Land Cover Quantification using Autoencoder based Unsupervised Deep Learning

Sandhya Manjunatha Bharadwaj

Thesis submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Master of Science
in
Computer Engineering

A. Lynn Abbott, Chair
Brian Lattimer, Co-Chair
Ryan K. Williams

August 4, 2020
Blacksburg, Virginia

Keywords: Deep Learning, Autoencoder, Land Cover, Hyperspectral Imagery, Spectral Unmixing, Reflectance Spectra

Land Cover Quantification using Autoencoder based Unsupervised Deep Learning

Sandhya Manjunatha Bharadwaj

(ABSTRACT)

This work aims to develop a deep learning model for land cover quantification through hyperspectral unmixing using an unsupervised autoencoder. Land cover identification and classification is instrumental in urban planning, environmental monitoring and land management. With the technological advancements in remote sensing, hyperspectral imagery which captures high resolution images of the earth's surface across hundreds of wavelength bands, is becoming increasingly popular. The high spectral information in these images can be analyzed to identify the various target materials present in the image scene based on their unique reflectance patterns. An autoencoder is a deep learning model that can perform spectral unmixing by decomposing the complex image spectra into its constituent materials and estimating their abundance compositions. The advantage of using this technique for land cover quantification is that it is completely unsupervised and eliminates the need for labelled data which generally requires years of field survey and formulation of detailed maps. We evaluate the performance of the autoencoder on various synthetic and real hyperspectral images consisting of different land covers using similarity metrics and abundance maps. The scalability of the technique with respect to landscapes is assessed by evaluating its performance on hyperspectral images spanning across 100m x 100m, 200m x 200m, 1000m x 1000m, 4000m x 4000m and 5000m x 5000m regions. Finally, we analyze the performance of this technique by comparing it to several supervised learning methods like Support Vector Machine (SVM), Random Forest (RF) and multilayer perceptron using F1-score, Precision

and Recall metrics and other unsupervised techniques like K-Means, N-Findr, and VCA using cosine similarity, mean square error and estimated abundances. The land cover classification obtained using this technique is compared to the existing United States National Land Cover Database (NLCD) classification standard.

Land Cover Quantification using Autoencoder based Unsupervised Deep Learning

Sandhya Manjunatha Bharadwaj

(GENERAL AUDIENCE ABSTRACT)

This work aims to develop an automated deep learning model for identifying and estimating the composition of the different land covers in a region using hyperspectral remote sensing imagery. With the technological advancements in remote sensing, hyperspectral imagery which captures high resolution images of the earth's surface across hundreds of wavelength bands, is becoming increasingly popular. As every surface has a unique reflectance pattern, the high spectral information contained in these images can be analyzed to identify the various target materials present in the image scene. An autoencoder is a deep learning model that can perform spectral unmixing by decomposing the complex image spectra into its constituent materials and estimate their percent compositions. The advantage of this method in land cover quantification is that it is an unsupervised technique which does not require labelled data which generally requires years of field survey and formulation of detailed maps. The performance of this technique is evaluated on various synthetic and real hyperspectral datasets consisting of different land covers. We assess the scalability of the model by evaluating its performance on images of different sizes spanning over a few hundred square meters to thousands of square meters. Finally, we compare the performance of the autoencoder based approach with other supervised and unsupervised deep learning techniques and with the current land cover classification standard.

Dedication

To my parents, Vijaya and Manjunath Bharadwaj, and my sister, Shreya Bharadwaj

Acknowledgments

I would like to thank my advisor, Dr. Brian Lattimer, for his continued guidance and encouragement throughout this research. This work would not have been possible without his support, mentorship and invaluable suggestions. I would like to thank Dr. Lynn Abbott and Dr. Ryan. K. Williams for serving on my committee and for all their support and valuable inputs.

I am grateful to the Extreme Environments and Materials Lab, Virginia Tech, for providing the laboratory infrastructure and computing resources and also for supporting me with graduate research assistantship. I am thankful to the Bradley Department of Electrical and Computer Engineering, Virginia Tech, for giving me an opportunity to pursue my Master's degree and for funding my study by providing teaching assistantship. I would like to thank the Advanced Research Computing at Virginia Tech for providing computational resources that were required for this work.

Finally, I would like to thank my parents and my sister for always believing in me and supporting my dreams.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Land covers are the physical materials such as trees, roads, lakes, wetlands, buildings and croplands that are present on the surface of the earth. The identification of different land covers on the earth's surface and generation of accurate land cover maps is of high ecological importance in environment monitoring and urban planning. The recent advancements in remote sensing have led to the development of land cover classification using high resolution hyperspectral images.

Hyperspectral Imaging (HSI) is an imaging technique which captures and processes images across a large number of wavelengths. This technique is becoming increasingly popular in modern remote sensing where hyperspectral images are used used to locate and identify the different objects present in the given scene. While traditional RGB images are captured in three wavelength bands and multispectral images are captured in up to twenty wavelength bands, the hyperspectral images are captured across hundreds of narrow wavelength bands conveying more spectral information.

HSI is gaining momentum in the areas of environmental monitoring, urban planning, military surveillance, agriculture, mining, astronomy and biomedical domains. The rapid developments in technology have to led to equipping earth observing satellites and aircrafts with high quality cameras and sensors which generate large volumes of hyperspectral data with high resolutions. Application of deep learning techniques to traditional remote sensing areas is on the rise due to the availability of large amounts of data.

1.1 Background

1.1.1 Reflectance Spectra

Hyperspectral imaging is a technique that captures images using multiple bands across the electromagnetic spectrum. Figure 1.1 shows an example of hyperspectral imaging where a hyperspectral camera is used to capture an image of a target leaf and the reflectance data is the reflectance pattern of the leaf.

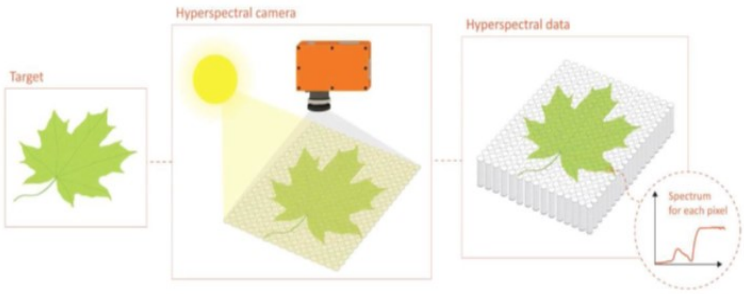

Figure 1.1: Hyperspectral Imaging [27]

The amount of reflectance of the incident electromagnetic wave from a surface as a function of the wavelength is known as *spectral reflectance* or *reflectance spectrum*. Different materials reflect and absorb the incident spectrum by different amounts and hence their reflectance patterns are different. Figure 1.2 illustrates the reflectance spectra of three materials, namely, water, soil and green vegetation and it can be observed that different spectral patterns are obtained for each of the materials.

Every surface reflects the incident electromagnetic spectrum with a pattern that is unique to the surface. Therefore, the reflectance spectra is also known as *reflectance signature*. These signatures can be analyzed to identify the reflecting material and learn about its structural and chemical properties. The different land covers present in a hyperspectral image can be identified by analyzing the shape and magnitude of the reflectance spectra.

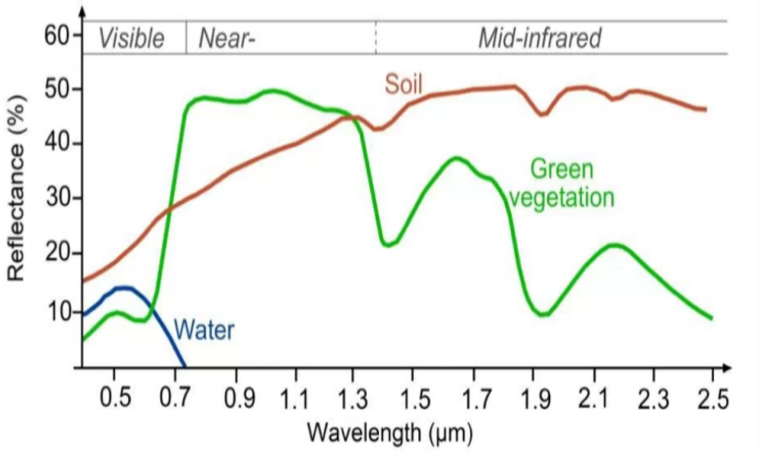

Figure 1.2: Reflectance Spectra [28]

1.1.2 Hyperspectral Imaging

The hyperspectral reflectance patterns typically span from 350nm to 2500nm wavelength of the electromagnetic spectrum across the ultraviolet, visible, near infrared and shortwave infrared regions. Airborne remote sensing systems like satellites and low flying air-crafts are used to capture hyperspectral images as shown in Figure 1.3 (a).

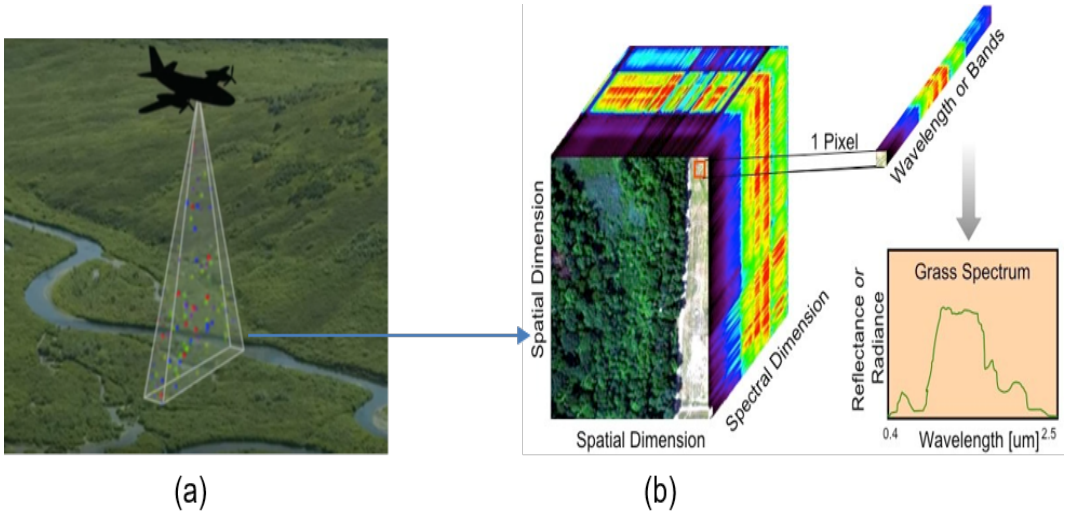

Figure 1.3: Hyperspectral Image Cube [21, 38]

The hyperspectral image data that is captured after imaging is shown in Figure 1.3 (b). Each pixel in a HSI is a high-dimensional vector corresponding to the spectral reflectance across different wavelengths resulting in a three-dimensional data cube of a two-dimensional spatial scene known as *hyperspectral image cube*. Every pixel in the image is the reflectance spectrum of the corresponding region in the image scene.

1.1.3 Spectral Unmixing

Each pixel in a hyperspectral image is usually a mixture of the reflectance spectra of more than one distinct material in the scene. This is because the electromagnetic spectrum is scattered by multiple materials in the scene and the observed reflectance pattern at the imaging sensor is a combination of the reflected spectral patterns of all the materials. Such an image pixel that is a combination of the reflectances of more than one material is called a *mixed pixel*.

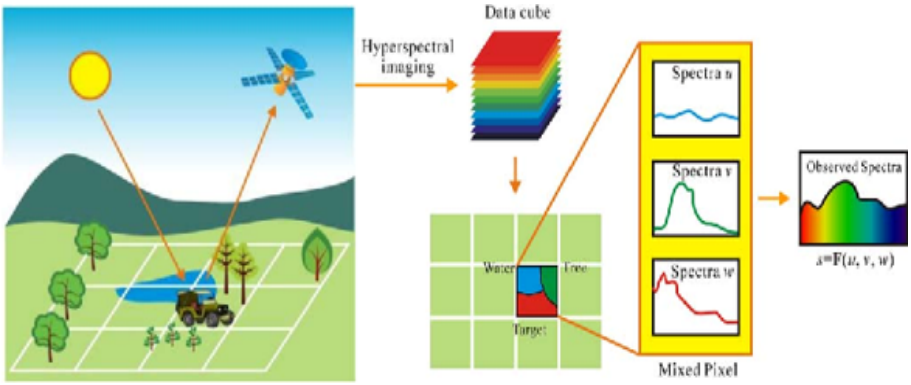

Figure 1.4: Hyperspectral Unmixing [37]

Figure 1.4 illustrates the formation of a mixed pixel consisting of water, tree and target materials. The observed spectra at the hyperspectral image sensor is a combination of the reflectances of the three materials.

The process of decomposing the mixed pixel into its constituent, pure spectra is known as

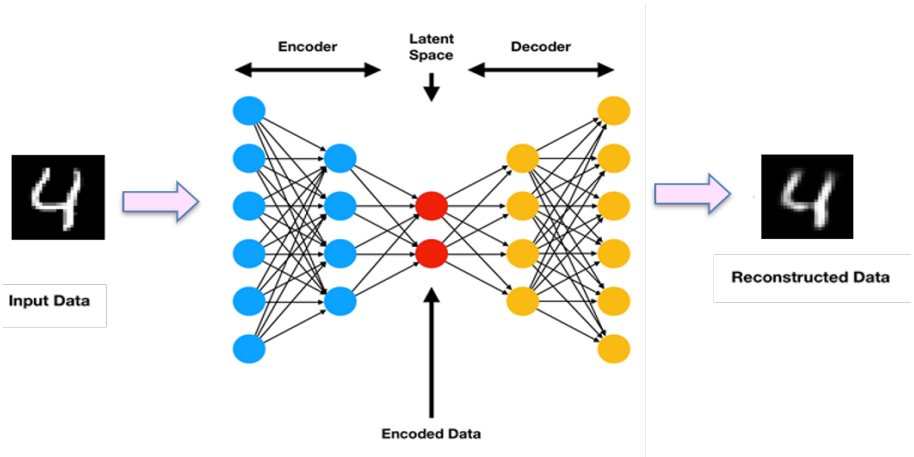

Figure 1.5: Working of an Autoencoder [6]

spectral unmixing. Each of the individual constituent spectra are called as *end members*. The process of predicting the percentage composition of each of the end member materials present in the hyperspectral image is known as *abundance estimation*. A land cover map representing the distribution of the different end members in the image is called as *abundance map*.

1.1.4 Autoencoder

An *Autoencoder* is an artificial neural network that compresses a high dimension input data into a lower dimension latent representation. The working of a basic autoencoder for compressing a MNIST image data into a lower dimension is illustrated in Figure 1.5. It is an unsupervised deep learning technique that does not require labelled data for training. The autoencoder tries to reconstruct the same data that is provided at the input layer as the output. The network consists of 2 parts, namely, encoder and decoder. The encoder compresses the input data into a reduced dimension known as the *latent-space representation* or the *encoded data*. The decoder is the second part of the network that tries to reconstruct the original data back from this latent representation based on a loss function.

1.2 Motivation

A land cover is defined as the observed biophysical cover on the earth's surface [10]. A land cover map is a spatial representation of the extent of the different types of land covers like forests, lakes, crops, roads, etc., present in a given region. Detailed land cover and vegetation maps are essential for development of improved fire spread models, land management and vegetation treatment modules. Accurately identifying the different land covers and the types of trees present in a given region requires extensive ground based field surveys which are expensive, time consuming and require a lot of manual labor. Further, natural disasters like forest fires and soil erosion often destroy forests and living habitats which in turn affect the land cover at that region. Therefore, these studies are not a one time process and the data needs to be updated regularly. Thus, there is a great need to develop fast, automated models which can easily identify and update the different types of land cover present in a region.

Owing to the growing demand of hyperspectral imaging, several independent studies are being carried out by researchers to document the pure spectra of materials. Several government and academic agencies like USGS Spectral Library [33] and EcoSIS [34] are providing open source collections of pure component spectra of various vegetation, land covers, minerals, organic compounds and other materials. Further, large amounts of multispectral and hyperspectral reflectance data captured by remote sensing satellites like Landsat and aircrafts with high resolution instruments like Airborne Visible InfraRed Imaging Spectrometer (AVIRIS) are publicly available. In spite of the abundant availability of data, several recent studies related to land cover quantification using hyperspectral imagery have performed independent, resource intensive field surveys and used their own drones and low-flying aircrafts to obtain hyperspectral reflectance data in the regions of their study. We noticed a gap in utilizing the already available spectral information from the spectral libraries and

the high resolution reflectance data from remote sensing imagery in land cover identification tasks. This thesis aims at developing deep learning models for automatically identifying the various land covers and vegetation present in a given region using hyperspectral reflectance data in an unsupervised manner which does not require field surveys.

1.3 Research Objectives

The main research objectives of this work are as follows:

1. To develop an unsupervised deep learning model for land cover identification and quantification through hyperspectral unmixing of reflectance data using an autoencoder

2. To qualitatively and quantitatively evaluate the performance of the developed model on synthetic and real data at various land scales and with different land cover and vegetation components

3. To assess the performance of the autoencoder model by comparison with other supervised and unsupervised deep learning techniques and existing land cover classification standard

Chapter 2

Literature Review

2.1 Literature Review

In this section, we look at some of the works related to the application of machine and deep learning techniques in identification of land covers using hyperspectral reflectance. The hyperspectral images contain high resolution spectral information which increases the accuracy of detecting target materials in classification tasks [7, 9].

Hyperspectral image analysis tasks can be grouped into two main categories: Classification and Spectral Unmixing [9]. Classification is the technique of identifying the material present at each pixel in a hyperspectral image. The aim of hyperspectral classification is to develop a map representing the distribution of different materials in the image. Land cover mapping [5, 12], plant species identification [7, 14, 30, 39] and mineral identification [26] are some of the applications of this technique. Recognizing the significance of vegetation identification on ecology, a data science challenge was conducted by National Institute of Standards and Technology (NIST) to identify plant species from airborne remote sensing data [14]. The authors of the work in [7] have developed an automated classifier using Convolutional Neural Network (CNN) for identifying seven dominant tree species in a mixed-conifer forest in California from hyperspectral images. This work further demonstrates that the use of high resolution hyperspectral imagery results in higher classification accuracy than RGB imagery.

The second category of HSI analysis is called spectral unmixing. This process involves identifying the proportions of different materials at each pixel of the hyperspectral image. It is very likely that each pixel in a hyperspectral image covers more than one material in the scene , that is, the spectrum measured at a pixel could be a combination of the reflectances of different materials. Such a pixel is called a *mixed pixel* and the process of recovering the fractions of the different materials at every pixel in the image is called spectral unmixing [9].

Hyperspectral unmixing using autoencoder is becoming increasingly popular due to the advancements in deep learning and the availability of large volumes of data. An autoencoder is an unsupervised deep learning technique that compresses the input data into a smaller dimension in latent space which achieves spectral unmixing. A denoising autoencoder with sparsity is implemented in [24] for unmixing synthetic and real datasets. A novel endmember extraction method using a two-staged autoencoder is presented in [22]. The authors of [29] propose a deep network with stacked autoencoders and a variational autoencoder for unmixing several land cover hyperspectral datasets. The advantage of using autoencoders is that it is an unsupervised technique and does not require labelled data for training. Classification tasks performed using supervised learning techniques such as support vector machine [16], artificial neural networks [30], convolutional neural networks [7, 8] require large amounts of labelled spectral data for every category of material present in a given region. This has led to the popularity of unsupervised feature learning in HSI tasks.

Several recent works [11, 22, 23, 24, 29, 35] on land cover and vegetation classification have been evaluated on hyperspectral classification datasets such as University of Pavia and Washington DC Mall [31], and the studies on hyperspectral unmixing have been performed on Urban, Samson and Jasper Ridge datasets [31] and there have not been many studies on any other real world datasets. This is because the ground truth formulation for identifying land covers and the field survey for labelling the tree species is an extremely complex task

and the ground truths for the above mentioned datasets are publicly available. Although having a common dataset is necessary for benchmarking and comparing the different techniques, most works have limited their study to these datasets and have not explored any other hyperspectral reflectance datasets. Further, these standard hyperspectral datasets are extremely small. The Urban dataset consists of 307 x 307 image pixels corresponding to a 4 square meter region and 210 wavelength bands, whereas the Jasper Ridge consists of 100 x 100 pixels and 224 wavelength channels with a spectral resolution of 9.46nm, the University of Pavia land cover data is an image of size of 610 x 610 pixels with 1.3m spatial resolution and 103 spectral bands. Due to lack of availability of labelled data for evaluation, the land cover and vegetation classification has been limited to very small spatial areas and the scalability of these approaches to larger land regions has not been explored in the previous works.

In our study, we aim to develop an autoencoder based hyperspectral unmixing method to obtain fine scale land cover and vegetation classification over larger regions spanning several square miles.

2.2 Research Contributions

This section presents the contributions of this research work.

1. This work demonstrates the performance of the autoencoder based unmixing on a novel, previously unexplored hyperspectral reflectance dataset for land cover quantification.

2. This research demonstrates the scalability of the autoencoder based hyperspectral unmixing technique for land cover quantification across large regions spanning several

square miles.

3. This work shows how to adapt the autoencoder architecture to various datasets of different sizes to perform land cover quantification through hyperspectral unmixing.

