4th Intercontinental Geoinformation Days

igd.mersin.edu.tr

Global scale-biomass estimation based on a deep learning method

Somayeh Talebiesfandarani [*1], **Ali Shamsoddini** [1]

[1]Tarbiat Modares University, Department of Remote sensing and GIS, Tehran, Iran

Keywords
Biomass
CNN
Microwave
Feature Selection

Abstract

Modeling accurate above-ground biomass (AGB) maps is a critical issue in remote sensing research. Since the relationship between biomass and environmental variables are usually complex, because of being affected by many factors, using non-parametric methods like Convolutional Neural Network (CNN) to estimate biomass on the global scale is convenient. To choose the most significant variables to enter to the AGB estimation model two feature selection techniques were applied, Support Vector Machine for Regression Feature Selection (SVRFS) and Random Forest Feature Selection (RFFS) techniques. The optimum AGB model was created using the training dataset and the predicted model was created using the test dataset. The results showed CNN with the SVRFS technique, achieved the highest RMSE values (31.22 Mg/ha). This study highlighted the capability of the deep learning algorithm to improve AGB estimates on a global scale.

1. Introduction

Biomass is a good measure of plant domination in research. Remote sensing technology is a powerful tool in biomass estimation in regional and global scale (Lu 2006). Passive microwave observations can provide data from both green and non-green vegetation components(Liu, Van Dijk et al. 2015, Talebiesfandarani, Zhao et al. 2019) and can take information from a deeper layer of vegetation, depending on the frequency. Besides, passive microwave observations are insensitive to cloud cover. Biomass estimation using passive microwave data is based on Vegetation Optical Depth (VOD) (Ulaby, Kouyate et al. 1986, Momen, Wood et al. 2017). The VOD retrieved at lower frequencies like L-band has special relationship to the vegetation features of the whole canopy. Beside VOD, there are many variables that are important in biomass estimation(Rodríguez-Fernández, Mialon et al. 2018, Vittucci, Laurin et al. 2019). In this research, beside VOD, precipitation, temperature, tree height, NDVI, EVI, climatic water availability (CWA) and evapotranspiration were used to enter to the models.

Recently, Non-parametric models like convolutional neural network (CNN) as a deep learning approach are a popular way to analyze complex environment relationships (Chen, Ren et al. 2018, Jin, Li et al. 2020, Kattenborn, Leitloff et al. 2021).

It's clear that in a deep learning model, all the input variables in the biomass estimation dataset are not helpful to build the model. Furthermore, adding many variables from different datasets to the model can cause model complexity and reduce the overall accuracy of the model. Efforts were made to mitigate these problems using some feature selection techniques to find the best series of features and build an effective biomass model. Here Random Forest Feature Selection (RFFS) and Support Vector Regression Feature Selection (SVRFS) were used to find how the accuracy of the biomass estimation is affected by the different combinations of feature selection techniques.

The main goals of this study were:

-Using feature selection techniques to select the most influential variables in biomass estimation.

-Investigating the ability of a deep learning method on biomass estimation in the global scale and comparing the models.

2. Method

2.1. Input Data

To take climatic variables the Climatic Research Unit Time Series (CRU TS) 4.05 diffusion (Harris et al., 2014) has been utilized in the present study. The CRU TS dataset that used here consists of precipitation,

* **Corresponding Author**

*(so.talebi@modares.ac.ir) ORCID ID 0000-0001-6141-6489
(ali.shamsoddini@modares.ac.ir) ORCID ID 0000-0002-4559-7563

Cite this study

Talebiesfandarani, S., & Shamsoddini, A. (2022). Global scale-biomass estimation based on a deep learning method. 4th Intercontinental Geoinformation Days (IGD), 207-210, Tabriz, Iran

temperature (air temperature at 2m above the soil) and evapotranspiration. In addition to CRU TS, the climatic water availability (CWA) was used as one of the inputs to the models. CWA [mm/yr.], show the amount of water lost during the dry season. Here is supposed CWA not to change during the period of this research (2000-2019).

