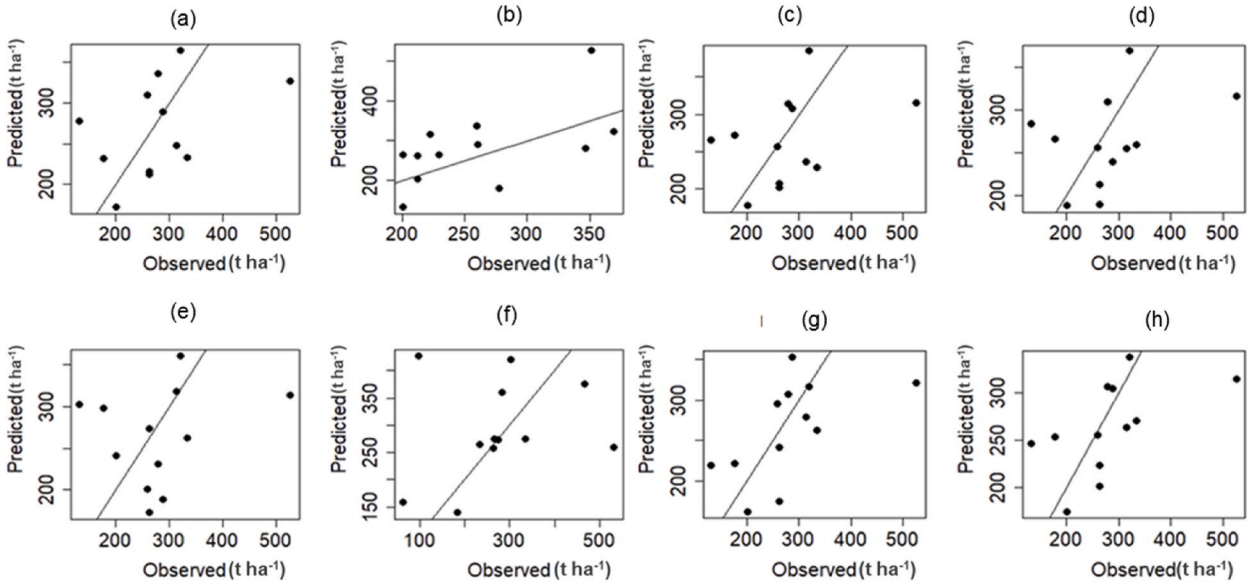
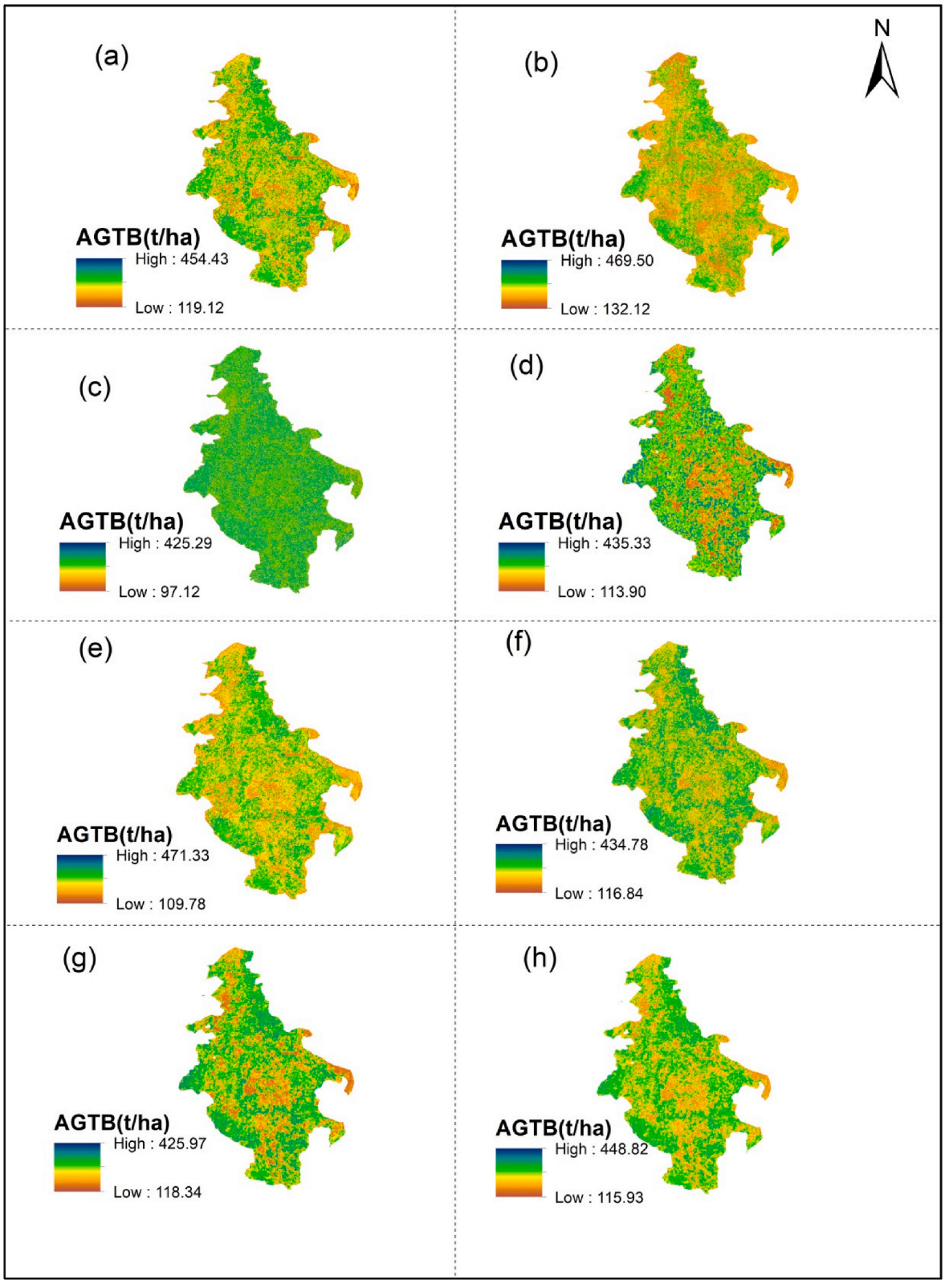