4. This research work uses open source hyperspectral reflectance data and spectral information databases for training the deep learning networks without performing any independent ground based field surveys.

Chapter 3

Methodology

This section describes the proposed approach, data sources and the implementation of the deep learning based land cover quantification.

3.1 Approach

The proposed approach for our deep learning based land cover quantification is illustrated in Figure 3.1. It involves two main steps: Spectral Unmixing and Spectral Identification.
The hyperspectral image cube is provided as the input to the autoencoder which is a deep learning model. Every pixel in the image is a high dimensional vector consisting of the mixed reflectances of multiple materials present in the image scene. The image cube is unfolded and the pixel wise input spectra are provided as the input to the autoencoder. Each input pixel from the hyperspectral image becomes a training data point. The model decomposes the complex input spectra into a set of pure spectra corresponding to the different materials in the image scene. This process is called *spectral unmixing* and the pure component spectra are known as *end members*.
Each of the unmixed, pure spectra is compared to a collection of spectral signatures of various materials known as the *spectral library*, and the library member it most closely matches to is identified. This process of matching every unmixed spectral component to the library and identifying the target material it corresponds to is called as *spectral identification*.

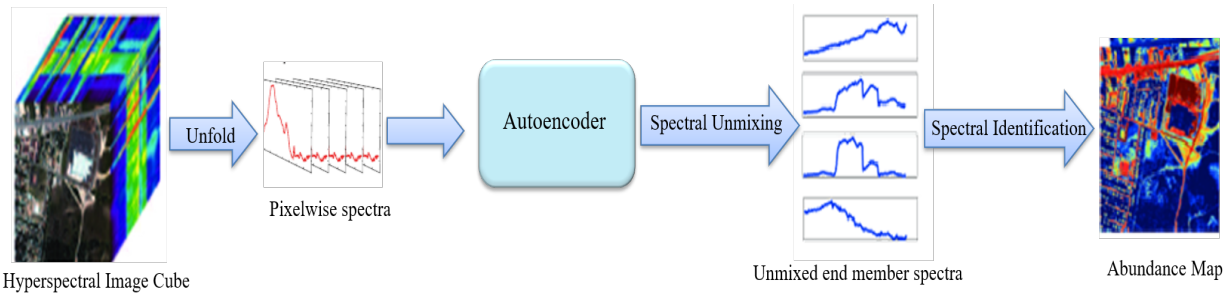

Figure 3.1: Approach

After the end members are identified, every input pixel in the image is classified into one of the land cover categories represented by the end members. The percent composition of each end member in the hyperspectral image is determined. This step is known as *abundance mapping* and the corresponding land cover map representing the distribution of the different end members in the image is known as the *abundance map*.

Thus, the deep learning model successfully identifies the different land covers present in the input hyperspectral image and further quantifies their percentage compositions in the image scene. The details of the working of the autoencoder are explained in the following sections.

3.1.1 Autoencoder for Spectral Unmixing

The autoencoder network takes the hyperspectral image spectra as the input, reduces it to a lower dimension and tries to reconstruct the reflectance patterns at the output. The mixed pixels present in the input image data get unmixed in the latent space and the hidden layer consists of the unmixed spectral components of the different end members present in the image scene. Traditionally, an autoencoder has been used for reducing the dimension of the input data where the hidden layer is the learned low-dimensional representation of the input. The unmixing process can also be explained as finding a low-dimensional representation. Autoencoders perform unmixing of the hyperspectral input image as demonstrated by the

works in [11, 29, 35]. Thus, we use an autoencoder for land cover quantification through hyperspectral unmixing of the reflectance data in this work.

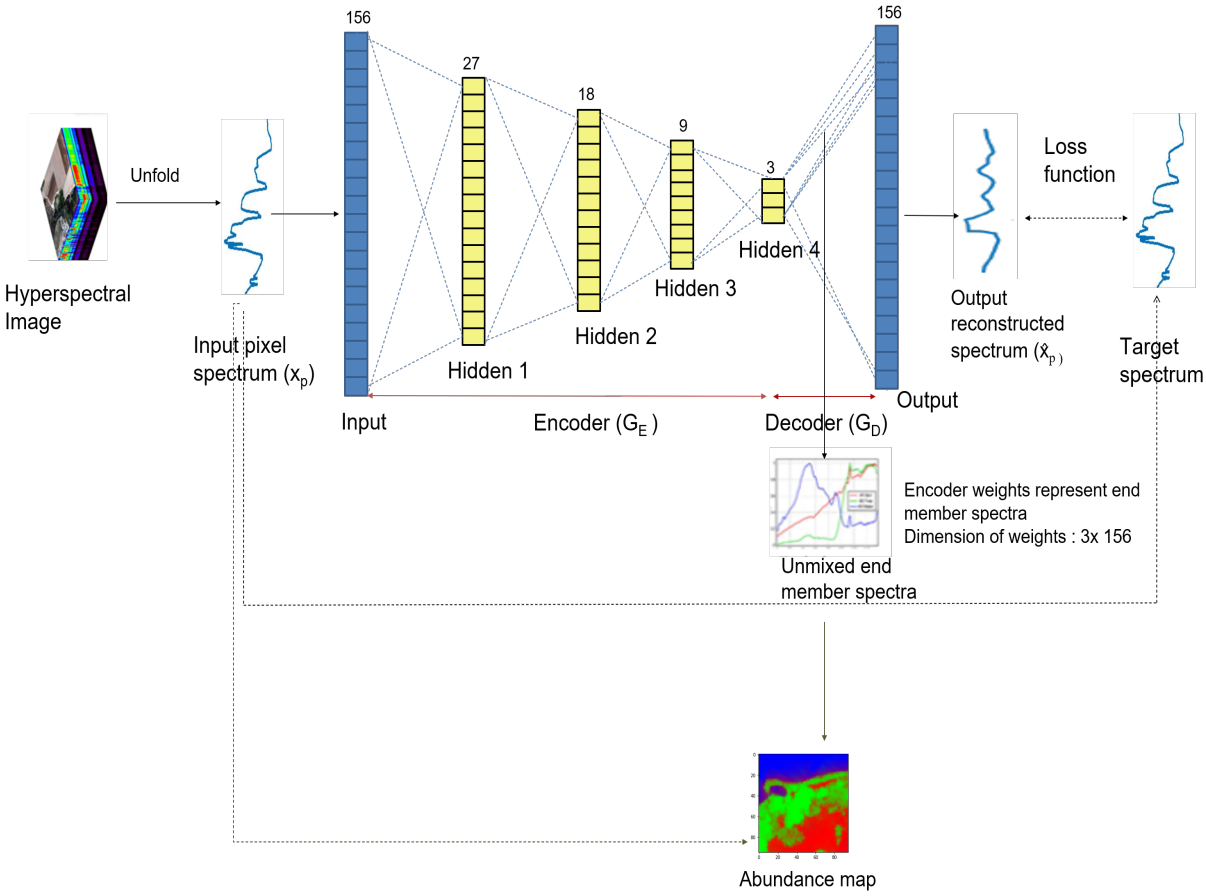

Figure 3.2: Spectral Unmixing using Autoencoder

Figure 3.2 illustrates hyperspectral spectral unmixing using an autoencoder. The hyperspectral image cube is a 3D image cube of a 2D spatial image scene. Each pixel is a high dimension vector representing the reflectance spectrum as a function of wavelength. The input hyperspectral image is unfolded and each input pixel is provided as an individual training data point to the autoencoder. The Samson hyperspectral dataset described in Section 3.2.5 is provided as the input in Figure 3.2. It is an image of size 95x95 pixels consisting of 156 wavelength bands forming an image cube of dimensions 95x95x156. Here each input pixel

corresponds to a reflectance spectrum of 156 dimensions. This image cube is reshaped into 9025x156 dimensions and the pixel-wise input data is fed into the autoencoder for training. The number of neurons in the input and output layer is equal to 156, which is equal to the number of reflectance wavelength bands. The autoencoder is asymmetric and it consists of a deep encoder network with continuously decreasing hidden units and a single layer in the decoder network. The encoder consists of 4 hidden layers with 27, 18, 9 and 3 neurons. All the layers in the network are fully connected layers. A Batch Normalization layer is added before the output layer. The last hidden layer, i.e. Hidden 4 in Figure 3.2, has the number of units equal the number of land cover types (end members) present in the input hyperspectral image.

The autoencoder model tries to reconstruct the spectrum of each input pixel at the output of the network using a loss function. At the end of the training process, the weights of the last hidden layer represent the unmixed end members. In Figure 3.2, the weights connecting Hidden Layer 4 to the output layer has a dimension of 3x156 which can be interpreted as three 156 dimension vectors. Each of these 156 dimension vectors represents the reflectance signature of one of the three end members present in the input hyperspectral image, namely, soil, tree and water. These vectors are known as the unmixed end member spectra.

After identifying all the end members, each input pixel present in the hyperspectral image is classified into one of the end member classes. This is done by calculating the cosine similarity between the spectrum of the input pixel and the end member spectra and classifying it to the class of the end member that it most closely matches. All the input pixels are identified and the percentage compositions of each end member in the image is calculated. An abundance map indicating the distribution of the various land cover types is generated.

The autoencoder achieves unmixing of the hyperspectral data as follows.

The input hyperspectral image is assumed to be a linear mixture model [11, 23, 29]. If there is more than one material within the boundaries of a pixel in a linear mixture model, the observed spectral vector of the pixel will be a weighted sum of the same number of end members representing the materials along with the noise that comes from the imaging instruments [23].

X is the input hyperspectral image with x_p representing each input pixel. Assuming a linear mixture model, we get,

$$X = MS + N \tag{3.1}$$

where M is the end member matrix, S is the abundance matrix, and N is the noise.

The autoencoder is a feedforward neural network that tries to reproduce its input at the output. The encoder maps the input spectral vectors into a latent space. The decoder then uses these latent space vectors to reconstruct the input spectrum as closely as possible.

The encoder $G_E : \mathbb{R}^{\mathrm{Bx1}} \rightarrow \mathbb{R}^{\mathrm{Rx1}}$ encodes the input spectra x_p to a hidden representation h_p as follows:

$$G_E(x_p) = h_p \, \epsilon \, \mathbb{R}^{\mathrm{Rx1}} \tag{3.2}$$

where R is the number of end members, and B is the number of bands.

The encoder in Figure 3.2 consists of 4 hidden layers followed by a Batch normalization layer. The input and output layers have a linear activation function and the hidden layers have Leaky ReLU as the activation function. The bias of the output layer is set to zero.

Every layer performs the transformation

$$a^l = g(W^l a^{\text{l-1}}) \tag{3.3}$$

where g is the activation function of the layer l, a^l is the activation of the layer l, $a^{\text{l-1}}$ is the activation of the layer (l-1) which is the input to layer l, and W^l is the weight of layer l. Therefore,

$$G_E(x_p) = BN(g(W^4 g(W^3 g(W^2 g(W^1 a^0))))) \tag{3.4}$$

where W^l and a^l represent the weights and activation of each layer respectively, a_0 is the input x_p,and BN represents Batch Normalization.
$BN(g(W^4 g(W^3 g(W^2 g(W^1 a^0)))))$ is the latent space representation (h_p) of the input x_p.

The decoder $G_D : \mathbb{R}^{\text{Rx1}} \rightarrow \mathbb{R}^{\text{Bx1}}$ decodes the latent space representation h_p to an approximation of the input as follows:

$$G_D(h_p) = W^L a^{\text{L-1}} = \hat{x}_p \, \epsilon \, \mathbb{R}^{\text{Bx1}} \tag{3.5}$$

where $\hat{x}_p$ is the reconstruction of x_p, L is the output layer, W^L and $a^{\text{L-1}}$ are the weight and activation of the output layer respectively.
The autoencoder's reconstruction of the input after the training can now be written as:

$$G_D(h_p) = W^L(BN(g(W^4 g(W^3 g(W^2 g(W^1 a^0)))))) \tag{3.6}$$

Therefore,

$$G_D(G_E(x_p)) = W^L(BN(g(W^4 g(W^3 g(W^2 g(W^1 a^0)))))) = \hat{x}_p \tag{3.7}$$

where the weight matrix $W^L \epsilon \mathbb{R}^{\mathrm{BxR}}$ is the weight of the output layer that contains the end members in its columns. Each of the R end members has a dimension of $BX1$ and is represented by one of the columns of W_L. Each $BX1$ column vector of W_L provides the spectral value for a particular wavelength in the end member. For example, consider a case when the first column of W_L represents soil from Samson dataset, then this column has a dimension of 156x1 which corresponds to the reflectance signature of soil with every value in the 156x1 vector representing the reflectance value at a particular wavelength band.

Since the input is linearly mixed, each pixel that is reconstructed at the output has one end member spectrum as the major contributor while the other two end members contribute less to the spectral value. The contribution of each member to the output is determined by the value of the activation function for the last hidden layer.

The network is trained using backpropagation to minimize the loss function $L(x_p, \hat{x}_p)$ to reduce the discrepancy between the output and the input. Mean square error and cosine proximity loss functions are used in this work.

At the end of training, the network's weights connecting the last hidden layer to the output layer (W^L) are the end members, namely, soil, tree and water in the case of Samson dataset.

3.1.1.1 Autoencoder Architecture

The generic architecture of the autoencoder used in this work is presented in Table 3.1 which is inspired by the work in [23]. The number of units in the input and the output layers is equal to the number of wavelength bands in the hyperspectral input data which is represented as B. The value of B is 315 for synthetic data and NEON hyperspectral reflectance data and 156 for Samson hyperspectral data. The number of units in the last hidden layer, i.e., Layer

5 is equal to the number of end members in the input image (N).

Leaky ReLU is used as the activation function for the hidden layers as it was found to perform better than ReLU and Sigmoid activations. Linear activation function is used for the input and output layers. L2 regularization is used for the output layer as it prevents overfitting. The Batch Normalization Layer, i.e., Layer 6, normalizes the activations of the previous layer in each batch and is known to speed up learning in neural networks [23]. The loss function, batch size, number of hidden layers and the kernel size of each layer are modified slightly in the different sets of experiments based on the input data and the size of the training data which are described in their respective sections.

Table 3.1: Autoencoder architecture

Layer Number	Type	Activation Function	Number of units
1	Input	Linear	B
2	Hidden	Leaky ReLU	9*N
3	Hidden	Leaky ReLU	6*N
4	Hidden	Leaky ReLU	3*N
5	Hidden	Leaky ReLU	N
6	Batch Normalization	-	-
7	Output	Linear L2 Regularization Min Max Constraint	B

3.2 Datasets

The deep learning model requires the hyperspectral reflectance data as the input. Further, it requires ground truth abundance data for evaluating the performance of the model. This technique also requires a spectral library consisting of the pure spectra for identifying the different land covers. The various data sources used in this work are described in detail in the subsequent subsections.

3.2.1 Hyperspectral Reflectance Data

The hyperspectral reflectance data required for training the deep learning models is obtained from the National Ecological Observatory Network (NEON) Database [21]. NEON is a large scale ecological project started by the National Science Foundation to measure and quantify various ecological processes. There are 81 terrestrial and airborne sites located all over the United States to measure the various changes in the atmosphere, land cover, ecohydrology, biogeochemistry in order to analyze the changes in the ecosystem.

High resolution hyperspectral reflectance data is collected using lightweight aircrafts that fly at low altitudes of around 1,000 meters above the ground level. The hyperspectral reflectance is captured across 426 bands in the 380 to 2500 nm wavelenth range by the AVIRIS Next-Gen Imaging Spectrometer. Each wavelength band is narrow with a high spectral resolution of 5 nanometers and a spatial resolution of 1 meter. The surveys are carried out annually over the NEON sites to collect data during peak greenness, typically between May to October, to capture information that accurately quantify the ecological changes. This data is used for training the deep learning models to identify the different types of land cover present in the image scene.

3.2.2 Spectral Library

A spectral library is a database of spectral signatures of different materials. They contain the reflectance patterns of pure materials that are generally obtained using high resolution spectrometers in laboratory conditions and field surveys. These spectral libraries are used as references to identify the different land cover and vegetation types in our study. There has been a rise in the studies related to detecting and mapping materials through remote sensing and laboratory image spectroscopy. Several well-known government and academic research institutes have identified this potential and are striving to provide open source spectral libraries for scientific and research purposes.

The following are some of the major spectral libraries:

1. **USGS Spectral Library** [33]

   The United States Geological Survey (USGS) is an agency of the government of the United States that has several disciplines for the scientific studies related to the landscape, natural resources, geology and geography of the United States. The USGS Spectroscopy Lab measures the reflectances of several different materials using laboratory techniques such as X-ray diffraction and Electron Probe Micro Analysis and airborne field spectroscopy and remote sensing satellite systems such as AVIRIS, Hyperion and Landsat and provides this data as an open source database. The reflectance data spans across the ultraviolet, visible, near-infrared, mid-infrared, and far-infrared regions in the 0.2 to 200 micrometer wavelength range. This spectral library consists of thousands of spectra of a variety of soil mixtures, minerals, vegetation, coatings, liquids and organic compounds among other materials.

2. **ECOSTRESS Spectral Library** [2, 15]

   The ECOSTRESS Spectral library is released by NASA's Jet Propulsion Laboratory

managed by California Institute of Technology. This consists of data from three sources, namely, Johns Hopkins University, Jet Propulsion Laboratory and United States Geological Survey. It consists of hyperspectral reflectance spectra of over 3400 natural and man made materials collected using Advanced Spaceborne Thermal Emission Reflection Radiometer over the 0.35 - 15.4 micrometer wavelength range.

3. **EcoSIS Spectral Library** [34]

   Ecological Spectral Information System (EcoSIS) is an open spectral library published by NASA and maintained by University of Wisconsin-Madison which allows scientists to share and publish spectral data. The main aim of this project is to provide free and open access of ecological spectra to the scientific community. It consists of more than 70,000 different types of spectra of different vegetation, landscapes and ecosystem types.

These spectral libraries are vast and they contain reflectance signatures of thousands of different minerals, soils, vegetation, organic compounds, etc., some of which are beyond the scope of this study. Further, the sampling rates are different in different spectral libraries. Therefore, we developed our own custom spectral library in this work. We compiled the spectra of several major types of land cover like soil, sand, road, water, rooftop and various vegetation species from the above spectral libraries using an uniform sampling rate across all the sources and eliminated redundancies of some spectral components and used it as the ground truth for identifying the types of land cover unmixed by the autoencoder. Limiting the spectral library to only the expected spectral components reduces misclassifications resulting in more accurate predictions.

3.2.3 Vegetation Data

The Forest Inventory Analysis (FIA) program [13] of the U.S Forest Service aims to provide a comprehensive inventory for forest and vegetation data in order to asses and analyze the changes in forests, grasslands, wildlife habitats, ecosystem management and urban planning. This database provides open access to various attributes such as the major forest types, existent tree species, understory vegetation, tree crown conditions, soil type and woody materials present in the forests and urban areas.

The data of the different tree species present in a given region for fine scale vegetation classification is obtained from the FIA data inventor. The data has a confidence level of 68% and is used as the ground truth for validating our model results.

3.2.4 Ground Truth Abundance Data

The hyperspectral images of the selected regions are annotated using the image segmentation tools of the arcGIS [1] software coupled with manual drawing of boundaries for labelling the various land covers. The percent distribution of each land cover type is calculated and used as the ground truth for evaluating the predicted abundances of the different end members.

3.2.5 Hyperspectral Benchmarking Data

The hyperspectral datasets such as Samson, Urban, Cuprite and Jasper [31] are used as the standard benchmarking datasets in hyperspectral image analysis. Several state-of-the-art hyperspectral unmixing techniques [11, 22, 29] are evaluated on these datasets due to the limited availability of other labelled datasets.

The Samson hyperspectral dataset is a small image composed of 95x95 pixels of a real scene

with three end members, namely, soil, tree and water. This data consists of 156 wavelength bands spanning across 401 to 889 nm with a spectral resolution of 3.13 nm. The performance of our autoencoder is evaluated on the Samson dataset and the results are compared to the state-of-the-art methods.

3.3 Implementation

This section explains the various steps in the implementation pipeline of the proposed unsupervised deep learning technique for land cover quantification which is illustrated in Figure 3.3.

1. Input data

   The hyperspectral image data is the input data for the deep learning model. Different sets of real and synthetic hyperspectral images consisting of various target materials are provided as the inputs. The reflectance data is collected from the NEON database for the analyses on the real data. The synthetic data is generated by linearly mixing the pure spectra from the spectral library. The required pure spectra are obtained from the spectral libraries. The vegetation ground truth data is obtained from the FIA database. The abundance ground truth is obtained through annotation using software tools combined with manual labelling.

2. Data preprocessing

   The data preprocessing is performed before training the deep learning models. This is done to remove the missing and invalid reflectance bands from the spectral reflectance data. The reflectance data from NEON Database consists of 426 spectral bands. The invalid, missing and the water vapor bands are removed and the preprocessed

reflectance consists of 315 wavelength bands.

3. Autoencoder Training

   The unsupervised autoencoder is trained on the preprocessed spectral reflectance data. The autoencoder is trained such that it can reconstruct the input pixel at the output using a loss function. In this work, we set the number of end members based on the ground truth data before training the model. The autoencoder is then trained the unmix the input hyperspectral data into the set number of end members. More details regarding the model architecture and the training parameters are described in the subsequent sections.