The VOD data (SMOS-IC), used in this study, at L band was obtained from SMOS which operate at 1.2 GHz. It is a global product with 25-km spatial and one-day temporal resolution.

The input data consist of yearly, seasonal and monthly data.

AGB map was derived as a part of the Climate Change Initiative, CCI project (CCI AGB D4.3, 2020)(Santoro, Cartus et al. 2021). NDVI, EVI and tree high also were used as the input data to the CNN model.

2.2. CNN

Our CNN model consists of 5 convolution layers. Each convolution layer extracts a special-spectral feature. To avoid overfitting, dropout function was used but no special improvement in the test data. Stepwise learning rate was used to avoid overfitting during training the model. All the layers of the network performed the feature extraction. In the first layers, low-level features and in the last layers high-level features was extracted. Here, a fully connected neural network is used. Adjusting the optimal hyperparameters for CNN was based on random search, because the grid search was time consuming. Table 1 shows the configurations of the tuning hyperparameters for CNN algorithm.

The original data are divided into training dataset (80%) and testing dataset (20%). The training data are classified into five folds (five-fold cross-validation), four folds chosen for training and one left signified for validation. Each of the five folds operates once as validation set and four times as training data. The minimum average RMSE in the five validation datasets is a key specifying the optimal combination of hyperparameters.

Table 1. The procedure of tuning the hyperparameters adjusted for each model

Algorithm	Hyperparameters Tuned	Hyperparameter Configurations
CNN	Epoch H Kernel Learning rate seed	10000,30000, 70000,100000 [299, 24, 5, 3, 1], [299, 24, 8, 2, 1], 3*3, 5*5, 7*7 0.1, 0.2, .0.3 10-100interval 10

2.3. Feature Selection Techniques

Here RFFS and SVRFS are chosen to apply to the original dataset to specify suitable variables. RFFS is an embedded algorithm that utilizes random forests as the base arranger(Rodríguez-Fernández, Mialon et al. 2018). Firstly, the random forest was fitted for all the AGB features. In every run, one variable is permuted by random permutation model and other variables are entered in to the model without permutation. This process was implemented for all input variables one by one. Paying attention to RMSE values from permutations and sorting them showed the most important variables were those with greater RMSE values. Then, 20% of the less important attributes were removed, and again, the model fitted with other variables. This step was repeated and the process was done until the least numbers of variables with the highest RMSE values were left. In the final step, the most important features were ranked(Shamsoddini, Trinder et al. 2013, Shamsoddini and Raval 2018).

In the SVRFS technique, the process was done the same as RFFS technique, and the SVR model was used as the base arranger to choose appropriate variables to enter the models(Lal, Chapelle et al. 2006, Rodriguez-Galiano, Luque-Espinar et al. 2018).

Biomass estimation has been done using two feature selection techniques with a combination of CNN model (RFFS-CNN SVRFS-CNN respectively).

2.4. Evaluation of the AGB estimation models

The AGB models trained with all training data (80%) and 20% of the initial data as an independent-test dataset were used to estimate the coefficient of determination (R^2), root mean square error (RMSE), relative RMSE (RMSE%), and bias. Paired samples t-test also was used for statistically comparing the efficiency of different AGB models.

3. Results

3.1. The primitive AGB estimation models result with all layers as inputs

Table 2 represented the distribution of the R^2, RMSE, RMSE%, and bias for CNN biomass estimation model in the case of entering all the input data to the models (without using feature selection techniques).

Table 2. The performance metrics for CNN algorithm with all input variables

	R^2	RMSE(Mg/ha)	RMSE%	bias (Mg/ha)
CNN	0.8795	40.5700	30.1041	0.50

3.2. Specifying the optimal number of input variables (implementation of feature selection techniques)

The trend of RMSE values of RFFS-CNN SVRFS-CNN models as the input variable numbers changed is represented in fig. 1. Input data started from about 300 variables and low-scoring features were removed until eight variables left. In general, in the figure, by decreasing the number of variables, RMSE values decreased.