Fig. 6. Scatter plot for predicted and observed biomass for test data in RF algorithm(a = Scatter plot for Red bands variable combination, b = Scatter plot for Traditional spectral indices variable combination, c = Scatter plot for Red edge spectral indices variable combination, d = Scatter plot for Gray-level co-occurrence matrix variable combination, e = Scatter plot for Red bands and Traditional spectral indices variable combination, f = Scatter plot for Red bands and Red edge spectral indices variable combination, g = Scatter plot for Red bands and Gray-level co-occurrence matrix variable combination and h = Scatter plot for Red bands, Traditional spectral indices, variable combination, Red edge spectral indices and Gray-level co-occurrence matrix**).**

SGB (Fig. 10(b)) and SVM (Fig. 10(c)). Thus, NDVI is most important variable in case of RF and SGB followed by entropy, MCARI and so on. But in case of SVM algorithm, B_04 of raw band shows most importance variable followed by B_12, NDVI and so on. The study carried out for AGTB estimation using multiple MLAs had shown that soil-adjusted vegetation index (SAVI) as most important predictive variable followed by NDVI in Yongshou Country. However, SAVI is not included in our study and thus NDVI has been most important variable for two models i.e., RF and SGB.

3.5. Comparative analysis of models

Overall, the RMSE value is found less in RF algorithm than other algorithms i.e., SVM and SGB (Fig. 11). Furthermore, the combination of GLCM texture of near infrared and RB in RF algorithm had shown better performance having less RMSE value of 78.81 t



(caption on next page)

Fig. 7. AGTB predicted map by RF algorithm for each variable combination (a = AGTB predicted map for Red bands variable combination, b = AGTB predicted map for Traditional spectral indices variable combination, c = AGTB predicted map for Red edge spectral indices variable combination, d = AGTB predicted map for Gray-level co-occurrence matrix variable combination, e = AGTB predicted map for Red bands and Traditional spectral indices variable combination, f = AGTB predicted map for Red bands and Red edge spectral indices variable combination, g = AGTB predicted map for Red bands and Gray-level co-occurrence matrix variable combination and h = AGTB predicted map for Red bands, Traditional spectral indices, variable combination, Red edge spectral indices and Gray-level co-occurrence matrix). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

Table 4
Model performances on train and test data for SGB algorithm.

S. N	Variables	Model performance		Train data			Test data		
		R^2	RMSE	R^2	RMSE	r	R^2	RMSE	R
1	Raw bands (RB)	0.69	124.54	0.44	134.09	0.66	0.14	92.15	0.37
2	Traditional spectral indices (TSI)	0.70	119.92	0.49	117.44	0.70	0.30	85.26	0.55
3	Red edge spectral indices (RESI)	0.72	119.92	0.42	121.31	0.65	0.18	87.56	0.42
4	Gray-level co-occurrence matrix (GLCM)	0.74	125.63	0.42	132.20	0.65	0.14	90.60	0.38
5	Traditional spectral indices and raw bands	0.71	121.23	0.58	115.69	0.76	0.25	84.24	0.50
6	Red edge spectral indices and raw bands	0.72	122.21	0.58	116.20	0.76	0.16	87.96	0.40
7	Gray-level co-occurrence matrix (GLCM) and raw bands	0.70	125.73	0.51	129.99	0.71	0.35	87.81	0.59
8	All	0.77	119.84	0.64	111.32	0.80	0.27	83.42	0.52