4. Testing and Evaluation

   After the training is complete, the weights of the last hidden layer connecting to the output layer represent the unmixed end members. Each of these end members correspond to a land cover class. These end member spectra that are unmixed by the autoencoder model are compared to the spectral signatures from the spectral library and are evaluated using metrics such as cosine similarity and mean square error to identify the different types of land cover. The estimated types of land cover are compared to the ground truth to verify if the model has correctly identified the various types of land cover present in the input hyperspectral image.

   Once the end member classes present in the input hyperspectral are determined, every pixel in the input hyperspectral image is classified into one of the end member classes using cosine similarity. The percentage composition of each member in the input hyperspectral image is determined and is evaluated using abundance maps and ground truth abundance values.

   The spectral library is compiled by obtaining an average of multiple spectra for each

land cover class from the EcoSIS, Ecostress and USGS spectral libraries. The spectral library used for evaluation of the results in all the experiments consists of 16 materials covering major urban land covers and a few vegetation species, namely, soil, pine tree, grass, incense cedar tree, sugar pine tree, asphalt road, concrete road, sand, asphalt roof, concrete roof, shingle roof, gravel roof, brick, soil with concrete- gravel and soil with asphalt-gravel which is shown in Figure 3.4.

The experiments are performed on three sets of data:

(a) Experiments on synthetic data

(b) Experiments on benchmarking hyperspectral data

(c) Experiments on real data

5. Hyperparameter Tuning

   If the unmixing results and the estimated abundances determined in Step 4 are not found to be satisfactory, then some of the hyperparameters of autoencoder such as the layer architecture, loss function, batch size, learning rate, etc., are changed and the Steps 3 and 4 are repeated until the model results are satisfactory.

6. Land Cover Quantification Output

   The weights of the last hidden layer of the trained autoencoder network are the unmixed output spectra which correspond to the various land covers present in the input hyperspectral image. Thus, the model is able to successfully identify and quantify the abundance distributions of the different end members present in the image scene which is the output of the implementation pipeline.

The various research objectives of this work are achieved through 3 sets of experiments, namely, experiments on synthetic data, experiments on benchmarking hyperspectral data

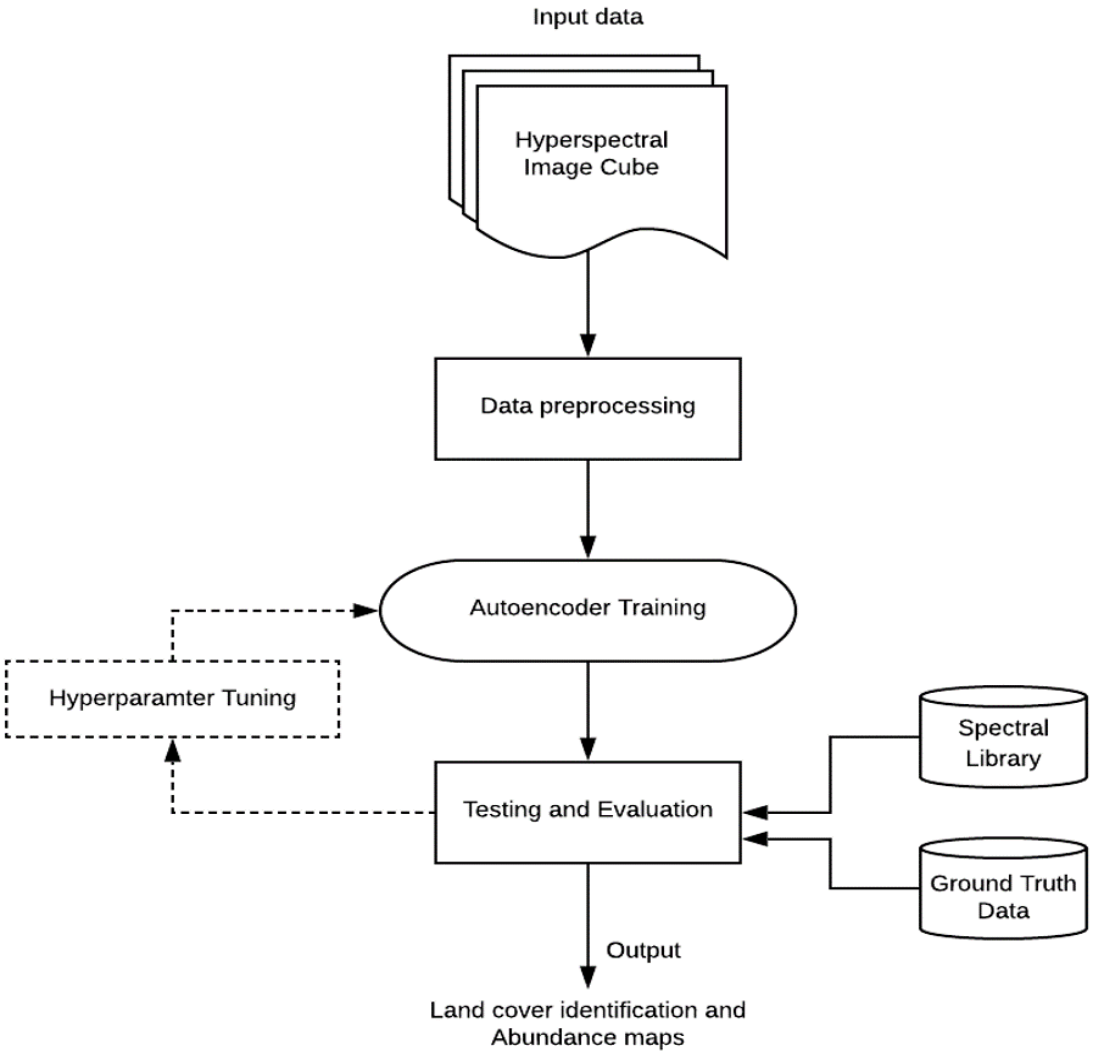

Figure 3.3: Implementation Pipeline

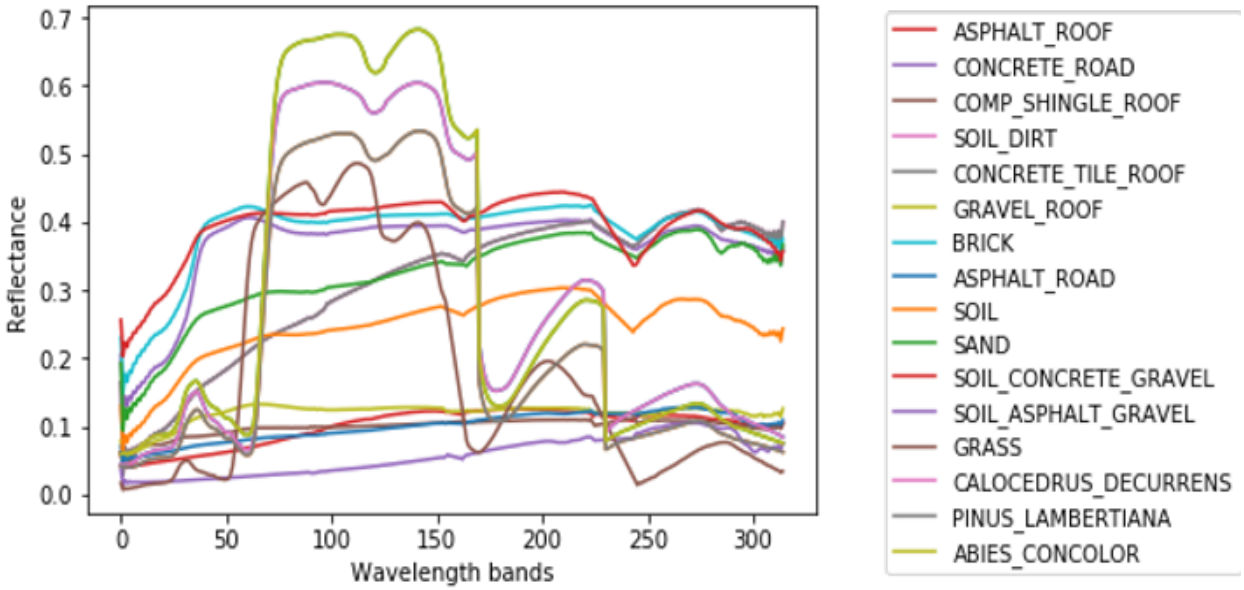

Figure 3.4: Spectral Library

and experiments on real hyperspectral data. Each of these sets of experiments follows the implementation pipeline shown in Figure 3.3 and are described in detail in the subsequent subsections. All the experiments are performed using Python programming language with Keras and Scikit-learn libraries.

3.3.1 Experiments on synthetic data

The feasibility of the autoencoder based spectral unmixing is first evaluated on synthetic data since it is less complex and free from noise unlike real data.

Generation of synthetic data:

The synthetic data is generated using the pure spectral components from the spectral library and linearly mixing them resulting in hyperspectral data with mixed pixels [29]. The pure end members are chosen from the USGS [33], Ecostress [15] and EcoSIS [34] Spectral Libraries and are resampled to contain 315 reflectance bands over 380-2150 nm wavelength

range.

For example, a synthetic dataset with two end members, soil and tree is generated as follows: Each training sample in the dataset is a mixed pixel composed of more than one material of different compositions. We first simulate 1000 training samples where the major end member is soil. Each of the data samples has soil composition varying between 90% to 100%. That is, a value p such that $0.9 < p < 1$, is randomly generated which constitutes the soil abundance in the pixel. Since the total abundance fraction in a pixel must sum up to 1, the remaining (1-p) composition of the pixel is tree. A linearly mixed pixel (x) consisting of soil and tree with p and 1-p abundances respectively is generated as follows :

$$x = p * spectra\_soil + (1 - p) * spectra\_tree \tag{3.8}$$

where spectra_soil and spectra_tree represent the 315 dimension pure, reflectance spectra of soil and tree from the spectral library.
If a training sample is generated with 100% abundance of soil and 0% abundance of tree, then it represents the pure soil spectra.
Similarly, the next 1000 training samples are generated with tree as the major end member. Thus, there are a total of 2000 mixed pixels with different, randomly generated proportions of soil and tree. These 2000 samples consist of 1000 samples for each end member.

For generating a synthetic dataset with more than two end members, the major end member is simulated with an abundance purity of p, where p varies between 0.8-1.0 (i.e., 80-100%). The remaining (1-p) fraction is constituted by all the different end members with randomly generated proportions such that the total composition of all the end members in each sample

sums up to 1.0. For example, consider a dataset with tree, soil and road end members with a total of 3000 training samples. A mixed pixel with soil as the major end member is generating with the soil composition p, such that p lies between 0.8-1.0. Suppose the value of p is 0.85, then the composition of tree in this pixel, i.e. q, is randomly generated generated such that q varies between 0-0.15. For example, if the value of q is 0.10, then the composition of road in this pixel should be 0.05 since the total composition of soil, tree and road should add up to 1.0. Similarly, a total of 1000 training data samples are generated with soil as the major end member. The same procedure is repeated to generate 1000 mixed pixels with tree as the major end member and another 1000 mixed pixels with road as the major end member. The major end member in a pixel has an abundance of at least 0.80 when there are more than two end members instead of 0.90 in order to have a better representation of all the end members in the mixed pixel.

We have considered scenarios with different types of land cover and different number of training samples in order to effectively assess the performance of the autoencoder in the land cover quantification task. The different studies performed on synthetic data are described below:

1. Different number of land covers

   The performance of the autoencoder is evaluated on simulated data with two, three, four and five end members with 1000 training samples per end member class to analyze how the model behaves when there are different number of land covers to be identified in a given hyperspectral image scene.

2. Different distribution of land covers

   In order to evaluate the model behavior in image scenes with different distribution of the land covers, datasets are simulated with different number of training samples for each end member. For a two end member dataset consisting of soil and tree land

covers with a total of 2000 samples, we simulate training data with 50:50, 60:40, 80:20 and 90:10 compositions of tree:soil. For example, in a dataset with 50:50 proportion of tree:soil, there are 1000 training samples that are mixed pixels with soil as the major end member. Each of these 1000 samples have a 90-100% composition of soil and 0-10% composition of tree. The same dataset also consists of another 1000 training samples that have tree as the major end member. Each of these pixels consist of 90-100% composition of tree and the remaining 0-10% is constituted by soil.

Similarly, a dataset with 60:40 proportion of tree:soil with a total of 2000 training samples consists of 1200 mixed pixels with tree as the major end member with 90-100% pixel abundance and 800 mixed pixels with soil as the major end member with 90-100% pixel abundance.

3. Different land covers

   In the Experiment 1 described above, the two end member dataset consists of tree and soil, the three end member set consists of 2 kinds of trees and soil, the four member dataset includes 2 kinds of trees, soil and road, and the five member dataset includes tree, soil, water, road and roof. These experiments help to analyze the autoencoder's performance on different land covers.

4. Different sizes of land covers

   The autoencoder is evaluated on training datasets with 100, 1000 and 10,0000 samples per class consisting of two materials, soil and tree, to asses the scalability of the model in regions with different sizes of land covers.

Evaluation Metrics:

The unmixed end member spectra are quantitatively evaluated by comparing them to the components in the spectral library using *Cosine similarity* and *Mean Square Error* metrics.

Further, qualitative evaluation is performed through visualization of the spectra.

Cosine similarity gives an indication of how close two vectors are by calculating the cosine of the angle between them. The cosine similarity between two n-dimensional vectors, A and B, is calculated as follows [3]:

$$Cosine\,similarity = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n}(A_i)^2}\sqrt{\sum_{i=1}^{n}(B_i)^2}} \tag{3.9}$$

Mean square error (MSE) is a metric that measures the average squared difference between the estimated value $\hat{y}_i$ and the actual value y_i is calculated using the following equation [4] :

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{3.10}$$

where n is the number of data points Since the shape of the reflectance pattern is different for different materials, cosine similarity is a suitable metric to differentiate between the various land cover types as it captures the variation in shapes. The shape of the spectra of different vegetation species are very similar to each other with only differences in their peak magnitudes. Since cosine similarity metric is invariant to the magnitude, we use mean square error as an additional metric to measure how close the unmixed spectra are to the pure components.

3.3.2 Experiments on Samson hyperspectral data

Samson dataset is one of the standard benchmarking datasets for hyperspectral unmixing as described in Section 3.2.5. We perform the autoencoder based unmixing on this dataset which consists of 9025 training samples of a real scene with soil, tree and water end members. The architecture of the autoencoder for training the Samson data is shown in Figure 3.2.

Evaluation Metrics:

(i) The unmixed end member spectra are quantitatively evaluated by comparing them to the ground truth end member spectra using *cosine similarity* and *mean square error* metrics. Qualitative evaluation is performed through visualization of the spectra.

(ii) The estimated percentage abundance of each end member is compared to the ground truth abundance. Further, the abundances are qualitatively evaluated using *abundance maps*. *Abundance map* is a land cover map representing the distribution of the different land covers in an image scene.

(iii) The performance of the proposed autoencoder is compared to other state-of-the-art autoencoders using *spectral angle distance* metric. *Spectral angle distance (SAD)* is a measure of the angle between two vectors that is calculated using the following equation:

$$SAD = \arccos \frac{ab}{||a||||b||} \tag{3.11}$$

where a and b represent the two vectors.

3.3.3 Experiments on real hyperspectral data

The performance of the unsupervised unmixing is evaluated on the real hyperspectral reflectance data obtained from the NEON airborne remote sensing source. We perform a comprehensive analysis on different sized regions covering various types of land covers and further compare the results with other supervised and unsupervised techniques for land cover identification.

The architecture of the autoencoder that is used for unmixing the NEON reflectance data for identifying the land covers is shown in Figure 3.5. A 100m x 100m image is provided as

the input in the Figure 3.5. The NEON reflectance data consists of 315 wavelength bands after preprocessing and removing the invalid bands. Therefore, the input image cube has a dimension of 100x100x315 where each input pixel corresponds to a reflectance spectra with 315 dimensions. The data cube is reshaped into 10000x315 dimensions and the pixel-wise input data is fed into the autoencoder for training. The number of units in the input and output layer is equal to 315. It consists of two hidden layers with 32 and 2 units. The Hidden Layer 2 consists of 2 units because the input hyperspectral image consists of two land cover classes, namely, soil and vegetation. The autoencoder tries to reconstruct the spectra of each input pixel at the output of the network using mean square error loss function. At the end of the training process, the weights connecting Hidden Layer 2 to the output layer which has a dimension of 2x315 represent the end member spectra of tree and soil.

The details of the experiments performed on the real reflectance data are provided below:

1. Different land covers

   The hyperspectral image data for land cover quantification is obtained from the following regions :

   (i) Teakettle Experimental Forest region in California, U.S, with the major land cover types being vegetation and soil for the year 2017 [19]. This region consists of three major tree species, namely, white fir, incense cedar and sugar pine.

   (ii) Smithsonian Environment Research Center in Maryland, U.S, consisting of vegetation, soil and water as the major land cover types for the year 2017 [18].

   (iii) ABBY field site in Washington, U.S, which is a forest region consisting of soil and vegetation as the major land covers for the year 2018 [20]. This region consists of two major species of trees, namely, douglas fir and western hemlock.

   These regions are selected as they are diverse and in different geographic locations

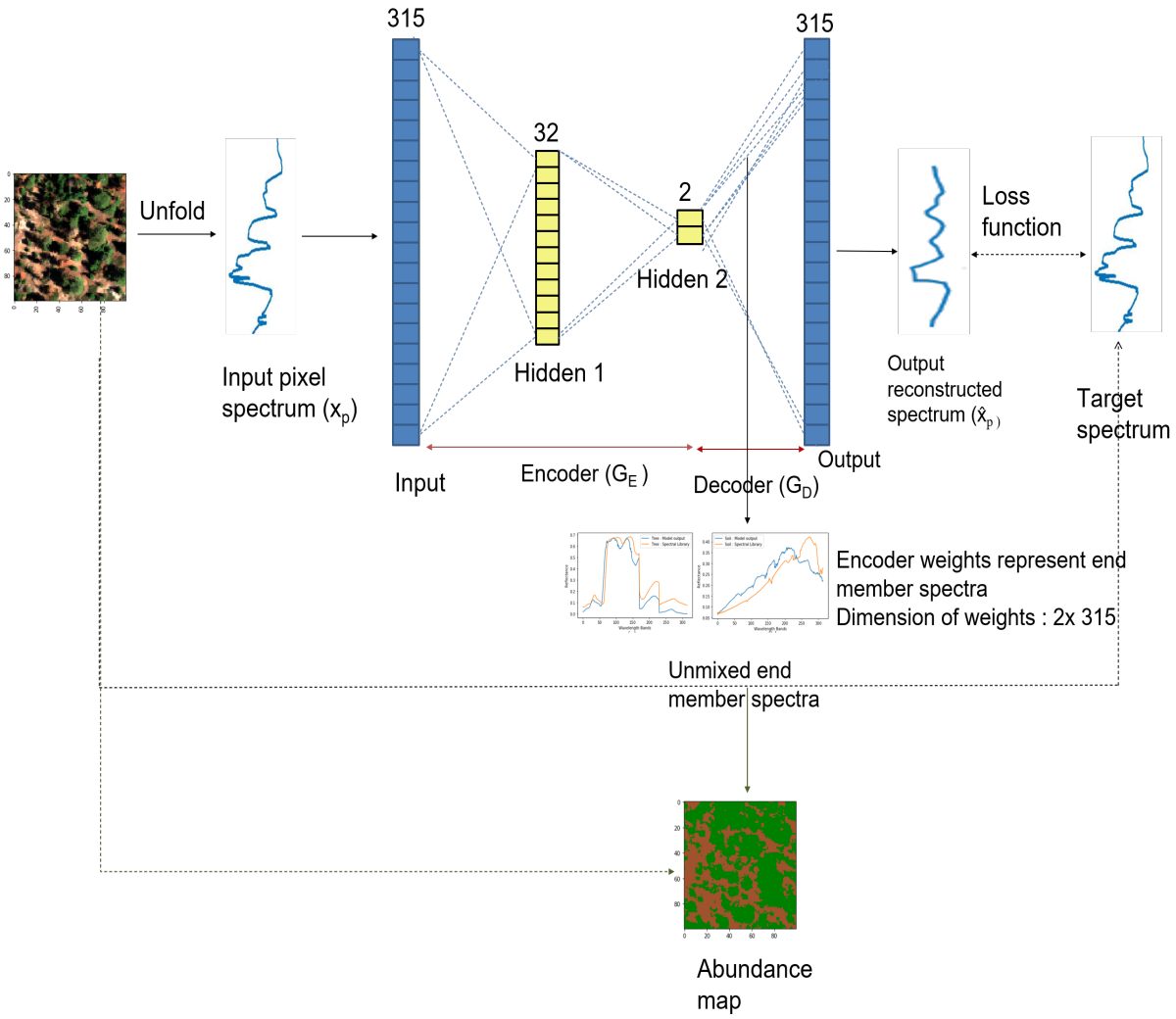

Figure 3.5: Autoencoder for unmixing real hyperspectral data

which helps to capture the variability in ecological conditions, land covers and vegetation.

