# A Disentangled Variational Autoencoder for Prediction of Above Ground Biomass from Hyperspectral Data

**Conference Paper** · October 2021

DOI: 10.1109/IGARSS47720.2021.9554415



**3 authors**, including:

Parth Naik
Helmholtz-Zentrum Dresden-Rossendorf


Michele Dalponte
Fondazione Edmund Mach - Istituto Agrario San Michele All'Adige

# A DISENTANGLED VARIATIONAL AUTOENCODER FOR PREDICTION OF ABOVE GROUND BIOMASS FROM HYPERSPECTRAL DATA

[1,2]Parth Naik, [2]Michele Dalponte, [1]Lorenzo Bruzzone

[1]Department of Information Engineering and Computer Science, University of Trento, Italy.
[2]Sustainable Agro-ecosystems and Bioresources Department, Fondazione Edmund Mach, Italy.

IGARSS 2021 - 2021 IEEE International Geoscience and Remote Sensing Symposium | 978-1-6654-0369-6/21/$31.00 ©2021 IEEE | DOI: 10.1109/IGARSS47720.2021.9554415

## ABSTRACT

The prediction of forest biophysical parameters is an important task in remote sensing for understanding global carbon cycle. Spectral remote sensing data are available globally at a relatively economical cost making them a viable resource for forest remote sensing. However, the main drawbacks associated with such data is the uncertainty of predictions and cluttered process of selecting band combinations from hyperspectral/multispectral data to produce spectral features for modelling. In this paper, we present an approach that exploits the latest developments in generative variational autoencoders (VAE) that produce disentangled representation from input data to assess the capability of hyperspectral data to model forest aboveground biomass (AGB). The proposed VAE generates a special kind of deep spectral features that are proportional to AGB. A modelling accuracy of $R^2 = 0.57$ (cross-validated) was obtained by the proposed approach, thus pointing out the potential of hyperspectral data to model AGB using disentangled deep spectral features. The proposed approach also enables in bypassing the unreliable process of selecting band combinations to produce spectral features and shows good prospects for mapping global level biomass.

***Index Terms*****—** variational autoencoders, disentangled representation, hyperspectral data, deep features, AGB.

## 1. INTRODUCTION

The process of feature extraction is instrumental for most applications in remote sensing in order to accurately classify data or precisely estimate a target parameter. There are numerous methods developed for manual and automatic extraction of features from remote sensing data using pixel or object based approaches [1]–[3]. The low or high level feature extraction approaches based on pixel or object based methods have been a subject of extensive research over several decades in the field of remote sensing. In recent times, deep learning based approaches have been found effective in characterizing complex distributions underlying remote sensing data. Deep learning methods extract high level vision information that relays abstract features and representations [4]–[6]. The integral process of feature extraction using deep convolutional neural networks trained on remote sensing datasets have delivered improved performance for various applications [7], [8]. Numerous studies have been performed to develop trained models for remote sensing image classification, denoising and change detection [9] however, their role for supervised regression has remained less explored.

Variational Autoencoders (VAE) are deep neural models that assume proportional dependence of training samples over a latent representation generated by an encoder unit and sampled from a Gaussian distribution by a decoder unit. The VAE has been used for various applications in the field of remote sensing that includes classification, object and anomaly detection [10]–[12]. However, the recent developments in the VAE architecture enables learning disentangled latent representations that are specifically proportional to the target variable for regression tasks [13], [14]. The advanced VAE framework integrates the dissociated processes of feature extraction and regression using a generative conditional distribution of latent space on the target variable instead of a single latent Gaussian that captures all the variance of the data. In short, the advanced generative VAE model leads to a target-specific distribution of latent representations that are optimal for regression tasks.

Hyperspectral remote sensing data is very widely used for forest resource management practices around the world [15], [16]. The primary approaches for biophysical parameter estimation attempts to relate spectral signature of hyperspectral data with actual parameter on ground using statistical modelling [17]–[20]. Vegetation indices are most popularly used spectral features but their dependence on implicit assumptions is a serious drawback of such features for modelling biophysical variables [21]. Such spectral features have been observed to underexploit the full potential of the hyperspectral imagining with tens to hundreds of contiguous narrow bands. Moreover, it is difficult to optimally define a band combination that are specifically proportional to target biophysical parameters and often fail to represent some physiological processes [22]. The latest developments in disentangled representation using VAE's shows huge potential to resolve this problem in remote sensing of forests by generating features from hyperspectral remote sensing data that are specific to the target forest parameter.