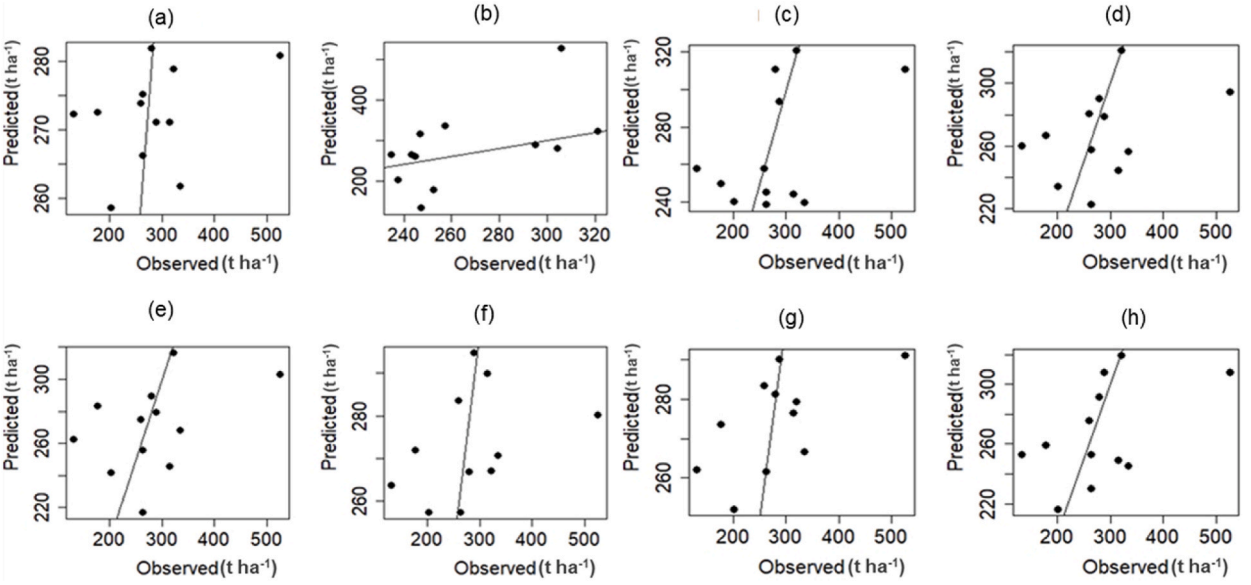
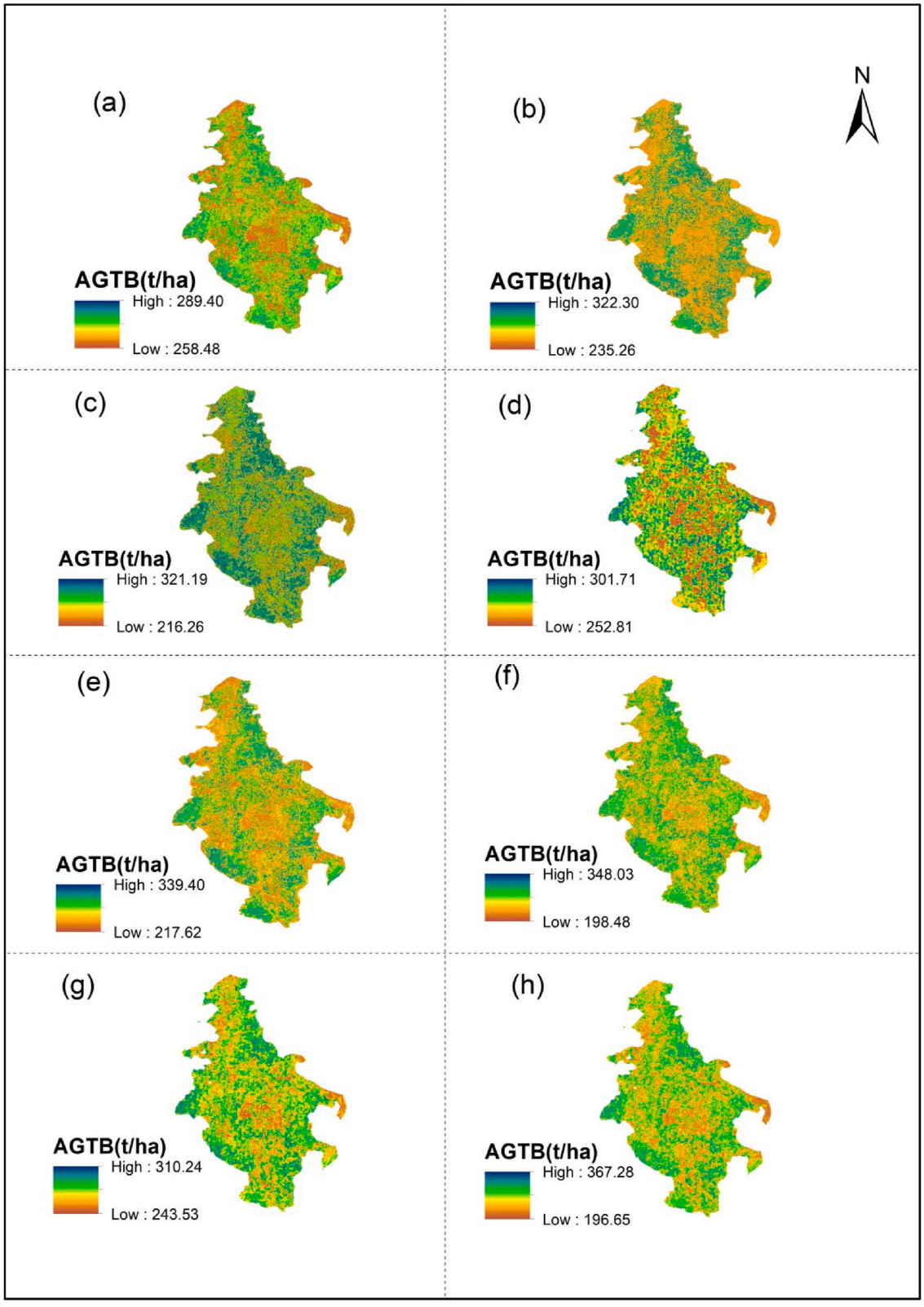


Fig. 8. Scatter plot for predicted and observed biomass for test data in SGB algorithm(a = Scatter plot for Red bands variable combination, b = Scatter plot for Traditional spectral indices variable combination, c = Scatter plot for Red edge spectral indices variable combination, d = Scatter plot for Gray-level co-occurrence matrix variable combination, e = Scatter plot for Red bands and Traditional spectral indices variable combination, f = Scatter plot for Red bands and Red edge spectral indices variable combination, g = Scatter plot for Red bands and Gray-level co-occurrence matrix variable combination and h = Scatter plot for Red bands, Traditional spectral indices, variable combination, Red edge spectral indices and Gray-level co-occurrence matrix).

ha^{-1}among all 24-model run. Similar type of comparative study were carried out in Zhejiang Province, China using five MLAs (KNN, SVM, RF and SGB) concluded that better prediction by RF algorithm [28] which is also resemble with our study output. Furthermore comparative modeling for above ground carbon stock estimation in Iran using MLAs (Boosted regression tree, SVM and RF) again concluded that better performance of RF model among all [52]. Study carried in Myanmar by Ref. [53] had demonstrated that RF based above ground biomass performed better compared to SGB model.