2. Different sizes of land covers

   The autoencoder model is evaluated on hyperspectral images of different sizes in order to study the scalability of the technique to regions with different sizes of land covers. Here a reasonable assumption is made that a larger hyperspectral image scene represents larger land covers as it spans over a wider area. Experiments are performed on regions with dimensions 100m x 100m, 200m x 200m, 1000m x 1000m, 4000m x 4000m and 5000m x5000m.

3. Comparison with supervised techniques

   The performance of the autoencoder is compared to several supervised deep learning methods such as Support Vector Machine (SVM), Random Forest (RF) and neural networks.

4. Comparison with unsupervised techniques

   The performance of the autoencoder is compared to other unsupervised techniques such as K-Means, N-Findr and VCA. K-Means is an unsupervised clustering algorithm that tries to classify the input spectra into different clusters and the centroid of each cluster corresponds to an end member spectral signature. N-Findr [36] and VCA [17] are traditional spectral unmixing algorithms based on geometric methods.

5. Comparison with land cover classification standard

   National Land Cover Database (NLCD) is a USGS standard for land cover classification [32]. It consists of 20 ecosystem classes for classifying the land covers. NLCD serves as the definitive Landsat-based, 30-meter resolution, land cover database for the US. It supports a variety of Federal, State, local, and nongovernmental applications

that seek to assess ecosystem status and health, understand the spatial patterns of biodiversity, predict effects of climate change, and develop land management policy. The different classes of the NLCD standard are:

(a) Open Water

(b) Perennial Ice/ Snow

(c) Developed, Open Space

(d) Developed, Low Intensity

(e) Developed, Medium Intensity

(f) Developed, High Intensity

(g) Barren Land (Rock/Sand/Clay)

(h) Deciduous Forest

(i) Evergreen Forest

(j) Mixed Forest

(k) Dwarf Scrub

(l) Shrub/Scrub

(m) Grassland/Herbaceous

(n) Sedge/Herbaceous

(o) Lichens

(p) Moss

(q) Pasture/Hay

(r) Cultivated Crops

(s) Woody Wetlands

(t) Emergent Herbaceous Wetlands

The land cover classes obtained by the autoencoder are compared with the existing NLCD classification to assess the accuracy of the autoencoder.

Evaluation metrics:

(i) The end member spectra unmixed by the autoencoder are qualitatively evaluated by visualization of the spectra. Further, quantitative evaluation is performed by comparing them to the pure spectra in the spectral library using *cosine similarity* and *mean square error* metrics.

(ii) The percentage composition of each end member is calculated and compared to the ground truth abundance value. *Abundance maps* are generated for qualitative evaluation of the land cover abundances.

(iii) The results of the autoencoder are compared to the results of the supervised learning methods using *precision*, *recall* and *F1 Score* metrics which are explained in the following equations [25]:

$$Precision = \frac{TruePositive}{TruePositive + FalsePositive} \tag{3.12}$$

$$Recall = \frac{TruePositive}{TruePositive + FalseNegative} \tag{3.13}$$

$$F1Score = \frac{2 * Precision * Recall}{Precision + Recall} \tag{3.14}$$

where True Positive is the correctly predicted positive value, True Negative is the correctly predicted negative value, False Positive is the actual negative class that is predicted as positive, and False Negative is the actual positive class that is predicted

as negative.

(iv) The performance of the autoencoder is compared with that of the unsupervised techniques using *cosine similarity*, *mean square error* and estimated percentage abundances.

Chapter 4

Results

In this chapter, we present the results and evaluations of the autoencoder based land cover quantification on all different datasets.

4.1 Results of Synthetic Data

The results of hyperspectral unmixing on synthetic data are presented below.

4.1.1 Experiments with different number of land covers

Datasets are generated with 2, 3, 4 and 5 different land covers.

1. Two end members

   A dataset with two end members, namely, tree and soil end members is simulated with 1000 training samples per class and the autoencoder based unmixing is performed. The model parameters are: autoencoder with 4 hidden layers with 2, 6, 12, 18 units with L2 regularization, Adam optimizer with 0.05 learning rate, cosine similarity loss function, batch size of 32 and 200 epochs. The ground truth spectra from the spectral library along with the unmixed results are shown in Figure 4.1. The cosine similarity and mean square error between the model output and the corresponding component from spectral library are shown in Table 4.1.

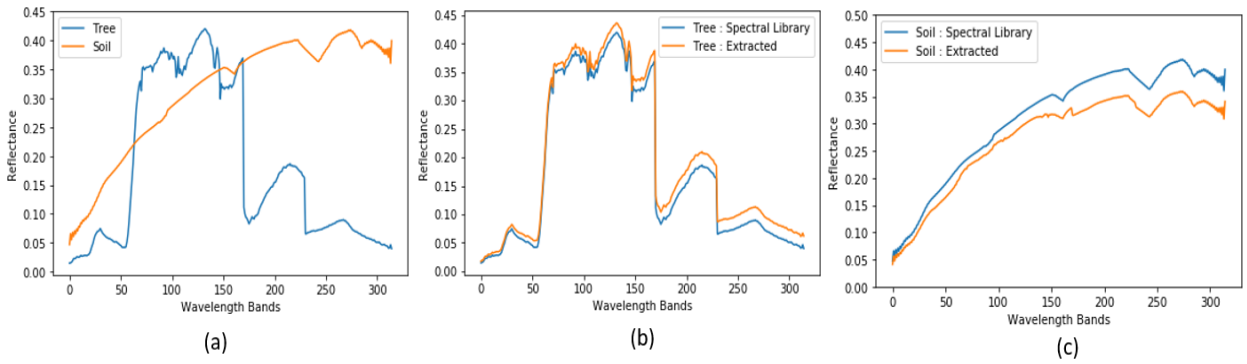

Figure 4.1: Result spectra of synthetic data with 2 end members (a) Ground truth spectra (b) Tree (c) Soil

2. Three end members

   A dataset with three end members consisting of two different trees species, namely, *Pinus Lambertiana* and *Liquidambar styraciflua*, and soil end members is generated with 1000 samples per class. The model parameters are: autoencoder with 4 hidden layers with 3, 9, 18, 36 units with L2 regularization, Adam optimizer with 0.01 learning rate, cosine similarity loss function, batch size of 32 and 200 epochs. The ground truth spectra from the spectral library along with the model results are shown in Figure 4.2. The cosine similarity and mean square error between the model output and the corresponding library components are shown in Table 4.1. It can be observed that the model is able to differentiate between the vegetation and soil land covers. It can also distinguish between the different species of vegetation which have similar spectral shapes and vary only in their magnitudes.

3. Four end members

   A dataset with 4 end members consisting of soil, road and 2 different trees species, namely, *Pinus Lambertiana* and *Liquidambar styraciflua*, is generated with 1000 samples per class. The model parameters are: autoencoder with 4 hidden layers with 4, 8, 16, 28 units with L2 regularization, Adam optimizer with 0.01 learning rate, cosine

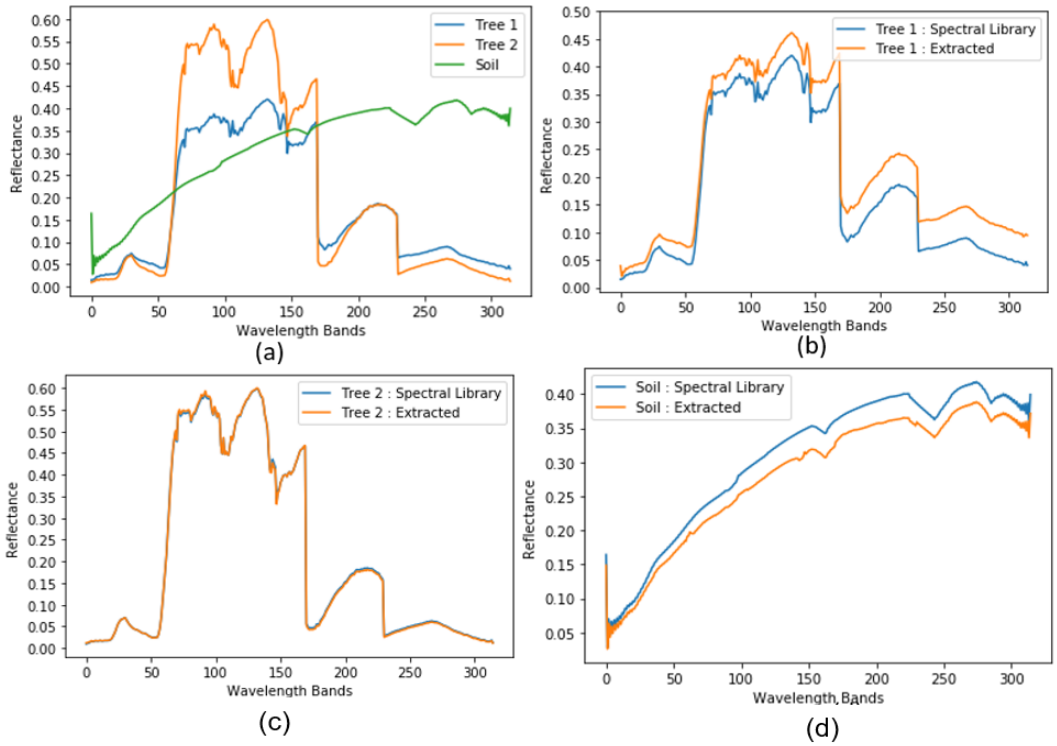

Figure 4.2: Result spectra of synthetic data with 3 end members (a) Ground truth spectra (b) Tree 1 (c) Tree 2 (d) Soil

similarity loss function, batch size of 16 and 2000 epochs. The ground truth spectra from the spectral library along with the model results are shown in Figure 4.3 and the cosine similarity and mean square error between the model outputs and their corresponding library components are shown in Table 4.1.

It can be observed that the model is able extract the shapes of all the end member spectra well. However, the magnitudes of only some of the end members such as tree 1 and tree 2 match closely with the ground truth where as the magnitudes of road and soil end members deviate slightly from the ground truth values. This does not cause problems in cases where the land cover spectra are very different from each other as it is sufficient to match only the spectral shapes in such cases. The magnitude is important only in applications such as vegetation classification where the spectral shapes are similar and the end members need to be differentiated based on their magnitudes. In this particular example, the magnitudes of the unmixed tree 1 and tree 2 spectra match closely with the ground truth and the model is able to distinguish between the two types of trees. Therefore, the magnitude deviation for soil and road end members does not cause any problems as they can be identified based on their different spectral shapes.

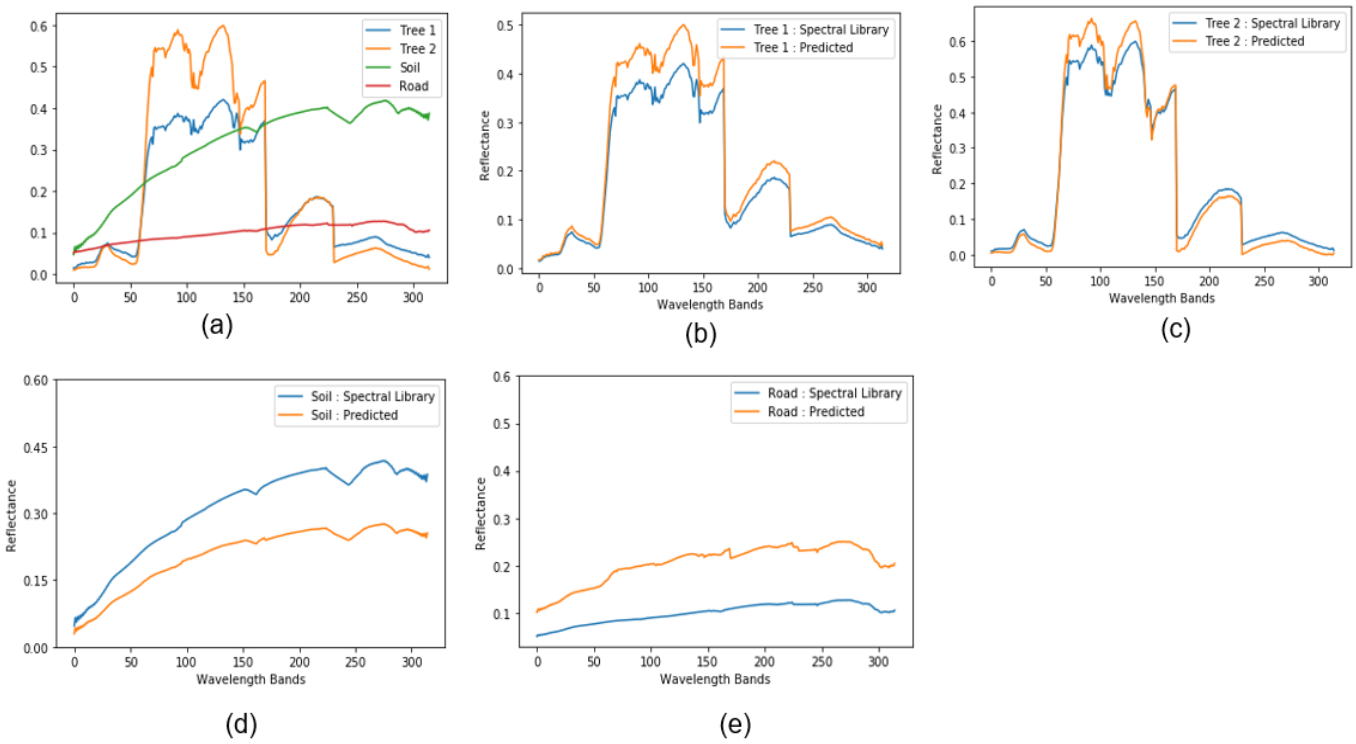

Figure 4.3: Result spectra of synthetic data with 4 end members (a) Ground truth spectra (b) Tree 1 (c) Tree 2 (d) Soil (e) Road

4. Five end members

   A synthetic dataset with 5 end members consisting of tree, soil, road, water and roof is generated with 1000 samples per class. The model parameters are: autoencoder with 4 hidden layers with 5, 15, 30, 45 units with L2 regularization, Adam optimizer with 0.001 learning rate, cosine similarity loss function, batch size of 8 and 100 epochs. The ground truth spectra from the spectral library along with the model results are shown in Figure 4.4. The cosine similarity and mean square error values of the unmixed spectra are shown in Table 4.1.

   The results show that the autoencoder is able extract the shapes of all the end member spectra well. However, the magnitudes of only some of end members such as tree, soil and water match closely with the ground truth where as the magnitudes of road and

roof deviate slightly from the ground truth values. This does not cause problems in this case as these land covers are very different from each other and matching only their spectral shapes is sufficient to identify them.

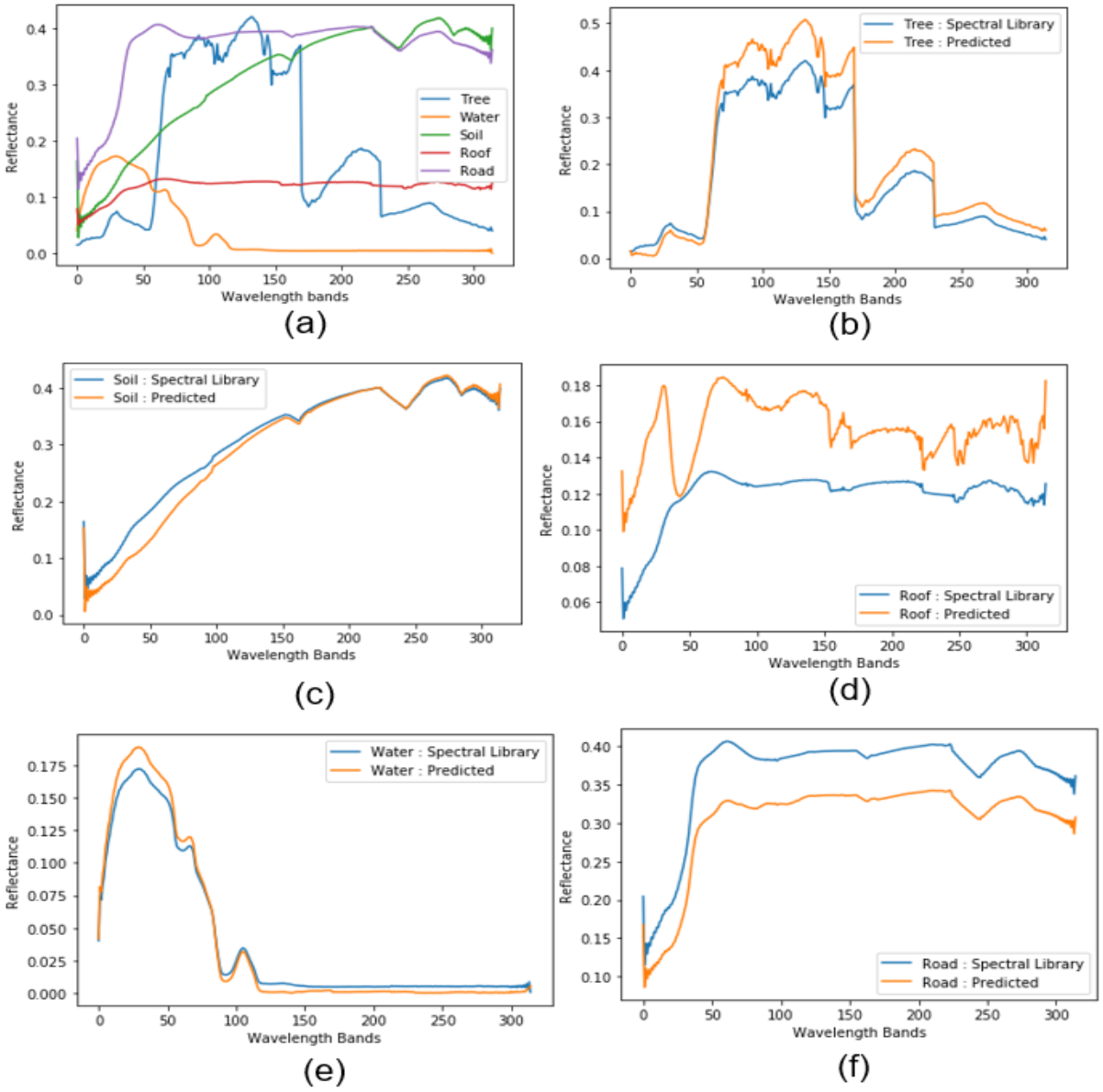

Figure 4.4: Result spectra of synthetic data with 5 end members (a) Ground truth spectra (b) Tree (c) Soil (d) Roof (e) Water (f) Road

Table 4.1: Results of synthetic data with different number of land covers

Number of End Members	End Members	Cosine Similarity	Mean Square Error
2	Tree	0.998688	3.2422×10^{-4}
	Soil	0.999552	1.5897×10^{-3}
3	Tree 1	0.993461	2.0684×10^{-3}
	Tree 2	0.999958	7.9507×10^{-6}
	Soil	0.999891	7.7879×10^{-4}
4	Tree 1	0.999982	1.7474×10^{-3}
	Tree 2	0.997108	1.0596×10^{-3}
	Soil	0.999894	1.1773×10^{-2}
	Road	0.998503	1.1567^{-2}
5	Tree	0.998823	2.2855×10^{-3}
	Soil	0.997883	4.8318×10^{-4}
	Roof	0.993942	1.6059×10^{-3}
	Water	0.997448	5.2822×10^{-5}
	Road	0.999498	3.6742^{-3}

4.1.2 Experiments with different sizes of land covers

Spectral unmixing is performed on synthetic datasets with different number of training samples which gives an indication of the performance of the model on regions with different sizes of land covers. Datasets with 2 end members, namely, soil and tree with 200, 2,000 and 20,000 training samples with equal number of samples per class in all the cases are synthetically generated. The visualizations of the output spectra are shown in Figure 4.5. The similarity of the unmixed spectra with the pure spectral components in the library is determined using cosine similarity and mean square error and is presented in Table 4.2. This analysis gives an indication of the scalability of the model. The results correspond well with the ground truth and prove that spectral unmixing using autoencoder can be performed on small as well as large datasets.

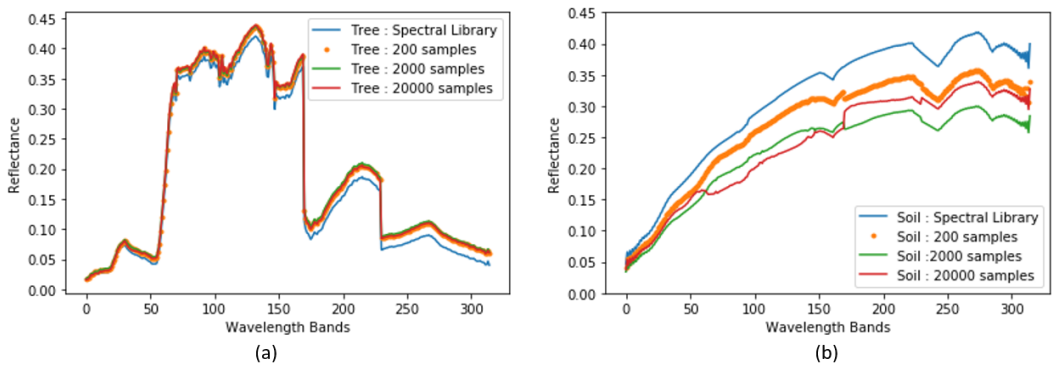

Figure 4.5: Result spectra of synthetic data with different sizes of land covers (a) Tree (b) Soil

Table 4.2: Results of synthetic data with different sizes of land covers

Number of Samples	Cosine Similarity	Mean Square Error
200	Tree: 0.999097 Soil: 0.999732	Tree: 2.6677×10^{-4} Soil: 2.6302×10^{-4}
2,000	Tree: 0.998688 Soil: 0.999555	Tree: 3.2422×10^{-4} Soil: 1.5897×10^{-3}
20,000	Tree: 0.999190 Soil: 0.998391	Tree: 2.7871×10^{-4} Soil: 5.5455×10^{-3}

4.1.3 Experiments with different distribution of land covers

Datasets with 2 end members, namely, soil and tree, with 50:50, 60:40, 70:30, 80:20 and 90:10 compositions of tree:soil for a total number of 2000 training samples are generated. The unmixed end member spectra and the estimated abundances for all the cases are shown in the Figure 4.6 and Table 4.3. In order to evaluate if the model is able to accurately estimate the abundances of the various classes even with skewed distributions, the percentage composition of each of the different end members is calculated for all the cases. The results indicate that the abundance estimations and the similarity scores of the end member spectra

match well with the ground truth for all the different ratios of the end members.