Tree height, average EVI index in autumn 2017, average EVI index in winter in 2017, average evapotranspiration in winter 2017, average temperature in spring 2004, average NDVI in August 2017, average optical depth in January 2017 and mean NDVI in October in 2017were included in 8 most effective variables when use RFFS respectively.

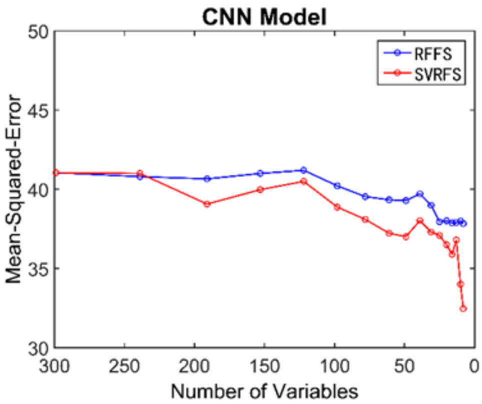

Fig.1. The trend of RMSE values for different number of variables for the CNN model

When SVRFS model was used as feature selection model tree height, Average spring rainfall in 2004, Average optical depth in November 2017, average NDVI in August 2017, average NDVI in autumn 2017, average optical depth in January 2017, average spring precipitation in 2016 and average optical depth in spring 2017 were the 8 most significant variables in AGB estimation. Tree Hight, mean NDVI in August 2017 and mean optical depth in January 2017 were the same variables in two feature selection models. The most important variable in two feature selection models was tree height. It shows tree height variables are fundamental variable in AGB estimation regardless of the method is used.

Since the most important target in using feature selection technique is reducing the RMSE value, precise attention to the accuracies, indicated SVRFS technique compared to RFFS could reduce RMSE value appropriately in AGB estimation model (31.2209 Mg/ha, for SVRFS-CNN, compare to 36.6764 Mg/ha, for RFFS-CNN). Here also SVRFS technique could reduce overestimation and underestimation of the models more significant than RFFS.

3.1. Evaluation of the Model

Statistical comparison of different models was done sing the paired sample t-test (Shamsoddini and Raval, 2018). Table 3 indicated the p-value derived from paired sample t-test. No statistical differences were between the CNN model and CNN in combination with two feature selection techniques. RFFS-CNN and SVRFS-CNN also have no statistically difference but SVRFS-CNN outperformed better than RFFS-CNN (Table 3).

Table3. P-value related to paired sample t-test to compare two models

	CNN	RFFS-CNN	SVRFS-CNN
CNN	*	0.94	0.20
RFFS-CNN		*	0.17
SVRFS-CNN			*

4. Discussion

CNN model could predict biomass well (Kussul, Lavreniuk et al. 2017). One of the important issues in taking the best results from the CNN model depended on choosing the optimal scale for it. The optimal scale depends on many factors like the appropriate number and type of samples, spatial and spectral resolution, sensor and land cover type(Dong, Du et al. 2020). Another important issue in CNN model precision for biomass monitoring is adjusting and tuning the hyperparameters properly(Dong, Du et al. 2020, Gupta, Rajnish et al. 2021). In this research, many efforts been made in choosing optimal scale and tuning the hyperparameters.

In this research SVRFS technique outperformed RFFS in the AGB mode(Tuong, Tani et al. 2020). It means that the SVRFS technique could retain more convenient information from the original data set. Also, compared to the AGB models with all variables as inputs, the least overestimations and underestimations related to the CNN model with combination of feature selection techniques(Li, Li et al. 2020). Here, tree height was the most practical feature in both feature selection techniques(Wang, Zhang et al. 2021). It is noteworthy that reliable reference biomass maps such as Saatchi(Saatchi, Harris et al. 2011) and Glob biomass(Santoro, Cartus et al. 2018, Santoro, Cartus et al. 2021) were produced based on tree height measurements(Nogueira, Engel et al. 2014). The other joint selected variables were VOD and NDVI. VOD has coarse special resolution but its high sensitivity to AGB makes it suitable indicator for biomass monitoring in large scale(Chaparro, Duveiller et al. 2019). AGB also has a significant relationship with NDVI, especially in shorter vegetation cover(Goswami, Gamon et al. 2015).