4. Discussion

Nowadays remote sensing-based variable are widely using for forest resource assessment like above ground biomass by application of machine learning algorithms [21,23,53]. This study is preliminary study in Nepal on comparative analysis of machine learning



(caption on next page)

Fig. 9. AGTB predicted map by SGB algorithm for each variable combination(a = AGTB predicted map for Red bands variable combination, b = AGTB predicted map for Traditional spectral indices variable combination, c = AGTB predicted map for Red edge spectral indices variable combination, d = AGTB predicted map for Gray-level co-occurrence matrix variable combination, e = AGTB predicted map for Red bands and Traditional spectral indices variable combination, f = AGTB predicted map for Red bands and Red edge spectral indices variable combination, g = AGTB predicted map for Red bands and Gray-level co-occurrence matrix variable combination and h = AGTB predicted map for Red bands, Traditional spectral indices, variable combination, Red edge spectral indices and Gray-level co-occurrence matrix). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

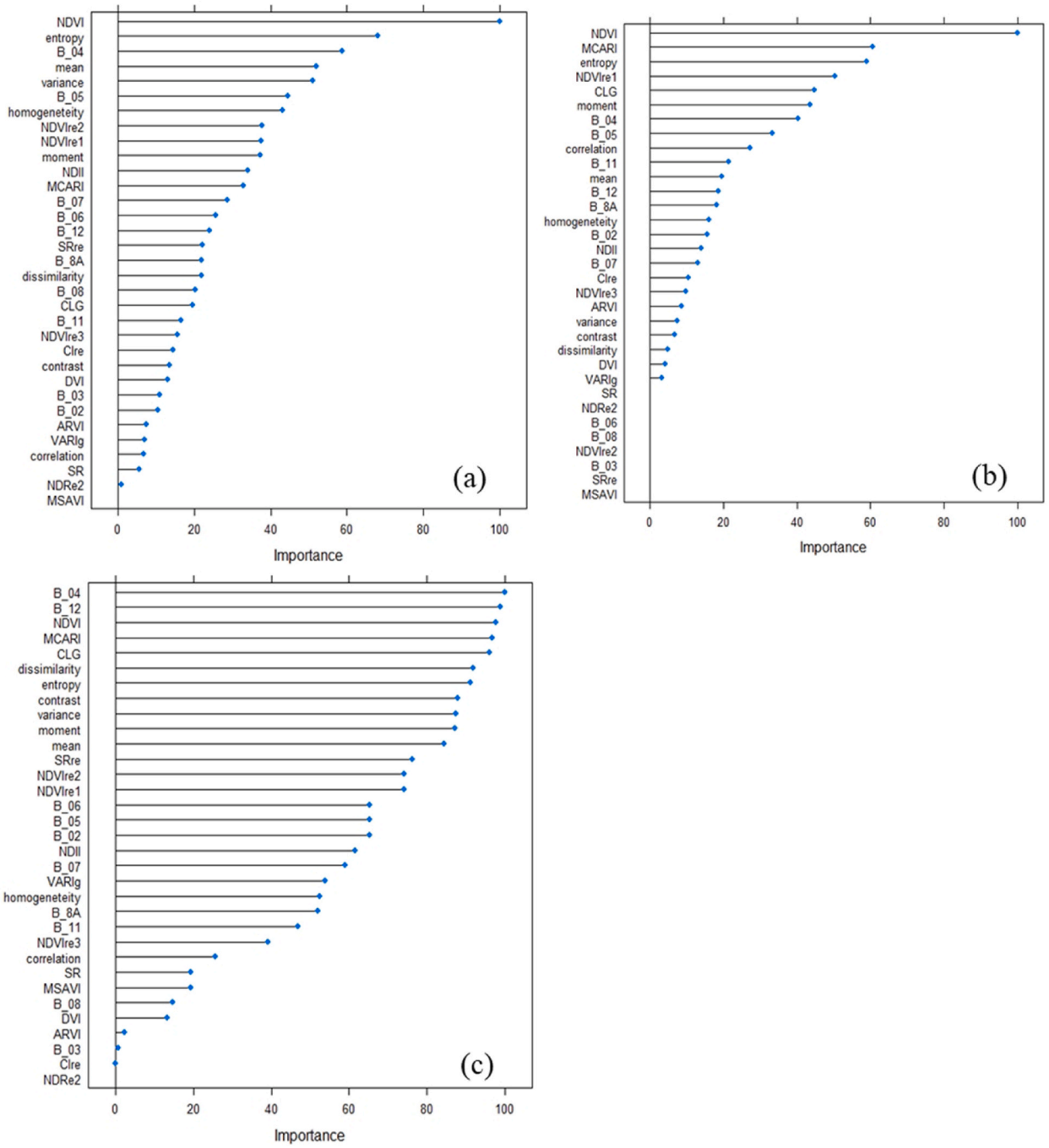


Fig. 10. Variables important; (a) RF, (b) SGB and (c) SVM.

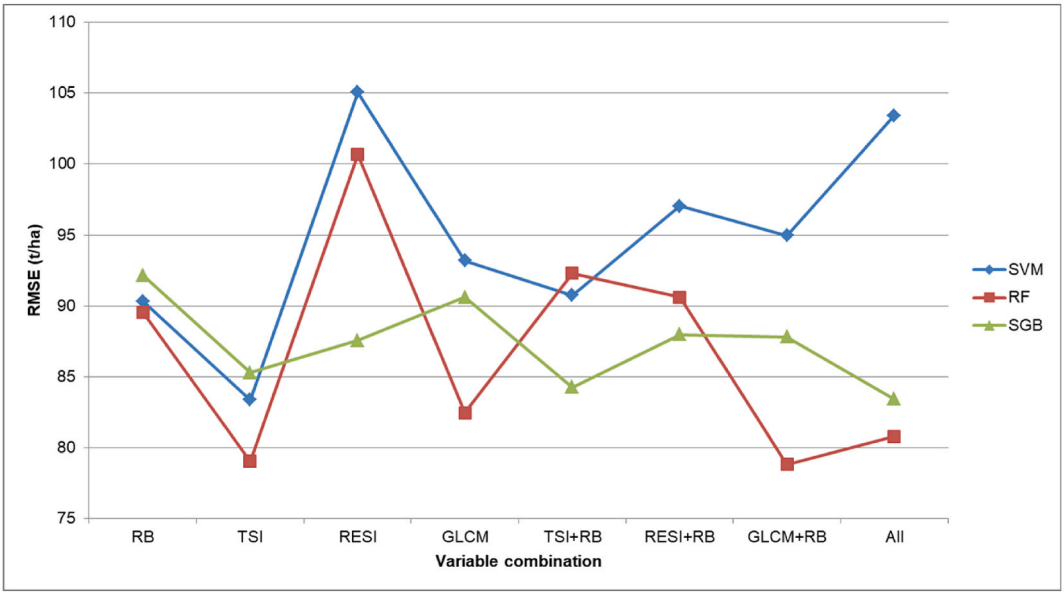


Fig. 11. Comparison of RMSE for each three algorithms across variable combination dataset model.