The main objective of this paper is to examine the distribution of disentangled latent space of a generative VAE and assess its capability to model forest aboveground biomass (AGB) from hyperspectral data. This paper utilizes advances in generative Variational Autoencoder architectures to produce disentangled representation as well as 3D element-wise operations on hyperspectral data for spatio-spectral feature learning.

## 2. DATASET DESCRIPTION

The spectral dataset used for the study are airborne hyperspectral data with a spatial resolution of 1.5 m and 96 spectral bands equally spaced between 354 nm to 1034 nm. Additionally, a high density multiple return airborne lidar data (10 points/m$^2$) are used with 41 field AGB plot samples of 706.5 m$^2$ area for generating training labels (plot level AGB) for 1200 random plots of approximately equal area as field plots. The 1200 training labels are selected to have an approximate normal distribution in the range of 36 Mg/ha to 784 Mg/ha.

## 3. METHODOLODY

### 3.1. Proposed VAE model

The proposed VAE model adopts the framework of disentangled VAE proposed in [13], with modified network parameters that enable operations with hyperspectral remote sensing data. A traditional VAE infers a distribution of data using a variational procedure with a probabilistic encoder *'q(r|i)'*, where *'i'* represents input training data and *'r'* represents the latent representation generated using a Gaussian prior. The data are generated by sampling *'r'* by a generational network *'p(i|r)'*.

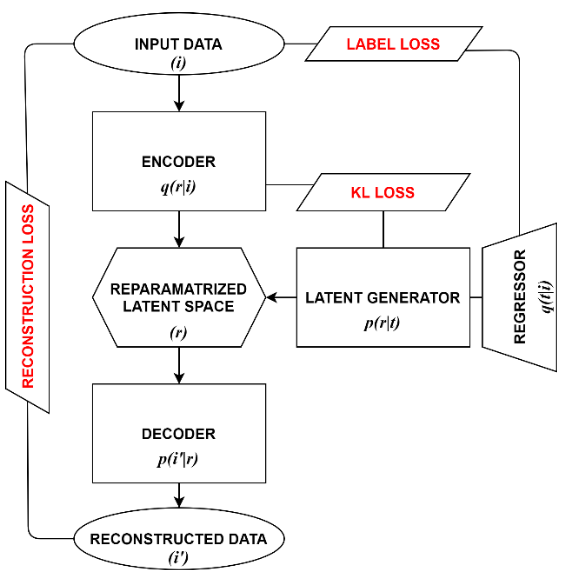


Fig. 1. Illustration of the proposed VAE model

In the proposed VAE model shown in Fig. 1, we explicitly frame the conditional distribution of latent representation *'r'* on the target parameter *'t'* (AGB) instead of a single Gaussian prior in addition to compressing the data dimensionality. The latent generator *'p(r|t)'* captures the AGB specific prior on latent representations. The non-linearity of the latent generator can be captured by the decoder network *'p(i|r)'* such that it essentially constructs a disentangled dimension associated with *'t'*. *'q(t|i)'* is a regular feed-forward regression network which delivers prediction and prediction uncertainty (standard deviation) as outputs which makes it a probabilistic regressor. The total loss of the proposed VAE network represented in equation (1), which is modified to include loss from latent generator and a probabilistic regressor.

$$\begin{aligned}\mathcal{L}(i) = & -D_{KL}\big(q(t|i)\,||\,p(t)\big) \\ & + \mathbb{E}_{q(r|i)}\left[\log\ p(i|r)\right] \\ & - \mathbb{E}_{q(t|i)}\Big[D_{KL}\big(q(r|i)\,||\,p(r|t)\big)\Big]\end{aligned} \tag{1}$$

In the above equation, the first term represents the KL-divergence *($D_{KL}$)* that regularizes prediction of *'t'*. The second term aligns the decoded data from latent representation to resemble to the input. The third term (1) regularizes network parameters so that the posterior *q(r|i)* resembles the AGB-specific prior *p(r|t)*. This type of inter-linked loss mechanism regulates latent representations and AGB such that the generated latent representation for the predicted *'t'* has to resemble the latent representations from input *'i'*. The encoder network with the regressor form an 'inference network' and the latent generator along with decoder form a 'generative network'. The total loss of the VAE network is the sum of the reconstruction loss, the KL loss and the label loss. This framework allows the exchange of low level features learned by the network layers of the autoencoder and regressor.