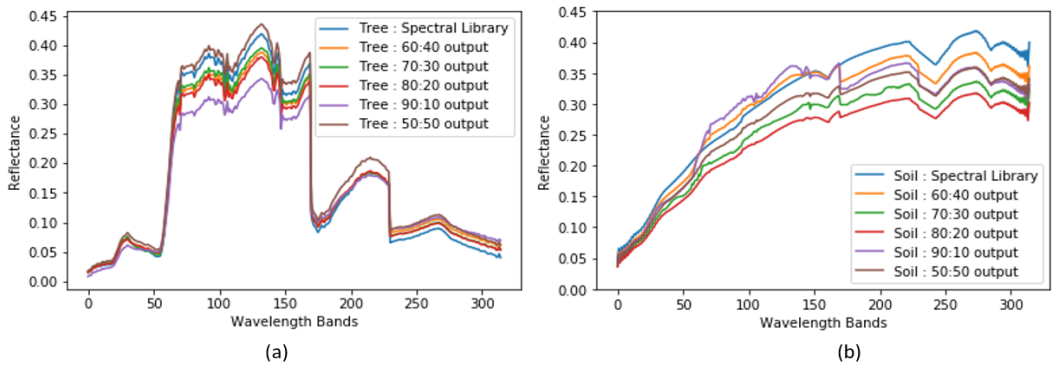

Figure 4.6: Result spectra of synthetic data with different distribution of land covers (a) Tree (b) Soil

Table 4.3: Results of synthetic data with different distribution of land covers

Ground Truth Tree:Soil	Cosine Similarity	Mean Square Error	Estimated Abundance
50%:50%	Tree: 0.998688 Soil: 0.999552	Tree: 3.2422×10^{-4} Soil: 1.5897×10^{-3}	Tree: 50% Soil: 50%
60%:40%	Tree: 0.998132 Soil: 0.998751	Tree: 3.055×10^{-4} Soil: 4.8673×10^{-4}	Tree: 60% Soil: 40%
70%:30%	Tree: 0.999427 Soil: 0.999658	Tree: 1.0573×10^{-3} Soil: 8.3955×10^{-4}	Tree: 70% Soil: 30%
80%:20%	Tree: 0.998992 Soil: 0.999784	Tree: 1.4111×10^{-4} Soil: 1.0047×10^{-4}	Tree: 80% Soil: 20%
90%:10%	Tree: 0.998992 Soil: 0.999784	Tree: 3.2111×10^{-3} Soil: 2.5237×10^{-3}	Tree: 90% Soil: 10%

4.2 Results of Samson Hyperspectral Data

The results of Samson benchmarking dataset, which is a hyperspectral image of a real scene, are shown below. This dataset consists of three end materials, i.e. Soil, Tree, Water. Figures

4.7 (a) and (b) show the hyperspectral image and the ground truth end members respectively. The autoencoder is trained with four hidden layers with 27, 18, 9, and 3 units, Adam optimizer with 0.01 learning rate, cosine similarity loss function, batch size of 16 and for 200 epochs. The output spectra of the model is shown in the Figure 4.8 and the evaluation results are given in Table 4.4.

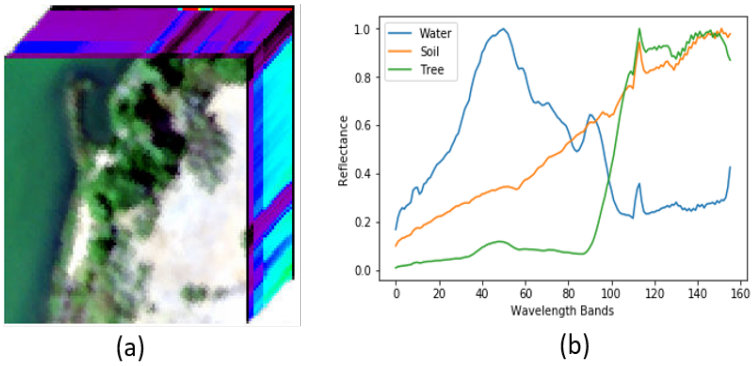

Figure 4.7: (a) Samson Hyperspectral Image (b) Ground truth spectra

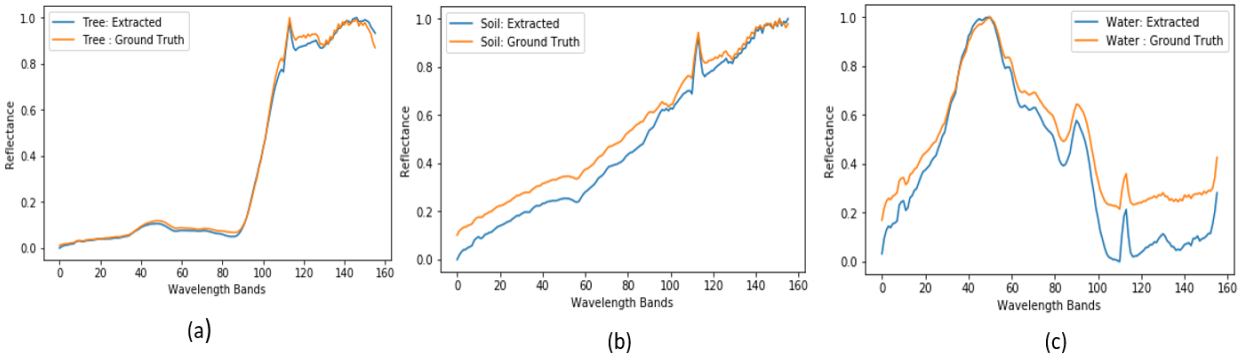

Figure 4.8: Result spectra of Samson hyperspectral dataset (a) Tree (b) Soil (c) Water

Table 4.4: Results of Samson hyperspectral data

End member	Cosine Similarity	Mean Square Error
Tree	0.999373	3.5587×10^{-4}
Soil	0.999555	4.9707×10^{-3}
Water	0.998636	1.6322×10^{-2}

The estimated percentage composition of each of the materials, i.e. soil, tree and water is calculated and compared with the ground truth abundances which can be seen in Table 4.5. The abundance maps of all the end members are shown in Figure 4.9. It can be observed that the model is able to extract all the three end members and quantify their abundances in accordance with the ground truth with an average of ~1.8% deviation from the ground truth.

Table 4.5: End member abundances

Material	Ground truth	Estimated
Tree	40.62%	39.11%
Soil	33.41%	36.10%
Water	25.97%	24.79%

4.2.1 Performance comparison with state-of-the-art autoencoder models

The performance of the proposed autoencoder on Samson dataset is compared with that of two state-of-the-art autoencoders. The first autoencoder is from the work in [22] which is a two-stage autoencoder with a novel, modified spectral angle distance loss function. The

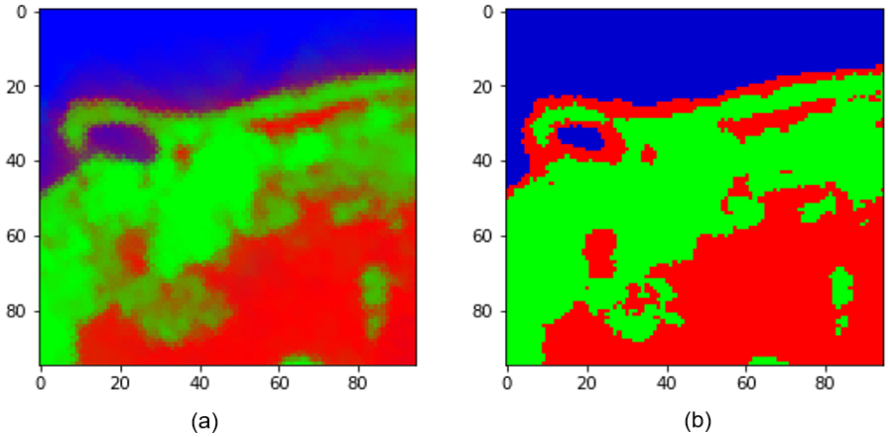

Figure 4.9: Abundance maps of Samson dataset: Blue- Water, Green- Tree, Red- Soil (a) Ground Truth(b) Estimated

model is initialized with VCA technique, has a dropout layer, l1 normalization layer and l2 regularizer, Adam optimizer and uses a batch size of 64 for the training data. The second autoencoder is from the research work in [23] which has a complex architecture with four hidden layers, dropout layer, batch normalization layer, spectral angle distance (SAD) loss function and a batch size of 20.

In order to compare our proposed method to other state-of-the-art methods using these data sets, we calculate the SAD between the unmixed end member spectra and the given ground truth end member spectra. The unmixing is performed 50 times for each method and the mean and standard deviation of the result SAD is calculated across all the end members and reported in Table 4.6. It can be observed that the proposed method which uses a simpler autoencoder with cosine similarity loss function provides a slight improvement in mean SAD value than the two state-of-the-art autoencoders.

Table 4.6: Performance comparison with state-of-the-art autoencoder models (All values in radians)

Model	Model 1	Model 2	Our autoencoder
Soil	0.0129±0.1	-	0.0253±0.038
Tree	0.0469 ±0.1	-	0.0207±0.034
Water	0.0295 ±0.3	-	0.0424±0.101
Mean SAD	0.0298 ±0.2	0.031 ±0.004	0.0294±0.0576

4.3 Results of Real Hyperspectral Data

The results on real hyperspectral reflectance data obtained from NEON database [21] for different regions are presented in this section.

The different sets of experiments performed are:

1. Regions with different land covers
2. Regions with different sizes
3. Comparison with Supervised Techniques
4. Comparison with Unsupervised Techniques

4.3.1 Regions with different land covers

Experiments are performed on different regions consisting of a variety of land covers and the results are presented below.

(i) Region 1: Soil and Vegetation

A 100m x 100m region in the Teakettle Experimental Forest described in Section 3.3.3 consisting of soil and vegetation land covers is considered which is shown in Figure

4.10 (a). The autoencoder model unmixes the input data into two land covers and identifies each of the target materials correctly from a spectral library consisting of major land covers and the vegetation species of the region.

The training data consists of 10,000 samples and the model parameters are: autoencoder with 2 hidden layers consisting of 32 and 2 hidden units, Adam optimizer with 0.001 learning rate, mean square error loss function, batch size of 16 and 50 epochs.

The result spectra are shown in Figures 4.10 (b) and (c). Table 4.7 shows the similarity scores and the mean square error values of the unmixed soil and tree spectra. The high cosine similarity values indicate that the shapes of the extracted end members match well with the ground truth spectra and the low mean square error values indicate that the magnitudes of the extracted end members match well with the ground truth spectra from spectral library. The ground truth and estimated abundance maps are illustrated in Figures 4.11 (a) and (b). The abundance percentages are shown in Table 4.8. It can be observed that the estimated abundance values for tree and soil are very close to the ground truth.

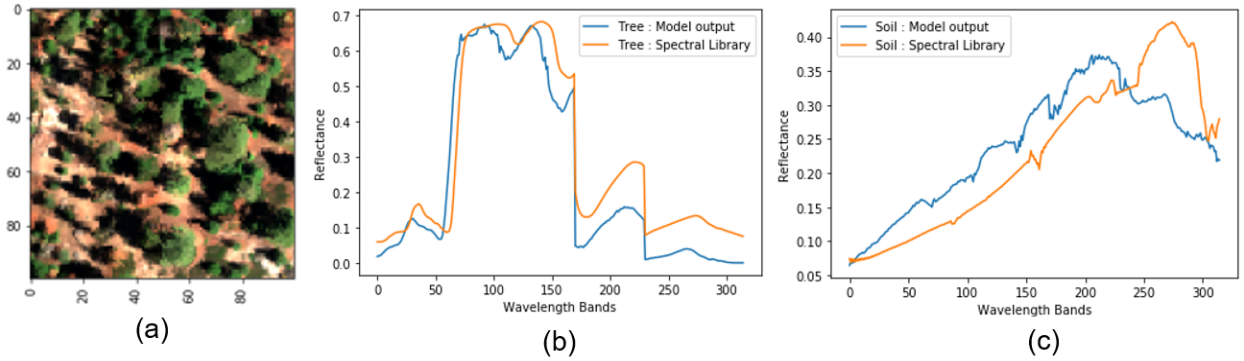

Figure 4.10: Result spectra of real hyperspectral data: Region 1 (a) Input (b) Tree (c) Soil

Table 4.7: Results of real hyperspectral data: Region 1

End member	Cosine Similarity	Mean Square Error
Tree	0.972957	8.0356×10^{-3}
Soil	0.962456	3.8219×10^{-3}

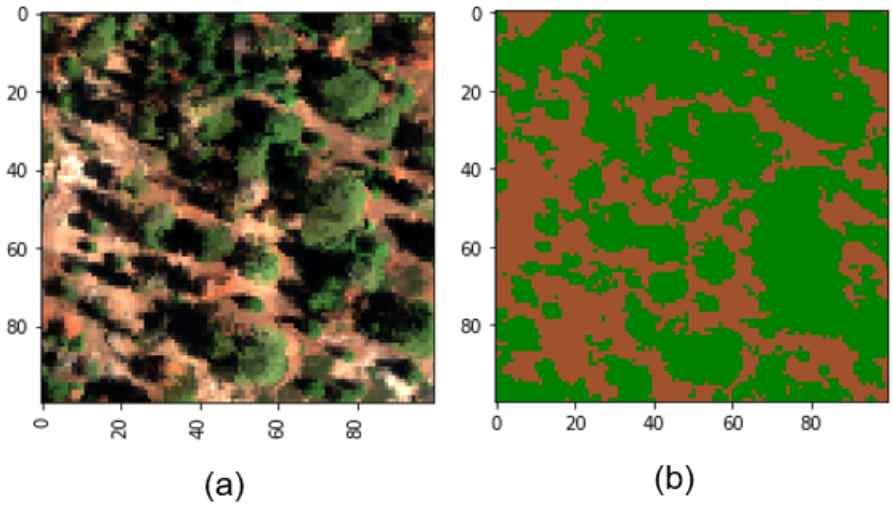

Figure 4.11: Region 1 Abundance Maps: Brown- Soil, Green- Tree (a) Ground truth (b) Estimated

Table 4.8: Abundances: Region 1

End member	Ground truth	Estimated
Tree	65.44%	61.59%
Soil	34.56%	38.41%

(ii) Region 2 : Vegetation and Road

A 200m x 200m region from the Smithsonian Environmental Research Center consisting of road and vegetation land covers shown in Figure 4.13 (a) is considered for analysis. The training data consists of 40,000 samples and the model parameters are: autoencoder with 2 hidden layers consisting of 6 and 2 hidden units, Adam optimizer with 0.01 learning rate, cosine similarity loss function, batch size of 512 and 15 epochs.

The results of the unmixed land cover spectra are shown in the Figures 4.12 (b) and (c) and Table 4.9. The high cosine similarity scores indicate that the shapes of the unmixed end members match well with the ground truth and the low mean square error values indicate that the magnitudes of the end members are close to the ground truth. The ground truth abundances and predicted abundances are shown in Figure 4.13 and Table 4.10. It can be observed that the estimated abundances match well with the ground truth abundance for both road and tree land covers.

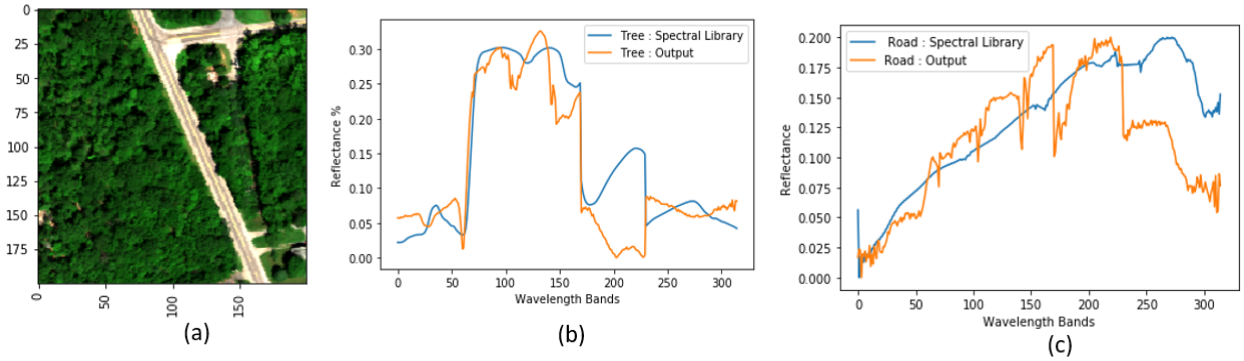

Figure 4.12: Results of real hyperspectral data: Region 2(a) Input (b) Tree (c) Road

Table 4.9: Result spectra of real hyperspectral data: Region 2

End member	Cosine Similarity	Mean Square Error
Tree	0.961507	2.9851×10^{-3}
Road	0.983616	4.5928×10^{-2}

Table 4.10: Abundances: Region 2

End member	Ground truth	Estimated
Tree	87.23%	85.88%
Road	12.77%	14.12%

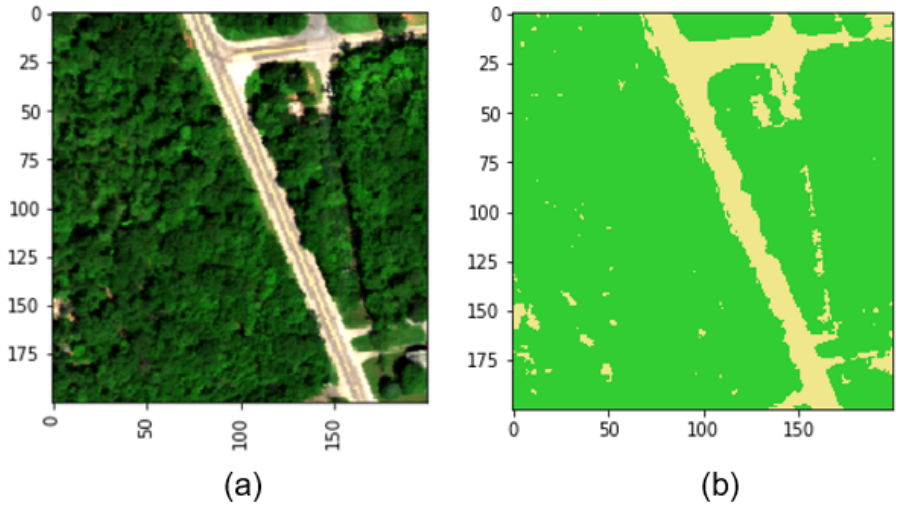

Figure 4.13: Region 2 Abundance Maps: Road - Brown, Tree- Green (a) Ground truth (b) Estimated

(iii) Region 3 : Vegetation and Water

A 200m x 200m region from the Smithsonian Environmental Research Center consisting of water and vegetation land covers shown in Figure 4.14 (a) is considered for analyses. The training data consists of 40,000 samples and the model parameters are: autoencoder with 3 hidden layers consisting of 12, 6 and 2 hidden units respectively, Adam optimizer with 0.01 learning rate, cosine similarity loss function, batch size of 32 and 25 epochs.

The results of the unmixed land cover spectra are shown in the Figures 4.14 (b) and (c) and Table 4.11. The ground truth and predicted abundances are presented Table 4.12 and the abundance maps are shown in Figure 4.15. The similarity score, mean square error and the percentage abundance values indicate that the model results match closely with the ground truth.