algorithms and performance of combination of geospatial variable for biomass estimation. Hence the main expected contribution of study was to build a baseline for preparation of sound approach of biomass estimation in Nepal for better quantification of forest resources. In this study three machine learning (RF, SVM and SGB) were used to estimate AGTB of western terai Sal forest of Nepal based on four forms of variables (raw bands, traditional spectral indices, red edge spectral indices and gray level co-occurrence matrix) of remote sensing data (Table 1) with their combinations. On comparison of SVM and SGB, RF performed better for AGTB estimation, had lowest RMSE (78.81 t/ha) and R^2 (0.34). Additionally, the combination of gray level co-occurrence matrix and raw bands has responded better output for AGTB estimation among eight combinations in RF algorithm (Table 3).

The study carried out in Yok Don National Park, Vietnam concluded that sentinel-2 data in conjunction with RF based model has good potential to predict spatial map of AGTB [54] which is also agreeable statement in case of our study. A study on AGTB using the RF algorithm in Himalayan foothills of India with a total 51 spectral indices and textural variables, showed the result having good coefficient of variation and RMSE = 62.56 $Mgha^{-1}$ and hence convinced us to use of remote sensing, field data, and RF algorithm to assess spatial distribution of AGTB [55].A comparative study for three MLAs (RF, SVM and ANN) for AGTB in Yongshou Country demonstrated that RF performed better for estimation [21] as our overall study conclusion also consistent with that RF based model had good performance. Interestingly a research was carried out in Norway nation forest for above ground biomass estimation using RF model which had concluded that model would give better accuracy if we combine multispectral data and Arctic canopy height model data [56]. Again, study carried out by recommended that application of RF model with the combination of synthetic aperture radar backscatters and vegetation indices could give best result for above ground biomass. Furthermore study carried in community forest in Nepal for biomass estimation using RF algorithm for comparison of potential of texture indices from Sentinel-2A data, resulted that biomass estimation based on texture indices plus vegetation indices model fit better than other [57].

Likewise the study carried out in Sierra Madre Occidental in Mexico for above ground estimation using SVM and RF algorithm showed that best performance of SVM with RMSE = 8.20 Mg/ha [5] whose result is different from our result in case of validation RF showed best among three. Furthermore study carried out in Madhya Pradesh, India on biomass prediction comparison for traditional regression models and support vector machine had found better performance of SVM over regression model [58].It has been concluded that integration of SVM algorithm and ALOS/PALSAR data in hilly terrain performed good for biomass estimation [59] and thus such technique can be used in mid-hills of Nepal. Additionally it was clearly demonstrated that support vector machine increased the performance for AGTB estimation using multi resource remote sensing data [60].

Furthermore the study carried out in dense tropical forest of India using SGB and RF had shown that less performance of SGB as compare to RF [61] which resembles with our result in case of test data. SGB model predicted the AGTB having coefficient of determination 0.6 and RMSE (79.45 t/ha) in study done by Ref. [62] however in our study best model predicted based on SGB gave R^2 (0.77) and RMSE (119.84 t ha^{-1}) from model cross validation. Similarly study done [53] in Myanmar by using MLAs (RF based kriging, RF and SGB) resulted that R^2 (0.35) and RMSE (32.02 t ha^{-1}) in case of evergreen forest.

The study also assessed the response of combination of variables for each MLA and thus found that significantly different performance for RF, SVM and SGB. Thus, GLCM and RB, TSI and ALL combination were better predictions for biomass estimation in RF, SVM and SGB respectively. It is also concluded that the variable combination performance for each three algorithms was significantly different. Hence TSI, GLCM + TSI and ALL variable combination have shown better performance for SVM, RF and SGB respectively. The comparison of results provided by previous studies with this study guided more study requirements in Nepal in order to get better

modeling results for biomass estimation. Eventually this type of approach of biomass estimation will definitely help for periodic monitoring for REDD + program in Nepal.

5. Conclusion

In this study, we utilized three machine learning algorithms (SVM, RF, and SGB) with various variable combinations derived from Sentinel-2 MSI satellite data to map the above ground tree biomass (AGTB) in the Sal Forest of western terai, Nepal. Moreover, we evaluated the performance of eight variable datasets for AGTB estimation, resulting in nine models out of 24 that demonstrated better performance through the correlation of predicted and observed data via scatter plot. These models include TSI of SVM, TSI, GLCM plus bands, all variables, GLCM only of RF, and TSI, TSI plus bands, GLCM plus bands, and All variable of SGB, which are expected to predict AGTB accurately in future Nepal. The best model in our study, with the lowest RMSE (78.81 t ha^{-1}) and AGTB range of 118.34–425.97 t ha^{-1}, was the datasets variable of GLCM plus bands of RF algorithms. Additionally, NDVI was found to be the most important variable in RF and SGB algorithms, followed by entropy, B_04, mean, and so on, while the B_04 band was significant in SVM algorithm. Our study concludes that the choice of algorithm and variable selection for modeling is significant for biomass estimation in any forest region.