### 3.2. Implementation Framework

The methodology flowchart for implementation of proposed model is shown in Fig. 2.

Deep learning models are excessively data dependent. Therefore, we integrate a systematic data pipeline for the VAE network in order to feed input data. Deep models require a huge amount of training data to learn their parameters. Due to the limited number of field AGB samples, we used a primary state-of-the-art lidar data model developed in [23] (point density = 10 points/m$^2$, $R^2$ = 0.8) to generate training labels for the network. As the main objective of the paper is to assess the capability of the proposed model to learn from spectral data and the primary lidar model is robust and dependable, it is reasonable to assume that the error propagated from lidar model in generating training labels is negligible. Therefore, the generated 1200 AGB plots could be used as training labels for the experiment. For the same 1200 plots, we extracted hyperspectral data cubes (18x18 pixels) and performed continuum removal as a part of the pre-processing of the data. The 1200 training labels were selected to have a Gaussian distribution and min-max scaling was performed. These pre-processed hyperspectral data cubes and the AGB plots were converted to numpy arrays and fed to the network.

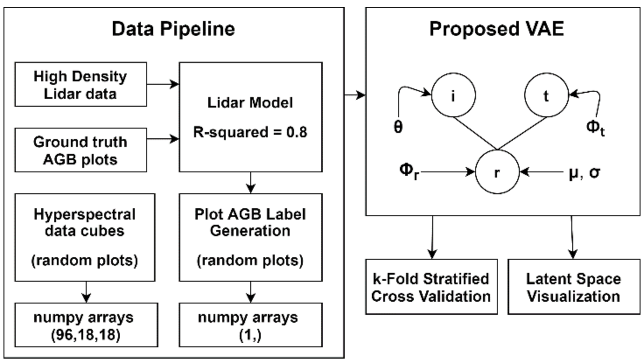


Fig. 2. Methodology Flowchart
θ=generative parameters, $\varphi_t$=inference parameters, $\varphi_r$=latent parameters, i=training data, r=latent space, t=target AGB

The encoder network consists two 3D convolution layers followed by 2 Dense layers to produce a latent space of size 16 from input array of size (96,18,18). The regressor had 2 dense layers and shared convolution layers of encoder. The decoder decoded the input from the latent space with an inverse structure of the encoder. L2 regularization was applied to all dense layers due to large number of model parameters.

The entire network was trained with a batch size of 64 and 150 epochs and a stratified k-fold cross-validation with 5 folds was applied for train-test data splitting. The $R^2$ (cross-validated) and the RMSE of the predictions were considered for testing accuracy of predictions. The latent space was transformed to a 2D space using t-distributed stochastic neighbor embedding (TSNE) and color-coded by target AGB to visualize the latent representation.

## 4. RESULTS

The AGB prediction results for the proposed VAE model along with the cross-validated $R^2$ and RMSE are shown in Fig 3. The results of TSNE visualization of the latent space for a traditional VAE and the proposed VAE are shown in Fig 4.

The modelling accuracy $R^2$ (cross-validated) of 0.57 and RMSE of 95.64 Mg/Ha was achieved by the proposed VAE model. A prediction bias was observed particularly for lowest and highest AGB values which can be attributed to lack of training samples at extremities. However, the results were reasonably accurate considering the size of training data used train the model parameters associated with network. The use of cross validation shrinks the chances of overfitting on the data and the proposed modelling approach presents a promising and robust basis for forest biophysical parameter estimation based on spectral data. Although the modelling accuracy does not surpass the state-of-the-art, the wide availability of spectral remote sensing data increases the utility of the proposed approach.