Finally, it should be noted that improving AGB estimation for landcovers and ecoregions can investigate more details in biomass estimation and provide stable results in contributions of the predictor variables to the AGB estimation model and is necessary in the future.

5. Conclusion

This study compared CNN AGB estimation model with and without feature selection techniques. The outcomes of this study were as follows:

- CNN model could predict biomass well on the global scale.

-Feature selection techniques were an effective tool in choosing the best variables to enter into the AGB model and improved estimation results. SVRFS technique outperformed RFFS.

-The jointly selected variables in two feature selection techniques were tree height, VOD and NDVI.

- The best models to estimate AGB with the combination of feature selection techniques were SVRFS-CNN with the lowest RMSE values.

References

Chaparro, D., G. Duveiller, M. Piles, A. Cescatti, M. Vall-Llossera, A. Camps and D. Entekhabi (2019). "Sensitivity of L-band vegetation optical depth to carbon stocks in tropical forests: a comparison to higher frequencies and optical indices." Remote sensing of environment **232**: 111303.

Chen, L., C. Ren, B. Zhang, Z. Wang and Y. Xi (2018). "Estimation of forest above-ground biomass by geographically weighted regression and machine learning with sentinel imagery." Forests **9**(10): 582.

Dong, L., H. Du, N. Han, X. Li, D. e. Zhu, F. Mao, M. Zhang, J. Zheng, H. Liu and Z. Huang (2020). "Application of convolutional neural network on Lei Bamboo above-ground-biomass (AGB) estimation using Worldview-2." Remote Sensing **12**(6): 958.

Goswami, S., J. Gamon, S. Vargas and C. Tweedie (2015). Relationships of NDVI, Biomass, and Leaf Area Index (LAI) for six key plant species in Barrow, Alaska, PeerJ PrePrints.

Gupta, M., K. Rajnish and V. Bhattacharjee (2021). "Impact of Parameter Tuning for Optimizing Deep Neural Network Models for Predicting Software Faults." Scientific Programming **2021**.

Jin, X., Z. Li, H. Feng, Z. Ren and S. Li (2020). "Deep neural network algorithm for estimating maize biomass based on simulated Sentinel 2A vegetation indices and leaf area index." The Crop Journal **8**(1): 87-97.

Kattenborn, T., J. Leitloff, F. Schiefer and S. Hinz (2021). "Review on Convolutional Neural Networks (CNN) in vegetation remote sensing." ISPRS Journal of Photogrammetry and Remote Sensing **173**: 24-49.

Kussul, N., M. Lavreniuk, S. Skakun and A. Shelestov (2017). "Deep learning classification of land cover and crop types using remote sensing data." IEEE Geoscience and Remote Sensing Letters **14**(5): 778-782.

Lal, T. N., O. Chapelle, J. Weston and A. Elisseeff (2006). Embedded methods. Feature extraction, Springer: 137-165.

Li, Y., M. Li, C. Li and Z. Liu (2020). "Forest aboveground biomass estimation using Landsat 8 and Sentinel-1A data with machine learning algorithms." Scientific reports **10**(1): 1-12.

Liu, Y. Y., A. I. Van Dijk, R. A. De Jeu, J. G. Canadell, M. F. McCabe, J. P. Evans and G. Wang (2015). "Recent reversal in loss of global terrestrial biomass." Nature Climate Change **5**(5): 470-474.

Lu, D. (2006). "The potential and challenge of remote sensing-based biomass estimation." International journal of remote sensing **27**(7): 1297-1328.