Combining field inventory data with spectral and textural responses calculated from Sentinel-2 MSI satellite imagery using machine learning algorithms can provide a quick, feasible, reliable, and effective technique to assess spatial AGTB of the Nepalese forest ecosystem. This technique can support monitoring, reporting, and verification of biomass/carbon stock and carbon accounting. Therefore, this type of study should be carried out in other ecological zones to develop better guidelines for estimating biomass/carbon at a spatial level to address the carbon measurement and monitoring issues of REDD+ in Nepal.

Data availability statement

The data used in the study are available in the Google Drive link provided here: https://drive.google.com/drive/u/1/folders/1WYJ71vs1WKDXfdJqOr2adZ1UskKqIm4r.

CRediT authorship contribution statement

Bikram Singh: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. **Amit Kumar Verma:** Project administration, Resources, Supervision, Visualization, Writing – original draft, Writing – review & editing. **Kasip Tiwari:** Formal analysis, Methodology, Visualization, Writing – original draft, Writing – review & editing. **Rajeev Joshi:** Data curation, Project administration, Resources, Supervision, Writing – original draft, Writing – review & editing.

Declaration of competing interest

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

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2023.e21485.

References

[1] G. Chirici, F. Giannetti, R.E. McRoberts, D. Travaglini, M. Pecchi, F. Maselli, M. Chiesi, P. Corona, Wall-to-wall spatial prediction of growing stock volume based on Italian National Forest Inventory plots and remotely sensed data, Int. J. Appl. Earth Obs. Geoinf. 84 (2020), 101959, https://doi.org/10.1016/j.jag.2019.101959.
[2] E.W. Mauya, L.T. Ene, O.M. Bollandsås, T. Gobakken, E. Næsset, R.E. Malimbwi, E. Zahabu, Modelling aboveground forest biomass using airborne laser scanner data in the miombo woodlands of Tanzania, Carbon Balance Manag 10 (2015) 1–16.
[3] E.W. Mauya, J. Koskinen, K. Tegel, J. Hämäläinen, T. Kauranne, N. Käyhkö, Modelling and predicting the growing stock volume in small-scale plantation forests of Tanzania using multi-sensor image synergy, Forests 10 (2019) 279.
[4] J. Morris, Recycle, Bury, or burn wood Waste biomass?: LCA Answer Depends on carbon accounting, emissions Controls, Displaced fuels, and Impact costs, J. Ind. Ecol. 21 (2017) 844–856, https://doi.org/10.1111/jiec.12469.
[5] P.M. López-Serrano, J.L. Cárdenas Domínguez, J.J. Corral-Rivas, E. Jiménez, C.A. López-Sánchez, D.J. Vega-Nieva, Modeling of aboveground biomass with Landsat 8 OLI and machine learning in temperate forests, Forests 11 (2019) 11.
[6] P. Wicaksono, P. Danoedoro, Hartono, U. Nehren, Mangrove biomass carbon stock mapping of the Karimunjawa Islands using multispectral remote sensing, Int. J. Remote Sens. 37 (2016) 26–52.
[7] T. Hu, Y. Su, B. Xue, J. Liu, X. Zhao, J. Fang, Q. Guo, Mapping global forest aboveground biomass with spaceborne LiDAR, optical imagery, and forest inventory data, Remote Sens 8 (2016) 565.
[8] FAO, Global Forest Resources Assessment 2020 (2020), https://doi.org/10.4060/ca9825en.

[9] M. Santoro, A. Beaudoin, C. Beer, O. Cartus, J.E.S. Fransson, R.J. Hall, C. Pathe, C. Schmullius, D. Schepaschenko, A. Shvidenko, M. Thurner, U. Wegmüller, Forest growing stock volume of the northern hemisphere: spatially explicit estimates for 2010 derived from Envisat ASAR, Remote Sens. Environ. 168 (2015) 316–334, https://doi.org/10.1016/j.rse.2015.07.005.

[10] Y. Hu, X. Xu, F. Wu, Z. Sun, H. Xia, Q. Meng, W. Huang, H. Zhou, J. Gao, W. Li, Estimating forest stock volume in Hunan Province, China, by integrating in situ plot data, Sentinel-2 images, and linear and machine learning regression models, Remote Sens 12 (2020) 186.

[11] R.A. Houghton, F. Hall, S.J. Goetz, Importance of biomass in the global carbon cycle, J. Geophys. Res. Biogeosciences. 114 (2009).

[12] S.M. Ghosh, M.D. Behera, B. Jagadish, A.K. Das, D.R. Mishra, A novel approach for estimation of aboveground biomass of a carbon-rich mangrove site in India, J. Environ. Manage. 292 (2021), 112816.

[13] M.A. Njana, H. Meilby, T. Eid, E. Zahabu, R.E. Malimbwi, Importance of tree basic density in biomass estimation and associated uncertainties: a case of three mangrove species in Tanzania, Ann. For. Sci. 73 (2016) 1073–1087.

[14] H. Astola, T. Häme, L. Sirro, M. Molinier, J. Kilpi, Comparison of Sentinel-2 and Landsat 8 imagery for forest variable prediction in boreal region, Remote Sens. Environ. 223 (2019) 257–273.

[15] S. Liang, Recent developments in estimating land surface biogeophysical variables from optical remote sensing, Prog. Phys. Geogr. Earth Environ. 31 (2007) 501–516, https://doi.org/10.1177/0309133307084626.

[16] R.E. McRoberts, E.O. Tomppo, Remote sensing support for national forest inventories, Remote Sens. Environ. 110 (2007) 412–419.

[17] S. Sinha, C. Jeganathan, L.K. Sharma, M.S. Nathawat, A review of radar remote sensing for biomass estimation, Int. J. Environ. Sci. Technol. 12 (2015) 1779–1792, https://doi.org/10.1007/s13762-015-0750-0.