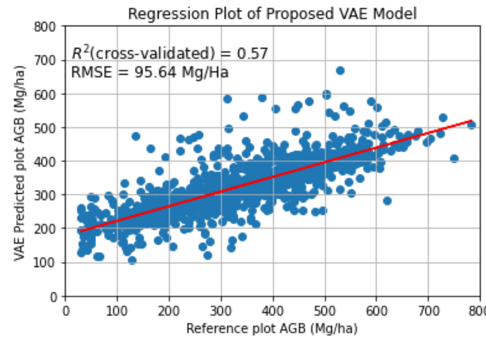


Fig. 3. Reference AGB vs VAE predicted AGB

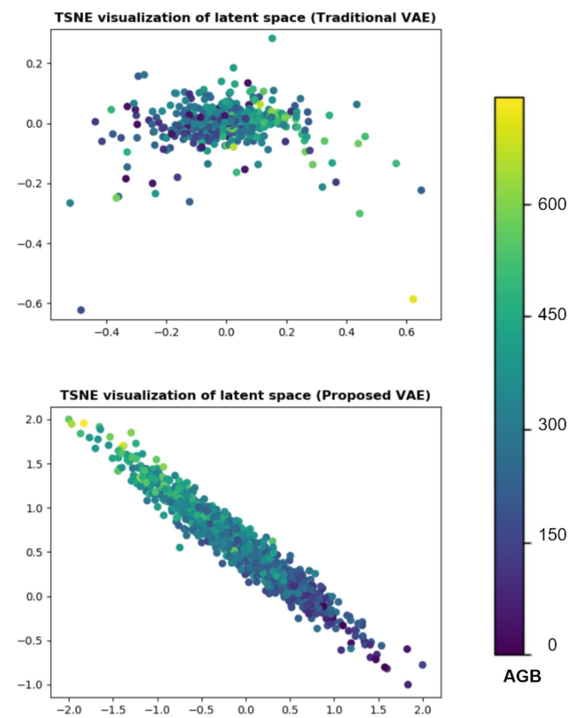


Fig. 4. TNSE visualization of latent space for a traditional VAE(top) and proposed VAE(bottom)

The TSNE visualization of latent space of a traditional VAE and proposed VAE clearly shows the level of disentanglement achieved in the latent representation of the proposed approach. The disentangled representation ensures unique correspondence between the range of representation and the range of the target variable i.e. AGB. This is very essential especially with spectral data, the prime limitation of which is a high uncertainty in correlation of traditional spectral features with the forest biophysical parameters. The results show that one direction of variation is associated with

AGB whereas the latent representation produced by an unsupervised traditional VAE does not achieve such variation and therefore no clear disentangled representations.

## 5. DISCUSSION AND CONCLUSION

In this paper, we used a variational autoencoder that generates disentangled representations from hyperspectral remote sensing data to predict forest biomass. The proposed approach achieved a reasonable modelling accuracy and more importantly the generative nature of the model enabled the variation of latent representations as a factor of target parameter i.e. AGB. The study does not surpass the modelling accuracy of the state-of-the-art lidar based models however presents a successful approach to diminish the uncertainties in prediction of biophysical parameters for spectral remote sensing data. This study provides a sophisticated and effective replacement for vegetation indices as features for modelling AGB in form of disentangled latent representations. Moreover, an implementation of the proposed method on satellite hyperspectral or multispectral data with a global coverage shall provide a solution to global biomass estimation problem.

## 6. REFERENCES


[1] S. Mei, J. Ji, Y. Geng, Z. Zhang, X. Li, and Q. Du, "Unsupervised spatial-spectral feature learning by 3D convolutional autoencoder for hyperspectral classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 9, pp. 6808–6820, 2019.

[2] T. Lu *et al.*, "Set-to-Set Distance-Based Spectral – Spatial Classification of Hyperspectral Images," pp. 1–13, 2016.

[3] P. Deepa and K. Thilagavathi, "Feature extraction of hyperspectral image using principal component analysis and folded-principal component analysis," *2nd Int. Conf. Electron. Commun. Syst. ICECS 2015*, no. Icecs, pp. 656–660, 2015.

[4] B. Petrovska, E. Zdravevski, P. Lameski, R. Corizzo, I. Štajduhar, and J. Lerga, "Deep learning for feature extraction in remote sensing: A case-study of aerial scene classification," *Sensors (Switzerland)*, vol. 20, no. 14, pp. 1–22, 2020.

[5] S. Mei, J. Ji, J. Hou, X. Li, and Q. Du, "Learning Sensor-Specific Spatial-Spectral Features of Hyperspectral Images via Convolutional Neural Networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 8, pp. 4520–4533, 2017.