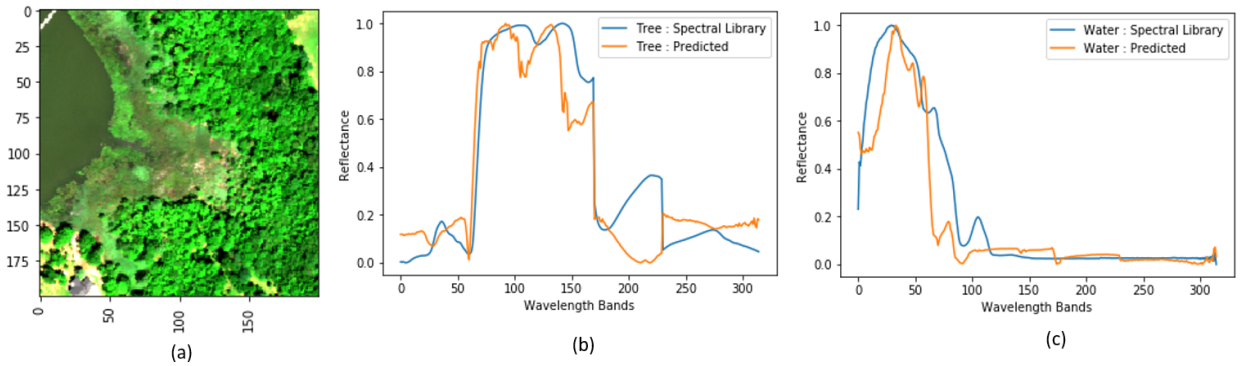

Figure 4.14: Result spectra of real hyperspectral data: Region 3 (a) Input (b) Tree (c) Water

Table 4.11: Results of real hyperspectral data: Region 3

End member	Cosine Similarity	Mean Square Error
Tree	0.960457	2.4558×10^{-2}
Water	0.950889	1.8366×10^{-2}

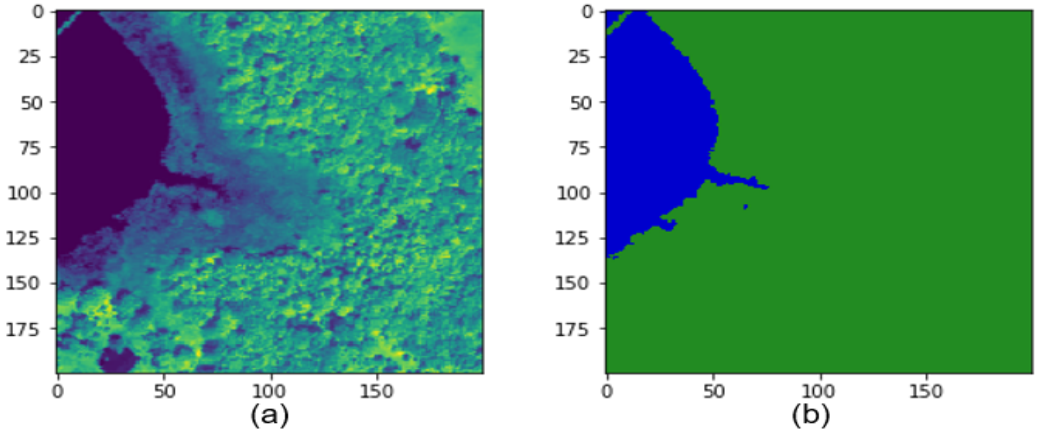

Figure 4.15: Region 3 Abundance Maps : Water - Blue, Tree - Green (a) Ground truth (b) Estimated

Table 4.12: Abundances: Region 3

End member	Ground truth	Estimated
Tree	85.94%	86.81%
Water	14.06%	13.19%

4.3.2 Regions of different sizes

In this set of experiments, we analyze the performance of the unsupervised autoencoder on hyperspectral images of different sizes which gives an indication of scalability of the model to larger regions. Here we use the data from Teakettle Experimental Forest region in California for the 100m x100m, 1000m x 1000m and 5000m x 5000m regions and the data from ABBY field site in Washington for the 4000m x 4000m region which consist of vegetation and soil as the major land covers.

(i) 100m x 100m region

The analysis of 100m x 100m region has already been performed in the previous section while analyzing regions with different land covers. The input image and the output spectra of the 2 end members, namely, vegetation and soil, are illustrated in Figure 4.10 and the output spectra are evaluated using cosine similarity and mean square error as shown in and Table 4.7. Figure 4.11 and Table 4.8 show the estimated abundances and the abundance maps respectively.

(ii) 1000m x 1000m region

A hyperspectral image of size 1000x1000 pixels covering 1 sq km forest region is used for training the autoencoder. The model parameters are as follows: autoencoder with 3 hidden layers consisting of 18, 6 and 2 hidden units, mean square error loss function, Adam optimizer with 0.001 learning rate, batch size of 512 and 10 epochs. The resultant spectra are shown in Figures 4.16 (b) and (c) and Table 4.13 and are found

to match well with the spectral library. The estimated abundance map is shown in Figure 4.17 and the abundant percentages are given in Table 4.14. It can be observed that the abundances of both soil and vegetation match closely with the ground truth.

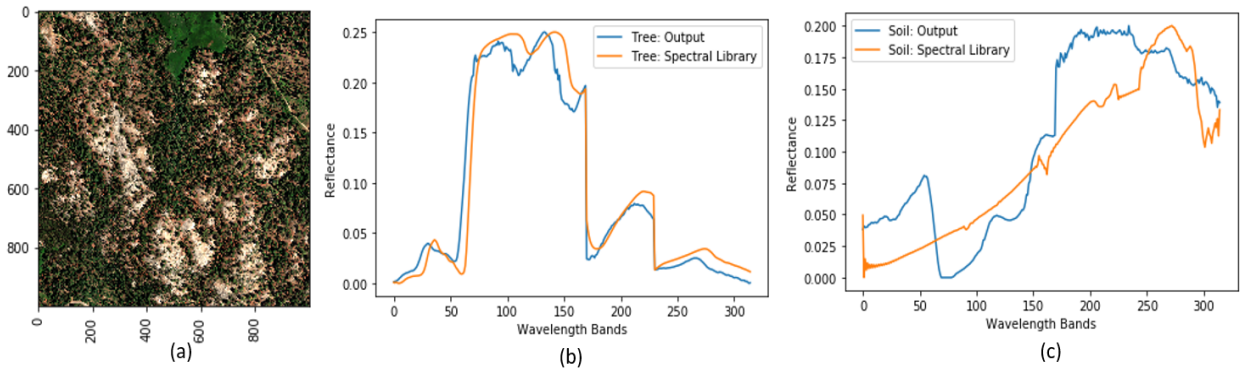

Figure 4.16: Result spectra of 1000m x 1000m region (a) Input (b) Tree (c) Soil

Table 4.13: Results of 1000m x 1000m region

End member	Cosine Similarity	Mean Square Error
Tree	0.983649	5.4087×10^{-4}
Soil	0.963692	1.3493×10^{-3}

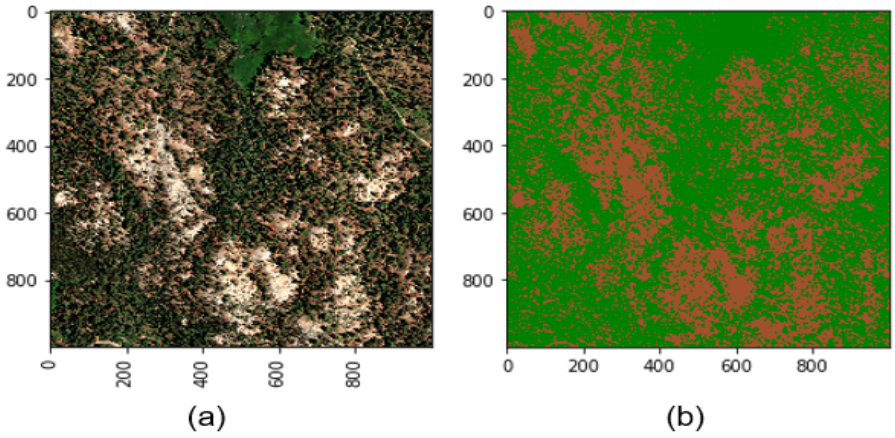

Figure 4.17: 1000m x 1000m Region Abundance Maps: Tree - Green, Soil - Brown (a) Ground Truth (b) Estimated

Table 4.14: Abundances: 1000m x 1000m region

End member	Ground truth	Estimated
Tree	68.32%	65.71%
Soil	31.68%	34.29%

(iii) 4000m x 4000m region

Here we consider a hyperspectral image of size 4000x4000 pixels covering a region of dimension 4000m x 4000m which spans 16 square kilometer area (approximately 6 square miles). This data is taken from ABBY Field site in Washington (iii) for the year 2018 and is shown in Figure 4.18 (a). This region consists of 2 end members, soil and vegetation. The vegetation in this region consists of grass and 2 other major species of trees, namely, Douglas Fir (*Scientific name:Pseudotsuga menziesii*) and Western Hemlock (*Scientific name:Tsuga heterophylla*). The data regarding the types of trees and their distribution is obtained from the Forest Inventory Analysis database described in Section 3.2.3.

Two sets of analyses are performed on this data, i.e. with two and four end members. The experiment with two end members is used to estimate soil and vegetation abundances, and the experiment with four end members is performed to identify a finer classification by estimating the abundances of all the three vegetation species along with soil.

(a) 2 end members

Here we try to unmix the input data into 2 land cover classes, namely, soil and vegetation. The model parameters are as follows: autoencoder with 3 hidden layers consisting of 18, 6 and 2 hidden units, mean square error loss function, Adam optimizer with 0.001 learning rate, batch size of 16 and 10 epochs. The result spectra of both the end members are shown in Figures 4.18 (b) and (c). The

soil and vegetation end members match closely with the ground truth spectra as indicated by the high similarity and low mean square error values in Table 4.15. The ground truth and estimated abundances are shown in Table 4.16 which show that the model results match well with the ground truth.

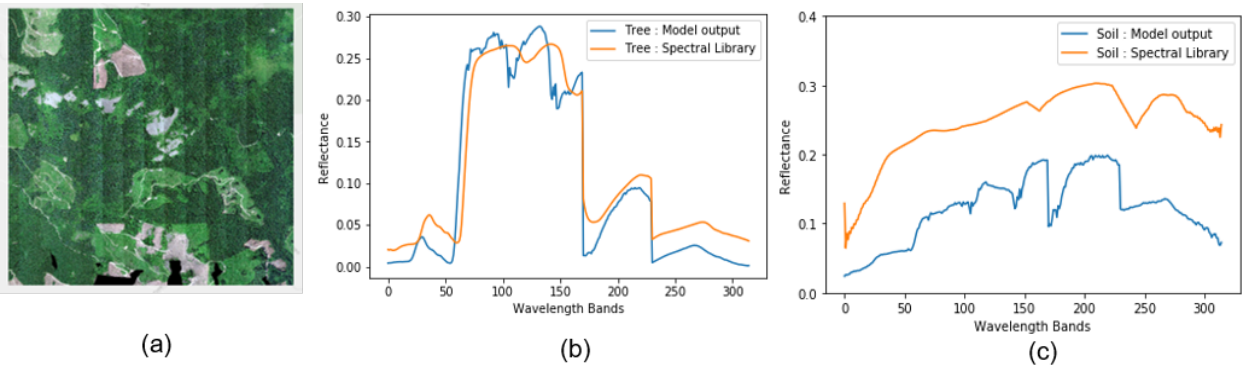

Figure 4.18: Result spectra for 4000m x 4000m region with 2 end members (a) Input (b) Tree (c) Soil

Table 4.15: Results of 4000m x 4000m region with 2 end members

End member	Cosine Similarity	Mean Square Error
Tree	0.976557	1.0363×10^{-3}
Soil	0.974478	5.9784×10^{-3}

Table 4.16: Abundances: 4000m x 4000m region with 2 end members

End member	Ground truth	Estimated
Tree	91.8%	89.02%
Soil	8.2%	10.98%

(b) 4 end members

Here we try to obtain a fine scale vegetation classification for the same 4000m x 4000m region shown in Figure 4.18 (a) to estimate the major types of trees. The input data is unmixed into 4 land cover classes, namely, soil, grass, douglas fir

tree and western hemlock tree. The model parameters are as follows: autoencoder with 3 hidden layers consisting of 18, 6 and 2 hidden units, mean square error loss function, Adam optimizer with 0.001 learning rate, batch size of 16 and 15 epochs. The result spectra of all the end members are shown in Figures 4.19 (a)-(d). The soil and the three vegetation end members match closely with the ground truth spectra both in shape and magnitude as indicated by their high similarity and low mean square error scores in Table 4.17. It is very important to match the peak magnitudes to be able to differentiate the different vegetation species since the shapes of all the different kinds of vegetation are almost the same. The ground truth and estimated abundances are shown in Table 4.18. It can be observed that the estimated abundances match with the ground truth values with an average deviation of ~4.7% which shows the autoencoder is able to perform fine scale vegetation classification.

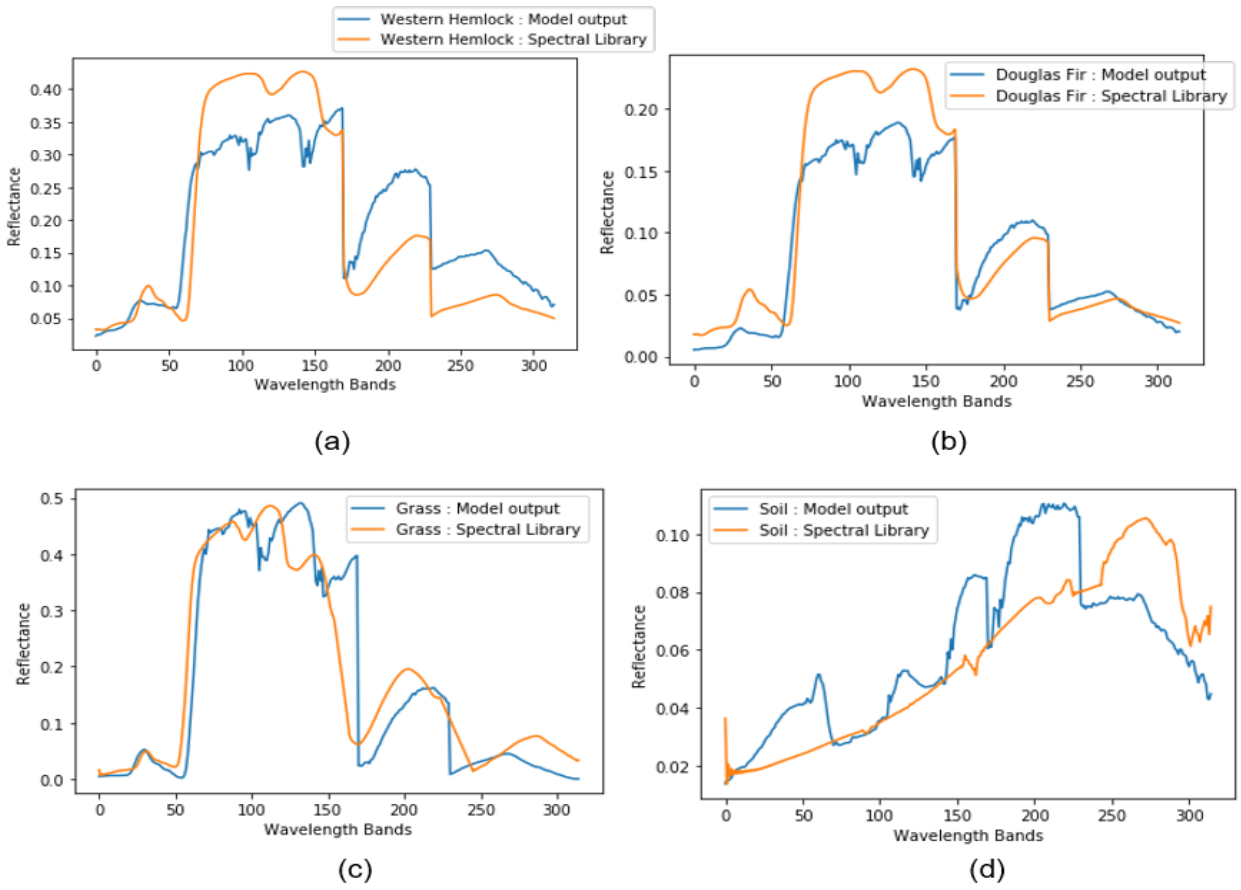

Figure 4.19: Result spectra for 4000m x 4000m region with 4 end members (a) Western hemlock tree (b) Douglas fir tree (c) Grass (d) Soil

Table 4.17: Results of 4000m x 4000m region with 4 end members

End member	Cosine Similarity	Mean Square Error
Soil	0.934322	3.6139×10^{-4}
Grass	0.955343	5.6270×10^{-3}
Douglas Fir	0.980355	1.0524×10^{-3}
Western Hemlock	0.951793	5.3262×10^{-3}

Table 4.18: Abundances: 4000m x 4000m region with 4 end members

End member	Ground truth	Estimated
Soil	8.2%	4.81%
Grass	31.72%	40.34%
Douglas Fir	45.12%	45.95%
Western Hemlock	14.96%	8.9%

- Comparison of 4000m x 4000m region with NLCD Standard

  The NLCD classification for the same 4000m x 4000m region is shown in Table 4.19. The NLCD standard provides Evergreen Forest as the major land cover class with 81.04% composition followed by 7.39% of Shrub, 3.1% of Developed Open Space, 3.04% of Grassland followed by small compositions of a few other classes. The different classes and their compositions provided by the autoencoder are shown in Table 4.20. While the NLCD standard is able to only identify evergreen forest as a major class, the autoencoder is able to identify the major tree types present in the region thus providing vegetation classification at a finer scale.

Table 4.20: Autoencoder Classification: 4000m x 4000m region

Class Name	Composition
Soil	4.81%
Grass	40.34%
Douglas Fir	45.95%
Western Hemlock	8.9%

Table 4.19: NLCD Classification: 4000m x 4000m region

Class Name	Composition
Evergreen forest	81.04%
Shrub/Scrub	7.39%
Mixed forest	3.35%
Developed Open Space	3.10%
Grassland/Herbaceous	3.04%
Developed Low Intensity	0.055%
Deciduous Forest	1.06%
Woody Wetlands	1.01%

(iv) 5000m x 5000m region

A large hyperspectral image of size 5000x5000 pixels covering a land surface of dimension 5000m x 5000m that spans a 25 square kilometer area (approximately 10 square miles) which is shown in Figure 4.20 is used as the input for spectral unmixing. This region consists of 2 end members, soil and vegetation. The vegetation in this region consists of grass and three major species of tree, namely, White Fir tree (*Sceintific name: Abies Concolor*, Incense cedar tree *(Scientific name: Pinus Lamertiana)* and Sugar pine tree *(Scientific name: Calocedrus Decurrens)*. The data regarding the types of trees and their distribution is obtained from the Forest Inventory Analysis database described in Section 3.2.3.

Two sets of analyses are performed on this data, i.e. with two and five end members. The experiment with two end members is used to estimate soil and vegetation abundances, and the experiment with five end members is performed to identify a finer classification by estimating the abundances of each of the four vegetation species along with soil.

(a) Two end members :

The autoencoder is trained to unmix the input data into two end member classes using the following model parameters: autoencoder with 3 hidden layers consisting of 18, 6 and 2 hidden units, mean square error loss function, Adam optimizer with 0.001 learning rate, batch size of 16 and 15 epochs. The unmixed result spectra are shown in the Figure 4.21. The cosine similarity scores are high and mean square error values are low as seen in Table 4.21 which indicate that the unmixed soil and tree spectra match well with the ground truth both in shape and magnitude. The results of abundance estimation are shown in Table 4.22 which indicate that the estimated abundances are in accordance with the ground truth.

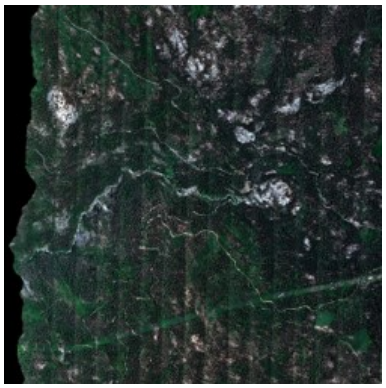

Figure 4.20: Input image: 5000m x 5000m region

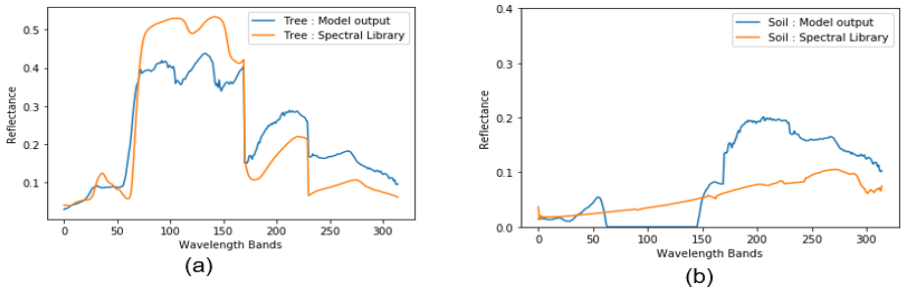

Figure 4.21: Result spectra for 5000m x 5000m region with 2 end members (a) Tree (b) Soil

Table 4.21: Results of 5000m x 5000m region with 2 end members

End member	Cosine Similarity	Mean Square Error
Tree	0.962064	7.0389×10^{-3}
Soil	0.919896	3.7209×10^{-3}

Table 4.22: Abundances: 5000m x 5000m region with 2 end members

End member	Ground truth	Estimated
Tree	79.45%	74.94%
Soil	20.55%	25.06%

(b) Five end members :

Here we attempt to perform a fine scale vegetation estimation for the same 5000m x 5000m region shown in Figure 4.20 to estimate the major types of trees. The input data is unmixed into 5 land cover classes, namely, soil, grass, white fir tree, incense cedar tree and sugar pine tree. The model parameters are as follows: autoencoder with 3 hidden layers consisting of 18, 6 and 5 hidden units respectively mean square error loss function, Adam optimizer with 0.001 learning rate, batch size of 16 and 15 epochs. The unmixed end member spectra are shown in Figures 4.22 (a)-(e) and the similarity and mean square error scores are shown in Table 4.23. It can be observed that the result spectra match with the ground truth spectra in shape and magnitude thus successfully discriminating between the different vegetation species. The abundance estimation results are presented in Table 4.24. It is observed that the abundances of some of the end members match with the ground truth values where as there is a large deviation from the ground truth for some end members. This is because the vegetation spectra of all the tree species and grass have the same shapes but differ only in their peak magnitudes which are very close to each other leading to misclassifications.