Momen, M., J. D. Wood, K. A. Novick, R. Pangle, W. T. Pockman, N. G. McDowell and A. G. Konings (2017). "Interacting effects of leaf water potential and biomass on vegetation optical depth." Journal of Geophysical Research: Biogeosciences **122**(11): 3031-3046.

Nogueira, L. R., V. L. Engel, J. A. Parrotta, A. C. G. d. Melo and D. S. Ré (2014). "Allometric equations for estimating tree biomass in restored mixed-species Atlantic Forest stands." Biota Neotropica **14**.

Rodríguez-Fernández, N. J., A. Mialon, S. Mermoz, A. Bouvet, P. Richaume, A. Al Bitar, A. Al-Yaari, M. Brandt, T. Kaminski and T. Le Toan (2018). "An evaluation of SMOS L-band vegetation optical depth (L-VOD) data sets: high sensitivity of L-VOD to above-ground biomass in Africa." Biogeosciences **15**(14): 4627-4645.

Rodriguez-Galiano, V. F., J. A. Luque-Espinar, M. Chica-Olmo and M. P. Mendes (2018). "Feature selection approaches for predictive modelling of groundwater nitrate pollution: An evaluation of filters, embedded and wrapper methods." Science of the total environment **624**: 661-672.

Saatchi, S. S., N. L. Harris, S. Brown, M. Lefsky, E. T. Mitchard, W. Salas, B. R. Zutta, W. Buermann, S. L. Lewis and S. Hagen (2011). "Benchmark map of forest carbon stocks in tropical regions across three continents." Proceedings of the national academy of sciences **108**(24): 9899-9904.

Santoro, M., O. Cartus, N. Carvalhais, D. Rozendaal, V. Avitabile, A. Araza, S. de Bruin, M. Herold, S. Quegan and P. Rodríguez-Veiga (2021). "The global forest above-ground biomass pool for 2010 estimated from high-resolution satellite observations." Earth System Science Data **13**(8): 3927-3950.

Santoro, M., O. Cartus, S. Mermoz, A. Bouvet, T. Le Toan, N. Carvalhais, D. Rozendaal, M. Herold, V. Avitabile and S. Quegan (2018). A detailed portrait of the forest aboveground biomass pool for the year 2010 obtained from multiple remote sensing observations. EGU general assembly conference abstracts.

Shamsoddini, A. and S. Raval (2018). "Mapping red edge-based vegetation health indicators using Landsat TM data for Australian native vegetation cover." Earth Science Informatics **11**(4): 545-552.

Shamsoddini, A., J. C. Trinder and R. Turner (2013). "Non-linear methods for inferring lidar metrics using SPOT-5 textural data." ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences **2**: 259-264.

Talebiesfandarani, S., T. Zhao, J. Shi, P. Ferrazzoli, J.-P. Wigneron, M. Zamani and P. Pani (2019). "Microwave vegetation index from multi-angular observations and its application in vegetation properties retrieval: Theoretical modelling." Remote Sensing **11**(6): 730.

Tuong, T. T. C., H. Tani, X. Wang, N. Q. Thang and H. M. Bui (2020). "Combination of SAR Polarimetric Parameters for Estimating Tropical Forest Aboveground Biomass." Polish Journal of Environmental Studies **29**(5).

Ulaby, F. T., F. Kouyate, B. Brisco and T. L. Williams (1986). "Textural infornation in SAR images." IEEE Transactions on Geoscience and Remote Sensing(2): 235-245.

Vittucci, C., G. V. Laurin, G. Tramontana, P. Ferrazzoli, L. Guerriero and D. Papale (2019). "Vegetation optical depth at L-band and above ground biomass in the tropical range: Evaluating their relationships at continental and regional scales." International Journal of Applied Earth Observation and Geoinformation **77**: 151-161.

Wang, Y., X. Zhang and Z. Guo (2021). "Estimation of tree height and aboveground biomass of coniferous forests in North China using stereo ZY-3, multispectral Sentinel-2, and DEM data." Ecological Indicators **126**: 107645.