[18] M. Longo, M. Keller, M.N. dos-Santos, V. Leitold, E.R. Pinagé, A. Baccini, S. Saatchi, E.M. Nogueira, M. Batistella, D.C. Morton, Aboveground biomass variability across intact and degraded forests in the Brazilian Amazon, Global Biogeochem. Cycles 30 (2016) 1639–1660.

[19] A. Berninger, S. Lohberger, M. Stängel, F. Siegert, SAR-based estimation of above-ground biomass and its changes in tropical forests of Kalimantan using L-and C-band, Remote Sens 10 (2018) 831.

[20] O.M. Opelele, Y. Yu, W. Fan, C. Chen, S.K. Kachaka, Biomass estimation based on Multilinear regression and machine learning algorithms in the Mayombe tropical forest, in the Democratic Republic of Congo, Appl. Ecol. Environ. Res. 19 (2021) 359–377.

[21] Q. Ye, S. Yu, J. Liu, Q. Zhao, Z. Zhao, Aboveground biomass estimation of black locust planted forests with aspect variable using machine learning regression algorithms, Ecol. Indic. 129 (2021), 107948.

[22] H. Sun, G. Qie, G. Wang, Y. Tan, J. Li, Y. Peng, Z. Ma, C. Luo, Increasing the accuracy of mapping urban forest carbon density by combining spatial modeling and spectral unmixing analysis, Remote Sens 7 (2015) 15114–15139.

[23] Y. Zhang, J. Ma, S. Liang, X. Li, M. Li, An evaluation of eight machine learning regression algorithms for forest aboveground biomass estimation from multiple satellite data products, Remote Sens 12 (2020) 4015.

[24] Y. Su, Q. Guo, B. Xue, T. Hu, O. Alvarez, S. Tao, J. Fang, Spatial distribution of forest aboveground biomass in China: estimation through combination of spaceborne lidar, optical imagery, and forest inventory data, Remote Sens. Environ. 173 (2016) 187–199.

[25] L. Breiman, J.H. Friedman, R.A. Olshen, C.J. Stone, Classification and Regression Trees, Routledge, 2017.

[26] S. Pandit, S. Tsuyuki, T. Dube, Estimating Above-Ground Biomass in Sub-tropical Buffer Zone Community Forests , Nepal , Using Sentinel 2 Data, 2018, https://doi.org/10.3390/rs10040601.

[27] C.T. de Almeida, L.S. Galvao, J.P.H.B. Ometto, A.D. Jacon, F.R. de Souza Pereira, L.Y. Sato, A.P. Lopes, P.M.L. de Alencastro Graça, C.V. de Jesus Silva, J. Ferreira-Ferreira, Combining LiDAR and hyperspectral data for aboveground biomass modeling in the Brazilian Amazon using different regression algorithms, Remote Sens. Environ. 232 (2019), 111323.

[28] C. Wu, H. Shen, A. Shen, J. Deng, M. Gan, J. Zhu, H. Xu, K. Wang, Comparison of machine-learning methods for above-ground biomass estimation based on Landsat imagery, J. Appl. Remote Sens. 10 (2016), 35010.

[29] Nepal National MoFE, REDD+ Strategy (2018-2022), 2018. http://redd.gov.np/post/nepal-national-redd-strategy-2018.

[30] DFRS, STATE of NEPAL ' S FORESTS, 2015.

[31] Y.K. Karna, Y.A. Hussin, H. Gilani, M.C. Bronsveld, M.S.R. Murthy, F.M. Qamer, B.S. Karky, T. Bhattarai, X. Aigong, C.B. Baniya, Integration of WorldView-2 and airborne LiDAR data for tree species level carbon stock mapping in Kayar Khola watershed, Nepal, Int. J. Appl. Earth Obs. Geoinf. 38 (2015) 280–291.

[32] Y.A. Hussin, H. Gilani, L. Van Leeuwen, M.S.R. Murthy, R. Shah, S. Baral, N.E. Tsendbazar, S. Shrestha, S.K. Shah, F.M. Qamer, Evaluation of object-based image analysis techniques on very high-resolution satellite image for biomass estimation in a watershed of hilly forest of Nepal, Appl. Geomatics. 6 (2014) 59–68, https://doi.org/10.1007/s12518-014-0126-z.

[33] DoFSC, Community Forest Inventory Guideline, 2005.

[34] J. Chave, M. Réjou-Méchain, A. Búrquez, E. Chidumayo, M.S. Colgan, W.B.C. Delitti, A. Duque, T. Eid, P.M. Fearnside, R.C. Goodman, Improved allometric models to estimate the aboveground biomass of tropical trees, Glob. Chang. Biol. 20 (2014) 3177–3190.

[35] ESA, User Handbook, ESA Stand. Doc. 64 (2015).

[36] J.W. Rouse, R.H. Haas, J.A. Schell, D.W. Deering, Monitoring vegetation systems in the Great Plains with ERTS, NASA Spec. Publ. 351 (1974) 309.

[37] A.A. Gitelson, Y. Gritz, M.N. Merzlyak, Relationships between leaf chlorophyll content and spectral reflectance and algorithms for non-destructive chlorophyll assessment in higher plant leaves, J. Plant Physiol. 160 (2003) 271–282.

[38] A.A. Gitelson, Y.J. Kaufman, R. Stark, D. Rundquist, Novel algorithms for remote estimation of vegetation fraction, Remote Sens. Environ. 80 (2002) 76–87.

[39] V. Klemas, R. Smart, The influence of soil salinity, growth form, and leaf moisture on-the spectral radiance of, Photogramm. Eng. Remote Sens. 49 (1983) 77–83.

[40] J. Qi, A. Chehbouni, A.R. Huete, Y.H. Kerr, S. Sorooshian, A modified soil adjusted vegetation index, Remote Sens. Environ. 48 (1994) 119–126.