[6] W. Li, G. Wu, F. Zhang, and Q. Du, "Hyperspectral Image Classification Using Deep Pixel-Pair Features," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 2, pp. 844–853, 2017.

[7] W. Hu, Y. Huang, L. Wei, F. Zhang, and H. Li, "Deep convolutional neural networks for hyperspectral image classification," *J. Sensors*, vol. 2015, 2015.

[8] L. Deng and D. Yu, "Deep learning: Methods and applications," *Found. Trends Signal Process.*, vol. 7, no. 3–4, pp. 197–387, 2013.

[9] X. X. Zhu *et al.*, "Deep learning in remote sensing: a review," *arXiv*, no. December, 2017.

[10] X. Wang, K. Tan, Q. Du, Y. Chen, and P. Du, "CVA2E: A Conditional Variational Autoencoder with an Adversarial Training Process for Hyperspectral Imagery Classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 8, pp. 5676–5692, 2020.

[11] S. Sinha *et al.*, "Detecting Avalanche Deposits using Variational Autoencoder on Sentinel-1 Satellite Imagery To cite this version : HAL Id : hal-02318407 Detecting Avalanche Deposits using Variational Autoencoder on Sentinel-1 Satellite Imagery," 2019.

[12] A. Detection, "LSTM-Based VAE-GAN for Time-Series Anomaly Detection," 2020.

[13] Q. Zhao, E. Adeli, N. Honnorat, T. Leng, and K. M. Pohl, "Variational AutoEncoder for Regression: Application to Brain Aging Analysis," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 11765 LNCS, pp. 823–831, 2019.

[14] Y. Yoo, S. Yun, H. J. Chang, Y. Demiris, and J. Y. Choi, "Variational autoencoded regression: High dimensional regression of visual data on complex manifold," *Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017*, vol. 2017-Janua, pp. 2943–2952, 2017.

[15] G. le Maire *et al.*, "Calibration and validation of hyperspectral indices for the estimation of broadleaved forest leaf chlorophyll content, leaf mass per area, leaf area index and leaf canopy biomass," *Remote Sens. Environ.*, vol. 112, no. 10, pp. 3846–3864, 2008.

[16] S. Huber, M. Kneubühler, A. Psomas, K. Itten, and N. E. Zimmermann, "Estimating foliar biochemistry from hyperspectral data in mixed forest canopy," *For. Ecol. Manage.*, vol. 256, no. 3, pp. 491–501, 2008.

[17] P. Naik, M. Dalponte, and L. Bruzzone, "A comparison on the use of different satellite multispectral data for the prediction of aboveground biomass," vol. 1153315, no. September 2020, p. 40, 2020.

[18] E. A. Addink, S. M. De Jong, and E. J. Pebesma, "The importance of scale in object-based mapping of vegetation parameters with hyperspectral imagery," *Photogramm. Eng. Remote Sensing*, vol. 73, no. 8, pp. 905–912, 2007.

[19] J. P. Rivera-Caicedo, J. Verrelst, J. Muñoz-Marí, G. Camps-Valls, and J. Moreno, "Hyperspectral dimensionality reduction for biophysical variable statistical retrieval," *ISPRS J. Photogramm. Remote Sens.*, vol. 132, pp. 88–101, 2017.

[20] E. Adam, O. Mutanga, E. M. Abdel-Rahman, and R. Ismail, "Estimating standing biomass in papyrus (Cyperus papyrus L.) swamp: Exploratory of in situ hyperspectral indices and random forest regression," *Int. J. Remote Sens.*, vol. 35, no. 2, pp. 693–714, 2014.

[21] R. Lasse, *Dissertations in Forestry and Natural Sciences*, no. March. 2016.

[22] J. Verrelst *et al.*, "Optical remote sensing and the retrieval of terrestrial vegetation bio-geophysical properties - A review," *ISPRS J. Photogramm. Remote Sens.*, vol. 108, pp. 273–290, 2015.

[23] S. Tonolli, M. Dalponte, L. Vescovo, M. Rodeghiero, L. Bruzzone, and D. Gianelle, "Mapping and modeling forest tree volume using forest inventory and airborne laser scanning," *Eur. J. For. Res.*, vol. 130, no. 4, pp. 569–577, 2011.