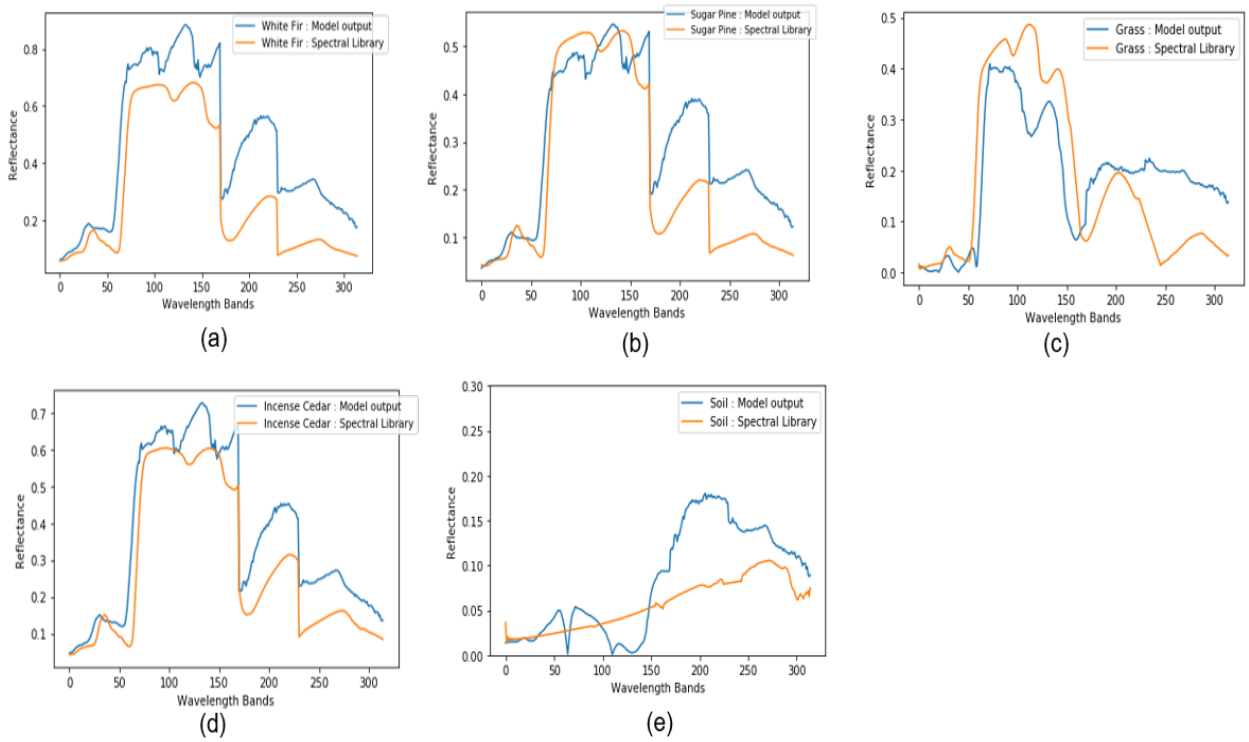

Figure 4.22: Result spectra for 5000m x 5000m region with 5 end members (a) White Fir (b) Sugar pine (c) Grass (d) Incense cedar (e) Soil

Table 4.23: Results of 5000m x 5000m region with 5 end members

End member	Cosine Similarity	Mean Square Error
Soil	0.946729	2.2658×10^{-3}
Grass	0.895806	1.1892×10^{-2}
White Fir	0.965229	3.4026×10^{-2}
Incense Cedar	0.981970	3.31452×10^{-2}
Sugar Pine	0.953882	1.0431×10^{-2}

- Comparison of 5000m x 5000m region with NLCD Standard

  The NLCD classification for 5000m x 5000m region is shown in Table 4.25.

  The NLCD standard provides Evergreen Forest as the major land cover class

Table 4.24: Abundances : 5000m x 5000m region with 5 end members

End member	Ground truth	Estimated
Soil	20.55%	23.65%
Grass	24.401%	34.37%
White Fir	46.901%	22.87%
Incense Cedar	2.642%	5.89%
Sugar Pine	5.449%	13.32%

with 93.39% composition, 5.2% of Shrub land cover class and small percentage compositions of a few other classes. The different classes and their compositions provided by the autoencoder are shown in Table 4.26. While the NLCD standard only identifies the major forest type of the region, the autoencoder is able to identify the major tree types present in the region thus providing vegetation classification at a finer scale.

Table 4.25: NLCD Classification: 5000m x 5000m region

Class Name	Composition
Evergreen forest	93.39%
Shrub/Scrub	5.20%
Deciduous Forest	0.39%
Grassland/Herbaceous	0.37%
Emergent Herbaceous Wetlands	0.28%
Mixed Forest	0.25%
Woody Wetlands	0.12%

Table 4.26: Autoencoder Classification: 5000m x 5000m region

Class Name	Composition
Soil	23.65%
Grass	34.37%
White Fir	22.87%
Incense Cedar	5.89%
Sugar Pine	13.32%

4.3.3 Comparison with supervised techniques

The performance of the unsupervised autoencoder is compared with supervised learning classifiers such as SVM, RF and neural network for hyperspectral images of different sizes. F1 score, precision, recall and estimated abundance percentages are used as the metrics for performance comparison.

i) 100m x 100m region

The models are trained on a 100m x 100m region of Teakettle Experimental Forest consisting of 2 end members, soil and vegetation which can be seen in Figure 4.23 (a). Classifiers based on supervised methods such as SVM, RF and neural network are trained on the hyperspectral reflectance data from the 100m x 100m region and their corresponding labels. The autoencoder is trained in an unsupervised manner using only the reflectance data without the ground truth labels. The model parameters of all the techniques are given below:

(a) SVM: Kernel- Linear, Regularization penalty (C)- 1

(b) RF: Number of trees- 10, Bootstrap- True

(c) Neural Network: 3 hidden layers with 50,30,10 units, Input layer- 315 units, Output layer- 2 units with softmax activation, Activation function- ReLU, Optimizer- Adam with 0.001 learning rate, Batch size- 200, Epochs- 200, Regularization- l2 regularizer (0.0001)

(d) Autoencoder: 2 hidden layers with 2 and 32 units, Regularization- l2 regularizer (0.0001), Optimizer- Adam with 0.001 learning rate, Batch size- 16, Epochs- 50, Loss function- Mean squared error, Activation function- LeakyReLU for hidden layers, Linear activation for input and output layers

The trained models are tested on another unseen 100m x 100m region from Teakettle Experimental Forest which is shown in Figure 4.23 (b) which was not part of the training data and consists of the same 2 end members. The results of all the supervised and unsupervised techniques on the testing data are presented in Table 4.27. A comparison of the estimated abundances of the test region using different methods is shown in Table 4.28. The precision, recall, F1 scores and estimated abundances of the autoencoder are found to be comparable with that of the supervised techniques. Although the results of the supervised learning techniques are more accurate than the autoencoder, they require a lot of labelled data for training the models where as the autoencoder can provide comparable results with ~2% deviation in abundances with respect to the ground truth.

ii) 1000m x 1000m

The models are trained on a 1000m x 1000m region of Teakettle Experimental Forest consisting of 2 end members, soil and vegetation shown in Figure 4.24(a). SVM, RF and neural network classifiers are trained on the hyperspectral image data from the

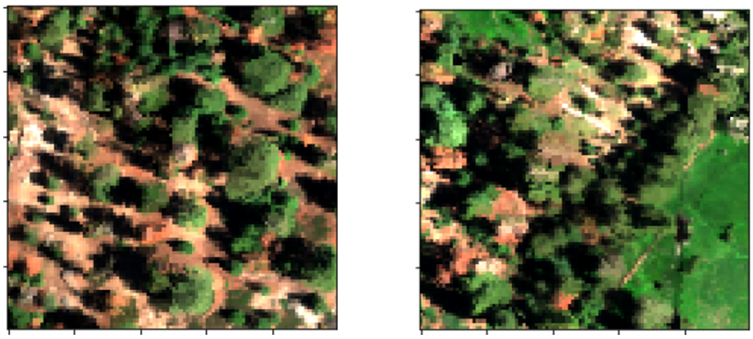

Figure 4.23: 100m x 100m region (a) Training data (b) Testing data

Table 4.27: Comparison with supervised techniques : 100m x 100m region

Model	Precision	Recall	F1 Score
SVM	0.9834	0.9834	0.9834
RF	0.989948	0.9848	0.987367
Neural Network	0.989501	0.9896	0.98955
Autoencoder	0.955684	0.930289	0.942815

1000m x 1000m region and their corresponding ground truth labels. The autoencoder is trained in an unsupervised manner using only the reflectance data without the labels. The model parameters of all the techniques are given below:

(a) SVM: Kernel- Linear, Regularization penalty (C)- 1

(b) RF: Number of trees- 10, Bootstrap- True

(c) Neural Network: 2 hidden layers with 30,10 units, Input layer- 315 units, Output layer- 2 units with softmax activation, Activation function- ReLU, Optimizer- Adam with 0.001 learning rate, Batch size- 200, Epochs- 200, Regularization- l2 regularizer (0.0001)

(d) Autoencoder: 2 hidden layers with 2 and 32 units, Regularization- l2 regularizer (0.0001), Optimizer- Adam with 0.001 learning rate, Batch size- 16, Epochs- 50,

Table 4.28: Comparison of estimated abundances of test data

Model	Soil Abundance	Vegetation Abundance
Ground truth	13.07%	86.93%
SVM	12.87%	87.13%
RF	13.89%	86.11%
Neural Network	13.28%	86.72%
Autoencoder	15.38%	84.62%

Loss function- Mean squared error, Activation function- Leaky ReLU for hidden layers, Linear activation for input and output layers

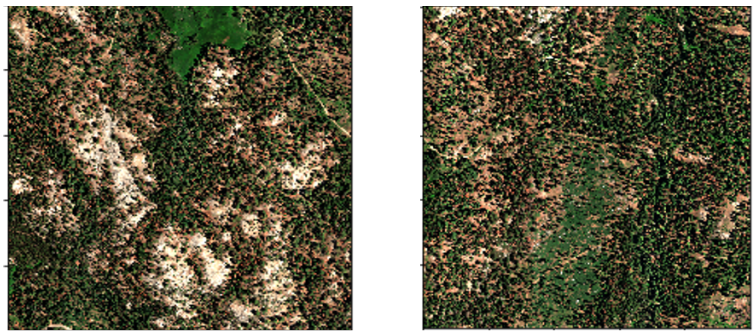

Figure 4.24: 1000x1000 region (a) Training data (b) Testing data

The testing data is another 1000m x 1000m region which was not part of the training data and consists of the same 2 end members and is shown in Figure 4.24(b). The trained models are evaluated on the test data and the results of the supervised and unsupervised models on the testing data are shown in Table 4.29. The estimated abundances of the different end members for test data is shown in Table 4.30. It can be observed that the precision, recall, F1 scores and estimated abundances of the autoencoder are less accurate than the supervised techniques. But the supervised learning model require a lot of labelled data for training. The autoencoder is able to provide

comparable results with ~2% deviation in abundances from the ground truth without the need of labelled data.

Table 4.29: Comparison with supervised techniques: 1000m x 1000m region

Model	Precision	Recall	F1 Score
SVM	0.967653	0.967653	0.967653
RF	0.969240	0.969026	0.969134
Neural Network	0.970089	0.970105	0.970097
Autoencoder	0.937365	0.884243	0.91003

Table 4.30: Comparison of estimated abundances of test data: 1000m x 1000m region

Model	Soil Abundance	Vegetation Abundance
Ground truth	20.99%	79.01%
SVM	21.25%	78.75%
RF	21.26%	78.74%
Neural Network	21.47%	78.53%
Autoencoder	23.22%	76.78%

4.3.4 Comparison with unsupervised techniques

In this section, the performance of the autoencoder is compared with other unsupervised techniques such as K-Means, VCA and N-Findr for hyperspectral images of different sizes.

1. 100m x 100m region

   The land cover classification accuracy of the unsupervised autoencoder is compared with the unsupervised K-Means clustering, VCA and N-Findr unmixing methods for a 100m x 100m region shown in Figure 4.25 (a) that consists of soil and vegetation end members. The unmixed soil and vegetation resultant spectra are compared to the ground truth spectra using cosine similarity and mean square error values as shown in Table 4.31 and the autoencoder spectra is found to match well with the the ground truth with lowest mean square error scores. The abundance estimation results are provided in Table 4.32. It can be observed that the estimated abundances of the autoencoder are closer to the ground truth values. Therefore, the autoencoder is found to provide more accurate land cover quantification than the other unsupervised techniques.

Table 4.31: Result spectra comparison with unsupervised techniques: 100m x 100m region

Technique	Cosine Similarity	Mean Square Error
K-Means	Tree: 0.981984 Soil: 0.943832	Tree: 7.9176×10^{-2} Soil: 1.5296×10^{-2}
N-Findr	Tree: 0.982563 Soil: 0.988073	Tree: 2.6687×10^{-2} Soil: 1.1163×10^{-1}
VCA	Tree: 0.982563 Soil: 0.988074	Tree: 3.1231×10^{-2} Soil: 9.3364×10^{-2}
Autoencoder	Tree: 0.972957 Soil: 0.962456	Tree: 8.0356×10^{-3} Soil: 3.8219×10^{-3}

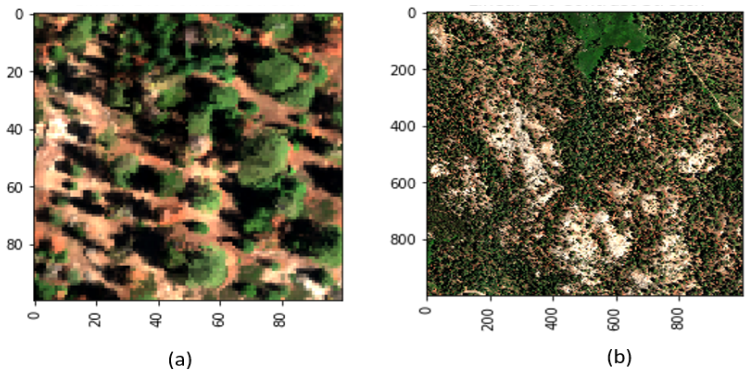

Figure 4.25: Comparison with unsupervised techniques (a) 100m x 100m region (b) 1000m x 1000m region

Table 4.32: Abundance comparison with unsupervised techniques: 100m x 100m region

Model	Soil	Vegetation
Ground truth	31.56%	68.44%
K-Means	40.75%	59.25%
N-Findr	45.59%	54.41%
VCA	50.01%	49.99%
Autoencoder	38.14%	61.84%

2. 1000m x 1000m region

   The spectral unmixing is performed on a 1000m x 1000m region consisting of soil and vegetation end members which is shown in Figure 4.25(b) using several unsupervised methods. The unmixed soil and vegetation spectra are compared to the ground truth spectra using cosine similarity and mean square error values as shown in Table 4.33 and the autoencoder spectra is found to match well with the the ground truth with lowest mean square error scores. The abundance estimation results are provided in Table 4.34. It can be observed that the autoencoder performs better than the other unsupervised techniques.

Table 4.33: Result spectra comparison with unsupervised techniques: 1000m x 1000m region

Technique	Cosine Similarity	Mean Square Error
K-Means	Tree: 0.9854223 Soil: 0.952289	Tree: 8.0699×10^{-2} Soil: 1.5266×10^{-2}
N-Findr	Tree: 0.969348 Soil: 0.972874	Tree: 4.6719×10^{-2} Soil: 2.0057×10^{-1}
VCA	Tree: 0.936535 Soil: 0.836466	Tree: 2.7445×10^{-2} Soil: 1.4554×10^{-1}
Autoencoder	Tree: 0.983649 Soil: 0.963692	Tree: 5.40876×10^{-4} Soil: 1.3493×10^{-3}

Table 4.34: Abundance comparison with unsupervised techniques: 1000m x 1000m region

Model	Soil Abundance	Vegetation Abundance
Ground truth	31.67%	68.32%
K-Means	52.98%	47.02%
N-Findr	58.75%	41.25%
VCA	37.94%	62.06%
Autoencoder	34.28%	65.72%

Chapter 5

Discussion

In this work, we have implemented an unsupervised deep learning technique using an autoencoder for land cover quantification from hyperspectral images. Here we discuss the results of the various experiments that have been performed.

5.1 Synthetic hyperspectral dataset

The results presented in Section 4.1 demonstrate that linearly mixed synthetic data can be unmixed using an autoencoder. The shape and the magnitudes of the output spectra are compared to the components of a spectral library and the land covers present have been identified correctly. We obtain a high cosine similarity score between the output spectra and its corresponding pure component in the library which indicates that their shapes are similar and that there is a match in their reflectance signatures. Further, we obtain a low mean square between the unmixed spectra and their counterparts in the spectral library which indicates a high degree of match in their magnitudes which is crucial while discriminating between various vegetation species. We have qualitatively evaluated the result spectra by visualization of the unmixed and library spectra for each of the experiments.

The results indicate that the proposed method has been successful in quantifying different kinds of land cover like soil, tree, water, road and roof with different compositions with two, three, four and five end members in a given scene. We observe that the number of hidden

layers in the autoencoder increases with the size of the image and the kernel size of the hidden layers increases with the number of end members. Training the model with a small batch size of 16, cosine similarity loss function and Adam optimizer is found to give good results in most cases. Experiments were first performed on the synthetic data since they are simpler and without noise unlike the real reflectance data. The results indicate that land cover quantification on synthetic data with mixed pixels can be successfully performed using the proposed autoencoder approach.

5.2 Samson hyperspectral dataset

The results of Section 4.2 show the performance of the autoencoder based spectral unmixing on a popular hyperspectral dataset, namely, Samson dataset, which consists of tree, water and soil land covers. An autoencoder model with 4 hidden layers with 3,9,18 and 27 units, cosine similarity loss function, Adam optimizer with 0.01 learning rate, batch size of 16 and 50 epochs is found to perform well. The autoencoder extracts the tree end member with a similarity score and mean square error of 0.999373 and 3.6687 x 10^{-4} respectively, soil with 0.999555 and 4.9707 x 10^{-3} respectively and water with 0.998636 and 1.6322 x 10^{-2} respectively. The ground truth and predicted abundances are found to be 33.41% and 36.10% respectively for soil, 40.62% and 39.11% respectively for tree and 25.97% and 27.79% respectively for water. The abundance maps that represent the distribution of the different land covers match well with the ground truth maps. This study indicates that land cover quantification can be successfully performed on real hyperspectral data for small regions.

5.3 Real hyperspectral dataset

The results in Section 4.3 show the performance of the autoencoder for land cover identification task on hyperspectral reflectance data from real forest regions. The unmixed land covers from the input data are found to have high similarity scores and low mean square with the pure components from the spectral library as shown in the Tables 4.7, 4.9, 4.11 and 4.13. The estimated compositions of the different end members are found to match well with the ground truth abundance data as shown in the Tables 4.11, 4.10, 4.12 and 4.17 and the abundance maps are shown in the Figures 4.11, 4.13, 4.15 and 4.17.

Further, the results of Section 4.3 indicate that the proposed technique works well in forest regions of dimensions 100m x 100m, 200m x 200m, 1000m x 1000m, 4000m 4000m and 5000m x 5000m, that is, it performs well in smaller regions spanning 10,000 square meter area to larger regions spanning across 25 square kilometer area (9.65 square miles). This shows that autoencoder based unmixing is scalable and can be used to quantify small and large landscapes.

The autoencoder provides a finer scale land cover classification than the existing USGS NLCD standard shown by the results of experiments on 4000m x 4000m and 5000m x 5000m regions. The NLCD uses a 30m, low resolution satellite data and identifies only the major ecosystem types in a given region where as the autoencoder uses a high resolution data of 1m and identifies the type of the land cover and the vegetation species in a given region.

This approach provides accurate quantification of different land cover classes when their spectral reflectance patterns are different from each other. For example, in the 100m x 100m region with soil and vegetation and in the 200m x 200m region with water and vegetation, all the end members are very different from each other. In such cases, these land covers can be easily distinguished from each other and their abundances can be accurately quantified

using the autoencoder. But the cases involving fine scale vegetation classification become challenging. This is because all kinds of vegetation have similar reflectance patterns with low reflectances in the visible and short wave infrared regions of the electromagnetic spectrum and high reflectances in the near infrared region as seen in Figure 5.1. The pure spectra of the different vegetation species present in the 5000m x 5000m region are shown in Figure 5.2. It can be observed that the different vegetation species differ only in their peak magnitudes which makes it difficult to unmix the hyperspectral image into different vegetation species. When there are less number of vegetation species like in the 4000m x 4000m region, the autoencoder is able to identify and estimate the different tree species accurately as seen in Figure 4.19 and Tables 4.17 and 4.18. But in the 5000m x 5000m region which has a higher number of vegetation species, the different vegetation classes are identified correctly but their abundance values are not very accurate as seen in Figure 4.22 and Tables 4.24 and 4.23. Thus, the autoencoder is able to unmix the input data into different vegetation species but the abundance estimation accuracy is limited by the number of species. Another reason for the low accuracy of abundance estimation could be the low accuracy of the FIA ground truth data that has a confidence level of only 68%. Using more accurate data sources for validating the models will result in better evaluations.