[41] C.F. Jordan, Derivation of leaf-area index from quality of light on the forest floor, Ecology 50 (1969) 663–666.

[42] C.J. Tucker, Red and photographic infrared linear combinations for monitoring vegetation, Remote Sens. Environ. 8 (1979) 127–150.

[43] Y.J. Kaufman, D. Tanre, Atmospherically resistant vegetation index (ARVI) for EOS-MODIS, IEEE Trans. Geosci. Remote Sens. 30 (1992) 261–270.

[44] C.S.T. Daughtry, C.L. Walthall, M.S. Kim, E.B. De Colstoun, J.E. McMurtrey Iii, Estimating corn leaf chlorophyll concentration from leaf and canopy reflectance, Remote Sens. Environ. 74 (2000) 229–239.

[45] A. Gitelson, M.N. Merzlyak, Spectral reflectance changes associated with autumn senescence of Aesculus hippocastanum L. and Acer platanoides L. leaves. Spectral features and relation to chlorophyll estimation, J. Plant Physiol. 143 (1994) 286–292.

[46] A. Fernández-Manso, O. Fernández-Manso, C. Quintano, SENTINEL-2A red-edge spectral indices suitability for discriminating burn severity, Int. J. Appl. Earth Obs. Geoinf. 50 (2016) 170–175.

[47] D.A. Sims, J.A. Gamon, Relationships between leaf pigment content and spectral reflectance across a wide range of species, leaf structures and developmental stages, Remote Sens. Environ. 81 (2002) 337–354.

[48] A.A. Gitelson, A. Viña, V. Ciganda, D.C. Rundquist, T.J. Arkebauer, Remote estimation of canopy chlorophyll content in crops, Geophys. Res. Lett. 32 (2005).

[49] E.M. Barnes, T.R. Clarke, S.E. Richards, P.D. Colaizzi, J. Haberland, M. Kostrzewski, P. Waller, C. Choi, E. Riley, T. Thompson, Coincident detection of crop water stress, nitrogen status and canopy density using ground based multispectral data, in: Proc. Fifth Int. Conf. Precis. Agric, . Bloom. MN, USA, 2000.

[50] R.M. Haralick, K. Shanmugam, I.H. Dinstein, Textural features for image classification, IEEE Trans. Syst. Man. Cybern. (1973) 610–621.

[51] M. Batta, Machine learning algorithms - a review, Int. J. Sci. Res. (IJ. 9 (2020) 381–undefined, https://doi.org/10.21275/ART20203995.

[52] A. Safari, H. Sohrabi, S. Powell, S. Shataee, A Comparative Assessment of Multi-Temporal Landsat 8 and Machine Learning Algorithms for Estimating Aboveground Carbon Stock in Coppice Oak Forests, 2017, pp. 6407–6432, https://doi.org/10.1080/01431161.2017.1356488.

[53] P. Wai, H. Su, M. Li, Estimating aboveground biomass of two different forest types in Myanmar from sentinel-2 data with machine learning and geostatistical algorithms, Remote Sens 14 (2022) 2146.

[54] A.T.N. Dang, S. Nandy, R. Srinet, N.V. Luong, S. Ghosh, A.S. Kumar, Forest aboveground biomass estimation using machine learning regression algorithm in Yok Don National Park, Vietnam, Ecol, Inform 50 (2019) 24–32.

[55] S. Purohit, S.P. Aggarwal, N.R. Patel, Estimation of forest aboveground biomass using combination of Landsat 8 and Sentinel-1A data with random forest regression algorithm in Himalayan Foothills, Trop. Ecol. 622 (62) (2021) 288–300, https://doi.org/10.1007/S42965-021-00140-X, 2021.
[56] S. Puliti, M. Hauglin, J. Breidenbach, P. Montesano, C.S.R. Neigh, J. Rahlf, S. Solberg, T.F. Klingenberg, R. Astrup, Modelling above-ground biomass stock over Norway using national forest inventory data with ArcticDEM and Sentinel-2 data, Remote Sens. Environ. 236 (2020), 111501.
[57] S. Pandit, S. Tsuyuki, T. Dube, Exploring the inclusion of Sentinel-2 MSI texture metrics in above-ground biomass estimation in the community forest of Nepal, Geocarto Int. 35 (2020) 1832–1849.
[58] D. Deb, S. Deb, D. Chakraborty, J.P. Singh, A.K. Singh, P. Dutta, A. Choudhury, Aboveground biomass estimation of an agro-pastoral ecology in semi-arid Bundelkhand region of India from Landsat data: a comparison of support vector machine and traditional regression models, Geocarto Int. 37 (2022) 1043–1058, https://doi.org/10.1080/10106049.2020.1756461.
[59] T. Sivasankar, J.M. Lone, K. S.K, A. Qadir, N. R.P.L, Estimation of above ground biomass using support vector machines and ALOS/PALSAR data, Vietnam J. Earth Sci. 41 (2019) 95–104, https://doi.org/10.15625/0866-7187/41/2/13690.
[60] Y. Guo, Z. Li, X. Zhang, E.X. Chen, L. Bai, X. Tian, Q. He, Q. Feng, W. Li, Optimal Support Vector Machines for forest above-ground biomass estimation from multisource remote sensing data, Int. Geosci. Remote Sens. Symp. (2012) 6388–6391, https://doi.org/10.1109/IGARSS.2012.6352721.
[61] S.M. Ghosh, M.D. Behera, Aboveground biomass estimation using multi-sensor data synergy and machine learning algorithms in a dense tropical forest, Appl. Geogr. 96 (2018) 29–40, https://doi.org/10.1016/j.apgeog.2018.05.011.
[62] S.M. Ghosh, M.D. Behera, Aboveground biomass estimation using multi-sensor data synergy and machine learning algorithms in a dense tropical forest, Appl. Geogr. 96 (2018) 29–40.