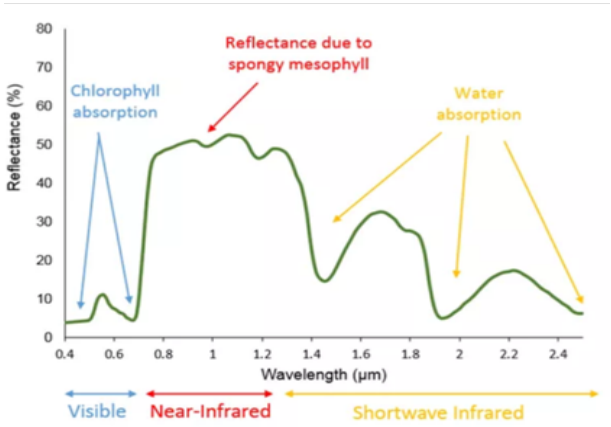

Figure 5.1: Vegetation Spectra [28]

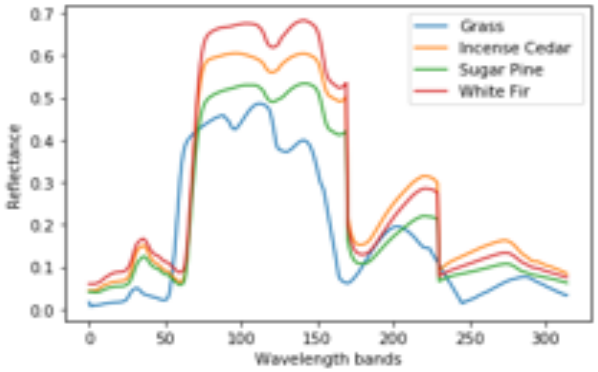

Figure 5.2: Different vegetation spectra in the 5000m x 5000m region

5.3.1 Performance comparison with other techniques

The results of the autoencoder are compared to the results of supervised classifiers as described in Section 4.3.3. The SVM, RF and the neural network are trained on a 100m x 100m forest region consisting of forest and vegetation land covers using the manually annotated ground truth data which consists of labels for every pixel in the image. The testing is performed on another unseen 100m x 100m region which consists of the same end members. We

obtain a F1 score of 0.91 for autoencoder, 0.981 for SVM, 0.99 for RF and 0.986 for neural network. Although the supervised techniques have higher F1 scores, these techniques require a large amount of training data and they require every pixel in the hyperspectral image to be labelled. Similar results are obtained even for a larger 1000m x1000m region. The autoencoder provides comparable abundance estimation results without requiring any annotated data. This is a huge advantage of the autoencoder technique as obtaining the ground truth labels is a time consuming process. Therefore, the autoencoder based approach is suitable for applications like town planning and environmental monitoring that require an estimate of the different land covers present in a region without the need to classify each and every pixel. Further, the autoencoder is able to perform a finer scale classification to identify the individual tree species in the experiments on 4000m x 4000m and 5000m x 5000m region using only the hyperspectral reflectance data whereas the supervised techniques are only able to identify the land cover as vegetation and not the individual tree species. The supervised methods would require training data with labels for each tree type to be able to obtain a fine scale vegetation classification.

The unsupervised autoencoder is compared with other unsupervised techniques, namely, K-Means, VCA and N-Findr in Section 4.3.4. For a 100m x 100m region, the autoencoder estimates the soil and vegetation land covers as 34.28% and 65.72% where as K-Means provides the estimations as 52.98% and 47.02% respectively, VCA provides the estimations as 37.94% and 62.06% respectively and N-Findr provides the estimations as 58.75% and 41.25% respectively with the ground truths of soil and vegetation being 31.67% and 68.32% respectively. This shows that the autoencoder performs better on data with mixed pixels since the K-Means technique tries to directly cluster the data into different classes where as the autoencoder performs unmixing of the spectral data through the latent representation

of the hidden layer. The autoencoder also outperforms traditional unmixing techniques. Similar results are obtained on the 100m x100 region where the autoencoder performs better than K-Means, VCA and N-Findr.

5.4 Autoencoder architecture

This section presents some discussions related to the effect of changing the autoencoder's architecture and hyperparameters on its performance and adapting the architecture to various datasets.

An asymmetric autoencoder with a deep encoder network and a single decoder layer is found to give the best spectral unmixing results. A very deep, complex encoder does not give any significant advantage. An autoencoder with 3-4 hidden layers provides the best results for all the datasets.

The number of hidden layers and hidden units increases with the dataset size and the number of end members. The examples provided below show how the number of layers and the hidden units vary with the number of end members.

1. 100m x 100m real hyperspectral data
    - 2 end members: 2 hidden layers with 2, 32 units
    - 3 end members: 3 hidden layers with 2, 6, 12 units
2. Synthetic data
    - 2 end members: 4 hidden layers with 2, 6, 12, 18 units
    - 3 end members: 4 hidden layers with 3, 9, 18, 36 units
    - 4 end members: 4 hidden layers with 4, 8, 16 , 28 units

- 5 end members: 4 hidden layers with 5, 15, 30, 45 units

It was experimentally determined that Leaky Rectified Linear Unit (LReLU) activation function for the hidden layers gives better results than Rectified Linear Unit (ReLU) and Sigmoid activations. This is because the LReLU has nonzero gradient for all inputs.

Training the model with Adam optimizer yields better results than with SGD and Adadelta optimizers.

The loss function was found to vary depending on the data and the appropriate loss function for each dataset was determined experimentally. Real hyperspectral dataset performs well with mean square error loss function and the synthetic hyperspectral dataset and Samson hyperspectral dataset perform well with cosine similarity loss function.

The batch size was found to be an important hyperparameter and varying the batch size significantly affected the model results. The batch size is effectively the sample size used to estimate the gradient for backpropagation, hence a small batch size results in a noisy gradient which could help the network by avoiding ending in saddle points or local minima [23]. A small batch size of 16 performed well in most cases.

The autoencoder based unmixing performed better with less number of end members in the input hyperspectral image. Different sets of analyses were performed by varying the number of end members from 2 to 5 on both synthetic and real data. It was found that with less number of end members, both the shape and magnitude of the unmixed result spectra matched well with the ground truth spectral library, where as, with higher number of end members only the shape matched closely with the ground truth while there were slight deviations in magnitudes from the ground truth.

Experiments showed that adding a Batch Normalization layer after the last hidden layer and before the output layer gives good unmixing results. Batch Normalization whitens the data which is well known for speeding up learning in neural networks by reducing the internal

covariance shift, which is a term used for the change of a layer's activations distribution over time [23].

Thus, automated models that identify the major land covers present in a given hyperspectral image scene can be developed using unsupervised autoencoders as indicated by our results.

Chapter 6

Conclusions

In this work, we have successfully developed a deep learning model using an unsupervised autoencoder for land cover identification and quantification from hyperspectral images. The unmixed output spectra and the estimated abundances correspond well with the ground truth which is evaluated quantitatively using metrics such as cosine similarity, mean square error and percentage abundances and qualitatively using abundance maps and spectral visualizations. This achieves research objective 1 of this work that is described in Section 1.3.

We have comprehensively analyzed the performance of this approach on distinct regions with different ecology consisting of a variety of land covers for both synthetic and real hyperspectral image data. We have assessed the scalability of this technique by showing that the method can be applied to small and large landscapes spanning from a few hundred square meters up to 10 square miles. The proposed approach is evaluated on synthetic and real hyperspectral reflectance data consisting of different land covers such as tree, soil, water, road, roof and grass and with different number of land covers in an image scene varying from 2 to 5. The real reflectance data is obtained from different geographical regions such as the Smithsonian Environmental Research Center in eastern United States, Teakettle Experimental Forest in western United States and ABBY field site in north western United States consisting of diverse land covers and vegetation species from different ecosystems. The results of the above analyses indicate that research objective 2 is achieved.

The proposed technique is evaluated by comparing its performance with supervised learning techniques such as SVM, RF and neural networks using precision, recall and F1 score metrics. Although the supervised techniques perform slightly better than the autoencoder, they require large amount of labelled data for training the models. Obtaining labels for the land covers requires high resolution images and expertise about the geography of the region. Generating labels for identifying the tree species for fine scale vegetation classification requires resource-intensive ground surveys. The advantage of using the autoencoder is that it provides performance comparable with that of supervised learning methods using only the hyperspectral images without requiring any training labels. Further, this technique performs better than other unsupervised techniques such as K-Means and traditional unmixing methods such as VCA and N-Findr by providing more accurate land cover abundance estimations. The autoencoder identifies the type of the land cover and the vegetation species thereby obtaining a finer scale land cover classification than the existing USGS NLCD standard which identifies only the major ecosystem types in a given region. The above analyses achieve research objective 3 of this work.

The autoencoder based approach provides accurate quantification of different land cover classes when their spectral reflectance patterns are different from each other. Fine scale vegetation classification becomes challenging since the spectra of all the vegetation types are similar differing only in their peak magnitudes. The autoencoder is able to unmix the input data into different vegetation species but the accuracy of abundance estimation is limited by the number of end members. The abundance estimation with higher number of vegetation species is not very accurate as there could be some misclassifications due to the similarity in the spectral signatures of different species.

The proposed autoencoder based land cover quantification is a simple and easy process where the hyperspectral image of the selected land area can be directly fed into an autoencoder

and the model can detect the major land covers present in the image scene. This technique is useful when there is a need to identify the major land cover surfaces and their approximate spread in applications like town planning to identify if the region in consideration is an urban area or forest land and in forestry applications to identify the main tree species in the given forest. Wildland fire spread applications require the knowledge of the physical materials in a given scene to predict accurate burn maps. The study of the effect of natural disasters and forest fires to land surfaces and living habitats require knowledge of the physical land covers in a region. The traditional methods used in these applications depend on ground surveys and low resolution images which are time consuming and require several days to generate land cover maps. The autoencoder technique can be used in such applications to generate fast, real-time maps. Therefore, this method can be used as an important tool for designing detailed land cover and vegetation maps for the development of fire spread models, land management, urban planning and vegetation treatment modules.

Our method is completely unsupervised which does not require field-based ground truth data which saves significant time and effort. This work uses open source spectral libraries, hyperspectral reflectance images and ground truth vegetation data without performing any independent field studies. This research demonstrates the performance of autoencoder based spectral unmixing for land cover quantification on a novel, less explored hyperspectral dataset named NEON dataset. This work performs assessment on large landscapes spanning across several square miles which has not been explored in prior works.

Thus, an automated model for identifying and quantifying the land covers from hyperspectral images has been developed using an unsupervised autoencoder.

Chapter 7

Future Work

The autoencoder based land cover quantification approach presented in this thesis can be extended by automating the end member selection process. Currently, we set the number of end members in an image scene during the training process based on the ground truth. Instead, the number of end members can be set as a hyperparameter which the model automatically determines based on the reconstruction error and the accuracy of output. This would make the technique completely automatic.

The autoencoder architecture can be enhanced by using techniques like layerwise pre-training, initialization of weights with VCA algorithm and cascading of autoencoders. The loss functions such as spectral information divergence and spectral angle distance that are popularly used in remote sensing applications can be used for training the autoencoder for better unmixing results.

Another future enhancement could be the development of several region specific spectral libraries. Separate libraries based on the geographic regions which includes the major land covers and tree types present in that particular region can be compiled instead of having one large, generic spectral library. This helps in obtaining classifications at a finer scale resulting in more accurate abundance estimations.

The models developed in this work are evaluated using ground truth data formulated by software based image segmentation and manual land covers annotations. The evaluation procedure can be improved by using the field survey data as the ground truth.

In this work, the pure spectra of each tree species is obtained by taking an average of several components from the EcoSIS, Ecostress and USGS spectral libraries. An improvement can be made to the spectral library by capturing the seasonal vegetation changes by including separate spectral components for healthy, dry, young and old vegetation for each species. This helps to capture the fine differences in the peak magnitudes of the different species thereby improving vegetation classification.

Bibliography

[1] ArcGIS. 2020. ArcGIS Pro. https://www.esri.com/en-us/arcgis/products/arcgis-pro/overview (Accessed on 08/16/2020).

[2] Alice M Baldridge, SJ Hook, CI Grove, and G Rivera. 2009. The ASTER spectral library version 2.0. *Remote Sensing of Environment* 113, 4 (2009), 711–715.

[3] Kevin Baragona. 2010. Cosine Similarity. https://deepai.org/machine-learning-glossary-and-terms/cosine-similarity (Accessed on 08/16/2020).

[4] Moshe Binieli. 2018. Machine learning: an introduction to mean squared error and regression lines. https://www.freecodecamp.org/news/machine-learning-mean-squared-error-regression-line-c7dde9a26b93/ (Accessed on 08/16/2020).

[5] L. Christovam, Guilherme Pessoa, Milton Shimabukuro, and Maria De Lourdes Galo. 2019. Land Use And Land Cover Classification Using Hyperspectral Imagery: Evaluating The Performance Of Spectral Angle Mapper, Support Vector Machine And Random Forest. (06 2019). https://doi.org/10.5194/isprs-archives-XLII-2-W13-1841-2019

[6] Steven Flores. 2019. Variational Autoencoders are Beautiful. (2019). https://www.compthree.com/blog/autoencoder (Accessed on 07/17/2020).

[7] Geoffrey A Fricker, Jonathan D Ventura, Jeffrey A Wolf, Malcolm P North, Frank W Davis, and Janet Franklin. 2019. A convolutional neural network classifier identifies

tree species in mixed-conifer forest from hyperspectral imagery. *Remote Sensing* 11, 19 (2019), 2326.

[8] Hongmin Gao, Shuo Lin, Yao Yang, Chenming Li, and Mingxiang Yang. 2018. Convolution Neural Network Based on Two-Dimensional Spectrum for Hyperspectral Image Classification. *Journal of Sensors* 2018, 2 (2018). https://doi.org/10.1155/2018/8602103

[9] Utsav B. Gewali, Sildomar T. Monteiro, and Eli Saber. 2018. Machine learning based hyperspectral image analysis: A survey. (2018). arXiv:1802.08701 http://arxiv.org/abs/1802.08701

[10] Antonio Di Gregorio and Louisa J.M. Jansen. [n.d.]. Land Cover Classification System (LCCS). ([n. d.]). http://www.fao.org/3/x0596e/x0596e01e.htm (Accessed on 07/17/2020).

[11] Rui Guo, Wei Wang, and Hairong Qi. 2015. Hyperspectral image unmixing using autoencoder cascade. In *2015 7th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*. IEEE, 1–4.

[12] Gaurav Hegde, J Mohammed Ahamed, R Hebbar, and Uday Raj. 2014. Urban land cover classification using hyperspectral data. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* 8, 8 (2014), 751–754.

[13] Forest Inventory and Analysis National Program. 2020. Tools and Data. https://www.fia.fs.fed.us/tools-data/index.php. (Accessed on 07/17/2020).

[14] Sergio Marconi, Sarah J Graves, Dihong Gong, Morteza Shahriari Nia, Marion Le Bras, Bonnie J Dorr, Peter Fontana, Justin Gearhart, Craig Greenberg, Dave J Harris, et al.

2019. A data science challenge for converting airborne remote sensing data into ecological information. *PeerJ* 6 (2019), e5843.

[15] Susan K Meerdink, Simon J Hook, Dar A Roberts, and Elsa A Abbott. 2019. The ECOSTRESS spectral library version 1.0. *Remote Sensing of Environment* 230 (2019), 111196.

[16] TA Moughal. 2013. Hyperspectral image classification using support vector machine. In *Journal of Physics: Conference Series*, Vol. 439. IOP Publishing, 012042.

[17] J. M. P. Nascimento and J. M. B. Dias. 2005. Vertex component analysis: a fast algorithm to unmix hyperspectral data. *IEEE Transactions on Geoscience and Remote Sensing* (2005).

[18] National Ecological Observatory Network. 2017. Data Products DP3.30006.001 SERC. http://data.neonscience.org (Accessed on 08/18/2020).

[19] National Ecological Observatory Network. 2017. Data Products DP3.30006.001 TEAK. http://data.neonscience.org (Accessed on 08/18/2020).

[20] National Ecological Observatory Network. 2018. Data Products DP3.30006.001 ABBY. http://data.neonscience.org (Accessed on 08/18/2020).

[21] The National Ecological Observatory Network. 2018. Airborne Remote Sensing. https://www.neonscience.org/data-collection/airborne-remote-sensing (Accessed on 07/17/2020).

[22] Savas Ozkan, Berk Kaya, and Gozde Bozdagi Akar. 2018. Endnet: Sparse autoencoder network for endmember extraction and hyperspectral unmixing. *IEEE Transactions on Geoscience and Remote Sensing* 57, 1 (2018), 482–496.

[23] Burkni Palsson, Jakob Sigurdsson, Johannes R Sveinsson, and Magnus O Ulfarsson. 2018. Hyperspectral unmixing using a neural network autoencoder. *IEEE Access* 6 (2018), 25646–25656.

[24] Ying Qu and Hairong Qi. 2018. uDAS: An untied denoising autoencoder with sparsity for spectral unmixing. *IEEE Transactions on Geoscience and Remote Sensing* 57, 3 (2018), 1698–1712.

[25] Christopher Riggio. 2019. What's the deal with Accuracy, Precision, Recall and F1? https://towardsdatascience.com/whats-the-deal-with-accuracy-precision-recall-and-f1-f5d8b4db1021 (Accessed on 08/17/2020).

[26] Saeed Sojasi, Bardia Yousefi, Kévin Liaigre, Clemente Ibarra-Castanedo, Georges Beaudoin, Xavier P. V. Maldague, François Huot, and Martin Chamberland. 2017. The role of the continuous wavelet transform in mineral identification using hyperspectral imaging in the long-wave infrared by using SVM classifier, Vol. 10214. International Society for Optics and Photonics, SPIE. https://doi.org/10.1117/12.2264580

[27] Spectral Imgaing Ltd. SPECIM. 2020. What is Hyperspectral Imaging? https://www.specim.fi/library/what-is-hyperspectral-imaging (Accessed on 08/12/2020).

[28] Koala Spotting. 2018. Spectral Properties and Unique Spectral Signatures. https://remotesensinginactionreflectivelearningblog.wordpress.com/2018/01/18/spectral-properties-and-unique-spectral-signatures/ (Accessed on 08/12/2020).

[29] Yuanchao Su, Jun Li, Antonio Plaza, Andrea Marinoni, Paolo Gamba, and Somdatta Chakravortty. 2019. DAEN: Deep autoencoder networks for hyperspectral unmixing. *IEEE Transactions on Geoscience and Remote Sensing* 57, 7 (2019), 4309–4321.

[30] G Rex Sumsion, Michael S Bradshaw, Kimball T Hill, Lucas DG Pinto, and Stephen R Piccolo. 2019. Remote sensing tree classification with a multilayer perceptron. *PeerJ* 7 (2019), e6101.

[31] Le Sun. [n.d.]. Datasets for Classification. http://lesun.weebly.com/hyperspectral-data-set.html (Accessed on 07/17/2020).

[32] U.S. Geological Survey. 2020. National Land Cover Database 2016 (NLCD2016) Legend. hhttps://www.mrlc.gov/data/legends/national-land-cover-database-2016-nlcd2016-legend (Accessed on 08/13/2020).

[33] USGS. 2018. Spectroscopy Lab. https://www.usgs.gov/labs/spec-lab/capabilities/spectral-library (Accessed on 07/17/2020).

[34] Erin Patricia Wagner, Justin Merz, and Philip A Townsend. 2018. Ecological Spectral Information System: An Open Spectral Library. *AGUFM* 2018 (2018), B41L–2878.

[35] Lloyd Windrim, Rishi Ramakrishnan, Arman Melkumyan, Richard J Murphy, and Anna Chlingaryan. 2019. Unsupervised feature-learning for hyperspectral data with autoencoders. *Remote Sensing* 11, 7 (2019), 864.

[36] Michael E Winter. 1999. N-FINDR: An algorithm for fast autonomous spectral endmember determination in hyperspectral data.

[37] Lefei Zhang, Liangpei Zhang, Dacheng Tao, Xin Huang, and Bo Du. 2013. Hyperspectral remote sensing image subpixel target detection based on supervised metric learning. *IEEE transactions on geoscience and remote sensing* 52, 8 (2013), 4955–4965.

[38] Amanda K Ziemann. 2015. A manifold learning approach to target detection in high-resolution hyperspectral imagery. (2015). Thesis. Rochester Institute of Technology. Accessed from https://scholarworks.rit.edu/theses/8617.

[39] Sheng Zou, Paul Gader, and Alina Zare. 2019. Hyperspectral tree crown classification using the multiple instance adaptive cosine estimator. *PeerJ* 7 (2019), e6405.